USING RESIDUAL ANALYSIS TO VALIDATE WATERMELON (*Citrullus lanatus*) DATE OF PLANTING EXPERIMENT MODELS

OKEKE, G. C.[*], IBRAHIM A., IBEGBULEM J. A., AYUBA, I. A., YUSUF A. A., HARUNA A. AND ORJI, P. O.

National Agricultural Extension and Research Liaison Services, Ahmadu Bello University, Zaria
* gilbertokeke@gmail.com; +2348081302387

ABSTRACT

This research was carried out at the Teaching and Research Farm of the Department of Crop Production and Landscape Management, Ebonyi State University, Abakaliki. The main purpose of this work is to apply the residual analysis to check the suitability of the series of similar experimental model to describe the effects of dates of planting on yield of watermelon with the view of predicting optimum planting date for the cultivation of watermelon. Four experiments were laid out as series of similar experiments representing the four planting dates (April 4, April 18, May 2 and May 16, 2014). Each experiment consists of three varieties of watermelon (Sugarbaby, Kaolack and F1 Koloss) and three levels of poultry manure (0, 5 and 10kg/ha). The data were separately analyzed for yield for each planting date, where variances were homogenous (i.e. where the experiments showed no significant Bartlett's test) and this permits the combined analysis of variance for each planting date. The normal probability plots using the residuals showed that it was reasonable to assume that the random errors for the residuals were drawn from approximately normal distributions. Also, the histogram was bell shaped showing that the functional part of the model was correctly specified. The coefficient of determination (R^2) was 65.5%. Generally, the results showed that series of similar experiments methodology was able to model the changes associated with different dates of planting. The questions associated with model adequacy were discussed. We recommend F1 Koloss watermelon variety and 5kg/ha poultry manure with early planting which should begin immediately after the first rain provided it is sufficient for field preparations, especially ploughing and harrowing.

Keywords: Planting date, watermelon, statistical model, residual analysis, scatter plot, normal probability plot, histogram.

INTRODUCTION

Studies on date of planting have been conducted extensively for several crops across different agro-ecological zones. A review of studies on watermelon date of planting indicates that the research was necessitated by environmental pattern (Chandra and Mama, 1988); pest management (Chand and Singh, 1985), disease management and fruit yield (Draper and Smith, 1998).

Several statistical procedures and steps have been adopted by researchers for date of planting studies to suit their objectives. Obi (2013a) proposed series of similar experiment models for date of planting experiments. Some researchers (Cirilo and Andrade, 1994; Sanders *et al*., 1999), used this series of similar experiment model in their studies. A survey of other works on planting date showed that randomized complete block design with four complete blocks has been used to study the effect of date of planting and residue rate effect on growth partitioning and yield of egusi-melon (Urkurkar and Chandrawashi, 1983). Furthermore, some other studies used randomized complete block design in their experimental trials on maize (Montgomery, 1991; Adeoti and Emechebe, 1996), while other researchers in their trial on response of soybean lines with juvenile traits to day length and date of planting combined both greenhouse and field work (Shapiro and Francia, 1972; Soroja and Raju, 1983). In both cases, they used factorial arrangement. Acikgoz (1987) in his studies on effect of date of planting and plant method on rice yield per day also used factorial experiment. Fakorede *et al*., (1985), in maize planting date trial, used a randomized complete block design (RCBD) with a split split-plot arrangement. He assigned dates to main plot, planting date to sub plot and genotypes to the sub-sub plot. Chand and Singh (1985) and Chaudhry (1984) used split-plot design in their study of effect of planting date on Stem Rot (SR) incidence in rice. Both Chand and Singh (1985) and Chaudhry (1984) assigned planting date to sub plot and variety to sub-sub plot. Obi *et al*., (2009), in rice planting date used residual analysis to validate rice sowing date experiment model.

Picking a model for a problem is a serious task. If the model fits well, then it can be used to increase the understanding and learning of the problem and/or for prediction. Several procedures and steps have been adopted by researchers for planting date studies to suit their objectives. The ever- proliferation of statistical steps available to researchers has given room for use of different statistical design for research into finding optimum planting date for crops. Considering the subtle differences, advantages and disadvantage that these statistical designs brings, the results of such analysis may lead to false conclusion or be less reliable at least for comparative purposes. There is need to look into planting date trials again to see the possibility of proffering a statistical model that could be commonly used by researchers. Walter T. Federer, (2001) proposed Federer Design giving his premise on date of planting during his personal communication with Obi, I.U. which is stated as follow: “Variety x Location, Variety x Year, and Variety x Location x Year interactions are mostly date of planting x Variety interaction”. Federer said, “if I were setting up an experiment on planting date, I would use a number of locations (Different), and Years and in each Year and Location, I would have a replicated experiment where the replicates were planting, say three (3) planting date with two (2) replicates, each, Whole-plots would be crops and the Split-plots would be Varieties within Crops”. Obi thinks that what these researchers are doing is “fitting data to a model” not “fitting model to data.” He did this by illustrating the field layout of those designs derived by those researchers.

Verifying the adequacy of the linear model

The outlier tests (Draper and Smith, 1998 and Herbek, Murdock and Blevins, 1986) as well as the variance homogeneity tests (Bartlett, 1983, Maluf *et at.*, 2005 and Cirilo and Andrade, 1994) have been used. Kirton, personal communication with Federer

(1977), department of Agriculture, New South Wales, Sydney, Australia, as cited by (Obi, 2011), suggested that after one has a "correct model," one should as follows in searching for a discrepant "treatment" or "block": (i) compute estimate residuals, (ii) use absolute values of the residuals , and (iii) perform a standard analysis of variance (ANOVA), the same one as used for the response, and/or multiply comparisons on the absolute values of the estimated residuals. If the model is "correct", then the null hypothesis should be true for all the categories in the ANOVA except for residuals of the absolute values of residuals. That is, the expected value for each F-test is "one". If the null hypothesis (hypotheses) is/are not true, then this procedure can be used to pinpoint discrepant treatments, blocks, etc., in the experiment (Federer *et al.*, 1983).

Federer *et al*, (1983) did "Studies of Residual Analysis" titled; "Analysis of Absolute Values of Residuals to test Distributional Assumptions of Linear Models for Balanced Designs". Although statistical models are not largely used in planting date data analysis, a verification of the suitability of the model is not always checked. Hence, indicating that the search for the *Model* for planting date experiment is continuing and the *Model* and *Field Layout* are yet to be confirmed.

The main purpose of this work is to apply the residual analysis to check the suitability of the conventional series of similar experiment proposed by (Obi, 2013 a and b) to describe the effect of planting date on watermelon cultivars. Additionally, the authors expect that this paper is capable of introducing a step-by-step procedure for the implementation of the residual analysis to any statistically significant model. Also, answers would be proffered to questions concerning model validity such as:-

(i) How can I tell if a model fits my data?

(ii) How can I assess the sufficiency of the functional part of the model?

(iii)How can I test whether or not the random errors are distributed normally?

Statistical Model

A statistical model is a mathematical model which contains error with a specific probability distribution. Usually, this model is used to predict the value of one of the variables when the other is known, under specific conditions (Marzouk and El-Bawab, 1999). In a statistical model, two or more variables are related using regression analysis equations. These equations are mainly used to predict the dependent variable, Y, as a function of the independent variable, X. In the analysis, some assumptions are necessary (Chatterjee and Price, 1977; Fakorede *et al.*, 1985).

The dependent variable, X, is considered free of errors because X is not a random variable. There is a linear relationship between Y and X and the statistical model that relates Y_i to X_i as given by:

$Y_i = A + B X_i + \mathcal{E}_i$ ……………………(Eq.1)

For i = 1, . . ., n, where n is the number of observations.

In equation 1, A and B are unknown constants to be estimated and they are called parameters of the regression model. The random value, $\mathcal{E}_i$ is the denominated random error. The value of $\mathcal{E}_i$ for any observation will depend on both a possible error of measurement and other variables different from X_i that were not measured that could affect Y_i. The values of $\mathcal{E}_i$ for any observation will depend on both a possible error of measurement and other variables different from X_i that were not measured that could affect Y_i. The values of $\mathcal{E}_i$ are random variables, assuming the following assumptions:

(i) The average of $\mathcal{E}_i$ values is equal to zero and its variance, is unknown and constant for $1 \leq I \leq n$;

(ii) $\mathcal{E}_i$ values are not correlated;

(iii)The distribution of $\mathcal{E}_i$ values is normal for $1 \leq I \leq n$.

Second and third assumptions imply that $\mathcal{E}_i$ values are mutually independent. The regression line is, in general, unknown and therefore must be estimated through the sampling data. In the particular case where the regression of Y in relation to X is linear, the best fit line can be written as:

$\hat{Y}_1 = \hat{A} + \dot{B}X_1$ ....(Eq.2)

Where the symbol "caret" (^) denotes estimate (estimator) $\hat{A}$ and $\hat{B}$ are determined by the least squares method and $\hat{Y}_1$ are the estimated values of Y_1 using Equation 2 such that the differences between Y_1 and $\hat{Y}_1$ shall be minimum. Generally, these differences are known as residuals, i.e., errors associated to the predicted values of Y_1 corresponding to each X_1 value and which can be calculated through the following expression:

$\hat{e}_1 = Y_1 - \hat{Y}_1$ ………………………….. (Eq.3)

A discussion of a simple linear model considering two variables X and Y would enhance our understanding of procedures used to examine the adequacy of a model.

$Y = a + bX + e$

The variable 'e' denotes random error, that is, if there were no error, Y would be a deterministic linear function of X.

When is a model good? At first, one might say when there is no error. However, for all the data that we consider in this class, there will always be error. Actually we will say a model is good if there is no connection between e and a + bX: that is, the random error is free of X. hence, for predicting Y, we have found the model that contains all the information based on X. Now there may be other variables which help in predicting Y. These will be contained in e.

Model Assumption: The assumption we want to verify on a model is: the random error component is independent of the X component. How would we check this assumption? If we knew the random errors, e, we would just plot them against a + bX. A random scatter plot would indicate that the errors do not depend on a a + bX; i.e., the errors are free of a + bX. Thus the model is good. However, we don't know the errors, we only know Y and X. But using Y and X we estimate a and b. This leads to an estimate of a + bX, the predicted value of Y, which we label as $\hat{Y}$. Our estimate of the error is $Y - \hat{Y}$. This is called the residual, literally, what's left. We will denote the residual by ê, that is, $\hat{e} = Y - (\hat{a} + \hat{b}X)$. Then we can check our model assumption by plotting ê versus $\hat{Y}$. This is called the residual plot. A random scatter indicates a good model. If it not a random scatter then we need to rethink the model (Obi, 2013b). The verification of residuals normality can also be analyzed by plots, such as normal score and normal probability graphs. In these graphs, the assumption of normality is valid if the points in the graph are localized approximately along a straight line. However, in case of doubt, the linearity can be confirmed using a statistical test of normality, such as the one proposed by (Marzouk and El-Bawab, 1999).

Experimental Procedures

The research was carried out at the Teaching and Research Farm of the Department of Crop Production and Landscape Management, Ebonyi State University, Abakaliki. The study region falls within the tropical rain forest zone of South-Eastern Nigeria and it is located between latitude 06^{o} 4` N and longitude 08$^{o}$65` E at the elevation of 71.44 mm above sea level. It has a bimodal rainfall pattern. It's rainfall per annum ranges between 1700-2000 mm in April to July and August to November for early and late season plantings, respectively. The relative humidity at dry season ranges between 60-80% and the soil belongs to the order ultisols (FDALR, 1985).

Four experiments were laid out as series of similar experiments representing the four planting dates (April 4, 18, May 2 and 16, 2014) using 1.0 m x 0.6 m plant spacing. Each experiment consists of three varieties of watermelon (Sugarbaby, Kaolack and F1 Koloss) and three levels or rates of poultry manure (0, 5 and 10 kg/ha). The experiments were laid out in a randomized complete block design (RCBD) with three blocks corresponding to a 3 x 3 factorial experiment for the four planting date. The total land area for the four planting date is 0.553ha. The designated plant distances were measured and marked out in their respective plots and seeds were hand-sown at the rate of two (2) seeds per hole on the bed at the depth of 2cm. Distance between seed beds was 1m for each of the planting date while distance between blocks was 1m. The form of analysis for each planting date showing sources of variation and degrees of freedom (General and Specific) are shown in Table 2. Figure 1 showed the field layout consisting of three varieties (Kaolack, Koloss F1 and Sugarbaby) of watermelon and four planting date. Seed beds were well prepared by ploughing and harrowing and plots were marked out. Weeds were controlled through manual hoeing and subsequently by hand pulling as the watermelon vines spread and cover the plots to thus suppress weed growth. The first experiment commenced on April 4, 2014 and the subsequent experiments followed after every two weeks.

1. DATE OF PLANTING D_1

Block I	V1P0	V2P2	V3P3	V2P0	V1P2	V3P2	V2P3	V3P0	V1P3
Block II	V2P2	V3P0	V1P2	V3P3	V2P3	V1P3	V3P2	V1P0	V2P0
Block III	V3P3	V1P2	V2P2	V1P0	V3P2	V2P0	V1P3	V2P3	V3P0

2. DATE OF PLANTING D_2

Block I	V1P0	V2P2	V3P3	V2P0	V1P2	V3P2	V2P3	V3P0	V1P3
Block II	V2P2	V3P0	V1P2	V3P3	V2P3	V1P3	V3P2	V1P0	V2P0
Block III	V3P3	V1P2	V2P2	V1P0	V3P2	V2P0	V1P3	V2P3	V3P0

3. DATE OF PLANTING D_3

Block I	V1P0	V2P2	V3P3	V2P0	V1P2	V3P2	V2P3	V3P0	V1P3
Block II	V2P2	V3P0	V1P2	V3P3	V2P3	V1P3	V3P2	V1P0	V2P0
Block III	V3P3	V1P2	V2P2	V1P0	V3P2	V2P0	V1P3	V2P3	V3P0

4. DATE OF PLANTING D4

Block I	V1P0	V2P2	V3P3	V2P0	V1P2	V3P2	V2P3	V3P0	V1P3
Block II	V2P2	V3P0	V1P2	V3P3	V2P3	V1P3	V3P2	V1P0	V2P0
Block III	V3P3	V1P2	V2P2	V1P0	V3P2	V2P0	V1P3	V2P3	V3P0

FIGURE 1: THE FIELD LAYOUT FOR DATE (TIME) OF PLANTING EXPERIMENT AS SERIES OF SIMILAR EXPERIMENTS (D_1 = FIRST DATE OF PLANT AND D_4 = DATE OF PLANTING 8 WEEKS AFTER)

TABLE 2: FORM OF ANALYSIS OF VARIANCE FOR A RANDOMIZED COMPLETE BLOCK DESIGN (RCBD) FOR EACH DATE OF PLANTING, SHOWING SOURCES OF VARIATIONS, DEGREES OF FREEDOM (GENERAL) AND (SPECIFIC,) ONLY.

Sources of Variation	D.F (General)	D.F (Specific)
Blocks	r-1	2
Variety (V)	v-1	2
Poultry manure (M)	m-1	2
V x M	(v-1)(m-1)	4
Error	(r-1)(vm-)	16
Total	Vrm-1	26

Data on yield was collected at maturity. Fruit yield per plot was converted to tonnes per hectare. The Analysis of Variance was performed using the procedure outlined by (Steel and Torrie, 1996) for each measured parameter. Means that had a significant F-test were separated using $LSD_{0.05}$ (Obi, 2011 and Obi *et al*., 2009). The data analysis was carried out in stages.

Test for homogeneity of variance

Bartlett's (1973) test for homogeneity of variance was conducted in order to determine whether or not the assumption of homogeneity was met (Obi, 2013a, Obi, 2013c, Obi, *et al.*, 2009 and Ogbonna, 1996).

Stage I: The data were separately analyzed for yield for each planting date, where variances were homogenous (i.e. where the experiments showed no significant Bartlett's test). This permits the combined analysis of variance for each planting date.

Stage II: A combined analysis of variance was done for yield measured for each planting date following the procedure of analysis of combined experiments as outlined by (Mclntosh, 1983). The form of combined analysis for each trait measured for planting date showing sources of variation and degrees of freedom (General and Specific) are shown in Table 3. The linear statistical model used for the analysis of variance is as

$$X_{ijk} = \mu + ß_i + Ʈ_j + \alpha_k + (Ʈ\alpha)_{jk} + €_{ijk}$$

Where;

X_{ijk}= any individual observation

μ = Overall mean

$ß_i$ = Blocking effect

$Ʈ_j$ = Effect of varieties

α_k = Effect of Poultry Manure

$(Ʈ\alpha)_{jk}$ = Interaction effect of Variety and Poultry Manure

$€_{ij}$ = experimental error

TABLE 3: FORM OF COMBINED ANALYSIS OF VARIANCE FOR FOUR DATES OF PLANTING, SHOWING SOURCES OF VARIATION, DEGREES OF FREEDOM (GENERAL) AND (SPECIFIC), ONLY.

Sources of Variation	D.F (General)	D.F (Specific)
Varieties (V)	v-1	2
Poultry Manure (M)	m-1	2
Date of planting (D)	d-1	3
D x V interaction	(d-1)(v-1)	6
D x M interaction	(d-1)(m-1)	6
M x V interaction	(m-1)(v-1)	4
V x D x M	(v-1)(d-1)(m-1)	12
Blocks within Date of planting	(r-1)(d)	8
Error	d(r-1)(v-1)	64
Total	rdvm-1	107

Stage III: A combined analysis of variance was done for yield measured over four planting date. Planting date was considered to have random effects while varieties were considered as fixed effects. The form of analysis of variance showing sources of variation and degrees of freedom is presented in table 3. The linear additive model for a Combined Analysis of Variance of plant spacings is stated as follows:

$$X_{ijkl} = \mu + \tau_i + \alpha_j + \beta_k + e_l + (\tau\alpha)_{ij} + (\tau\beta)_{ik} + (\alpha\beta)_{jk} + (\tau\alpha\beta)_{ijk} + \epsilon_{ijkl}$$

Where X_{ijk} = Observation made on the i^{th} variety within the l^{th} replication in the j^{th} Poultry Manure and k^{th} planting date

μ = The population or general mean

τ_i = Effect of the i^{th} Variety

α_j = Effect of the j^{th} Poultry Manure rates

β_k = Effect of the k^{th} date of planting

e_l = Effect of the l^t block/replication within planting date

$(\tau\alpha)_{ij}$ = Interaction effect of i^{th} Variety x j^{th} Poultry Manure

$(\tau\beta)_{ik}$ = Interaction effect of i^{th} Variety x k^{th} Planting Date

$(\alpha\beta)_{jk}$ = Interaction effect of j^{th} Poultry Manure and k^{th} Planting Date

$(\tau\alpha\beta)_{ijk}$ = Interaction effect of i^{th} Variety, j^{th} Poultry Manure and k^{th} Planting Date

ϵ_{ijk} = Experimental Error of i^{th} Variety Variation, within j^{th} Poultry Manure, k^{th} Planting Date and l^{th} Block within Date of Planting.

RESULTS

The result shows experiment model analyzed in stages. The example of the result of mean square of the stage I of the analysis of variance and degrees of freedom for yield component is presented in Table 4. The example of the results of variance for traits measured from the four plant spacing and the Bartlett's test for homogeneity of variance is shown in Table 6. The Bartlett's test was not significant. Based on the non-significant Bartlett's test for homogeneity of variance, the procedure of analysis of combined experiments as outlined by Mclntosh (1983) was used to combine the four plant spacing. This is the stage II of the analysis of the experiment model.

TABLE 4: SOURCES OF VARIATION, DEGREES OF FREEDOM AND MEAN SQUARES FROM ANALYSIS FOR YIELD OF THREE VARIETIES OF WATERMELON PLANTED ON EACH OF THE FOUR PLANTING DATE

Sources of Variation	Degrees of Freedom	**Mean Squares**			
		April 4, 2014	April 18, 2014	May 2, 2014	May 16, 2014
Block	2	8.111	10.81	58.93	2.48
Manure	2	36.111**	207.81**	330.48**	38.37**
Varieties	2	61.000**	75.26**	14.37	14.37
V x M	4	29.444**	28.15	5.43	2.26
Error	16	7.028	13.69	17.80	11.06
Total	26	-	-	-	-

*** = significant at 5% and 1% probability levels respectively.

TABLE 5: COMBINED ANALYSIS OF VARIANCE SHOWING SOURCES OF VARIATION, DEGREES OF FREEDOM AND MEAN SQUARES FROM ANALYSIS FOR YIELD PARAMETER OF THREE WATERMELON PLANTED ON FOUR PLANTING DATES

Sources of Variation	Degrees of Freedom	**Mean Squares**			
		50% Flowering	Fruit Diameter (cm)	Fruit Weight (kg)	Fruit Yield (kg/ha)
Variety (V)	2	51.39815	146.246**	0.94494*	7.25926**
Date of Planting (D)	3	45.12037	111.278**	4.8849**	111.0864**
Poultry Manure (M)	2	170.009**	215.785**	6.9885**	0.67592
M x D	6	88.0833**	103.546**	2.3385**	0.614198
M x V	4	12.8704	17.6476	0.2252	2.578704
D x V	6	15.0278	33.9864**	0.3759	5.864198
V x D x M	12	38.34656*	37.728**	0.7262**	2.811508*
Block/Date	8	17.16667	3.80132	0.46967	0.763889*
Error	64	25.25490	10.11789	0.29804	4.533496
Total	107	-	-	-	-

*,** = significant at 5% and 1% probability levels respectively.

TABLE 6: BARTLETT'S TEST FOR HOMOGENEITY OF VARIANCE FOR NUMBER OF FRUIT PER PLOT OF WATERMELON PLANTED UNDER FOUR PLANTING DATES

Planting Dates	d.f	Error Var.	S^2	$LogS^2$	(d.f x $LogS^2$)	1/d.f
First Planting	16	56.833	9.472	0.977	5.859	0.167
Second Planting	16	72.830	12.138	1.084	6.505	0.167
Third Planting	16	35.500	5.917	0.772	4.633	0.167
Fourth Planting	16	77.830	12.972	1.113	6.678	0.167
Totals	64	242.993			23.675	0.250
Pooling			10.125	1.005	24.129	

Chi Square calculated = 0.884NS

Critical value of X^2 at 5%, 16 d.f = 24.996; critical value of X^2 at 1%, 16 d.f = 30.578 , NS = Not significant

The R^2 for this planting date series of similar experiment models was calculated to be 65.5% and was observed to be very close the to 61.30% gotten by Obi *et al.,* 2009 in his model (using residual analysis to validate rice sowing dates experiment model). The graphical residual analysis of the study which is given by the graph of residuals plotted against predicted values is presented in figure 2 below. The plots did not revealed any particularly troublesome pattern other than a random pattern, although the largest positive residual value observed was slightly above 15 and stood out from the others. From the graph below, it could be observed that the residuals appear to behave randomly and that suggest that the model fits the data well.

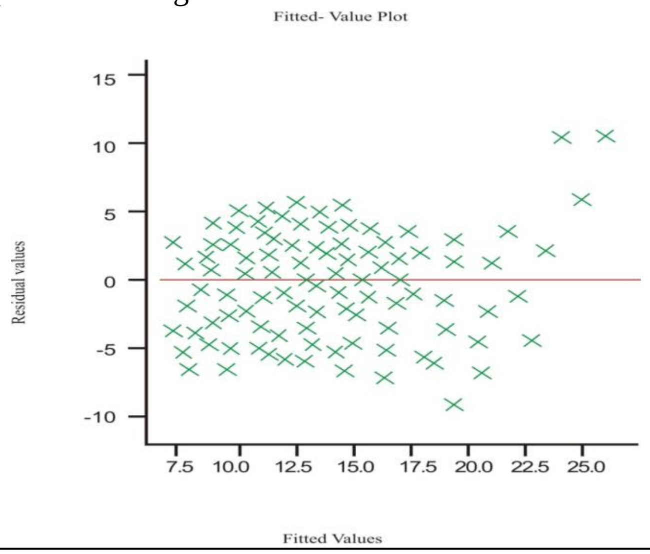

FIGURE 2: PLOT OF PREDICTED YIELD PER PLOT VERSUS

The question of how to know whether or not the random errors are distributed normally is answered by the normal probability plot. It is used to check whether or not it is reasonable to assume that the random errors inherent in the process have been drawn from a normal distribution. The normal probability is constructed by plotting the sorted values of the residuals against the associated theoretical values of standard normal distribution. The distinct curvature or other significant deviations from a straight line as shown below indicate that the random errors are probably not normally distributed. The normal probability plots for the experiment indicate that it is reasonable to assume that the random errors for these processes are drawn from approximately normal distributions. However, since none of the points in the plots deviate much from the linear relationship defined by residuals, it is also reasonable to conclude that there are no outliers in any of these data sets. The normal probability plot is presented in Fig. 3.

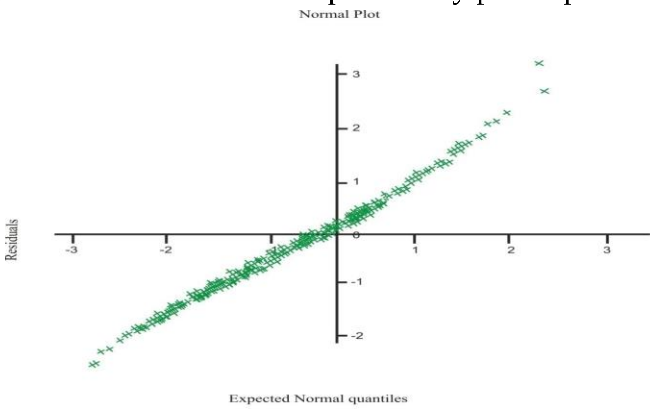

FIGURE 3: PLOT OF RESIDUAL AGAINST NORMAL SCORES

The question of how to know whether or not the random errors are distributed normally is answered by the histogram. The histogram which is more or less a bell shaped, provides a clearer picture of the shape of the distribution. The bell-shape of the histogram confirms the conclusions from the normal probability plots.

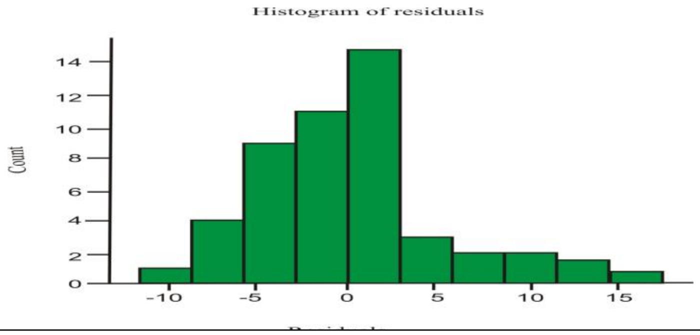

FIGURE 4: HISTOGRAM OF RESIDUAL FOR YIELD/PLOT

DISCUSSION

According to Montgomery (1991), before the conclusions from the analysis of variance of a design are adopted, the adequacy of the model should be checked. Often the validation of a model seems to consist of nothing more than quoting the R^2 (Coefficient of Determination) statistic from the fit (which measures the fraction of the total variability in the response that is accounted for by the model). Unfortunately, a high R^2 value does not guarantee that the model fits the data well. Use of a model that does not fit the data well cannot provide good answer to the underlying scientific question of how one can know if a model fits the data under investigation. Even though the R^2 for planting date model accounted for 65.5%, the primary statistical tool for most process modeling application is graphical residual analysis (Obi, 2002). In addition, the normal probability plot also serves as to confirm the adequacy of a model (Mclntosh, 1983 and Obi, 2002).

Different types of plots of the residuals from a fitted model provide information on the adequacy of different aspects of the model. Numerical methods for model validation, such as the R^2 statistic, are also useful but usually to a lesser degree than graphical methods. Graphical methods have an advantage over numerical methods for model validation because they readily illustrate a broad range of complex aspects of the relationship between the model and the data. Numerical methods for model validation tend to narrowly focus on a particular aspect of the relationship between the model and the data and often try to compress information into a single descriptive number or test result. If the model's fit to the data were correct, the residuals would approximate the random errors that make the relationship between the explanatory variables and the response variable a statistical relationship. Therefore, if the residuals appear to behave randomly, it suggests that the model fits the data well. On the other hand, if non-random structure is evident in the residuals, it is a clear sign that the model fits the data poorly. The plot did not reveal any particularly destructive pattern other than a random pattern, although the largest positive residual value observed slightly above 10 stands out from the others for planting date. It is not enough in the scattered plot to indicate unsuitability of the model for the study. According to (Obi, 2002), it is possible that a particular treatment combination produces slightly more erratic response than the others. The problem moreover is not severe enough to have a negative impact on the analysis and conclusion (Obi, 2002).

The assessment of the sufficiency of the functional part of a model also depends on the scatter plot of the residuals against the predictor variables in the model and against potential predictors that are not included in the model. These are the primary plots used to assess sufficiency of the functional part of the model. Plots in which the residuals do not exhibit any systematic structure indicate that the model fits the data well. Plots of the residuals against other predictor variables, or potential predictors, which exhibit systematic structure, indicate that the form of the function can be improved in some way. In this study, *Figure* 2 did not show a systematic structure.

The question of how to know or check whether or not the random errors are distributed normally is answered by the histogram and the normal probability plot. These are used to check whether or not it is reasonable to assume that the random errors inherent in the process have been drawn from a normal distribution. The normality assumption is needed for the error rates we are willing to accept when making decisions about the process.

The normal probability plot is constructed by plotting the sorted values of the residuals against the associated theoretical values from the standard normal distribution. Unlike most residual scatter plots, however, a random scatter of points does not indicate that the assumption being checked is met in this case. Distinct curvature or other significant deviations from a straight line indicate that the random errors are probably not normally distributed. A few points that are far off the line suggest that the data has outliers in it.

The normal probability plot in *Figure 2* indicates that it is reasonable to assume that the random errors for these processes are drawn from approximately normal distributions. In this case, there is a strong linear relationship between the residuals and the theoretical values from the standard normal distribution. Of course the plots do show that the relationship is not perfectly deterministic and it will never be but the linear relationship is still clear. Since none of the points in the plot deviate much from the linear relationship defined by residuals, it is also reasonable to conclude that there are no outliers in the data set.

The graph of residuals plotted against predicted values and the normal probability plot did not reveal any particularly destructive pattern, although the largest positive residual value observed slightly above 4 and 10 stands out from the others and the normal plot indicated few points at the extreme. These are not enough in the scattered plot to indicate unsuitability of the model for the study. According to Obi, 2002, it is possible that a particular treatment combination produces slightly more erratic response than the others. The problem more over is not severe enough to have a dramatic impact on the analysis and conclusions (Obi, 2002).

The normal probability plot helps us to determine whether or not it is reasonable to assume that the random errors in a statistical process can be assumed to be drawn from a normal distribution. An advantage of the normal probability plot is that the human eye is very sensitive to deviations from a straight line that might indicate that the errors come from a non-normal distribution. However, when the normal probability plot suggests that the normality assumption may not be reasonable, it does not give us a very good idea what the distribution does look like.

A histogram of the residuals from the fit on the other hand, can provide a clearer picture of the shape of the distribution. The fact that the histogram provides more general distributional information than does the normal probability plots suggests that it will be harder to discern deviations from normality than with the more specifically oriented normal probability plot.

The histogram for the study shown in *Figure 4* showed that the histogram is more or less bell-shaped, confirming the conclusions from the normal probability plots. One important detail to note about normal probability plot and the histogram according to Obi 2006, is that they provide information on the distribution of the random errors from the process only if: The functional part of the model is correctly specified, the standard deviation is constant across the data, there is no drift in the process and the random errors are independent from one run to the next.

CONCLUSION

A residual analysis procedure was successfully applied to analyze experiment model for plant spacing research on Watermelon using Watermelon yield as a parameter. The procedure proved to be very simple and easy to implement and it can be applied to any statistical model. The residual analysis showed that the conventional experiment model can be adequately used to study the effect of plant spacing on yield of Watermelon in particular and any other annual crop, generally. The plots were able to verify all questions pertaining to model validity. However, the R^2 was compared with that of Onuoha's (1995) experiment and it was discovered that they are similar which further proved Obi Design II to be true (Obi, 2013a). The scatter plot, residual plot and the histogram are similar with that of (Onuoha, 1995). Generally, the results showed that series of similar experiments methodology was able to model the changes associated with different planting date.

REFERENCES

Acikgoz, N. (1987). Effect of Sowing Time and Planting Method on Rice Yield per Day. *International Rice Research Notes* (IRRN) 12: 1-34.

Adeoti, A.A. and Emechebe, A.M. (1996). Effect of Sowing Date and Type of Fertilizer on Kenaf yield and it's reaction to *Coniella* leaf spot. *Samaru Journal of Agricultural Research*, 13:13-18.

Anyaeze, M.C. (1995). Studies on Time of Planting and on Some Agronomic Characteristics of Sorghum (*Sorghum bicolor* (L.) Monech) in the Nsukka Plains of Southeastern Nigeria. *Unpublished MSc Dissertation* – Department of Crop Science. University of Nigeria, Nsukka.

Bartlett, M.S. (1973). Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. *Journal of the Royal Statistical Society, Series A*, 4:137-185.

Chand, H. and Singh, R. (1985): Effect of Planting Time on Stem Rot (SR) Incidence. *International Rice Research Notes* (IRRN) 10:6-18.

Chandra, D. and Mama, G.B. (1988). Effect of planting dates, seedlings age and planting density on late planted wet season rice. *International Rice Research Notes* (IRRN) 13: 6-30

Chatterjee, S. and Price, B. (1977). *Regression Analysis by Example*. – New York: John Wiley and Sons, Inc., pp.19-22.

Chaudhry, M. (1984): Effect of Sowing Date on Growth and Performance of Six Rice Varieties in Wertern Turkey. *International Rice Research Notes* (IRRN), 9: 2-24.

Cirilo, A. G. and F. H. Andrade, (1994). Sowing date and maize productivity:1. Crop growth and dry matter partitioning. *Crop Science*, 34:1039-1043.

Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*. New York. John Wiley and Sons, Inc., 1998.

Fakorede, M.A.B., Kim, S.K., Mareck, J.H. and Iken, J.E. (1985). Breeding Strategies and Potentials of available varieties in relation to attaining self-sufficiency in maize production in Nigeria. Presented at the *National Symposium "Towards the attainment of Self-Sufficiency of Maize Production in Nigeria*, 9 March, 1985 I.A.R.& T., Ibadan. 15-30 pp

FDALR, (1985). Reconnaissance Soil Survey of Anambra State, Nigeria. Soils Report FDALR.

Federer, W.T. (1977). Sampling, blocking and model consideration for split plot and split block designs. *Biometrical Journal* 19, 181-200.

Federer, W.T., Obi, I.U. and Robson, D.S. (1983). Analysis of absolute values of Residuals to Test Distributional Assumption of Linear Model from Balanced Design. In: *Biometrics Unit*, Paper No. Bu-22-P, Cornell University, Ithaca, New York, U.S.A. 8pp.

Herbek, J. H., V. L. Murdock and R. L. Blevins, (1986). Tillage system and date of planting effects on yield of corn on soils with restricted drainage. *Agronomy Journal*, 78:580-582.

Maluf, O., Milan, M.T., Spinelli, D. and Martinez, M.E. (2005). Residual analysis applied to S-N data of a surface rolled cast iron – Mat. Res. 8(3) Sao Carlos Jully/Sept. 2005

Marzouk, I.A. and El-Bawab, A.M.O. (1999). Effect of sowing date of barley on its infestation with the corn leaf Aphid, *Ropalosiphum maides* (Fitch) (Homoptera Aphididae) and yield components. *Egyptian Journal of Agricultural Research*, 77(4):1493-1499.

Mclntosh, M.S. (1983). Analysis of combined experiments. *Agronomy Journal*, 75(1): 153-155.

Montgomery, D.C. (1991). *Design and Analysis of Experiments*. Third Edition. – John Willey and Sons, Inc. N.Y. 210-214 pp.

Obi, I.U. (2002). *Statistical Methods of Detecting Differences between Treatment Means and Research Methodology Issues in Laboratory and Field Experiments*. AP Express Publishers Limited, 3 Obollo Road, Nsukka – Nigeria. pp.117

Obi, I.U. (2006). What have I done as an Agricultural Scientist? (Achievement, Problems and Solution Proposals). *An Inaugural Lecture of the University of Nigeria, Nsukka, Nigeria*. Delivered on July 25, 2006. Published by the University of Nigeria Senate Ceremonials Committee.

Obi, I.U. (2011). *Introduction to Regression, Correlation and Covariance Analysis (with worked examples)*. Optimal International Ltd., 113 Agbani Rd. Enugu. 2nd Ed. Pp. 105.

Obi, I.U. (2013a). *Introduction to Factorial Experiments for Agricultural, Biological and Social Sciences Research*. Optimal International Ltd., 113 Agbani Rd. Enugu. 3rd Ed. Pp. 113.

Obi, I.U. (2013b). Post Graduate Lecture Notes on Biometrical Genetics. Department of Crop Production and Landscape Management, Ebonyi State University. *Unpublished Memo*.

Obi, I.U. (2013c). *Introduction to Factorial Experiment for Agricultural, Biological and Social Science Research (Third Edition)*. Optimal Computer Solution Ltd, Agbani Road, Enugu, Nigeria VI +47.

Obi, I.U., T. Vange and P.E. Chigbu, (2009). Using residual analysis to validate rice sowing dates experimental model.

Applied Ecology and Environmental Research, 149-163

Ogbonna, P.E. (1996): Studies on time of planting and some agronomic characteristics of the local seed type of "Egusi" melon (*Cucumis melo*) in the derived Savanna Zone of South-Eastern Nigeria. *Unpublished MSc. Dissertation*, -Department of Crop Science, University of Nigeria, Nsukka..

Onuoha, E. (1995): Response of maize (*Zea mays* L.) to time of planting, type and time of fertilizer application in Southeastern Nigeria. *Unpublished PhD Thesis*, Department of Crop Science, University of Nigeria, Nsukka.

Pearson, E.S. and Hartley, H.O. (1954): *Biometrika Tables for Statisticians*. (4th ed.) Cambridge (Eng): published for the Biometrika Trustee at the University press, 1954

Quesemberry, C.P. and David, H.A. (1961). Some test of outliers. *Biometrika*, 48(3-4): 379-390.

Sanders, D.C., Cure, J.D. & Schultheis, J.R., (1999). Yield response of watermelon to planting density, planting pattern and polyethylene mulch. *HortScience*, 34(7): 1221–1223.

Shapiro, S.S. and Francia, R.S. (1972). Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, 67 (337): 215-223.

Soroja, R. and Raju, N. (1983). Influence of planting time on rice leaf folder incidence. *International Rice Research Notes* (IRRN), 8: 6-17.

Steel, R.G.D. and Torrie, J.H. (1996). *Principles and procedures of Statistics: A Biometrical Approach*. McGraw-Hill N.Y. 195-233pp.

Urkurkar, J.S. and Chandrawashi, B.R. (1983). Optimum Transplanting Time for Rice: An Agrometrological Approach. *International Rice Research Notes* (IRRN), 8: 6-24.