USING RESIDUAL ANALYSIS TO VALIDATE RICE SOWING DATES EXPERIMENT MODEL

I.U. OBI[1] – T. VANGE[2*] – P.E. CHIGBU[3]

[1]*Department of Crop Science, University of Nigeria, Nsukka, Enugu State Nigeria*

[2]*Department of Plant Breeding and Seed Science, University of Agriculture, Makurdi, 970001, Benue State, Nigeria*

[3]*Department of Statistics, University of Nigeria, Nsukka, Enugu State, Nigeria*

**Corresponding author*
e-mail: t_vange@yahoo.com

(Received 21st August 2007 ; accepted 6th July 2009)

Abstract. Picking a model for a problem is a major undertaking. If the model fits well then it can be used to increase understanding of the problem and/or for prediction. Several statistical procedures have been used by researchers for planting date studies to suit their objectives. The ever-proliferation of Statistical procedures available to researchers has given room for use of diverse statistical design for research into finding optimum planting time for crops. Considering the subtle differences, advantages and disadvantage that these statistical designs pose, the results of such analysis may lead to false conclusion or be less reliable at least for comparative purposes. There is need to look at date of planting trials again to see the possibility of proffering a statistical model that could be commonly used by researchers. 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 sowing dates on yield of upland rice with the view of predicting optimum sowing date. Results show that the series of similar experiment methodology is able to model the changes associated with different sowing dates. The questions associated with model adequacy were discussed.

Keywords: *Sowing dates, rice, statistical model, residual analysis*

Introduction

Studies on dates of planting have been extensively conducted for several crops across different ecological zones. A review of studies on rice planting dates indicates that these research were necessitated by weather pattern [6, 38]; pest management [5, 18, 32, 36, 37]; disease management [2, 20]; and grain quality [9].

Several statistical procedures have been used by researchers for planting date studies to suit their objectives. Obi [23] proposes series of similar experiment models for time of planting experiments. Some researchers [3, 25, 27], used this series of similar experiment model in their study. A survey of other works on planting date show that [35] used a randomized block design with four complete blocks to study the effects of planting dates and residue rate effect on growth partitioning and yield of corn. Also [2, 18] used randomised complete block design in their planting date trials, while [28] in their trial on response of Soybean lines with “juvenile” trait to day length and sowing date combined both greenhouse and field work. In both cases they used factorial arrangement. Acikgoz [1] in his studies on effect of sowing time and planting method on rice yield per day also used factorial experiment. Fakorede [12], in maize planting date trial, used a randomised complete block design (RCBD) with a split-plot arrangement. He assigned dates to main plot, plant density to sub plot and genotypes to the sub-sub plot. Chand and Singh [5], used split-plot design in their study of effect of planting time on Stem Rot (SR) incidence in rice, [8] also used split-plot design for his studies. Both [5] and [8] assigned dates to main plot and variety to sub-plot.

Federer Design: Walter T. Federer, Professor of Biological Statistics, Biometrics Unit, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York (2001) (Personal communication with Obi, I. U.) gave his premise for Time (Date) of Planting which is Stated as follows: “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 Date of Planting, 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 Dates of Planting, say three (3) Dates with two (2) Replicates, each, Whole-plots would be Crops and the Split-plots would be Varieties within Crops”. Some form of an Analysis of Variance (ANOVA) as suggested by Federer is as presented in *Table 1*.

Table 1. *Form of partial analysis of variance, showing source of variation only (Walter T. Federer, April 2001, Personal communication in writing).*

Source of Variation
“Location = L
Year = Y
L x Y
Dates of Planting = D
Reps within D within L and within Y
D x L
D x Y
D x L x Y
Crops = C
C x L
C x Y
C x L x Y
C x D
C x D x L
C x D x Y
C x D x Y x D
Varieties within Crops, *etc.*

Verifying the adequacy of the linear model

The outlier tests [10, 15, 30, 34] as well as the variance homogeneity tests [4, 16, 29] have been used. Kirton, personal communication with Federer (1977), Department of Agriculture, New South Wales, Sydney, Australia, as cited by [23], suggested that after one has a “correct model,” one should proceed as follows in searching for a discrepant “treatment” or “block”: (i) *compute estimated 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 multiple comparisons on the *absolute* values of the estimated residuals. If the model is “correct”, then the null hypothesis should be true for all categories in the ANOVA except for residuals of 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 [13].

Federer *et al.,* [14] 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 largely used in Sowing dates data analysis, a verification of the suitability of the model is not always checked. Hence indicating that the search for the *Model* for the time (date) of planting 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 [23] to describe the effect of sowing dates on rice 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 a random 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 [31]. 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 [7, 11].

The independent 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 is given by:

$$Y_i = A + B\,X_i + \varepsilon_i \tag{Eq.1}$$

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, ε_i is the denominated random error. The value of ε_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 ε_i are random variables, assuming the following assumptions:

(i) The average of ε_i values is equal to zero and its variance, σ^2, is unknown and constant for $1 \leq i \leq n$;
(ii) ε_i values are not correlated;
(iii) The distribution of ε_i values is normal for $1 \leq i \leq n$.

Second and third assumptions imply that ε_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:

$$\hat{Y}_i = \hat{A} + \hat{B}X_i \tag{Eq.2}$$

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

$$\hat{e}_i = Y_i - \hat{Y}_i \quad \text{(Eq.3)}$$

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. But 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: So 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 could just plot them against a + bX. A random scatter would indicate that the errors do not depend on 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 $\hat{e}$, that is $\hat{e} = Y - (\hat{a} + \hat{b}X)$. Then we can check our model assumption by plotting $\hat{e}$ versus $\hat{Y}$. This is called the residual plot. A random scatter indicates a good model. If it is not a random scatter then we need to rethink the model [26]. The verification of residuals normality can also be analysed 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 [31].

Experimental Procedures

Field experiments were conducted during the 2002 and 2003 cropping season at three (3) locations within Benue State viz Makurdi ($7^0\,41'00''N;\ 8^0\,37'00''E$; 100 m above sea level), Otobi near Otukpo ($7^0\,13'\,00''N;\ 8^0\,9'\,00''E$; 105 m above sea level) and Yandev near Gboko ($7^0\,18'\,00''N;\ 8^0\,41'\,00''E$; 180 m above sea level). The plots were sown on June 15 and June 30 (early planting dates) and July 15 and July 30 (late planting dates).

Five (5) lines of rainfed upland rice comprising four newly developed New Rice for Africa (NERICA) lines (WAB 450-11-1-P-31-1-HB, WAB 450-1-B-P-38-I-HB, WAB 450-1-B-P-105-HB, WAB 450-1-B-P-160-HB) and a recommended variety ITA 150 (Standard check) obtained from the West Africa Rice Development Association (WARDA), IITA Ibadan were used for the experiment.

The treatments consist of five rice lines laid out in a Randomised Complete Block designs (RCBD) of three blocks/replications. Each treatment was assigned to an experimental unit of $12m^2$ size (3m x 4m) but data were taken from the inner $6m^2$ (2m x 3m) with 1m left as perimeter border. Replicates and experimental plots were separated length-wise by 1m walk-way and adjacent experimental plots were separated by 0.5m. This arrangement was repeated for each of the four planting dates at the three locations of study. Experimental layout of this form was described as series of similar experiment model (Obi, 2006). The form of analysis for each date of planting showing sources of variation and degrees of freedom (General) and (specific) are shown in *Table 2*. *Figure 1.* showed the field layout consisting of five rice lines, four date of planting as repeated in each of the three locations.

The experimental areas were ploughed and harrowed twice. The seeds were drilled at the rate of 50kg/ha (60g/plot) at 20cm apart; this gave a total of 16 rows of 4m length per plot. Fertilizer was applied at the rate of 75kg N, 60kg P, 60kg K per hectare at three split doses. Basal fertilizer application was done using NPK 15:15:15 brand of compound fertilizer at 30kg N/ha just before the second harrowing. The first top dressing was carried out at five (5) weeks after planting (5 WAP) at the same rate as basal application and using the same brand of fertilizer while the second top dressing was carried out with urea at 15kg N/ha (12 WAP) which corresponded with panicle initiation stage. All the three split fertilizers doses were applied using broadcasted method.

1. Date of Planting d_1

BLOCK I	v_1	v_2	v_4	v_3
BLOCK II	v_3	v_4	v_2	v_1
BLOCK III	v_2	v_3	v_4	v_1

2. Date of Planting d_2

BLOCK I	v_4	v_2	v_1	v_3
BLOCK II	v_1	v_4	v_3	v_2
BLOCK III	v_4	v_3	v_1	v_2

3. Date of Planting d_3

BLOCK I	v_2	v_4	v_3	v_1
BLOCK II	v_1	v_2	v_4	v_3
BLOCK III	v_4	v_3	v_2	v_1

4. Time of Planting d_4

BLOCK I	v_2	v_1	v_3	v_4
BLOCK II	v_4	v_2	v_1	v_3
BLOCK III	v_3	v_3	v_4	v_2

Figure 1. *The field layout for Date (Time) of Planting Experiment as Series of Similar Experiments (d_0 = First Date of Planting and d_4 = Date of Planting 8 weeks after)*

Table 2. *Form of Analysis of Variance for each Date of Planting Showing: Sources of Variations and Degrees of Freedom (General) and (Specific), Only*

Source of Variation	d.f. (General)	d.f. (Specific)
Block/Replication	r – 1	2
Varieties	v – 1	4
Experimental Error = Block x Variety	(r – 1)(v – 1)	8
Total	Vr – 1	14

A combination of two post-emergence herbicides (propanil and 2,4-D.) were applied at 4 WAP at a mixture ratio of 2:1 with propanil at 1.6L/ha and 2,4-D at 1.1L/ha using CP 3 knapsack sprayer. Supplemental weed control was manually done using the local hoe and hand picking as was necessary during the season.

Data on yield was collected at maturity; the plants from the inner 2m x 3m (6m^2) of each experimental unit (i.e. with 1m left as perimeter border) were harvested, threshed carefully, winnowed and the grains weighed and recorded in kilograms. Grain yield per plot was converted to tonnes per hectare.

The Analysis of Variance was performed using the procedure outlined by [33] for each measured parameter. Means that had a significant F-test were separated using LSD 0.05 [24]. The data analysis was carried out in stages.

Test of Homogeneity of variance

Bartlett's [4] test for homogeneity of variance was conducted in order to determine whether or not the assumption of homogeneity of variance was met [23].

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

Stage II: A combined analysis of variance was done for yield measured over date of planting in each location following the procedure of analysis of combined experiments as outlined by [19]. The form of combined analysis for each trait measured for date of planting 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

$$Y_{ijkl} = \mu + \rho_j + \alpha_i + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijkl}$$

Y_{ijkl} = Observation made on the ith variety within the jth replication in the kth planting date within the lth location

μ = overall experimental mean

ρ_j = effect of the jth Block/replication within time

α_i = effect of the ith variety

β_k = effect of the kth date of planting

$(\alpha\beta)_{ik}$ = Interaction effect of ith variety with the kth planting date

ε_{ijkl} = random unit of ith variation within the jth block within the kth planting date on the ith variety component of experimental error.

i = 1,2, 3, 4 and 5 rice lines.

j = 1,2 and 3 block.

k = 1,2,3 and 4 planting dates.

l = 1,2 and 3 locations

Table 3. *Form of Combined Analysis of Variance for Four Date of Planting Showing Sources of Variance, Degrees of Freedom (General) and (Specific) only*

Source of Variation	**d.f. (General)**	**d.f. (Specific)**
Varieties (V)	v – 1	(4)
Date of Planting (D)	d – 1	(3)
V x D	(v – 1)(d – 1)	(12)
Block within Date	(r-1)(d)	(8)
Experimental Error = Block x DV	d(r – 1)(v – 1)	32
Total	rdv – 1	59

Stage III: A combined analysis of variance was done for yield measured over the three locations. Date of plantings as well as locations were 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 4*.
The linear statistical model used for the analysis of variance is as follows:-

$$Y_{ijkl} = \mu + \alpha_i + \rho_j + \beta_k + \gamma_l + (\alpha\beta)_{ik} + (\alpha\gamma)_{il} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{ikl} + \varepsilon_{ijkl}$$

Y_{ijkl} = Observation made on the ith variety within the jth replication in the kth planting date within the lth location
μ = overall experimental mean
α_i = effect of the ith variety
ρ_j = effect of the jth block/replication
β_k = effect of the kth planting date
γ_l = effect of the lth location
$(\alpha\beta)_{ik}$ = Interaction effect of ith variety with the kth planting date
$(\alpha\gamma)_{il}$ = Interaction effect of the ith variety with lth location
$(\beta\gamma)_{kl}$ = Interaction effect of the kth planting date with the lth location
$(\alpha\beta\gamma)_{ikl}$ = Interaction effect of the ith variety on the kth planting date within lth location.
ε_{ijkl} = random unit of variation within the jth replication within the kth planting date within the lth location on the ith variety component of experimental error.
i = 1,2, 3, 4 and 5 rice lines.
j = 1,2 and 3 blocks/replications.
k.= 1,2,3 and 4 planting dates.
l.= 1,2 and 3 for locations.

Table 4. *Form of combined analysis of variance for three locations and four dates of planting showing Sources of variation (Source), Degrees of freedom (d.f)*

Source	**d.f, (General)**	**d.f (Specific)**
Location (L)	(l-1)	2
Dates (D)	(d-1)	3
L x D	(l-1)(d-1)	6
Block/Location	l(r-1)	6
Rice Lines (V)	(v-1)	4
L x V	(l-1)(v-1)	8
D x V	(d-1)(v-1)	12
L x D x V	l(dv-1)(r-1)	24
Residual	lv(r-1)	114

where, l, d, v and r are number of locations, dates of planting, rice lines, and number of blocks, respectively

Results

Series of similar experiment model involve analyses of variance in stages. The example of the results of mean square of the stage I of the analysis of variance and degree of freedom for yield on June 15 is presented in *Table 5*. The example of the results of variance for traits measured from the four dates of planting at Makurdi and the results of Bartlett's test of homogeneity of variance is shown in *Table 6*. The Bartlett's test was not significant. Based on non significant Bartlett's test of homogeneity of variance, the procedure of analysis of combined experiments as outlined by McIntosh (1983) was used to combine the four planting dates at each location. This is the stage II of the analysis of series of similar experiment model. An example of the results of the combined analyses of the date of planting for each of the location is presented in *Table 7*.

Table 5. *Sources of Variation, Degrees of freedom (D.F) and Mean Squares from analyses of variance for yield of five rice lines grown at June 15 at Makurdi, Otobi and Yandev during 2002 and 2003 cropping seasons*

Source of Variation	**d.f**	**Mean squares**					
		2002			**2003**		
		Makurdi	Otobi	Yandev	Makurdi	Otobi	Yandev
Block/Replication	2	1.03	3.39	0.40	1.39	3.38	0.43
Varieties	4	0.43	0.49	0.51	0.27	0.45	0.54
Experimental Error	8	0.34	0.73	0.27	0.37	0.71	0.29
Total	14						

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

Table 6. *Error Squares (Variances) for Yield of Five Varieties of Rice Planted on fourth-nightlJy interval (15th June – 30th July, 2002 and 2003) in Makurdi, Otobi and Yandev*

Date of Planting	**d.f.**	**Mean squares**					
		2002			**2003**		
		Makurdi	Otobi	Yandev	Makurdi	Otobi	Yandev
June 18	8	0.3369	0.7278	0.2743	0.7264	0.2569	0.2966
July 2	8	1.023	0.4921	0.2390	0.4688	0.2007	0.06184
July 16	8	1.023	0.1208	1.2220	0.09375	0.0938	0.6448
July 30	8	0.8663	1.562	0.5014	0.3070	0.4492	0.7734
X^2 Calculated		6.03NS	11.05NS	6.96NS	7.435NS	4.611NS	11.33NS

Critical values of χ^2 at 5%, 8 d.f. = 15.51; Critical values of χ^2 at 1%, 8 d.f. = 20.09
NS = Not significant.

The Results of the stage III analysis which is the combined analysis of variance for yield of five varieties of rice planted over four sowing/planting dates at Makurdi, Otobi and Yandev locations in 2002 and 2003 are presented in *Table 8.* and their Bartlett's test presented in *Table 9.*, respectively. The Bartlett's test was not significant. The results of combined analysis of variance done for each character measured for the two years of study across the four date of planting among the three locations was pooled and is presented in *Tables 10*. The model for this last stage of combined analysis was validated.

Table 7. *Combined analysis of variances showing sources of variation, degrees of freedom (d.f) and mean squares from analyses of variance for yield of five varieties of rice planted on fourth-nightly interval (15th June – 30th July, 2002 and 2003) at Makurdi, Otobi and Yandev*

Date of Planting	**d.f.**	**Mean Squares**					
		Locations					
		2002			**2003**		
		Makurdi	Otobi	Yandev	Makurdi	Otobi	Yandev
Varieties (V)	4	1.68*	2.49*	3.42**	1.411*	0.66	0.71*
Date of Planting	3	2.92**			7.54**	7.39**	0.15
(D)			13.52**	11.36**			
V x D	12	0.86NS	0.65NS	1.92**	0.76	0.23	0.19
Block within Date	8	2.31**	3.32**	3.09**	1.71**	1.67**	0.46
Error	32	0.61	0.73	0.56	0.44	0.42	0.25
Total	59						

*, ** = significant at 5% and 1% respectively.
NS = non significant.

Table 8. *Combined analysis of variance for yield of five varieties of rice planted over four sowing/planting dates at Makurdi, Otobi and Yandev locations in 2002 and 2003*

Date of Planting	**d.f.**	**Mean Squares**	
		2002	**2003**
Location (L)	2	0.84**	19.35**
Dates (D)	3	0.16**	8.56**
L x D	6	0.42**	3.26**
Block/Location	6	0.06	1.11
Rice Lines (V)	4	0.24**	2.23**
L x V	8	0.03	0.28
D x V	12	0.05	0.67
L x D x V	24	0.03	0.26
Residual	114	0.04	0.52
Total	179	0.07	0.98

*, ** = significant at 5% and 1% respectively. NS = non significant.

Table 9. *Error squares (variances) for yield five varieties of rice planted on fourth-nightly interval (15th June – 30th July,) in Makurdi, Otobi and Yande in 2002 and 2003*

Date of Planting	**Error d.f.**	**Error Mean Squares (variances)**	
		2002	2003
Otobi	32	1.227	0.42
Yandev	32	1.058	0.25
Makurdi	32	0.9026	0.44
X^2 Calculated		0.4596	2.17

Critical values of χ^2 at 5%, 32 d.f. = 43.77

Critical values of χ^2 at 1%, 32 d.f. = 50.89

The R^2 for this model accounted for 61.3%. The graphical residual analysis of this trial which is given by the graph of residuals plotted against predicted values is presented in *Figure 1*. The plot did not revealed anything particularly troublesome pattern other than a random pattern, although the largest positive residual value observed slightly above 3 stands out from the others.

The normal probability plot is presented in *Figure 2*. The normal probability plots for this indicate that that it is reasonable to assume that the random errors for these processes are drawn from approximately normal distributions.

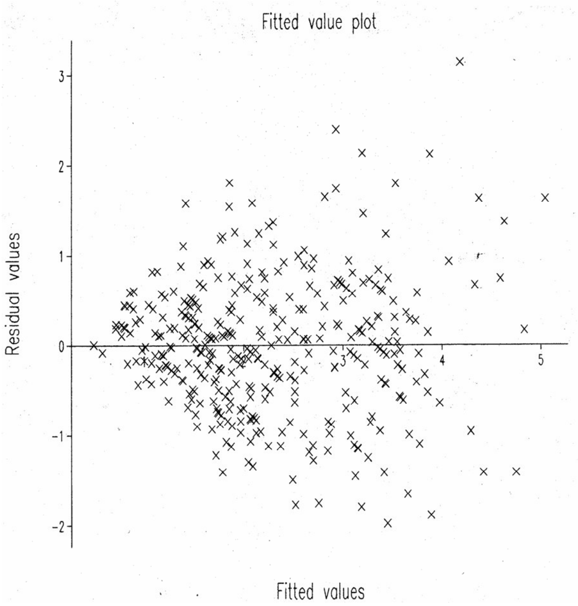

Figure 2. *Plot of Predicted Yield (t/ha) versus Yield (t/ha) residual*

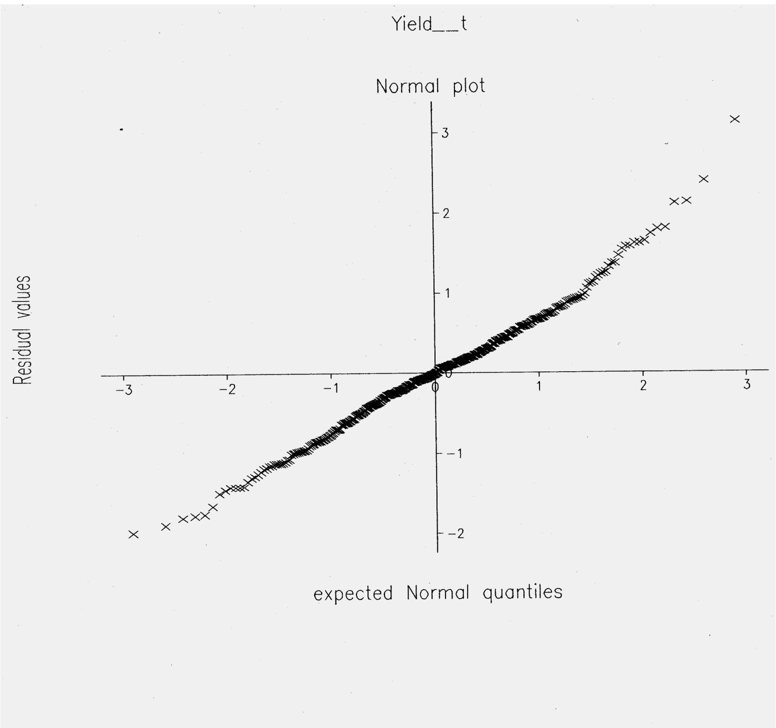

Figure 3. *Plot of residual against normal scores for Yield (t/ha)*

The Histogram for this study is shown in Figure 4. The histogram is more-or-less bell-shaped.

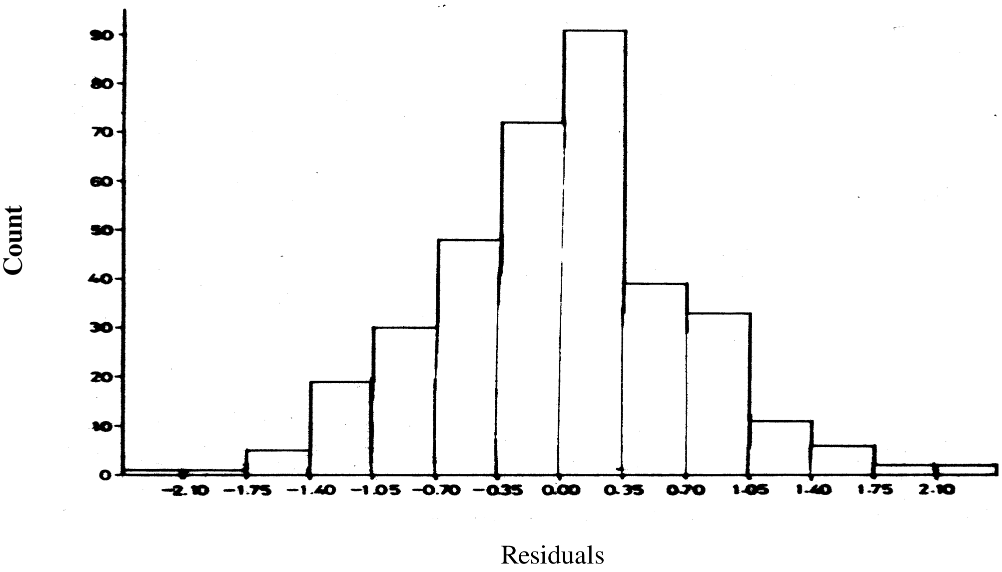

Figure 4. *Histogram of residual for yield (t/ha)*

Discussion

Model validation is possibly the most important step in the model building sequence. It is also one of the most overlooked. 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 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 answers to the underlying scientific question of how one can know if a model fits the data under investigation. Even though the R^2 for this model accounted for 61.3%, the primary statistical tool for most process modeling applications is graphical residual analysis [21]. In addition, the normal probability plot also serves as to confirm the adequacy of a model. [21, 26].

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 be narrowly focused on a particular aspect of the relationship between the model and the data and often try to compress that 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 revealed anything particularly troublesome pattern other than a random pattern, although the largest positive residual value observed slightly above 3 stands out from the others. It is not enough in the scattered plot to indicate unsuitability of the model for the study. According to [21], 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 [21].

The assessment of the sufficiency of the functional part of a model also depends on the scatter plot of the residuals versus the predictor variables in the model and versus 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 versus 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 indicate a systematic structure.

The question of how to 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. If the random errors are not from a normal distribution, incorrect decisions will be made more or less frequently than the stated confidence levels for our inferences indicate.

The normal probability plot is constructed by plotting the sorted values of the residuals versus 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. Instead, if the random errors are normally distributed, the plotted points will lie close to straight line. 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 some outliers in it.

The normal probability plot in *Figure 3.* indicated that 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 never will be), but the linear relationship is still clear. Since none of the points in these plots deviate much from the linear relationship defined by the residuals, it is also reasonable to conclude that there are no outliers in any of these data sets.

The graph of residuals plotted against predicted values and the normal probability plot did not reveal anything particularly troublesome pattern, although the largest positive residual value observed slightly above 3 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 [21] 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 [21]

The normal probability plot helps us 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 plot suggests that it will be harder to discern deviations from normality than with the more specifically-oriented normal probability plot.

The Histogram for this study shown in *Figure 4.* indicated that the histogram is more-or-less bell-shaped, confirming the conclusions from the normal probability plots. One important detail to note about the normal probability plot and the histogram according to [22] is that they provide information on the distribution of the random errors from the process only if

1. the functional part of the model is correctly specified,
2. the standard deviation is constant across the data,
3. there is no drift in the process, and
4. the random errors are independent from one run to the next.

If the other residual plots indicate problems with the model, the normal probability plot and histogram will not be easily interpretable.

Conclusion

A residual analysis procedure was successfully applied to analyze series of similar experiment model for sowing date studies in rice using rice 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 series of similar experiment model can be adequately used to study effect of sowing dates on yield of rice in particular and any other annual crop, generally. The plots verify all questions pertaining to model validity.

REFERENCES

[1] Acikgoz, N. (1987): Effect of sowing time and planting method on rice yield per day. – International Rice Research Notes (IRRN) 12: 1-34.

[2] 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 J. Agric. Res. 13: 13-18.

[3] Anyaeze, M.C. (1995): Studies on time of planting and on some agronomic characteristics of Sorghurn (*Sorghurm bicolor* (L.) Monech) in the Nsukka plains of Southeastern Nigeria.(Unpublished M.Sc. Dessertation – Department of Crop Science, University of Nigeria, Nsukka).

[4] Bartlett, M.S. (1937): Some Examples of Statistical Methods of Research in Agriculture and Applied Biology. – J. Roy. Stat. Soc. Suppl**.** 4: 137-185.

[5] Chand, H. and Singh, R. (1985): Effect of planting time on stem rot (SR) incidence. – International Rice Research Notes (IRRN) 10: 6-18.

[6] Chandra, D. and Manna, 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.

[7] Chatterjee, S. and Price, B. (1977): *Regression Analysis by example*. – New York: John Wiley & Sons, Inc., 1977.

[8] Chaudhry, M. (1984): Effect of sowing date on growth and performance of six rice varieties in Western Turkey – IRRN 9: 2-24.

[9] Cuevas-Perez, F. and Benwrio, L.E. (1987): Influence of planting date on milling performance of rice varieties under delayed harvesting. – International Rice Research Notes (IRRN) 12: 5-10.

[10] Dixon, W.J. (1953): Processing Data for Outliers. – Biometrica 9(1): 74-79.

[11] Draper, N.R. and Smith, H. (1998): *Applied Regression Analysis*. – New York, John Wiley & Sons, Inc., 1998.

[12] Fakorede, M.A.B. (1985): Response of maize to planting dates in a tropical rainforest location. – Expl. Agric. 21: 19-30.

[13] Federer, W.T. (1977): A Problem in Residual Analysis, An M.S. Thesis Problem. Bu-629 M. – Biometrics Units Memo. Series, Cornell University, Ithaca New York, 1483., U.S.A.

[14] Federer, W.T., Obi, I.U. and Robson, D.S. (1983): Analysis of Absolute Values of Residuals to Test Distributional Assumptions of Linear Models from Balanced Design. Paper No. Bu-22-P, – In the Biometrics Unit, Cornell University, Ithaca, New York, U.S.A. 8pp.

[15] Grubbs, F.E. (1969): Procedures for Detecting Outlying Observations in Samples. – Technometrics 11(1): 1-21.

[16] Levene, H. (1960): In Contributions to Probability and Statistics. – Stanford Univ. Press, Stanford, Calif., U.S.A. p. 278.

[17] Martínez, M.E (2001): *Desenvolvimento de um modelo estatístico para aplicação no estudo de fadiga em emendas dentadas de madeira*. [PhD Thseis]. – São Carlos: University of São Paulo; 2001.

[18] Marzouk, I.A. and El-Bawab, A.M.O. (1999): Effect of sowing date of barley on it's infestation with the corn leaf Aphid, *Ropalosiphum maides* (Fitch)(Homoptera: Aphididae) and yield components. – Egypt. J. Agric. Res. 77(4): 1493-1499.

[19] McIntosh, M.S. (1983): Analysis of Combined experiments. – Agron. J. 75(1): 153-155.

[20] Mohan, S., Vasuderan, P. and Gururajan, B. (1984): Influence of planting time on blast (B1) incidences. – International Rice Research Notes (IRRN) 9: 5-13.

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

[22] NIST/SEMATECH (2006): NIST/SEMATECH e-Handbook of Statistics Methods, http://www.itl.nist.gov/div898/handbook/ assessed on November 30, 2006 and confirmed on January 30, 2007.

[23] Obi, I.U. (2006): What have I done as an agricultural scientist? (Achievements, 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.

[24] Obi, I.U. (2002): Statistical methods of detecting differences between treatment means. – Department of Crop Science, University of Nigeria, Nsukka – Nigeria, Monographs.

[25] 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 southeastern Nigeria.(Unpublished M.Sc. Dessertation, – Department of Crop Science, University of Nigeria, Nsukka).

[26] Maluf, O., Milan, M.T., Spinelli, D. and Martínez, M. E. (2005): Residual analysis applied to S-N data of a surface rolled cast iron – Mat. Res. 8(3) São Carlos July/Sept. 2005

[27] Onuoha, E. (1995): Response of maize (*Zea mays* L.) to time of planting, type and time of fertilizer application in southeastern Nigeria (Unpublished Ph.D. Thesis, – Department of Crop Science, University of Nigeria, Nsukka).

[28] Parvez, A.Q. and Gardner, F.P. (1987): Day length and sowing date responses of soybean lines with "juvenile" trait. – Crop Sci. 27: 305-310.

[29] Pearson, E.S. and Hartley, H.O. (1954): Biometrika Tables for Statisticians.

[30] Quesenberry, C.P. and David, H.A. (1961): Some Test of Outliers. – Biometrika 48(3-4): 379-390.

[31] 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.

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

[33] Steel, R.G.D. and Torrie, J.H. (1980): Principles and Procedures of Statistics: A Biometrical Approach. – Mc Graw-Hill N.Y. 195-233pp.

[34] Stefansky, W. (1972): Rejecting Outliers in Factorial Design. – Technometrics 14(2): 469-479.

[35] Swanson, S.P. and Wilhelm, W.W. (1996): Planting date and residue rate effects on growth, partitioning and yield of corn – Agron. J. 88: 205-210.

[36] Ukwungwu, M.N. (1984): Planting time and stem borer incidence in Badeggi, Nigeria. – International Rice Research Notes (IRRN) 9: 1-22.

[37] Ukwungwu, M.N. (1987): Influence of age of crop and time of planting on gall midge (GM) incidence. – International Rice Research Notes (IRRN) 12: 3-32.

[38] Urkurkar, J.S. and Chandrawashi, B.R. (1983): Optimum transplanting time for rice: an agrometeorological approach. – International Rice Research Notes (IRRN) 8: 6-24.