

*Annals Of Agric. Sc., Moshtohor,*  
*Vol. 45(3): 1005-1020, (2007).*

**CHECKING VIOLATION OF ASSUMPTIONS UNDERLYING THE  
ANALYSIS OF VARIANCE  
BY**

**Tageldin, M.H.A.**

Agronomy Department, Faculty of Agriculture, Moshtohor, Banha University

**ABSTRACT**

The strategy for checking any violations of the analysis of variance assumptions has rarely been adopted among researchers and data analysts especially with models that contain more than one error term. The objective of this work is to set out steps to check these assumptions of a split-plot design. Analysis of variance assumptions have been checked for a field experiment laid out in a split-plot design. Three whole-plot (WP) nitrogen fertilizer rates were laid out in four randomized blocks, and maize (*Zea mays* L.) cultivars were in four split plots (SP). To fulfill this objective a five-stage strategy was followed. These checking stages were for: fit of the model, outliers, independence of the error variable, equality of error variances, and finally normality assumption. This strategy depends mainly on calculating the residual values of both the main- and the sub-plot factors. In addition, the coefficients of multiple determination,  $R^2$ , were calculated for whole- and sub-plots as measures of model lack of fit. Some assumptions seemed to be violated during the two years of the study. In Year 1, the pattern of residuals of the three N levels resembled a funnel-like shape; this warrants a possible violation of equality of error variance assumption. In Year 2, plotting of residuals indicates the presence of an extreme residual point at -2.72078 standard units that lies within Block 2 with both N Level 1 and Cultivar 2. Regarding the independence issue, the residual spatial pattern, represented here by spatial arrangement (block factor), did not seem to exhibit apparent problems in both years. Based on the residuals associated with only subplots (SP), in Year 1, the variations in residuals within Cultivars 2 and 4 were more pronounced compared to those within Cultivars 1 and 3. In Year 2, a relatively more spread exists within Cultivars 4. Coefficients of determination were about 9.0 and 25.0 % for WP and SP in Year 1, and were quite improved in Year 2 to become 39.0 and 58.0 %, respectively. For the WP sub designs in both year, the extremely low coefficient values indicate that a small proportion of the variability in the data is contributed by the WP nitrogen fertilizer factor included in the model, whereas these values were relatively higher for SP sub designs. The WP N factor should not be replaced, but this warrants more attention be paid when applying nitrogen fertilizer rates to main plots. Nearly all non significant effects had extremely small (<1.0) F ratios. These <1.0 values, however, may indicate violations of one or more of ANOVA assumptions. Both error variance equality and normality assumptions seemed not to be violated in both years.

## INTRODUCTION

The analysis of variance, ANOVA, requires certain assumptions be fulfilled for the validity of both significance testing and inference(s) making. Ways of assessing aptness of the ANOVA model assumptions have become a necessary step to be implemented by researchers and data analysts.

In agricultural research, these assumptions are very often assumed satisfied in the data being analyzed and rarely checked for. "The data analyst is responsible for checking that the principal assumptions seem reasonably well satisfied in the data being analyzed" (Snedecor and Cochran, 1980, p. 274). These assumptions are often either mentioned briefly or even entirely overlooked in introductory statistical literature. In the 1947 *Biometrics Journal*, different, but related, topics concerning these assumptions were fully discussed in three classical consecutive articles written by Cochran, Bartlett and Eisenhart. In these three papers, Eisenhart discussed the ANOVA assumptions, Cochran addressed consequences when assumptions are not met, and finally Bartlett dealt with transformation as a remedial tool.

The assumptions needed in the analysis of variance are mainly: i) both the effects of treatments and environment must be additive; and the experimental errors must be ii) independent; iii) of equal variance and mutually uncorrelated; and iv) normally distributed (Cochran, 1947; Wiesberg, 1980, p. 119; Draper and Smith, 1981, p. 22).

In his article, Eisenhart (1947, p. 2) argued why statistics books, of his time, had failed concerning this issue since they did not "state explicitly the several assumptions underlying the analysis of variance, and to indicate the importance of each from a practical viewpoint." He also added that books "have not generally indicated in sufficient detail the actual functions of the respective assumptions—1) which can be dispensed with for certain purposes; 2) which are absolutely necessary, and what are likely to be the consequences if these are not fulfilled; and 3) what can be done "to bring into line," for purposes of the analysis, data which in their original form are not amenable to analysis of variance".

This implies that all these ANOVA assumptions, in real-life data, are not quite equal in their negative impact on analysis of variance and significance tests, and possible remedies should be thought of for problematic data sets. Minor departures from the underlying assumptions do not disturb the conclusions or significance tests to any important degree, but major violations are most likely to invalidate conclusions and must be avoided (Damon and Harvey, 1987, p. 75).

Cochran (1947) emphasized mainly various violation effects. Non additivity causes error variances be heterogeneous. Non independence leads to biases in estimating standard errors. Heterogeneous error variances, as well as non normality cause a loss in the efficiency of estimating treatment effects, and

the latter causes a drop in power in both F and student's t tests. In addition, in t tests, estimated pooled error variance which is based on heterogeneous errors and non normality lead to serious distortion of significance levels. The tabular probability, in non normal distributions, is an underestimate, that is by using F or t tables we intend to err in the direction of announcing too many significant results. ANOVA, however, is robust to non normality, but not quite so robust with respect to heterogeneous error variance (Milliken and Johnson, 1984, p.17). In many recent studies the impacts of failure of one/more of the assumptions on the validity of significance tests have extensively been considered (see Bathke, 2003; Maas and Hox, 2003; Chiarotti, 2004; Zimmerman, 2005; and Meek, Ozgur, and Dunning, 2007).

Generally in practice several assumptions may fail to hold simultaneously. In non normal distributions there is usually a correlation between the variance and the means, so that failure of Assumption 4, normality, is likely to be accompanied by failure of Assumption 3, homogeneous error variance (Cochran, 1947). The assumptions altogether are needed to get an unbiased estimate of error variance and exact tests of significance. For the latter, both independence and normality are especially required. If additivity, common variance are fulfilled, but neither independence nor normality, this results in getting different estimates of error variance in the expected mean squares in a two-way ANOVA, which are complex weighted averages of variances and covariances of the random variable at hand (Eisenhart, 1947).

The likely disruption(s) of model assumptions --in real-life situations-- that might be caused by one or more of these violations, though quite crucial to be familiar with their possible effect on the validity of significance testing, yet gaining knowledge of various diagnostic tools is rather equally important. In Neter, Wasserman, and Kutner (1985, pp 602-635) they generally discussed some of these diagnostic tools especially residual analyses. In regression models, Draper and Smith (1981) also emphasized residual diagnostic tools relatively more deeply.

For one-way ANOVA model, Dean and Voss (1999) have set out a sequential strategy for checking model assumptions; however, it can be extended to any other statistical models. First, the fit of the model is checked by plotting the standardized residuals versus the levels of each of the independent variable (treatment factor, block factor) included in the model. Second, outliers – any unusual observations. Outliers are easy to detect from a plot of the standardized residuals versus the levels of the treatment factors. Third, do the error variables  $\mathcal{E}_{ijk}$  appear to be independent? It is checked for by plotting the standardized residuals against any temporal order (time) or spatial arrangement (blocks) by which the corresponding observations were handled. Fourth, does the error variable  $\mathcal{E}_{ijk}$  have similar variances for each treatment? In this case the standardized residuals are plotted against the fitted values  $\hat{y}_{ijk}$ . Fifth, does the error variable  $\mathcal{E}_{ijk}$  appear to be a random sample from a normal distribution? It can be checked by applying the *normal probability plot* procedure (Draper and

Smith, 1981, p.117; and Dean and Voss, 1999, p.119). It is a plot of the standardized residuals against their normal scores. Recently Almimi, Kulahci, and Montgomery (2007) have argued that the property of designs having more than one error term should be taken into consideration when computing measures of adequacy of model fit. They proposed the computation of two,  $R^2$ ,  $R^2$ -Adjusted, prediction Error Sums of Squares (PRESS), and  $R^2$ -Prediction statistics to measure the adequacy of fit for the whole and the subplot sub-models in a split-plot design. This is complemented with the graphical analysis of the two types of errors to check for any violation of the underlying assumptions.

Assessing these tools in models having more than one error term needs to be addressed, since most literature have mainly focused on one-way models only. This paper, therefore, aims merely at presenting some essential diagnostic statistical tools that have been applied to check up for model assumption disruptions in a split plot design. Suggesting ways of fixing any violation(s), if present, is beyond the scope of this work.

## MATERIALS AND METHODS

### Experimental data

Data used in the analyses came from a two-year maize field trial that was initially laid out in 2002 in a split plot design. Nitrogen fertilizer was the main factor in a four randomized complete blocks, and maize cultivars were four splits on three nitrogen rates. For more details of the materials, refer back to Tageldin (2005).

### Checking model assumptions

A split-plot linear additive model is represented by

$$Y_{ijk} = \mu + \rho_k + \alpha_i + \gamma_{ik} + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad (1)$$

let  $Y_{ijk}$  represents the observation in the  $k$ th block of a randomized complete block design on the  $i$ th whole-unit treatment with  $j$ th subunit treatment. Let  $k=1, \dots, r$  blocks,  $i=1, \dots, a$  a whole-unit treatment, and  $j=1, \dots, b$  subunit treatments. Let  $\gamma_{ik}$  and  $\varepsilon_{ijk}$  be normally and independently distributed with zero mean with  $\sigma^2_\gamma$  as the common variance of the  $\gamma_{ik}$ 's, the whole-unit random components, and with  $\sigma^2_\varepsilon$  as the common variance of  $\varepsilon$ 's, the subunit random components. Here, the  $\rho_k$  is assumed random, and both  $\alpha_i$  and  $\beta_j$  are the fixed components of the model.

### Examination of Residuals

The assumptions on the model involves the error variable,  $\varepsilon_{ijk} = Y_{ijk} - E(Y_{ijk})$ , and can be checked by examination of the residuals. The  $ijk$  th residual  $e_{ijk}$  is defined as the observed value of  $Y_{ijk} - \hat{Y}_{ijk}$  where  $\hat{Y}_{ijk}$  is the least squares estimator of  $E(Y_{ijk})$ , that is  $e_{ijk} = Y_{ijk} - \hat{Y}_{ijk}$ .

Then, the residuals are transformed into standardized residuals, since standardization facilitates the identification of outliers (Dean and Voss, 1999). The standardization is achieved by dividing the residuals by their standard deviation that is by the square root of mean square error estimate  $\hat{\sigma}^2_e$ . The standardized residuals have variance equals 1.0, and calculated by

$$z_{ijk} = \frac{\hat{\epsilon}_{ijk}}{\sqrt{\frac{SS_e}{df}}} \tag{2}$$

**Residual plots**

A residual plot is a plot of the standardized residuals  $Z_{ijk}$  against the levels of another variable. The choice of a particular factor on the assumption being checked (Dean and Voss, 1999, p105).

**Sequence of checking model assumptions**

Dean and Voss (1999) suggested a strategy for checking any violations in model assumptions. They are:

- i) Model adequacy -- it deals with exploring the form of the chosen model -- are the mean response for the treatments adequately described by

$$E(Y_{ijk}) = \mu + \rho_k + \alpha_i + \gamma_{ik} + \beta_j + (\alpha\beta)_{ij} \tag{3}$$

The fit of the model is checked by plotting the standardized residuals versus the levels of each of the independent variable (treatment factor, block factor) included in the model.

- ii) Outliers -- any unusual observations. An outlier is an observation that is much larger or much smaller than expected. This is indicated by a residual that has an unusually large or negative value, Outliers are easy to detect from a plot of the standardized residuals versus the levels of the treatment factors.
- iii) Independence -- do the error variables  $\epsilon_{ijk}$  appear to be independent? It is checked for by plotting the standardized residuals against any temporal order (time) or spatial arrangement (blocks) by which the corresponding observations were handled.
- iv) Constant variance -- do the error variables  $\epsilon_{ijk}$  have similar variances, i.e. homoscedastic? In this case the standardized residuals are plotted against the fitted values  $\hat{Y}_{ijk}$ .
- v) Normality -- do the error variable  $\epsilon_{ijk}$  appear to be a random sample from a normal distribution? It can be checked by applying the *normal probability plot* procedure (Draper and Smith, 1981, p.117; and Dean and Voss, 1999, p.119; and Almimi, Kulahci, and Montgomery, 2007). It is a plot of the standardized residuals against their normal scores. Normal scores are percentiles of the standard normal distribution. These normal scores are calculated by using Blom's  $q$  th normal score technique (Dean and Voss, 1999, p.119). It is the value  $\xi_q$  for which

$$P(Z \leq \xi_q) = (q - 0.375) / (n + 0.25) \tag{4}$$

where  $Z$  is a standard normal random variable. Blom's  $q$ th normal score is

$$\xi_q = \phi^{-1} [(q - 0.375) / (n + 0.25)] \tag{5}$$

Where  $\phi$  is the cumulative distribution function (cdf) of the standard normal distribution. The normal scores possess a symmetry about zero mean that is the  $j$  th smallest and the  $j$  th largest scores are always equal in magnitude but opposite in sign.

**Calculations of the whole (WP) and the subplot (SP) residuals**

Almimi, Kulahci, and Montgomery (2007) outlined the procedure to calculate the WP and the SP residuals. Once the model is fitted, obtain the residual values. The WP residuals are calculated by averaging the residuals corresponded to the replications of each WP. All the replications within a WP have the same residuals. The SP residuals, on the other hand, are obtained by subtracting the WP residuals from the whole-model residuals.

In addition, to calculate  $R^2$  for both the WP and the SP sub designs, the values of the all-model sum of squares are divided into two groups: the first has the sums of squares for the WP and the second has the sums of squares for the SP effects. The values of  $R^2$  for both the WP and the SP sub designs are calculated as follows:

$$R^2_{WP} = \frac{SS_{Model(WP)}}{SS_{Total(WP)}} \text{ and } R^2_{SP} = \frac{SS_{Model(SP)}}{SS_{Total(SP)}} \quad (6)$$

**RESULTS AND DISCUSSION**

The fit of the chosen model is generally checked by plotting the standardized residuals versus the levels of each independent variable (Figs 1-3) in both years of the study. Figures 1a&b show the pattern of the standardized residuals in relation to the three N fertilizer levels in 2002 and 2003. In Fig. 1a, the three N levels exhibited a similar residual pattern around mean zero. They did not exhibit a non random pattern, i.e. too often positive values for some levels and too often negative for others. However, in Level 2 the residuals were nearly clustered around zero compared to the other two levels.

In 2003, the pattern was entirely different (Fig. 1b). The pattern among the three N levels resembled a funnel-like shape; this implies that the variance was increasing towards the lower N level. This inconsistency in variance among N levels indicates a 'red sign' in the data. It, therefore, tells a problem does exist (Draper and Smith, 1981, p. 147). Moreover, at N level 1 there appears to be an unusual observation at -2.72078, which is nearly about 1.4 standard units away from the nearest point at -1.33809 (Fig. 1b). This does not indicate that it is an outlier, since an outlier, according to Draper and Smith (1981, p. 152) is one that is far greater than the rest of the residuals in absolute value and perhaps lies three or four standard deviations or further from the mean of the residuals. Points of this kind, yet warrant further investigation. This might reveal a likely error in recording the data, or the error variable not being normally distributed, or having different variances, or incorrect specification of the model (Dean and Voss, 1999, p.107).

By plotting the overall standardized residuals against the maize cultivar factor in both years (Figs. 2a&b), a relatively more spread of the residuals of three cultivars, especially those of Cultivar 4, compared to those of Cultivar 3 in 2002 (Fig. 2a). This might indicate heterogeneous error variances. In 2003,

initially there is, however, no obvious problematic pattern of the residuals except for the extreme value at  $-2.72078$  associated with Cultivar 2 (Fig. 2b). The pattern appears to take the 'band' shape which indicates a nearly constant variance (Draper and Smith, 1981).

Based on the residuals associated with only subplots (SP) (Almimi *et al.*, 2007), the pattern of the residual plot vs. maize cultivars has become more indicative in both years (Figs. 3a&b). In 2002, the variations in residuals within Cultivars 2 and 4 were more pronounced compared to those within Cultivars 1 and 3 (Fig. 3a). In 2003, a relatively more spread exists within Cultivars 4 (Fig. 3b). Within relatively small plot area that contains few rows, any unusual-induced influence, of the random variation due to collection of factors not included in the model, may exhibit high variability within these particular plot(s). For example, any missing within-row plants at harvesting may particularly cause residuals to inflate, and any uncaredful management practices as well.

Regarding the dependence issue, the residual spatial pattern, represented here by spatial arrangement (block factor), did not seem to exhibit apparent problems (Figs. 4a&b) in both years. Except for the extreme point at  $-2.72078$  in 2003 that lies within Block 2 (Fig. 4b). It is the very same point associated with both N Level 1 (Fig. 1b) and Cultivar 2 (Fig. 2b), as we mentioned earlier. This relatively extreme point is not too far from the rest of the points in standard deviation units to be considered an outlier as Draper and Smith (1981) have defined the term 'outlier'. This point does not, however, fit in with the remaining ones.

Experiment wise, it is not unlikely to obtain one or more cases of observation values that fail to conform to the rest of the observed data. Weisberg (1980, p. 113) has stated several possibilities. First, an improbable but perfectly compliant observation was made. This means that the observed value satisfies the fitted linear model, but the associated  $e_i$  happens to be large. A case like this should not be right away discarded from the estimation process. Second, following careful checking of the conditions and circumstances under which the data were collected, the case was found to correspond to exceptional yet explainable circumstances, such as failure of a measuring scale used in recording data in the field, or switching between technicians. In this case, one is probably justified in eliminating the case from the data set and estimating the model without it.

Third, an exceptional event happened as in the second case, but no specific reason for it can be detected. Here again, one would probably delete the case from the analysis. Fourth, the case is perfectly satisfactory since neither exceptional nor improbable event had occurred. However, the model was inappropriate choice. This extreme point may be the most important in a study, as it could represent new and unexpected information. The researcher may wish to study the conditions of this point separately. A linear model may still be appropriate for the rest of the data. Draper and Smith (1981) have called such point "*influential*".

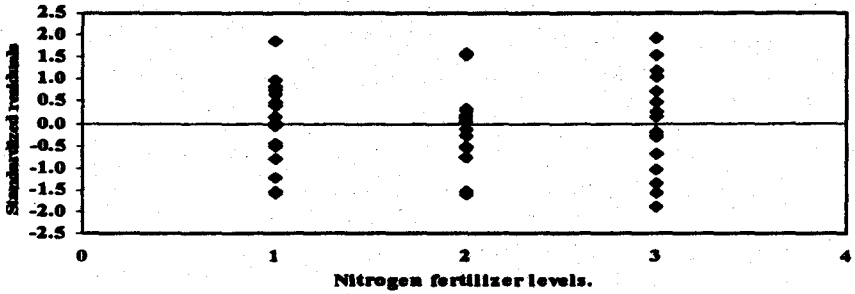


Fig. 1a. Seed yield Standardized residuals related to N fertilizer levels in 2002.

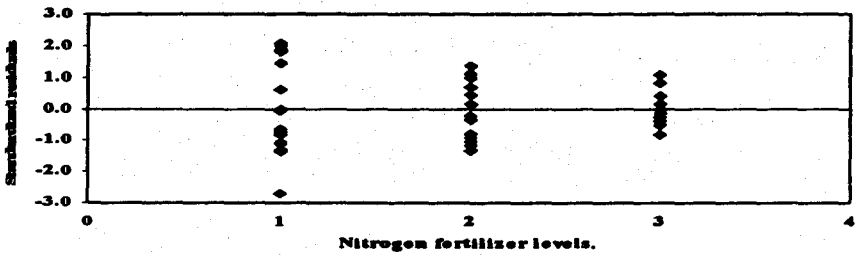


Fig. 1b. Seed yield Standardized residuals related to N fertilizer levels in 2003.

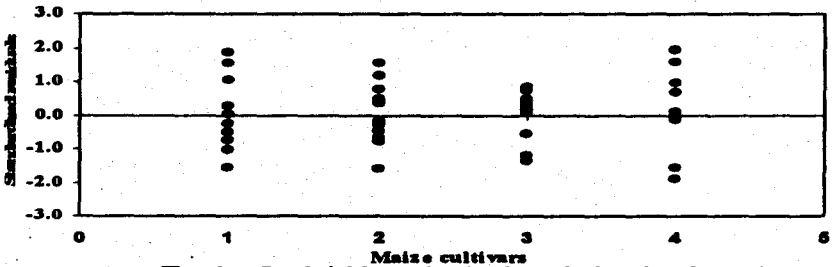


Fig. 2a. Seed yield standardized residuals related to maize cultivars in 2002.

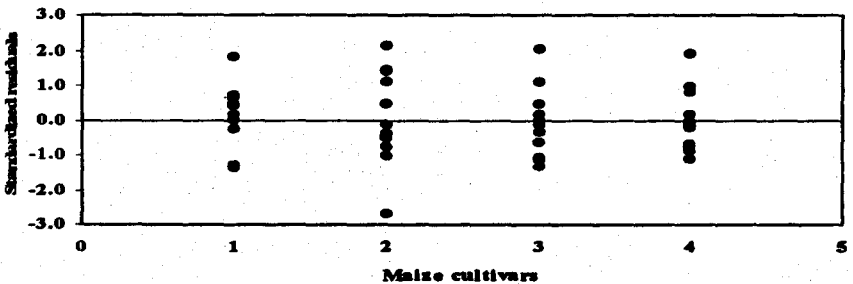


Fig. 2b. Seed yield standardized residuals related to maize cultivars in 2003.



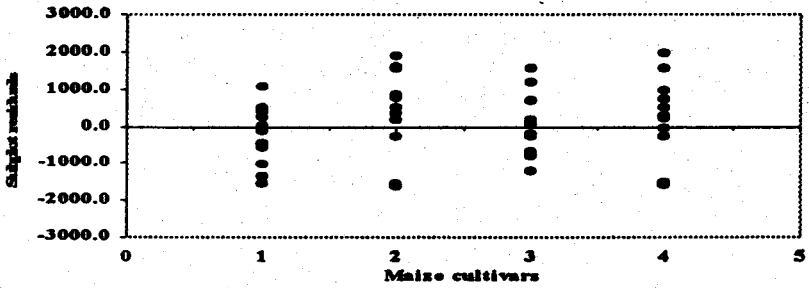


Fig. 3a. Subplot residuals related to maize cultivars in 2002.

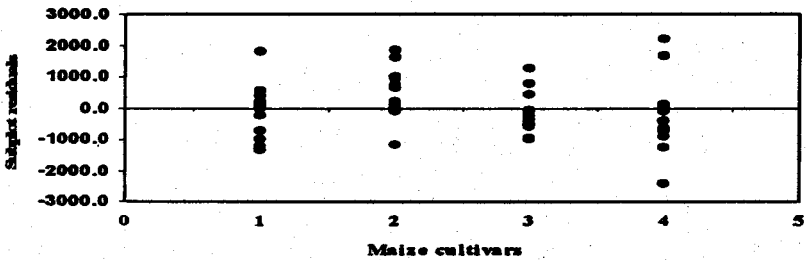


Fig. 3b. Subplot residuals related to maize cultivars in 2003.

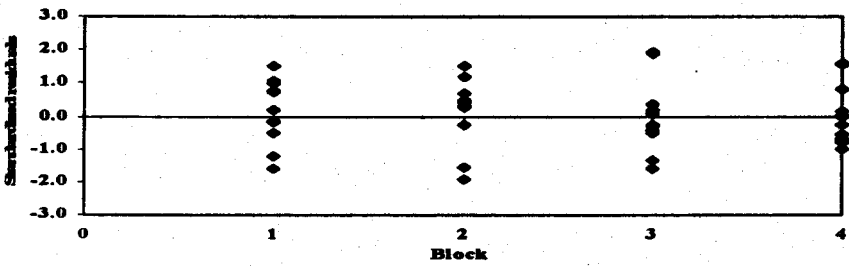


Fig. 4a. Seed yield standardized residuals in different blocks in 2002.

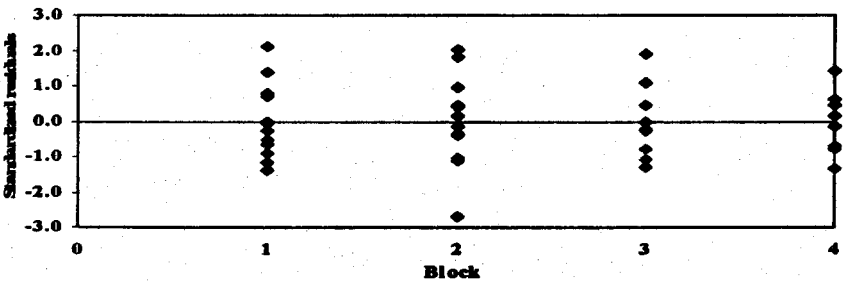


Fig. 4b. Seed yield standardized residuals in different blocks in 2003.

Another perspective of model adequacy of fit is achieved by calculating  $R^2$  values for the full model as well as for both WP and SP sub designs in 2002 (Tables 1&2) and in 2003 (Tables 3&4). The full-model small  $R^2$  values in 2002 (Table 1) has relatively improved in 2003 (Table 3) indicating that a little proportion of the variability in mean seed yield is explained by the factors included in the ANOVA model regardless of the way these factors had been arranged in split plot layout.

Table (1): ANOVA of 3 nitrogen rates x 4 maize cultivars in a four randomized complete blocks for maize seed yield in 2002.

Source	DF	SS	MS	F	p
Model	20	45994200.24	2299710.012	1.31	0.25
Error	27	47329880.22	1752958.527		
Total	47	93324080.46			

$R^2 = 49.284\%$

C.V.=19.879%

Table (2): ANOVA partitioned by both whole (WP) and subplots (SP) for maize seed yield in 2002.

ANOVA for the WP Sub Design					ANOVA for the SP Sub Design						
Source	DF	SS	MS	F	p	Source	DF	SS	MS	F	p
Model	5	2690483.76	—	—	—	Model	9	15720886.67	—	—	—
Block	3	2229807.54	743269.18	0.42	0.73	Cv	3	7355980.26	2451993.42	1.40	0.26
N	2	460676.22	230388.11	0.05	0.95	N x Cv	6	8364906.39	1394151.06	0.80	0.58
B x N	6	27582829.80	4597138.30	2.62	0.03	Residual	27	47329880.22	1752958.52	—	—
Total	11	30273313.57	2752119.41	—	—	Total	35	63050766.89	1801450.48	—	—

$R^2_{WP} = 8.88\%$

$R^2_{SP} = 24.93\%$

Table (3): ANOVA of 3 nitrogen rates x 4 maize cultivars in a four randomized complete blocks for maize seed yield in 2003.

Source	DF	SS	MS	F	P
Model	20	80766886.00	4038344.30	2.93	0.005
Error	27	37172625.00	1376763.88		
Total	47	117939511.00			

$R^2 = 68.48\%$

C.V.=15.30%

Table (4): ANOVA partitioned by both whole (WP) and subplots (SP) for maize seed yield in 2003.

ANOVA for the WP Sub Design					ANOVA for the SP Sub Design						
Source	DF	SS	MS	F	p	Source	DF	SS	MS	F	p
Model	5	11150508.57	—	—	—	Model	9	52347499.12	—	—	—
Block	3	5772911.47	1924303.82	1.40	0.26	Cv	3	27426044.42	9142014.80	6.64	0.001
N	2	5377597.09	2688798.54	0.93	0.44	N x Cv	6	24921454.70	4153575.78	3.02	0.02
B x N	6	17268878.31	2878146.38	2.09	0.08	Residual	27	37172625.00	1376763.88	—	—
Total	11	28419386.88	2583580.62	—	—	Total	35	89520124.12	2557717.83	—	—

$R^2_{WP} = 39.23\%$

$R^2_{SP} = 58.47\%$

For the WP sub designs in both years (Tables 2&4), the extremely low  $R^2$  values assure that a small proportion of the variability in the data is contributed by the WP nitrogen fertilizer factor included in the model, whereas these values were relatively higher for SP sub design. The latter high values indicate that a larger proportion of the variability in the mean seed yield was explained by the SP effects: Cv and N x Cv in the ANOVA model. Neither effect was significant in 2002; yet both were significant in 2003 ( $p < 0.05$ ). Almimi *et al.* (2007) indicated that the low values for the WP sub design do not necessarily mean that the WP plot factor should be replaced with another factor. This is due to —they argued— the presence of any significant SP interaction effects which include the WP factor since these effects contribute to explaining the variability in the data. In the present study, the 2002 N x Cv interaction effect was highly non effective ( $p=0.58$ ), in 2003, however, it was quite significant ( $p=0.02$ ). Based upon Almimi *et al.*'s (2007) argument the N factor should not be replaced, but this warrants more attention be paid when applying nitrogen fertilizer rates to main plots.

In the above ANOVA tables, nearly all non significant effects had extremely small ( $<1.0$ ) F ratios. Each of block, N, and N x Cv effects in 2002, and N effect in 2003 had F-ratios less than one (Tables 2&4). All of the F-ratios in fixed-effect linear models are constructed as the mean square (MS) for the effect of interest divided by an estimate of the error variance. If the null hypothesis is true and all assumptions are satisfied then the F-ratio is expected to  $\cong 1.0$ . If the null hypothesis is false and all assumptions satisfied then the mean square for the effect of interest contains both an estimate of the error variance and a sum of squared terms attributable to the effect of interest (Meek *et al.*, 2007). They pointed out that small ( $<1$ ) F-ratios should not be overlooked since they may indicate a miss-specified model, or in other words, lack of fit, and violation of the normality or homogeneity of variance assumptions.

If a factor should have initially been included in the model, the sums of the squares (SS) for this factor and its interaction with the other factors already in the model will be included in the error term. If the missing factor(s) has/have significant effect(s) then the error mean square can be greatly inflated, resulting in significantly small F-ratios (Meek *et al.*, 2007). In addition, any covariate factors that may have been omitted from the model may conceal differences among factor levels so as to inflate error variance leading also to small F-ratios. In case of crops that depend on individual plants per unit area as in maize, any fluctuation in plant population at harvest is one of the most crucial covariate factors that should be adjusted for in the model had it been a problem. Away from results of the same model fit to historical data under similar conditions, any omitted factor(s) are practically hard to find out in real experimental situations unless statistical analyses are carried out, and reveal suspicions of model fit adequacy.

Another point worth considering in the above ANOVA tables is the little  $p$  values—= 0.03 in 2002 and =0.08 in 2003— of the error mean square (Block x N) interaction effect of the WP part in both years (Tables 2& 4). In addition to the

usual assumptions underlying the ANOVA procedures, Meek *et al.* (2007) hold that the assumption of no interaction is necessary for constructing the test statistics for evaluating both WP and block effects when the effects are fixed.

According to Eq. 1, the mean square associated with WP factor, MSA, has an expected mean square  $\sigma^2_e + b\sigma^2_\gamma + rb(a-1)^{-1} \sum_i \alpha_i^2$  while the term used for the mean square 'error', MSAR, has an expected value of  $\sigma^2_e + b\sigma^2_\gamma$  only if the interaction term is zero. On the other hand, if the assumption of no interaction is violated the WP sub model is given by:  $\mu + \rho_k + \alpha_i + (\alpha\rho)_{ik} + \gamma_{ik}$ . The expected value for the MS(AR) turns out to be:  $\sigma^2_e + (r-1)(a-1)^{-1} \sum_{i,k} (\alpha\gamma)_{ik}$ .

If error (a), i.e. the Block x WP factor interaction in a split plot design turns out to be significant, this means that an interaction likely exists since there are remarkable differences among the whole plots within the same block. Unfortunately, this interaction can not be estimated from a split plot design in which the whole-plot factor laid out in randomized block design; yet, it can be evaluated if it had been arranged in a generalized complete block design (GCBD). In the latter, there are repeat runs for each level of WP factor within each block. In designing experiments—especially in case of suspicion of the presence of such kind of interaction—it is advisable to allow for repeat runs, so model lack of fit can be estimated by estimating both pure error and lack of fit components (Draper and Smith, 1981).

Following checking for the dependence problem, diagnosis for error  $\varepsilon_{ijk}$  unequal variances is applied to data as Dean and Voss (1999) suggested. The most common pattern of nonconstant variances as the mean response increases. This situation is presented when the plot of the standardized residuals versus the fitted values resembles a megaphone in shape. In Fig. 5a, the plots resulted in a 'band' shape which implies the absence of any apparent heterogeneity problem (Weisberg, 1981) in the 2002 data. In 2003 data, by including the extreme point of -2.72078 standard units the band-like shape occurs; however, excluding this point from the data has changed the plot shape to 'parabolic' one (Fig. 5b). This latter unsatisfactory residual behavior would indicate that model was inadequate and needs for extra terms in the model (e.g., square or cross-product terms), or needs for a transformation on the observations of the dependent variable before analysis (Draper and Smith, 1980, p. 147). On the other hand, the presence of this extreme point in the data indicates how this point seems very influential in the current data set and discarding it without careful scrutiny of the circumstances that had entailed its occurrence would certainly lead to loss of important information.

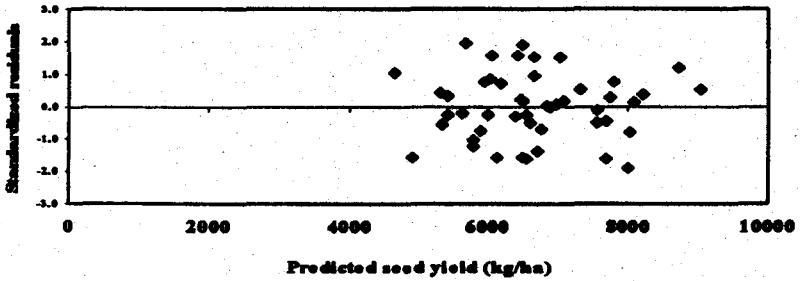


Fig. 5a. Seed yield standardized residuals against predicted seed yield in 2002.

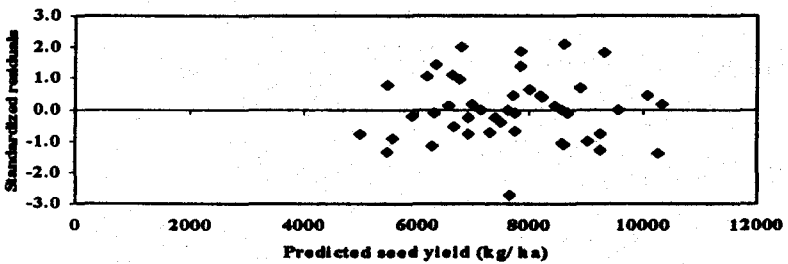


Fig. 5b. Seed yield standardized residuals against predicted seed yield in 2003.

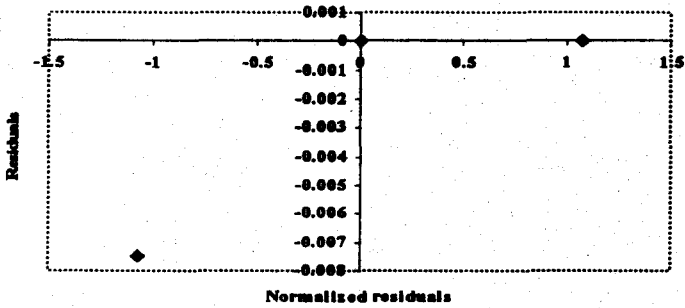


Fig. 6a. Normal probability plot of whole plots in 2002.

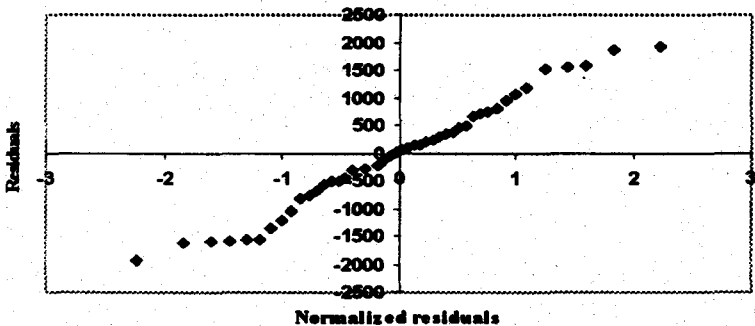


Fig. 6b. Normal probability plot of sub plots in 2002.

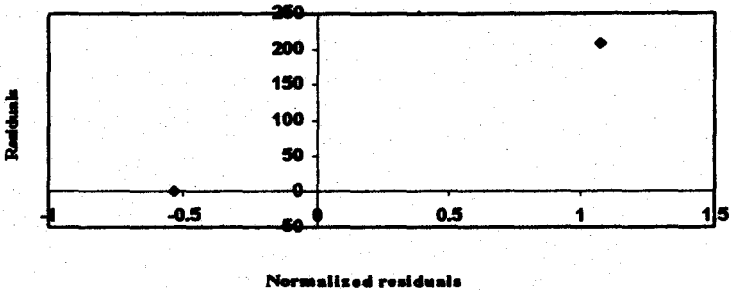


Fig. 7a. Normal probability plot of whole plots in 2003.

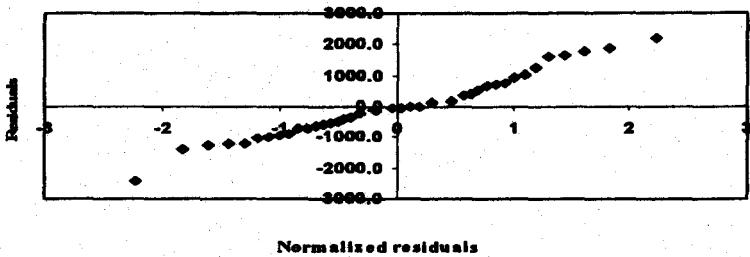


Fig. 7b. Normal probability plot of whole plots in 2003.

Cochran (1947) warned of the possibility of several assumptions to fail simultaneously. In non normal distributions the problem of heterogeneous errors is likely to arise. Normal probability plots of the WP and the SB residuals, for both years, are presented in Figs. 6&7. Split-plot designs most often contain a little number of whole-plots; this certainly causes a difficulty in deriving conclusions concerning normality checking of whole plots. Almimi *et al.* (2007) faced the same problem since they used four levels of the WP factor, and in the present study there are three levels. This difficulty appears by checking Figs. 6a & 7a of main-plot normal probability plots in both years. In both figures, it is not clear enough to interpret the plots as if a non normality problem exists in whole plots especially in Fig. 7a where two WP residuals have the same coordinates. This situation, therefore, recommends that using enough main plots in split-plot designs is urgently needed so that interpreting and concluding from such figures becomes easier for readers. For the SP residuals, on the other hand, normality appears satisfactory in both years of the study (Figs. 6b&7b). Almost all points fit to a straight line (Draper and Smith, 1981; and Dean and Voss, 1999) in both figures.

## CONCLUSION

Considering that ANOVA assumptions are always robust to any violations is a strategy of data analysts that should be abandoned, and checking for any violations should become a required step whether data suspicion arises or

not. Checking for these assumptions of a split-plot design needs calculating both whole- and sub-plot residuals since there exist two error terms in the model. To check for whole-plot normality, a difficulty with this technique often comes out as a result of the limited number of the whole-plot factor levels. This limitation does not give too many residual points to effectively interpret the whole-plot residual behavior during checking the normality assumption. Therefore, researcher should adopt using as enough whole-plot levels as possible. Small  $F$ -ratios ( $<1$ ), in addition, usually lead to failing to reject the null hypothesis and generally are not investigated any further. These values may, however, indicate a miss specification of the model, and/or violations of normality or homogeneity of variance assumptions.

### REFERENCES

- Almimi, A.A.; Kulahci, M. and Montgomery, D.C. (2007): Checking the adequacy of fit of models from split-plot designs. Available online at: <http://ie.fulton.asu.edu/research/workingpaper/pdf/ashrafalmimiarticle3.pdf>
- Bartlett, M.S. (1947): The use of transformations. *Biometrics* 3(1):39-52.
- Bathke, A. (2004): The ANOVA  $F$  test can still be used in some balanced designs with unequal variances and non normal data. *Journal of Statistical Planning and Inference* 126:413-422.
- Chiarotti, F. (2004): Detecting assumption violations in mixed-model analysis of variance. *Ann Ist super Sanità* 40(2):165-171.
- Cochran, W.G. (1947): Some consequences when the assumptions underlying the analysis of variance are not satisfied. *Biometrics* 3(1): 22-38.
- Damon, R.A. and Harvey, W.R. (1987): *Experimental design, ANOVA, and regression*. p. 75. Harper & Row Publishers, Inc., New York, USA.
- Dean, A., and Voss, D. (1999): *Design and analysis of experiments*. Chapter 5. pp 103-133. Springer Texts in Statistics. Springer Verlag, New York Inc.
- Draper, N.R. and Smith, H. (1981): *Applied regression analysis*. 2<sup>nd</sup> ed. John Wiley & Sons. USA.
- Eisenhart, C. (1947): The assumptions underlying the analysis of variance. *Biometrics* 3(1): 1-21.
- Maas, C.M., and Hox, J.J.. (2004): The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis* 46: 427-440.
- Meek, G.E.; Ozgur, C.O. and Dunning, K.A. (2007): Small  $F$ -ratios: red flags in the linear models. *Journal of Data Science* 5:199-215.
- Milliken, G.A. and Johnson, D.E.. (1984): *Analysis of messy data*. Vol. 1: Designed experiments, p.17. Van Nostrand Reinhold Co. New York, USA.
- Neter, J.; Wasserman, W. and Kutner, M. (1985): *Applied linear statistical models*. 2<sup>nd</sup> ed. Chapter 18. pp 602-635. McGraw-Hill.
- Snedecor, G.W., and Cochran, W.G. (1980): *Statistical methods*. p.274. the 7<sup>th</sup> ed. The Iowa State University Press, USA.
- Tageldin, M.H. (2005): Nonparametric analysis of a two-factor ANOVA using rank-based methods. The 11<sup>th</sup> Agronomy Conf., Assiut Univ., 15-16 November, Egypt.
- Weisberg, S. (1980): *Applied linear regression*. p. 119. John Wiley & Sons. USA.

Zimmerman, D.W. (2005): Two separate effects of variance heterogeneity on the validity and power of significance tests of location. Statistical Methodology: available online at: <http://sciencedirect.com/>

## التحقق من انتهاكات الفروض التي تحكم تحليل التباين

محمد هاني أحمد تاج الدين

قسم المحاصيل، كلية الزراعة بمشهور — جامعة بنها

إن التأكد من صحة فروض تحليل التباين كثيرا ما تهمل من جانب الباحثين والقائمين بتحليل البيانات، حيث يسلم مبدئيا بصحتها. ولقد تناولت مراجع كثيرة ما يترتب على انتهاكات تلك الفروض، وتناولت أخرى بعض الاختبارات الخاصة بتلك الفروض ولكن لتحليل التباين في اتجاه واحد فقط. وبناء على ذلك، يهدف هذا البحث الى استراتيجية لاختبار تلك الفروض الخاصة بتحليل التباين لتصميمات القطع المنشقة مرة واحدة التي تحتوي بدورها على أكثر من خطأ تجريبي بحيث من الممكن أن تمتد لمستويات أخرى من تلك التصميمات الاحصائية. استخدمت لاختبار تلك الفروض تجربة حقلية، لمدة عامين، في تصميم قطع منشقة مرة واحدة وضعت فيها القطع الرئيسية الثلاث في أربع قطاعات كاملة العشوائية، حيث شملت تلك القطع على معدلات التسميد النيتروجيني، بينما شملت القطع المنشقة أربعة أصناف من الذرة الشامية. تتكون استراتيجية اختبار أية انتهاكات لفروض تحليل التباين خمس خطوات لاختبار: مطابقة النموذج الرياضي للبيانات المتاحة، المشاهدات المتطرفة، وكل من استقلالية وتساوي تباينات الأخطاء التجريبية، وأخيرا التوزيع الطبيعي لتلك الأخطاء. تعتمد تلك الخطوات في الأساس على حساب قيم الأخطاء المتبقية لكل من القطع الرئيسية والفرعية. كم تم حساب قيم معامل التحديد لكل تلك القطع كوسيلة لاختبار حسن مطابقة النموذج الرياضي للبيانات. وقد بين التحليل بوجه عام بعض انتهاكات لتلك الفروض الخاصة بتحليل التباين. فقد بين الشكل العام لمنحنى الأثر المتبقي مقابل عامل السماد النيتروجيني أن هناك عدم ثبات في قيم تباينات الأخطاء التجريبية. كما كان هناك قيمة متطرفة بلغت قيمتها 2.72078- وحدة عيارية. أما بخصوص فروض الاستقلالية و تساوي التباين والتوزيع الطبيعي، فلم يتبين أن هناك أية انتهاكات في بيانات العاميين. أما فيما يخص قيم معامل التحديد لكل من القطع الرئيسية والفرعية فقد كانت منخفضة بصورة ملحوظة في العام الأول حيث بلغت ٩% فقط للقطع الرئيسية و حوالي ٢٥% للقطع الفرعية، وتحسنت تلك القيم نسبيا في العام الثاني لتبلغ ٣٩% و ٥٨%، على التوالي. تلك القيم المنخفضة كثيرا بالنسبة لعامل السماد النيتروجيني في القطع الرئيسية تدل على المساهمة الضعيفة لتلك العامل تحت ظروف تلك التجربة في المتوسط الحسابي للعامل التابع ألا وهو محصول الفدان من الذرة الشامية. ومما يسترعي الانتباه في جداول تحليل التباين في العاميين بشكل عام، أن غالبية التأثيرات الغير معنوية كانت قيم 'ف' لها دون الواحد الصحيح مما يبعث إنذارا لاحتمال وجود خلل ما في حسن مطابقة النموذج الرياضي، وكذلك قد ينبى عن انتهاكات لفرض أو أكثر من فروض تحليل التباين.