

Assessing The Adequacy Of Split-Plot Design Models

David, I. J., Asiribo, O. E., Dikko, H. G.

Abstract: This paper assesses the adequacy of model fit of the split-plot design models that is the whole plot (WP) sub design model with WP error and the split-plot (SP) sub design model with SP error using the four measures of adequacy of fit for the WP and SP sub designs proposed by Almimi *et al.* [3] which are the R^2 , R^2 -Adjusted, Prediction Error Sum of Square (PRESS) and R^2 -Prediction statistics and we proposed five other methods which are the Modeling Efficiency (MEF), Resistant Coefficient of Determination (r_r^2), Mean Square Error (MSE), Mean Square Error Prediction (MSEP) and Adequate Precision (AP) statistics. A $2^{(1+3)}$ replicated two-level SP design and $3^1 \times 4^2$ replicated mixed level SP design were used for computing the measures of model adequacy of fit for each WP and SP sub design models. These measures describe the predictive performance of each WP and SP sub design models.

Key Words: AP, MEF, MSE, MSEP, Model Adequacy, PRESS, R^2 , R^2 -Adjusted, R^2 -Prediction, r_r^2 , SP error, WP error

1 INTRODUCTION

Experiments are performed by investigators in virtually all fields of enquiry, usually to discover something about a particular process or system. Literarily, an experiment is a test. A designed experiment is a test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reasons for changes in the output responses [15]. The split-plot experimental design has a model with two experimental errors and the adequacy of the model can be diagnosed for its fitness, using methods such as the coefficient of determination (R^2), adjusted coefficient of determination (R^2 - Adjusted), prediction error sum of squares (PRESS), mean bias, graphical approach, etc. The manner in which the measures of R^2 , R^2 - Adjusted, PRESS, etc are obtained from a design with one type of error in its model is quite different from that of the split-plot design with two types of errors in its model, which are whole plot or main plot error and the split-plot or sub-plot error. The calculation of these measures should be for both the main plot error and the split-plot error due to the assumption that the whole-plot errors are independent and identically distributed as $N(0, \sigma_w^2)$, and the split-plot errors are independent and identically distributed as $N(0, \sigma^2)$, which are mutually independent. These are, guaranteed by the system of randomization in split-plot experimental design [3]. Almimi *et al.* [3] stated that although R^2 is a useful measure of goodness of fit, it has some limitations [15]: R^2 can increase by adding terms to the model, the magnitude of R^2 depends on the range of variability in the regressor variable, and R^2 does not measure the magnitude of the slope of the regression line.

Nonetheless, the use of the least-square method to derive a linear regression of observed on model-predicted values for model evaluation has little interest since the predicted value is useless in evaluating the mathematical model; therefore, the R^2 is irrelevant since one does not intend to make predictions from the fitted line [13]. Then R^2 is weakly increasing in the number of regressors in the model as such, R^2 models with different numbers of independent variables. This conforms to the statement by [3] that the limitations about R^2 mentioned by [15] call for the use of alternative measures. He further stated that the R^2 - adjusted is preferred to R^2 because its value only increases if a variable that is being added to the model reduces the residual mean square. Hence, unlike R^2 , R^2 - adjusted does not increase if irrelevant terms are being added to the model. Masoud and Rahim [11] stated that when outliers occur in a data the assumption of normally distributed errors is violated and reduces the performance of the linear regression models. Further they stated that detecting outliers is much easier than deciding what to do with them because after detecting, the investigator should consider whether the outliers are recording errors. They advised that after these cases are detected they should never be blindly deleted and the investigator should analyze the full data set including the outliers as well as the data set after the outliers have been removed (either by deleting the cases or the variables that contain the outliers). Kvalseth [8] proposed the use of resistant coefficient of determination (r_r^2), which uses the medians instead of the means and results in a coefficient that is highly resistant to outliers or extreme data points. Almimi *et al.* [3] stated that the prediction error sum of squares, PRESS and R^2 - prediction which are useful measures of model adequacy were proposed by [1], [2], is used to evaluate the performance of the model in predicting the responses in new and future experiments. Moreover, PRESS is very useful when comparing models because a model with a small value of PRESS is generally better than a model with a large value of PRESS. Similarly, R^2 -Prediction, which is based on PRESS, is a measure of the predictive performance of a model. Tedeschi [17] presented the modelling efficiency statistic (MEF), in which he stated that the method is similar to ρ , which is interpreted as the proportion of variation explained by the fitted line whereas the MEF statistic is the proportion of variation explained by the line $Y = f(X_1, \dots, X_p)$. He further stated that the statistic has been extensively used in hydrology models, but can certainly be used in biological models. He concluded that in a perfect fit, both the ρ (correlation coefficient) and the

- David, I. J. is currently a PhD student of Statistics in Department of Mathematics, Ahmadu Bello University, Zaria-Nigeria. PH- +2348060581391. Email: ikwuoches8@gmail.com
- Asiribo, O. E. is a Professor of Statistics, Department of Mathematics, Ahmadu Bello University, Zaria-Nigeria. PH- +8033526707.
- Dikko, H. G. holds a PhD in Statistics and is a Lecturer in the Department of Mathematics, Ahmadu Bello University, Zaria-Nigeria. PH- +2348083544070. Email: hgdikko@yahoo.com

MEF would result in a value equal to one and that the upper bound of MEF is one and the (theoretical) lower bound of MEF is negative infinity. Loague and Green [10] stated that if MEF is lower than zero; the model predicted values are worse than the observed mean. The MEF statistic may be used as a good indicator of goodness of fit [12]. Tedeschi [17] stated that the mean square error (MSE) and the mean square error of prediction (MSEP) are probably the most common and reliable estimate to assess the precision and measure the predictive accuracy of a model respectively. Tedeschi [17] stated that the MSE assesses the precision of the fitted linear regression using the difference between observed values (Y_i) and regression-predicted values (\hat{Y}_i). While MSEP consist of the difference between observed values (Y_i) and model – predicted values ($f(X_1, \dots, X_p)_i$) rather than regression-predicted value. Sergio and Joan [16] stated that like variance, MSE has the disadvantage of heavily weighting outliers. They said it is because of the squaring of each term, which effectively weights large errors more heavily than small ones. This property, undesirable in many applications, has led researchers to use alternatives such as the mean absolute error, or those based on the median. Some drawbacks of the MSE and MSEP analysis are notable. Mitchell and Sheehy [14] stated that the MSEP (or its root) removes the negative sign and weights the deviation by their squares, thus giving more influence to larger data points, and does not provide any information about model precision. Tedeschi [17] stated that the MSEP estimate is a reliable measure of model accuracy. Nonetheless, the reliability will decrease as n decreases. Therefore, the estimate of MSEP variance has an important role. Tedeschi [17] further stated that a different scenario arises when the model parameters are adjusted to the data set; the real MSEP will be underestimated because the model will reproduce more closely the data that have been modeled than it would for the entire population of interest. A simple and easy approach to check model adequacy between two (or more) models is computing the MSEP of each model and choosing the one with smaller MSEP estimates. Wallach and Goffinet (1989) presented a different approach that makes use of the difference in MSEP (Δ MSEP) values between two models. They stated that if the estimated difference (Δ MSEP) is positive then model $g(Z_1, \dots, Z_k)_i$ is preferred, while if Δ MSEP is negative then model $f(X_1, \dots, X_p)_i$ is preferred since it has the smaller MSEP estimate.

$$\Delta \text{MSEP} = \frac{\sum_{i=1}^n \left[\left(Y_i - f(X_1, \dots, X_p)_i \right)^2 - \left(Y_i - g(Z_1, \dots, Z_k)_i \right)^2 \right]}{n}$$

Montgomery [15] presented and discussed the adequate precision (AP) statistic which is computed by dividing the difference between the maximum predicted (max \hat{Y}_i) response and the minimum predicted (min \hat{Y}_i) response by the average standard deviation of all predicted responses. He stated that large values of this quantity are desirable and values that exceed four ($AP > 4$) usually indicate that the model will give reasonable performance in prediction. This paper discussed potential weakness in some of the methods used in checking adequacy of models that uses the mean which is not resistant to outliers and their strength if the median is used instead of the mean.

2 METHODOLOGY

Nine methods and their procedures to compute model adequacy of fit from a split-plot design as discussed in the literature are explored. Four out of the nine methods were used, being applied by Almimi *et al.* [3] in both the replicated and unreplicated split-plot design. However, we introduced the application of the remaining four out of which we proposed the use of difference in mean square error (Δ MSE) values between two models that is the whole plot and split-plot models and the method proposed by Wallach and Goffinet [18]: difference in mean square error prediction (Δ MSEP). In this research paper, we only considered the replicated split-plot design.

2.1 Procedure for Computing Model Adequacy of Fit from Split-Plot Design

Detailed procedures for computing the measures of model adequacy of fit from split-plot design for R^2 , R^2 – Adjusted, PRESS and R^2 -Prediction are been given by Almimi *et al.* [3]. However, we presented procedures for the other introduced measures for model adequacy of fit from split-plot design. The introduced measures are modeling efficiency statistic (MEF), resistant coefficient of determination (R_r^2), difference in mean square error (Δ MSE), difference in mean square error prediction (Δ MSEP) and adequate precision (AP).

2.1.1 MEF (Modeling Efficiency) Statistic

1. From the WP sub-design, subtract the model-predicted values ($f(X_1, \dots, X_p)_i$) from the observed values (Y_i), take the summation and square. Divide the value obtained by the total sum of squares (SS_{Total}) for the WP sub-design and subtract the value obtained from one to obtain the MEF for the WP sub-design model.

$$MEF_{WP} = 1 - \frac{\sum_{i=1}^n \left(Y_i - f(X_1, \dots, X_p)_i \right)_{(WP)}^2}{\sum_{i=1}^n \left(Y_i - \bar{Y} \right)_{(WP)}^2} \quad (1)$$

2. Use the ANOVA column for the SP sub-design, subtract the model-predicted values ($f(X_1, \dots, X_p)_i$) from the observed values (Y_i), take the summation and square. Divide the value obtained by the total sum of squares (SS_{Total}) and subtract the value obtained from one to obtain the MEF for the SP sub-design model.

$$MEF_{SP} = 1 - \frac{\sum_{i=1}^n \left(Y_i - f(X_1, \dots, X_p)_i \right)_{(SP)}^2}{\sum_{i=1}^n \left(Y_i - \bar{Y} \right)_{(SP)}^2} \quad (2)$$

2.1.2 R_r^2 (Resistant Coefficient of Determination)

1. For the WP sub-design, the R_r^2 is calculated using the absolute WP residual ($|e_{i(wp)}|$) and obtain its median ($Med\{e_{i(wp)}\}$).

- Obtain the absolute difference of the mean value of the WP observed values from the WP observed values ($|Y_{i(wp)} - \bar{Y}_{(wp)}|$) and obtain its median

$$(\text{Med}|Y_{i(wp)} - \bar{Y}_{(wp)}|).$$

- Divide ($\text{Med}|e_{i(wp)}|$) by ($\text{Med}|Y_{i(wp)} - \bar{Y}_{(wp)}|$) square the result obtained and subtract from 1 to obtain r_r^2 .

$$r_{r(wp)}^2 = 1 - \left(\frac{\sum_{i=1}^n (|e_i|)_{WP}}{\sum_{i=1}^n (|Y_i - \bar{Y}|)_{WP}} \right)^2 \quad (3)$$

- For the SP sub-design, the r_r^2 is calculated using the absolute SP residual ($|e_{i(sp)}|$) and obtain its median ($\text{Med}|e_{i(sp)}|$).
- Obtain the absolute difference of the mean value of the SP observed values from the SP observed values ($|Y_{i(sp)} - \bar{Y}_{(sp)}|$) and obtain its median ($\text{Med}|Y_{i(sp)} - \bar{Y}_{(sp)}|$).
- Divide ($\text{Med}|e_{i(sp)}|$) by ($\text{Med}|Y_{i(sp)} - \bar{Y}_{(sp)}|$) square the result obtained and subtract from 1 to obtain r_r^2 .

$$r_{r(sp)}^2 = 1 - \left(\frac{\sum_{i=1}^n (|e_i|)_{SP}}{\sum_{i=1}^n (|Y_i - \bar{Y}|)_{SP}} \right)^2 \quad (4)$$

2.1.3 Δ MSE (Difference in Mean Square Error) and Δ MSEP (Difference in Mean Square Error Prediction)

- Use the ANOVA column of the WP sub-design with the complete whole plot factors (i.e. significant and insignificant factors) to obtain the residual sum of square ($SS_{Residual}$) and divide by (n-2) to obtain the MSE for WP as follows:

$$MSE_{WP} = \frac{SS_{Residual(WP)}}{(n-2)_{(WP)}} \quad (5)$$

- Use the ANOVA column of the SP sub-design with the complete split-plot factors (i.e. significant and insignificant factors) to obtain the residual sum of square ($SS_{Residual}$) and divide by (n-2) to obtain the MSE for SP as follows:

$$MSE_{SP} = \frac{SS_{Residual(SP)}}{(n-2)_{(SP)}} \quad (6)$$

- Obtain the difference between the MSE_{WP} and MSE_{SP} as follows:

$$\Delta MSE_{(wp, sp)} = MSE_{wp} - MSE_{sp} \quad (7)$$

- Use the ANOVA column of the WP sub-design to obtain the residual sum of square ($SS_{Residual}$) and divide by (n) to obtain the MSEP for WP as follows:

$$MSEP_{WP} = \frac{SS_{Residual(WP)}}{(n)_{(WP)}} \quad (8)$$

- Use the ANOVA column of the SP sub-design to obtain the residual sum of square ($SS_{Residual}$) and divide by (n) to obtain the MSE for SP as follows:

$$MSEP_{SP} = \frac{SS_{Residual(SP)}}{(n)_{(SP)}} \quad (9)$$

- Obtain the difference between the $MSEP_{WP}$ and $MSEP_{SP}$ as follows:

$$\Delta MSEP_{(wp, sp)} = MSEP_{wp} - MSEP_{sp} \quad (10)$$

2.1.4 Adequate Precision (AP)

- Obtain the fitted values corresponding to the whole plot sub model.
- Select the maximum predicted response ($\text{Max } \hat{Y}$) and minimum predicted response ($\text{Mini } \hat{Y}$) corresponding to the whole plot.
- Obtain their difference ($\text{Max } \hat{Y} - \text{Mini } \hat{Y}$).
- Divide the estimated difference ($\text{Max } \hat{Y} - \text{Mini } \hat{Y}$) by the average standard deviation of all predicted response corresponding to the whole plot

$$AP_{wp} = \frac{(\text{Max } \hat{Y} - \text{Mini } \hat{Y})_{WP}}{(S_{\hat{Y}} / \sqrt{n})_{WP}} \quad (11)$$

- Obtain the fitted values corresponding to the split-plot sub model.
- Select the maximum predicted response ($\text{Max } \hat{Y}$) and minimum predicted response ($\text{Mini } \hat{Y}$) corresponding to the split plot.
- Obtain their difference ($\text{Max } \hat{Y} - \text{Mini } \hat{Y}$).
- Divide the estimated difference ($\text{Max } \hat{Y} - \text{Mini } \hat{Y}$) by the average standard deviation of all predicted response corresponding to the split plot

$$AP_{sp} = \frac{(\text{Max } \hat{Y} - \text{Mini } \hat{Y})_{SP}}{(S_{\hat{Y}} / \sqrt{n})_{SP}} \quad (12)$$

2.1.5 Whole Plot and Split-Plot Residuals Calculation

Almimi *et al.* [3] presented the procedure in calculating the whole plot (WP) and split-plot (SP) residuals, the steps are been given as follows:

- Once a model is fitted and a regression formula is obtained for a split-plot design, use this formula to calculate the fitted values.

2. Subtract the fitted values from the response values to obtain the residuals.
3. To calculate the WP residuals, average the residuals that correspond to the runs in each WP. Now, all the runs within a WP have the same WP residual.
4. To calculate the SP residuals, subtract the WP residuals obtained in step 3 from the residuals obtained in step 2 for all the runs in the design.

3.0 ANALYSIS AND DISCUSSION OF RESULTS

The data used for illustrations for the various methods discussed in this paper are of two sets where the first set is one of the data used by Almimi *et al.* [3] which is a $2^{(1+3)}$ replicated two-level split-plot design data. The data consist of four factors each at two levels, the WP sub design is made of one factor (Z) and the remaining three factors (A, B and C) are for the SP sub design while the second set of data is a $3^1 \times 4^2$ replicated mixed level split-plot design data on the effect of four nitrogen fertilizer levels on the yield of four varieties of rice at three irrigation levels.

3.1 Presentation and Interpretation of Results for the $2^{(1+3)}$ Replicated Two-Level Split-Plot Design

After we applied all the procedures for the reviewed methods and proposed methods for checking model fitness from a $2^{(1+3)}$ replicated two-level split-plot design. Below are the results.

Table 1: ANOVA for the $2^{(1+3)}$ Replicated Two-Level Split-Plot Design

ANOVA for the WP Sub Design					
Source	DF	SS	MS	F	P-Value
Model	1	59.13	59.13	2.94	0.228
Z	1	59.13	59.13	2.94	0.228
Error	2	40.17	20.08		
Total	3	99.30			
ANOVA for the SP Sub Design					
Source	DF	SS	MS	F	P-value
Model	5	2136.94	427.388	155.8	0.0001
A	1	597.72	597.72	217.92	0.0001
B	1	1226.36	1226.36	447.09	0.0001
ZA	1	14.72	14.72	5.37	0.030
ZB	1	285.01	285.01	103.90	0.0001
AB	1	13.13	13.13	4.79	0.039
Error	23	63.08	2.743		
Total	28	2200.02			

Source: Author's computation

Table 2: Summary of the Results for the Four Measures of Model Adequacy of Fit for the $2^{(1+3)}$ Replicated Two-Level Split-Plot Design by Almimi *et al.* [3]

Whole Plot Sub Design		Subplot Sub Design	
R^2_{WP}	0.595	R^2_{SP}	0.971
R^2_{WP} -Adjusted	0.393	R^2_{SP} -Adjusted	0.965
PRESS _{WP}	40.293	PRESS _{SP}	63.866
R^2_{WP} Predicted	0.594	R^2_{SP} Predicted	0.971

Source: Author's computation

Table 3: Summary of the Results for the Four Proposed Measures of Model Adequacy of Fit for the $2^{(1+3)}$ Replicated Two-Level Split-Plot Design by Almimi *et al.* [3]

Whole Plot Sub Design	Subplot Sub Design		
MEF_{WP}	-21.16	MEF_{SP}	0.9531
$r^2_{r(wp)}$	0.3261	$r^2_{r(sp)}$	0.9537
AP _{WP}	11.135	AP _{SP}	19.054
$\Delta MSE_{(wp, sp)} = 17.347$		$\Delta MSEP_{(wp, sp)} = 65.5239$	

Source: Author's computation

Table 1 above present the ANOVA table for the WP sub design significant effects and the SP sub design significant effects while table 2 present results for the four measures of model adequacy of fit from a $2^{(1+3)}$ replicated two level split-plot design presented by [3] and table 3 present the results for the four introduced measures for this paper for model adequacy of fit from a $2^{(1+3)}$ replicated two level split-plot design presented by [3]. The measures proposed by [3] and those we proposed provide some notable difference though conclusion is the same. The R^2 , adj- R^2 , and pred- R^2 values for the WP and SP sub design model are (0.595 and 0.971), (0.393 and 0.965) and (0.594 and 0.971) respectively. The r_r^2 and MEF values for the WP and SP sub design model are (0.3261 and 0.9537) and (-21.6 and 0.9531) respectively. We observed that the R^2 and pred- R^2 values for the WP sub design model is about 60% but the adj- R^2 value for the WP sub design model is 39% comparing both values the difference is large and this is an indication that the R^2 and pred- R^2 values could be misleading but for the SP sub design model the difference between the R^2 and pred- R^2 value of 97% and adj- R^2 value of 96.5% is small, this shows that for the SP sub design model R^2 and pred- R^2 are reliable as the adj- R^2 . Comparing this values with the r_r^2 and MEF values we observed that the r_r^2 value for the WP sub design model of 32.6% is even smaller than the adj- R^2 value of 39% this is an indication that the r_r^2 is not affected by outliers in the data so its value is not influenced by it and the R^2 and pred- R^2 are not to be relied upon. Likewise, the negative value of the MEF is a clear indication for worse predictive values by the WP sub design effects because the value is less than zero. The values of r_r^2 (95.37%) and MEF (95.31%) for the SP sub design model are less than the R^2

and $pred-R^2$ values (97%) and $adj-R^2$ value (96.5%) this is another indication that the r_r^2 is not affected by outliers and is more reliable likewise the MEF value is similar to the r_r^2 value. The $PRESS$ value for the WP sub design model is smaller compared to the SP sub design model $PRESS$ value which is large indicating that the WP sub design model fit well in predicting new responses compared to the SP sub design model but this is not in agreement with other measures of fit which indicates that the SP sub design model is adequately fit compared to the WP sub design model. Likewise, the AP values for the WP and SP sub design model are large because they exceed four but the SP sub design model value is larger. The $PRESS$ and AP values can be misleading as that of the R^2 and $pred-R^2$ values because we observed that a small value of $PRESS$ or large value of AP does not really tell that the model fit adequately besides the $PRESS$ is worse because there is no bench mark or cut point to say a $PRESS$ value is small enough or too large to say a model fit adequately or worse while the AP value could be that, it is affected by outliers that is why we observed an AP value greater than four in the WP sub design model when it is been observed by other measures (r_r^2 , $adj-R^2$ and MEF) values indicated that the WP sub design model is not adequately fit.

3.2 Presentation and Interpretation of Results for the 3¹x4² Replicated Mixed Level Split-Plot Design

In this section we applied all the reviewed methods and proposed methods for checking model fitness from a 3¹ x 4² replicated mixed level split-plot design. Below are the results.

Table 4: ANOVA for the 3¹x4² Replicated Mixed Level Split-Plot Design

ANOVA for the WP Sub Design					
Source	Df	SS	MS	F	P-Value
Model	5	860.931	172.19	15.75	
I	2	742.922	371.46	33.99	0.0001
V	3	118.009	39.34	3.6	0.034
Error	18	196.699	10.93		
Total	23	1057.63			
ANOVA for the SP Sub Design					
Source	Df	SS	MS	F	P-value
Model	18	542.68	30.15	8.469	
N	3	198.16	66.05	18.56	0.0001
I*N	6	156.62	26.10	7.34	0.0001
V*N	9	187.90	20.88	5.87	0.0001
Error	54	192.17	3.56		
Total	72	734.85			

Source: Author's computation

Table 5: Summary of the Results for the Four Measures of Model Adequacy of Fit for the 3¹x4² Replicated Mixed Level Split-Plot Design

Whole Plot Sub Design		Subplot Sub Design	
R^2_{WP}	0.8140	R^2_{SP}	0.7385
R^2_{WP} -Adjusted	0.7624	R^2_{SP} -Adjusted	0.6513
$PRESS_{WP}$	114.116	$PRESS_{SP}$	278.583
R^2_{WP} Predicted	0.8921	R^2_{SP} Predicted	0.6209

Source: Author's computation

Table 6: Summary of the Results for the Four Proposed Measures of Model Adequacy of Fit for the 3¹ X 4² Replicated Mixed Level Split-Plot Design

Whole Plot Sub Design	Plot Sub	Subplot Sub Design	
MEF_{WP}	0.9229	MEF_{SP}	0.7307
$r_r^2(wp)$	0.8824	$r_r^2(sp)$	0.8477
AP_{WP}	31.8567	AP_{SP}	41.1480
$\Delta MSE_{(wp, sp)} = 7.37$		$\Delta MSE_{(wp, sp)} = -1.1558$	

Source: Author's computation

Table 4 above present the ANOVA table for the WP sub design significant effects and the SP sub design significant effects while table 5 above present the results for the four measures (used by [3]) of model adequacy of fit from a 3¹ x 4² replicated mixed level split-plot design and table 6 above present the results for the four introduced measures for the research paper of model adequacy of fit from a 3¹ x 4² replicated two level split-plot designs. The measures proposed by [3] and those we proposed provides some notable differences though conclusion is the same. The R^2 , $adj-R^2$, and $pred-R^2$ values for the WP and SP sub design model are (0.8140 and 0.7385), (0.7624 and 0.6513) and (0.8921 and 0.6209) respectively. The r_r^2 and MEF values for the WP and SP sub design model are (0.8824 and 0.8477) and (0.9229 and 0.7307) respectively. We observed that the $pred-R^2$ value for the WP is greater than the R^2 WP and the $pred-R^2$ value for SP sub design model is less than that of the R^2 value for the SP sub design model, these differences could be due to the presence of outliers in the data. The $adj-R^2$ value for the WP sub design model is lower than that of the R^2 and $pred-R^2$ and its SP sub design model value is less than the R^2 value but slightly above the $pred-R^2$ value this shows its ability to withstand unnecessary increment when a factor is removed or added from a model but it is not free from the effect of outliers. Comparing this values with the r_r^2 and MEF values we observed that the r_r^2 value for the WP sub design model of 88.24% is greater than the R^2 and the $adj-R^2$ values but there is no much difference with the WP $pred-R^2$ value while the MEF value for the WP sub design model is greater than all the other measures of model adequacy of fit. Thus, the r_r^2 value for the SP sub design model is higher than all the other measures of model adequacy of fit for the SP sub design

model. There is no much difference in r_r^2 value for both the WP and SP sub design models and base on these values from the r_r^2 we could say that the WP and SP sub design models predict well the observed values because both values explain large variability in the data. This shows that the r_r^2 is not affected by outliers in the data. The PRESS statistic value for the WP sub design model support the claim that the WP is adequate for prediction but, the AP value do not support the claim rather it support the SP sub design model to be more adequately precise in navigating its design space though both AP values for the WP sub design and SP sub design are far greater than four and this could be its inability of withstanding outliers or extreme data point.

4 CONCLUSION

The results obtained from the $2^{(1+3)}$ replicated two-level split-plot design showed that the r_r^2 result is unbiased because its value is not affected by outliers, its SP sub design model value is high but not as high as the R^2 , adj- R^2 and pred- R^2 values which are affected by outliers as sited in the literature. The results from the methods applied on the second data ($3^1 \times 4^2$ replicated mixed level split-plot design) showed that the WP sub design model attained high values of R^2 , adj- R^2 , pred- R^2 , MEF and r_r^2 with negative value of Δ MSEP but the AP and Δ MSE values showed that the SP sub design model have a better predictive Precision and from our observation the r_r^2 value for the WP sub design model (0.8824) is not adequately larger than the SP sub design model value (0.8477). Based on these results we cannot completely say that the WP sub design model is better than the SP sub design model rather they are operating in the same rate. This is an indication that the other measures of model adequacy of model fit could be misleading due to none resistance to outliers, unnecessary increase or reduction of values due to introduction of a variable or variables into the model or removal of a variable or variables from the model. Based on our observations we recommend the following for assessing the adequacy of model fit from SP designs and checking the performance of two or more models either for present or future use.

1. Measures of adequacy of fit of models that uses the mean could be misleading because their magnitudes are not consistently related to the accuracy of prediction, that is, where accuracy is defined as the degree to which model-predicted observations approach the magnitudes of their observed counterparts [19]. So, care should be taking in using measures that uses the mean.
2. We recommend the use of robust objective functions, which are based on the median instead of the mean because the resistant coefficient of determination r_r^2 which uses the median shows that its reliability is adequate due to its ability of withstanding outliers or extreme data points.
3. The Δ MSE and the Δ MSEP values could be questionable because as N samples become smaller their results becomes biased, so, we recommend that large sample size should be used, if sample size is

small then they should not be considered as reliable measures for model adequacy of fit.

4. We recommend also that the Δ MSE is not an adequate measure of model fit for SP designs and should not be relied upon because it is often expected that the WP sub design model will always have the largest MSE compared to the MSE of the SP sub design model only in rare and special cases where the reverse is the case because the SP design usually sacrifices precision in estimating the average effects of the treatments assigned to WP sub design by improving the precision for comparing the average effects of treatments assigned to SP sub design. Hence, we recommend the use of Δ MSE for assessing the precision of regression models.
5. We also recommended that the PRESS and AP statistic could be misleading because they don't have reliabel bench mark or point interval for recommending a best PRESS or AP value. Although, the AP value is desirable or best when its value is > 4 [15] but we observed that we obtained AP values that are adequately > 4 for both the WP sub design and SP sub design models for both data.

Finally, it is our hope that this research paper presents the strength and fault of each methods explored either as a review or introduction to assessing adequacy of model fit from split-plot designs with objective basis for model selection by researchers in the field of Statistics, Mathematics, Econometrics, Agronomy, Biological Sciences, etc.

REFERENCES

- [1]. Allen, D. M. (1971). Mean Square Error of Prediction as a Criterion for Selecting Variables. *Technometrics* 13(3): 469-475.
- [2]. Allen, D. M. (1971). The Relationship between Variable Selection and Data Augmentation and a Method for Prediction. *Technometrics* 16(1): 125-127.
- [3]. Almini, A. A., Kulahci, M. and Montgomery, D. C. (2006). Checking the Adequacy of fit of Models from split-plot Designs. *Journal of Quality Technology* 38: 58 – 190.
- [4]. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag. PP. 60.
- [5]. DeGroot, M. H. (1980). *Probability and Statistics*. 2nd ed. Addison-Wesley.
- [6]. Dent, J. B. and Blackie, M. J. (1979). *Systems Simulation in Agriculture*. Applied Science, London.
- [7]. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*. Wiley-Interscience. ISBN 0-417-17082-8.
- [8]. Kavalseth, T. O. (1985). Cautionary Note about R^2 . *The American Statistician* 39: 279 – 285.
- [9]. Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. 2nd ed. New York: Springer.

MR1639875 (<http://www.ams.org/mathscinet-getitem?mr=1639875>). ISBN 0-387-98502-6.

- [10]. Loague, K. and Green, R. E. (1991). statistical and Graphical Methods for Evaluating Solute Transport Models: Overview and Application. *Journal of Contaminant Hydrology* 7: 51-73.
- [11]. Masoud, Y. and Rahim, M. (2010). The Effect of Outliers on Robust and Resistant Coefficient of Determination in the Linear Regression Models. *International Journal of Academic Research* 2(3): 133-138.
- [12]. Mayer, D.G. and Butler, D.G. (1993). Statistical Validation. *Ecological Modeling* 68: 21 - 32.
- [13]. Mitchell, P. L. (1997). Misuse of Regression for Empirical Validation of Models. *Agricultural Systems* 54: 313-326.
- [14]. Mitchell, P. L. and Sheehy, J. E. (1997). Comparison of Predictions and observations to assess Model Performance: A Method of Empirical Validation. *Kluwer Academic Publishers, Boston, MA, PP. 437-451.*
- [15]. Montgomery, D. C. (2005). *Design and Analysis of Experiments*, 5th ed. John Wiley and Sons New York, PP. 76-104 and 392 – 422.
- [16]. Sergio, B. and Joan, C. (2001). Oriented Principal Component Analysis for Large margin Classifiers. *Neural Networks* 14(10): 1447-1461.
- [17]. Tedeschi, L. O. (2006). Assessment of the Adequacy of Mathematical Models. *Agricultural systems* 89: 225-247.
- [18]. Wallach, D. and Goffinet, B. (1989). Mean Squared Error of Prediction as a Criterion for Evaluating and Comparing System Models. *Ecological Modeling* 44: 299-306.
- [19]. Willmott, C. J. (1982). Some Comments on the Evaluation of Model Performance. *Bulletin American Meteorological Society* 63: 1309-1313.