VOL. 03 NO. 9,
FEBRUARY, 2024
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Detection and Correction of Violations of Linear Model Assumptions by Means of Residuals

Reuben Friday Osemeke[1], John Nwabueze Igabari[2] & Nwabenu Dominic Christian[3]

[1]Department of Statistics, University of Delta, Agbor, Nigeria. [2]Department of Mathematics, Delta State University, Abraka, Nigeria. [3]Department of Mathematics and Statistics, Ogwashi-Uku. Delta State Polytechnic.

osemekereuben@gmail.com

Abstract

The examination of detection and correction of linear model violations by means of residuals is an important step in statistical analyses. Sometimes regression results are non-optimal, biased estimates, increase in Type 1 error rates, poor standard errors, untrustworthy confidence intervals, and insignificance F and T Test. This is as a result of assumption violations arising from non-normality, non-linearity, serial correlation, and multicollinearity All statistical analysis as well as graphical plots was carried out with SPSS 25 statistical package. The chi-squared test of two selected statistical test is greater than chi squared critical or sig value is less than 0.05. This shows evidence of heteroscedasticity(non-constant variance). The variance inflation factor (VIF) of predictors X_1 and X_3 are 50.603 and 49.286 which are appropriately high. This shows evidence of multicollinearity in the model. The ANOVA F test sig value of 0.112 is greater than 0.05. This shows overall non-significance relationship between the variables in the model and provides insight to other violations present. In all, the standardized residuals were used in the analysis instead of ordinary residuals. The Durbin Watson test statistics are 1.470 less than 1.5 specified benchmark showing evidence of serial correlation. Some adjustment such as increase in sample size with a linear and quadratic inheritance were made to correct all model violations and improve on the model through its test statistic, critical values, sig value and residual visual assessment. The best optimal transformation is when VIF is less than 2.500.The

VIF<2.500 satisfies the classical regression of normality, constant variance, linearity, independent, with no outlier present and no multicollinearity present in the model.

Keywords: Detection, Correction of Violations, Linear Model Assumptions, Means of Residuals

Introduction

Regression analysis is one of the important statistical tools and methods used in the process of statistical analysis [1]. It is widely accepted [2] that regression analysis is concerned with deriving parameters that describe the relationship between a dependent variable Y_i and set of independent variables, $X_1,\ X_2,\ X_3,\ \ldots, X_p$. Regression models have good predictive ability by estimating coefficients using least squares method (or otherwise) as well as the investigation of four classical regression assumptions such as normality, linearity, constant-variance, and independent of disturbance term [3].

One of the most frequently employed models to express the influence of several predictors $X_1,\ X_2,\ X_3,\ \ldots, X_p$ on a continuous outcome variable is the linear regression model which is mathematically defined by:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \in_i \quad (1)$$

It is proposed by [4] that standard regression models of equation (1) takes to measures without errors.

The various values of parameters $\beta_0,\ \beta_1,\ \beta_2, \ldots, \beta_p$ are each a measure of association between respectively independent variables and the dependent variable, Y_i .

Apart from the statistical purpose of making predictions, regression can be used to find a model as seen in equation (1). Before the fitted linear model of equation (1) is used for prediction purpose, analysis of residuals is used to test for classical regression assumptions.

These classical assumptions associated usually with the residuals have been extensively studied and applied to model real life data. Subjecting the analyses to serious classical assumption testing may shed bad light on fitted linear regression model.

Nevertheless, there are situations in statistical analysis where the underlying classical regression analysis provides poor model fit in data analysis. This is as a result of violations of model assumption emanating from non-normality of data distribution,

non-linearity (curvilinear effects), non-constant variance, serial correlation, collinearity pairwise and multicollinearity among the independent variables. It is observed by [2] that parameters of the simple linear regression model are not scale invariant but quite sensitive to data transformations.

In multiple linear regressions, much work has not been done on the use of residuals in their various capacities of test statistic, critical value, significant level and residual visualization's to detect various levels of assumption violations and not providing feasible control measures to remedy the problems of model violations.

In the view of [5], diagnostic techniques can be graphical, which is more flexible, but harder to definitely interpret or numerical (statistical tests), which are narrower in scope, but require no intuition. Montgomery, Peck & Vining (2001) . [4] is of the opinion that both types of diagnostic techniques are usually based on analysis of residuals which is an effective way to discover several types of model inadequacies including assumption violations. It was argued by [6] that final judgment on model adequacy and model violations should be based on statistical tests of residuals, instead of their graphical displays. However, [7] insists that residual plots are much more useful than formal testing procedures, when assessing model adequacy. In an empirical study, [8] used a time series regression model to predict infant mortality rate in Nigeria. The study showed that the variables were auto-correlated with decaying time lags.

However, despite the difference in opinions as presented in literature, both approaches remain the most commonly used regression diagnostic tools. Violations of these four classical regression assumptions, in line with the present of multicollinearity and outliers have to be corrected using trial and error transformations.

Different transformations are tried until the residuals seem feasible. Regression model is termed optimal, and are used for prediction purposes when four classical regression assumptions of residuals have been satisfied, in addition to no outlier present and as well variance inflationary factor below 2.500 for each of the independent variables in the model.

To this end, this study introduced a dynamic ways of the Detention and Corrections of violations of linear model assumptions by means of residuals. It is to investigate and analyze the corrections of violations of linear model assumptions by means of residuals.

The specific objectives of this study, therefore are to use examples to examine the effects of White test, and Breusch-Pagan test in detecting non-constant variance through the diagnoses of standardized residuals instead of ordinary residuals as used in literature , provide measures through statistical test and visual assessment to detect various model violations and validations in multiple settings at certain levels of significance, employing ANOVA F test to detect non-normality and provides insight

to other model violations and use values of variance inflation factor (VIF) to explore the performance of transformations of optimal regression models with classical regression assumptions.

Materials and Methods

The Graphical Plots of Heteroscedasticity Detention

It is widely accepted [4, 9, 10] that in multiple linear regression, the usual recommended method of detecting heteroscedasticity is to plot regression residuals against the predicted value $\hat{Y}_i$ or against set of independent variables and see if there is a pattern of horned shaped in residual plot indicating that variance of observations is increasing or decreasing with other variables.

However, when checking for heteroscedasticity variance, it is important to use standardized residuals, because ordinary residuals as used in literature for heteroscedasticity tendency. Before they are used in checking for equality of variance in regression analysis, they need to be standardized.

The graphical methods are only a means of preliminary analysis and it is quite subjective. In order to get more precise answer to questions of whether heteroscedasticity is present or not, it is better to use appropriate statistical test which has a specified benchmark and convincing interpretation.

Statistical Tests of Heteroscedasticity Detection

Breusch - Pagan Test for Standardized Residuals

Breusch-Pagan statistic [11] is used to detect heteroscedasticity (non-constant variance) present in a multiple linear regression model. Its test the linear associations between model standardized residuals and model predictors. In Breusch-Pagan test, the squared standardized residuals ($Z \in_i^2$) are regressed against some set of independent variables. The auxiliary equations now becomes

$$Z \in_i^2 = \sigma_0 + \sigma_1 X_1 + \sigma_2 X_2 + \cdots + \sigma_p X_p + e_i \qquad (2)$$

The coefficient of determination (R^2) derived from equation (2) are used to multiplied the given sample size (n) to obtain

$$nR^2_{z\in_i^2} = n * R^2_{z\in_i^2} \qquad (3)$$

Under the null hypothesis, linear model is distributed asymptotically as $\chi^2_{p-1,\alpha}$ (4)

The decision is to rejects null hypotheses of constant variance if test statistic is greater than the chi - squared critical value.

Otherwise, do not. Alternatively, if p-value is less than 0.05 level of significance, the null hypotheses of homoscedasticity is rejected.

White Test for Standardized Residuals

The immediate drawback of Breusch–Pagan test is that it can only detect linear associations between model standardized residuals and model predictors. In order to generalize it further, the white test [12] looks at higher –order, non-linear function forms of the X terms (i.e., quadratic and cross-product interactions among the predictors). In this case, the regression for (squared) standardized error terms, assume for four (4) independent variables would be defined as

$Z \in_i^2 = \sigma_0 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 + \sigma_4 X_4 + \sigma_5 X_1^2 + \sigma_6 X_2^2 + \sigma_7 X_3^2 + \sigma_8 X_4^2 + \sigma_9 X_1 X_2 X_3 X_4 + e_i$ (5)

The coefficients of determination (R^2) derived from equation (5) are used to multiply the sample size (n) to have

$$nR^2_{ze_i^2} \qquad (6)$$

And this follows a chi squared critical value of $\chi^2_{p-1,\alpha}$

The decision is to rejects null hypotheses if test statistic is greater than the chi - square critical value. Otherwise, do not reject.

Techniques for Detecting Multicollinearity with Adjusted Coefficients of Determination Diagnostics

Variance Inflationary Factor (VIF)

Variance inflationary factor (VIF) arises when some of independent variables in a multiple settings are highly correlated with a single independent variable. In such situations, collinear variables do not provide new information, and it becomes difficult to separate the effects of such variables on the dependent variables(Y). One method of measuring multicollinearity between one independent variable and other independent variables is by variance inflationary factor (VIF). It is defined as

$$VIF = \frac{1}{1-R_j^2} = \frac{1}{tolerance_{R_j^2}} \qquad (7)$$

where R_j^2 is the coefficient of multiple determination.

$$VIF_{adj} = \frac{1}{1- R^2_{adjusted}} = \frac{1}{tolerance_{adj}} \qquad (8)$$

Where $R^2_{adjusted}$ is the adjusted coefficient of determination?

The tolerance value is the inverse of VIF. The VIF value between 5 and 10 indicates very high correlated variables, and VIF values greater than 10 are an indication that the regression coefficients are feeble estimates with presence of multicollinearity [13]. Equation (8) has a better predictive power because the adjusted coefficient of determination select only significant optimal variables into the model than the ordinary coefficient of determination which increases as the sample size increases, so it is better to use the adjusted coefficient of determination as a better method for determining the variation of the independent variables over the dependent variable.

TABLE 1: VIF Presentation and Interpretation

VIF value	Decision	Conclusion
VIF = 1	Not Correlated	No Multicollinearity present
$1 < VIF \leq 5$	Moderately Correlated	Partial Multicollinearity
VIF>5	High Correlated variables	Full Present of Multicollinearity
VIF>10(More than 10)	100% Highly Correlated variables	100% Full present of Multicollinearity

Assessing Non-Normality of Residuals

According to [14], violations of assumption of normality of residuals may result in incorrect inferences in small samples, while [15] asserted that violations of normality assumption do not have much impact on inference. For the scope of this study, two methods will be diagnosed which are the standardized residual plots and ANOVA F test diagnostics for statistical plots.

Graphical Plots of Assessing Non-Normality of Standardized Residuals

The scatter plots of standardized residuals against predicted value are used to detect non-normality of the residuals. Evidence of non-randomness and uneven distribution is a sign of non-normality.

ANOVA F Test for Assessing Non-Normality

The ANOVA F test is used to test for overall significance of multiple independent variables; the null and alternative hypotheses are set up as follows

H_0: Estimated regression model is non - normal.

H_1 : Estimated regression model is quite normal.

The F test statistic for the entire regression model is $F = \frac{MSR}{MSE}$ (9)

where MSR stands for Mean Square Regression and MSE is Mean Square Error.

The ANOVA F critical value is defined as

$F_{(p-1,n-p-1)}\alpha$

(10)

The decision is to reject null hypotheses if the F test calculated is greater than the F critical. The Sig value can also be used to assess the stated hypotheses.

Detecting Serial Correlation (Non-Independence)

It is well known that even cross-sectional data (i.e., data collected on different units at about the same time) may be correlated because of spatial proximity, family relationships, or other reasons. Such autocorrelation, due to spatial clustering of sampling units, needs also to be taken into account appropriately when building the model.

The Graphical Test of Assessing Serial Correlation

The residuals plots against time or observations or against predicted value helps to detect serial correlation in the model. The scatter plots showing evidence of oscillation that is data set moving up and down in a serially correlated pattern.

Statistical Test for Detecting Serial Correlation with Specified Benchmark

The Durbin Watson statistic has a value ranging between 0 and 4. A value of 2.0 indicates there is no autocorrelation detected in the sample. A value below 1.5 shows evidence of auto-correlation. As a rule of thumb, Durbin Watson should be between 1.5 and 2.5 to indicate independent of the observation. Any value below or outside this range shows non-independent, which is an indication that one of the assumption has been violated.

Results and Discussion

Data Presentation

A Bread bakery is interested in establishing the effect of Trademark (X1), Bread Texture (X2) and Bread aroma (X3) on Consumers' attitude(Y). Twenty three customers were sampled, and their ratings are presented in Table 2, and the bivariate correlations matrix in Table 3.

Table 2 Consumers' attitude survey

n	Y	X1	X2	X3
1	6.52	26.02	26.02	25.02
2	7.92	20.02	16.02	52.02
3	6.82	16.02	10.02	29.02

4	6.52	17.02	12.02	28.02
	.	.	.	.
	.	.	.	.
20	6.82	35.02	38.02	39.02
21	6.52	15.02	12.02	22.02
22	8.62	28.02	32.02	65.02
23	8.62	32.02	38.02	49.02

Table 3: Bivariate Correlation matrix

	Y	**X1**	**X2**	**X3**
Y	1	.366	.367	.511**
X1	.366	1	.990*	.635**
X2	.367	.990*	1	.647**
X3	.511*	.635**	.647**	1

The Bivariate Correlations of Table 3 suggests that a strong correlation exist between Trademark(X1) and Bread Texture(X2) of 0.990 (r = 0.990) . This is a sign of multicollinearity present in the data set. The bivariate correlations among the remaining independent variables do not appear to be so strong. They are fair correlations which we may not consider as causing multicollinearity.

Measuring Heteroscedasticity with Standardized Residuals Graphical Plots

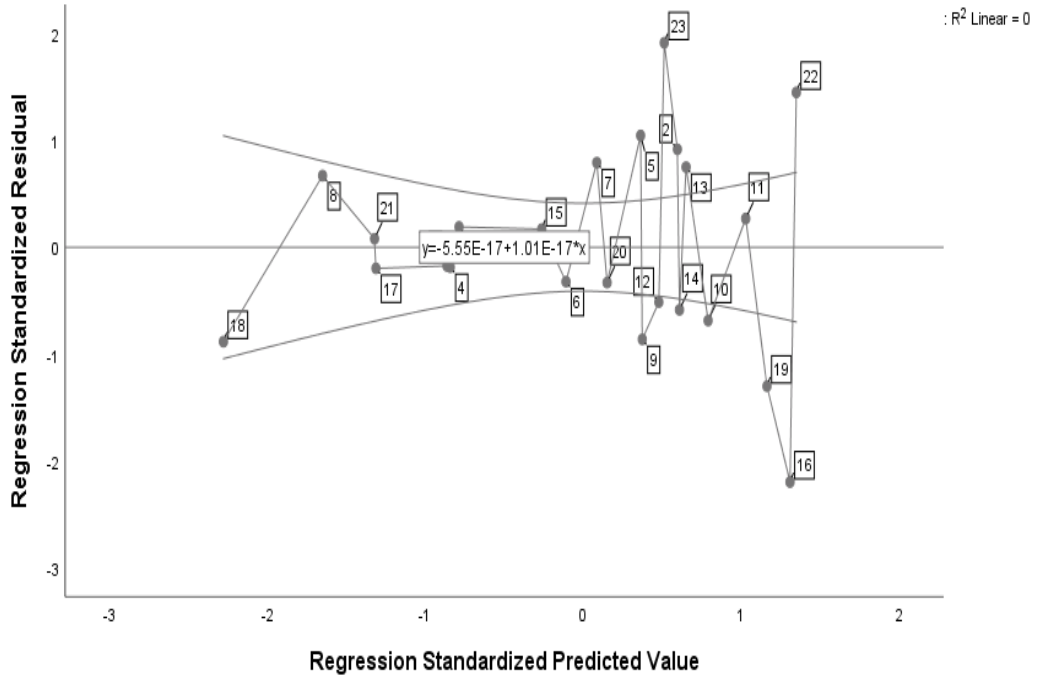


Figure 1. *Regression Analysis of Standardized Residuals.*

The scatter plot of Figure 1 shows an increasing and decreasing trend of standardized residuals. It is horned shaped in pattern. This finding indicates heteroscedasticity in multivariate set of independent variables. This provides general overview on the detention of all model violations.

Statistical Methods of Measuring Heteroscedasticity (Non-Constant Variance) with Standardized Residuals diagnostics

Table 5. *Numerical Values of two (2) Selected Statistical test in identifying Heteroscedasticity in multivariate regression.*

Statistical Test	Chi Square test	Sig Value	F crit	Decision	Conclusion
Breusch-Pagan	5.389>F_{Crit}	0.020<0.05	3.84	Rejects Ho	Heteroscedasticity is present
White test	17.408>F F_{Crit}	0.043<0.05	16.92	Rejects Ho	Heteroscedasticity is present

Measuring the Degree and Impact of Multicollinearity

Table 6. *Testing for Multicollinearity: Assessing Sig value, Tolerance and VIF values with Standardized Residuals Diagnostics.*

Variable	Sig	Tolerance	Variance Inflation	Remark
X1 Trademark	.855>0.05	.020>0.01	50.603>10	Multicollinearity Present
X2 Bread Texture	.892>0.05	.579>0.01	1.726<10	Multicollinearity Present
X3 Bread Contentment	.083>0.05	.020>0.01	49.286>10	Multicollinearity Present

In Table 6, variance inflationary factor (VIF) of Trademark(X1) and Bread Contentment(X3) are 50.603 and 49.286 which are appropriately high. It is above the rule of thumb ten (10) as said in literature. Also, this shows great evidence of multicollinearity in regression model.

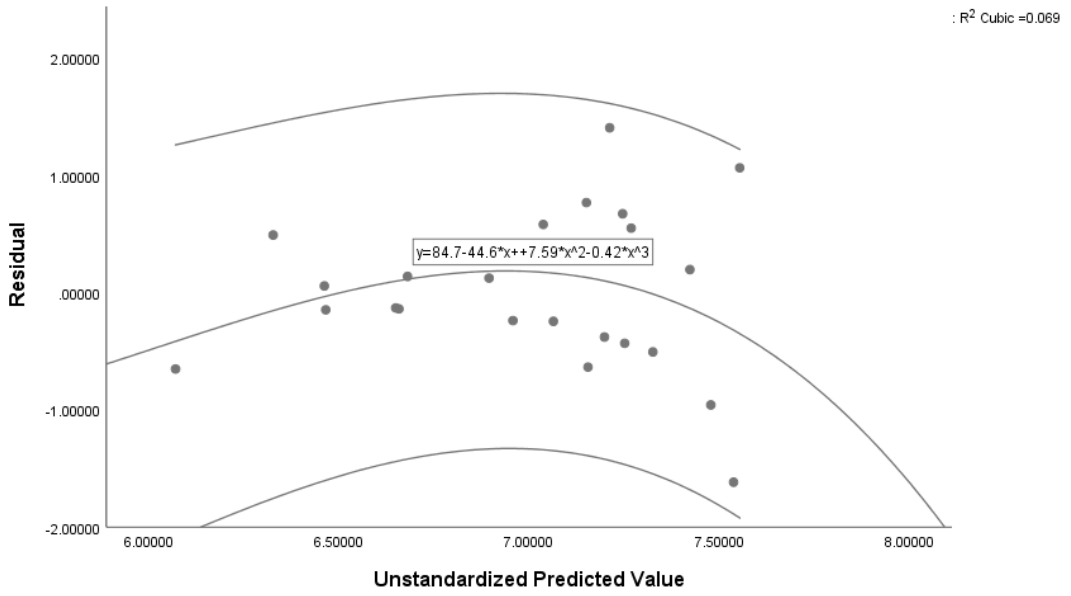
Identifying Non-Normality of Error Term

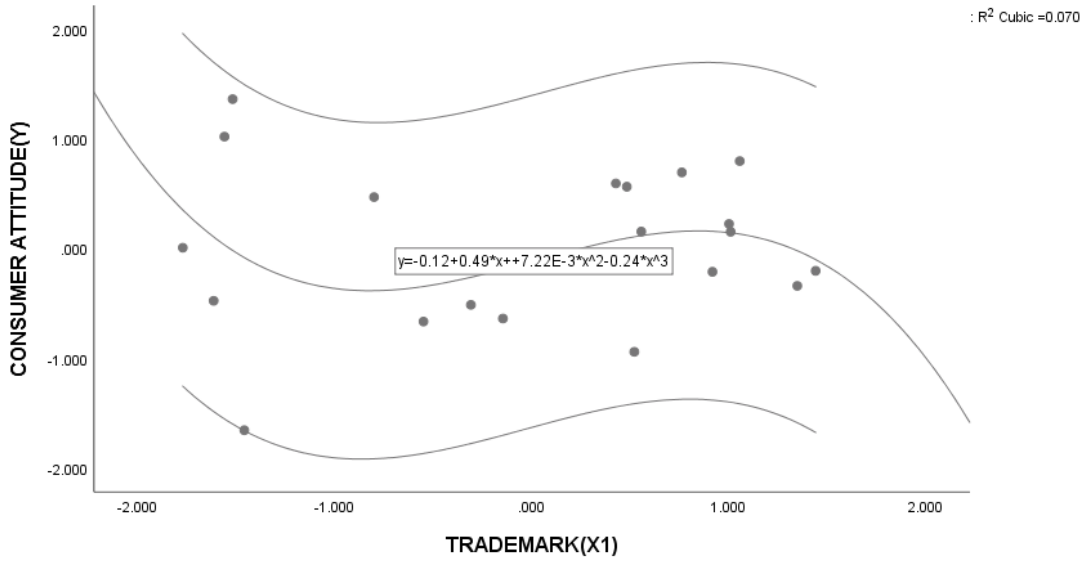
Table 7. *ANOVA F test and Sig Value of each Predictor variables.*

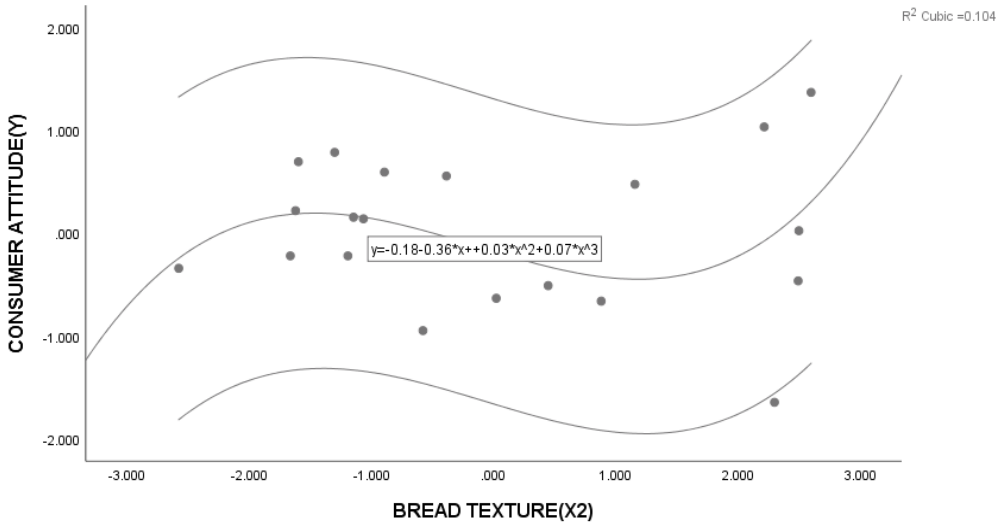
Source	Type III SS	F	SIG	F Crit	Decision	Conclusion
Corrected Model	3.685[a]	2.278<3.13	.112>0.05	3.13	Accepts Ho	General Model non-normality
x1	.019	.034<3.13	.855>0.05	3.13	Accepts Ho	Individual Model non-normality
x2	.010	.019<3.13	.892>0.05	3.13	Accepts Ho	Individual Model non-normality
x3	1.800	3.338>3.13	.083>0.05	3.13	Accepts Ho	Individual Model non-normality

The analysis of Table 7 shows that p-value of 0.112 is greater than 0.05. The individual variables also shows evidence of non-normality. The null hypotheses of model inadequacy is accepted.

General Analysis of Residual and Partial Regression Plots




: R2 Cubic =0.070
CONSUMER ATTITUDE(Y)
2.000
1.000
.000
-1.000
-2.000
y=-0.12+0.49*x++7.22E-3*x^2-0.24*x^3
-2.000
-1.000
.000
1.000
2.000
TRADEMARK(X1)


R2 Cubic =0.104
CONSUMER ATTITUDE(Y)
2.000
1.000
.000
-1.000
-2.000
y=-0.18-0.36*x++0.03*x^2+0.07*x^3
-3.000
-2.000
-1.000
.000
1.000
2.000
3.000
BREAD TEXTURE(X2)

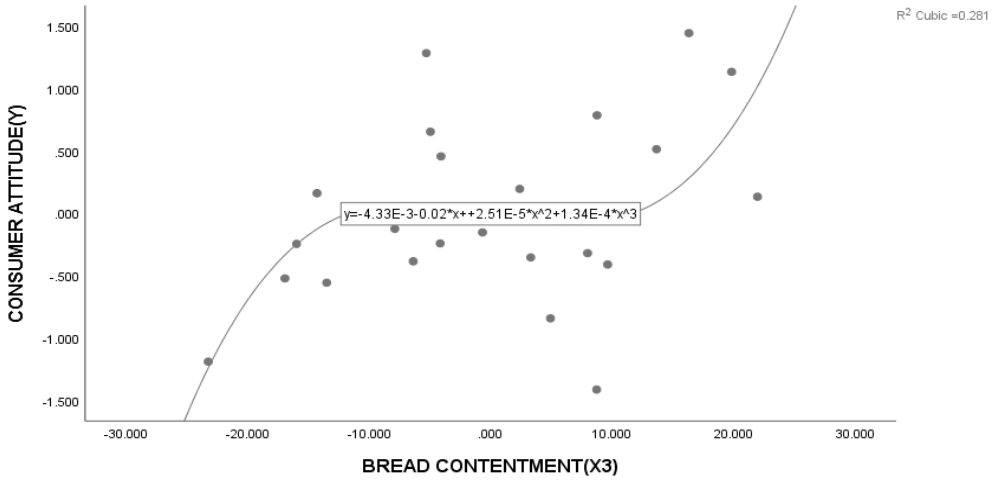

R2 Cubic =0.281
CONSUMER ATTITUDE(Y)
1.500
1.000
.500
.000
-.500
-1.000
-1.500
y=-4.33E-3-0.02*x++2.51E-5*x^2+1.34E-4*x^3
-30.000
-20.000
-10.000
.000
10.000
20.000
30.000
BREAD CONTENTMENT(X3)

Figure 2. *Standardized partial regression plots.*

Figure 2 shows visual assessment of residuals and partial regression plots. The above plots shows evidence of curvilinear effects which is an indication of nonlinearity in the residuals and provides insights to other assumption violation. The above regression analysis through its statistical test and graphical plots shows that the four classical regression of error term have been violated. There was no evidence of any outliers present. This is where some researcher makes mistakes. Publishing a work that has assumption model violations is like treating an ailment without going through medical laboratory test. This is called statistical error because it will not lead to optimal model solution.

A transformation

Table 8.: *VIF values at various Transformation Level: An Insight Guide to Optimal and Acceptable Regression model.*

Variables	VIF of Log Transformation (n≤23)	VIF of Increase Sample Sizes (n≤69)	VIF of Square Root Transformation (n≤23)
X1	49.286	2.028<2.500: Optimal Model	52.434
X2	50.603	2.238<2.500: Optimal Model	52.706
X3	1.726	1.179<2.500: Optimal Model	1.776

In Table 8, the VIF values of increase in sample sizes ($n \leq 69$) with linear and quadratic trends seems feasible and satisfactory than the other transformation because the (VIF) are very minimal (less than 2.500). The presence of multicollinearity has been removed to the barest minimum.

Table 7. *Homoscedasticity (Constant-variance): Assumption validation after Transformation (n≤69).*

Statistical Test	Chi Square	Sig Value	Critical value	Decision	Conclusion
White Test	9.639	0.381 >0.05	15.507	Accepts Ho	Homoscedasticity Present

The White test after transformation is the best statistical test used to test for heteroscedasticity. The reason is that it has a linear trend, quadratic trend and interaction term in one linear model. It can also be used to test for heteroscedasticity

nd homocedasticity at different model trends and can aslo be used to detect other violations present in the regression model.

Table 8. *Non-normality Assumption validation after Transformation (n≤69)*

ANOVA F Test Statistic

Model		SS	Df	MS	F	Sig.	F Critical	Remark
1	Regression	12.496	3	4.165	8.806	.000<0.05	3.34	Rejects Ho and errors are Normality Distributed

Finally, the above analysis with require transformation of VIF less than 2.500 is satisfactory as we can see a great significant improvement in the regression analysis in terms of Test statistics, critical value, sig value and the graphical plots.

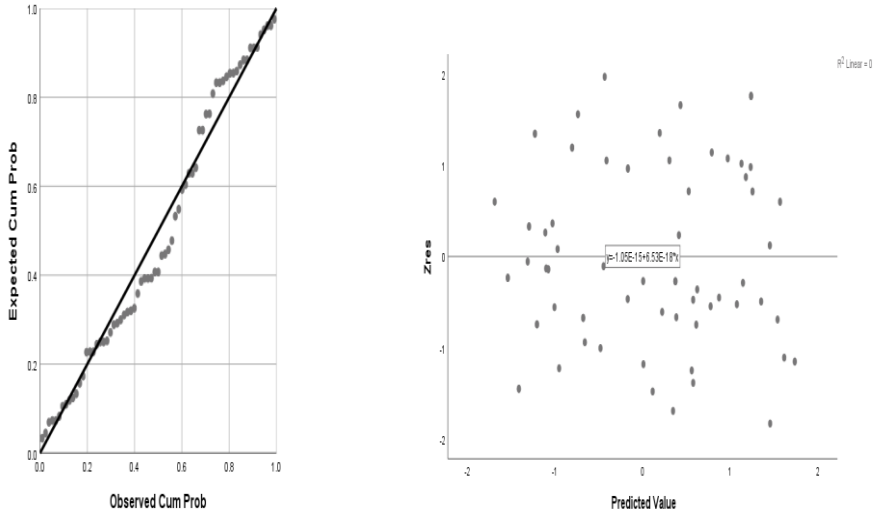


Figure 4: Standardized Residuals Graphical plots and Normal P-P plot.

This are both used to show general overview and provides insight into all model validations ranges from normality, constant variance, independent and linearity of the standardized residuals. This shows great evidence of optimal regression model and predictive ability. The Normal P-P plots shows a 45 degree linearity pattern, and with a moving from (0,0) to (1,1). The Graphical plot of standardized residual against predicted value shows hovering around the mean of zero, are evenly distributed and shows no sign of outlier.

Findings from the Study

1) The regression standardized residuals against regression predicted value shows general evidence of violations and validations of classical regression assumptions before and after transformation. It gives an idea into the analysis of partial regression plots.
2) The two selected statistical test with a standardized residuals diagnostics of various values of chi square test ($n*R^2$) are greater than chi square critical values or sig values are less than 0.05 level of significance. This shows evidence of heteroscedasticity present in the multivariate model.
3) The variance inflation factor (VIF) of Trademark(X1) and Bread Texture(X2) are high, 50.603 and 49.286 respectively, showing evidence of multicollinearity.
4) The ANOVA F test shows evidence of overall non-normality and individual non-normality and provides insight to other model violations.
5) The variance inflation factor (VIF) of increase in sample sizes below 2.500 for each of the independent variables shows evidence of optimal model transformation method, and satisfaction of the six classical regression assumptions.

Conclusion

The following conclusion were reached

1) The White test statistic of standardized residuals diagnostics drives heteroscedasticity (non-constant variance) and homoscedasticty(constant variance) for both small and large sample sizes in terms of chi square (Test statistic), critical value and Sig value at 0.05 level of significance.
2) The scatter plots of regression standardized residuals against the regression predicted value shows general overview about assumption model violations and validations. It serve as a road map to show all model violations and validations in a single settings.
3) The ANOVA F test through its non-significance linear relationships shows evidence of general non-normality and individual non-normality as well as providing insight to some model violations.
4) VIF below 2.500 as the best feasible and satisfactory transformation of optimal regression models and satisfaction of classical assumption.
5) Using VIF greater 10 or less than 2.500 as the road map to non-optimal and non-optimal models increase in sample sizes with linear and quadratic features have a variance inflation factor less than 2.500 and should be used.

Conflicts of Interest

The authors hereby affirm that there is no conflict of interest whatsoever

References

Adenomen, M. O and Micheal, V. A., Model Violations Diagnostics. International Journal of Innovative Finance and Economics Research. 6(3): 74-89 (2017).

Basero O., Modelling transformed Health care cost with unknown heteroscedasticity. Value of Health. Vol 13 No 3. Online ISSN: 1524-4733. Print ISSN: 1098-3015 (2007).

Belsley, D. A. conditioning Diagnostics: Collinearity and Weak Data in Regression, John Wiley and sons, Inc, New york (1991).

Belsley, D. A. conditioning Diagnostics: Collinearity and Weak Data in Regression, John Wiley and sons, Inc, New york (1991).

Breusch, T & Pagan, A. A simple Test for heteroscedasticity and Random Coefficients Variation. Econometrical, 47, 1287-1294 https://doi.org/10-2307/19111963.(1980)

Breusch, T & Pagan, A. A simple Test for heteroscedasticity and Random Coefficients Variation. Econometrical, 47, 1287-1294 https://doi.org/10-2307/19111963.(1980)

Chatterjee, S & Hadi, A. S. Sensitivity Analysis: Linear Regression. New York: John Wiley & Sons, Inc (1998).

Chatterjee, S & Hadi, A. S. Sensitivity Analysis: Linear Regression. New York: John Wiley & Sons, Inc (1998).

Cohen, J., Cohen, P., West, S G., & Aiken, L. S. Applied multiple regression/ correlation analysis for the behavioural sciences (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates (2003).

Cohen, J., Cohen, P., West, S G., & Aiken, L. S. Applied multiple regression/ correlation analysis for the behavioural sciences (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates (2003).

Cook and Weisberg. Diagnostics for heteroscedasticity in Regression. Bometrika. 70 (1): 1-10 doi 10. 1093/biometric 70.1.1 (2013).Mendenhall, W. M & Sincich, T. A second course in statistics: Regression analysis (6th ed). Upper Saddle River, NJ: Prentice Hall (2003).

Cook and Weisberg. Diagnostics for heteroscedasticity in Regression. Bometrika. 70 (1): 1-10 doi 10. 1093/biometric 70.1.1 (2013).

Faraway, J. Linear Models with R. Boca Raton, FLC: Chapman & Hall (2004).

Fox, J. Regression Diagnostics. Newbury Pack. C. A Sage (1997).

Fox, J. Regression Diagnostics. Newbury Pack. C. A Sage (1997).

Igabari, J. N. Computational Implications for the Least Squares Parameters of the Simple Linear Regression Model under Data Transformation. Nigeria Journal of Science and Environment. 11(1): 15 – 17.

Mendenhall, W. M & Sincich, T. A second course in statistics: Regression analysis (6th ed). Upper Saddle River, NJ: Prentice Hall (2003).

Montgomery, D. C, Peck, E. A, and Vining, G. E Introduction to Linear Regression Analysis, 3rd Edition, Wiley New York (2012).

Onyemarin, N. N; Ojarikre, H. I. & Igabari, J. N. An ARMA Modelling Approach on Infant Mortality in Nigeria. Nigeria Journal of Science and Environment. 21(1): 94 - 114.

Onyemarin, N. N; Ojarikre, H. I. & Igabari, J. N. An ARMA Modelling Approach on Infant Mortality in Nigeria. Nigeria Journal of Science and Environment. 21(1): 94 - 114.

Schmidt, A. F., & Finan, C.. Linear Regression and Normality Assumption. Journal of Clinical Epidemiology, Vol, 98, pp. 146-151 (2018).

Schmidt, A. F., & Finan, C.. Linear Regression and Normality Assumption. Journal of Clinical Epidemiology, Vol, 98, pp. 146-151 (2018).

White, H. A heteroskedasticity consistent covariance matrix best estimator and a direct test for heteroskedasticity. Econometrical 48: 817-838 (1980).

White, H. A heteroskedasticity consistent covariance matrix best estimator and a direct test for heteroskedasticity. Econometrical 48: 817-838 (1980).