African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

DISCRIMINATING BETWEEN SECOND-ORDER MODEL WITH/WITHOUT INTERACTION BASE ON CENTRAL TENDENCY ESTIMATION

P. S. Owhondah[1], D. Enegesele[2*], O.E. Biu[3], D.S.A. Wokoma[4]

[1]Department of Mathematics/Statistics, Ignatius Ajuru University of Education, Rivers State, Nigeria

[2]Department of Mathematical & Computing Sciences, KolaDaisi University, Ibadan, Nigeria

[3]Department of Mathematics & Statistics, Faculty of Science, University of Port Harcourt, Nigeria.

[4]Department of Mathematics & Statistics, Captain Elechi Amadi Polytechnic, Rumuola, Port Harcourt, Nigeria

Email:biu.emmanuel@uniport.edu.ng,emmanuelbiu@yahoo.com,dennis.enegesele@koladaisiuniversity.edu.ng, ndukawonu@gmail.com

**Corresponding Author: den4_js@yahoo.com*

Cite this article:

Owhondah P.S. , Enegesele D., Biu O.E., Wokoma D.S.A. (2021), Discriminating Between Second-Order Model With/Without Interaction Base on Central Tendency Estimation. African Journal of Mathematics and Statistics Studies 4(3), 47-63. DOI: 10.52589/AJMSS-71MQSBGZ.

Manuscript History

Received: 6 June 2021

Accepted: 5 July 2021

Published: 8 Oct 2021

Copyright © 2020 The Author(s). This is an Open Access article distributed under the terms of Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0), which permits anyone to share, use, reproduce and redistribute in any medium, provided the original author and source are credited.

ABSTRACT: *The study deals with discriminating between the second-order models with/without interaction on central tendency estimation using the ordinary least square (OLS) method for the estimation of the model parameters. The paper considered two different sets of data (small and large) sample size. The small sample size used data of unemployment rate as a response, inflation rate and exchange rate as the predictors from 2007 to 2018 and the large sample size was data of flow-rate on hydrate formation for Niger Delta deep offshore field. The R^2, AIC, SBC, and SSE were computed for both data sets to test for adequacy of the models. The results show that all three models are similar for smaller data set while for large data set the second-order model centered on the median with/without interaction is the best base on the number of significant parameters. The model's selection criterion values (R^2, AIC, SBC, and SSE) were found to be equal for models centered on median and mode for both large and small data sets. However, the model centered on median and mode with/without interaction were better than the model centered on the mean for large data sets. This study shows that the second-order regression model centered on median and mode are better than the model centered on the mean for large data set, while they are similar for smaller data set. Hence, the second-order regression model centered on median and mode with or without interaction are better than the second-order regression model centered on the mean.*

KEYWORDS: Second-Order Model With/Without Interaction, Central Tendency Estimation, Ordinary Least Square, Test For Adequacy, Small Sample Size, and Large Sample Size.

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63) www.abjournals.org

INTRODUCTION

Simple linear regression is an approach in statistics that is employed in the modeling of linear surfaces (Shalabh, 2012). Regression analysis can be a linear, nonlinear, and second-order (quadratic or polynomial) model. There is a major problem in deciding whether a model is linear or nonlinear as some literature will say that if the highest power of the unknown is one, it is linear and if the highest power is two, the model is quadratic, and if more than two is polynomial. All of the above definitions and classifications of a regression model are now misleading. A regression model is linear when it is linear in parameters, irrespective of the fact that it is linear, quadratic, or polynomial. The linear regression model has only one independent variable and states that the mean of the dependent variable changes at a constant rate as the value of the independent variable increases or decreases. The nonlinear model is a process where data are modeled by a function which is a nonlinear combination of the model parameters and depend on one or more independent variable. A second-order model is a regression model with k predictors. A second-order (k = 2) forms a quadratic expression, a third-order (k = 3) polynomial forms a cubic expression. Central tendency according to Manikandan (2011) is defined as the statistical measure that identifies a single value as representative of an entire distribution. It aims to provide an accurate description of the entire data. The mean, median, and mode are the three commonly used measures of central tendency.

The rationale for this paper is to compare the second-order quadratic model with or without interaction using the central tendency estimate. The specific objectives of the study include. (1) Estimating arithmetic mean, median, and mode of a given set of data (2) Estimating the parameters of quadratic (centre on mean, median, and mode) model with or without interaction for large and small sample sizes. (3) Compare the estimated parameters and model adequacy criterion of three models to determine the best model for large and small sample sizes.

LITERATURE REVIEW

Comparison of the second-order quadratic model of central tendency estimation with/without interaction has not been so evident in the literature. Iwundu (2016b) considered the behavior of equiradial designs under changing model parameters for reduced and full quadratic models. The work did not consider the quadratic model with the central tendency (mean, median, and mode). Sameera (2014) considered the comparison of models with/without intercept which was seen as a full and reduced model, but the research centered its findings on a first-order linear regression model. Jeffery *et al.*, (2012) presented a comparison of methods of estimating quadratic effects in nonlinear structural equation models, it was discovered that quadratic effects between non-discrete variables are often hypothesized in the social sciences, as a result of the fact that when quadratic terms are coupled with a linear component, they become more adequate in approximately many curvilinear behaviours. Comparisons of estimation approaches of nonlinear effects in Standard Error of the Mean (SEM) have been conducted through many simulation studies. Delphine and Olivier (2014) study on Robust analysis of the central tendency showed that outliers can sometimes be very difficult to detect and that the full inferential procedure is to some extent biased by such a procedure. A more appropriate and modern approach is to use a robust procedure that makes

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

available the estimation, inference, and testing that are not influenced by outlying observations but describes correctly the structure of the data

Second-order quadratic models are employed in comparing models with or without interaction using central tendencies (mean $\bar{x}$, median x' and mode x''). The data of unemployment, exchange rate, and inflation from 2007 to 2018 was used as illustration 1 (small data). Also, the data of flow-rate on hydrate formation was used as illustration 2 (large data). The secondary data of unemployment rate, inflation rate and the exchange rate used was obtained from the Central Bank of Nigeria, Statistical Bulletin (2017), and National Bureau of Statistics (2017). The flow-rate on hydrate formation data was from Niger Delta deep offshore, obtained from the University of Port Harcourt Petroleum Department. It consists of four predictors and one response. The inflation rate is denoted by x_1, the exchange rate is denoted by x_2, and y represent the unemployment rate.

MATERIALS AND METHODS

The statistical software used are Matlab, Micro-Excel, and Minitab18. The research considered two different sets of data; a small sample size (n < 30) which is data on the unemployment rate as a response variable, inflation rate and exchange rate as the predictors' variables from 2007 to 2018 (n=12). The large sample size (where n > 30) is data of flow-rate on hydrate formation for Niger Delta deep offshore field (n=130). The small data was used as illustration 1 and the large data was used as illustration 2.

Estimating the mean, median, and mode of ungrouped data for illustration 1(small sample size)

Regression Analysis on Central Tendency:

The first-order regression model without interaction is given as:

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + e_i \quad (1)$$

The second order regression model without interaction is given as.

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + e_i \quad (2)$$

The first-order regression model with interaction is given as

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e_i \quad (3)$$

The second-order regression model with interaction is given as

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + e_i \quad (4)$$

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

The center of mean, median, and mode are as follows:

For mean:

where $x_1 = (x_1 - \bar{x}_1)$, $x_2 = (x_2 - \bar{x}_2)$

For median:

where $x_1 = (x_1 - x_1')$ $x_2 = (x_2 - x_2')$

For mode:

where $x_1 = (x_1 - x_1'')$ $x_2 = (x_2 - x_2'')$

A quadratic model with arithmetic mean $\bar{x}$ without interaction is given as:

$$Y = \beta_O + \beta_1\left(x_i - \bar{x}\right) + \beta_2\left(x_i - \bar{x}\right)^2 + e_i \tag{5}$$

The quadratic model with median x' without interaction is given as:

$$Y = \beta_O + \beta_1\left(x_i - x'\right) + \beta_2\left(x_i - x'\right)^2 + e_i \tag{6}$$

The quadratic model with mode x'' without interaction is given as:

$$Y = \beta_O + \beta_1\left(x_i - x''\right) + \beta_2\left(x_i - x''\right)^2 + e_i \tag{7}$$

The quadratic model with interaction for mean is given as:

$$Y = \beta_0 + \beta_1\left(x_1 - \bar{x}_1\right) + \beta_2\left(x_2 - \bar{x}_2\right) + \beta_{12}\left(x_1 x_2 - \bar{x}_1 \bar{x}_2\right) + \beta_{11}\left(x_1 - \bar{x}_1\right)^2 + \beta_{22}\left(x_2 - \bar{x}_2\right)^2 + e_i \tag{8}$$

The quadratic model with interaction for the median is given as:

$$Y = \beta_0 + \beta_1\left(x_1 - x_1'\right) + \beta_2\left(x_2 - x_2'\right) + \beta_{12}\left(x_1 x_2 - x_1' x_2'\right) + \beta_{11}\left(x_1 - x_1'\right)^2 + \beta_{22}\left(x_2 - x_2'\right)^2 + e_i \tag{9}$$

The quadratic model with interaction for mode is given as:

$$Y = \beta_0 + \beta_1\left(x_1 - x_1''\right) + \beta_2\left(x_2 - x_2''\right) + \beta_{12}\left(x_1 x_2 - x_1'' x_2''\right) + \beta_{11}\left(x_1 - x_1''\right)^2 + \beta_{22}\left(x_2 - x_2''\right)^2 + e_i \tag{10}$$

Estimating the mean, median, and mode of ungrouped data for Illustration 2 (large sample size)

The first-order regression model without interaction is given as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \ell_i \tag{11}$$

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63) www.abjournals.org

The second-order regression model without interaction is given as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 + \ell_i \qquad (12)$$

The first-order regression model with interaction is given as:

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 \\ & + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + e_i \end{aligned} \qquad (13)$$

The second-order regression model with interaction is given as

$$\begin{aligned} Y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{44} x_4^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 \\ & \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + e_i \end{aligned}$$
(14)

The center of mean, median, and mode are as follows:

For Mean:

$$where \;\; x_1 = (x_1 - \bar{x}_1), \;\; x_2 = (x_2 - \bar{x}_2), \; x_3 = (x_3 - \bar{x}_3), \; x_4 = (x_4 - \bar{x}_4)$$

Then the interaction is given as:

$$(x_1 - \bar{x}_1)(x_2 - \bar{x}_2), (x_1 - \bar{x}_1)(x_3 - \bar{x}_3), \ldots (x_3 - \bar{x}_3)(x_4 - \bar{x}_4)$$

For Median:

where $x_1 = (x_1 - x_1'), \; x_2 = (x_2 - x_2'), \; x_3 = (x_3 - x_3'), \; x_4 = (x_4 - x_4')$

Then the interaction is given as:

$$(x_1 - x_1')(x_2 - x_2'), (x_1 - x_1')(x_3 - x_3') \ldots (x_3 - x_3')(x_4 - x_4')$$

For Mode:

where $x_1 = (x_1 - x_1''), \; x_2 = (x_2 - x_2''), \; x_3 = (x_3 - x_3''), x_4 = (x_4 - x_4'')$

Similarly, the interaction is given as:

$$(x_1 - x_1'')(x_2 - x_2''), (x_1 - x_1'')(x_3 - x_3'') \ldots (x_3 - x_3'')(x_4 - x_4'')$$

The flow-rate on hydrate formation data sets baseline was developed and was used to define the multiple linear regression relationship, the interaction between all the variables causing hydrate formation and the Cobb Douglass model was fitted. This baseline model (multiple linear regression) was used to determine the needed variations to be made on that field to effectively manage hydrate before agglomeration to the point of creating a blockage. The response variable (Qoil) is the flow-rate of oil and the predictor variables are Basic Sediment and Water (BSW), Gas Oil Ratio (GOR), Well Head Pressure (WHP), and the Well Head Temperature (WHT).

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

Estimating the Mean for Ungrouped Data

Given ungrouped data without frequency or repetition of values as seen below

$$x_1\ ,\ x_2,\ x_3,\ x_4, \ldots,\ x_n$$

The mean as expressed in Nwagozie (2011), Manikandan (2011), and Kellr and Warrack (2003) is given as,

$$\text{Mean } (\bar{x}) = \frac{x_1 + x_2 + \ldots x_n}{n} \tag{15}$$

Estimating the Median for Ungrouped Data

Egbule (2008) and Manikandan (2011) define the median of ungrouped data as:

$$\text{Median} = \frac{x_2 + x_3}{2}$$

If the data is given as $x_1, x_2, x_3 x_4$ and x_3 if the data is orderly arrange as $x_1\ ,\ x_2,\ x_3,\ x_4,\ .x_5$

Estimating the Mode for Ungrouped Data

Manikandan (2011) defines the mode for ungrouped as the value that repeats itself most often in data. For ungrouped frequency distribution data given as $x_1\ ,\ x_2,\ x_3,\ x_2,\ .x_4$, the mode is x_2

Obtaining the Deviation for each of Mean, Median, and Mode from the Explanatory Variables x_i

Obtaining the deviation from the mean ($\bar{x}$)

Given the data $x_1, x_2, x_3, \ldots, x_n$

The deviations from the mean are

$$(x_1 - \bar{x}), (x_2 - \bar{x}), (x_3 - \bar{x}), \ldots, (x_n - \bar{x})$$

We also obtain the square of the deviations given as

$$(x_1 - \bar{x})^2, (x_2 - \bar{x})^2, (x_3 - \bar{x})^2, \ldots, (x_n - \bar{x})^2$$

Obtaining the Deviations from the Median (x^1)

The deviations from the median are

$$(x_1 - x'), (x_2 - x'), (x_3 - x'), \ldots, (x_n - x')$$

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

And the square deviations are given as

$$(x_1 - x')^2, (x_2 - x')^2, (x_3 - x')^2, \ldots, (x_n - x')^2$$

Obtaining the Deviation from the Mode (x'')

The deviations from the Mode are:

$$(x_1 - x''), (x_2 - x''), (x_3 - x''), \ldots, (x_n - x'')$$

And the square deviations are given as

$$(x_1 - x'')^2, (x_2 - x'')^2, (x_3 - x'')^2, \ldots, (x_n - x'')^2$$

Obtaining parameter estimates for models of mean, median, and mode.

Using the model of mean, median, and mode and data of response variable $Y_1, Y_2, \ldots, Y_n$, and explanatory variables $x_1, x_2, \ldots, x_n$, the system of equations is given as:

$$\begin{aligned}
Y_1 &= \beta_0 + \beta_1 r_1 + \beta_2 r_1^2 + e_1 \\
Y_2 &= \beta_0 + \beta_1 r_2 + \beta_2 r_2^2 + e_2 \\
&. \\
&. \\
&. \\
Y_n &= \beta_0 + \beta_1 r_n + \beta_2 r_n^2 + e_n
\end{aligned}$$

(16)

Where $r = (x_i - \bar{x})$ and $r^2 = (x_i - \bar{x}),^2$ and writing (16) in matrix form,

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_n \end{pmatrix} \; and \; \underline{x} = \begin{pmatrix} 1 & r_1 & r_1^2 \\ 1 & r_2 & r_2^2 \\ . & & \\ . & & \\ . & & \\ 1 & r_n & r_n^2 \end{pmatrix} \quad \underline{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} \quad \underline{e} = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix}$$

The above is in the form

$$\underline{Y} = \underline{\beta}\ \underline{x} + \underline{e} \qquad (17)$$

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

We obtain the transpose of x written as x' and is given as

$$x' = \begin{pmatrix} 1 & 1. & . & . & 1 \\ r_1 & r_2 . & . & . & r_n \\ r_1^2 & r_2^2 . & . & . & r_n^2 \end{pmatrix}$$

The least square equation is given as follows

$$\underline{\hat{\beta}} = \left(x^1 x\right)^{-1} \left(x^1 Y\right) \quad (18)$$

Testing for model adequacy

AIC approach for mean, median and mode

The AIC for mean model as shown in Kutner et al (2005) is given as.

$$AIC(\bar{x}) = n / nSSE(\bar{x}) - n / nn + 2P \quad (19)$$

The AIC for the median model as shown in Kutner et al (2005) is given as.

$$AIC(x') = n / nSSE(x') - n / nn + 2P \quad (20)$$

The AIC for mode model as shown in Kutner et al (2005) is given as.

$$AIC(x'') = n / nSSE(x'') - n / nn + 2P \quad (21)$$

$$SSE = \sum_{i=i}^{n} \sum_{i=i}^{n} \left(xij - \bar{xj} \right)^2 \quad (22)$$

Schwarz' Bayesian criterion

It is given as

$$SBC_P = n \ln SSE - n \ln n + [\ln n] p \quad (23)$$

This is also applied to further test for the adequacy of the model. The smaller the SBC the better the model.

Coefficient of Determination (R^2)

The R^2 statistic is characterized as

$$R^2 = \frac{SSR}{SST}$$

where $SSR = \sum(\hat{y} - \bar{y})^2$ $and\ SST = \sum(y - \bar{y})^2$

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63) www.abjournals.org

RESULTS AND DISCUSSION

This section is divided into two parts; (1) Second order regression with/without interaction for small sample size n< 30. (2) Second-order regression with/without interaction for large sample size n> 30. That is two illustrations

Illustration 1

The data of Small sample size with an unemployment rate as the response (y), inflation rate (x_1), and exchange rate (x_2) are the predictors from 2007 to 2018 (i.e. p = 2) were collected. The estimate of the three models with/without interaction parameters (i.e. Centered on Mean, Median, and Mode models) are shown in Tables 1 and 3. The information criterion parameters (R^2, R^2 ⁻Adjusted, MSE, AIC, and SBC) are shown in Table 2 and 4 respectively

Table 1: Estimated Parameters for second- order models without interaction (Small Sample Data)

		Centre on mean		Centre on median		Centre on mode	
Variable	*Parameters*	*Coefficients*	*P-value*	*Coefficients*	*P-value*	*Coefficients*	*P-value*
Intercept	β_0	9.993553	0.000**	7.234447	0.000**	7.40567	0.000**
x_1	β_1	-0.38355	0.146	-0.57973	0.074	-0.51129	0.049*
x_2	β_2	0.086041	0.002**	0.075676	0.027*	0.075712	0.027*
$(x_1)^2$	β_{11}	-0.11406	0.073	-0.11406	0.043*	-0.11406	0.073
$(x_2)^2$	β_{22}	0.000179	0.409	0.000179	0.409	0.000179	0.4087

Footnote: **= sig. at 1%, *= sig. at 5%

Table 2: Second Order (without interaction) Models Adequacy Comparison (Small Sample Data)

criterion for model Selection	*Centre on mean*	*Centre on median*	*Centre on mode*
Total Parameters fits (k)	5	5	5
Number of significant parameters	2	3	3
R^2	92.99%	92.99%	92.99%
R^2 ⁻Adjusted	88.98%	88.98%	88.98%
MSE	3.6596	3.6596	3.6596
AIC	25.5684	25.5684	25.5684
SBC	27.9929	27.9929	27.9929

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

Table 3: Estimated Parameters for second- order models with interaction (Small Sample Data)

		Centre on mean		Centre on median		Centre on mode	
Variable	*Parameters*	*Coefficients*	*P-value*	*Coefficients*	*P-value*	*Coefficients*	*P-value*
Intercept	β_0	10.3380	0.0002**	7.3410	0.0001**	7.6109	0.0002**
X_1	β_1	-0.4275	0.1548	0.9311	0.2630	0.8139	0.2624
X_2	β_2	0.0840	0.0042**	0.0824	0.0405*	0.0807	0.0374*
$(X_1)^2$	β_{11}	-0.1944	0.2936	-0.1944	0.2936	-0.1944	0.2936
$(X_2)^2$	β_{22}	0.0001	0.6683	0.0001	0.6683	0.0001	0.6683
X_1X_2	β_{12}	0.0058	0.6312	0.0058	0.6312	0.0058	0.6312

Footnote: **= sig. at 1%, *= sig. at 5%

Table 4: Second Order (with interaction) Models Adequacy Comparison (Small Sample Data)

Criterion for Model Selection	*Centre on mean*	*Centre on median*	*Centre on mode*
Total Parameters fits (k)	6	6	6
Number of significant parameters	2	2	2
R^2	93.27%	93.27%	93.27%
R^2 -Adjusted	87.67%	87.67%	87.67%
MSE	4.0952	4.0952	4.0952
AIC	28.9176	28.9176	28.9176
SBC	31.8271	31.8271	31.8271

Illustration 2

A large sample size was collected from the University of Port Harcourt, Petroleum Department flow-rate on hydrated formation data sets. The Basic Sediment and Water (BSW) "Z_1", Gas Oil Ratio (GOR)"Z_2", Well Head Pressure (WHP)"Z_3", Well Head Temperature (WHT)"Z_4" (i.e. p = 4) are the predictors and Flow-Rate of Oil (BOPD) as the response "y".

The estimates of the three models with/without interaction parameters (i.e. Centered on Mean, Median, and Mode) are shown in Table 5 and 7 respectively. Also, the computed criterion for model selection(R^2, R^2 -Adjusted, MSE, AIC, and SBC) are shown in Table 6 and 8 respectively.

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

Table 5: Estimated Parameters for second- order models without interaction(Large Sample Data)

second order regression model with interaction	Centre on mean		Centre on median		Centre on mode	
	Coefficients	*P-value*	*Coefficients*	*P-value*	*Coefficients*	*P-value*
Intercept	5546.36	8.68E-34**	7302.836	2.98E-52**	6222.597	5.56E-28**
Z_1	-87.2263	1.06E-17**	-55.3705	0.001545**	-55.3705	0.001545**
Z_2	0	0	0.809075	0.283996	-2.89352	0.000563**
Z_3	0	0	32.76547	1.73E-09**	-4.64253	0.539812
Z4	122.7465	0	133.165	7.70E-07**	125.8102	1.29E-06**
z_1^2	0.197983	0.564792	-0.28452	0.349767	-0.28452	0.349767
z_2^2	-0.00295	0.0075**	-0.00104	0.013358*	-0.00104	0.013358**
z_3^2	0.001399	0.03097*	-0.40428	1.32E-05**	-0.40428	1.32E-05**
z_4^2	-6.65788	0.0027**	-3.67739	0.058001	-3.67739	0.058001

Footnote: ***= sig. at 1%, **= sig. at 5%

Table 6: Second Order (without interaction) Models Adequacy Comparison (Large Sample Data)

criterion for model Selection	Centre on mean	Centre on median	Centre on mode
Total Parameters fits (k)	9	9	9
Number of significant parameters	5	6	6
R^2	63.0%	73.3%	73.3%
R^{2-}Adjusted	63.0%	71.5%	71.5%
MSE	3986669	2927638	2927638
AIC	1978.60	1938.77	1898.41
SBC	2004.34	1938.77	1898.41

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

Table 7: Estimated Parameters for second-order models with interaction (Large Sample Data)

second-order regression model with interaction	Centre on mean		Centre on median		Centre on mode	
	Coefficients	*P-value*	*Coefficients*	*P-value*	*Coefficients*	*P-value*
Intercept	5283.798	6.00E-34**	7115.399	1.45E-35**	5986.128	4.28E-29**
Z_1	0	0	-68.3398	2.38E-05**	-20.9592	0.4265
Z_2	0	0	0.08446	0.9278	-0.66797	0.4542
Z_3	0	0	21.22672	0.0209*	-50.9722	0.00016**
Z4	0	0	153.0696	0.02512*	6.081218	0.96203
z_1^2	0.349861	0	0.272391	0.312136	0.272391	0.3121
z_2^2	-8.50E-05	0.946583	0.000194	0.685578	0.000194	0.6856
z_3^2	0.000112	0.872352	-0.25468	0.0684	-0.25468	0.0684
z_4^2	-2.98425	0.24025	-5.36896	0.0437*	-5.36896	0.0437*
Z_1Z_2	0.214659	9.23E-17**	0.001776	0.91275	0.001776	0.91275
Z_1Z_3	-0.15384	9.88E-16**	0.917855	0.0020**	0.917855	0.0020**
Z_1Z_4	-2.14911	0.001913**	1.745713	0.0623	1.745713	0.0623
Z_2Z_3	0	0	-0.02841	0.03392*	-0.02841	0.03392*
Z_2Z_4	-0.46052	0	-0.13018	0.1632	-0.13018	0.163267
Z_3Z_4	0.25031	1.47E-06**	2.077392	0.0031**	2.077392	0.0031**

Footnote: **= sig. at 1%, *= sig. at 5%

Table 8: Second Order (with interaction) Models Adequacy Comparison (Large Sample Data)

criterion for model Selection	Centre on mean	Centre on median	Centre on mode
Total Parameters fits (k)	16	16	16
Number of significant parameters	5	8	6
R^2	69.3%	83.3%	83.3%
R^{2-}Adjusted	62.8%	81.3%	81.3%
MSE	3388666	1920863	1920863
AIC	1971.64	1898.41	1898.41
SBC	2017.39	1944.17	1944.17

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

DISCUSSION OF RESULTS

Illustration 1(small sample size)

Tables 1 and 2 showed the results of the parameters estimated and model adequacy comparison (Centre on mean, median, and mode) of the second-order models without interaction when the data set is a small sample size. The second-order models without interaction developed as shown in Table 2 have an equal coefficient of determination R^2, R^2 – Adjusted, MSE and Model Adequacy criteria (AIC and SBC). The number of significant parameters centered on the mean is two (2), while the models centered on median and mode have three (3) parameters each that are significant at 1% and 5% respectively. Hence, the model centered on median and mode performed better than the model centered on mean in terms of significant parameters.

Similarly, Tables 3 and 4 showed the results of the parameters estimated and model adequacy comparison (Centre on mean, median, and mode) of the second-order models with interaction when the data set is a small sample size. The coefficient of determination R^2, R^2 –Adjusted, MSE and Model Adequacy criteria (AIC and SBC) with the number of significant parameters are all equal as shown in Table 4. This result suggested that all the models are equal base on all the estimations computed. It may be because the values for the mean, mode, and median were equal, thus making their deviations to be equal as well.

Illustration 2(large sample size)

Tables 5 and 6 showed the results of the parameters estimated and model adequacy comparison (Centre on mean, median, and mode) of the second-order models without interaction when the sample size is large. The second-order models built without interaction shows that the coefficient of determination R^2, R^2 –Adjusted, MSE and Model Adequacy criteria (AIC and SBC) and the number of significant parameters centered on median and mode performed better than the model centered on meanas shown in Table 6. The number of significant parameters for the model centered on mean is five (5) while the model centered on median and mode has six (6) parameters each that are significant at 1% and 5%. Hence, the model centered on median and mode performed better than the model centered on mean in terms of significant parameters and model adequacy comparison. Tables 7 and 8 showed the results of the parameters estimated and model adequacy comparison (Centre on mean, median, and mode) of the second-order models with interaction when the sample size is large. Table 8 showed that the coefficient of determination R^2, R^2 –Adjusted, MSE and Model Adequacy criteria (AIC and SBC) and the number of significant parameters centered on median and mode performed better than the model centered on the mean.

Comparison of the mean, median, and mode models with interaction based on parameters estimations (small sample size)

Table 3 shows that the slopes of exchange rate and inflation in a linear setting for median and mode models are better than the slopes of exchange rate and inflation for the mean model. The slope of the inflation rate of the mean model has a negative contribution to the unemployment rate in Nigeria, while it has a positive contribution to the unemployment rate in Nigeria for median and mode models. The exchange rate contributed more to the unemployment in the mean model than in the median and mode models

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

Comparison of the mean, median, and mode models without interaction based on parameter estimates (small sample)

Table 1 shows that the quadratic terms of inflation and exchange rates for mean, median, and mode models have equal contributions to the unemployment rate in Nigeria. The linear terms for inflation in the three models show negative contribution and the linear terms for the exchange rate in the three models are all positive to the unemployment rate, but the mode and median model contributions to unemployment are higher.

Illustration 2(large sample size)

Model without interaction

Table 6 shows that the second-order model centered on median and mode have an equal coefficient of determination R^2 which is higher than that of the model centered on mean and a lower AIC and SBC. The mean square value for the model centered on the median is equal to that centered on the mode. For the model centered on the mean, the Z_2 and Z_3 values which represent the coefficients of the linear terms are equal to zero (0). This indicates that the Z_2(GOR) and Z_3(WHP) have no contribution in the flow-rate on hydrate formation as expressed in the quadratic functions containing four predictors. These two variables Z_2(GOR) and Z_3(WHP) improved slightly in the quadratic terms, but their improvement was insignificant, the Z_2^2 showed a negative contribution while Z_3^2 showed a positive contribution.

Model with interactions

The value of the coefficient of determination R^2 for model centered on the median is equal to the value obtained for model centered on mode. The median and mode models are better in analysis than the mean model, which is evident in the value of R^2 in Table 8. The linear model parameters in Table 7 for the model centered on mean from Z_1 to Z_4 have no contribution to the fluid-flow rate on hydrate formation of the Niger Delta deep offshore field, wherein the linear model parameters for the model centered on median and mode have some amount of contributions to the fluid-flow. The interaction terms for all the three models contributed meaningfully to the fluid-flow rate on hydrate formation except for the interaction between Z_2Z_3, which represents GOR and WHP respectively for the model centered on mean in Table 7.

SUMMARY AND CONCLUSION

The better performance of the median and mode models to the contribution of the unemployment rate in Nigerian simply means that the mean model favors the employment rate in Nigeria. Thus, the higher the intercept of the models the higher the unemployment rate, the lower the intercept the higher the employment rate in Nigeria. The inflation rate has a negative value on the unemployment rate for the mean model. This means that the inflation rate for the mean model favors the employment rate, while the median and mode models favor the unemployment rate. Akaike's information criterion and Schwarz criterion for the mean model favor the employment rate over the unemployment rate for models with or without interaction. The model with interaction for AIC and SBC favor employment rate for

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

mean, median, and mode models. This is evident in the high value of AIC and BIC recorded in Table 4. The model of the median and mode are better in estimating the inflation rate, exchange rate, and unemployment rate in Nigeria than the mean model as shown in Table 3. The model centered on median proved to be the best in modeling the Qoil on the predictors in the flow-rate on hydrate formation, followed by the mode model and then the mean model. All the three models according to the value of $R^2$and the intercept terms are good for the analysis. The results showed that the GOR and WHP have no contribution in the model centered on mean in its linear terms, but had a slight improvement in the quadratic terms, which also showed insignificant contributions.

CONCLUSION

The linear term of the inflation rate for the mean model does not favor unemployment and the linear terms for GOR and WHP do not favor Qoil in the model centered on the mean. The quadratic term of the inflation rate for the mean, median, and mode models does not favor unemployment. Furthermore, the quadratic terms for GOR and WHP for all the three models do not favor Qoil. The interaction between inflation and exchange rate has values approximately equal to zero for all three models. This implies that the joint effect of inflation and the exchange rate does not affect the unemployment rate in Nigeria. The mean model without interaction performed better than the mean model with interaction while the median and mode model with interaction is better than the median and mode model without interaction. Hence, the second-order regression model centered on median and mode with or without interaction are better than the second-order regression model centered on the mean for large data.

RECOMMENDATIONS

1. The model without interaction should be built in the modeling of Nigeria's economy.
2. A higher-order model should be used to investigate the contributions of GOR and WHP for the flow-rate on hydrate formation.

REFERENCES

CBN (2017). Central Bank of Nigeria Annual report—December 2017 Retrieved March 29, 2020, from https://www. cbn.gov.ng/out/2017/fprd/fsr%20december%20 2017%20(2).pdf

CBN (2017). Nigerian Economy in relation to Global Economy report-December 2017 Retrieved March 29, 2020, from https://www. cbn.gov.ng/out/2017/fprd/fsr%20december%20 2017%20(2).pdf

Delphine C. and Olivier R. (2014): Robust analysis of the central tendency simple and multiple regression and ANOVA. A step by step tutorial. *International journal of psychological research 3*(1), 78-87.

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

Egbule J.F. (2008). Statistics for Researchers in the Behaviourial Sciences and Education Research-Oriented edition: Ethiope Publishing cooperation, Benin City, Edo State Nigeria, 135 – 147.

Iwundu, M. P. (2016b). On the behaviour of second-order N-point equiradial designs under varying model parameters. International Journal of Statistics and Applications, Vol. 6. No. 5, pg. 276-292.

Jeffrey A., McFarlane C., Vasudevan A. (2012). Rethinking enclosure: space, subjectivity and the commons, Antipode. 44 (4): 21247-1267.

Keller G. & Warrack B. (2003). Statistics for management and economics (6th ed). Cole: Thomson Brooks.

Kutner,M. H, Neter, J. Nachtsheim, C.J, & Wasserman, W. (2005). Applied linear statistical models. *International Journal of Production.*

Manikandan S. (2011). Measures of central tendency: The mean, Journal of Pharmacology & Pharmacotherapeutics, 140-142.

National Bureau of Statistics (2016). Data of inflation, exchange, and unemployment rate in Nigeria.

NBS. (2016). Inflation index: April 2016. Retrieved March 11, 2020, http://www.nigerianstat.gov.ng/download/518

Renaud, O., & Victoria-Feser, M.-P. (2009). A robust coefficient of determination for regression submitted. San Diego, CA: Academic Press.

Sameera (2014). Comparison between model with or without intercept. *Gen Math notes, 21*(1); 118-127.

Shalabh, J.S. (2012). Design of Experiments: an Introduction based on Linear Models.*Journal of the Royal Statistical Society*: Series A (Statistics in Society), 175, Issue 3.

Wilcox, R.R., Keselman, H.J. (2003). Modern robust data analysis methods: Measures of central tendency. Psychological Methods, 8, 254–274.

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 4, Issue 1, 2021 (pp. 47-63)
www.abjournals.org

APPENDIX A

Secondary data used for first illustration (small sample size)

YEARS	UNEMPLOYMENT RATE	INFLATION RATE	EXCHANGE RATE
2018	22.6	12.1	306.1
2017	17.5	16.5	305.8
2016	13.4	15.7	253.5
2015	9.0	9.0	192.4
2014	7.8	8.0	158.6
2013	10.0	8.5	157.3
2012	10.6	12.2	157.5
2011	6.0	10.8	153.9
2010	5.1	13.7	150.3
2009	4.9	12.5	148.9
2008	4.9	11.6	118.5
2007	3.5	5.4	125.8

Article DOI: 10.52589/AJMSS-71MQSBGZ
DOI URL: https://doi.org/10.52589/AJMSS-71MQSBGZ