International Journal of Science Academic Research
Vol. 02, Issue 05, pp.1478-1487, May, 2021
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

DISCRIMINATING BETWEEN SECOND-ORDER MODEL WITH VARIATION IN MODEL PARAMETERS BASED ON CENTRAL TENDENCY ESTIMATION

*ONU Obineke Henry, OWHONDAH Providence Samuel and NELSON Maxwell

Mathematics and Statistics Department, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria

Received 07th March 2021; **Accepted** 10th April 2021; **Published online** 24th May 2021

Abstract

Study of discriminating between second-order model with or without interaction for central tendency estimation was presented using ordinary least square for estimation of the model parameters. The research considered two different sets of data small sample size which is data of unemployment rate as response, inflation rate and exchange rate as the predictors from 2007 to 2018 and large sample size which is data of flow-rate on hydrate formation for Niger Delta deep offshore field. The R^2, AIC, SBC and SSE were applied for both data sets to test for the adequacy of the models. The small data was used as illustration 1 and the large data was used as illustration 2. It was revealed that, model centered on mean with interaction proved better than mean model without interaction. Median and mode values were found to be equal, as a result, the estimates of the median models were equal to the estimates of the mode models in all cases for both large and small data. The models centered on median and mode with interaction were better than those without interaction for both illustrations. Mean and mode models with or without interactions were found better than the mean models with or without interactions for both illustrations. The joint effect of inflation and exchange rate were found to be insignificant to the unemployment rate in Nigeria, while that of the interaction in median and mode models were seen better than that of mean model with a percentage difference of 57.1 for models with interaction. The intercept for mean model with interaction was found less than that of median and mode models with a percentage difference of 61 approximately. The intercept for median and mode model without interaction are better than mean model with a percentage difference of 55.8.

Keywords: Second-order model, Central tendency, Interaction model, Without interaction model.

INTRODUCTION

A simple linear regression is an approach in statistics According to Shalabh (2012) that is employed in the modeling of a linear surfaces. Regression analysis can be linear, nonlinear, second-order (quadratic or polynomial) regression. The model that is linear or nonlinear have been a major problem to decide as many 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 it is polynomial. All of the above definitions and classifications of a regression model is now misleading. Currently, a regression model is linear when it is linear in parameters, irrespective of the fact that it is linear, quadratic or polynomial or not. See literature review of this work. Models having any of its variables with power greater than two is called polynomial regression. The methods for fitting linear regression with the highest power of the unknown one (1) is also used for the model with highest power of the unknown greater than one (1).

RESPONSE SURFACE METHODOLOGY

Response surface methodology is defined as the collection of some scientific methods in building models for the purpose of experimentation. In model building or design of experiment, the experimenter's interest is generally aimed at maximizing the gain and minimizing the cost of running a system. The response variable usually depends on the amount of the explanatory variables that were invested in the system, hence response surface methodology can be used in wide range of fields which may include manufacturing, agriculture, government and multi-national companies etc, where some information as concerns the response are made available (Onu and Iwundu, 2017; Khuri and Mukhopadhyay, 2010). The response surface methodology is a generally used statistical and mathematical method for analyzing and modeling a process such that the response of interest is influenced by various variables and the objective of this method is to maximize or minimize (optimize) the response. The parameters that influence the process are called independent variables, while the responses are called dependent variables. For example, the hardness of a meat is affected by cooking time X_1 and cooking temperature X_2. The meat hardness can be changed under any combination of treatment X_1 and X_2. Therefore, time and temperature can vary continuously. If treatments are from a continuous range of values, response surface methodology is useful for improving, developing and optimizing the response variable. Comparison of second-order quadratic model of central tendency estimation with or without interaction have not be 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 or without intercept which is seen as full and reduced model, but the research centered its findings on a first-order linear regression model. All the above stated literatures fell short of using central tendency estimation for quadratic model, comparison with or without interaction. It is against this backdrop this work was presented. The work is aimed at comparing second order quadratic model with or without interaction using central tendency estimation for small and large sample sizes. The study is targeted at exposing the effect of omitting the interaction term in a quadratic model with central tendency

Corresponding Author:** ***ONU Obineke Henry,
Mathematics and Statistics Department, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria.

estimation and this will be found helpful to researchers as Akaike's information criterion (AIC), the Schwarz' Bayesian criterion (SBC) will be applied to study the model specification. They provide an alternative to the adjusted coefficient of multiple determination $R^2\alpha,\rho$ since $R^2\alpha,\rho$ is a model selection criterion that does not agree with models having large numbers of parameters. It is important to apply AIC and SBC and search for models that have small value of AIC and SBC. 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 study will apply a data of unemployment rate, exchange rate and inflation from 2007 to 2018 used as illustration 1 (small data). Also, the data of flow-rate on hydrate formation will be used as illustration 2 (large data). The secondary data of unemployment rate, inflation rate and exchange rate used was obtained from Central Bank of Nigeria, Statistical Bulletin (2017) and National Bureau of Statistics (2017). While the flow-rate on hydrate formation was from Niger Delta deep offshore, obtained from University of Port Harcourt Petroleum Department. It consists of four predictors and one response. The inflation rate shall be the x_1, the exchange rate shall be the $x_2$and y shall be the unemployment rate.

Inflation rate

Nigeria is currently experiencing high inflation level which grew to 13.7 percent in April 2016 and it is 0.9 percent higher than the previous month which was 12.8 percent. Scarcity of petroleum products are the severe drivers of cost-push inflation, which in turn forces the increase in transportation costs and therefore causes arbitrary increase in the cost of all other goods and services consistently for several months. Keeping the prices of goods and services stable at some rates that would not be harmful to the economic system is one of the fundamental goals of a modern economic system.

Exchange Rate

Crude oil provided approximately 90 percent of Nigeria's foreign exchange earnings, about 80% of federal revenue and contributes to the growth rate of Gross Domestic Product (GDP). Since oil contributes to 90 percent of foreign exchange, the fall in oil price affected foreign exchange, which devalues naira. Nigeria imports most of its consumable items, including refined petroleum, food items, raw materials and spare parts. The masses are bearing the burden of the increase in prices of imported goods and services in the form of high inflation. The government in trying to embark on a policy to control foreign exchange affected some firms, which led to their closure. NBS (2016) reported that in the second quarter of 2016, the nation's Gross Domestic Product (GDP) decline by -2.06 (year on year) in real terms.

Unemployment Rate

The international Labor Organization (1982) defined unemployment as group of persons who are without job from government, private or self-employed with a defined age level that are ready to work. These set of persons have made efforts to get work, but could not. Central tendency defined in Manikandan (2011) as a statistical means used in identifying a unique value as a representative of a whole distribution. It gives an accurate description of the data. This measure is used in making comparison of data. Nwagozie (2011) viewed central tendency as the average value in a given set of values when arranged in ascending or descending order of their magnitude. The average is a value representing a set of data and this is because, it lies centrally within the set of data when arranged in order. Also see Keller and Warrack (2003). The central tendency can be measured by the arithmetic mean test statistics represented by $\bar{x}$, the median test statistic represented in this research as x' and the mode test statistic represented by x'' others may include the geometric mean, harmonic mean and the root mean square test statistic. Each of these averages offer some level of advantages as well as disadvantages in any research as seen in Kutner *et al.* (2005), Manikandan (2011), Keller and Warrack (2003) and Egbule (2008) as seen in Shalabh (2012) and Kutner *et al.* (2005). Sameera (2014) studied the comparison between models with intercept term and that without intercept term in a linear process and here leverage point was applied and it was observed that evaluating the leverage in the new points was equal to the evaluation when the linear regression model was forced through the origin, i.e$m_0 = 0$ in the full model. This was achieved by augmenting the data. For augmentation of design point, see Iwundu and Onu (2017). It was also discovered that intercept was significant in the full model, but it becomes insignificant when the leverage point was added, thereby forcing the model through the origin.

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.

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + e_i$$

The second order regression model without interaction is given as.

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

The first order regression model with interaction is given as.

$$Y = \beta_O + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e_i$$

The second order regression model with interaction is given as

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

The center of mean, median and mode are as follows

For mean

Where $x_1 = (x_1 - \bar{x}_2)$

$x_2 = (x_2 - \bar{x}_2)$

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:

$$Y = \beta_O + \beta_1\left(x_i - \bar{x}\right) + \beta_2\left(x_i - \bar{x}\right)^2 + e_i \quad (1)$$

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

$$Y = \beta_O + \beta_1\left(x_i - x'\right) + \beta_2\left(x_i - x'\right)^2 + e_i \quad (2)$$

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

$$Y = \beta_O + \beta_1\left(x_i - x''\right) + \beta_2\left(x_i - x''\right)^2 + e_i \quad (3)$$

The quadratic model with interaction for mean is given as.

$$Y = \beta_O + \beta_1\left(x_1 - \bar{x}_1\right) + \beta_2\left(x_2 - \bar{x}_2\right) + \beta_{12}\left(x_1x_2 - \overline{x_1x_2}\right) + \beta_{11}\left(x_1 - \bar{x}_1\right)^2 + \beta_{22}\left(x_2 - \bar{x}_2\right)^2 + e_i \quad (4)$$

Also, quadratic model with interaction for median is given as:

$$Y = \beta_O + \beta_1\left(x_1 - x_1'\right) + \beta_2\left(x_2 - x_2'\right) + \beta_{12}\left(x_1x_2 - x_1'x_2'\right) + \beta_{11}\left(x_1 - x_1'\right)^2 + \beta_{22}\left(x_2 - x_2'\right)^2 + \ell_i \quad (5)$$

The quadratic model with interaction for mode is given as:

$$Y = \beta_O + \beta_1\left(x_1 - x_1''\right) + \beta_2\left(x_2 - x_2''\right) + \beta_{12}\left(x_1x_2 - x_1''x_2''\right) + \beta_{11}\left(x_1 - x_1''\right)^2 + \beta_{22}\left(x_2 - x_2''\right)^2 + \ell_i \quad (6)$$

Illustration 2 (large sample size)

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.

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \ell_i$$

The second order regression model without interaction is given as.

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_3^2 + \beta_{44}x_4^2 + \ell_i$$

The first order regression model with interaction is given as

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{12}x_1x_2$$
$$+ \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4\ell_i$$

The second order regression model with interaction is given as

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{33}x_{3i}^2$$
$$+ \beta_{44}x_4^2 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{14}x_1x_4 + \beta_{23}x_2x_3 + \beta_{24}x_2x_4 + \beta_{34}x_3x_4 + \ell$$

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 are

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

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 are

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

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 are

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

The flow-rate on hydrate formation data sets, using production data from Niger Delta deep offshore field was obtained from the University of Port Harcourt petroleum department. A base line model was developed that will define the multiple linear regression relationship, interaction between all the variables causing hydrate formation and Cobb Douglass model was fitted. This baseline model (multiple linear regression) will now be used in determining the needed variations to be made on that field to effectively manage hydrate before agglomeration to the point of creating a blockage. The main variables are Qoil which means the flow-rate of oil and it is the response variable. The predictors are BSW, GOR, WHP and WHT. The statistical software to be applied in this study is the Microsoft-Excel.

Estimating the Mean for Ungrouped Data

All the processes above on mean are followed with median and mode, just that for median model, the mean $\bar{x}$ is replaced with x'' and for mode model the mean $\bar{x}$ is replaced by x'

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

$x_1x_2, \ldots, x_n,$

The mean $\bar{x}$ = summing all the individual values up to the nth value divided by the total observations. It is given as expressed in Nwagozie (2011), Manikandan (2011) and Keller and Warrack (2003) as,

$$\therefore \text{ Mean } \bar{x} = \frac{x_1 + x_2 + \ldots x_n}{n} \quad (7)$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (8)$$

For a frequency distribution ungrouped,

$$\text{mean} = \frac{\sum f(x_i)}{\sum f} \quad (9)$$

where x_i represent $x_i, x_2 \ldots, x_n$

Where $\sum$ is the Greek letter Sigma (summation), $x_i = x_1\ x_2, \ldots, x_i$

And n = number ofx_i 's

Estimating the Median for Ungrouped Data

As seen in Egbule (2008) and Manikandan (2011), the median of ungrouped data given as

$$x_1, x_2, x_3 x_4, x_5$$

is the middle value which is x_3 provided the data is orderly. If the data is given as $x_1, x_2, x_3 x_4$ the median is given as

$$\text{Median} = \frac{x_2 + x_3}{2} \quad (10)$$

For a frequency distribution ungrouped median = $\left(n/2\right)^{th}$ (11)

If n is even or $\left(\frac{n+1}{2}\right)^{th}$ if n is odd where n is the number of frequency.

Estimating the Mode for Ungrouped Data

Mode is defined as seen in Manikandan (2011) as the value that repeats itself most often in a data. For a frequency distribution ungrouped, given as:

$$x_1, x_2, x_4, x_2, x_3$$

The mode is x_2 , which occurred most in the data.

Obtaining parameter estimates for models of mean, median and mode

From the model of mean, median and mode and data of response variable $Y_1, Y_2, \ldots, Y_n$ and explanatory variables $x_1, x_2, \ldots, x_n$, we form the system of equations as seen

$$Y_1 = \beta_o + \beta_1 r_1 + \beta_2 r_1 + e_1$$
$$Y_2 = \beta_o + \beta_1 r_2 + \beta_2 r_2^{\,2} + e_2$$
$$Y_n = \beta_o + \beta_1 r_n + \beta_2 r_n^{\,2} + e_3$$

Where $r = \left(x_i - \bar{x}\right)$ and $r^2 = \left(x_i - \bar{x}\right),^2$ we put 3.1 in matrix form, we have

$$\underline{Y}\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \underline{x} = \begin{pmatrix} 1 & r_1 & r_1^{\,2} \\ 1 & r_2 & r_2^{\,2} \\ \vdots & \vdots & \vdots \\ 1 & r_n & r_n^{\,2} \end{pmatrix}$$

$$\underline{\beta} = \begin{pmatrix} \beta_o \\ \beta_1 \\ \beta_2 \end{pmatrix} \text{ and } \underline{e} = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

The above is in the form

$$\underline{Y} = \underline{\beta}\ \underline{x} + \underline{e} \quad (12)$$

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

$$x' = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ r_1 & r_2 & \ldots & r_n \\ r_2 & \vdots & \ldots & r_n^{\,2} \end{pmatrix} \quad (13)$$

This is obtained by interchanging the first column with the first now, the second column with the second row and so on. We obtain $x'Y$ The alias matrix is applied to obtain the parameters and is as expressed in Kutner*et al* (2005), Sameera (2014), Huaglin and Welsch (1978) Iwundu and Onu (2017). The least square equation as seen below is applied

$$\hat{\underline{\beta}} = \left(x'x\right)^{-1}\left(x'Y\right) \quad (14)$$

The determinant of $\left(x'x\right)$ is obtained as expressed in Odili (2000) and consequently, the inverse of $x'x$ obtained and given as -1

$$\left(x'x\right)^{-1} = \frac{Adj\left(x'x\right)}{\left|x'x\right|} \quad (15)$$

The matrix of inverse obtained from 3.4 and $x'Y$ are substituted into 3.11 to obtain $\hat{\underline{\beta}} = \begin{pmatrix} \beta_o \\ \beta_1 \\ \beta_2 \end{pmatrix}$

Similar approach is applied for the model of deviations and square deviations of media and mode.

Testing for model adequacy

AIC approach for mean, median and mode

The AIC is given as seen in Kutner et al (2005) as.

$$AIC\bar{x} = nInSSE\bar{x} - nInn + 2p \quad (16)$$

Were AIC representing the Akaike's Information Criterion for a mean model.

Note that, the first term is $nInSSE$, which becomes reduced as p increases. The second term is fixed for a given sample size n and the third term increases with the number of parameters P. The models with small SSE do well by this criterion as long as the penalties 2P for AICx is concerned. The smaller the value of AIC the better the model.

The AIC for median and modal model are obtained as explained above.

The sum of square error is given as

$$SSE = \sum_{i=i}^{n}\sum_{i=i}^{n}\left(xij - \bar{xj}\right)^2 \quad (17)$$

Schwarz' Bayesian criterion

It is given as

$$SBC_P = n\ln SSE - n\ln n + [\ln n]p \quad (18)$$

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 coefficient of determination, denoted by R^2 or r^2 and pronounced "R squared", is a number that indicates the proportion of the variance in the dependent variable that is predictable from the independent variable (Richard , 1994). R^2 is also a measure of the proposition of variability in the data set that is accounted for by a regression model. It assumes that every independent variable in the model helps to explain variation in the dependent variable (y) and thus gives the percentage of explained variation if all independent variables in the model affect the dependent variable (y). The R^2statistics is defined as

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

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

The R^2 lies between -1 and +1, if -1 there is a high negative relationship between the variables, if +1 there is a high positive relationship between the variables and if 0, there is no relationship between the variable(s). The second illustration in this work contains some terms like z1, z2, z3, z4, z1z1, z2z2, z3z3, z4z4 for parameters of model center on mean without interaction, w1,w2,w3,w4, w1w1, w2w2, w3w3, w4w4 also z1, z2, z3, z4, z1z1, z2z2, z3z3, z4z4 are parameters of the model center on median and p1,p2,p3,p4,p1p1,p2p2,p3p3,p4p4 are parameters of the model centered on mode all without interaction. Where z1,w1, and p1, z2,w2 and p2, z3,w3 and p3, z4,w4, and p4 are the parameters of the main effects, in other words, they are the parameters of the linear variables, devoid of interactions. Z1z1, w1w1 and p1p1, z2z2, w2w2, p2p2 up to the fourth z4z4, w4w4 and p4p4 are the parameters of the quadratic terms for models without interactions. While z1, z2, z3, z4, z1z1, z2z2, z3z3, z4z4, z1z2, z1z3, z1z4, z2z3, z2z4, z3z4, w1, w2, w3, w4, w1w1, w2w2, w3w3, w4w4, w1w2, w1w3, w1w4, w2w3, w2w4, w3w4 and p1, p2, p3, p4, p1p1, p2p2, p3p3, p4p4, p1p2, p1p3, p1p4, p2p3, p2p4, p3p4 are respectively the parameters for models centered on mean, median and mode. Z1z2, w1w2, p1p2 are the interaction between the variables 1 and 2 centered on mean, median and mode. So also the other.

The symbol *** indicates significant at 1%, ** indicates significant at 5%, and * indicates significant at 10%.

Application of analysis of variance (ANOVA)

Analysis of variance popularly known as ANOVA is applied to both the model with intercept and that without intercept. We obtain the sum of squares of the regression, between treatment, error sum of square and the sum of square total (Keller and Warrack, 2003). Sum of square treatment is the test statistic that is used to measure the similarities of the mean samples to each other. It is given as;

$$SS_{Treat} = \sum_{i=1} n_i\,(\bar{x}_i - \bar{\bar{x}})^2$$

If a large difference is experienced in the between treatment means known as the sum of square treatment, it means that one and above sample means will considerably differ from the grand mean as seen in Keller and Warrack (2003) in order to know whether or not to reject the null hypothesis, it is advisable to know how much variation that exist within treatments variation, and this in order word called sum of square error denoted as SSE. It is given as;

$$SSE = \sum_{j=1}^{n}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$$

Which can also be written by expansion as;

$SSE = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 . \; . \; . + (n_K - 1)S_k^2$, this is as expressed in Keller and Warrack (2003), Nwaogazie (2011) and Egbule (2008). We proceed to computing the mean squares, for which mean square for treatment is obtain as;

$MS_{treat} = \frac{SS_{treat}}{n-1}$, that is to say, the sum of square treatment is divided by the number of treatments in the sample minus 1.
While mean square error

$$MSE = \frac{SSE}{\mathrm{N}-n} \quad (19)$$

Where N is the total sample and n is the number of treatments. We finally compute the F statistic given as

$$F = \frac{MS_{treat}}{MSE}$$

The hypothesis is built and conclusion drown as shown;
$H_0: m_0 = 0$
$H_1: m_0 \neq 0$ (20)

The Fstatistic tell us whether the value of SS_{treat} is large such that the null hypothesis can be rejected.

We reject null hypothesis if

$$F > F\alpha, n-1, \mu - n \quad (21)$$

t test statistic was applied when interest shifted from observing whether their means μ_1 equal μ_2 or not to testing ifμ_1 is greater than μ_2 and verse versa.
The t statistic used in this research is given as;

$$t = \frac{\widehat{m}_0}{S(\widehat{m}_0)} \qquad (22)$$

Where $\widehat{m}_0$ is the estimate of the intercept term in a model with intercept and $S(\widehat{m}_0)$ is the standard deviation of the intercept term. For slope term $\widehat{m}_1$, the T statistic is given as;

$t = \frac{\widehat{m}_1}{S(\widehat{m}_1)}$ (Kutner *et al.,* 2005)

But

$$S(\widehat{m}_0) = MSE\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum(xi-\bar{x})^2}\right] \qquad (23)$$

According to Kutner *et al.* (2005).

$$SSE = \sum (y_i - \hat{y}_1)^2$$

$$SStotal = \sum (y_i - \bar{y})^2$$

And

$$SSR = \sum (\hat{y}_1 - \bar{y})^2$$

And

$$MSE_{Error} = \frac{\sum(y_i - \hat{y}_{i)}}{n-2} = \frac{SSE}{n-2}$$

$$MSE_{Reg} = \frac{\sum(\widehat{y_1} - \bar{y}_{i)}}{n-2} = \frac{SSE}{1} = SSR$$

$$MSE_{totla} = \frac{\sum(y_i - \bar{y})^2}{n-1} = \frac{SSE_{total}}{n-1}$$

A typical example of a one-way ANOVA is as shown

Table 1. One Way ANOVA

Source of Variation	Df	SS	MS	Fcal
Treatment (B/W)	(k-1)	SS_{treat}	MS_{treat}	$\frac{MS_{treat}}{MSE}$
Error (within)	$(\mu - k)$	SSE	MSE	
Total	$(\mu - k)$	SST		

RESULTS AND DISCUSSION

Estimating parameters for models centered on Mean, Median and Mode with interaction

Estimating the parameters of the model centered on Mean with interaction using MATHLAB

We have the 12x6 matrix inputted as seen

$$X = \begin{bmatrix} 1 & 0.5 & 120.17 & 60.1 & 0.25 & 14440.8 \\ 1 & 5.5 & 120.17 & 660.9 & 30.25 & 14440.8 \\ 1 & 4.5 & 68.17 & 306.8 & 20.25 & 4647.1 \\ 1 & -2.5 & 6.17 & -15.4 & 6.25 & 38.1 \\ 1 & -3.5 & -26.83 & 93.9 & 12.25 & 719.8 \\ 1 & -2.5 & -28.83 & 72.1 & 6.25 & 831.2 \\ 1 & 0.5 & -27.83 & -13.9 & 0.25 & 774.5 \\ 1 & -0.5 & 31.83 & 15.9 & 0.25 & 1013.1 \\ 1 & 2.5 & -35.83 & -89.6 & 6.25 & 1283.8 \\ 1 & 1.5 & -36.83 & -55.2 & 2.25 & 1356.4 \\ 1 & 0.5 & -66.83 & -33.4 & 0.25 & 4466.2 \\ 1 & -6.5 & -59.83 & 388.9 & 42.25 & 3579.6 \end{bmatrix}$$

The inputted matrix is displayed with each value multiplied by 1. 10e+004 as seen

$$X = \begin{bmatrix} 0.0001 & 0.0001 & 0.0120 & 0.0060 & 0.000 & 1.4441 \\ 0.0001 & 0.0006 & 0.0120 & 0.0661 & 0.0030 & 1.4441 \\ 0.0001 & 0.0004 & 0.0068 & 0.0307 & 0.0020 & 0.4647 \\ 0.0001 & -0.0003 & 0.0006 & 0.0015 & 0.0006 & 0.0038 \\ 0.0001 & 0.0004 & -0.0027 & 0.0094 & 0.0012 & 0.0720 \\ 0.0001 & 0.003 & 0.0029 & 0.0072 & 0.0006 & 0.0831 \\ 0.0001 & 0.0001 & -0.0028 & -0.0014 & 0.0000 & 0.0775 \\ 0.0001 & -0.0001 & 0.0032 & 0.0016 & 0.0000 & 0.1013 \\ 0.0001 & 0.0003 & -0.0036 & -0.0090 & 0.0006 & 0.1284 \\ 0.0001 & 0.0001 & -0.0037 & -0.0055 & 0.0002 & 0.1356 \\ 0.0001 & 0.0001 & -0.0067 & -0.0033 & 0.0000 & 0.4466 \\ 0.0001 & -0.0006 & -0.0060 & -0.0389 & 0.0042 & 0.3580 \end{bmatrix}$$

We transpose matrix X to obtain X'

We standardize the matrix, by multiplying X' by X, to make it a square matrix, we have

$X'X = X'* X$

We then have

$$X'X = \begin{bmatrix} 0.0000 & 0 & 0.0000 & 0.0000 & 0.0000 & 0.0005 \\ 0 & 0.0000 & 0.0000 & 0.0000 & -0.0000 & 0.0009 \\ 0.0000 & 0.0000 & 0.0005 & 0.0009 & 0.0000 & 0.0315 \\ 0.0000 & 0.0000 & 0.0009 & 0.0071 & 0.0004 & 0.1302 \\ 0.0000 & -0.0000 & 0.0000 & 0.0004 & 0.0000 & 0.0071 \\ 0.0005 & 0.0009 & 0.0315 & 0.03102 & 0.0071 & 4.7775 \end{bmatrix}$$

The determinant of the standardized matrix is obtained as seen

$|X'X|$=det($X'X$)

We have

$|X'X|$= 4.1367e+023

The inverse of the matrix is obtained as

$(X'X)^{-1}$ =inv($X'X$)

We have

$$(X'X)^{-1} = \begin{bmatrix} 0.2818 & -0.0108 & 0.0007 & 0.0012 & -0.0190 & -0.0000 \\ -0.0108 & 0.0139 & -0.0001 & -0.0002 & 0.0033 & -0.0000 \\ 0.0007 & -0.0001 & 0.0001 & -0.0000 & 0.0003 & -0.0000 \\ 0.0012 & -0.0002 & -0.0000 & 0.0000 & -0.0003 & -0.0000 \\ -0.0190 & 0.0033 & 0.0003 & -0.0003 & 0.0047 & 0.0000 \\ -0.0000 & -0.00000 & -0.0000 & -0.0000 & 0.0000 & 0.00000 \end{bmatrix}$$

The response variable (unemployment rate) was inputted as seen

Y=[22.6; 17.5; 13.4; 9; 7.8; 10; 10.6; 6; 5.1; 4.8; 4.4; 4]

We multiply the transpose of X by Y, each value is multiplied by 1.0e+005, it was inputted as seen

$X'Y$=X'*Y

We have
$X'Y =$

The parameters were estimated using the least square equation as seen

$\underline{\hat{\beta}}=(X'X)^{-1}*X'Y$

We have

$$\underline{\hat{\beta}}=\begin{bmatrix} 8.5830 \\ -0.2955 \\ 0.0592 \\ 0.0032 \\ -0.1177 \\ 0.0004 \end{bmatrix}$$

- Estimating the parameters of the model centered on Median with interaction using MATHLAB
- The approaches in 4.1 shall be followed strictly as seen in appendix B and the result is shown in table (a)
- Estimating the parameters of the model centered on Mode with interaction using MATHLAB
- The approaches in 4.1 shall be followed strictly as seen in appendix B and the result is shown in table (a)

Estimating parameters for models centered on Mean, Median and Mode without interaction

- Estimating the parameters of the model centered on Mean without interaction using MATHLAB
- The approaches in above sections shall be followed strictly as seen in appendix C and the result is shown in table (a)
- Estimating the parameters of the model centered on Median without interaction using MATHLAB
- The approaches in above sections shall be followed strictly
- Estimating the parameters of the model centered on Mode without interaction using MATHLAB
- The approaches in above sections shall be followed strictly as seen in appendix C and the result is shown in table (a)

Model Adequacy Approach for Mean, Median and Mode with interaction

Coefficient of Determination R^2 for mean, median and mode model with interaction

The R^2 for model centered on mean is given as

$$R^2 = \frac{1808.926}{3638.801} = 0.4970$$

Where SSR=1808.926, SSE= 1829.875, but
SST= SSR+SSE
SST=1808.926+1829.875=3638.801

It follows strictly for median and mode model and in each case, we obtained

$R^2 = 0.4997$

Akaike's information criterion for mean, median and mode model with interaction

For mean model using Akaike's information criterion given as

$$AIC_P = n \ln SSE_P - n \ln n + 2P$$

$$AIC_6 = 12 \ln(1829.875) - 12 \ln 12 + 2 \times 6$$

=72.3251

For median and mode model with interaction, it follows from above, we obtain in each case the value of 64.4843

Schwarz' Bayesian criterion for mean, median and mode model with interaction

We have for mean median and mode model respectively as 75.2345, 67.3937,67.3937.

Model Adequacy for Mean, Median and Mode without interaction

Coefficient of Determination R^2 for mean, median and mode model without interaction

The R^2 for model centered on mean, median and mode are respectively

0.5424,0.5494, 0.5494.

Akaike's information criterion for mean, median and model without interaction

For mean, median and mode model without interaction we obtain the values respectively as

67.9749, 62.7825, 62.7825

Schwarz' Bayesian criterion for mean and median model without interaction

Using the Schwarz' Bayesian criterion for mean, median and mode model without interaction, we obtain

=70.3992,65.207, 65.207.

Illustration 2

Large sample size from the flow-rate on hydrated formation data sets from University of Port Harcourt Petroleum Department.

Estimating parameters for models centered on Mean, Median and Mode without interaction

Parameters for model centered on Mean without interaction
The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.793746
R Square	0.630032
Adjusted R Square	0.595444
Standard Error	1996.664
Observations	129

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	8	8.28E+08	1.04E+08	34.62644	9.16E-28
Residual	122	4.86E+08	3986669		
Total	130	1.31E+09			

Parameters for model centered on Median without interaction
The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.856017
R Square	0.732766
Adjusted R Square	0.71495
Standard Error	1711.034
Observations	129

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	8	9.63E+08	1.2E+08	41.1305	6.53E-31
Residual	120	3.51E+08	2927638		
Total	128	1.31E+09			

Parameters for model centered on Mode without interaction
The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.856017
R Square	0.732766
Adjusted R Square	0.71495
Standard Error	1711.034
Observations	129

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	8	9.63E+08	1.2E+08	41.1305	6.53E-31
Residual	120	3.51E+08	2927638		
Total	128	1.31E+09			

Estimating parameters for models centered on Mean, Median and Mode with interaction

Parameters for model centered on Mean with interaction
The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832623
R Square	0.693261
Adjusted R Square	0.628045
Standard Error	1840.833
Observations	129

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	14	9.11E+08	65099077	29.88351	1.16E-31
Residual	119	4.03E+08	3388666		
Total	133	1.31E+09			

Parameters for model centered on Median with interaction
The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.912924
R Square	0.833431
Adjusted R Square	0.812975
Standard Error	1385.952
Observations	129

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	14	1.1E+09	78261422	40.74284	9.94E-38
Residual	114	2.19E+08	1920863		
Total	128	1.31E+09			

3. See Appendix G for the detailed analysis.
4. **4.6.3** Parameters for model centered on Mode with interaction
5. The excel analysis is shown below

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.912924
R Square	0.833431
Adjusted R Square	0.812975
Standard Error	1385.952
Observations	129

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	14	1.1E+09	78261422	40.74284	9.94E-38
Residual	114	2.19E+08	1920863		
Total	128	1.31E+09			

DISCUSSION OF RESULTS

Illustration 1(small sample size)

Comparison of the mean, median and mode models with interaction on the basis of parameters estimations

The intercept for mean model is less than the intercept of the median and mode models with about 61 percent difference. It is also observed that the slopes of exchange rate and inflation in a linear setting for median and model models are better than the slopes of exchange rate and inflation for mean model. The contributions of the interaction of exchange rate and inflation is approximately zero (0), but that of median and model models are equal but better than the mean model with a percentage difference of 57.1. Small value of the interactions for the both models suggest that, it is better to analyze the data with models without interaction. The slope of inflation rate of mean model has negative contribution to the unemployment rate in Nigeria, while it has positive contribution to the unemployment rate in Nigeria for median and mode models. The exchange rate contributed more to the unemployment in mean model than in median and mode models. The parameters for median model and mode model are equal, since the value of the mode and the value of median were found equal, hence making their deviations to be equal as well.

Comparison of the mean, median and mode models without interaction on the basis of parameter estimates

The intercept for median and mode model are better than the intercept for mean model with a percentage difference of 55.8.

The quadratic terms of inflation and exchange rates for mean, median and mode model show approximately equal contribution 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 unemployment rate, but the mode and median models contributions to unemployment are higher. The mode and median models gave equal parameters in all the estimations.

Comparison of the mean, median and mode models on the basis of AIC and SBC

The AIC for median and mode models are equal and less than that obtained in the mean model, meaning that, the median and mode models favors unemployment rate for models with interactions. Also, the SBC for median and mode models are equal and less than that obtained in the mean model, meaning that, the median and mode models favors unemployment rate for models with interactions. For models without interaction, the inflation and exchange rates for median and mode models contributes more to the unemployment rate as revealed by the values of AIC and SBC.

Illustration 2(large sample size)

Model without interaction

The second-order model centered on median has equal coefficient of determination R^2 with the second-order model centered on mode, but it was found greater than the second-order model centered on mean by 14.02%. the intercept or the grand mean for the second-order response function centered on median is greater than that centered on mode and also greater than that centered on mean by 14.79% median-mode model, 24.05% median-mean model and 10.87% mode-mean model. The second-order regression equation centered on median and that centered on mode are better than that centered on mean, as revealed by the values of the mean squares. Since the higher the mean square value, the better the model. Consequently, the mean square value for model centered on median is equal to that centered on mode. For model centered on mean, the z2 and z3 values which represents the coefficients of the linear terms for variable 2 and 3, are equal to zero (0). This indicates that the z2(GOR) and z3(WHP) have no contribution in the flow-rate on hydrate formation as expressed in the quadratic functions containing four predictors. These two variables z2(GOR) and z3(WHP) improved slightly in the quadratic terms, but their improvement was insignificant, the z2z2 showed a negative contribution while z3z3 showed a positive contribution. The parameter w1 for the model centered on median is equal to the parameter p1 for the model centered on mode and all the parameters of the quadratic terms for both models centered on median and that centered on mode are all equal. This shows that the estimation of model parameters for a quadratic model centered on median has same estimation to that of quadratic model centered on mode. That is to say, there exists greater agreement between median model and mode model, than for mean model. These results are agreeing with the result of the illustration 1, for small sample size.

Model with interactions

The value of coefficient of determination R^2 for model centered on median is equal to the value obtained for model centered on mode. The model centered median has greater R^2 than the model centered on mean with the percentage of 16.81. The intercept term for model centered on median is greater than that centered on mode and that centered on mode is greater than that centered on mean, with the percentages 15.87 for median-mode, 25.74 for median-mean and 11.73 for mode-mean model. The median and mode models are better in analysis than mean model, which is evident by the in the value of R^2 in the ANOVA. The linear model parameters for model centered on mean from z1 to z4 have no contribution on the fluid-flow rate on hydrate formation of the Niger Delta deep offshore field, where the linear model parameters for model centered on median and mode have some amount of contributions to the system under study. The interaction terms for all the three models contributed meaningfully to the fluid-flow rate on hydrate formation except for the interaction between variables 2 and 3 written as z2z3, which represents GOR and WHP respectively for model centered on mean which has no contribution to the fluid-flow rate on hydrate formation. The quadratic terms of the models centered on median and mode yielded equal parameters, while the model centered on mean differs.

Summary

The analysis and the discussions carried out so far lead to summarizing as follows;

1. The better performance of the median and mode models to the contribution of unemployment rate in Nigerian simply means that the mean model favors the employment rate in Nigeria.
2. It is hereby stated that the higher the intercept of the models the higher the unemployment rate, the lower the intercept the higher the employment rate in Nigeria.
3. The inflation rate having a negative value on the unemployment rate for mean model and positive value for median and mode models means that the inflation rate for mean model favors employment rate, while the median and mode models favor the unemployment rate.
4. The Akaike's information criterion and the Schwaz' criterion for mean model favor employment rate than the unemployment rate both for models with or without interaction.
5. The models with interaction for AIC and SBC favor employment rate both for mean, median and mode models. This is evident in the high value of AIC and BIC recorded in table 4.2.
6. The model of mean with interaction is better than that of mean without interaction and the model of median and mode with interaction are better than those without interaction for the study of inflation and exchange rate on the unemployment rate in Nigeria, as revealed by AIC, SBC and the SSE.
7. The model of the median and mode are better in estimating the inflation rate, exchange rate and the unemployment rate in Nigeria than the mean model.
8. 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.
9. The results showed that the GOR and WHP no contribution in the model centered on mean in its linear terms, but had slight improvement in the quadratic terms, which also showed insignificant contributions.

CONCLUSION

We conclude from the analysis, that the linear term of the inflation rate for mean model does not favor unemployment and the linear terms for GOR and WHP does not favor Qoil in model centered on mean, also, the quadratic term of the inflation rate for both mean, median and mode model does not favor unemployment, the quadratic terms for GOR and WHP for all the three models does not favor Qoil. The interaction between inflation and exchange rate has values approximately equal to zero for all the three models, this shows that the joint effect of inflation and exchange rate has approximately equal to zero value to the unemployment rate in Nigeria. Also, the quadratic term of the exchange rate has value almost equal to zero. Generally, the mean model without interaction proved better than mean model without interaction and the median and mode model with interaction are better than the median and mode model without interaction. Finally, Median and mode models with or without interaction are better than the mean model with or without interaction.

Recommendation

It was recommended that

1. The model without interaction should be built while modeling the economy of Nigeria.
2. The mean model is not a good fit for modeling the unemployment rate in Nigeria.
3. A higher order model should be used to investigate the contributions of GOR and WHP for the flow-rate on hydrate formation.

REFERENCES

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.*

Hoaglin D. C. and Welsch R.E. 1978. The hat matrix in regression and ANOVA: The American statistician 32(1) – 16 – 21.

Keller G. and Warrack B. 2003. Statistics for management and economics (6^{th}ed): *Thomson Brooks /Cole – 603-641.*

Kutner M. H., Nschtsheim C. J., Neter J. and Li W. 2005. Applied linear statistical model, fifth edition, McGraw-Hilit:a Irwin, Boston Burr RIdge, IL Dubuque, IA MadIson, WI New York San FrancIsco St LoUIs Bangkok Bogota Caracas Kuala Lumpur LIsbon London MadndMexIcoCItyMIlan Montreal New DeihlSantIago Seoul Smgapore Sydney TaIpeI Toronto

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

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

Nwaogazie I.L. 2011. Probability and statistics for science and engineering practice. *De-Adroit Innovation 13 AnnangStr.Ogui N/L out Enugu-101-115.*

Onu O.H. and Iwundu M.P. 2017. Preferences of equiradial designs with changing axial distances, design sizes and increase center points and their relationship to the N-point central composite design: *International journal of advanced statistics and probability, 5(2)-77-82.*

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

****