[Volume 4 issue 11 November 2016]
PageNo.1650-1658 ISSN : 2320-7167

INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTER RESEARCH

On Comparative Study of Efficiency of Standardized and Unstandardized Regression Model for the Effect of Climatic Variability on Maize Production in Nigeria. (Forest Region)

Atanlogun S.K, Edwin O.A, and Omotosho S.A

Mathematics and Statistics Department, Rufus Giwa Polytechnic, Owo, Nigeria.

Abstract:

In this study, USMLRM and SMLRM were compared. To achieve the goal, USMLRM and SMLRM estimating techniques were considered on simultaneous equation models where maize production served as response variable and Rainfall, Temperature, Humidity served as explanatory variables. Data were collected from the Central Bank of Nigeria (CBN) Statistical Bulletin, December 2009 and the record office of Food Agriculture organisation (FAO) which covers the period of 1985 to 2007. Result from the analysis showed that USMLRM and SMLRM produced different values of coefficient and standard error in the two models. The SPSS statistical package was adopted to carry out the analysis of the results and the study however conclude that SMLRM was considered to be the most efficient model compared with that of USMLRM.

Keywords: Sub-Saharan Africa (SSA), Standardized Multiple Linear Regression Model (SMLRM), Unstandardized Multiple Linear Regression Model (USMLRM), Dependent Variable (DV), Independent Variables (IVs), Maize production (MP), Rainfall (RF), Temperature (Temp), Humidity (HUM).

Introduction:

Modelling the relationship between explanatory and response variables is a fundamental activity encountered in statistics. The statistical techniques for modelling and analysing variables called Regression analysis focus on the relationship between a response (outcome) variable and one or more explanatory (predictor) variables. Simple regression model entails the process of investigating the relationship between a single explanatory (predictor) variable and single response predictant variable. Multiple Regression model entails the regression of more than two variables; in this case we have one response variable and several explanatory variables. In the case of Standardized and Unstandardized regression model, standardized coefficients simply represent regression results with standard scores. Standardized coefficients are often called beta, beta weights, beta coefficient, or path coefficient in path analysis. While unstandardized coefficient simply represent regression result with unstandardized scores. These represent general result of regression analysis when the response variable and explanatory variables are yet to be converted to Z scores.

Efficiency is one of the properties that give a better description of modelling in statistics. In this study the output results of the standardized coefficients and the Unstandardized coefficients would be considered for observation. Hence the most efficient model would be determined.

Maize:

ize or corn is a cereal crop that is grown widely throughout the world in a range of agro ecological environments. More maize is produced annually than any other grain. About 50 species exist and consist of the different colours, texture grain shapes and sizes. White, yellow and red are the most common types. The white and yellow varieties are preferred by most people depending on the region. Maize is a leafy stalk whose kernel has seeds inside. It is an angiosperm, which means that its seed are enclosed inside a fruit or shell. It is used as a food staple by many people in Mexico, Central and South America and parts of Africa.

Importance:

Maize is the most important cereal crop in Sub-Saharan Africa (SSA) and an important staple food for more than 1.2 billion people in SSA and Latin America. All parts of the crop can be used for food and non-food products. In industrialized countries, maize is largely used as livestock feed and as a raw materials for industrial products. https://www.iita.org/maize.

Production:

Worldwide production of maize is 785 million tons, with the largest producer, the United State producing 42% Africa produces 6.5% and the largest African producer is Nigeria with nearly 8 million tons, followed by South Africa. Africa imports 28% of the required maize from countries outside the continent. https://www.iita.org/maize. Most maize production in Africa is rain fed. Irregular rainfall can trigger famines during occasional droughts.

Harvest:

According to 2007 FAO estimates 158 million hectares of maize are harvested worldwide. Africa harvests 29 million hectares with Nigeria, the largest producer in SSA, harvesting 3%, followed by Tanzania https://www.iita.org/maize.

Consumption:

Worldwide consumption of Maize is more than 116 million tons. Maize is processed and prepared in various forms depending on the country. Ground maize is prepared into porridge in eastern and southern Africa, while maize flour is prepared into porridge in West Africa. Ground maize is also fried or baked in many countries. In all parts of Africa, green (fresh) maize is boiled or roasted on its cob and served as a snack. Popcorn is also a popular snack. https://simple.m.wikipedia.org/wiki/maize.

MODEL SPECIFICATION:

Standardized Multiple Linear Regression Model (SMLRM):

In this type of model, we have the results represent what happens after all of the variables (predictors and outcome) have initially been converted into $Z-scores$ (formula).

Standardized coefficient simply represents regression results with standard scores. By default, most statistical software automatically converts both criterion (DV) and predictors (IVs) to $Z-scores$ and calculates the regression equation to produce standardized coefficients. Many staticians argue that standardized coefficient offer no or little advantage over unstandardized coefficients, and often offer confusing information. In some disciplines researchers routinely prefer standardized coefficient over Unstandardized because they believe that standardized coefficients are more interpretable, provide an assessment of predictor importance. The larger the standardized coefficient in absolute value is, the more important the predictor.

Standardized coefficients are dependent upon the sample standard deviation and if that value is inflated or deflated relative to the population standard deviation, then standardized coefficients will produce an incorrect inference for the population value. http://www.bwgriffin.com/gsu/course/edu8132/notes/king standardized coefficients.pdf.

A variable is said to be standardized if we subtract the mean value of its variable from its individual values and divide the difference by the standard deviation of that variable. Hence, we have $Z = \frac{X - M}{sd}$

Where X is the raw score, M is the mean and sd is the standard deviation.

Thus, the regression of Y on X, if we redefine these variables as

$$Y_i^* = \frac{Y_i - \overline{Y}}{S_Y}$$

$$X_i^* = \frac{X_i - \overline{X}}{S_X}$$

Where $\overline{Y}$ =Sample mean of Y, S_Y = Sample standard deviation of Y, $\overline{X}$ = Sample mean of X and S_X = Sample standard deviation of X; the variables Y_i^* and X_i^* are called standardized variables, with mean zero and variance one. However instead of running the regression $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_{p-1} x_{i,p-1} + \varepsilon_i$

we could run regression on the standardised variables as

$$Y_i^* = \beta_0^* + \beta_1^* x_{i1}^* + \beta_2^* x_{i2}^* + ... + \beta_{p-1}^* x_{i,p-1}^* + \varepsilon_i$$

Therefore we have

$$\beta_0^* = Y_i^* - \beta_1^* X_{i1}^* - \beta_2^* X_{i2}^* - ... - \beta_{p-1}^* X_{i,p-1}^*$$

Since $Y_i^* = X_{i1}^* = ... = X_{i,p-1}^* = 0$ (standardized variables has mean zero (0)) and $\beta_0^* = 0$

Hence we have

$$Y_i^* = \beta_1^* x_{i1}^* + \beta_2^* x_{i2}^* + ... + \beta_{p-1}^* x_{i,p-1}^* + \varepsilon_i$$

Is called a standardized multiple linear regression model with $P-1$ predictor variables. It can also be written as: $Y_i = \sum_{k=1}^{p-1} \beta_k^* X_{ik}^* + \varepsilon_i$

Since $E(\varepsilon_i) = 0$, the response function for SMLRM above becomes: $E(Y) = \beta_1^* x_1^* + \beta_2^* x_2^* + ... + \beta_{p-1}^* x_{,p-1}^*$

Interpretation:

β_1^* up to β_{p-1}^* are known in the literature as beta coefficient. The interpretation is that if the (standardized) regressor increases by one standard deviation, on average, the (standardized) regressand increases by β_1^* standard deviation units. Thus, unlike the traditional model, we measure the effect not in terms of the original units in which Y and X are expressed, but in standard deviation units.

Unstandardized Multiple Linear Regression Model (USMLRM):

For each predictor variable in a multiple regression analysis, the output will provide an Unstandardized regression coefficient (usually depicted with the letter B). Unstandardized results are probably more straight forward to understand. Unstandardized relationships are expressed in terms of the variables, original, raw units. These are considered when the coefficient simply represents results with unstandardized scores.

In this type of model, we have response variable Y which is determined by two or more predictor variables $X_1,\ldots,X_{p-1}$. Assume that a linear relationship exist between a response variable Y and X_{p-1} predictor variables.

"The regression model":

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$$

is called Unstandardized Multiple Linear Regression Model with $P-1$ predictor variables. It can also be written as:

$$Y_i = \beta_0 + \sum_{k=1}^{p-1} \beta_k x_{ik} + \varepsilon_i$$

E (ε_i) =0, the response function for regression model above is:

$$\text{E(Y)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{p-1} x_{p-1}.$$

General Linear Regression Model in Matrix Term

The model is: $Y = X\beta \ + \varepsilon$

Where:

Y is a vector of response

β is a vector of parameter

X is a matrix of constants

Ɛ is a vector of independent normal random variables with expectation. $E(\varepsilon_i) = 0$ and covariance matrix

$$\sigma^2(\varepsilon) = \begin{bmatrix} \sigma^2 & 0 & . & . & . & 0 \\ 0 & \sigma^2 & . & . & . & 0 \\ . & . & & . & & . \\ . & . & & . & & . \\ . & . & & . & & . \\ 0 & 0 & . & . & . & \sigma^2 \end{bmatrix}$$

Where σ^2 (Ɛ) is an nx1

Consequently, the random vector Y has expectation $E(Y)_{nx1} = X\beta$ and the variance-covariance matrix of Y is the same as that of Ɛ: $\sigma^2[Y]_{nxn} = \sigma^2 I$.

However in matrix term we need to define the following matrices:

$$Y_{nx1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \qquad X_{nxp} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ 1 & X_{31} & X_{32} & \cdots & X_{3,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix}$$

$$\beta_{px1} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix} \qquad \varepsilon_{nx1} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

$$X^I X = \begin{bmatrix} n & \Sigma X_{1i} & \Sigma X_{2i} & \cdots & \Sigma X_{ni} \\ \Sigma X_{1i} & \Sigma X_{1i}^2 & \Sigma X_{1i} X_{2i} & \cdots & \Sigma X_{1i} X_{ni} \\ \Sigma X_{2i} & \Sigma X_{1i} X_{2i} & \Sigma X_{2i}^2 & \cdots & \Sigma X_{2i} X_{ni} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Sigma X_{ni} & \Sigma X_{1i} X_{ni} & \Sigma X_{2i} X_{ni} & \cdots & \Sigma X_{ni}^2 \end{bmatrix}$$

$$X^I Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1,p-1} & X_{2,p-1} & \cdots & X_{n,p-1} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_{1i} Y_i \\ \Sigma X_{2i} Y_i \\ \vdots \\ \Sigma X_{ni} Y_i \end{bmatrix}$$

Estimation of $\beta_i (i = 0,1,2,...,n)$ in multiple regressions.

The least square criterion is generalised as follow for general linear regression model stated earlier:

Q= $\Sigma_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{i1} - ... - \beta_{p-1} X_{i,p-1})^2$

The least square estimators are those values of $\beta_0, \beta_1, ... \beta_{p-1}$ that minimize Q. We let b denotes the vector of the least square estimated regression coefficients $b_0, b_1, ..., b_{p-1}$

$$b_{px1} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{bmatrix}$$

The Least Square Normal equations for the general linear regression model are: $X^1 Xb = X^1 Y$ and the least square estimators are: $\hat{b} = \left(X^1 X\right)^{-1}_{pxp} X^1 Y_{px1}$

The method of maximum likelihood leads to the same estimator for normal error regression model as those obtain by the method of least squares.

The Maximum likelihood Estimator (MLE):

$$L\left(\beta, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} \exp\{-\frac{1}{2\sigma^2}\sum_{1=1}^{n}\left(Y_i - \beta_0 - \beta_1 X_{i1} - \ldots - \beta_{p-1} X_{i,p-1}\right)^2\}$$

Maximizing this likelihood function with respect to $\beta_0, \beta_1, \ldots \beta_{p-1}$ leads to the estimators given earlier such that $b = \left(X^1 X\right)^{-1}_{pxp} X^1 Y_{px1}$

The Models of Interests are:

Standardized model;

$MP^* = \beta_1^* Rf^* + \beta_2^* Temp^* + \beta_3^* Hum^*$ OR

$Z{\sim}MP = \beta_1^* Z{\sim}Rf + \beta_2^* Z{\sim}Temp + \beta_3^* Z{\sim}Hum$ .... (i)

Unstandardized model;

$MP = \beta_0 + \beta_1 Rf + \beta_2 Temp + \beta_3 Hum$ .... (ii)

Definition of Terms:

$MP \Rightarrow$ Maize production (measured in tonnes)

$Rf \Rightarrow$ Rainfall (measured in mm)

$Temp \Rightarrow$ Temperature (measured in ^{0}c)

$Hum \Rightarrow$ Humidity (measured in %)

$Z \sim MP \Rightarrow$ Standardized maize production (measured in tonnes)

$Z \sim Rf \Rightarrow$ Standardized Rainfall measured (measured in mm)

$Z \sim Temp \Rightarrow$ Standardized Temperature (measured in ^{0}c)

$Z \sim Hum \Rightarrow$ Standardized Humidity (measured in %)

MATERIAL:

The data used for this study was extracted from Central Bank of Nigeria (CBN) Statistical bulletin, December 2009 and the record office of Food Agriculture Organisation (FAO). It covered the period 1985 to 2007.

DISCUSSION OF RESULTS:

The two different statistical models fitted for this research work with their estimated parameters are SMLRM and USMLRM. The SPSS statistical software package was adopted to obtain the necessary results for discussion.

Table (1): Output of Unstandardized Regression Analysis:
MP. versus RF., TEMP, HUM

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6927.105	29964.385		-.231	.820
	RF	3.270	2.068	.354	1.581	.130
	TEMP	1104.045	678.516	.375	1.627	.120
	HUM	-273.899	210.306	-.282	-1.302	.208

a. Dependent Variable: MP

Source: Authors computation from SPSS software

From the USMLRM output in the table 1 above, the model becomes:

MP=-6927.105+3.270RF+1104.045TEMP-273.899HUM.

Table (2): Output of Standardized Regression Analysis:
$Z \sim MP.$ **Versus** $Z \sim RF,\ Z \sim TEMP.,\ Z \sim HUM$

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.194		-.001	.999
	Z-RF.	.355	.224	.355	1.587	.129
	Z-TEMP.	.375	.231	.374	1.624	.121
	Z-HUM.	-.283	.216	-.283	-1.307	.207

a. Dependent Variable: Z-MP.

Source: Authors computation from SPSS software

From the SMLRM output in table 2 above the model becomes:

$$Z \sim MP = 0.355Z \sim RF + 0.375Z \sim TEMP - 0.283Z \sim HUM$$

Interpretation:

The coefficient of Z-score in the above model in table 2 could be interpreted as follows;

$\beta_1^* = 0.355$: A 1 standard deviation increase in $Z \sim RF$ is predicted to result in a 0.355 standard deviation increase in $Z \sim MP$ holding $Z \sim TEMP.$ and $Z \sim HUM.$ constant.

$\beta_2^* = 0.375$: A 1 standard deviation increase in $Z \sim TEMP.$ is predicted to result in a 0.375 standard deviation increase in $Z \sim MP$ holding $Z \sim RF$ and $Z \sim HUM.$ constant.

$\beta_3^* = -0.283$: A 1 standard deviation decrease in $Z \sim HUM.$ is predicted to result in a -0.283 standard deviation decrease in $Z \sim MP$ holding $Z \sim RF$ and $Z \sim TEMP.$ constant.

However considering the standard error of the two models USMLRM and SMLRM from the output in table 1 and 2 above the standard error of SMLRM are smaller than that of USMLRM. The result revealed that SMLRM is the most efficient. Hence, the humidity as one of the factors from the two models above gave negative sign which implies that the factor (Humidity) is a risk factor that reduces the maize production for the period covered.

CONCLUSION:

In this research, the two models - USMLRM and SMLRM were compared. Result from the analysis showed that USMLRM and SMLRM produce different values of coefficients and standard errors in the simultaneous equations. This study therefore concluded that SMLRM was considered to be the most efficient model in the analysis due to the smallest standard error deduced from the output results of the research work.

REFERENCES:

1. Atanlogun S.K, Edwin O.A and Afolabi Y.O (2014): on comparative modelling of GLS and OLS estimating techniques. International journal of scientific & technology research volume 3, ISSN 2277-86. Pg 125-128
2. Atanlogun S.K, Aliu A.H and Edwin O.A (2015): on comparative modelling of Multiple Linear Regression model (MLRM) and Binary Logistic Regression Model (BLRM) for Hypertension in Human Body System. International journal of mathematics and computer research volume 3, ISSN 2320-7167. Pg 1110-1116
3. Beldona and Varnon E. (2007): Regression Analysis for equipment auditing, Emerald Group publicity limited, Texas, United State of America.
4. Damodar N. Gujarati and Dawn C. Porter (2009) Basic econometrics, published by Mc Graw-Hill/Irwin, a business unit of Mc Graw-Hill companies. Inc.1221. Avenue of the Americans New York (5th edition).
5. Damodar N. Gujarati (2003) Basic econometrics, Tata Mc Graw-Hill publishing company limited New Delhi (4th edition).
6. Draper N.R and Smith H. (1998): Applied regression analysis, John Wesley and Sons, inc., New York.
7. Faraway J. (2002): Practical regression and ANOVA using R, Newton Publisher United Kingdom.
8. Mc Cullagh, P & Neider, J. A. (1989): Generalised Linear Models 2nd ed. London: Chapman & Hall. Pg 46-58.

9. Murray R.S., John J.S and Srinvasa R.A (2000): Probability and statistics Mc Graw-Hill publisher, New publisher, New York, Chicago.
10. Murphy J.L (1973): Introductory Econometrics, Richard D. Irwin, Homewood, IL.
11. Netter J., Kutner M.H., Nachtsheim C.J and Wasserman W. (1996): Applied linear statistic model, Allyn and Bacon publication, New York.
12. Netter J and Wasserman W. (1974): Applied linear statistical model, Richard D. Irwin, Homewood, IL.
13. Schroeder I.A (1986): Applied Regression Analysis, Smith Publisher, United Kingdom.