[Volume 3 issue 8 August 2015]
PageNo.1110-1116 ISSN : 2320-7167

INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTER RESEARCH

On Comparative Modelling Of Multiple Linear Regression Model (MLRM) and Binary Logistic Regression Model (BLRM) For Hypertension in Human Body System

[1]*Atanlogun S.K,* [2] *Aliu A.H,* [3] *Edwin O.A*

[1, 2&3]Mathematics and Statistics Department, Rufus Giwa Polytechnic, Owo, Nigeria.

E-mail: [1] atanlogunkola@yahoo.com

E-mail: [2] ahaliu@ymail.com

E-mail: [3] toyinedwin@yahoo.com

ABSTRACT

In this study MLRM and BLRM were compared. To achieve the goal, MLRM and BLRM estimating approach were considered on a simultaneous equation models where gender is been classified as CLASS coded with an indicator, taking on the value 0 and 1 respectively. Data were collected based on Age, Sex, Weight, Height and Blood pressure of patients at two groups: 'Group A and Group B from the file record office of Federal Medical Centre Owo, Ondo state. The body mass index (BMI) was calculated from the patient's weight and height. Result from the analysis showed that MLRM and BLRM produced different values of coefficient and standard error in the two models. The MINITAB statistical software package was adopted to carry out the analysis of the results and the study however concluded that MLRM was considered to be the best contributory and most efficient model compared with that of BLRM.

Keywords: Multiple Linear Regression Model (MLRM), Binary Logistic Regression Model (BLRM), Hypertension, Weight (WT), Body Mass Index (BMI), Blood Pressure (BP), Gender->Class, Age.

INTRODUCTION

Modelling the relationship between explanatory and response variables is a fundamental activity encountered in statistics. Regression analysis entails finding out the relationship that exit between economic variables. Regression analysis involves the use of the linear classical model which gives direction to the process of regression. Simple regression model entails the process of investigating the relationship between a single explanatory (predictor) variable and a single response predictant variable. Multiple regression model entails the regression of more than two variables; in this case we have one dependent variable and several independent or explanatory variables. However the general linear regression model encompasses not only quantitative predictor variable but qualitative ones, designation of one or two possible outcomes are observed in a binary form of response such as alive or dead, success or failure. Although response may be accumulated to provide the number of success and the number of failure, the quantitative nature of the response still remains.

Efficiency and level of contribution are parts of properties that give a better description of the modelling in statistics. In this study the Multiple Linear Regression Model (MLRM) and Binary Linear Regression Model (BLRM) would be adopted on a simultaneous equation models. Hence the model with the best level of contribution and efficiency would be determined.

Hypertension

Hypertension or high blood pressure is a condition in which the blood pressure in the arteries is chronically elevated. With every heartbeat, the heart pumps blood through the arteries to the rest of the body (J. Ray soc med 74:896, 1981). It is an increase in blood pressure above the normal range, usually diagnosed on the systolic above 120mmHg and diastolic above 80mmHg. However the normal blood level is below 120/80; where 120 represent the systolic measurement (peak pressure in the arteries) and 80 represent the diastolic measurement (minimum pressure in the arteries). Blood pressure between 120/80 and 139/89 is called pre-hypertension (to denote an increased risk of hypertension) and a blood pressure of 140/90 or above is considered hypertension. Previous studies on hypertension generally emphasized on reducing the concentration on the efficacy of diastolic and systolic blood pressure. However, emphases were not much concentrated on the eating habit of the patients. Even though, most hypertensive patient required at least two different recommended drug classes to achieve the targeted blood pressure of 130/85mmHg, as a single dose might not easily bring it down.

Causes of Hypertension

Though the exact courses of hypertension are usually unknown, there are several factors that have been highly associated with the condition. These includes: *Smoking *Obesity or Being over-weight * Diabetes * Sedentary lifestyle * Lack of physical activity * High level of salt intake (sodium sensitivity) *Insufficient calcium, potassium and magnesium consumption * Vitamin D-deficiency * High level of alcohol consumption * Stress *Aging * Medicine such as birth control pills * Genetic and a family history of hypertension * chronic kidney disease * Adrenal and thyroid problems and tumors.

Classification of Hypertension

Essential hypertension: - Though essential hypertension remains somewhat mysterious, it has been linked to certain risk factors. High blood pressure tends to run in families and is more likely to affect men than women. Age and race also play a role. Essential hypertension is also greatly influenced by diet and lifestyle. The link between salt and high blood pressure is especially compelling. By contrast, people who add no salt to their food show virtually no traces of essential hypertension. The majority of people with high blood pressure are “salt sensitive” meaning that anything more than the minimal bodily need for salt is too much for them and increase their blood pressure. Other factors that have been associated with essential hypertension include obesity; diabetes; stress; insufficient intake of potassium, calcium and magnesium; lack of physical activity; chronic alcohol consumption. Secondary hypertension: - when a direct causes for high blood pressure can be identified, the condition is describe as secondary hypertension. Among the know causes of secondary hypertension, kidney disease ranks highest. Hypertension can also be triggered by tumors or other abnormalities that causes the adrenal glands (small glands that sit atop the kidney) to secret excess amount of the hormones that elevate blood pressure.

Classification of high blood pressure

Normal blood pressure: - less than 120/80
Pre-hypertension: -120-139/80-89
Hypertension: - greater than 140/90
Stage 1 Hypertension: - 140-159/90-99
Stage 2 Hypertension: - 160 or greater/100 or greater

MODEL SPECIFICATION

Multiple Linear Regression Model (MLRM)

In this type of model, we have response variable Y which is determine by two or more predictor variables x_i ,..., 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 a 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(Ɛ)=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:

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

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

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

$$X^I Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ X_{1,p-1} & X_{2,p-1} & \cdots & X_{n,p-1} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma X_1 \\ \Sigma X_{1i} Y_i \\ \Sigma X_{2i} Y_i \\ \vdots \\ \Sigma X_{ni} Y_i \end{bmatrix}$$

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:

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

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: $\hat{b} = \left(X^1 X\right)_{pxp}^{-1} X^1 Y_{px1}$

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):

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

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

Logistic Regression Model (LRM)

Logistic regression is part of a category of statistical methods called generalised linear model. The logistic regression model is simply a non linear transformation of linear regression. The logistic distribution is an S-shaped distribution function (cumulative density function) which is similar to the standard normal distribution and constrains the estimated probabilities to lie between 0 and 1. (http:/en.wikipedia.org/wiki/binary logistic regression).

Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. In most cases the response variable is dichotomous, such as presence/absence or success/failure.

There is pertinent condition that should be noted about this function. First of all, it is bounded between zero (0) and one (1). Secondly, there is a linear model hidden in the function that can be revealed with a proper transformation of the response. Finally, the sign association with the coefficient β indicates the direction of the curve.

The Model:

The response variable in logistic regression is usually dichotomous, this type of variable is called Bernoulli (or binary) variable. When the response variable is binary, there would be presence of an indicator, taking on the values 1 and 0 with probabilities π and $1-\pi$ respectively. Y is a Bernoulli random variable with parameter $E(Y) = \pi$ (John Neter, Michael H. Kutner, Christopher J. Nachtshein & William Wasserman, 1996).

$$E\left[Y/X\right] = \pi = \frac{\exp\left(\beta^1 X_i\right)}{1+\exp\left(\beta^1 X_i\right)} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \text{(i)}$$

The equation (i) for multiple now becomes:

$$\pi = \frac{e^{\left(\beta_0+\beta_1 x_1+\beta_2 x_{2+}...+\beta_k x_k\right)}}{\left[1+e^{\left(\beta_0+\beta_1 x_1+\beta_2 x_{2+}...+\beta_k x_k\right)}\right]} \quad \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \text{(ii)}$$

Where β_0 is the constant of the equation and β_i is the coefficient of the predictor variables.

The models of interest are:

$$GENDER = \beta_0 + \beta_1 \Delta BMI + \beta_2 \Delta BP + \beta_3 AGE$$

$$CLASS = \frac{e^{\left(\beta_0+\beta_1 \Delta BMI+\beta_2 \Delta BP+\beta_3 AGE\right)}}{[1+e^{\left(\beta_0+\beta_1 \Delta BMI+\beta_2 \Delta BP+\beta_3 AGE\right)}]}$$

Here we have two equations with one endogenous variables (GENDER) coded with an indicator taking on the values 1 and 0 with probabilities π and $1-\pi$ respectively, and three exogenous variables (ΔBMI, ΔBP and AGE).

Definitions of Terms

GENDER- Sex coded with an indicator of 1 and 0 such that

$$GENDER = \begin{cases} 1 \; if \; patient \; female \\ o \; if \; patient \; male \end{cases}$$

ΔBMI - changes in Body Mass Index
ΔBP - changes in Blood Pressure
AGE - Denotes the Age of individual patient.

MATERIAL

The data used for this study was extracted from the record office of Federal Medical centre Owo, Ondo state. The patients were randomly grouped into two groups: Group A and Group B for different consultant for eight consecutive weeks and the result of the two groups were now compared after this period. The Gender called CLASS was coded with an indicator taking on the value 1 and 0 respectively. The Age, Sex, Weight, Height and blood pressure of the patients were considered at various groups. Hence, body mass index (BMI) was calculated from the patient's weight and height. 40 observations were considered, out of which 30 were female and the remaining 10 were male.

DISCUSSION OF RESULTS

Two different statistical models fitted for this research work with their estimated parameters are MLRM and BLRM. The MINITAB statistical software package was adopted to obtain the results necessary for discussion.

Table (1): Regression Analysis: GENDER versus ΔBMI, ΔBP, AGE

Predictor	Coef	SE Coef	Z	P
Constant	0.7375	0.5345	1.38	0.176
ΔBMI	-0.06006	0.04589	-1.31	0.199
ΔBP	0.6762	0.6510	1.04	0.306
AGE	0.002245	0.009279	0.24	0.810

Source: Authors computation from MINITAB software

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

$GENDER=0.738-0.0601\Delta BMI+0.676\Delta BP+0.00224AGE$

Table (2): Binary Logistic Regression: GENDER versus ΔBMI, ΔBP, AGE

Predictor	Coef	SE Coef	Z	P
Constant	0.886431	1.70701	0.52	0.604
ΔBMI	-0.229338	0.165261	-1.39	0.165
ΔBP	2.56921	2.38260	1.08	0.281
AGE	0.0049619	0.0294072	0.17	0.866

Source: Authors computation from MINITAB software

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

$$Class = \frac{e^{\{0.886-0.229\Delta BMI+2.569\Delta BP+0.005AGE\}}}{[1+e^{\{0.886-0.229\Delta BMI+2.569\Delta BP+0.005AGE\}}]}$$

Interpretation

The Body Mass Index (BMI) predictor from the two models above gave negative sign which implies that the two factors are the higher risk factors in reducing the blood pressure among the hypertensive patients. The estimated parameters coefficients of the MLRM are all less than the coefficients of the BLRM, out of which the blood pressure (BP) increases by 0.6762 in MLRM and 2.56921 in BLRM respectively. Hence, it shows that the model MLRM contributes better than the BLRM in the analysis. Considering the standard error of the two models (0.6510 and 2.38260) - MLRM and BLRM from the output in table 1 & 2 above, the standard errors of MLRM are smaller than that of the BLRM. The result reveals that MLRM is the most efficient.

CONCLUSION

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

REFERENCES

1. Agresti, Alan (1996): An introduction to categorical data analysis. NY: John Wiley. An Excellent, Assessable introduction pg 67.
2. Allison, Paul D. (1999): Comparing logistic and probit coefficients of cross groups. Sociological methods and research, 28(2): pg 186-208
3. 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
4. Berglung G. Wilhelmson L, Sannerte dt R,Hansson L, Anderson O, Siwertsson R, Wedel H. Wikstrand J: Coronary Heart Disease after treatment of hypertension. Lancet I 1978
5. Buipih C J: sodium excess or potassium lack as a cause of hypertension: A discussion paper. J. Ray soc med 74:896, 1981,
6. Damodar N. Gujarati (2003) Basic econometrics, Tata Mc Graw-Hill publishing company limited New Delhi (4th edition).
7. Damoras Alfred (1992): Logit Modeling: pratical applications. Thousand Oaks, C A: Sage publication series: Quantitative Application in social science, No 106
8. Estrella, A. (1998): A new measure of fit for equations with dichotomous dependent variables. Journal of Business and Economic Statistics 16(2): pg 198-205
9. Menard, Scott (2002) Applied logistic regression analysis, 2nd edition. Thousand Oaks, C A: Sage publications Series Quantitative Applications in the social Sciences, No. 106. First ed., 1995.
10. Mc Cullagh, P & Neider, J. A. (1989): Generalised Linear Models 2nd ed. London: Chapman & Hall. Pg 46-58.
11. Pampel Fred C. (2000): Logistic regression: A primer sage quantitative applications in the social sciences series 132. Thousand Oaks ,, C A: Sage publications. Pg 35-38
12. Pedhazur, E. J. (1997): Multiple regression in behavioural research, 3rd ed. Orlando, F L: Harcourt Brace.
13. Press, S.J. and S. Wilson (1978) Choosing between logistic regression and discriminat analysis. Journal of The American Statistical Association, vol. 73: pg 699-705
14. Rice, J.C. (1994) Logistic regression: “An introduction”. In B. Thompson, ed., Advance in social sciences methodology, Greenwhich, CT: JAI press. Vol.3 pg 191-245
15. Watsky AL, Freidman GD, Siegelaub AB, Gerard M: Alcohol consumption and blood pressure. N Engl s Med 296: 1194, 1977.