# MULTIVARIATE REGRESSION ANALYSIS OF BABIES BIRTH CHARACTERISTICS IN BAYELSA AND RIVERS STATES, NIGERIA

**[1]Dagogo, Joseph., [2]Biu, Emmanuel Oyinebifun, [3]Wonu, Nduka**

*[1,2]University of Port Harcourt, Choba, Rivers State, Nigeria*

*[3]Department of Mathematics/Statistics, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria*

*Email: Goodluck4all2003@yahoo.com, emmanuelbiu@yahoo.com and ndukawonu@gmail.com*

## Abstract

This research work used multivariate regression analysis to analyze babies' birth measurements which is appropriate since there are several numbers of dependent variables and independent variables, usually the case in a real-life situation. The dependent variables [Lean Body Mass (LBM) $Y_1$ and Body Mass Index(BMI) $Y_2$] for some fixed values of the independent variables Head Circumference (CM) $X_1$, Age (Days) $X_2$, Height (CM) $X_3$ and Weight (Kg) $X_4$]were collected. Furthermore, the work was able to fit a multivariate regression model (MRM) to the data sets collected, estimated the model parameters and obtained matrices of the fitted coefficients with its residual matrices. Evaluating the significance of the multivariate regression model parameters, by using Bartlett's approximation test statistic to determine if any single variable $Y_i$ ($Y_{i,}$ i=1 and 2) is to be predicted from all the independent variables in the system. The research work was able to fit several multivariate regression models to the data obtained. From the models built, we notice that $X_2$ (Age in Days) for model $Y_1$ and also $X_1$ [Head Circumference (CM)] and $X_2$ (Age in Days) for model $Y_2$ are not significant for female and male in Bayelsa State, while $X_2$(Age in Days) for model $Y_2$ is not significant for female and male in Rivers State. This result confirms that there are some independent variables, that is $\beta_{ij} \neq 0$ and shows that some independent variables explain the dependent variables.

**Keywords:** *Multivariate Regression, Bartlett's approximation, Lean Body Mass, Body Mass Index and Head Circumference*

## 1. Introduction

In this research, we considered a multivariate regression; a situation where we obtained a functional relationship between two dependent variables and two or more independent variables. These kinds of situations where we have two dependent variables and independent variables are in the realm of Multivariate Regression Model (MRM), while one dependent variable and two or more independent variables is a multiple regression.

At this point, we emphasize the difference between multiple regression and multivariate regression, because there can be some confusion as to what constitutes a multivariate system in the context of a predictive model. An MRM is governed by the dimensionality of number p dependent variables. This is because the only distributions that play an important role in the theory of multivariate regression are the joint distributions of the dependent variables. When we have just one dependent (p=1) variable, multivariate regression is reduced to a multiple regression [Udom, (2010)].

The general multivariate regression model is given in the following matrix form;

$$Y = X\beta + \varepsilon \qquad (1.1)$$

*where* Y is the $n \times p$ matrix whose ij$^{th}$ elements are $y_{ij}$ the observed response of the i$^{th}$ individual on the dependent variable, X is the $n \times (q+1)$ matrix of the independent variables whose i$^{th}$ row is $\underline{x}_i = (1, x_{1i}, x_{2i} \; \ldots \; x_{qi})$, $\beta$ is the $(q+1) \times p$ matrix of unknown parameters and $\varepsilon$ is an $n \times p$ matrix of errors (random variables) whose rows $\underline{e}_i$ are independent observations [random errors associated with the i$^{th}$ response] from a multivariate normal distribution with zero and covariance matrix, (Sen and Srivastava, 2000). We have

$Y = (\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_p)$, $\varepsilon = (\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_p)$ and

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} \; \ldots \; x_{q1} \\ 1 & x_{12} & x_{22} \; \ldots \; x_{q2} \\ \cdot & \cdot & \cdot \qquad \cdot \\ & \cdot & \cdot \qquad \cdot \\ & \cdot & \cdot \qquad \cdot \\ 1 & x_{1n} & x_{2n} \; \ldots \; x_{qn} \end{bmatrix} \qquad (1.2)$$

The assumption of multivariate normality is only necessary if inferences are to be made either for the predicted values or the parameters.

The multivariate regression analysis deal with more than one dependent variable or where the dependent variables can be combined to identify what (or if any) changes in the independent variables have a significant effect on the dependent variables.

The aim of this study is to construct a suitable multivariate regression model that can predict the dependent variables for some fixed values of the independent variables [i.e. Dependent variables are Lean Body Mass (LBM) $Y_1$ and Body Mass Index (BMI) $Y_2$; and Independent variables are Head Circumference (CM) $X_1$, Age (Days) $X_2$, Height (CM) $X_3$ and Weight (Kg) $X_4$ ] .

1. Fit a multivariate regression model (MRM) to the data collected and its parameters estimates.
2. Obtain the matrices of fitted values and residuals
3. Test the significance of the multivariate regression model parameters.
4. Using Bartlett’s approximation test statistic to test if any single variable $y_j$ is to be predicted from all the independent variables in the system.

The data used in this research work were gotten from the Federal Medical Centre (FMC), Yanagoa, Bayelsa State, Nigeria and University of Port Harcourt Teaching Hospital (UPTH) Babies birth and Labour ward, Port Harcourt, Rivers State, Nigeria.

## 2.Multivariate Regression

Alexopoulos (2014), stated that multivariate regression analysis is used to forecast dependents or criterion variables from two or more predictor variables. Consider the sample regression model for $n$ observations with $k$ independent variables: $\mathrm{Y} = \mathrm{XB} + e$,

where $\mathrm{Y}$ is an $n \times 1$ column vector containing data on the dependent variables; $\mathrm{X}$ is an $n \times (k+1)$ matrix containing data on the predictor (independent) variables; $\mathrm{B}$ is a $(k+1) \times 1$ column vector containing estimated regression coefficients; and $e$ is an $n \times 1$ column vector containing residual terms (errors) where $e = \mathrm{Y} - \mathrm{XB} = \mathrm{Y} - \hat{\mathrm{Y}}$, then the n multivariate regression model is expressed as

$$\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ .y_n \end{pmatrix}_{n\times 1} = \begin{bmatrix} 1\ x_{11}\ \ x_{12}\ \ldots x_{1k} \\ 1\ x_{21}\ \ x_{22}\ \ldots x_{2k} \\ . \\ . \\ . \\ 1\ x_{n1}\ \ x_{n2}\ \ldots x_{nk} \end{bmatrix}_{n\times(k+1)} \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ .\beta_n \end{pmatrix}_{(k+1)\times 1} + \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ .e_n \end{pmatrix}_{n\times 1} \tag{2.1}$$

Fahrmeir *et al., (2007)* stated that, multivariate regression is a scientific procedure that allows additional factors to enter the analysis separately so that the effect of each can be estimated. It is valuable for quantifying the impact of various simultaneous influences upon dependent variables. In addition, because of bias omitted variables with simple regression and multiple regressions, multivariate regression is often extremely important.

### 2.1Body Mass Index and Lean Body Mass

The **body mass index** (**BMI**), or **Quetelet index**, is a measure of an individual's relative weight based on mass and height. It is also defined as the individual's body mass divided by the square of their height – with the value universally being given in units of kg/m$^2$ (Eknoyan, 2007). Mathematically, written as

$$\text{BMI} = \frac{mass\ (kg)}{\left[ height(cm) \right]^2}$$

The lean Body mass (LBM) is a measure of an individual's relative weight based on height.

$$\text{LBM} = 3.8\times\left( 0.0215\left[Weight(kg)\right]^{0.6469}\left[height\,(cm)\right]^{0.7236}\right)$$

or

$$\text{LBM}=3.8\times\text{ECV}$$

*where* ECV is called Extra Cellular Volume and it is calculated as

$$\text{ECV}= 0.0215\left[Weight(kg)\right]^{0.6469}\left[height\,(cm)\right]^{0.7236}$$

According to MacKay (2010); BMI is used differently for children. Although, evaluated the same way as for adults, it is compared to typical values for other children of the same age instead of set thresholds for overweight and underweight. In a report by Center for Disease Control, (2013), the BMI percentile was used to compare children of the same sex and age.

## 2.2 Review Works of BMI and LBM Predictive Equations

Jinhua *et al.* (2019), established a predictive equation of lean body mass suitable for Chinese adults. These adults were grouped by sex and then subdivided according to their body mass index (BMI). The female group was divided into another two subgroups: the premenopausal and postmenopausal subgroups (BMI, height, weight, and age). A stepwise Multilinear regression analysis of height, weight, age, and BMI were used to developed Equations. Their results are showed that age was proved to have no influence on LBM in the female group while the regrouping according to BMI or menopause did not increase the predictive ability of equations. Good agreement between LBM evaluated by equation Lean Body Mass Prediction Equation and LBM measured by dual-energy X-ray absorptiometry was observed in both the male and female groups. Their research was able to establish a Predictive equation of LBM suitable for healthy southern Chinese adults using a large sample.

Several works previously developed anthropometric prediction equations for estimation of LBM. However, most of them estimated LBM indirectly by estimating the percentage of body fat content [Heitmann (1990); Deurenberg *et al.,* (1991); Gallagher *et al.,* (2000);], while only some or not many studies predicted LBM with specifically developed equations [Lee *et al.,* (2017); Salamat *et al.,* (2015); Yu *et al.,* (2013); Kulkarni *et al.,* (2013)]

Yu *et al.,* (2013), stated that a regression equation designed based on morphological statistics, such as height and weight, is simple, fast, and inexpensive in calculating LBM and especially suitable for studies with large samples. Estimating equation is a way of calculating LBM when dual-energy X-ray absorptiometry is not available for babies' birth measurements.

## 3.Method

The data sets collected from the Federal Medical Centre (FMC), Yanagoa, Bayelsa State, Nigeria and University of Port Harcourt Teaching Hospital (UPTH) Babies Birth and Labour Ward, Port

Harcourt, Rivers State, Nigeria; consists of 600 males and females birth characteristics for each state, a total of 1,200 babies' measurement from each state. The characteristics variables are Lean Body Mass (LBM) ($Y_1$) and Body Mass Index (BMI) ($Y_2$) as the dependent variables; Head Circumference ($X_1$), Age of the babies in Days ($X_2$), Height or Length of the babies ($X_3$) and Weight of the babies ($X_4$), are the independent variables.

To build a suitable multivariate regression model that can predict the dependent variables for some fixed values of the independent variables. We developed a program that can solve, analyze, predict, and determine the outputs of multivariate regression problems using $C^{\#}$ programming language, since the data sets are large $C^{\#}$ programming language is a powerful tool for analysis and manipulation of large data sets, using multivariate statistical method.

Also, to check the efficiency and accuracy of our program in building a suitable multivariate regression model; we have chosen multi-purpose mathematics and a statistical package called Microsoft office Excel particularly designed for higher dimensional matrix analysis. To analyze, predict, and determine the outputs of multivariate regression problems when the sample size is 30. The result of the developed program and Microsoft office Excel are the same. The Methods used in handling Multivariate data were employed,

In this research, regressing two dependent variables on four independent variables was required. Therefore, our assume multivariate regression model for the two dependent variables and four independent variables is

$$\begin{aligned} Y_{j1} &= \beta_{01} + \beta_{11}X_{j1} + \beta_{21}X_{j2} + \beta_{31}X_{j3} + \beta_{41}X_{j4} + e_{j1} \\ Y_{j2} &= \beta_{02} + \beta_{12}X_{j1} + \beta_{22}X_{j2} + \beta_{32}X_{j3} + \beta_{42}X_{j4} + e_{j2} \end{aligned} \tag{3.1}$$

and the matrix form is

$$\begin{pmatrix} Y_{j1} \\ Y_{j2} \end{pmatrix} = \left( X_{j0} \;\; X_{j1} \;\; X_{j2}, X_{j3} \;\; X_{j4} \right) \begin{pmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix} + \begin{pmatrix} e_{j1} \\ e_{j2} \end{pmatrix} \tag{3.2}$$

*where*

$X_{j0} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$, $X_{j1}$ $X_{j2}, X_{j3}$ $X_{j4}$ are the Independent variables, which as Head Circumference (CM) $X_{j1}$, Age (Days) $X_{j2}$, Height (CM) $X_{j3}$ and Weight (Kg) $X_{j4}$;

$\underline{Y}_j = \begin{pmatrix} Y_{j1} \\ Y_{j2} \end{pmatrix}$ are the dependent variables, we have Lean Body Mass (LBM) $Y_{j1}$ and Body Mass Index (BMI) $Y_{j2}$ and the error is $\underline{e}_j = \begin{pmatrix} e_{j1} \\ e_{j2} \end{pmatrix}$

We considered the multivariate extension of the multiple linear regression modelling the relationship between m response $Y_{j1}$, $Y_{j2}$,. . .,$Y_{jq}$ and a single set of p predictor variables $X_{j1}$ $X_{j2}, . . . , X_{jp}$ . Each of the q response is assumed to follow its own regression model, i.e

$$\begin{aligned} Y_{j1} &= \beta_{01} + \beta_{11}X_{j1} + \beta_{21}X_{j2} + . . . + \beta_{p1}X_{jp} + e_{j1} \\ Y_{j2} &= \beta_{02} + \beta_{12}X_{j1} + \beta_{22}X_{j2} + . . . . + \beta_{p2}X_{jp} + e_{j2} \\ &. \\ &. \\ &. \\ Y_{jq} &= \beta_{0q} + \beta_{1q}X_{j1} + \beta_{2q}X_{j2} + . . . + \beta_{pq}X_{j4} + e_{jq} \end{aligned} \tag{3.3}$$

*where*

$$E(\underline{e}_j) = E\left[\begin{pmatrix} e_{j1} \\ e_{j2} \end{pmatrix}\right] = \underline{0} \quad , \quad Var(\underline{e}_j) = \underline{\Sigma}$$

Conceptually, we can let

$$X_{jo}, X_{j1}\ X_{j2}, . . . , X_{jp}$$

and denote the predictor variables for the j$^{th}$ trial as

$\underline{Y}_j = \begin{pmatrix} Y_{j1} \\ Y_{j2} \\ . \\ . \\ . \\ Y_{jq} \end{pmatrix}$ be the responses and, $\underline{e}_j = \begin{pmatrix} e_{j1} \\ e_{j2} \\ . \\ . \\ . \\ e_{jq} \end{pmatrix}$ is the errors for the $j^{th}$ trial. Thus, we have an $n \times (p+1)$ design matrix

$$\underline{X} = \begin{bmatrix} X_{10} \ X_{11} \ \ldots X_{1p} \\ X_{20} \ X_{21} \ \ldots X_{2p} \\ . \\ . \\ . \\ X_{n0} \ X_{n1} \ \ldots X_{np} \end{bmatrix}$$

If we now set

$$\underline{Y} = \begin{bmatrix} Y_{11} \ Y_{12} \ \ldots Y_{1q} \\ Y_{21} \ Y_{22} \ \ldots Y_{2q} \\ . \\ . \\ . \\ Y_{n1} \ Y_{n2} \ \ldots Y_{nq} \end{bmatrix} = \left[\underline{Y}_1 \ \ \underline{Y}_2 \ \ldots \underline{Y}_q\right] \text{ and also}$$

$$\underline{\beta} = \begin{bmatrix} \beta_{01} \ \beta_{02} \ \ldots \beta_{0q} \\ \beta_{11} \ \beta_{12} \ \ldots \beta_{1q} \\ \beta_{21} \ \beta_{22} \ \ldots \beta_{2q} \\ . \\ . \\ . \\ \beta_{p1} \ \beta_{p2} \ \ldots \beta_{pq} \end{bmatrix} = \left[\underline{\beta}_1 \ \ \underline{\beta}_2 \ \ldots \underline{\beta}_q\right], \text{ then}$$

$$\underline{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \ldots Y_{1q} \\ Y_{21} & Y_{22} & \ldots Y_{2q} \\ . & & \\ . & & \\ . & & \\ Y_{n1} & Y_{n2} & \ldots Y_{nq} \end{bmatrix} = \begin{bmatrix} \underline{Y}_1 & \underline{Y}_2 & \ldots \underline{Y}_q \end{bmatrix}$$

$$\underline{e} = \begin{bmatrix} e_{11} & e_{12} & \ldots e_{1q} \\ e_{21} & e_{22} & \ldots e_{2q} \\ . & & \\ . & & \\ . & & \\ e_{n1} & e_{n2} & \ldots e_{nq} \end{bmatrix} = \begin{bmatrix} \underline{e}_1 & \underline{e}_2 & \ldots \underline{e}_q \end{bmatrix} = \begin{bmatrix} \underline{e}_1{}' \\ \underline{e}_2{}' \\ . \\ . \\ . \\ \underline{e}_q{}' \end{bmatrix}$$

The multivariate linear regression model is

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{e} \qquad (3.4)$$

With

$$\mathrm{E}[\underline{e}_i] = \underline{0} \qquad (3.5)$$

and

$$Cov(e_i\ , e_k) = \sigma_{ik} I\ , \qquad i\ , k = 1, 2, \ldots q \qquad (3.6)$$

Note also that the q observed response on the j$^{th}$ trial have covariance matrix

$$\underline{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots \sigma_{1q} \\ \sigma_{21} & \sigma_{22} & \ldots \sigma_{2q} \\ . & & \\ . & & \\ . & & \\ \sigma_{q1} & \sigma_{q2} & \ldots \sigma_{qq} \end{bmatrix}$$

The ordinary least square estimates are found in a manner analogous to the univariate case- we begin by taking

$$\underline{\hat{\beta}}_{(i)} = \left(\underline{X}'\,\underline{X}\right)^{-1}\left(\underline{X}'\,\underline{Y}_{(i)}\right) \tag{3.7}$$

Collecting the univariate least square estimates yields

$$\hat{\beta} = \left[\underline{\hat{\beta}}_{(1)}\ \ \underline{\hat{\beta}}_{(2)}\ \ldots \underline{\hat{\beta}}_{(q)}\right] = \left(\underline{X}'\,\underline{X}\right)^{-1}\underline{X}'\left[\underline{Y}_{(1)}\ \ \underline{Y}_{(2)}\ \ldots \underline{Y}_{(q)}\right] = \left(\underline{X}'\,\underline{X}\right)^{-1}\underline{X}'\,\underline{Y} \tag{3.8}$$

Now for any choice of coefficients

$$\beta = \left[\underline{b}_{(1)}\ \ \underline{b}_{(2)}\ \ldots \underline{b}_{(q)}\right]$$

The resulting errors matrix is

$$\underline{Y} - \underline{X}\underline{\beta} \tag{3.9}$$

The resulting Error Sums of Squares and Cross-Products is

$$\left(\underline{Y} - \underline{X}\underline{\beta}\right)'\left(\underline{Y} - \underline{X}\underline{\beta}\right) = \begin{bmatrix} \left(\underline{Y}_{(1)} - \underline{X}\underline{b}_{(1)}\right)'\left(\underline{Y}_{(1)} - \underline{X}\underline{b}_{(1)}\right) & \ldots & \left(\underline{Y}_{(1)} - \underline{X}\underline{b}_{(1)}\right)'\left(\underline{Y}_{(q)} - \underline{X}\underline{b}_{(q)}\right) \\ . & . & . \\ . & . & . \\ . & . & . \\ \left(\underline{Y}_{(q)} - \underline{X}\underline{b}_{(q)}\right)'\left(\underline{Y}_{(1)} - \underline{X}\underline{b}_{(1)}\right) & \ldots & \left(\underline{Y}_{(q)} - \underline{X}\underline{b}_{(q)}\right)'\left(\underline{Y}_{(q)} - \underline{X}\underline{b}_{(q)}\right) \end{bmatrix}$$

We can show that the selection $\underline{b}_{(i)} = \underline{\hat{\beta}}_{(i)}$ minimizes the $i^{th}$ diagonal sum of squares

$$\left(\underline{Y}_{(i)} - \underline{X}\underline{b}_{(i)}\right)'\left(\underline{Y}_{(i)} - \underline{X}\underline{b}_{(i)}\right) \tag{3.10}$$

i.e.

$tr\left[\left(\underline{Y} - \underline{X}\underline{\beta}\right)'\left(\underline{Y} - \underline{X}\underline{\beta}\right)\right]$ and $\left|\left(\underline{Y} - \underline{X}\underline{\beta}\right)'\left(\underline{Y} - \underline{X}\underline{\beta}\right)\right|$ is called generalized variance; Note that both minimized. So we have matrices of predicted values

$$\underline{\hat{Y}} - \underline{X}\underline{\hat{\beta}} = \left(\underline{X}'\,X\right)^{-1}\underline{X}'\,Y \tag{3.11}$$

and we have a resulting matrices of residuals

Note that the orthogonality conditions among errors (residuals), predicted values, and columns of the design matrix which hold in the univariate case are also true in the multivariate case because

$$\underline{X}'\left[\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\right] = \underline{X}' - \underline{X}' = \underline{0} \tag{3.12}$$

This means that the errors are perpendicular to the columns of the design matrix.

$$\underline{X}'\underline{\hat{e}} = \underline{X}'\left[\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\right] = \underline{X}' - \underline{X}' = \underline{0} \tag{3.13}$$

and to the predicted values

$$\underline{Y}'\underline{\hat{e}} = \underline{\hat{\beta}}'\underline{X}'\left[\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\right]\underline{Y} = \underline{0} \tag{3.14}$$

Furthermore, because

$$\underline{Y} = \underline{\hat{Y}} + \underline{\hat{e}} \tag{3.15}$$

we have

$$\underline{Y}\,\underline{Y} = \underline{\hat{Y}}\,\underline{\hat{Y}} + \underline{\hat{e}}'\,\underline{\hat{e}} \tag{3.16}$$

where $\underline{Y}\,\underline{Y}$ is total sums of squares and cross-products, $\underline{\hat{Y}}\,\underline{\hat{Y}}$ is the predicted sums of squares and cross-products and $\underline{\hat{e}}'\,\underline{\hat{e}}$ is the residual (error) sums of squares and cross-products.

Furthermore, we can test hypothesis about the model parameters. To test the null hypothesis

$$\begin{aligned} &H_0 : \beta_{ij} = 0 \quad \textit{for all } i \textit{ and } j \\ &\textit{against} \\ &H_1 : \beta_{ij} \neq 0 \quad \textit{for all } i \textit{ and } j \ \textit{alternative hypothesis} \end{aligned} \tag{3.17}$$

The test statistic is the Wilk's $\wedge$ given by

$$\wedge = \prod_{i=1}^{p} (1 + \lambda_i)^{-1} \tag{3.18}$$

*where*

$\lambda_i$ is the i^th^ eigenvalues of $(\underline{\hat{e}}'\,\underline{e})^{-1}\,H$

$$H = \left(\hat{Y}'\hat{Y} - n\underline{\bar{y}}\,\underline{\bar{y}}'\right) = \left[\left(YY - n\underline{\bar{y}}\,\underline{\bar{y}}'\right) - \underline{\hat{e}}'\,\underline{\hat{e}}\right] \tag{3.19}$$

*where* $\underline{\bar{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_p)'$

The Bartlett's approximate test statistic is

$$B_{\wedge} = -\left\{n - q - 1 - \frac{1}{2}(p - q + 1)\right\}\log_e \wedge \qquad (3.20)$$

This test statistic is asymptotically distributed according to the chi-square distribution on pq degrees of freedom if the null hypothesis is true. Therefore, the null hypothesis is rejected if $B_{\wedge}$ is greater than $\chi^2_{pq}(\alpha)$.

## 4.Results

This section deals with data analysis and results presentation from the computerized method of solving systems of multivariate regression equations which was compared to multiple regression analysis of BMI and LBM on the independent variables.

### Multivariate Regression Equations for Female Bayelsa State Babies

We estimated the multivariate regression; we have the following Matrices outputs

**Outputs**

| **X'X** | | | |
|---|---|---|---|
| 600 | 205.06 | 6480 | 307.35 |
| 205.06 | 70.695 | 2227.44 | 105.0626 |
| 6480 | 2227.44 | 158020 | 3445.9 |
| 307.35 | 105.0626 | 3445.9 | 158.2606 |
| 2006.45 | 686.747 | 25154.55 | 1039.415 |

| **(X'X)-1** | | | |
|---|---|---|---|
| 0.609962 | -0.56489 | 0.000745 | -0.9131 |
| -0.56489 | 1.641413 | -0.00013 | 0.037741 |
| 0.000745 | -0.00013 | 1.79E-05 | -0.00089 |
| -0.9131 | 0.037741 | -0.00089 | 2.137051 |
| 0.013292 | -0.00421 | -0.00013 | -0.05529 |

| **(X'Y)** | |
|---|---|
| 7609.518 | 1844.869 |
| 2602.809 | 631.2041 |
| 87605.81 | 22551.16 |
| 3896.729 | 955.6079 |
| 26210.18 | 6453.107 |

| **β = (X'X)-1(X'Y)** | |
|---|---|
| 26.80707 | -1.24051 |
| -0.44744 | 0.005863 |
| -0.00281 | 8.38E-05 |
| -49.6061 | 4.603093 |
| 3.429774 | 0.584451 |

From the estimated values Matrix above, Equation 3.3 becomes

$$Y_1 = 26.807 - 0.447X_1 - 0.003X_2 - 49.606X_3 + 3.450X_4$$
$$Y_2 = -1.241 + 0.006X_1 + 8.38E^{-5}X_2 + 4.603X_3 + 0.584X_{j4}$$

From the models built, we notice that $X_2$ for model $Y_1$ and also $X_1$ and $X_2$ for model $Y_2$ are not significant.

The estimated values of the dependent variables and residual vectors for female babies, in BMI, Bayelsa State were obtained. Then, the residual matrix (2×2) of the dependent variables and estimated dependent variables is

| **E=(e1,e2)'(e1,e1)** | |
|---|---|
| 99171.58 | 23838.45 |
| 23838.45 | 5887.28 |

Then, $(\hat{\underline{e}}' \underline{e})^{-1}$ H is

| | |
|---|---|
| 165.286 | 39.73075 |
| 39.73075 | 9.812133 |

| | |
|---|---|
| 0.226736 | -0.91809 |
| -0.91809 | 3.81939 |

we compute the eigenvalues for adequacy of the model

$$\begin{vmatrix} 0.2267-\lambda & -0.9181 \\ -0.9181 & 3.8194-\lambda \end{vmatrix} = 0$$

We have $\lambda_1 = 4.04 \ \ and \ \lambda_2 = 0.0057$ ; next we obtained the Wilk's $\wedge$ given in Equation (3.20)

$$\wedge = \prod_{i=1}^{p} (1 + \lambda_i)^{-1} = \frac{1}{(1 + 4.04)} \times \frac{1}{(1 + 0.0057)} = 0.1973$$

n=600, q=2 and p=4. Then, the Bartlett's approximate test statistic in Equation 3.24 is

$$B_{\wedge} = -\left\{600 - 2 - 1 - \frac{1}{2}(4 - 2 + 1)\right\} \log_e(0.1973) = 966.4965$$

The chi-square distribution on pq (4)(2) degrees of freedom if the null hypothesis is true. Therefore, the null hypothesis is rejected if $B_{\wedge}$ (996.4965) is greater than $\chi_8^2(0.05) = 15.5$. This result confirms that there are some independent variables that is $\beta_{ij} \neq 0$ and shows that some independent variables explain the dependent variables.

**Multivariate Regression Equations for Male Bayelsa State Babies**

We estimated the multivariate regression; we have the following Matrices outputs

**Outputs**

| X'X | | | | |
|---|---|---|---|---|
| 600 | 205.94 | 5461 | 305.331 | 2008.3 |
| 205.94 | 71.2488 | 1870.64 | 104.7259 | 688.1155 |
| 5461 | 1870.64 | 92553 | 2883.223 | 20851.55 |

| | | | | |
|---|---|---|---|---|
| 305.331 | 104.7259 | 2883.223 | 156.5375 | 1037.02 |
| 2008.3 | 688.1155 | 20851.55 | 1037.02 | 7157.395 |

**(X'X)-1**

| | | | | |
|---|---|---|---|---|
| 0.536573 | -0.67314 | 0.00084 | -0.66848 | 0.008566 |
| -0.67314 | 1.792913 | -0.00027 | 0.097411 | 0.003179 |
| 0.00084 | -0.00027 | 3.67E-05 | -0.0009 | -0.00019 |
| -0.66848 | 0.097411 | -0.0009 | 1.587878 | -0.04925 |
| 0.008566 | 0.003179 | -0.00019 | -0.04925 | 0.00511 |

**(X'Y)**

| | |
|---|---|
| 7768.482 | 1838.061 |
| 2665.05 | 629.8662 |
| 74185.32 | 18726.66 |
| 3927.095 | 949.5884 |
| 26758.66 | 6477.397 |

**β = (X'X)-1(X'Y)**

| | |
|---|---|
| 40.75155 | -1.29951 |
| -3.51064 | 0.054552 |
| 0.01028 | 0.000122 |
| -82.0464 | 4.690833 |
| 4.499173 | 0.58438 |

From the estimated values Matrix above, Equation 3.3 becomes

$$Y_1 = 40.752 - 3.511X_1 + 0.010X_2 - 82.046X_3 + 4..499X_4$$
$$Y_2 = -1.299 + 0.054X_1 + 0.0001X_2 + 4.690X_3 + 0.584X_{j4}$$

From the models built, we notice that $X_2$ for model $Y_1$ and also $X_1$ and $X_2$ for model $Y_2$ are not significant.

The estimated values of the dependent variables and residual vectors for male babies, Bayelsa State were also obtained. Then, the residual matrix (2×2) of the dependent variables and estimated dependent variables is

| **E=(e1,e2)'(e1,e1)** | |
|---|---|
| 106172.3 | 24117.78 |
| 24117.78 | 5887.687 |

Then, $(\underline{\hat{e}}' \ \underline{e})^{-1}$ H is

| | |
|---|---|
| 176.9538 | 40.19629 |
| 40.19629 | 9.812811 |

| | |
|---|---|
| 0.081318 | -0.3331 |
| -0.3331 | 1.466405 |

we compute the eigenvalues for adequacy of the model

$$\begin{vmatrix} 0.0813-\lambda & -0.3331 \\ -0.3331 & 1.466-\lambda \end{vmatrix} = 0$$

We have $\lambda_1 = 1.542$ *and* $\lambda_2 = 0.00555$ ; next we obtained the Wilk's $\wedge$ given in Equation (3.20).

$$\wedge = \prod_{i=1}^{p} (1+\lambda_i)^{-1} = \frac{1}{(1+1.542)} \times \frac{1}{(1+0.00555)} = 0.3913$$

n=600, q=2 and p=4. Then, the Bartlett's approximate test statistic in Equation 3.24 is

$$B_{\wedge} = -\left\{600 - 2 - 1 - \frac{1}{2}(4 - 2 + 1)\right\} \log_e (0.3913) = 558.7$$

The chi-square distribution on pq (4)(2) degrees of freedom if the null hypothesis is true. Therefore, the null hypothesis is rejected if $B_{\wedge}$ (=558.7) is greater than $\chi^2_8(0.05)$=15.5. This result confirms that there are some independent variables that is $\beta_{ij} \neq 0$ and shows that some independent variables explain the dependent variables.

**Multivariate Regression Equations for Female Rivers State Babies**

We estimated the multivariate regression model; we have the following Matrices outputs

**Outputs**

| **X'X** | | | | |
|---|---|---|---|---|
| 600 | 199.94 | 1842 | 272.48 | 1881.74 |
| 199.94 | 68.0666 | 629.91 | 91.1465 | 635.0541 |
| 1842 | 629.91 | 6968 | 850.14 | 6226.49 |
| 272.48 | 91.1465 | 850.14 | 125.9054 | 863.9866 |
| 1881.74 | 635.0541 | 6226.49 | 863.9866 | 6158.779 |

| **(X'X)-1** | | | | |
|---|---|---|---|---|
| 0.156182 | -0.19786 | 0.00443 | -0.17378 | -0.00742 |
| -0.19786 | 0.847119 | -0.00357 | -0.03079 | -0.01896 |
| 0.00443 | -0.00357 | 0.001928 | 0.003127 | -0.00337 |
| -0.17378 | -0.03079 | 0.003127 | 0.555954 | -0.02488 |
| -0.00742 | -0.01896 | -0.00337 | -0.02488 | 0.011284 |

| **(X'Y)** | |
|---|---|
| 9539.761 | 1621.143 |
| 3206.801 | 545.9546 |
| 30888.69 | 5282.587 |
| 4183.605 | 751.1737 |
| 30733.09 | 5267.232 |

| **β = (X'X)-1(X'Y)** | |
|---|---|
| 37.29282 | -1.03524 |

| | |
|---|---|
| 6.999274 | -0.16288 |
| -0.22305 | -0.00152 |
| -98.7879 | 4.543285 |
| 6.958037 | 0.552516 |

From the estimated values Matrix above, Equation 3.3 becomes

$$Y_1 = 37.293 + 6.999X_1 - 0.22305X_2 - 98.788X_3 + 6.958X_4$$
$$Y_2 = -1.035 - 0.163X_1 + 0.0015X_2 + 4.543X_3 + 0.553X_{j4}$$

From the models built, we notice that $X_2$ for model $Y_2$ are not significant.

The estimated values of the dependent variables and residuals vectors for female babies in Rivers State were obtained. Then, the residual matrix (2×2) of the dependent variables and estimated dependent variables is

| **E=(e1,e2)'(e1,e1)** | |
|---|---|
| 171872.6 | 25542.74 |
| 25542.74 | 4547.82 |

Then, $(\hat{\underline{e}}'\ \underline{e})^{-1}$ H is

| | |
|---|---|
| 286.4543 | 42.57123 |
| 42.57123 | 7.5797 |

| | |
|---|---|
| 0.021118 | -0.11861 |
| -0.11861 | 0.798083 |

we compute the eigenvalues for adequacy of the model

$$\begin{vmatrix} 0.0211-\lambda & -0.1186 \\ -0.1186 & 0.7980-\lambda \end{vmatrix} = 0$$

We have $\lambda_1 = 0.8158$ *and* $\lambda_2 = 0.0034$ ; next we obtained the Wilk's $\wedge$ given in Equation (3.20).

$$\wedge = \prod_{i=1}^{p}(1+\lambda_i)^{-1} = \frac{1}{(1+0.8158)} \times \frac{1}{(1+0.0034)} = 0.5488$$

n=600, q=2 and p=4. Then, the Bartlett's approximate test statistic in Equation 3.24 is

$$B_{\wedge} = -\left\{600 - 2 - 1 - \frac{1}{2}(4 - 2 + 1)\right\}\log_e(0.5488) = 359.1$$

The chi-square distribution on pq (4)(2) degrees of freedom if the null hypothesis is true. Therefore, the null hypothesis is rejected if $B_{\wedge}$ (359.1) is greater than $\chi^2_8(0.05)$=15.5. This result confirms that there are some independent variables that is $\beta_{ij} \neq 0$ and shows that some independent variables explain the dependent variables.

**Multiple regression analysis of BMI and LBM on the independent variables**

Minitab 18 statistical software was used to evaluate each dependent variable on the independent variables in Table 4.1, in other to compare the significant parameters estimated with the results of multivariate regression equations parameters**.**

**Table 4.1:** Multiple Regression Analysis Results

| Baby Sex/Location | Dependent variable | Independent variables Term | Coefficients | Standard Error Coefficients | t-Value | p-Value | R-square | R-square (adj) |
|---|---|---|---|---|---|---|---|---|
| Female/Bayelsa State | Body Mass Index (Kg) $Y_1$ | Constant | -1.2355 | 0.0158 | -78.150 | 0.000** | 93.20% | 93.16% |
| | | Head Circumference (M) $X_1$ | -0.0007 | 0.026 | -0.030 | 0.978 | | |
| | | Age (Days) $X_2$ | 0.000087 | 0.000086 | 1.010 | 0.312 | | |
| | | Length (M) $X_3$ | 4.5894 | 0.0296 | 155.050 | 0.000** | | |
| | | Wight (Kg) $X_4$ | 0.58547 | 0.00148 | 395.670 | 0.000** | | |
| | Lime Body | Constant | -1.235 | 0.0158 | -78. | 0.000* | 99.9 | 99.89 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mass (Kg) $Y_2$ | | 5 | | 150 | * | 0% | % |
| | | Head Circumference (M) $X_1$ | -0.0007 | 0.026 | -0.030 | 0.978 | | |
| | | Age (Days) $X_2$ | 0.000087 | 0.000086 | 1.010 | 0.312 | | |
| | | Length (M) $X_3$ | 4.5894 | 0.0296 | 155.050 | 0.000** | | |
| | | Wight (Kg) $X_4$ | 0.58547 | 0.00148 | 395.670 | 0.000** | | |
| Male/Bayelsa State | Body Mass Index (Kg) $Y_1$ | Constant | 40.72 | 1.340 | 30.460 | 0.000** | 73.80% | 73.62% |
| | | Head Circumference (M) $X_1$ | -3.57 | 2.440 | -1.460 | 0.144 | | |
| | | Age (Days) $X_2$ | 0.01 | 0.011 | 0.910 | 0.365 | | |
| | | Length (M) $X_3$ | -81.87 | 2.300 | -35.640 | 0.000** | | |
| | | Wight (Kg) $X_4$ | 4.491 | 0.130 | 34.440 | 0.000** | | |
| | Lime Body Mass (Kg) $Y_2$ | Constant | -1.2934 | 0.0176 | -73.530 | 0.000** | 99.87% | 99.87% |
| | | Head Circumference (M) $X_1$ | 0.0465 | 0.0322 | 1.450 | 0.149 | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Age (Days) $X_2$ | 0.000158 | 0.000146 | 1.080 | 0.279 | | |
| | | Length (M) $X_3$ | 4.6835 | 0.0302 | 154.970 | 0.000** | | |
| | | Wight (Kg) $X_4$ | 0.58424 | 0.00172 | 340.610 | 0.000** | | |
| Female/Rivers State | Body Mass Index (Kg) $Y_1$ | Constant | 37.267 | 0.826 | 45.140 | 0.000** | 88.61% | 88.53% |
| | | Head Circumference (M) $X_1$ | 7.01 | 1.92 | 3.650 | 0.000** | | |
| | | Age (Days) $X_2$ | -0.2229 | 0.0917 | -2.430 | 0.015** | | |
| | | Length (M) $X_3$ | -98.74 | 1.56 | -63.420 | 0.000** | | |
| | | Wight (Kg) $X_4$ | 6.957 | 0.222 | 31.350 | 0.000** | | |
| | Lime Body Mass (Kg) $Y_2$ | Constant | -1.0422 | 0.0122 | -85.600 | 0.000** | 99.66% | 99.66% |
| | | Head Circumference (M) $X_1$ | -0.1563 | 0.0283 | -5.520 | 0.000** | | |
| | | Age (Days) $X_2$ | -0.00159 | 0.00135 | -1.180 | 0.239 | | |
| | | Length (M) | 4.554 | 0.023 | 198 | 0.0 | | |

| | | | | |
|---|---|---|---|---|
| $X_3$ | 7 | | .38 | 00* |
| | | | 0 | * |
| | | | 168 | 0.0 |
| Wight (Kg) $X_4$ | 0.552 | | .86 | 00* |
| | 48 | 0.00327 | 0 | * |

**Footnote:** ** significant at 5%.

The results in table 4.1 confirms that there are some independent variables explain the dependent variables as determine by the multivariate regression method. Multiple regression technique was used to check efficiency and accuracy of our program in building a suitable multivariate regression model.

## 5. Summary and Conclusion

Multivariate regression was used to analyze baby's birth measurements. This tool is judged to be appropriate since there are two dependent variables and several independent variables as usually in real life situation. We also discovered during the course of this study that with the manual method of calculation, there is bound to be error in calculation. Also, if the dependents and independents exceed two each, the solution becomes too cumbersome when applying multiple regression technique. Furthermore, this work was able to fit a multivariate regression model (MRM) to the data collected and its coefficients (parameters). Obtain the matrices of fitted values and residuals. Test the significance of the multivariate regression model (MRM) parameters. Used Bartlett's approximation test statistic to test if any single variable $y_j$ is to be predicted from all the independent variables in the system.

## 6.Conclusion

Several multivariate regression models were established to the data sets collected. From the models built, we notice that $X_2$ (Age in Days) for model $Y_1$ and also $X_1$ [Head Circumference (CM)] and $X_2$ (Age in Days) for model $Y_2$ are not significant for female and male in Bayelsa State, while $X_2$(Age in Days) for model $Y_2$ are not significant for female and male in Rivers State.This result confirms that there are independent variables that is $\beta_{ij} \neq 0$, some of which explain the dependent variables.

## 7.References


1.Alexopoulos 2014 E. C. (2014). *"Introduction to Multivariate Regression Analysis"* Review Article PASCHOS KA; Hippokratia, 14 (1): 23-28.

2.Center for Disease Control (2013). "*Body Mass Index: BMI for Children and Teens"* Retrieved 2013-12-16.

3.Deurenberg P., Weststrate J. A., and Seidell J. C., (1991). "*Body mass index as a measure of body fatness: age- and sex-specific prediction formulas,"* British Journal of Nutrition, vol. 65, no. 2, pp. 105–114.

4.Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*, John Wiley and Sons Inc, New York.

5.Draper, N. R. and Smith, H. (1998). "*Applied Regression Analysis,"* Second Edition, Wiley, New York.

6.Eknoyan, Garabed (2007). "*Adolphe Quetelet (1796–1874*)—*The Average Man and Indices of Obesity*". *Nephrology Dialysis Transplantation* **23** (1): 47–51.

7.Fahrmeir, L., Kneib T., and S. Lang (2007). *Multivariate Statistical Modelling Based on Generalized Linear Models*, (2nd ed.). New York: Springer-Verlag.

8.Gadzik, James (2006). "*'How much should I weigh?' Quetelet's equation, upper weight limits, and BMI prime*". *Connecticut Medicine* **70** (2): 81–8. PMID 16768059.

9.Gallagher D., Heymsfield S. B., Heo M., Jebb S. A., Murgatroyd P. R., and Sakamoto Y., (2000). *"Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index,"* American Journal of Clinical Nutrition, vol. 72, no. 3, pp. 694–701.

10.Heitmann B. L., (1990) "*Evaluation of body fat estimated from body mass index, skinfolds and impedance. a comparative study," European Journal of Clinical Nutrition*, vol. 44, no. 11, pp. 831–837.

11.Jinhua L., Jingjie S., Bin Guo, J. G. and Hao X. (2019). *Establishment of Prediction Equations of Lean Body Mass Suitable for Chinese Adults*. Creative Commons Attribution License.

12.Lee D. H., Keum N., Hu F. B. *et al.,* (2017). "*Development and validation of anthropometric prediction equations for lean body mass, fat mass and percent fat in adults using the National Health and Nutrition Examination Survey (NHANES) 1999-2006*," British Journal of Nutrition, vol. 118, no. 10, pp. 858–866.

13.MacKay, N. J. (2010). *"Scaling of human body mass with height: The body mass index revisited". Journal of Biomechanics* **43** (4): 764–6.

14.Nduka, E. C. (1999). *Principle of Applied Statistics 1*, Crystal Publishers, Imo State, Nigeria.

15.Salamat M. R., Shanei A., Salamat A. H. *et al.*, (2015.) *"Anthropometric predictive equations for estimating body composition,"* Advanced Biomedical Research, vol. 4, p. 34.

16.Sen, A.and Srivastava, M. (2000). *Regression Analysis*, Springer Verlag. The Radical Statistician by Jim Higgins, Ed.D. Copyright 2005.

17.Udom, A. U. (2010); *Elements of Applied Mathematical Statistics*, icidreserach@gmail.com; ICIDR Publishing House, Ikot Ekpene, Akwa-Ibom State, Nigeria.

18.Wikipedia (2018); *Body Mass Index and Lean Body Mass; http://en.wikipedia.org/wiki/Body Mass Index and Lean Body Mass*

19.World Health Organization (2006, 2012); "*BMI Classification*" *Global Database on Body Mass Index*. Retrieved July 27, 2012.

20.Yu S., Visvanathan T., Field J. et al., (2013) "*Lean body mass: the development and validation of prediction equations in healthy adults*," BMC pharmacology & toxicology, vol. 14, p. 53.

21.Yu S., Visvanathan T., Field J. et al., (2013). *"Lean body mass: the development and validation of prediction equations in healthy adults*," BMC pharmacology & toxicology, vol. 14, p. 53.

22.Kulkarni B., Kuper H., Taylor A. *et al.*, (2013). "*Development and validation of anthropometric prediction equations for estimation of lean body mass and appendicular lean soft tissue in Indian men and women*," Journal of Applied Physiology, vol. 115, no. 8, pp. 1156–1162.