VOL. 30 NO. 9
ISSN: 2878-1894
SEPTEMBER, 2023

*African Scholar Publications & Research International*

*African Scholars Journal of Science Innovation & Tech. Research (JSITR-9)*

# Comparison of Simple Regression Models via Criteria Measures

***Esemokumo Perewarebo Akpos; & **Opara Jude**

*Department of Statistics, School of Applied Science, Federal Polytechnic Ekewe, Yenagoa, Bayelsa State, Nigeria. **Department of Mathematics and Statistics, Ignatius Ajuru University of Education Rivers State P.M.B. 5047, Port Harcourt, Rivers Nigeria

## *Abstract*

The study is on comparison of simple regression models via criteria measures. The source of data set used for this study was secondary, on weight and pulse rate of 90 patients. The response variable is pulse rate, while the explanatory variable is weight. Ten regression models; Linear, Quadratic, Polynomial, Logarithmic, Hyperbolic, power, Exponential growth, Square root, Sinusoidal and Arctangent were stated and employed for the study. For ease of data analysis, E-views package was implemented. Three model selection criteria measures for comparison known as; Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC) with Hannan-Quinn Information Criterion (HQIC) was employed. The result revealed that the polynomial regression model outperforms the other nine models studied to examine the relationship between weight and pulse rate of patients. Hence; other models that were not employed in this study should be studied by researchers and as well compare results with other goodness of fit measures other than the criteria measures employed in this study.

**Key words:** AIC, SIC, HQIC, Simple Linear Regression, Simple Nonlinear Regression, Model Comparison

## Introduction

Fitting simple regression models to data is normally employed within all fields of science; pharmaceutical and biochemical assay quantification, even though fitting a simple linear model to data seldom arises, because most data tend to follow nonlinear models (Duong & Lim, 2023). Nonlinear models exist, and the choice of selecting the right model for the data is a mixture of experience, knowledge about the underlying

process and statistical interpretation of the fitting outcome (Esemokumo et al, 2020). It is of paramount important in quantifying the validity of a fit by some measure which discriminates a 'good' from a 'bad' fit. Many researchers usually employ a common measure known as the coefficient of determination ($R^2$) used in linear regression when conducting calibration experiments for samples to be quantified (Montgomery et al, 2006).

Hence, in the linear perspective, this measure is very intuitive as values between 0 and 1 produce an easy interpretation of how much of the variance in the data is explained by the fit (Chicco et al, 2021). Even though for some time, it has been established that $R^2$ is an inadequate measure for nonlinear regression, many scientists and researchers still make use of it in studies dealing with nonlinear data analysis (Berk, 2020). Several initial and older descriptions for $R^2$ being of no avail in nonlinear fitting had pointed out this issue but have probably fallen into oblivion (Bartlett et al, 2020). This observation might be due to differences in the mathematical background of trained statisticians and researchers who often employ statistical methods but lack detailed statistical insight (Spiess and Neumeyer, 2010).

Having stated that researchers indiscriminately employ $R^2$ as a means of assessing the validity of a particular model when dealing with nonlinear data fit, it is stated that $R^2$ is not an optimal choice in a nonlinear regime as the total sum-of-squares (TSS) is not equal to the regression sum-of-squares (REGSS) plus the residual sum-of-squares (RSS), as is the case in linear regression, and hence it lacks the appropriate interpretation. The rationale behind a high occurrence in solely using $R^2$ values in the validity of nonlinear models could be as a result of researchers not being aware of this misconception.

Since the use of only $R^2$ to access the performance of nonlinear data analysis has been discouraged, this study would employ only the three criteria measures known as; Akaike Information Criterion, Schwarz Information Criterion, and Hannan-Quinn Information Criterion for model selection, proper interpretation and conclusion.

**Statement of Problem**

Medically speaking, it has been established that there is a linear relationship between pulse rate and weight of patients. However, many researchers especially those in other areas who probably do not have sufficient knowledge of statistics usually employed the linear regression technique to establish a relationship between these two variables, without looking at the nonlinear models. It is as a result of the situation that this study intends to look at various non-linear models versus linear model to establish the best model for pulse rate and weight of patients for the data gathered for this study.

**Methodology**

**Regression Models**

Eight Regression models are considered in this study, which are Linear, Quadratic, Polynomial, Logarithmic, Hyperbolic, power, Exponential growth, Square root, Sinusoidal and Arctangent model as written in Equations (1), (2), (3), (4), (5), (6), (7), (8), (9) and (10) respectively

$$Y = \phi_0 + \phi_1 Z + \varepsilon \quad (1)$$

$$Y = \phi_0 + \phi_1 Z + \varphi_2 Z^2 + \varepsilon \quad (2)$$

$$Y = \phi_0 + \phi_1 Z + \varphi_2 Z^2 + \varphi_3 Z^3 + \varepsilon \quad (3)$$

$$Y = \phi_0 + \phi_1 \ln(Z) + \varepsilon \quad (4)$$

$$Y = \phi_0 + \phi_1 (1/Z) + \varepsilon \quad (5)$$

$$Y = \phi_0 Z^{\hat{\phi}_1} + \varepsilon \quad (6)$$

$$Y = \phi_0 + \exp(\phi_1 z) + \varepsilon \quad (7)$$

$$Y = \phi_0 + \phi_1 \sqrt{z} + \varepsilon \quad (8)$$

$$Y = \phi_0 + \phi_1 Sin(Z) + \varepsilon \quad (9)$$

$$Y = \phi_0 + \phi_1 \arctan(\phi_2 Z + \phi_3) + \varepsilon \quad (10)$$

**Simple Linear Regression**

This is a regression line involving only two variables as it is applicable in this study. A widely used procedure for obtaining the regression line of Y and Z is the least square method.

The linear regression of Y on Z is stated in Equation (1)

If there are n pairs of sample observations $(Z_1, Y_1), (Z_2, Y_2), \cdots, (Z_n, Y_n)$, then we get

$$Y_i = \phi_0 + \phi_1 Z_i + \varepsilon_i, \ \ i = 1,2,\cdots,n \quad \ldots \quad (11)$$

Then seeking for the estimators $\hat{\phi}_0$ and $\hat{\phi}_1$ of $\phi_0$ and $\phi_1$ respectively in such a way that K is minimized.

Let $K = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \phi_0 - \phi_1 Z_i)^2 \quad \ldots \quad (12)$

Differentiate (12) partially w.r.t. $\phi_0$ and $\phi_1$, to get Equations (13) and (14) respectively

$$\sum_{i=1}^{n} Y_i - n\phi_0 - \phi_1 \sum_{i=1}^{n} Z_i = 0 \quad \ldots \quad (13)$$

$$\sum Z_i Y_i - \phi_0 \sum Z_i - \phi_1 \sum Z_i^2 = 0 \quad \ldots \quad (14)$$

Solving Equations (13) and (14) simultaneously, we get

$$\hat{\phi}_1 = \frac{n\Sigma Z_i Y_i - \Sigma Z_i Y_i}{n\Sigma Z_i^2 - (\Sigma Z_i)^2} \quad \ldots \quad (15)$$

$$\hat{\phi}_0 = \frac{\Sigma Z_i^2 \Sigma Z_i - \Sigma Z \Sigma Z y}{n\Sigma Z_i^2 - (\Sigma Z)^2} \quad \ldots \quad (16)$$

The calculation is usually set out in ANOVA form as shown (See Table A)

Table 1: Regression ANOVA Table

| Variance | Degree of freedom | Sum of square | Mean square |
|---|---|---|---|
| Regression | 1 | $RSS = \phi_1 \sum zy$ | $\text{RMS} = \frac{\text{RSS}}{1}$ |
| Error | n – 2 | ESS = TSS – RSS | $EMS = \frac{ESS}{n-2}$ |
| Total | n – 1 | TSS = $\sum y^2$ | |

In the same procedure, the parameters of other nonlinear models can be obtained.

Akaike Information Criterion (AIC)

The degree of goodness of fit for an assessed measurable equation is known as AIC (Maguilla et al, 2021) and it can be employed for model choice. It is scientifically characterized as;

$$AIC = \exp^{\frac{2p}{n}} \frac{\sum \hat{e}_i^2}{n} = \exp^{\frac{2p}{n}} \frac{SS_R}{n} \quad (17)$$

where *p*i s the number of parameters with the inclusion of the intercept. Equation (17) is stated mathematically for convenience sake as;

$$\ln(AIC) = \left(\frac{2p}{n}\right) + \ln\left(\frac{SS_R}{n}\right) \quad (18)$$

**Schwarz Information Criterion (SIC)**

The degree of goodness of fit for an evaluated measurable equation is known as SIC (Obaji & Nwagor, 2021) and it can be employed for modelchoice. It is mathematicallycharacterized as

$$SIC = n^{\frac{p}{n}} \frac{\sum \hat{e}_i^2}{n} = n^{\frac{p}{n}} \frac{SS_R}{n} \quad (19)$$

The log of (19) gives (20);

$$\log_e(SIC) = \frac{p}{n}\log_e(n) + \log_e\left(\frac{SS_R}{n}\right) \quad (20)$$

Hannan-Quinn Information Criterion (HQIC)

The degree of goodness of fit for an evaluated measurable equation is known as HQIC (Obaji & Nwagor, 2021) and it can be utilized for model choice. It is mathematically characterized as;

$$HQIC = n\ln\frac{SS_E}{n} + 2p\ln(\ln n) \quad (21)$$

The equation with least AIC, SIC or HQIC value is chosen as the best model.

**Analysis of Data**

The data set used for this study was extracted from a study by Esemokumo (2023), and presented in Table 2.

Table 2: Weight and Pulse Rate of Patients

| S/N | Pulse Rate(bpm) | Weight(kg) | S/N | Pulse Rate(bpm) | Weight(kg) |
|---|---|---|---|---|---|
| 1 | 59 | 67.9 | 46 | 54 | 62.6 |
| 2 | 60 | 69.3 | 47 | 57 | 65.8 |
| 3 | 57 | 65.7 | 48 | 61 | 55.3 |
| 4 | 62 | 70.5 | 49 | 56 | 65.4 |
| 5 | 68 | 76.7 | 50 | 62 | 70.5 |
| 6 | 66 | 74.7 | 51 | 56 | 65.2 |
| 7 | 62 | 70.8 | 52 | 86 | 95.3 |
| 8 | 52 | 60.6 | 53 | 60 | 69.4 |
| 9 | 55 | 63.5 | 54 | 69 | 77.5 |
| 10 | 61 | 69.6 | 55 | 70 | 78.9 |
| 11 | 55 | 64.4 | 56 | 58 | 66.5 |
| 12 | 77 | 85.5 | 57 | 66 | 74.7 |
| 13 | 57 | 65.6 | 58 | 61 | 70.3 |
| 14 | 61 | 69.7 | 59 | 51 | 60.3 |
| 15 | 66 | 74.8 | 60 | 54 | 63.2 |
| 16 | 62 | 70.5 | 61 | 56 | 54.3 |
| 17 | 60 | 68.6 | 62 | 55 | 64.4 |
| 18 | 61 | 70.4 | 63 | 55 | 64.1 |
| 19 | 56 | 65.4 | 64 | 57 | 65.6 |
| 20 | 60 | 68.6 | 65 | 64 | 73.2 |
| 21 | 62 | 70.5 | 66 | 54 | 62.6 |
| 22 | 57 | 65.5 | 67 | 52 | 60.5 |
| 23 | 61 | 70.4 | 68 | 50 | 58.9 |
| 24 | 62 | 70.5 | 69 | 51 | 60.4 |
| 25 | 65 | 73.6 | 70 | 56 | 65.3 |
| 26 | 61 | 70.2 | 71 | 52 | 61.4 |

| | | | | | |
|---|---|---|---|---|---|
| 27 | 65 | 73.6 | 72 | 58 | 67.3 |
| 28 | 62 | 70.5 | 73 | 57 | 65.5 |
| 29 | 56 | 65.4 | 74 | 61 | 70.4 |
| 30 | 56 | 65.3 | 75 | 62 | 70.5 |
| 31 | 69 | 77.6 | 76 | 58 | 67.4 |
| 32 | 73 | 81.8 | 77 | 56 | 65.1 |
| 33 | 65 | 74.3 | 78 | 57 | 65.5 |
| 34 | 71 | 80.4 | 79 | 51 | 60.4 |
| 35 | 62 | 70.5 | 80 | 53 | 61.6 |
| 36 | 55 | 63.8 | 81 | 56 | 65.2 |
| 37 | 71 | 80.2 | 82 | 55 | 63.8 |
| 38 | 56 | 65.2 | 83 | 57 | 65.7 |
| 39 | 59 | 67.8 | 84 | 61 | 69.7 |
| 40 | 52 | 60.5 | 85 | 60 | 68.9 |
| 41 | 52 | 60.6 | 86 | 61 | 70.3 |
| 42 | 49 | 57.8 | 87 | 58 | 66.8 |
| 43 | 61 | 55.8 | 88 | 56 | 64.6 |
| 44 | 56 | 65.4 | 89 | 51 | 60.3 |
| 45 | 59 | 67.9 | 90 | 59 | 67.8 |

Source: Esemokumo (2023).

Table 3: Computer Output for Linear Regression Model

| **Variable** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.** |
|---|---|---|---|---|
| C(1) | 0.189914 | 2.521545 | 0.075317 | 0.9401 |
| Z | 0.872418 | 0.036992 | 23.58394 | 0.0000 |
| R-squared | 0.864739 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | 0.863185 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 2.288640 | Akaike info criterion | | 4.516008 |
| Sum squared resid | 455.6950 | Schwarz criterion | | 4.571933 |
| Log likelihood | -198.9624 | Hannan-Quinn criter. | | 4.538550 |
| Durbin-Watson stat | 2.253396 | | | |

Table 4: Computer Output for Quadratic Regression Model

| **Variable** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.** |
|---|---|---|---|---|
| C(1) | 66.23939 | 13.46168 | 4.920587 | 0.0000 |
| Z | -1.001952 | 0.378142 | -2.649669 | 0.0096 |
| Z^2 | 0.013155 | 0.002644 | 4.975522 | 0.0000 |
| R-squared | 0.894972 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | 0.892530 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 2.028401 | Akaike info criterion | | 4.285499 |
| Sum squared resid | 353.8393 | Schwarz criterion | | 4.369386 |
| Log likelihood | -187.7047 | Hannan-Quinn criter. | | 4.319312 |

| | | | | |
|---|---|---|---|---|
| Durbin-Watson stat | 2.083950 | | | |

Table 5: Computer Output for Polynomial Regression Model

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 550.2054 | 65.31300 | 8.424133 | 0.0000 |
| Z | -21.27569 | 2.716489 | -7.832055 | 0.0000 |
| Z^2 | 0.292646 | 0.037285 | 7.848929 | 0.0000 |
| Z^3 | -0.001267 | 0.000169 | -7.507597 | 0.0000 |
| R-squared | 0.936848 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | 0.934620 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 1.582099 | Akaike info criterion | | 3.799284 |
| Sum squared resid | 212.7581 | Schwarz criterion | | 3.911133 |
| Log likelihood | -165.0682 | Hannan-Quinn criter. | | 3.844367 |
| Durbin-Watson stat | 1.963304 | | | |

Table 6: Computer Output for Logarithmic Regression Model

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | -192.6868 | 12.30470 | -15.65961 | 0.0000 |
| LOG(Z) | 59.83387 | 2.920060 | 20.49063 | 0.0000 |
| R-squared | 0.828357 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | 0.826384 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 2.578129 | Akaike info criterion | | 4.754221 |
| Sum squared resid | 578.2674 | Schwarz criterion | | 4.810145 |
| Log likelihood | -209.5628 | Hannan-Quinn criter. | | 4.776762 |
| Durbin-Watson stat | 2.259601 | | | |

Table 7: Computer Output for Hyperbolic Regression Model

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 118.7580 | 3.391405 | 35.01734 | 0.0000 |
| 1/Z | -3993.392 | 227.1429 | -17.58097 | 0.0000 |
| R-squared | 0.780353 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | 0.777829 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 2.916447 | Akaike info criterion | | 5.000825 |
| Sum squared resid | 739.9926 | Schwarz criterion | | 5.056749 |
| Log likelihood | -220.5367 | Hannan-Quinn criter. | | 5.023366 |
| Durbin-Watson stat | 2.255381 | | | |

Table 8: Computer Output for Power Regression Model

| Y=C(1)*(Z)^C(2) | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 0.827435 | 0.141819 | 5.834456 | 0.0000 |
| C(2) | 1.013246 | 0.040493 | 25.02262 | 0.0000 |
| R-squared | 0.864862 | Mean dependent var | | 59.37778 |

| | | | |
|---|---|---|---|
| Adjusted R-squared | 0.863326 | S.D. dependent var | 6.152697 |
| S.E. of regression | 2.274616 | Akaike info criterion | 4.503471 |
| Sum squared resid | 455.3011 | Schwarz criterion | 4.559022 |
| Log likelihood | -200.6562 | Hannan-Quinn criter. | 4.525872 |
| Durbin-Watson stat | 2.254300 | | |

Table 9: Computer Output for Exponential Growth Regression Model

| **Y=C(1)+EXP(C(2)*Z)** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.** |
|---|---|---|---|---|
| C(1) | 42.24638 | 0.546862 | 77.25229 | 0.0000 |
| C(2) | 0.041278 | 0.000412 | 100.1110 | 0.0000 |
| R-squared | 0.870334 | Mean dependent var | | 59.37778 |
| Adjusted R-squared | 0.868861 | S.D. dependent var | | 6.152697 |
| S.E. of regression | 2.228084 | Akaike info criterion | | 4.462132 |
| Sum squared resid | 436.8634 | Schwarz criterion | | 4.517684 |
| Log likelihood | -198.7960 | Hannan-Quinn criter. | | 4.484534 |
| F-statistic | 590.6691 | Durbin-Watson stat | | 1.918719 |
| Prob(F-statistic) | 0.000000 | | | |

Table 10: Computer Output for Square Root Regression Model

| **Y=C(1)+C(2)*@SQRT(Z)** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.** |
|---|---|---|---|---|
| C(1) | -59.93388 | 5.387667 | -11.12427 | 0.0000 |
| C(2) | 14.50103 | 0.654083 | 22.17001 | 0.0000 |
| R-squared | 0.848147 | Mean dependent var | | 59.37778 |
| Adjusted R-squared | 0.846422 | S.D. dependent var | | 6.152697 |
| S.E. of regression | 2.411184 | Akaike info criterion | | 4.620084 |
| Sum squared resid | 511.6150 | Schwarz criterion | | 4.675636 |
| Log likelihood | -205.9038 | Hannan-Quinn criter. | | 4.642486 |
| F-statistic | 491.5094 | Durbin-Watson stat | | 2.260743 |
| Prob(F-statistic) | 0.000000 | | | |

Table 11: Computer Output for Sinusoidal Regression Model

| **Variable** | **Coefficient** | **Std. Error** | **t-Statistic** | **Prob.** |
|---|---|---|---|---|
| C(1) | 59.36729 | 0.669407 | 88.68637 | 0.0000 |
| SIN(Z) | 0.114116 | 0.885768 | 0.128833 | 0.8978 |
| R-squared | 0.000191 | Mean dependent var | | 59.38202 |
| Adjusted R-squared | -0.011301 | S.D. dependent var | | 6.187424 |
| S.E. of regression | 6.222289 | Akaike info criterion | | 6.516368 |
| Sum squared resid | 3368.369 | Schwarz criterion | | 6.572293 |
| Log likelihood | -287.9784 | Hannan-Quinn criter. | | 6.538910 |
| Durbin-Watson stat | 1.514698 | | | |

Table 12: Computer Output for Arctangent Regression Model

| Y=C(1)+C(2)*@ATAN(C(3)*Z+C(4)) | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C(1) | 68.32450 | 1.380258 | 49.50126 | 0.0000 |
| C(2) | 16.55727 | 2.274517 | 7.279468 | 0.0000 |
| C(3) | 0.074639 | 0.013887 | 5.374913 | 0.0000 |
| C(4) | -5.735048 | 1.016512 | -5.641888 | 0.0000 |
| R-squared | 0.910801 | Mean dependent var | | 59.37778 |
| Adjusted R-squared | 0.907689 | S.D. dependent var | | 6.152697 |
| S.E. of regression | 1.869355 | Akaike info criterion | | 4.132490 |
| Sum squared resid | 300.5260 | Schwarz criterion | | 4.243593 |
| Log likelihood | -181.9621 | Hannan-Quinn criter. | | 4.177294 |
| F-statistic | 292.7114 | Durbin-Watson stat | | 2.149646 |
| Prob(F-statistic) | 0.000000 | | | |

Table 13: Summary Result of Different Models

| Model | AIC | SIC | HQIC |
|---|---|---|---|
| Linear Regression | 4.5160 | 4.5719 | 4.5386 |
| Quadratic Regression | 4.2855 | 4.3694 | 4.3193 |
| Polynomial Regression | 3.7993 | 3.9111 | 3.8444 |
| Logarithmic Regression | 4.7542 | 4.8101 | 4.7768 |
| Hyperbolic Regression | 5.0008 | 5.0567 | 5.0234 |
| Power Regression | 4.5035 | 4.5590 | 4.5259 |
| Exponential Growth Regression | 4.4621 | 4.5177 | 4.4845 |
| Square Root Regression | 4.6201 | 4.6756 | 4.6425 |
| Sinusoidal Regression | 6.5164 | 6.5723 | 6.5389 |
| Arctangent Regression | 4.1325 | 4.2436 | 4.1773 |

*Source*: *E-views software*

It is noticed from Table 13, that polynomial regression model gave the lowest criteria measures for AIC (3.7993), SIC (3.9111), HQIC (3.8444), which implies that the polynomial regression model is the best based on the dataset employed in this study. The second best model is the Arctangent Regression model, which gave its criteria measures for AIC as 4.1325, BIC as 4.2436 and HQIC as 4.1773. Again, Sinusoidal regression model is the least performed equation with highest AIC (6.5164), SIC (6.5723), HQIC (6.5389).

**Conclusion and Recommendation**

The result revealed that the polynomial regression model outperforms the other nine models studied to examine the relationship between weight and pulse rate of patients. Hence; other models that were not employed in this study should be studied by researchers and as well compare results with other goodness of fit measures other than the criteria measures employed in this study.

## References


Bartlett, P. L., Long, P. M., Lugosi, G. & Tsigler, A. (2020). Benign over-fitting in linear regression. Proceedings of the National Academy of Sciences of the USA 117(48):30063–30070.

Berk, R. A. (2020). Statistical learning as a regression problem. In: statistical learning from a regression perspective. Berlin: Springer International Publishing, 1–72.

Chicco, D., Warrens, M. J. & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7(2021), 1-24.

Duong, C. M. & Lim, T.T. (2023). Use of regression models for development of a simple and effective biogas decision-support tool. *Scientific Report*, 13(2023), 1-11.

Esemokumo, A. P., Bekesuoyeibo, M. & Nwobi, A. C. (2020). Model selection in bivariate regression models. *International Journal of Applied Science*, 3(4), 1-8.

Esemokumo, P. E. (2023). Asymmetric distributions and nonlinear functions in a canonical correlation analysis using simulated and real-life medical data. An unpublished PhD Thesis submitted to the department of Mathematics and Statistics, Ignatius Ajuru University of Education Rivers State.

Maguilla, E., Escudero, M., Jiménez-Lobato, V., Díaz-Lifante, Z., Andrés-Camacho, C. & Arroyo, J. (2021). Polyploidy expands the range of centaurium (Gentianaceae). *Frontiers in Plant Science*, 12(2021), 1-12.

Montgomery, D. C., Peck, E. A. & Vining, G. G. (2006). Introduction to Linear Regression Analysis. Wiley & Sons, Hoboken.

Obaji, I. & Nwagor, P. (2021). Multiple regression model selection via birth weight, mother age and gestation variables. International Journal of Statistics and Applied Mathematics, 6(6), 83-90.

Spiess, A. & Neumeyer, N. (2010). An evaluation of $R^2$ as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol*, 10(2010), 34-45.