International Journal of Research (IJR)

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

The Statistical Bias in Genetic Model Analysis with Varying Model Parameters

[1]Onu, Obineke Henry, [2]George, Daniel Sokari, [3] Uzoamaka, Esther Chinyem and [4] Okerengwu, Blessing

[1,2,3,4] Department of Mathematics/Statistics, Faculty of Natural and Applied Sciences. Ignatius Ajuru University of Education, Port Harcourt, NIGERIA
Corresponding email: onuobinekehenry@gmail.com

Abstract

This study was presented by considering models with or without interactions for large and small sample sizes. The paper, therefore seeks to analyze the relationships between heredity (response) and the predictors (age and sex) for large and small sample sizes and for model with interaction (full) or without interaction (reduced). Ordinary Least Square (OLS) was employed to analyze the data of heredity, age and sex of human beings obtained from Sameera (2014). Correlation was considered to obtain the degree of relationships among the variables. Grand Mean Absolute Deviation ($GM - AD$) was proposed as a measure of the statistical bias that exists in the relationship between parents and the offspring in genetic studies. It was observed that the grand mean of the model with interaction for small sample size is greater than the grand mean of the model without interaction for small sample size by 95.89%. The grand mean for model with interaction for large sample size was also greater than the grand mean of the model without interaction for large sample size, but with just percentage level of 26.61. The difference between the grand mean for model with interaction for small and large sample size was 8.51%, while that of model without interaction for small and large sample size was 94.87%. This shows that the model without interaction becomes stronger as the sample size increases, even more than the strength gained by model with interaction for increased sample sizes. The correlations between all the variables, heredity, age and sex are all positive, this reveals that, the relationships between parents and offspring in terms of genetic behavior is positive, irrespective of full or reduced models. It also shows that interaction of age and sex in genetic analysis of heredity is encouraged. Heredity and sex has the highest relationship than other pair for both models. The application of $GM - AD$ reveals that large sample size reduces the statistical bias in genetic model analysis, irrespective of whether full or reduced model used. Also, full model reduces the statistical bias in genetic model analysis, irrespective of the sample size. Hence, the proposed measure suggests that for bias to be reduced in genetic analysis, the sample size should be large and full model (with interaction) be used.

Keywords: Genetic Model; OLS; Full and Reduced Models; Bias; Grand Mean; Absolute Deviation; $GM - AD$.

1. Introduction

Heredity, Genetic Changes and Formation

Heredity also known as biological inheritance or simply inheritance according to Wikipedia is the process of passing on traits from parents to their offspring through either of sexual or asexual reproduction. The organisms or the offspring cells acquire genetic information relating to their parents by means of heredity, changes between individuals can accumulate and cause species to evolve by means of natural selection. The genetics is the study of heredity in biology.

Genome contains all the information that are required to shape and endure living being. This information is contained in the genome in coded form. DNA that is known as deoxyribonucleic acid is made of long molecules and they are found in chromosomes. In the chromosomes, the DNA is made up of two strands known as nucleotides and they are the orderly placement of

smaller molecules. Alberts et al (2014). Each of these nucleotide in the order entails four dissimilar bases known as adenine-A Thymine-T, Guanine-G or the Cytosine-C. Hydrogen bonds keep the two strands together and this is done between two conflicting bases. Pairs of bases are produced from bonds like G with C and A with T and they are opposite.

The Biological Process Underlying Inheritance

Meiosis is one of the biological processes and it involves the method of cell disunion which results in formation of sperm in males and in the cells of egg in females. This process is generally called gametes. Somatic cells are made of two copies of each chromosome known as diploid and then gametes have only a copy of each of these chromosome known as haploid. The diploid is generated by the union or the fusion of a sperm and an egg and it is called the zygote and is always, the foremost of all new organism. The meiosis always begins with the team up of homologous chromosomes in a diploid. The chromatids which is in pair is formed by the duplication of each chromosome. As the process ends, each chromosome will have four homologous chromatids. The chromatids are divided into sister otherwise called identical chromatid and non-sister also called non-identical chromatid. The non-sister chromatid sometimes exchanges segments of genetic elements and this process is called crossover. Laird and Lange (2010).

The statistical analysis of genetic information is challenging in spite of the steady appreciation of sample sizes. The Genome Wide Association Study (GWAS) methodology has a limitation in the science that several composite traits are with highly polygenic architectures with numerous feeble effect variants, see Frazer et al (2009) and this is the result of disappeared heritability as seen in Maher, (2008). The part of phenotypic change described by additive genetic effects is known as heritability. As a result, other methodologies were employed as seen latter in this work.

The Linear Mixed Model (LMM) is now serving as the customary back ground for several genetic investigations. They make available much control for confounding features and allows for coming together of inherited effects from various variants and hence allows for combined analysis of multiple traits. The study considered a linear model which is used to study the effect of age and sex on the heredity of human beings which is the regression analysis of the linear model with or without interaction. The linear model with interaction called full model and that without interaction called reduced model are considered for large and small sample sizes.

The search for the statistical bias in the genetic behavior of human beings using a full and reduced linear regression model and the effect of age and sex on the heredity with small and large sample sizes has not been much talked about in the literature. However, Piasecka *et al* (2018) studied the distinctive roles of age, sex and genetics in shaping transcriptional variation of human immune responses to microbial challenges. According to Franz *et al* (2011) behavioral genetics has revealed that apprehensive trait which is like individual difference that arises freely and interacts with environmental influence such as child ill-treatment is heritable. These studies fell short of considering the bias in modelling full and reduced model on small and large sample sizes using age and sex as the predictors and heredity as the response. It was against this backdrop this research was presented.

The work is aimed at determining the statistical bias that exist in the genetic regression analysis and the work seeks to expose the relationship between the parents and the offspring in a genetic process. This will be revealed by building models with or without interaction known as full or reduced linear models respectively, also, the work will consider large and small sample sizes on each of these two models in order to see how the relationship between the parents and the offspring will look like as sample size increases and how it will look like if interaction effect is omitted in the model. The principle of Grand Mean Absolute Deviation ($GM - AD$) will be proposed to investigate the statistical bias that exists between the full and reduced models and also between the large and small sample sizes. This methodology was in line with Iwundu and Onu (2017) for D-AD and G-AD in design and analysis of experiment for D-efficiency Absolute Deviation and G-efficiency Absolute Deviation respectively.

The study employed full linear (with interaction) model and reduced linear (without interaction) model to study the genetic behavior of human beings for large and small sample sizes. Statistically, a set of data is said to be large if $N \geq 30$, if not, it is small. On this note, the large sample size used in this research is 45 persons and the small sample size used is 15 persons. Their age as x_1, sex as x_2 are the predictors while heredity of these set of persons were recorded as the response as well as the age and sex. The data was obtained such that if 0 there is presence of heredity if 1 there is no heredity. If 1 it is female and if 0 it is male. As observed by Gillespie and Martin (2005), that it is hard to conceive that genetic variation in character is completely determined by age 12 and that minor genetic innovations are witnessed for male Neuroticism at age 14 and 16, as well as female Neuroticism at age 14. These minor genetic innovations possibly hint at age-specific genetic effects associated to hormonal or developmental changes during puberty and psychosexual development. The above literature reveals that age and sex have some contributions to the genetic processes.

Piasecka *et al* 2018 studied the distinguishing roles of age, sex and genetics in shaping transcriptional variation of human immune responses to microbial challenges. They created fungal-, bacterial-, and viral-induced resistant transcriptional shapes in a sex- and age-balanced cohort of 1,000 in good physical shape individuals and examined the causes of immune response variation. It was discovered that age and sex had effects on the transcriptional response of most immune-related genes, with the effect of age being extra stimulus-specific in relation to the effect of sex that were mostly shared through the state of affairs of things.

According to Casale (2016), the linear mixed model (LMM) is now serving as the customary structure for many genetic studies. They offer much regulator for confounding factors and allows for coming together of inherited effects from numerous variants and hence allows for combined studies of numerous characters.

The linear regression as expressed in Alberts *et al* (2014) stated that Genome contains all the information that are required to build and endure an organism. This information is contained in the genome in coded form. The DNA known as deoxyribonucleic acid is made of long molecules and they are found in chromosomes. In the chromosomes, the DNA is made up of two strands known as nucleotides and they are the orderly placement of smaller molecules. Each of these nucleotide in the order consists of four dissimilar bases known as Thymine-T, Adenine-A, Cytosine-C, or the Guanine-G. The hydrogen bonds keep the two strands together and this is

done between two opposite bases. These bonds can produce pairs of bases like G with C and A with T and they are complementary. The chromatids which is in pair is formed by the duplication of each chromosome. As the process ends, each chromosome will have four homologous chromatids.

According to Laird and Lange (2010), the chromatids are divided into sister otherwise called identical chromatid and non-sister also called non-identical chromatid. The non-sister chromatid sometimes exchanges segments of genetic elements and this process is called crossover.

Frazer *et al* (2009) asserted that the statistical analysis of genetic information is challenging in spite of the steady appreciation of sample sizes. The GWAS methodology has a limitation in the sense that numerous complex characters have very high polygenic structural design with several feeble effects variants and this is the result of missing heritability as seen in Maher, (2008). The fraction of phenotypic variance explained by additive genetic effects is known as heritability.

Pohlman and Leitner (2003) compared Ordinary Least Square Regression and Logistic Regression using two data sets, each with binary codes of 1 and 2. Where 1 represents the school drop outs and 2 represents those attending private college. The result of both analyses showed that the OLS and the Logistic regression were seen similar. Even the test of significance showed in a like manner an identical result, also, the logistic and OLS values that were independently predicted were correlated in a high level. They concluded that OLS and logistic can be used to test the relationships among the variables with a binary criterion. But after all logistic showed superiority over OLS while predicting the probability of an element and hence, it was recommended as the choice of model for such application. See also Menard 1995, pp 6. Though, the OLS may predict values beyond (0, 1) range, but such analysis can still be useful in testing hypothesis and for classification.

Hellevik (2009), argued against the widespread belief that linear regression should not be used when the dependent variable is dichotomous. The statistical relevance of the arguments against linear analyses, which was stated that the significance tests are inappropriate and that, one risk obtaining results that are not meaningful, are disputed. The violation of homoscedasticity assumptions by linear regression is of little importance practically, this is because, Hellevik in his study stated that an empirical comparison of results reveals almost identical outcomes for the two kinds of significance tests.

The above three literatures are used to support the choice of Ordinary Least Square in a dichotomous (binary) dependent variables in this study, for which most literatures may suggest the use of logistic regression (Generalized least square).

2. Materials and Method

The linear model with interaction called full model can be given as seen in Casale (2016) as

$$y_i = \sum_{f=1}^{f} x_{if}\, \beta_{1\rho} + \sum_{f=1}^{f} x_{i\rho}\beta_f\beta_{\rho+1} + \varphi_e \qquad (1)$$

Which can be reduced as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varphi_e \qquad (2)$$

For model without interaction, we make a difference to (3.2) by introducing the interaction term given as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varphi_e \quad (3)$$

On that note, from equation (3.2) and (3.3)

We have the output vector which is the victor of heredity as seen in Iwundu and Onu (2017) as

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad (4)$$

the input matrix which is the matrix of age and sex is given by

$$X = \begin{pmatrix} x_{11}\ x_{12} \ldots x_1 p \\ x_{21} x_{22} \ldots x_{2p} \\ \vdots \\ x_{n1} x \ldots . . x_{np} \end{pmatrix} \quad (5)$$

The weight vector β which is vector of model parameters is given as

$$\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad (6)$$

And the residual vector φ is given as

$$\underline{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_n \end{pmatrix}, \quad (7)$$

The model in (3.1) can be stated in matrix from as seen in Kutner *et al* (2005) and Iwundu and Onu (2017) as

$$y = x\beta + \varphi, \text{ where } \varphi \sim N(0, \delta_e^2 I_n) \quad (8)$$

And I_n represents an $N \times N$ identity matrix.

Applying least square equation which is given as

$$\underline{\beta} = (X'X)^{-1} X'y \quad (9)$$

On (3.2 and 3.3), in each case, we first obtain $X'X$ by transposing X to get X' and multiply X' by the X in (3.5) to obtain $X'X$.

Obtain the inverse of $X'X$, by finding its determinant $|X'X|$ and then get the cofactors of $X'X$ and transpose the cofactors to get the adjoined matrix of $X'X$ written as Adj $(X'X)$. Now $(X'X)^{-1}$ is given as:

$$(X'X)^{-1} = \frac{Adj(X'X)}{|X'X|}$$

We obtain $X'y$ by multiplying the transpose of X by y given as $X'y$, hence we proceed to finding the parameters $\underline{\hat{\beta}}$ given as:

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

Application of correlation

The coefficient of correlation between two variables x and y is given as

$$r_{yx} = \frac{n\sum xy - (\sum y)(\sum x)}{\sqrt{(n\sum y^2 - (\sum y)^2)(n\sum x^2 - (x)^2)}} \quad (10)$$

Application of the proposed Criterion

Grand Mean Absolute Deviation

Grand Mean Absolute Deviation (GM-AD) is proposed as a measure of checking minimum bias in genetic model analysis. It is given generally as

$$\text{GM} - \text{AD} = |\beta_{01} - \beta_{02}|$$

Where β_{01} is the grand mean for model 1 or the intercept for model 1 and β_{02} is the grand mean for model 2 or the intercept for model 2.

For two genetic models considered in this research, called full model (model with interaction and Reduced model (model without interaction) the Grand Mean Absolute Deviation is given as

$$\text{GM} - \text{AD} = |\beta_{0F} - \beta_{0R}| \quad (11)$$

Where β_{0F} is the grand mean for Full model and β_{0R} is the grand mean for Reduced model.

Searching for bias with large sample size, for Full and Reduce models, we have the Grand Mean Absolute Deviation given as

$$\text{GM} - \text{AD} = \left|\beta_{0F/L} - \beta_{0R/L}\right| \quad (12)$$

Where $\beta_{0F/L}$ is the grand mean of a Full model given large sample size and $\beta_{0R/L}$ is the grand mean of a Reduced model given small sample size.

While Searching for bias with small sample size, for Full and Reduce models, we have the Grand Mean Absolute Deviation given as

$$\text{GM} - \text{AD} = \left|\beta_{0F/S} - \beta_{0R/S}\right| \quad (13)$$

Searching for bias across sample sizes, for Full and Reduce models, we have the Grand Mean Absolute Deviation given as

$$\text{GM} - \text{AD} = \left|\beta_{0F/L} - \beta_{0F/S}\right| \quad (14)$$

OR

$$\text{GM} - \text{AD} = \left|\beta_{0R/L} - \beta_{0R/S}\right| \quad (15)$$

Selection conditions

If $\text{GM} - \text{AD} = |\beta_{01} - \beta_{02}|$ is minimum, means a minimum bias, if maximum, is a maximum bias. Every good estimator is that with minimum bias.

Therefore,

$$\text{if } \left|\beta_{0F/L} - \beta_{0R/L}\right| \leq \left|\beta_{0F/S} - \beta_{0R/S}\right| \quad (16)$$

then, we say that, large sample size reduces statistical bias than the small sample size, irrespective of the models used, if not, then the opposite is true.

Also,

$$\text{If } \left|\beta_{0F/L} - \beta_{0F/S}\right| \leq \left|\beta_{0R/L} - \beta_{0R/S}\right| \quad (17)$$

then we say that full model reduces the statistical bias than the reduced model irrespective of the sample sizes. The nonnegative value lies between 0 and 1. If 0 it shows that there is no bias, if 1, shows that the bias is maximum.

Results and Discussion

For the reduced genetic model with small sample size

From the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e$$

we have the matrix

$$x'x = \begin{pmatrix} 15 & 594 & 5 \\ 594 & 24846 & 205 \\ 5 & 205 & 5 \end{pmatrix}$$

The inverse of the matrix $(x'x)$ is obtained as;

$$(x'x)^{-1} = \begin{pmatrix} 1.26 & -0.03 & -0.04 \\ -0.03 & 0 & -0.002 \\ -0.04 & -0.002 & 0.30 \end{pmatrix}$$

The transpose of x multiplied by the response y is obtained as

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

$$(x'y) = \begin{pmatrix} 8 \\ 329 \\ 4 \end{pmatrix}$$

Using the formula

$$\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (x'x)^{-1}x'y$$

we obtain

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0.03 \\ -0.248 \\ 0.222 \end{pmatrix}$$

Other results follow from the above and they are summarized in table 1.

For full genetic model with small sample size

The results are as shown in table 1

For reduced genetic model with large sample size

The results are summarized in table 1

For full genetic model with large sample size

The results are as shown in table 1

Applying correlation to the study to obtain the degree of the relationships that exist among the variables under study

The correlation coefficient between hereditary (y) and age (x_1) as seen in Nwagozie (2011) and Kutner et al (2005) is given as applied to equation (3.9).

For small sample size

The results are summarized in table 2

For large sample size

The results are as shown in table 2

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

Table 1: Comparing the parameters of full model with the reduced model for large and small sample sizes

Model Type		**Small sample size**		
		Parameters		
	β_0	β_1	β_2	β_{12}
Full model	0.7296	-0.0085	-1.6214	0.00497
Reduced model	0.03	-0.248	0.222	
		Large sample size		
		Parameters		
Full model	0.7975	-0.0058	-0.5660	0.0122
Reduced model	0.5853	0.0008	0.2478	

Table 2: Comparing the correlation of the variables with full and Reduced models

Correlations			
Sample size	x_1x_2	yx_1	yx_2
Large	0.19	0.06	0.21
Small	0.105	0.174	0.378

Table 3: Comparison of the Grand Mean Absolute Deviations of full-reduced models for large and small sample sizes

Large sample			Large sample	$\beta_{0F/L} - \beta_{0R/L}$	$\beta_{0F/S} - \beta_{0R/S}$
$\beta_{0F/L}$	Full model 0.7975	$\beta_{0R/L}$	Reduced model 0.5853	0.2122	

$\beta_{0F/S}$	Small sample 0.7296	$\beta_{0R/S}$	Small sample 0.03		0.6996

Table 4: Comparison of Grand Mean Absolute Deviation of full-full and reduced-reduced models for large and small sample sizes

	Large sample		Large sample
$\beta_{0F/L}$	Full model 0.7975	$\beta_{0R/L}$	Reduced model 0.5853
$\beta_{0F/S}$	Small sample 0.7296	$\beta_{0R/S}$	Small sample 0.03
$\beta_{0F/L} - \beta_{0F/S}$	0.0679		
$\beta_{0R/L} - \beta_{0R/S}$			0.5553

4. Discussion of Results

It was found as seen in table 1, that the grand mean of the model with interaction for small sample size is greater than that of reduced model for small sample size by 95.89%. The grand mean for model with interaction for large sample size was also greater than that for model without interaction for large sample size, but with just percentage level of 26.61. The difference between the grand mean for model with interaction for small and large sample size was 8.51%, while that of model without interaction for small and large sample size was 94.87%. This shows that the model without interaction becomes stronger as the sample size increases, even more than the strength gained by model with interaction for increased sample sizes. The relationship between heredity and sex, heredity and age and age and sex was found to be in ascending order from heredity and sex to age and sex. The contribution of age for model with interaction was greater than that for model without interaction for small sample size, both showing negative contribution. For large sample size, the contribution of age reduced negatively for model with interaction and reduced significantly in positive direction for model without interaction. Sex for model with interaction was found to be negative while for model without interaction, it was positive for small sample size and similar trend was observed for large sample size. Interaction effect was found more favorable for model with interaction than for that without interaction with 59.26%. The correlations between all the variables, as seen in table 2 shows that, heredity, age and sex are all positive, this reveals that, the relationships between parents and offspring in terms of genetic behavior is positive, irrespective of full or reduced models. The correlation between

age and sex in full model was found greater than the correlation between age and sex for reduced model by 0.085, while the correlation between heredity and age for reduced model was found greater than that of full model by 0.114. Also, the correlation between heredity and sex for full model was found greater than that of reduced model by 0.168. The contribution of heredity and sex was found superior for both full and reduced model, showing that the correlation of heredity and sex is stronger irrespective of the model type used. While, the correlation between heredity and age is insignificant for full model but increases by 0.114 in reduced model. Heredity and sex has the highest relationship than other pair for both models. The application of $GM - AD$ reveals that large sample size reduces the statistical bias in genetic model analysis, irrespective of full or reduced model used as seen in table 3. Also, table 4 shows that full model reduces the statistical bias in genetic model analysis, irrespective of the sample size.

Conclusion

Some statistical bias exists in genetic models between large and small sample sizes and for model with interaction and that without interaction. The increase in the sample size favors model without interaction more than it favors model with interaction. The relationship between heredity and sex is stronger than that between heredity and age, the latter is stronger than that between age and sex. The grand mean for the both models considered were all positive for both small and large sample sizes. The bias that exist between the grand mean of model with interaction and the model without interaction for small sample size is about 0.6996, while the bias reduced for a large sample size to about 0.2122. Generally, the work suggests that the larger the sample size the smaller the bias for both full and reduced models. That is to say, it is conjecturally believed that at a certain large sample size, the grand mean of the model with interaction will be equal to that of model without interaction. The bias that exists between large and small samples for model with interaction is 0.0679, while that of large and small samples for model without interaction is 0.5553. This means that full model offers a better estimation for large sample than reduced model.

Recommendation

In this research, it is recommended that, in studying the effect of age, sex and other factors on the heredity, model with interaction should be built when the sample size is small, while the model without interaction be built for a large sample size in order to reduce the statistical bias in the analysis. But generally, large sample size reduces the level of statistical bias in a genetic model analysis.

References

Alberts, B., Alexander J., Julian L., David M., Martin R., Keith R., and Peter W. (2014). Molecular Biology of the Cell. 6th ed. *Garland Science.*

Casale F. P., (2016). *Multivariate Linear Mixed Models for Statistical Genetics, European Bioinformatics Institute. Doctor of Philosophy Dissertation.*

Fink, G., Sumner, B., Rosie, R., Wilson, H., and McQueen, J. (1999). Androgen actions on central serotonin neurotransmission: Relevance for mood, mental state and memory. *Behav. Brain Res. 105(1):53–68.*

Frazer, K. A., Sarah S. M., Nicholas J. S., and Eric J. T. (2009). Human genetic variation and its contribution to complex traits." *Nat. Rev. Genet. 10.4, pp. 241–51.*

Gillespie N. A. and Martin N. G. (2005). Multivariate Genetic Analysis Volume 3, pp. 1363 – 1370 *in Encyclopedia of Statistics in Behavioral Science Brian S. Everitt & David C. Howell John Wiley & Sons, Ltd, Chichester.*

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

Kutner M. H., Nschtsheim C. J., Neter J. and Li W. (2005); *Applied linear statistical model, fifth edition, McGraw-Hili t:a Irwin, Boston Burr RIdge, IL Dubuque, IA MadIson, WI New York San FrancIsco St LoUIs Bangkok Bogota Caracas Kuala Lumpur LIsbon London Madnd MexIco CIty MIlan Montreal New Deihl SantIago Seoul Smgapore Sydney TaIpeI Toronto*

Laird, N. M. and Christoph L. (2010). The Fundamentals of Modern Statistical Genetics. *Springer Science & Business Media.*

Maher, B. (2008). Personal genomes: The case of the missing heritability." Nature 456.7218, pp. 18–21.

Menard, S. (1995). *Applied Logistic Regression Analysis. Sage University Paper Series on Qualtitative Applications in the Social Sciences, 07-106, Thousand Oaks, CA: Sage.*

Nwaogazie I. L. (2011); Probability and statistics for science and engineering practice. *De-Adroit innovation 13 Annang Str. OguiN/L out Enugu.*

Piasecka, B., Duffy, D., Urrutia, A., Quach,, H., Patin, E., Posseme, C., Bergstedt, J., Charbit, B., Rouilly, V., MacPherson, C. R., Hasan, M., Albaud, B., Gentien, D., Fellay, J., Albert, M. L., Quintana-Murci, L., and Intérieur M. C., PNAS (2016), 2018 115 (3) E488-E497.

Pohlman J. T. and Leitner D. W. (2003), A comparison of Ordinary Least Squares and Logistic Regression. *The Ohio Journal of Science 103(5).*

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

Table 5: Genetic data used in the study for small sample size($N < 30$)

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Vol. 8 Issue 6
June 2021

The values in table 6, from 1 to 15 make up table 5.

Table6: Genetic data used in the study for large sample size ($N \geq 30$)

S/N	Hereditary (y)	Age (x_1)	Sex (x_2)
1	0	24	1
2	1	49	1
3	1	34	0
4	1	48	1
5	1	40	1
6	0	46	0
7	0	21	0
8	1	44	1
9	1	46	0
10	0	31	0
11	0	45	0
12	0	48	0
13	1	41	0
14	1	27	0
15	0	50	0
16	1	32	0
17	1	40	0
18	1	100	1
19	1	55	1
20	1	62	0
21	0	80	0
22	1	60	0
23	1	25	0
24	1	70	0
25	1	50	0
26	1	45	0
27	1	50	0
28	1	65	1
29	0	28	0
30	1	60	0
31	1	33	0
32	1	35	0
33	1	45	0
34	1	46	0
35	0	50	0
36	1	55	0
37	0	54	0
38	0	44	0
39	0	50	0
40	0	60	0
41	1	50	0
42	1	47	0
43	1	45	0
44	0	60	0
45	1	40	0

Source: Sammera (2014)