International Journal of Innovative Research and Advanced Studies (IJIRAS)
Volume 6 Issue 6, June 2019

ISSN: 2394-4404

Logistic Regression Model For Determining The Gender Of A Child Using Age Of The Mother And Month Of Conception

Egemba Chukwuka Raymond

Victoria Adah

Department of Mathematics/Statistics/Computer Science, College of science, University of Agriculture, Makurdi, Nigeria

Bashiru Omeiza Sule

Department of Mathematical Sciences, Faculty of Natural Science, Kogi State University, Anyigba, Kogi State, Nigeria

***Abstract:** The issue of sex determination (before conception) is beyond the study of Science. Though current studies have developed means to determine gender, it is not until the baby has been conceived. After conception, scientific methods can be used to determine whether the baby will be a boy or a girl. This study seeks to use the age of the mother and the month in which she conceives to predict the gender of the baby, using appropriate statistical methods. The data used for this study was extracted from the delivery books of the Maternity Department of General hospital Okehi local government in kogi state of Nigeria from 2010-2014. Logistic regression was used to analyze the data since the dependent variable has only two possible values (Male or Female). The results reveal that the odds of estimating correctly the gender of a baby improves by only 0.3% if one knows the age of the mother and reduces by 0.6 % if one knows the month within which the child was conceived.*

***Keywords:** Sex Determination, Logistic Regression, Age, Month of Conception, Prediction, Odds*

IJIRAS

I. INTRODUCTION

Gender is the state of being male or female. The underlying mechanism by which an individual develops into a male or female is called sex determination (Roberk J.B.2012). Over the course of human history, humans have been investigating how to determine the gender of a baby. Religious (i.e Islam and Christianity) beliefs have it that is only God that knows the gender of the baby that will be born.

According to biologists, the both parents have sex chromosomes. According to Hemophilia of Georgia, the father usually have one X and one Y chromosomes and while the mother have two X chromosomes. When an egg or sperm is made, it only gets one of the sex chromosomes from the parent, that is the mother can only make eggs with an X chromosome but the father can make either X or Y sperm. This means that the father has a higher likelihood of determining the sex of a baby but the father does not have control on the determination of the gender of the baby since he cannot determine whether he is going to produce an X or Y chromosome.

This Global age, scientists have also found a way to determine the sex of a baby using ultra sound scanning method, the ultrasound method uses high-frequency sound waves to produce an image on a screen of the baby in the mother’s uterus. The scans are typically done twice during pregnancy, but the one done between 18 and 22 weeks is when the sonographer (ultrasound technician) might identify the gender of the baby (Cari N 2014).

Due to these search for answers on the factors that are used to determine the gender of a baby, the current effort enable us use a statistical tool in determining the gender of a baby using the age of the mother and month of conception and to the best of the researchers knowledge, this study has not been conducted in the area under study.

ISSN: 2394-4404

II. THE DATA

The data used for this study was extracted from the delivery books of the maternity Department of Kogi State Specialist Hospital, Obangede which is located at Okehi Local government area of kogi State. The hospital provides a 24 hours service to the people of Okehi and beyond.

The maternity department of the hospital has records on the vitals of new born babies whose birth occurred at the hospital. The records show detailed information on deliveries at the hospital. The information include the age of the mother at birth, age of the mother at birth, sex of the baby, the date and time of delivery, address of the mother, time in which the labor commenced, perineum, parity, height of fundus, membrane, type of delivery, weight and height of the baby APGAR Score. According to American Pregnancy Association, Apgar is a measurement of the newborn's response to birth and life outside the womb. The ratings, APGAR, are based on Appearance (color), Pulse (heartbeat), Grimace (reflex), Activity (muscle tone), and Respiration (breathing). The scores, which are taken at 1 and 5 minutes following birth, range from 1 to 10.

Only data in the period 2010 – 2014 was used.

A. LOGISTIC REGRESSION

Its salient feature is that there is a binary response of interest and the predictor variables are used to model the probability of that response. Here the binary response variable, Y, is:

- ✓ The female is coded as 0; if the child given birth to is female and
- ✓ The male is coded as 1; if the child given birth to male

Logistic regression analysis extends the techniques of multiple regression to research situations in which the outcome variable is categorical. Applications abound in the fields of medicine (Sharareh *et al.* 2010), social sciences (Chuang 1997) and education (B.O. Sule & F.W.O. Saporu 2015).

Menard and Scott (1995) showed that the relationship between the predictor and response variables is not a linear function in logistic regression, instead, the logistic function is used, which is the logit transformation of θ

$$\theta = \frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}}$$

Kleinbaum (1994) presented an alternative way to write the logistic model called the logit form of the model by transforming the logistic model i.e

Logit $P(x) = In_e\left[\frac{P(x)}{1-P(x)}\right]$

Where $P(x) = \frac{1}{1+e^{\alpha+\sum \beta_i X_i}}$

B. LOGISTIC REGRESSION ANALYSIS

Logistic regression is a form of regression which is used when we want to predict probabilities of the presence or absence of a particular disease, characteristics, or an outcome in general, based on a set of independent explanatory variables of any kind. In other words, logistic regression is used to model the relationship between a binary response variable and one or more predictor variables, which may be either discrete or continuous.

For instance, assume that we have a set of independent variables (regressors) that can be used for prediction in each of the following situations: (1) predicting whether a company's dealer(s) will soon be mired in dire financial straits; (2) predicting whether a hospital patient will survive until been discharged; (3) predicting if a person is likely to develop heart disease. Notice that in each of these scenarios the response variable has only two possible outcomes. That is, the hospital patient will either survive or not survive, a person will either develop heart disease or not develop it, and a dealer will either have or not have financial problem (Ryan 1997, p 225)

To elucidate the popularity of logistic regression, we show here the logistic function, which describes the mathematical form on which the logistic model is based.

The basic form of the logistic function is;

$$P(Z) = \frac{1}{1+e^{-z}} = \frac{e^{z}}{1+e^{z}} \quad (1)$$

C. THE LOGISTIC MODEL

To obtain the logistic model from the logistic function, we let Y be a dichotomous random variable denoting the outcome of some experiment, and let X = $(X_1, X_2, \ldots, X_K)$ be a collection of predictor variables.

Denote the conditional probability that the outcome is present by

P(Y=1/x) = P(X)…………….. (2)

Where P(x) has the form:

$$P(X) = \frac{1}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}} = \frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}} \quad (3)$$

If the values of X are varied, and the n values $y_1, y_2, \ldots, y_n$ of Y are observed, we write

$$P_i = \frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_n X_n)}} \quad (4)$$

Alternatively, if we define P_1 = P(y=1/x) = P(X_1) of an object belonging to group 1 and P_0 = P(Y=0/X) = P(0/x)=P(X_0) as the probability of an object belonging to group 0, the logistic regression model has the form:

Logit $p(X) = Log\left[\frac{p(x)}{1-p(x)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X$ k (5)

Where

$$P(X) = \frac{1}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+...+\beta_k X_k)}}$$

Equation (5) is called the logit form of logistic regression model, the logit transformation denoted as Logit P(X) enables us to compute the log odd ratio of probability of belonging to

IJIRAS

 www.ijiras.com | Email: contact@ijiras.com

group 1, p(y=1/x) to probability of belonging to group 0, P(y=0/x), which is equally given by

$$\log\left(\frac{p_i}{1-p_i}\right)=\sum_{n=0}^{k} X_{ik}\beta_k \qquad (6)$$

i= 1,2,…N

i.e. the logistic regression model equated the logit transform, the log-odds of probability of a success, to a linear component

STATISTIC	AGE
N	805
Mean	27.2559
Median	27.0000
Mode	25.00
Std. Deviation	5.64379
Skewness	.556
Std. Error of Skewness	.086
Minimum	13.00
Maximum	55.00

Table 1: Descriptive Statistics for the Age of Mothers from 2010-2014

A GENERAL DESCRIPTIVE ANALYSES OF DATA FOR THE PERIOD 2010 – 2014

Table 1 showed that a total of 805 children were born in the five years period under consideration. As in other years, most of the women are in their twenties with the minimum age being 13 years and the maximum 55 years.

The age distribution has a standard deviation of 5 years with a range of 36 years. Further, Table 1 shows that the oldest woman to give birth was at the age of 50 while the youngest was 14 years old. There was a general Skewness of about 0.556.

Age	Frequency	Percent	Cumulative Percent
13	1	.1	.1
15	1	.1	.2
17	5	.6	.9
18	18	2.2	3.1
19	18	2.2	5.3
20	67	8.3	13.7
21	27	3.4	17.0
22	43	5.3	22.4
23	34	4.2	26.6
24	30	3.7	30.3
25	116	14.4	44.7
26	38	4.7	49.4
27	45	5.6	55.0
28	48	6.0	61.0
29	24	3.0	64.0
30	94	11.7	75.7
31	12	1.5	77.1
32	50	6.2	83.4
33	16	2.0	85.3
34	11	1.4	86.7
35	51	6.3	93.0
36	11	1.4	94.4
37	9	1.1	95.5
38	15	1.9	97.4
39	2	.2	97.6
40	11	1.4	99.0
42	1	.1	99.1
43	1	.1	99.3
45	4	.5	99.8
48	1	.1	99.9
55	1	.1	100.0
Total	805	100.0	

Table 2: Distribution of Ages of Women who gave Birth (2010-2014)

In Table 3, 451babies who are boys were correctly classified as boys when one does not know the age of the mother and the month of delivery, and 354 were wrongly classified as boys. Thus, if one simply guesses the gender of a baby, one would classify 56% of them correctly by chance.

Classification Table

Observed			Predicted gender Female	Predicted gender male	Percentage Correct
Step 0	gender	Female	0	354	.0
		Male	0	451	100.0
	Overall Percentage				56.0

a. Constant is included in the model.

b. The cut value is .500

Table 3: A Classification Table for Gender

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.242	.071	11.631	1	.001	1.274

Table 4: Effects of Variables in the Equation

Table 5 shows that the two variables (age and month) are individually not significant predictors of the gender of a baby, age has a score of 0.074, 1 degree of freedom and significance of 0.786, while month has a score of 0.074, degree of freedom of 1 and significance of 0.785. however, the month of conception does a better job than the age of the mother in predicting the gender of a baby because the significance level is lesser.

Variables not in the Equation

			Score	Df	Sig.
Step 0	Variables	age	.074	1	.786
		month	.074	1	.785
	Overall Statistics		.150	2	.928

Table 5: Effects of Variables not in the Equation

The omnibus test of model coefficients in table 6 indicates that, when both age and month are entered together, the model or the equation is still not significant ($\chi2$ =0.15, d.f = 2, N = 805, P < 0.928). Thus when both predictor variables (age and month) are considered together, they do not significantly predict whether or not a child’s gender will be male or female. This is evident from table 6.

IJIRAS

ISSN: 2394-4404

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	.150	2	.928
	Block	.150	2	.928
	Model	.150	2	.928

Table 6: Omnibus Test of Model Coefficients

Table 7 presents the odds ratios, which suggest that the odds of estimating correctly the gender of a baby improves by only 0.3% if one knows the age of the mother and worsen by 0.6% if one knows the month within which the child was conceived. The model summary estimates the percentage variance accounted for in the use of the model. It indicates only -0.6% or 0.3% of the variance, in whether a baby is a girl or a boy, can be predicted by the linear combination of the two variables.

Variables in the Equation

B	S.E.	Wald	df	Sig.
.003	.013	.075	1	.784
-.006	.022	.076	1	.783
.187	.377	.246	1	.620

a. Variable(s) entered on step 1: age, month

Table 7: Results of Logistic Regression

The final classification table (Table 8) indicates how well the combination of the variables predicts gender. Overall, 56.6% of the babies were predicted correctly. The independent variables were better at helping to determine who would be a female (87.4%) correctly than who would be a female (17.5%). The odds ratios in Table 8 indicates that the odds of estimating correctly the gender of a baby improve by 0.3% if one knows the age of the mother and worsen by 0.6% if one knows the month of conception of the baby.

Classification Table[a]

			Predicted		
			gender		Percentage Correct
Observed			female	male	
Step 1	gender	female	62	292	17.5
		male	57	394	87.4
	Overall Percentage				56.6

a. The cut value is .500

Table 8: Final Classification Table

III. CONCLUSION

The success story is that there is a relationship between the age of the mother, the month of conception and the gender of her baby based on the percentages of births of males and females. The research has revealed that the gender of a child is generally independent of the month of conception. Also the gender of a baby is independent of the age of the mother. On using the age of the mother and the month of conception to predict the gender of the baby, the month of conception does a better prediction than the age of the mother. Thus in predicting the gender of a baby, though both are statistically independent, the month of conception gives a better prediction than the age of the mother. Hence the gender of a child can be predicted using the month of conception with about 50% success. However, there can also be a 50% success if one guesses the gender of the baby.

ACKNOWLEDGMENT

The authors wish to thank the Maternity Department of General hospital Okehi Local Government in Kogi State of Nigeria for data access, helps and advice.

REFERENCES

[1] American Pregnancy Association, Labor and Birth terms to know. http://americanpregnancy.org/labor-and-birth/labor-and-birth-terms-to-know/ 2014
[2] Cari Nierenberg "Am I Having a Boy or Girl? – Ultrasound and Sex Prediction" 2014
[3] Chuang, H.L, "High school youth's drop out and re-enrollment behavior", Economics of Education review. Vol 16 no. 2, pp 171 – 186.1997
[4] Hemophilia of Georgia(201) Basic Genetics/ http://www.hog.org/handbook/section/2/basic-genetics visited on 06/25/16
[5] Kleinbaum, D. G., Kupper, L. L. and Morgenstern H., "Epidemiologic research: Principles and quantitative methods."Toronto lifetime learning, 461-470, 1982.
[6] Menard, S.," Applied Logistic Regression Analysis". Sage publications Ltd United Kingdom, 1995
[7] Robert J.B, Concepts of Genetics (companion site) 2nd Edition Chapter 4. www.mhhe.com/brookerconcepts. 2012
[8] Ryan, T. P., "Modern regression methods". John Wiley and Sons,New York USA pp. 255, 1997.
[9] Sharareh, R., Niakan, K., Mahshid., N. & Zeng, X. "A logistic regression model to predict high risk patients to fail in tuberculosis treatment course completion", International Journal of Applied Mathematics. IJAM vol 40 no. 2, 2010
[10] Sule O.B & Saporu F.W.O, "A Logistic Regression Model of Students' Academic Performance in University of Maiduguri, Maiduguri, Nigeria." Mathetical Theory and Modelling. IISTE Journal Vol.5, no. 1, pp. 124-136, 2015.