COUNT DATA MODELLING APPLICATION

JECINTA UGOCHUKWU IBEJI

218086403

UNIVERSITY OF KWAZULU-NATAL

2019

COUNT DATA MODELLING APPLICATION

JECINTA UGOCHUKWU IBEJI

218086403

A thesis submitted in fulfilment of the academic requirements for the degree of Master of Science (M.Sc. in Statistics)

UNIVERSITY OF KWAZULU-NATAL

SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE

WESTVILLE CAMPUS, DURBAN, SOUTH AFRICA

Thesis advisor: **Professor T. Zewotir**
Professor D. North

Declaration

The research thesis was completed by the candidate as part of a master's degree while based in the Discipline of Statistics, School of Mathematics, Statistics and Computer Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville campus, South Africa.

The contents of this work have not been submitted in any form to another university except where the work of others is referenced in the text. The results reported are studies by the candidate.

As the candidate's supervisor, I have approved this thesis for submission.

------------------------------------- --------

Jecinta Ugochukwu Ibeji (Student) Date

------------------------------------- --------

Professor Temesgen Zewotir (Supervisor) Date

------------------------------------- --------

Professor Delia North (Co-Supervisor) Date

Disclaimer

This document describes work undertaken as part of a master's program of study at the University of KwaZulu-Natal(UKZN). All views and opinions expressed therein remain the sole responsibility of the author and do not necessarily represent those of the University of KwaZulu-Natal.

Abstract

The rapid increase of total children ever born without a proportionate growth in the Nigerian economy has been a concern and making prediction with count data requires applying appropriate regression model.. As count data assumes discrete, non-negative values, a Poisson distribution is the ideal distribution to describe this data, but it is deficient due to equality of variance and mean. This deficiency results in under/over-dispersion and the estimation of the standard errors will be biased rendering the test statistics incorrect. This study aimed to model count data with the application of total children ever born using a Negative Binomial and Generalized Poisson regression The Nigeria Demographic and Health Survey 2013 data of women within the age of 15-49 years were used and three models applied to investigate the factors affecting the number of children ever born. A predictive count modelling was also carried out based on the performance evaluation metrics (root mean square error, mean absolute error, R-squared and mean square error). In the inferential modeling, Generalized Poisson Model was found to be superior with age of household head ($P < .0001$), age of respondent at the time of first birth ($P < .0001$), urban-rural status ($P < .0001$), and religion ($P < .0001$) being significantly associated with total children ever born. In the predictive modeling, all the three models showed almost identical performance evaluation metrics but Poisson regression was chosen as the best because it is the simplest model. In conclusion, early marriage, religious belief and unawareness of women who dwell in rural areas should be checked to control total children ever born in Nigeria.

Table of contents

List of Figures

List of Tables

Acknowledgements

All thanks to God Almighty who has made this day a reality. I could not have gone this far if not for Your mercy.

I would like to thank my supervisors Professor Temesgen Zewotir and Professor Delia North for being supportive.

I specifically appreciate my husband and my daughter for encouraging me throughout this journey. God bless you.

Finally, to all my colleagues especially Lateef, Partson, Samuel and Chigozie, Thank you very much.

Chapter 1: Introduction

1.1 Background of the study

Count data is a type of statistical data in which the observations can take only non-negative integer values $\{0,1,2,\ldots\}$, which arise from counting rather than ranking (Cameron and Trivedi, 2013). Examples of statistical analyses that involve count data are simple counts that include the number of thunderstorms that occurred in a one-year calendar, fatal vehicle accidents per day, customers arriving at the shopping mall per hour, with categorical data being the counts denote the number of items belonging to each category (Adhikari, 2010).

Count data can be applied to different fields, including medicine, agriculture, life science, business, social, behavioural science, and demographic and health survey data. For example, in medicine, Du et al. (2011) opined that researches have shown that the effect and worth of health information technology (HIT) frequently used on outcome measures that are counts of things, such as hospital admissions, the number of laboratory tests per patient, adverse drug events (ADEs), and rates of cardiac arrest. This kind of data gives several analytic challenges, such as a large and perhaps disproportionate number of zero values, slightly high frequency of small integer values, and non-constant variance (where the variance of the residuals differs for different ranges of independent variables) (Zhou et al., 2014). In agriculture, count data was used to describe the implementation of agricultural and natural resource management technologies by small farmers in Central America (Ramirez and Shultz, 2000).

Count data models are most appropriate for a certain type of adoption data. They are applied to investigate the impact of key socio-economic, biophysical, and institutional factors on the implementation of integrated pest management, agroforestry, and soil conservation technologies among small farmers. Factors affecting farmers' quantity decisions related to farming precision technology can also be determined using count data models (Isgin et al., 2008). Furthermore, (Ozmen and Famoye, 2007) in their work "count regression models with an application to zoological data containing structural zeros" the authors opined that count data models can be applied in life science to predict the number of *C. caretta* hatchlings dying from exposure to the sun. Applying count data models in business centres focuses on the consumption of a product. Tuzen and Erbas, (2018) compared count data models with the application of daily consumption of cigarette by young people in Turkey, and found Zero-inflated negative binomial (ZINB) and Negative binomial hurdle (NBH) to be preferable for analysis. The outcome variable in certain financial studies is a count that takes a non-negative integer value. Examples include the number of takeover bids received by a target firm, unpaid credit instalments (for scoring credit), accidents or accident claims (for insurance premia determination) and mortgage loans prepaid (mortgage-backed securities pricing). Cameron and Trivedi, (1996) applied Poisson and Negative Binomial Poisson and Negative Binomial Models for count data which had prominence on the underlying count process and links to dual data on durations. Likewise, modelling count data is a common task in economics and social sciences. According to Zeileis et al., (2008), Hurdle and Zero-inflated model are capable of handling over-dispersion and excess zeros (which are two problems that mostly occur in count data sets of economics and social sciences).

Notwithstanding different models used in modelling count data, Poisson regression has been reported as the main methods for count data modelling, but violates equidispersion hypothesis and confines its use in several real-world applications due to under/over-dispersion (Osuji et al., 2016). This superfluous disparity could lead to inaccurate inferences in the standard errors, tests, confidence intervals and parameter estimates, over-dispersion frequently surfaces for several reasons as well as mechanisms that cause too much zero counts (Guikema and Goffelt, 2008; Maxwell et al., 2018). Accordingly, in various areas, over-dispersed count data are common, which in turn has led to the development of a statistical methodology for modelling these over-dispersed data. Studies on dealing with under-dispersion and over-dispersion issues have been reported (AA and Naing, 2012). Kim and Jun (2016) tried to overcome over-dispersion by using Zero-inflated Poisson and Negative Binomial regression to analyze the death rate of patients infected with AIDs.

Negative Binomial regression and Generalized Poisson regression were used to model count data involving the number of cervical cases to overcome the problem of overdispersion (Melliana et al., 2013). Muluneh et al. (2016) investigated the effects of demographic and socio-economic factors on the number of children ever born by married women of gage 15-49rs in Ethiopia using the ‘quasi-likelihood’ in a generalized linear model to overcome the problem of over and under-dispersion. In some countries, especially where marriage is recognized as a major medium for procreation, data on children ever born are only available for ever married or currently married women. In Nigeria, the current average fertility rate per woman is 5.5 children compared with 5.7 children in 2003 and 2008. Residence and region are major determinants of fertility variation. In urban areas, women have 4.7 children on average, compared with 6.2 children per woman in rural areas. The North-West Zone has the highest fertility rate with and average of 6.7 children per woman, while the South South Zone has the lowest fertility rate of 4.3 children on the average. Fertility also varies with mother’s educational level and economic status (Macro and Commission, 2014). In view of this, the appropriate model to describe the total number of children ever born in Nigeria is still a subject of study.

1.2 Statement of problem

The application of suitable regression model for the analysis of count data has been a subject of concern. Count models allow for regression-type analyses when the dependent variable of interest is a numerical count. These can be used to estimate the effect of a policy intervention either on the average rate or on the probability of no event, single event, or multiple events. Poisson regression has been widely used in recent years. Nevertheless, it is identified that count data in demographic survey often display over or under dispersion. The inappropriate imposition of the Poisson may result in underestimation of the standard error and exaggerate the significance of the regression parameter thus giving false inference about the regression parameter. The research question is will Negative Binomial and Generalized Poisson regression serve as a substitute for handling count data?

1.3 Aim and Objectives

This study is aimed to use Negative Binomial and Generalized Poisson regression model as an alternative for handling count data with the interest in the number of children ever born to respondents in Nigeria.

The study therefore had the following objectives

1. To explore the condition of total children ever born to respondents in Nigeria.
2. To study the socio-economic and demographic factors affecting the total children ever born to respondents in Nigeria using Negative Binomial and Generalized Poisson regression.

1.4 Significance of the study

Poisson distribution has been the most commonly used distribution for modelling count data, but it assumes equal variance with the mean which makes it less appropriate since count data usually show over or under dispersion. The developing and applying statistical models as a substitute for modelling over-dispersed data are important . Some work has been reported on the application of Poisson and Negative Binomial on over-dispersed count data, to the best of our knowledge there is little work on the use of Negative Binomial and Generalized Poisson regression as an alternative in handling count data. Generalized Poisson regression is a useful model for fitting both over-dispersed and under-dispersed count data because it allows for more variability and it is more flexible in analyzing independent variables. In this work Negative Binomial and Generalized Poisson regression are applied as an alternative for handling over-dispersed and under-dispersed count data considering the number of children ever born in Nigeria.

Chapter 2: Data

This study uses secondary data from the individual's questionnaire of Nigeria Demographic and Health Survey 2013 which covers all the regions in the country. In the 2013 Nigeria Demographic and Health Survey demographic, socio-economic and health information were collected for both men and women.

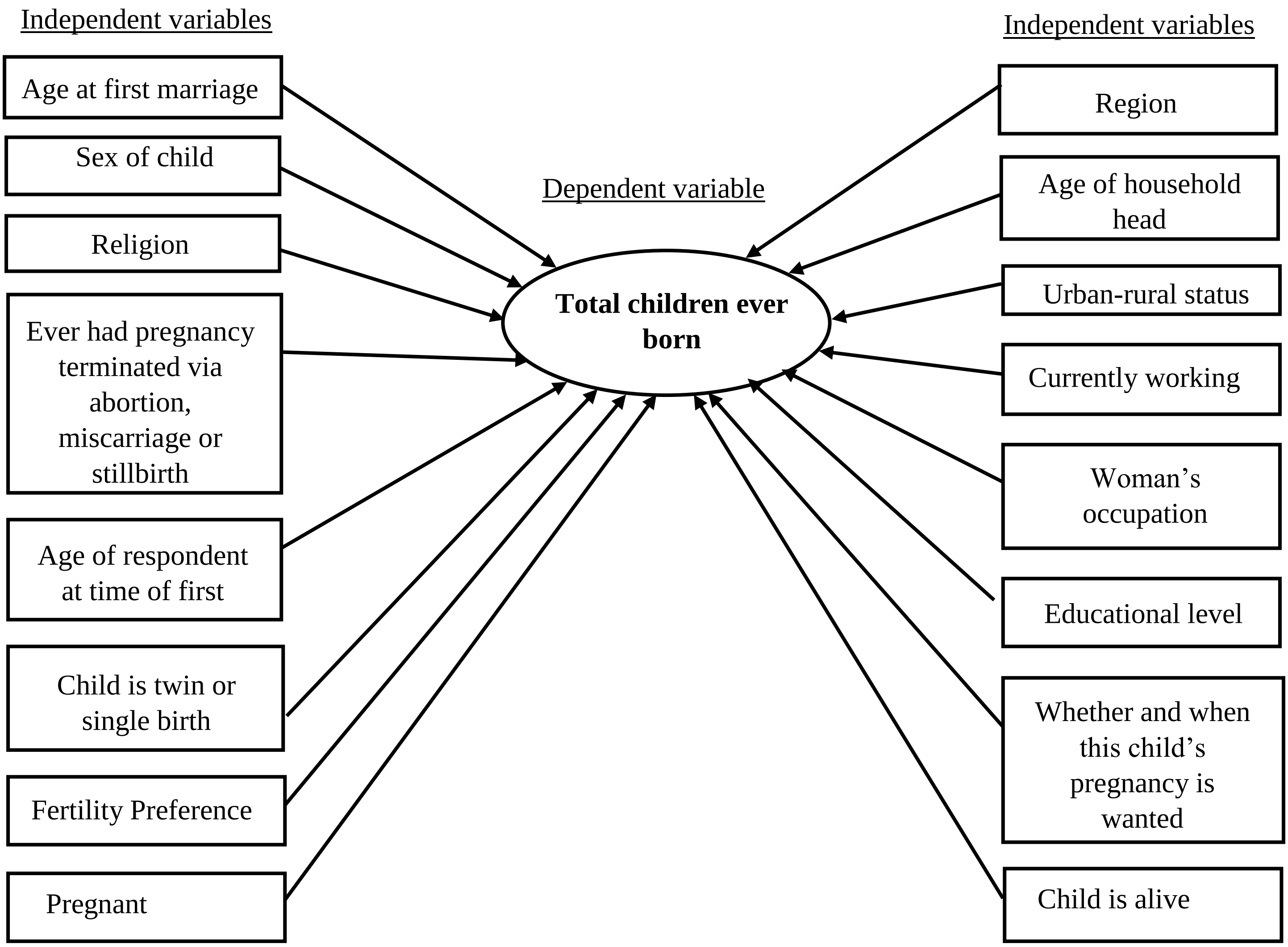


Figure 2.1: Schematic representation of the conceptual framework

Figure 2.1 displays the variables of interest with explanatory variables (called independent variables) consisting of age at first marriage or cohabitation, sex of child, religion, age of respondent at first birth, child is twin or single birth, child is alive, currently pregnant, whether and when this child's pregnancy is wanted, educational level, region, age of household head, ever had pregnancy terminated via abortion, miscarriage or stillbirth, woman's occupation, fertility preference and urban-rural dwelling status . The explanatory variables are chosen based on previous research that they might have some effect on the total children ever born by women in Nigeria. Muluneh et al. (2016) studied the determinant factors of fertility among married women in Ethiopia and found that increased household economic status, residing in urban areas, younger age at first birth and not using contraceptive were significantly associated with high fertility.

Similarly, Adhikari (2014) investigated the number of children ever born in Dayanagar VDC of Rupendehi district, Nepal, witih illiteracy and not using family planning were a major determinant of a high number of children ever born. Furthermore, Upadhyay and Bhandari (2017) studied factors related with children ever born and concluded that the roles of age at first marriage, occupation of husband and knowledge of contraceptive/family planning were statistically significant in varying children ever born by women of Somadi VDc of Palpa district of Nepal. Alaba et al. (2017) recently studied spatial patterns and determinants of fertility levels among women of childbearing age in Nigeria and found that age at first birth, staying in rural place of residence, the number of daughters in a household, being gainfully employed, married and living with a partner, community and household effects contributed to high fertility patterns. As it can be observed from Figure 2.1, this study is based on many explanatory factors and this is one of its strength.

Figure 2.2 displays the box and whisker of total children ever born with regards to urban-rural status, region, currently pregnant, religion, ever had pregnancy terminated via abortion, miscarriage or stillbirth, currently working, woman's occupation, fertility preferences, whether and when this child's pregnancy wanted, sex of child, child is alive, child is twin or single birth and educational level. The boxplots give a better understanding of the data by its distribution, outliers, mean, median and spread. For example, as shown in Figure 2.2, in urban-rural status, women in rural area have a higher spread of total children ever born than women in urban area. Also, for religion, the box and whisker plot indicates that women who belong to Muslim/Islamic religion have a higher number of children ever born than women who belong to Christian religion. North East has the highest number of children ever born while South West has the least total children ever born. The box and whisker plot for sex of child show that there is no much difference between the total number of male and female child born by the women.

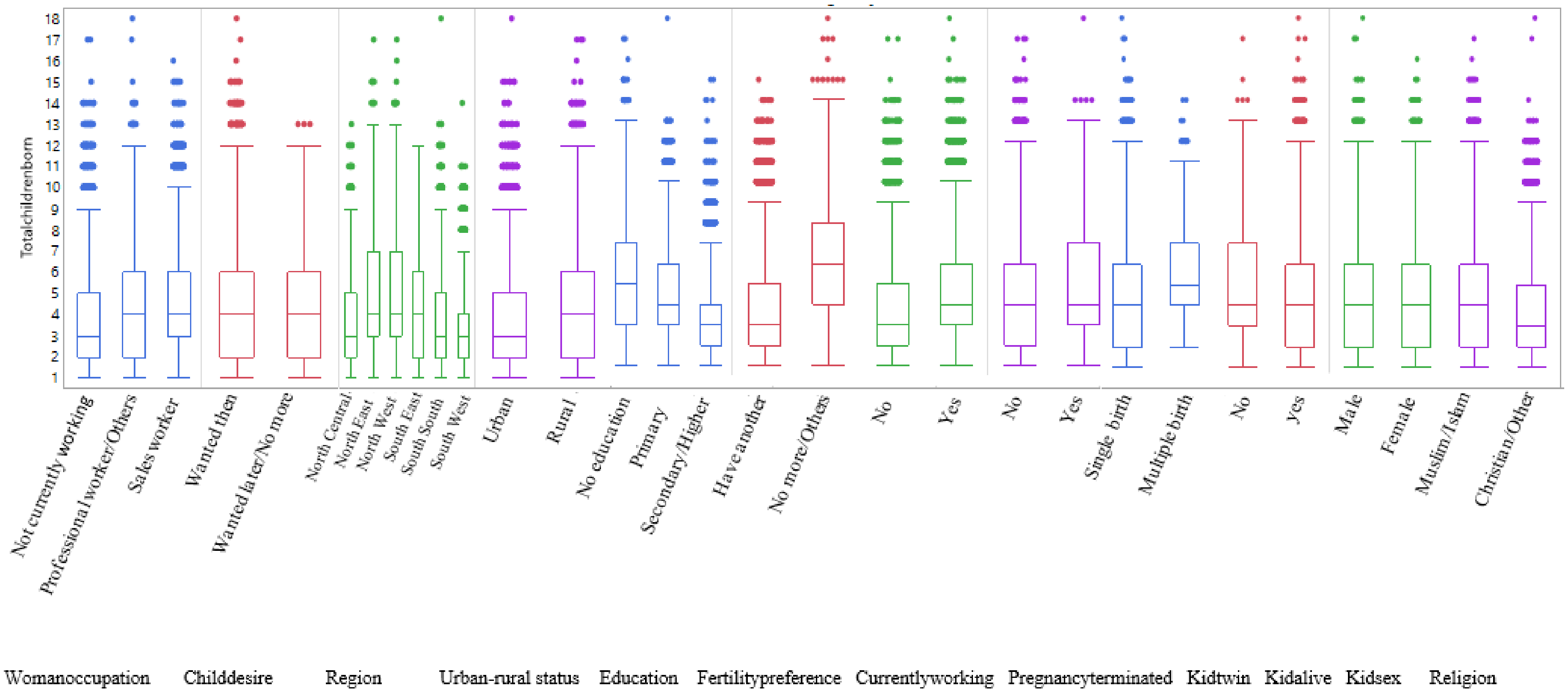


Figure 2.2: Box and whisker plot for total children ever born by categorical explanatory variables

The descriptive statistics of the age of household head, age at first marriage or cohabitation and age of respondent at time of first birth are shown in Table 2.1. The mean age of household head is approximately 41 years with standard deviation 12.03 and the mean age at first marriage or cohabitation is approximately 28 years with a standard deviation of 4.45. Meanwhile, the mean age of respondent at time of first birth is approximately 19 years with a standard deviation of 4.26, which means that the standard deviation is less spread out from the mean.

Table 2.1: Summary statistics of continuous explanatory variables

Variables	Mean	Standard Deviation	N	First Quartile	Third Quartile
Age of household head	41.4086	12.0343	31424	33	48
Age at first marriage or cohabitation	17.6189	4.4501	30878	14	20
Age of respondent at time of first birth	19.3642	4.2556	31482	16	22

Pearson correlation coefficient, Spearman correlation coefficient and Kendall's tau coefficient were used to measure the association between the dependent variable (total children ever born) and each of the continuous explanatory variables (age of household head, age at first marriage or cohabitation and age of respondent at time of first birth). The association between total children ever born and continuous explanatory variable is presented in Table 2.2. It shows that there is a significant association (p-value<.0001) between total children ever born and each of the predictors (age of household head, age at first marriage or cohabitation and age of respondent at time of first birth). The measure of association for all the coefficients for age of household is positive while age at first marriage or cohabitation and age of respondent at time of first birth are negative. This means that as the age of household head increases, the total children ever born increases but with a weak association, while the association of age at first marriage or cohabitation, and age of respondent at time of first birth, respectively, with total children ever born are in opposite direction with a weak relationship.

Table 2.2: Measure of association between total children ever born and continuous explanatory variables.

	Pearson	Spearman	Kendall	P-Value
Age of household head	0.31635	0.38368	0.29453	<.0001
Age at first marriage or cohabitation	-0.25668	-0.26524	-0.19423	<.0001
Age of respondent at time of first birth	-0.26246	-0.26187	-0.19169	<.0001

Chapter 3: Poisson regression model and application to total number of children ever born in Nigeria

3.1 Poisson distribution

Poisson distribution is known as a popular model for count data (Gschlößl and Czado, 2008). A major supposition in the Poisson model is the equality of the variance and the mean which is very limiting for over-dispersed data where the variance in the data is higher than the excepted one from the model. In other words, the observed variance is higher than the theoretical model (Gschlößl and Czado, 2008). The Poisson distribution in probability theory and statistics is a discrete probability distribution that states the probability of a given number of events occurring in a fixed interval of time or space.

An event can occur 0,1,2,… times in an interval. The average number of events in an interval is designated μ, which is the event rate, also called the rate parameter. The probability of observing y events in an interval is given by the equation

$$P(y\ events\ in\ the\ interval) = e^{-\mu}\frac{\mu^y}{y!}$$

This is the probability mass function (PMF) for a Poisson distribution.

Where,

- μ is the average number of events per interval
- e is the number 2.71828… (Euler's number) the base of the natural logarithms
- y takes values 0,1,2, …
- $y! = y \times (y-1) \times (y-2) \times ... \times 2 \times 1$ is the factorial of y.

From the probability mass function, μ(the positive real number) is equal to the expected value of Y also to its variance (Grimmett and Welsh, 2014).

$$\mu = \mathrm{E}(Y) = Var(Y).$$

The mean, variance, skewness and kurtosis of the Poisson distribution are

Mean, $\mathrm{E}(Y) = \mu$

Variance, $\sigma^2 = \mu$

Skweness, $\alpha_3 = \frac{1}{\sqrt{\mu}}$

Kurtosis, $\alpha_4 = 3 + \frac{1}{\mu}$

Characterization of location and variability of a data set is a major task in several statistical analyses. Other characterization of data includes skewness and kurtosis (Doty, 2017). Skewness can be defined as the measure of symmetric (that is lack of symmetric) while Kurtosis measures the heaviness of the tail or the lightness of the tail. In other words, heavy tails imply that the datasets have high kurtosis while light tails connote that the datasets have low kurtosis (Doty, 2017).

Using the probability mass of Poisson distribution, $\mathrm{P}(y\ events\ in\ the\ interval) = e^{-\mu}\frac{\mu^y}{y!}$

Let $\mu = 1,2,3,\dots,10$ and $y = 0,1,2,\dots,20$

taking $\mu = 1$ for all values of $y = 0,1,2,\dots,20$, the graphical representation of the skewness and kurtosis of Poisson distribution is shown in Figure 1.1

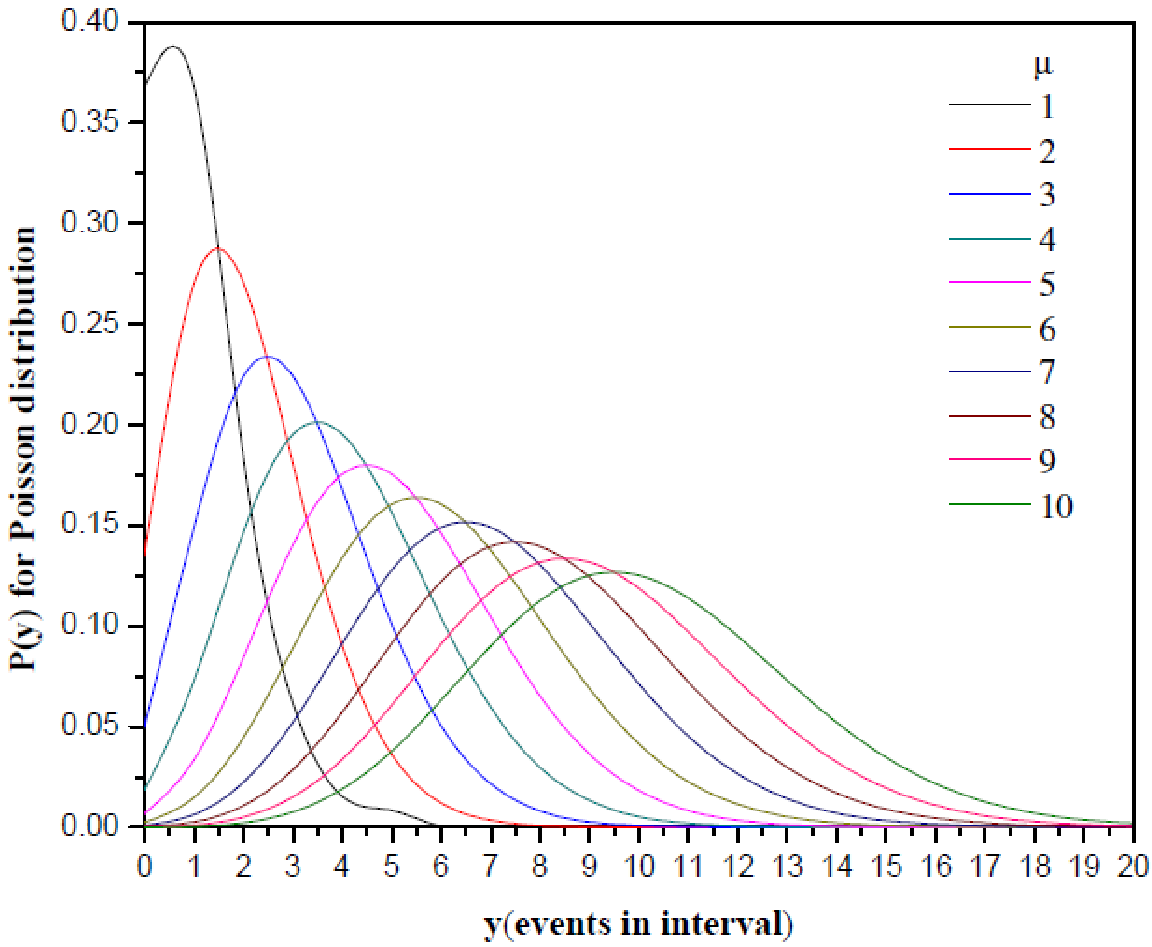


Figure 3.1: The Graphical plot of probability mass function (PMF) for Poisson distribution

Figure 3.1 shows several Poisson distributions that are portrayed concurrently. It can be observed that as the μ gets smaller, the degree of skewness increases while the kurtosis gets pointy. On the other hand, as μ gets larger, the degree of skewness decreases while the kurtosis gets less pointy, that is the values get more spread-out.

To derive Poisson regression from Poisson distribution, the rate parameter μ depends on covariates.

3.2 Poisson regression

In a situation where numbers are counted by events in time intervals, discrete count data is bound to surfaces. There is a natural choice of model based on the Poisson distribution of probability, such data being essentially non-negative integers. Poisson regression is a generalized linear model form of regression analysis, and is the reference line model for count data analysis (Long, 1997; Cameron and Trivedi, 1998; Winkelmann, 2008; Chatfield et al., 2010). It presumes that the response variable has a Poisson distribution and deduces the logarithm of its expected value, which can be modelled by a linear combination of unknown parameters. A Poisson regression model is seldom known as a log-linear model, especially when used to model contingency tables. On the contrary, its function is inadequate in real life because of its restrictive assumptions. It shows which independent variables have a significant effect on the response variable, and is mostly used for rare events, as these tend to follow a Poisson distribution (Crawley, 2012).

The Poisson regression model is

$$\mathrm{E}(Y_i|\boldsymbol{x}_i) = exp\left(\beta_0 + \sum_{k=1}^{k} \beta_k \boldsymbol{x}_{k_i}\right)$$

where β_0 is the regression coefficient for the intercept and β_k is the regression coefficient for each of the explanatory variable.

Note the restriction of the Poisson regression model is that the conditional variance is identical to the conditional mean.

In this work, a Poisson regression from the family of a generalized linear model was applied to Nigeria Demographic and Health Survey 2013 data to study the demographic, socio-economic and geographic factors affecting total children ever born. The general Poisson regression model includes log-linear, quasilinear and essentially nonlinear models. Frome (1983) noted that the IRLS algorithm corresponds to using the method of scoring to obtain maximum likelihood(ML) estimate when the events of interest follow Poisson distribution.

The alternative approach is the generalized linear model (GLM):

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} = \boldsymbol{x}_i' \beta,\ i = 1, \dots, n$$

where $\mathrm{E}(y) = \mu$.

Systematic component $X = (X_1, X_2, \dots X_k)$ is an explanatory variables and g(.) is the link function. The requirement for the GLM is that the distribution of Y should be a member of the exponential family

$$f(y; \theta, \phi) = Exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$$

where θ is the natural parameter and ϕ is the scale parameter.

Taking Y_1 ... Y_n independent and assume that Y_i has pdf

$$f(y_i|\theta_i, \phi) = exp\left[\frac{y_i\theta_i - b(\theta_i)}{\phi}\right] \times expc(y_i, \phi)$$

Assuming $E(Y_i) = \mu_i$ and there exist a known function $g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$

where $x_i' = [1\ x_{i1}\ x_{i2} \dots x_{ip}]$ and $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_p \end{pmatrix}$

$$l(\beta) = \sum_{i=1}^{n} \left(y_i\theta_i - b(\theta_i)\right)\big/\phi + \sum_{i=1}^{n} c(y_i, \phi)$$

using chain rule of differentiation,

$$\frac{\partial l(\beta)}{\partial \beta} = \frac{\partial l(\beta)}{\partial \theta_i} . \frac{\partial \theta_i}{\partial \beta}$$

$$\frac{\partial l(\beta)}{\partial \theta_i} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi}$$

then,

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)}{\phi} . \frac{\partial \theta_i}{\partial \beta}$$

But $g(\mu_i) = x_i'\beta$ *and* $\mu_i = b'(\theta_i)$ then $g\big(b'(\theta_i)\big) = x_i'\beta$

also,

$$\frac{\partial g}{\partial \mu_i}\frac{\partial \mu_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \beta} = x_i$$

therefore,

$$\frac{\partial \theta_i}{\partial \beta} = \frac{x_i}{g'(\mu_i)b''(\theta_i)}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n}\frac{\big(y_i - b'(\theta_i)\big)}{\phi}.\frac{x_i}{g'(\mu_i)b''(\theta_i)}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n}\frac{\big(y_i - b'(\theta_i)\big)x_i}{\phi g'(\mu_i)b''(\theta_i)}$$

where $\phi b''(\theta_i) = var(Y_i) = V_i$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n}\frac{(y_i - \mu_i)x_i}{g'(\mu_i)V_i}$$

$\hat{\beta}$ is found as the solution of $\frac{\partial l(\beta)}{\partial \beta} = 0$. But this set of equation needs to be found iteratively, so we need $\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}$, the matrix of second derivatives of the loglikelihood. In fact, glm works with $E\left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}\right)$. To find this we use

$$E\left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}\right) = -E\left(\frac{\partial l(\beta)}{\partial \beta}.\frac{\partial l(\beta)}{\partial \beta'}\right)$$

Taking the second derivative of $\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n}\frac{\big(y_i - b'(\theta_i)\big)x_i}{g'\big(b'(\theta_i)\big)V_i}$, we have

$$E\left(\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'}\right) = -E\left(\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{(g'(\mu_i)V_i)^2}x_i x_i'\right)$$

$$= -\sum_{i=1}^{n}\frac{V_i}{\big(g'(\mu_i)\big)^2 V_i^2}x_i x_i'$$

$$= -\sum_{i=1}^{n} w_i\, x_i x_i'$$

where $w_i \equiv 1\Big/\Big(V_i\big(g'(\mu_i)\big)^2\Big)$

W is written as a diagonal matrix

$$\begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}$$

and thus

$$E\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right) = -X'WX$$

Hence, we can say that if $\hat{\beta}$ is the solution of $\frac{\partial l(\beta)}{\partial\beta} = 0$, then $\hat{\beta}$ is asymptotically normal, with mean β and covariance matrix having as inverse

$$-E\left(\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta'}\right) = X'WX$$

Hereafter in all the glm models, the estimate of β has such a maximum likelihood property; that is

$$\hat{\beta} \sim N(\beta, X'WX)$$

Poisson distribution is shown to belong to the exponential family with probability mass function(pmf);

$$f(y) = e^{-\mu}\frac{\mu^y}{y!}$$

The logarithm of the probability distribution function of a Poisson random is

$$log f(y) = y log\mu - \mu - log(y!)$$

Which can be rewritten as;

$$f(y) = Exp\{y log\mu - \mu - log(y!)\}$$

Therefore, Poisson regression is an exponential family with:

$$\theta = log\mu$$

$$\mu = e^{\theta}$$

$$b(\theta) = e^{\theta} = \mu$$

$$a(\phi) = \phi = 1$$

$$c(y, \phi) = -log(y!)$$

In GLM, the canonical link is θ. Hence the natural link function is;

$$\theta = log\mu$$

Therefore, the GLM for Poisson distribution is

$$log\mu_i = x_i'\beta$$

The method of maximum likelihood is used to estimate the regression coefficient (β) and the logarithm of the likelihood function is given as;

$$logf_i(y_i) = y_i(x_i'\beta) - \exp(x_i'\beta) - log(y!)$$

The likelihood equation may be obtained by differentiating about each regression coefficient and setting the product equal to zero. This gives an outcome in a set of nonlinear equations that accept no closed-form solution. As a result, an iterative algorithm must be used to obtain the set of regression coefficients that maximize the loglikelihood.

Deviance which is twice the difference between the maximum achievable log-likelihood and the log-likelihood of the fitted model is a vital idea associated with a fitted GLM. It can be used in several ways, examples; to test the fit of the link function and linear predictor to the data, or to test the significance of a predictor variable in the model (Li, 1991). Under normality, the deviance is known as the residual sum of squares in multiple regression. While in Poisson regression, the deviance is

$$D = 2[\ell(y; y_i) - \ell(y\,; \mu_i)]$$

To derive the deviance for Poisson GLM, let $Y_1, \ldots, Y_n$ be samples for the model of interest, then the loglikelihood is

$$\ell(y; y_i) = \sum_{i=1}^{n} y_i log y_i - \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} log y_i!,$$

and the saturated model is

$$\ell(y; \mu_i) = \sum_{i=1}^{n} y_i log \mu_i - \sum_{i=1}^{n} \mu_i - \sum_{i=1}^{n} log y_i\,!$$

Since the deviance is twice the difference between the maximum achievable log-likelihood and the log-likelihood of the fitted model, then the deviance is

$$D = [2\sum_{i=1}^{n}\big(y_i \log(y_i) - y_i - log(y_i!)\big) - 2\sum_{i=1}^{n}\big(y_i log(\mu_i) - \mu_i - log(y_i!)\big)]$$

$$= 2\sum_{i=1}^{n}\Big(y_i\big(log(y_i) - log(\mu_i)\big) - (y_i - \mu_i)\Big)$$

Recall that $log(y_i) - log(\mu_i) = log\left({}^{y_i}/_{\mu_i}\right)$, so substituting back we have;

$$= 2\sum_{i=1}^{n}\left(y_i log\left({}^{y_i}/_{\mu_i}\right) - (y_i - \mu_i)\right)$$

A complete residual analysis should be incorporated in any regression analysis, and involves plotting the residuals against the regressor variables (as to check the outliers and curvature) and the response variable. There are various types of residuals;

- Raw residual defined as the difference between the actual response and the estimated value from the model. It is represented as

$$r_i = y_i - \hat{\mu}_i$$

- Pearson residual is used to correct unequal variance in the residuals by dividing the standard deviation. The formula is

$$p_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\emptyset}\hat{\mu}_i}}$$

- Deviance residual is well known because the deviance statistics is the sum of squares of these residuals and it is shown as

$$d_i = sign(y_i - \hat{\mu}_i)\sqrt{2\{y_i ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)}$$

- In residual diagnostic, Hat residual is used to compute the effect of each observation. The hat values, h_{ii}, are the diagonal entries of the Hat matrix which is calculated using

$$\boldsymbol{H} = \boldsymbol{W}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{1/2},$$

where $W,$ is a diagonal matrix made up of $\hat{\mu}_i$.

It is worthy of note to study the hat values in order to comprehend the observations that have large effect on the fitted regression coefficients. Those that are larger than $2k/_n$ are large hat values and they are further used to standardize residuals.

- Studentized Pearson residual is shown as

$$sp_i = \frac{p_i}{\sqrt{1 - h_{ii}}}$$

- Studentized Deviance residual is represented as

$$sd_i = \frac{d_i}{\sqrt{1 - h_{ii}}}$$

From the explanation of residual, dispersion parameters in terms of descriptive statistics describe the scattering of individual data around the mean.

The variance of y_i is $Var(y_i|\mu_i, \phi)$. This is the variance expected given a μ_i and dispersion parameter ϕ.

Generally, using Pearson chi-squared statistic, $\chi^2 = \sum_{i=1}^{n} \frac{(Y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$,

where $Var[Y_i] = V(\mu_i)\phi$. The scaled Pearson chi-squared statistic is defined as $X_s^2 = \frac{X^2}{\phi}$, if the model is specified correctly, $\chi_s^2 \sim \chi_{n-p}^2$. Asymptotically, where n is the sample size and p is the number of unknown regression coefficients (the $\beta_j's$) in the model. Then knowing the mean of a χ_{n-p}^2 random variable is $n - p$, we can use the approximation $\chi_s^2 \approx n - p$
and hence the estimator

$$\phi = \frac{X^2}{n-p},$$

While in terms of deviance, since the limiting chi-square distribution of the scaled deviance $D^* = {}^{D}/_{\phi}$ has $n - p$ degrees of freedom, equating D^* to its mean and solving for ϕ gives

$$\phi = {}^{D}/_{(n-p)}$$

Over/under dispersion can be handled formally by defining $Var(y) = \phi\mu$. From the expression, it can be said that $Var(y)$ is some multiple of the mean (μ) rather than being equal to μ and ϕ is known as dispersion parameter. If $\phi > 1$ or $\phi < 1$ then over or underdispersion exists respectively.

3.3 Application of Poisson regression model to the total number of children ever born in Nigeria

An examination of the effect of urban-rural status, region, age of respondent at time of first birth, religion, age of household head, age at first marriage or cohabitation, child is alive, woman's occupation, educational level, fertility preferences, ever had pregnancy terminated via abortion, miscarriage or stillbirth, whether and when this child's pregnancy is wanted, sex of child and child is twin or single birth on total children ever born by each respondent was performed using Poisson regression model analysis. It is necessary to examine the joint effects of some factors acting together over and above their main effects. We evaluated all possible two-interactions among the explanatory variables by including the interactions one at a time to the main effects model.

In the first round, all the possible interactions level of significance were recorded, and the highly significant ones were entered into the model with the main effect. The same procedure was used for the second-round two-way interaction term entry to the model with the main effects of the interaction selected in the previous round. The process continued until no more significant interaction effect was left to be included. Fertility preferences interacted with three variables: age at first marriage or cohabitation, region and child is twin. Education interacted with whether and when this child's pregnancy is wanted, and child is alive. While ever had pregnancy terminated via abortion, miscarriage or stillbirth interacted with woman's occupation.

It is very paramount to ascertain the validity of all the necessary assumption of a model before carrying out inference which is done after fitting the regression model. Model diagnostics is a set of procedures available to evaluate the cogency of a model in any of several different ways.

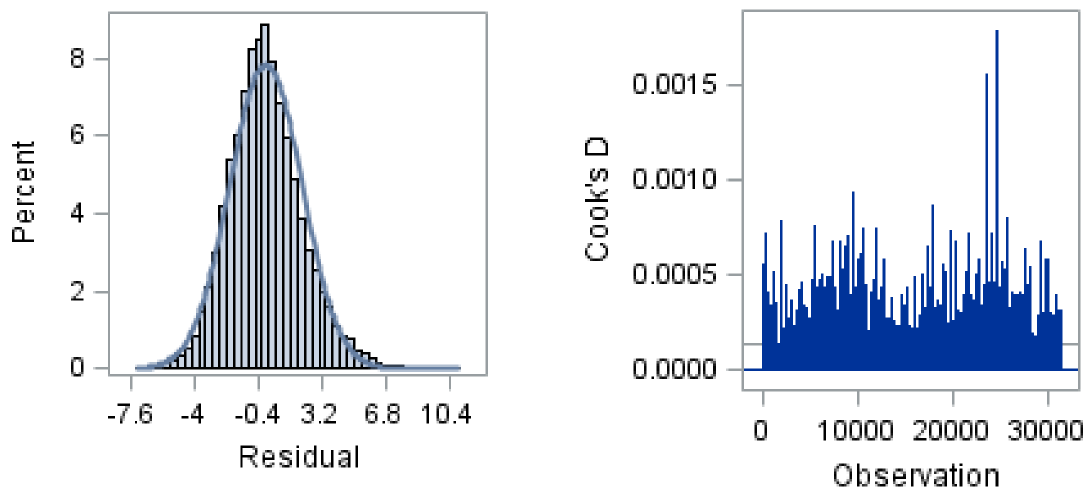


Figure 3.2: Model diagnostics plots

From the model diagnostics shown in Figure 3.2, the index plot of the residual indicates that all the observations are properly accounted for by the model. The index plot of the diagonal elements of the hat matrix suggests that there are no extreme points in the design space that one needs to consider. Also, since the D_i value is not close to 1, the index plot of Cook's distance suggests that no influential observation possibly had an adverse effect on the model parameter estimates and consequently on the goodness of fit model.

Table 3.1 contains statistics that summarize the fit of the model which are necessary for deciding on the appropriateness of the model compared to other models. It can be deduced from the deviance value in Table 3.1 that the data is not well fitted by the model since the ratio of deviance to the degree of freedom is less than 1, indicating under-dispersion. A scale option is specified (scale=dscale) to force the scaled deviance to be equal to one which makes our model to be optimally dispersed.

Table 3.1: Evaluation of over/under dispersion in Poisson regression

Criterion	Log Link DF	Value	Value/DF
Deviance	30000	27673.058	0.92
Scaled Deviance	30000	29985.000	1.00
Pearson Chi-Square	30000	27127.112	0.90
Scaled Pearson Chi-Square	30000	29393.442	0.98
Full Log Likelihood		-62229.792	
AIC		124527.584	
AICc		124527.663	
BIC		124810.110	

Table 3.2 presents the final model parameters along with the risk ratio. From Table3.2, the risk ratio (RR) can be used to interpret the significance of the main effects that were not included in the interaction while the other main effect variables included in the interaction are carefully interpreted using graphical aid. For example, as region and fertility preferences variables have significant interaction, the effect of fertility preference on total children ever born will be determined by the resident region of a woman. The effect of region will differ according to fertility preferences, and to elaborate on the effect of each pair of factors on mean number of children ever born, the interaction plots were used.

Table 3.2: Parameter estimates for the Poisson regression model

Predictors	Categories	Estimated	Standard error	Chi-square	Risk ratio	Pr > Chi-square
Urban-rural status (Reference=Urban)	Rural	-0.0161	0.0065	6.12	0.984	0.0133
Region (Reference=South West)	North Central	0.0190	0.0166	1.31	1.019	0.2528
	North East	0.1471	0.0179	67.64	1.158	<.0001
	North West	0.1595	0.0170	87.75	1.173	<.0001
	South East	0.1987	0.0182	119.47	1.220	<.0001
	South South	0.1162	0.0178	42.74	1.123	<.0001
Religion (Reference= Muslim/Islam)	Christian/Others	-0.0124	0.0089	1.95	0.988	0.1629
Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference=Yes)	No	-0.0533	0.0126	17.86	0.948	<.0001
Woman's occupation (Reference=Sales worker)	Not currently working	-0.1395	0.0201	48.18	0.870	<.0001
	Professional worker/Others	-0.0365	0.0175	4.37	0.964	0.0367
Fertility preferences (Reference=Undecided/Others)	Have another	-0.2716	0.0351	59.83	0.762	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted then)	Wanted later /No more	-0.0176	0.0169	1.09	0.983	0.2973
Sex of child (Reference=Male)	Female	0.0038	0.0053	0.53	1.004	0.4663
Child is alive (Reference=Yes)	No	0.1557	0.0224	48.38	1.168	<.0001
Child is twin or single birth (Reference= Single birth)	Multiple birth	0.0705	0.0205	11.85	1.073	0.0006
Educational level (Reference=Secondary/Higher)	No education	0.2531	0.0094	730.94	1.288	<.0001
	Primary	0.2022	0.0093	471.81	1.224	<.0001
Age of household head		0.0146	0.0002	4480.29	1.015	<.0001
Age at first marriage or cohabitation		0.0047	0.0014	10.44	1.005	0.0012
Age of respondent at time of first birth		-0.0293	0.0011	736.22	0.971	<.0001
Age at first marriage or cohabitation*Fertility preferences (Reference= Undecided/Others)	Have another	-0.0089	0.0015	33.14	0.991	<.0001
Region (Reference= South West)*Fertility preferences (Reference=Undecided/Others)	North Central*Have another	-0.0177	0.0223	0.63	0.982	0.4286
	North East*Have another	0.0732	0.0220	11.05	1.076	0.0009
	North West*Have another	0.0632	0.0207	9.32	1.065	0.0023
	South East*Have another	-0.0795	0.0244	10.63	0.924	0.0011
	South South*Have another	-0.0482	0.0248	3.76	0.953	0.0526

Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference= Yes)*Woman's occupation (Reference=Sales worker)	No* Not currently working No* Professional worker	-0.0517 0.0006	0.0212 0.0187	5.94 0.00	0.950 1.001	0.0148 0.9737
Fertility preferences (Reference=Undecided/Others)*Child is twin (Reference=Single birth)	Have another*Multiple birth	0.2199	0.0263	70.18	1.246	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted later)*Education(Reference=Secondary/Higher)	Wanted later/No more*No education Wanted later/No more*Primary	0.1127 0.0475	0.0254 0.0242	19.66 3.87	1.119 1.049	<.0001 0.0493
Child is alive (Reference=Yes)*Education (Reference=Secondary/Higher)	No*No education No*Primary	-0.1065 -0.0563	0.0249 0.0298	18.28 3.57	0.899 0.945	<.0001 0.0587

Table 3.3 shows the risk ratios and the 95% confidence intervals of the six main effects that were not included in the interaction, and the risk ratios of each factor will therefore be interpreted. The mean total children ever born at a given category will only be equivalent to the mean total children ever born from a reference category if the confidence interval includes 1, which is the condition for non-significant.

Table 3.3: The Poisson regression risk ratios extracted for main effects which were not involved in the interaction

Factor	Risk ratios	95% C Lower	Upper
Age of household head (Reference>=68)			
42-67	1.242	0.1871	0.2468
<=41	0.783	-0.2744	-0.2147
Age of respondent at time of first birth (Reference=23-33)			
<=22	1.292	0.2411	0.2711
>=34	0.553	-0.6967	-0.4873
Urban-rural status (Reference=Urban)			
Rural	1.068	0.0543	0.0777
Religion (Reference=Muslim/Islam)			
Christian/Others	0.876	-0.1441	-0.1206
Sex of Child (Reference=Male)			
Female	1.004	-0.0066	0.0146
Child is alive (Reference=Yes)			
No	1.312	0.2460	0.2974

In Table 3.3, considering age of household head, the risk ratio of children ever born by a household head from age 42-67 is 1.242 times that by others from age 68 and above. While a household head from age 41 and below had the risk ratio 0.78 of children ever born compared to others within the age of 68 and above. Which means that a household head from age 42-67 had 24.2% more children ever born than otherwise, a household head from age >=68 while a household head of age<=41 had 21.7% less children ever born than a household head of >=68. For age of respondent at time of first birth, a woman who had first birth on or before 22 years had 29.2% more children ever born than a woman who had her first birth at the age of 23 to 33 years. On the other hand, a woman who had first birth from age 34 and onwards had 44.7% less than the expected number of children of the otherwise identical characteristic mother of age 23 years to 33 years. With regards to urban-rural status, a rural woman 6.8% more children ever born compared to an urban woman. For Religion, a Christian woman had 12.4% fewer children than a Muslim/Islam woman. On average, a woman who gives birth to female children had 0.4% more children than her contemporary who gave birth to male children. Using child is alive variable, a woman whose child is not alive had 31.2% more children compared to her contemporary.

Figure 3.3 indicates that the effect of fertility preferences on the predicted mean of children ever born differs with age at first marriage or cohabitation and region. With respect to whether a woman chooses to have another or undecided/others, the predicted mean of children ever born decreases consistently across all ages of a woman as her age at first marriage increases. Regarding region, the difference between the predicted mean of children ever born by a woman who is in the fertility group of have another and a woman in the group of undecided/others is significantly different in all regions ($\rho < .0001$). In all the regions, the predicted mean of children ever born from a woman who is undecided/others is more than a woman with the choice of have another.

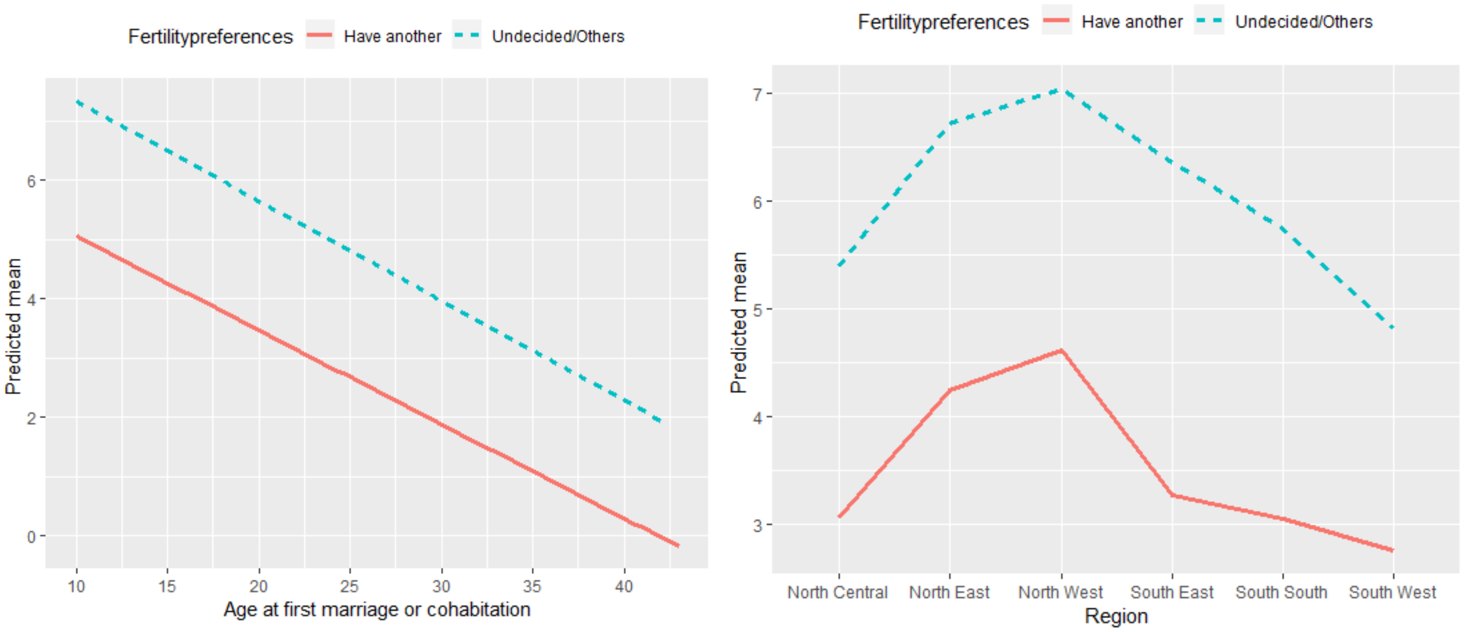
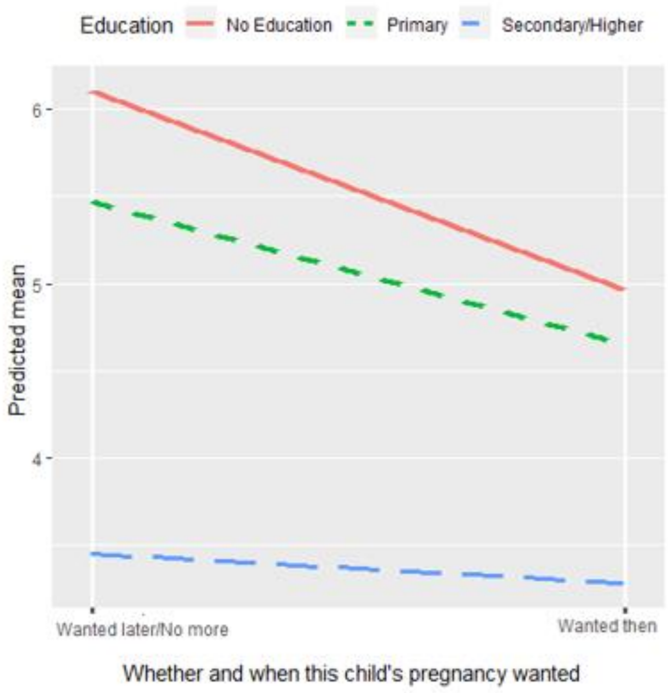


Figure 3.3: The mean number of children ever born by fertility preferences, age at first marriage or cohabitation and region

The relationship between education, whether and when this pregnancy is wanted, and child is alive is shown in Figure 3.4. The difference between the predicted mean of children ever born by a woman who has any of the educational level (whether no education, primary or secondary/others) is significant in all levels of child is alive ($\rho < .0001$). In all levels of child is alive, the predicted mean of children ever born by a woman with no education is more than her contemporaries.



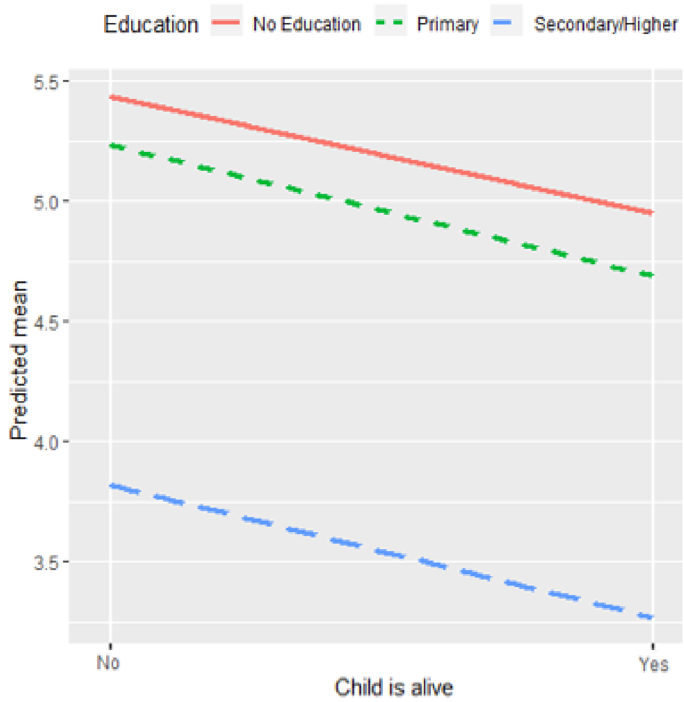


Figure 3.4: The mean number of children ever born by education, whether and when this child's pregnancy is wanted, and child is alive

Figure 3.5 presents the relationship between woman's occupation and ever had pregnancy terminated via abortion, miscarriage or stillbirth. The effect of woman's occupation on the mean of children ever born is significant for all group in ever had pregnancy terminated via abortion, miscarriage or stillbirth. The predicted mean of children ever born by a woman who has experienced abortion, miscarriage or stillbirth is more than a woman who has not.

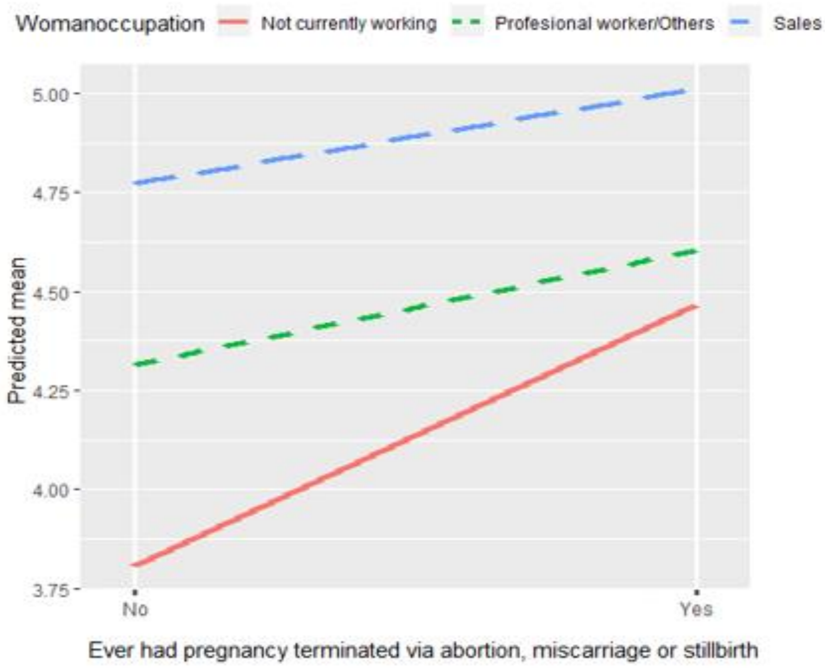


Figure 3.5: The mean number of children ever born by woman's occupation and ever had pregnancy terminated via abortion, miscarriage or stillbirth

Figure 3.6 shows the relationship between child is twin and fertility preferences, and reveals that the effect of child is twin on the predicted mean of children ever born is significant for all groups of fertility preferences. It is noted that the predicted mean of children ever born by a woman who has had multiple births is more than a woman who has had a single birth.

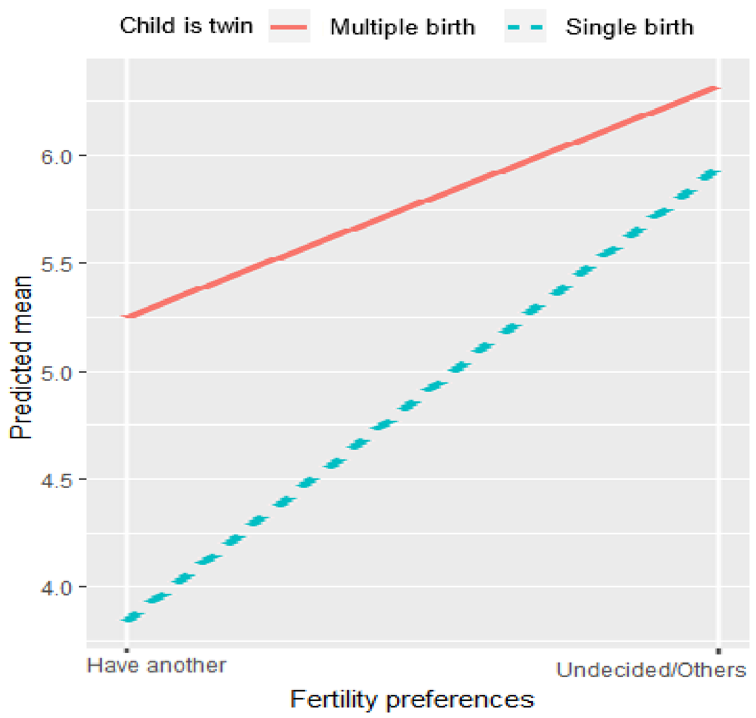


Figure 3.6: The mean number of children ever born by child is twin and fertility preferences

Conclusively, from the Poisson regression analysis, the estimated mean $(\bar{Y} = 4.33)$ and variance $\left(s_y^2 = 6.79\right)$ of the outcome show the presence of over-dispersion property of the data. This is against the key feature of Poisson model which assumptively maintains that the mean and variance of the count data should be equal. To checkmate the over-dispersion, the Negative Binomial is used.

Chapter 4: Negative Binomial regression model and application to total number of children ever born in Nigeria

4.1 Negative Binomial distribution

For a sequence of independent and identically distributed Bernoulli trials, Negative Binomial distribution is the number of successes before the kth failure (Agresti, 2002).

Probability mass function of Negative Binomial distribution is given by:

$$f(y) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \frac{(k\mu)^y}{(1+k\mu)^{y+1/k}} \text{ for y=0,1,2, …}$$

where $1/_k$ = number of failures which is real and positive number

y = number of successes and

$\mu = \frac{\frac{1}{k}(1-p)}{p}$ = mean

The mean, variance, skewness and kurtosis of the Negative Binomial distribution are

Mean, $E(y) = \mu = \frac{\frac{1}{k}(1-p)}{p}$, where p is the probability of success

Variance, $Var(y) = \mu + \mu^2 k$

Skweness, $\alpha^3 = \frac{2\mu+\frac{1}{k}}{\left(\mu+\frac{1}{k}\right)\sqrt{\left(\frac{1}{k}\right)\left(\frac{\mu}{\mu+\frac{1}{k}}\right)}}$

Kurtosis, $\alpha^4 = 3 + \frac{1}{\mu+\frac{1}{k}} + 5k + \frac{1}{\frac{1}{k}\left(\frac{2/_k+\mu}{\mu+1/_k}\right)}$

Since $\mu > 0$ and $k > 0$, the conditional variance is larger than the conditional mean.

Recall that $\mu = \frac{\frac{1}{k}(1-p)}{p}$, using the Negative Binomial expression,

$f(y) = \frac{\Gamma\left(y+\frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \frac{(k\mu)^y}{(1+k\mu)^{y+1/k}}$, for $1/_k = 1,2,3,\dots,10$, $y = 0,1,2,\dots,30$ and $\rho = 0.50$, the graphical representation of the probability mass function of Negative Binomial distribution is shown in Figure 4.1

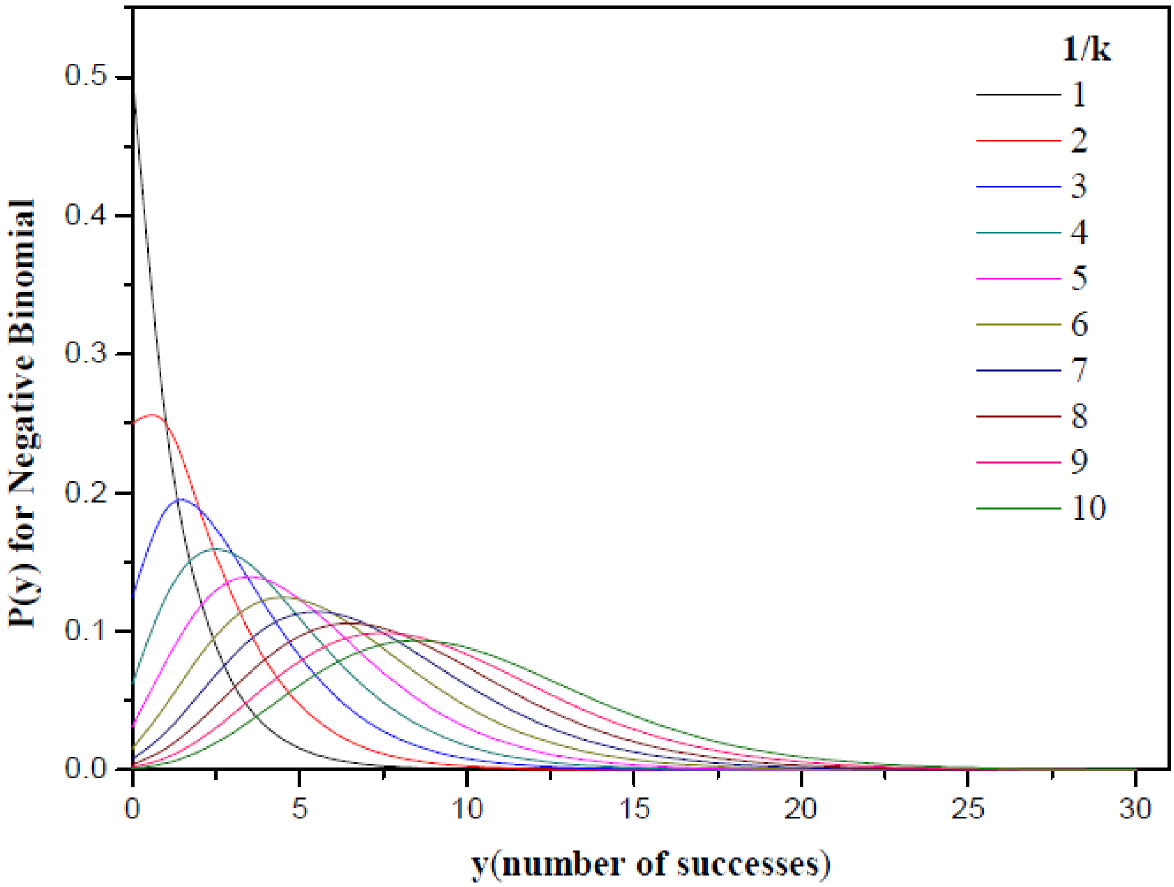


Figure 4.1: The Graphical plot of probability mass function (PMF) for Negative Binomial distribution

From Figure 1.2, as ${}^{1}/{}_{k}$ gets smaller the skewness of the distribution increases while the kurtosis becomes more peaked. But as ${}^{1}/{}_{k}$ gets larger, the skewness of the distribution gets smaller while the kurtosis gets flattened. It can be seen from Figures1.1 and 1.2 that there is sia milarity between Poisson distribution and Negative Binomial distribution, this is because as μ and ${}^{1}/{}_{k}$ increases, the skewness of their distribution become symmetric, but as μ and ${}^{1}/{}_{k}$ decreases, their respective distributions become positively skewed.

The relationship between Negative Binomial and Poisson distribution is seen when a gamma prior is used for a Poisson distribution. In other words, μ is distributed as a gamma distribution with shape= r and scale $\beta = \frac{(1-p)}{p}$,when μ is itself a random variable (Cook, 2009).

Theoretically, assuming Y has a Poisson distribution with mean λ, and λ has a Gamma distribution, $G(k, \mu)$. According to (Agresti, 2002), the Gamma probability density function can be expressed as;

$$f(\lambda; k, \mu) = \frac{\left({}^{k}\!/_{\mu}\right)^{k}}{\Gamma(k)} exp\left({}^{-k\lambda}\!/_{\mu}\right)\lambda^{k-1}, \quad \lambda \geq 0$$

The pdf of Poisson gamma mixture distribution is

$$h(y|\mu, k) = \int f(y|\lambda)g(\lambda|k, \mu)d\lambda$$

$$= \int e^{-\lambda}\frac{\lambda^{y}}{y!} \times \frac{\left({}^{k}\!/_{\mu}\right)^{k}}{\Gamma(k)} exp\left({}^{-k\lambda}\!/_{\mu}\right)\lambda^{k-1}\, d\lambda$$

$$= \frac{\left({}^{k}\!/_{\mu}\right)^{k}}{y!\,\Gamma(k)} \int \lambda^{y+k-1}\, exp\left(-(1 + \frac{k}{\mu})\lambda\right) d\lambda$$

Recalling that $y! = \Gamma(y+1)$, then after substituting the expression, we have

$$= \frac{\left({}^{k}\!/_{\mu}\right)^{k}}{\Gamma(y+1)\Gamma(k)} \int \lambda^{y+k-1}\, exp\left(-(1 + \frac{k}{\mu})\lambda\right) d\lambda$$

Directly from (Cameron and Trivedi, 2013), we have

$$h(y|\mu, k) = \frac{\left({}^{k}\!/_{\mu}\right)^{k}}{\Gamma(y+1)\Gamma(k)} \left(1 + \frac{k}{\mu}\right)^{-(y+k)} \Gamma(y+k),$$

which can be rewritten as;

$$= \frac{\Gamma(y+k)}{\Gamma(y+1)\Gamma(k)} \left(\frac{k}{\mu + k}\right)^{k} \left(1 - \frac{k}{\mu + k}\right)^{y} \text{ for } y = 0,1,2, \ldots,$$

which is equivalent to

$$f(y) = \frac{\Gamma\left(y + \frac{1}{k}\right)}{\Gamma(y+1)\Gamma\left(\frac{1}{k}\right)} \frac{(k\mu)^{y}}{(1 + k\mu)^{y + {}^{1}\!/_{k}}}$$

a pmf of Negative Binomial distribution.

4.2 Negative Binomial regression

Currently, studies on biological data that varies slightly from Poisson distribution have brought Negative Binomial distribution to the limelight. Accident statistics and insect counts in which relatively complex factors are at work are examples of data that considers the Negative

Binomial description. Negative Binomial regression and Poisson regression are similar, but the only difference that exists between them is that the dependent (Y) variable is an observed count which follows the Negative Binomial distribution. Accordingly, y possible values are nonnegative integers: 0,1,2,3, and so on (Dick et al., 2007).

Negative Binomial regression as a generalization of Poisson regression loosens the restrictive assumption of Poisson regression that the variance is equal to the mean. Negative Binomial is a Poisson-gamma mixture distribution. This formulation is prevalent because it permits the modelling of Poisson heterogeneity using a gamma distribution.

The Negative Binomial distribution is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures (which is denoted as $1/_k$) occurs. That is the probability distribution of the number of successes before the kth failure occurs is a Negative Binomial distribution.

Pascal distribution and Polya distribution are unique cases of Negative Binomial distribution. On a general note, "Negative Binomial" or "Pascal" for the case of an integer-valued stopping-time parameter r, and "Polya" for the real-valued cases are used by engineers, climatologists and others (Sakamoto et al., 1986).

It can be explained that if there is a sequence of independent Bernoulli trials then each trial has two potential outcomes called "success" and "failure". If a Bernoulli trial outcome sequence is observed until the predestined number of failures, $1/_k$, has occurred then y which is the random number of successes that is seen before the kth failure will have a Negative Binomial (or Pascal) distribution.

- Negative Binomial:

$$f(y) = \frac{\Gamma\left(y + 1/_k\right)}{\Gamma(y+1)\Gamma\left(1/_k\right)} \frac{(k\mu)^y}{(1+k\mu)^{y+1/_k}} \qquad for\ y = 0,1,2,\ldots$$

Negative Binomial has an advantage over Poisson regression due to its ability to possess one extra parameter that helps to adjust the variance independently from the mean.
To show that Negative Binomial regression is a member of the exponential family, we need to recall that Y as a stochastic variable is said to have a distribution belonging to the exponential

family if its probability density function(pdf), or probability mass function(pmf) of Y is discreet and can be written as

$$f(y;\theta,\phi) = Exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

where θ – canonical parameter

ϕ – dispersion parameter

The functions $a(\theta)$ and $c(y,\phi)$ are specified for each distribution (Bayarri and DeGroot, 1987). For known k, Negative Binomial distribution is shown to belong to the exponential family with probability mass function(pmf);

$$f(y) = \frac{\Gamma\left(y + {}^{1}/_{k}\right)}{\Gamma(y+1)\Gamma\left({}^{1}/_{k}\right)} \frac{(k\mu)^{y}}{(1+k\mu)^{y+{}^{1}/_{k}}}$$

$$logf(y) = ylog(k\mu) - \left(y + {}^{1}/_{k}\right)log(1+k\mu) + log\frac{\Gamma\left(y + {}^{1}/_{k}\right)}{\Gamma(y+1)\Gamma\left({}^{1}/_{k}\right)}$$

$$= exp\left\{ylog(k\mu) - \left(y + {}^{1}/_{k}\right)log(1+k\mu) + log\frac{\Gamma\left(y + {}^{1}/_{k}\right)}{\Gamma(y+1)\Gamma\left({}^{1}/_{k}\right)}\right\}$$

Since $\theta = logk\mu$, we made k the subject of the formula then substitutes the value in $(1 + k\mu)$to obtain $(1 + e^{\theta})$. Then bringing everything together, this can be rewritten as;

$$= exp\left\{ylog(k\mu) - \left(y + {}^{1}/_{k}\right)log\left(1+e^{\theta}\right) + log\frac{\Gamma\left(y + {}^{1}/_{k}\right)}{\Gamma(y+1)\Gamma\left({}^{1}/_{k}\right)}\right\}$$

therefore, Negative Binomial is an exponential family expression with

$$\theta = logk\mu$$

$$b(\theta) = -{}^{1}/_{k}\,\log\left(1 - e^{\theta}\right)$$

$$a(\phi) = 1, \qquad \phi = 1$$

$$c(y,\phi) = log\left(\frac{\Gamma(y + {}^{1}/_{k})}{\Gamma({}^{1}/_{k})\Gamma(y+1)}\right)$$

The twice difference between the maximum achievable log-likelihood and the log-likelihood of the fitted model is the deviance.

Under normality, the deviance is the residual sum of squares in Multiple regression while in Negative Binomial regression, deviance is the generalization of the sum of squares. By replacing μ_i with y_i, the maximum possible log likelihood is computed. Therefore, we have

$$D = 2[\ell(y; y_i) - \ell(y; \mu_i)]$$

To derive the deviance for Negative Binomial GLM, let the loglikelihood of model of interest be

$$\ell(y; y_i) = \sum_{i=1}^{n} y_i log(ky_i) - \sum_{i=1}^{n} \left(y_i + {}^{1}\!/_{k}\right) log(1 + ky_i) + \sum_{i=1}^{n} \frac{\Gamma\left(y_i + {}^{1}\!/_{k}\right)}{\Gamma\left({}^{1}\!/_{k}\right)}$$

and the saturated model as

$$\ell(y; \mu_i) = \sum_{i=1}^{n} y_i log(k\mu_i) - \sum_{i=1}^{n} \left(y_i + {}^{1}\!/_{k}\right) log(1 + k\mu_i) + \sum_{i=1}^{n} \frac{\Gamma\left(y_i + {}^{1}\!/_{k}\right)}{\Gamma\left({}^{1}\!/_{k}\right)}$$

$$D = 2\sum_{i=1}^{n} y_i log(ky_i) - \sum_{i=1}^{n} \left(y_i + {}^{1}\!/_{k}\right) log(1 + ky_i) + \sum_{i=1}^{n} \frac{\Gamma\left(y_i + {}^{1}\!/_{k}\right)}{\Gamma\left({}^{1}\!/_{k}\right)}$$
$$- \left(2\sum_{i=1}^{n} y_i log(k\mu_i) - \sum_{i=1}^{n} \left(y_i + {}^{1}\!/_{k}\right) log(1 + k\mu_i) + \sum_{i=1}^{n} \frac{\Gamma\left(y_i + {}^{1}\!/_{k}\right)}{\Gamma\left({}^{1}\!/_{k}\right)}\right)$$
$$= 2\sum_{i=1}^{n} \left(y_i\left(log(ky_i) - log(k\mu_i)\right) - \left(y_i + {}^{1}\!/_{k}\right)\left(log(1 + ky_i) - log(1 + k\mu_i)\right)\right)$$

$$= 2\sum_{i=1}^{n} \left[y_i log\frac{y_i}{\mu_i} - \left(y_i + {}^{1}\!/_{k}\right) log\frac{1 + ky_i}{1 + k\mu_i}\right]$$

It is worthy of note that a complete residual analysis should be incorporated in any regression analysis. It includes plotting the residuals against various other quantities such as the regressor variables which helps to examine outliers and curvature and the response variable. The following are types of residuals;

- Raw residual. In this case, because there is an expectation that the variances of the residuals will be unequal, interpretation of the raw residuals becomes difficult. The raw residual is expressed as;

$$r_i = y_i - \hat{\mu}_i$$

- Pearson residual obtained by dividing the standard deviation of y. It can be expressed as;

$$p_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i + r\hat{\mu}_i^2}}$$

4.3 Application of Negative Binomial regression model to the total number of children ever born in Nigeria

GLM using a Negative binomial distribution with its natural link function is fitted. Before some inferences were made and compared with Poisson regression, the model diagnostics were performed. Figure 4.2 displays the histogram of the predicted value from the model diagnostics performed. Like the diagnostic plot of Poisson regression, it is noted that the index plot for the residuals of this model accounts for all the observations and the index plot of the diagonal elements of hat matrix and proposes no exciting points in the design space that one needs to consider. In addition, no observation in the index plot of the cook's distance could affect the estimated coefficients and the goodness of fit.

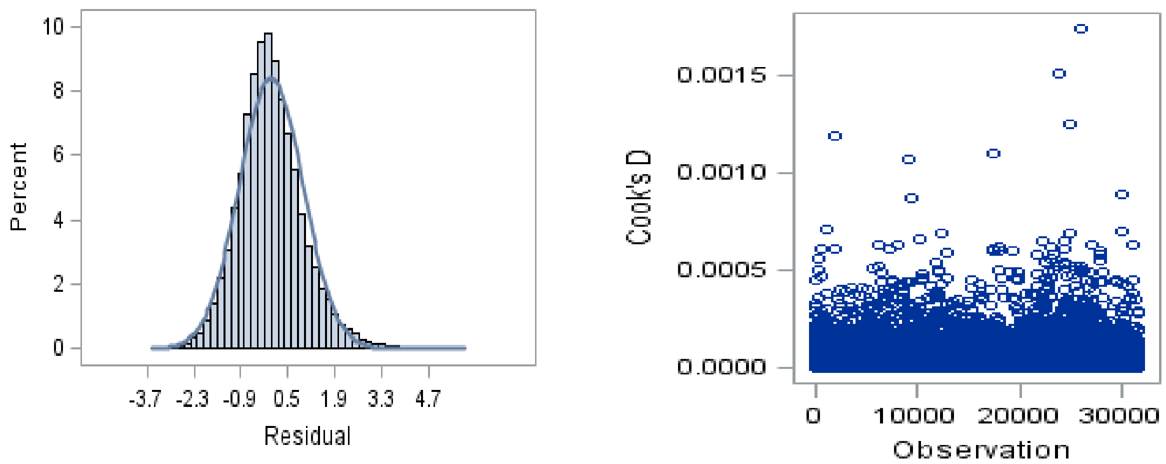


Figure 4.2: Model diagnostics plots

Table 4.1 contains statistics that summarize the fit of the model which are necessary for deciding the appropriateness of the model compared to other models. It can be deduced from the deviance value that the data is not well fitted by the model, as the ratio of deviance to the degree of freedom is less than 1, indicating under-dispersion. A scale option is specified (scale=dscale) to force the scaled deviance to be equal to one which makes our model to be optimally dispersed.

Table 4.1: Evaluation of over/under dispersion in Negative Binomial regression

Criterion	**Log Link DF**	**Value**	**Value/DF**
Deviance	30000	27319.201	0.91
Pearson Chi-Square	30000	26877.900	0.90
Full Log-Likelihood		-61479.395	
AIC		123028.789	
AICc		123028.873	
BIC		123319.625	

Table 4.2: Parameter estimates for the Negative Binomial regression model

Predictors	Categories	Estimated	Standard error	Chi-square	Risk ratio	Pr > Chi-square
Urban-rural status (Reference=Urban)	Rural	0.0032	0.0068	0.22	1.003	0.6370
Region (Reference=South West)	North Central	0.0141	0.0173	0.66	1.014	0.4165
	North East	0.1701	0.0181	88.40	1.185	<.0001
	North West	0.1336	0.0186	51.58	1.143	<.0001
	South East	0.1912	0.0189	102.23	1.211	<.0001
	South South	0.1302	0.0174	56.16	1.139	<.0001
Religion (Reference= Muslim/Islam)	Christian/Others	-0.0179	0.0092	3.81	0.982	0.0511
Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference=Yes)	No	-0.0454	0.0134	11.56	0.956	0.0007
Woman's occupation (Reference=Sales worker)	Not currently working	-0.1460	0.0212	47.27	0.864	<.0001
	Professional worker/Others	-0.0243	0.0182	1.77	0.976	0.1830
Fertility preferences (Reference=Undecided/Others)	Have another	-0.2829	0.0368	59.17	0.754	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted then)	Wanted later /No more	-0.0215	0.0170	1.60	0.979	0.2058
Sex of child (Reference=Male)	Female	0.0103	0.0055	3.47	1.010	0.0624
Child is alive (Reference=Yes)	No	0.1536	0.0226	46.30	1.166	<.0001
Child is twin or single birth (Reference= Single birth)	Multiple birth	0.0775	0.0208	13.88	1.081	0.0002
Educational level (Reference=Secondary/Higher)	No education	0.2520	0.0096	687.39	1.287	<.0001
	Primary	0.1934	0.0094	419.70	1.213	<.0001
Age of household head		0.0135	0.0002	3565.13	1.014	<.0001
Age at first marriage or cohabitation		0.0035	0.0015	5.40	1.004	0.0202
Age of respondent at time of first birth		-0.0287	0.0012	619.99	0.972	<.0001
Age at first marriage or cohabitation*Fertility preferences (Reference= Undecided/Others)	Have another	-0.0091	0.0016	32.66	0.991	<.0001
Region (Reference= South West)*Fertility preferences (Reference=Undecided/Others)	North Central*Have another	-0.0160	0.0236	0.46	0.984	0.4977
	North East*Have another	0.0419	0.0228	3.37	1.043	0.0663
	North West*Have another	0.0816	0.0227	12.92	1.085	0.0003
	South East*Have another	-0.0626	0.0259	5.82	0.939	0.0159
	South South*Have another	-0.0327	0.0245	1.79	0.968	0.1815

Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference= Yes)*Woman's occupation (Reference=Sales worker)	No* Not currently working No* Professional worker	-0.0390 -0.0191	0.0224 0.0194	3.03 0.97	0.962 0.981	0.0818 0.3253
Fertility preferences (Reference=Undecided/Others)*Child is twin (Reference=Single birth)	Have another*Multiple birth	0.2190	0.0270	65.84	1.245	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted later)*Education(Reference=Secondary/Higher)	Wanted later/No more*No education Wanted later/No more*Primary	0.1065 0.0571	0.0255 0.0244	17.40 5.50	1.112 1.059	<.0001 0.0190
Child is alive (Reference=Yes)*Education (Reference=Secondary/Higher)	No*No education No*Primary	-0.1040 -0.0524	0.0253 0.0300	16.86 3.05	0.901 0.949	<.0001 0.0807

Table 4.2 shows the Negative Binomial regression relative ratio of all the main effects of their interactions. It is also observed that rural, North Central, Christian, wanted later/no more, female, North Central and have another, South South and have another, no and professional worker, no and primary including no and not currently working variables are insignificant. To account for the main effects which were not involved in the interaction, the result in Table 4.3 is displayed.

Table 4.3: The Negative Binomial regression risk ratios extracted for main effects which were not involved in the interaction

Factor	**Risk ratios**	**95% C**	
		Lower	**Upper**
Age of household head (Reference>=68)			
42-67	1.257	0.1972	0.2603
<=41	0.802	-0.2518	-0.1889
Age of respondent at time of first birth (Reference>=34)			
<=22	2.384	0.7615	0.9760
23-33	1.845	0.5048	0.7204
Urban-rural status (Reference=Urban)			
Rural	1.076	0.0599	0.0859
Religion (Reference=Muslim/Islam)			
Christian/Others	0.886	-0.1335	-0.1085
Sex of Child (Reference=Male)			
Female	1.009	-0.0029	0.0203
Child is alive (Reference=Yes)			
No	1.130	0.1033	0.1415

Table 4.3 indicates that in terms of age of household head, a household head from age 42-67 has a risk ratio of 1.257 of children ever born compared to others from age 68 and above. While a household head from the age of 41 and below has a risk ratio of 0.802 of children ever born compared to those within the age of 68 and above. For age of respondent at time of first birth, a woman who had her first birth on or before 22 years has a risk ratio of 2.384 of children ever born than a woman who had her first birth at the age of 23 - 33 years. On the other hand, a woman who had her first birth from age 23-33years has a risk ratio of 1.845 compared to a woman whose first birth is at age 34 and above. Furthermore, in urban-rural status, the risk ratio of children ever born by a rural woman is 1.076 compared to an urban woman. With respect to religion, a Christian woman has a risk ratio of 0.886 of children ever born compared to a Muslim/Islamic woman. In sex of child, the risk ratio of female gender is 1.009 times the male gender, while the risk ratio of no is 1.130 times yes.

Figure 4.3 indicates that the effect of fertility preferences on the predicted mean of children ever born differs with age at first marriage or cohabitation and region. It shows that whether a woman chooses to have another or undecided/others, the predicted mean of children ever born decreases consistently across all ages of a woman as her age at first marriage increases. Regarding region, the difference between the predicted mean of children ever born by a woman who is in the fertility group of have another and a woman in the group of undecided/others is significant in all regions ($\rho < .0001$). In all the regions, the predicted mean of children ever born by a woman that chooses undecided/others is more than a woman that chooses have another.

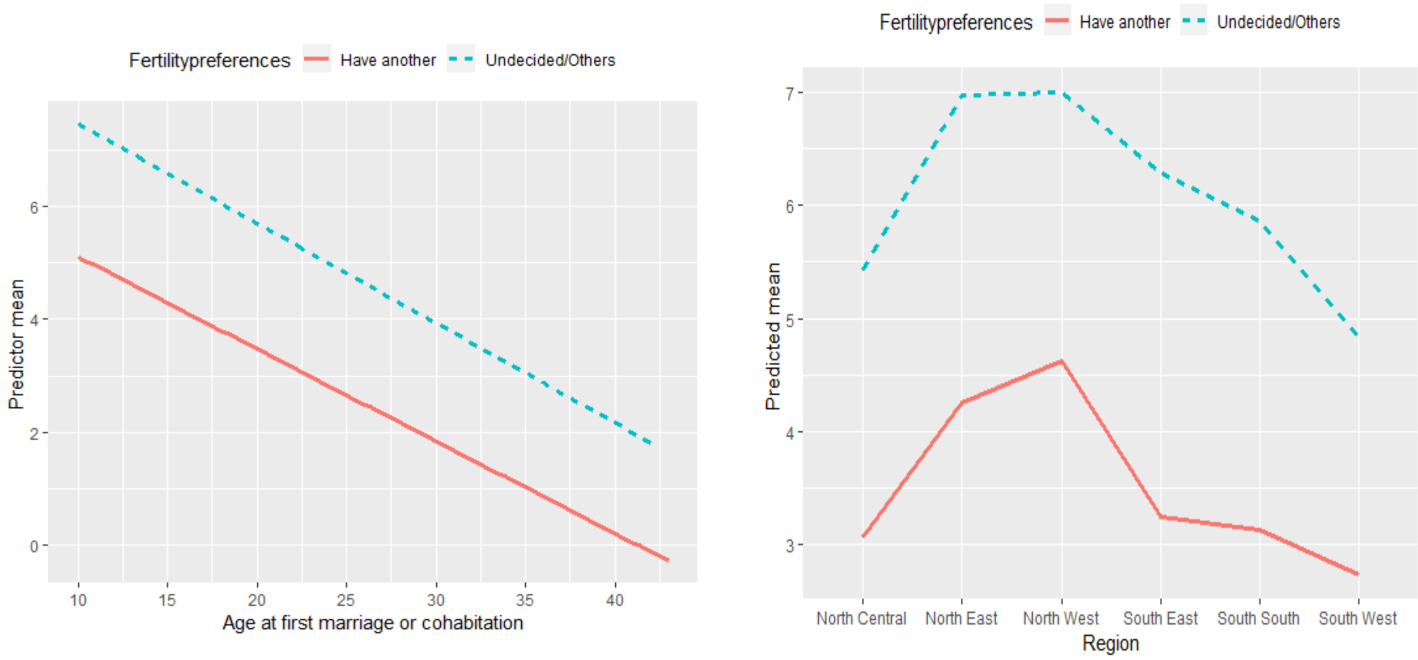


Figure 4.3: The mean number of children ever born by fertility preferences, age at first marriage or cohabitation and region

The relationship between education, whether and when this pregnancy is wanted, and child is alive is shown in Figure 4.4. Regarding whether and when this child's pregnancy is wanted, it is seen that unlike Poisson regression, the difference in the effect of education on the predicted mean of children ever born is significant for all levels with a woman who has no education giving more birth than her mates. Notably, the difference between the predicted mean of children ever born by a woman who falls under any of the educational level (whether no education, primary or secondary/others) is significant in all levels of child is alive ($\rho < .0001$). In all levels of child is alive, the predicted mean of children ever born by a woman with no education is more than her contemporaries. Generally, it can be deduced that the more educated a woman becomes the less the number of children she will give birth to.

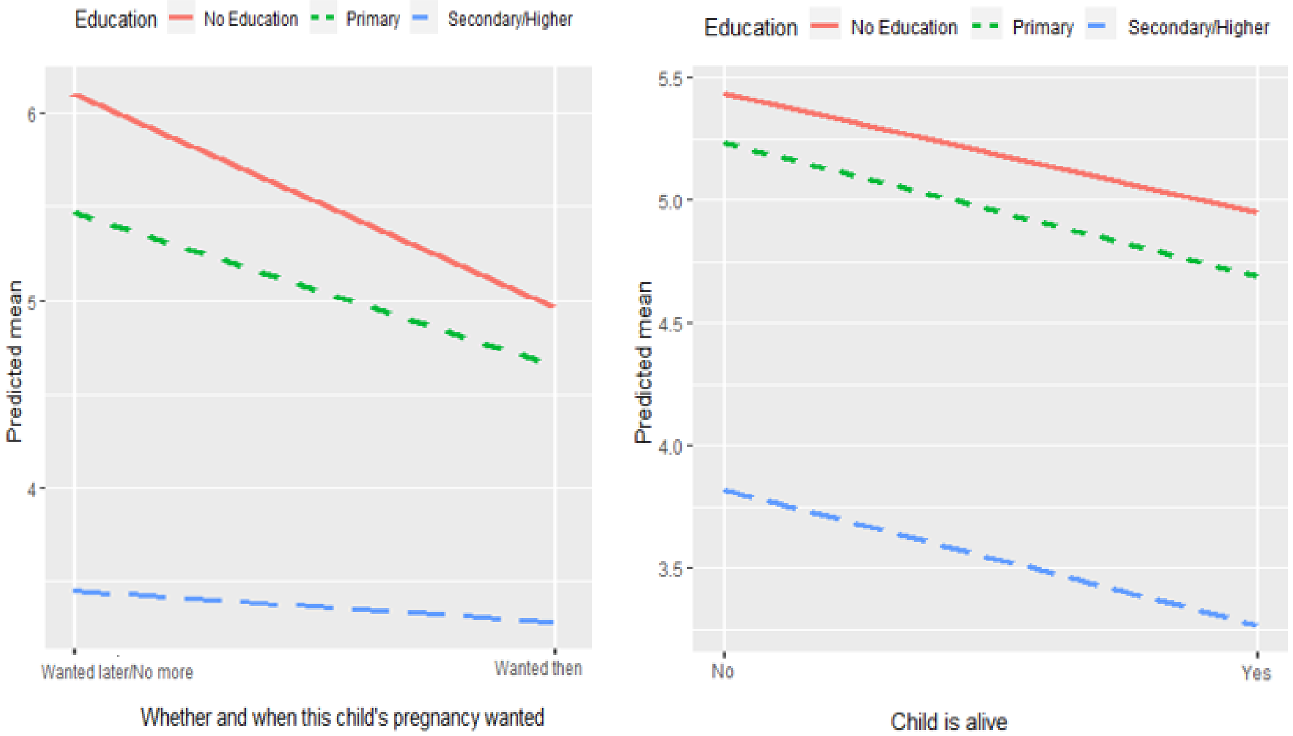


Figure 4.4: The mean number of children ever born by education, whether and when this child's pregnancy is wanted, and child is alive

Figure 4.5 shows the relationship between woman's occupation and ever had pregnancy terminated via abortion miscarriage or stillbirth. The effect of woman's occupation on the mean of total children ever born is significant for all group in ever had pregnancy terminated via abortion, miscarriage or stillbirth. In fact, the predicted mean of children ever born from a woman who has had abortion, miscarriage or stillbirth is more than a woman who has not.

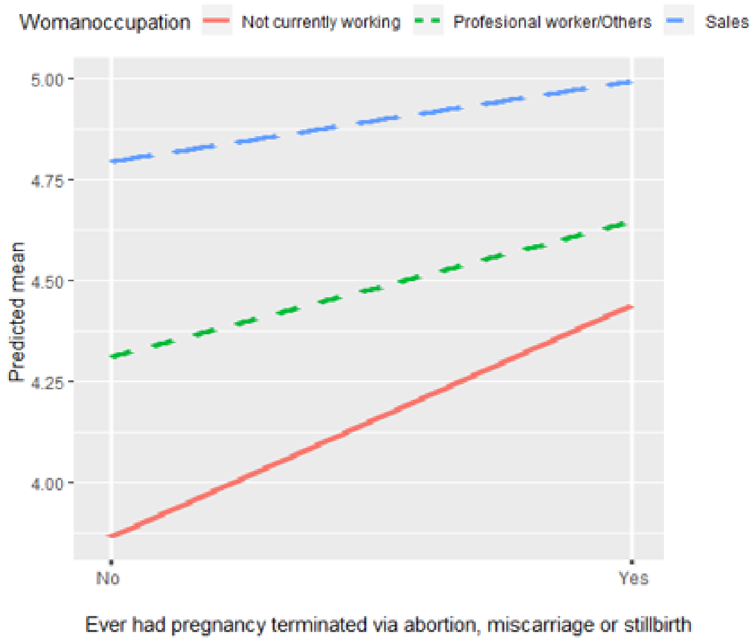


Figure 4.5: The mean number of children ever born by woman's occupation and ever had pregnancy terminated via abortion, miscarriage or stillbirth

The relationships between kidtwin and fertility preferences are shown in Figure 4.6. The relationships between kidtwin and fertility preferences in Figure 4.6 reveals that the effect of kidtwin on the predicted mean of children ever born is significant for all group of fertility preferences. It is noted that the predicted mean of children ever born by a woman who has multiple birth is more than a woman who had single birth.

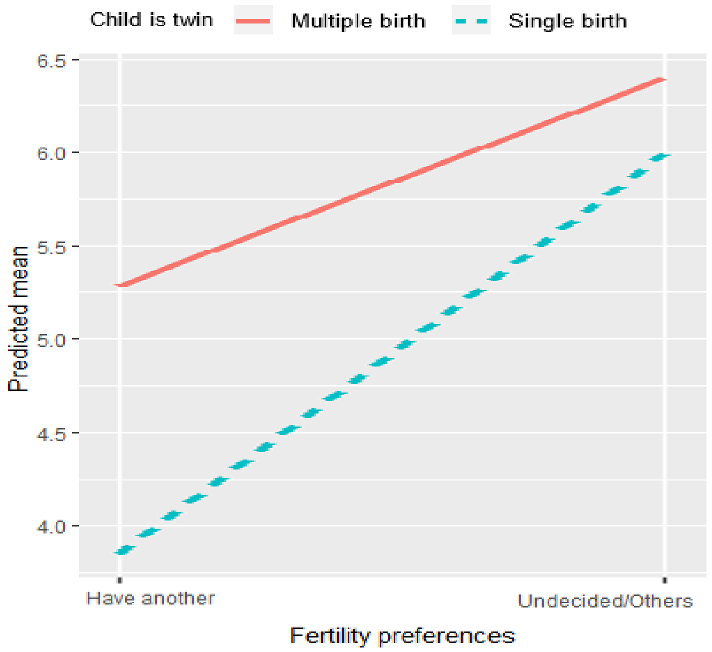


Figure 4.6: The mean number of children ever born by kidtwin and fertility preferences

In conclusion, notwithstanding the improvement on the result, the deviance of Negative Binomial is not up to 1. To overcome this shortcoming, Generalized Poisson regression is used as a strategy.

Chapter 5: Generalized Poisson regression model and application to total number of children ever born in Nigeria

5.1 Generalized Poisson distribution

Lagrange Poisson distribution is another name for Generalized Poisson distribution (Singh and Famoye, 1993). Simple Poisson distribution is known as a distinct example of Generalized Poisson distribution. In Generalized Poisson regression, Generalized Poisson distribution is used as an extension of Poisson regression that accounts for over-dispersion (Ntzoufras et al., 2005).

The probability mass function of the Generalized Poisson distribution is

$$p(y;\omega,\theta)=\frac{\theta(\theta+\omega y)^{y-1}}{y!}e^{-\theta-\omega y} \qquad y=0,1,2,\dots;0\le\omega<1;\theta>0$$

Where ω is a number or parameter in the range (0, 1) that specifies the shape the first shape parameter and θ is the positive number or parameter that specifies the second shape parameter.

The mean, variance, skewness and kurtosis of Generalized Poisson distribution is given by;

Mean, $\mu=\frac{\theta}{1-\omega}$

Variance, $\sigma^2=\frac{\theta}{(1-\omega)^3}$

Skewness, $\alpha^3=\frac{(1+2\omega)^2}{\theta(1-\omega)}$

Kurtosis, $\alpha^4=3+\frac{\left(1+8\omega+6\omega^2\right)}{\theta(1-\omega)}$

It has been shown by (Consul and Jain, 1973) that the probability mass function of Generalized Poisson distribution is a probability distribution since it has the property $\sum_{n=0}^{\infty}P_n=(\theta,\omega)=1$. This is achieved by using the identity $\sum_{n=0}^{\infty}\frac{(\theta+\omega n)^n}{n!}e^{-\theta-\omega n}=\frac{1}{1-\omega},\ for\ -\omega_0<\omega<1$ found in (Jensen, 1902). Using the probability mass function expression, let $\omega=0.0,0.002,\dots,0.01,\theta=1.0,1.2,\dots,2$. The graphical representation of the skewness and kurtosis of Generalized Poisson distribution is shown in Figure 5.1

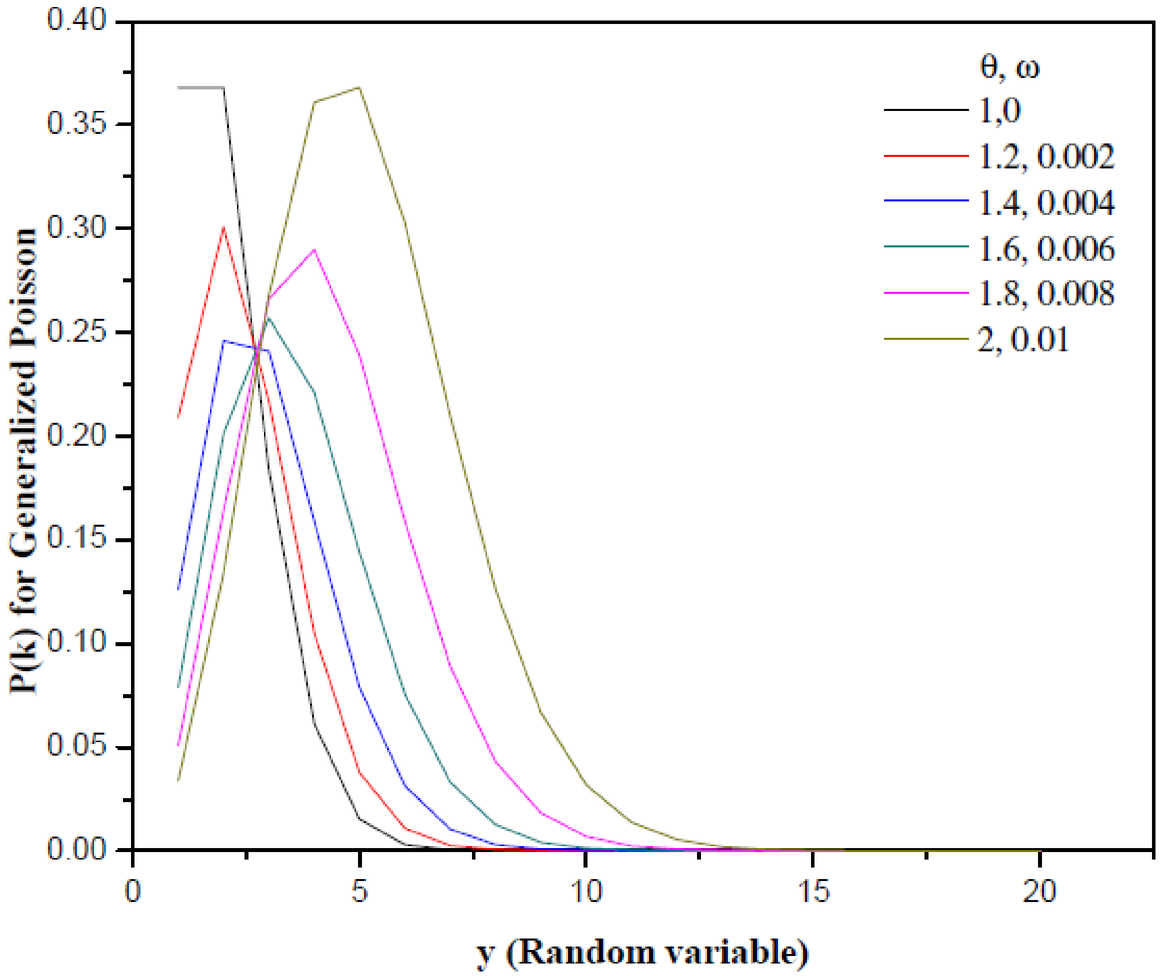


Figure 5.1: The Graphical plot of probability mass function (PMF) for Generalized Poisson distribution

In Figure 5.1, it is observed that as θ and ω decreases, the degree of skewness gets smaller while the kurtosis gets flattened. On the other hand, as θ and ω increases, the degree of skewness gets larger while the kurtosis becomes more peaked.

From the probability mass function expression, Generalized Poisson distribution tends to Poisson distribution when $\omega = 0$ and tends to Negative Binomial distribution when $\omega \to \infty$.

5.2 Generalized Poisson regression

For fitting both over-dispersed and under-dispersed count data, Generalized Poisson regression is a very useful model, this is because it allows for more variability. To study the outcome variable, the Generalized Poisson regression is preferred over Poisson regression and Negative Binomial because of its under-dispersion property (variance<mean). With our interest in Nigerian total children ever born, Poisson regression model, Negative Binomial regression

model and Generalized Poisson regression model as types of regression based on Poisson distribution have been applied to model these type of count variables.

Notwithstanding, the application of these models is centred on pure assumptions. For example, the Poisson regression model presumes equal mean and variance of the dependent variable which is not obtainable in real life because variance can be greater than mean (over-dispersion property) or lesser than mean (under-dispersion property). Should the estimates of the standard Poisson model remain consistent, biased standard errors and ineffective estimates of regression parameters are inevitable if there is a lack of knowledge of these properties. Negative Binomial regression model, which is often used to analyze an independent variable with over-dispersion, is more flexible than the standard Poisson regression model. However, Generalized Poisson regression can be used to analyze an independent variable with both over and under-dispersion, which is the reason it is said to be more flexible.

The Generalized Poisson distribution is given by

$p(y;\omega,\theta)=\frac{\theta(\theta+\omega y)^{y-1}}{y!}e^{-\theta-\omega y}$

For $y=0,1,2,\ldots$ and $\theta>0$ and $\max(-1,-\theta/4)\leq\omega\leq1$. While the mean and variance of y are $\mu=\theta(1-\omega)^{-1}$ and $\sigma^2=\theta(1-\omega)^{-3}=\mu(1-\omega)^{-2}$.

Since the dispersion parameter ω from equation (6) influences the mean as well as the variance, the following parameterization by (Consul and Jain, 1973, Zamani and Ismail, 2012) is used

$$\omega=\frac{\varphi\mu^{\rho-1}}{1+\varphi\mu^{\rho-1}},$$

which gives the following density

$$f(y|\mu,\varphi,\rho)=\frac{\mu\left(\mu+\varphi\mu^{\rho-1}y\right)^{y-1}}{(1+\varphi\mu^{\rho-1})^{y}y!}exp\left(-\frac{\mu+\mu\varphi^{p-1}y}{1+\mu\varphi^{\rho-1}}\right)$$

For $y=0,1,2,\ldots$ we assume $\varphi\geq0$

The link function of mean and variance of y are given as

$$\mathrm{E}(y|.)=\mu \text{ and } Var(y|.)=\mu(1+\varphi\mu^{\rho-1})^2$$

and the mean is linked to a linear predictor by

$$\mu=\exp(\eta)$$

5.3 Application of Generalized Poisson regression model to the total number of children ever born in Nigeria

Table 5.2 presents the result of the independent variables on total children ever born by the women which shows that some of the main effect and some of the Two 2-way interactions are insignificant. The incidence rate ratio was used to interpret the main effects that were not included in the interaction. Alternatively, the main effect variables that participated in the interaction and are significant should be carefully explained. For example, the interaction between fertility preferences and child is twin shows that since relating have another as a category in fertility preferences and multiple birth as a category in child is twin are significant, the total children ever born will be influenced if a woman's fertility preferences choice is have another or undecided/others. This will, however, vary according to child is twin or single birth with Table 5.1 displaying the fit statistics of the model

Table 5.1: Fit Statistics

-2Log Likelihood	148584.000
AIC	148654.000
AICc	148654.000
BIC	148945.000

5.31 Model diagnostics

From the graphical representation of the data (Figure 5.2), it is observed from the residual that the model accurately accounts for all observations but for the cook's distance, for which there is no methodologythat has yet beeb developed.

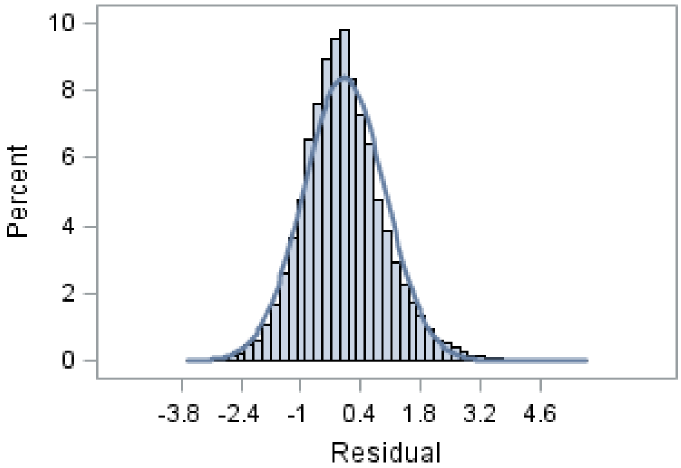


Figure 5.2: Model diagnostics plots

Table 5.2 displays the parameter for the Generalized Poisson regression, with North West, no, professional worker/others, female, multiple birth, North Central and have another, South East and have another, no and not currently working, no and professional worker/others, wanted later/no more and primary are insignificant.

Table 5.2: Parameter estimates for the Generalized Poisson regression model

Predictors	Categories	Estimated	Standard Error	z Value	Estimated IRR	Pr > |z|
Urban-rural status (Reference=Urban)	Rural	0.0255	0.0056	4.53	1.023	<.0001
Region (Reference=South West)	North Central	-0.0925	0.0186	-4.97	0.912	<.0001
	North East	0.0750	0.0187	4.02	1.078	<.0001
	North West	0.0169	0.0234	0.71	1.017	0.4807
	South East	0.1345	0.0228	5.88	1.144	<.0001
	South South	0.2267	0.0166	13.68	1.254	<.0001
Religion (Reference= Muslim/Islam)	Christian/Others	-0.0888	0.0079	-11.19	0.915	<.0001
Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference=Yes)	No	-0.0009	0.0143	-0.07	0.999	0.9475
Woman's occupation (Reference=Sales worker)	Not currently working	-0.0823	0.0206	-4.00	0.921	<.0001
	Professional worker/Others	-0.0059	0.0182	-0.32	0.994	0.7485
Fertility preferences (Reference=Undecided/Others)	Have another	-0.6407	0.0333	-19.27	0.579	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted then)	Wanted later /No more	-0.1152	0.0169	-6.82	0.891	<.0001
Sex of child (Reference=Male)	Female	0.0038	0.0025	1.54	1.004	0.1247
Child is alive (Reference=Yes)	No	0.0989	0.0224	4.41	1.104	<.0001
Child is twin or single birth (Reference= Single birth)	Multiple birth	-0.0319	0.0236	-1.35	0.969	0.1771
Educational level (Reference=Secondary/Higher)	No education	0.3255	0.0084	38.70	1.385	<.0001
	Primary	0.3443	0.0070	49.15	1.411	<.0001
Age of household head		0.0096	0.0001	68.89	1.010	<.0001
Age at first marriage or cohabitation		-0.0092	0.0013	-6.59	0.991	<.0001
Age of respondent at time of first birth		-0.0195	0.0011	-17.02	0.981	<.0001
Age at first marriage or cohabitation*Fertility preferences (Reference= Undecided/Others)	Have another	0.0115	0.0013	9.01	1.012	<.0001
Region (Reference= South West)*Fertility preferences (Reference=Undecided/Others)	North Central*Have another	-0.0345	0.0254	-1.36	1.005	0.1735
	North East*Have another	0.1294	0.0233	5.55	0.939	<.0001
	North West*Have another	0.1692	0.0280	6.04	0.889	<.0001
	South East*Have another	-0.0499	0.0306	-1.63	0.926	0.1029
	South South*Have another	0.1889	0.0221	8.56	1.148	<.0001

Ever had pregnancy terminated via abortion, miscarriage or stillbirth (Reference= Yes)*Woman's occupation (Reference=Sales worker)	No* Not currently working No* Professional worker	-0.0060 0.0118	0.0216 0.0194	-0.28 0.61	0.994 1.012	0.7795 0.5427
Fertility preferences (Reference=Undecided/Others)*Child is twin (Reference=Single birth)	Have another*Multiple birth	0.1967	0.0305	6.45	1.217	<.0001
Whether and when this child's pregnancy is wanted (Reference=Wanted later)*Education(Reference=Secondary/Higher)	Wanted later/No more*No Education Wanted later/No more*Primary	0.1807 -0.0218	0.0232 0.0201	7.78 -1.08	1.198 0.978	<.0001 0.2795
Child is alive (Reference=Yes)*Education (Reference=Secondary/Higher)	No*No Education No*Primary	-0.0833 0.1051	0.0246 0.0247	-3.38 4.26	0.909 1.037	0.0007 <.0001

To check for the effect of the main effects which were not involved in the interaction, further examination was done, as displayed in Table 5.3. The results of Generalized Poisson regression analysis show that, variables such as: religion, child is alive, sex of child, age of respondent at time of first birth, age of household head and urban-rural status are statistically associated with total children ever born in Nigeria. There is a strong relationship between urban-rural status and the mean of children ever born. The risk ratio of children ever born by mothers who reside in rural area is 1.023 times those who reside in urban area. The risk ratio of children ever by mothers who are Christian/Others is 0.981 times the risk ratio of children ever born by Muslim/Islam mothers. For child is alive, a woman whose child is not alive had 1.104 risk ratio compared to children ever born by a woman whose child is alive. For sex of child, women who gave birth to female children had risk ratio of 1.004 compared to those who gave birth to male children. Age of respondent at time of first birth is also positively related to total children ever born.

Table 5.3: The Generalized Poisson regression risk ratios extracted for main effects which were not involved in the interaction

Factors	Risk ratios	Z-Value
Age of household head	1.010	68.89
Age of respondent at time of first birth	0.981	-17.02
Urban-rural status (Reference=Urban)		
Rural	1.023	4.53
Religion (Reference=Muslim/Islam)		
Christian/Others	0.915	-11.19
Sex of child (Reference=Male)		
Female	1.004	0.1247
Child is alive (Reference=Yes)		
No	1.104	4.41

Figure 5.3 reveals that the effect of fertility preferences on the predicted mean of total children ever born differs with age at first marriage or cohabitation and region, with whether a woman choses to have another or undecided/others, the predicted mean of children ever born decreases as the age of a woman at first marriage increases. It can therefore be concluded that unlike Poisson regression and Negative Binomial, the difference in the effect of fertility preferences on the predicted mean of children ever born is significant for all ages of a woman at time of first marriage showing that a woman in the group of undecided/others give birth to more children. Regarding region, the difference between the predicted mean of children ever born by a woman who is in the group of have another in fertility and a woman in the group of undecided/others is significantly different in all regions ($\rho < .0001$). In all the regions, the

predicted mean of children ever born by a woman that is undecided/others is more than those of have another woman. This is consistent with the report of NDHS 2013.

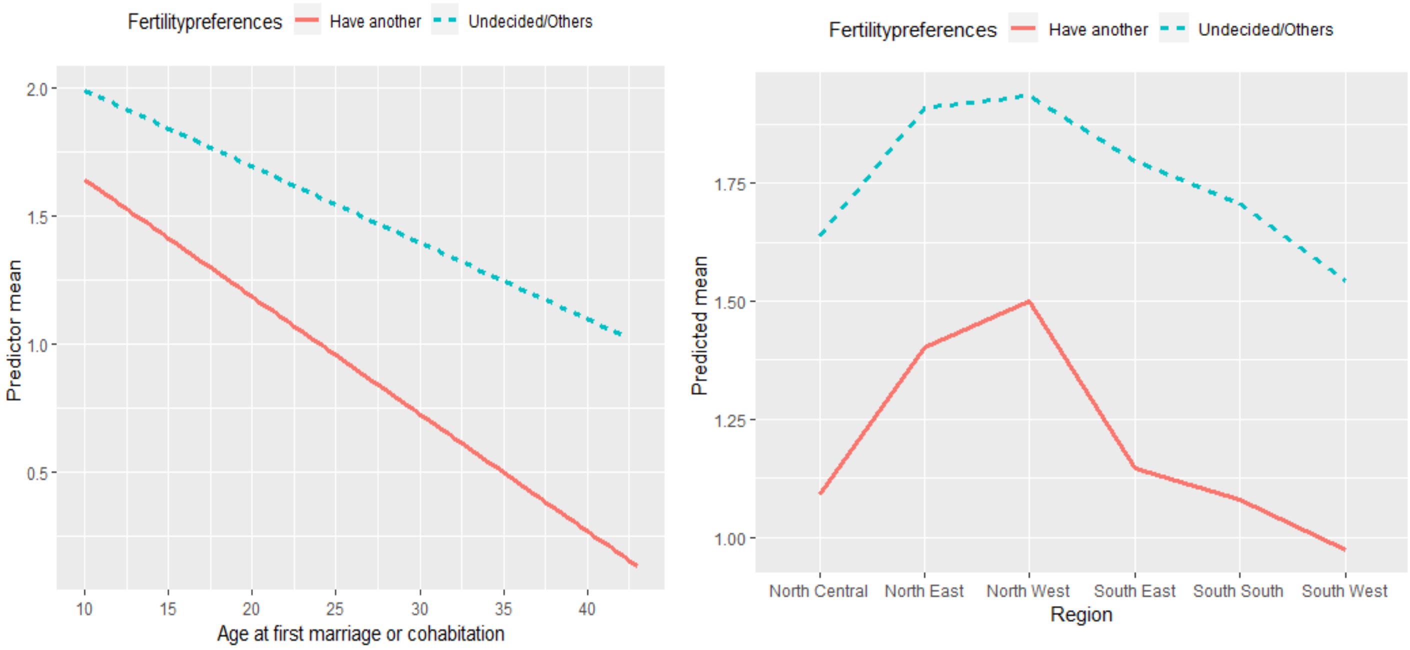


Figure 5.3: The mean number of children ever born by fertility preferences, age at first marriage or cohabitation and region

The relationships between education, whether and when this pregnancy is wanted, and child is alive are shown in Figure 5.4. Regarding whether and when this child's pregnancy is wanted, the difference in the effect of education on the predicted mean of children ever born is significant for all levels, with a woman who has no education giving more birth than her compatriates. The difference between the predicted mean of children ever born by a woman who has any of the educational level (whether no education, primary or secondary/others) is significant in all levels of child is alive ($\rho < .0001$). In all levels of child is alive, the predicted mean of children ever born by a woman with no education is higher compared to her contemporaries.

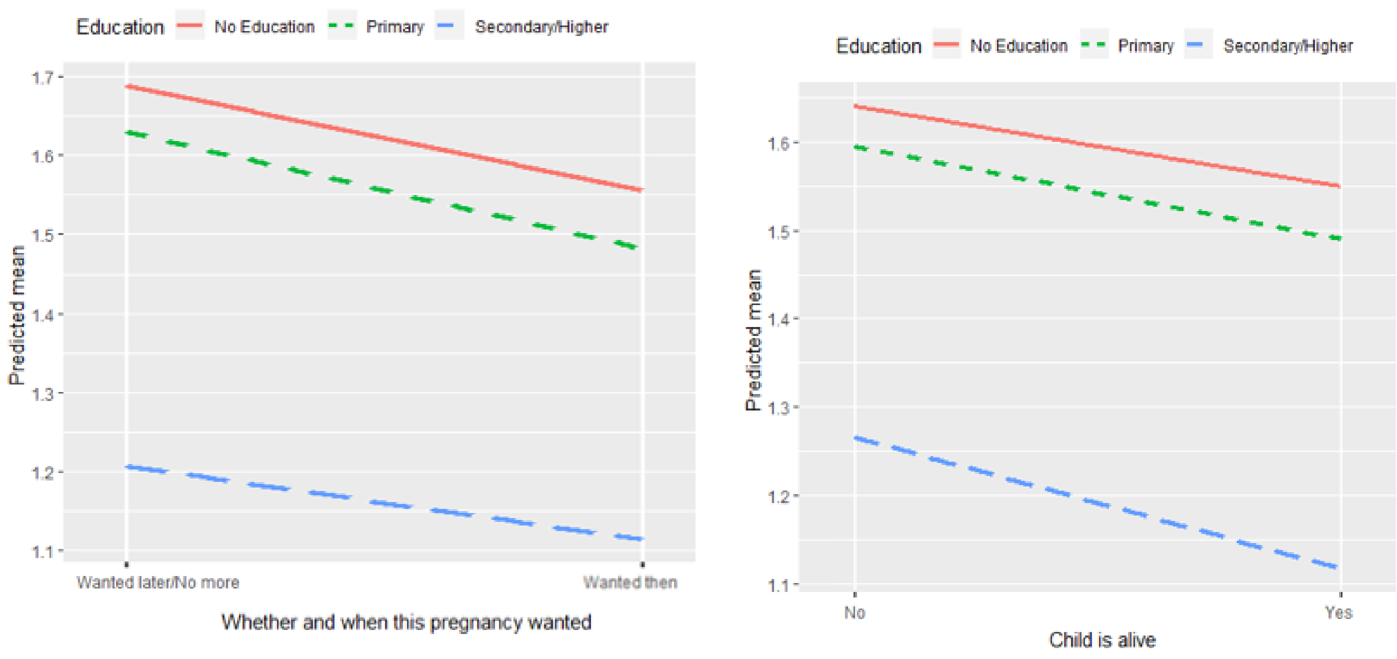


Figure 5.4: The mean number of children ever born by education, whether and when this child's pregnancy is wanted, and child is alive

Figure 5.5 presents the relationship between woman's occupation and ever had pregnancy terminated via abortion, miscarriage or stillbirth. The effect of woman's occupation on the mean of children ever born is significant for all group in ever had pregnancy terminated via abortion, miscarriage or stillbirth. The predicted mean of children ever born from a woman who has had abortion, miscarriage or stillbirth is more than one who has not.

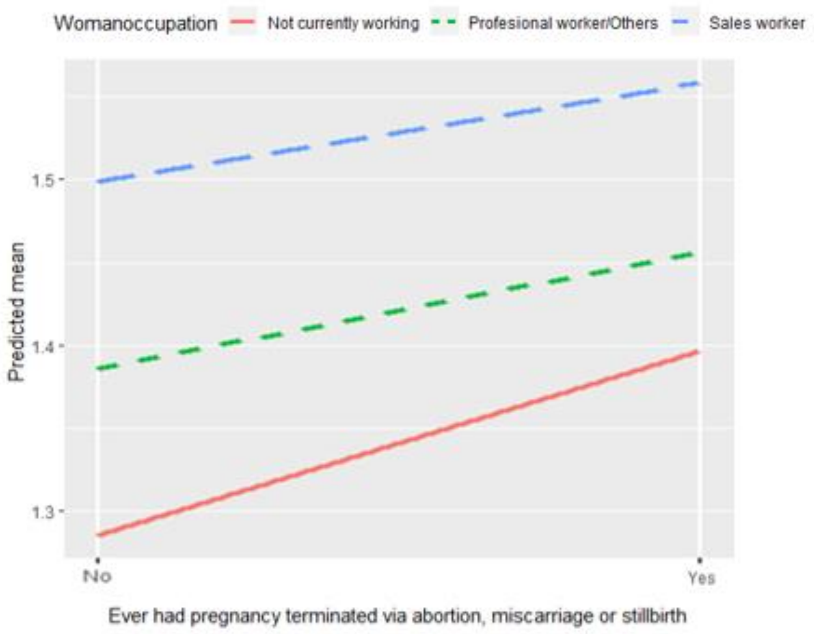


Figure 5.5: The mean number of children ever born by woman's occupation and ever had pregnancy terminated via abortion, miscarriage or stillbirth

Figure 5.6 presents the relationships between kidtwin and fertility preferences and shows that the effect of kidtwin on the predicted mean of children ever born is significant for all group of fertility preferences. It is noted that the predicted mean of children ever born from a woman who has multiple birth is more than one who had single birth.

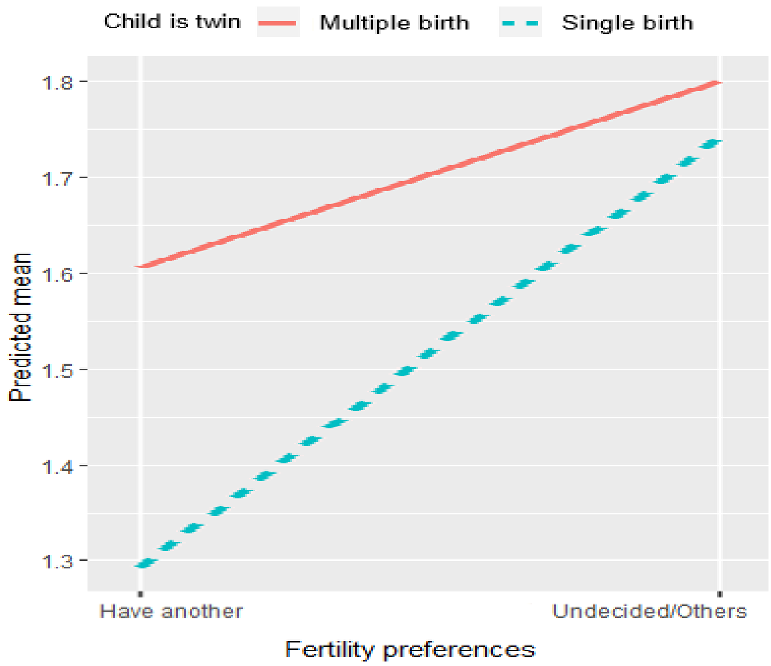


Figure 5.6: The mean number of children ever born by kidtwin and fertility preferences

The Generalized Poisson regression can account for under-dispersion displayed by the deviance of the Negative Binomial regression. The result indicate that this model accommodates both over-dispersion and underdispersion in count data which is in line with the findings of Islam et al. (2013).

Chapter 6: Predictive count modelling

One major objective in statistical analysis is to make predictions and to provide appropriate measures of uncertainty related to them. Therefore, estimates are expected to be probabilistic in nature, taking the form of probability over future quantities and events (Dawid, 1984). In view of this, predictive modelling is known as the process of creating, testing and validating a model to best predict the probability of an outcome (Liu et al., 2008). It can also be said to be a procedure for using known results to create, process and validate a model that can be used to predict future outcomes. Depending on the defined boundaries, predictive modelling which is synonymous with the field of machine learning is more commonly referred to in academic or research and development contexts (Finlay, 2014). When predictive modelling is deployed commercially it is often referred to as predictive analytics.

Predictive analytics as a data mining technique uses its tool in an effort to give answers to the question “what might possibly happen in the future?” In trying to determine the probability of a set of data belonging to another set of data, models can use one or more classifier. The available models on the modelling portfolio of predictive analytics software help to derive new information about the data and to develop predictive models (Archak et al., 2011). In general, the essence of this model is to test, validate and evaluate the model using the detection theory to guess the probability of an outcome in a given set of input data (Liu et al., 2008). This is done to be able to predict the future, and to enhance and enable rapid decision making at the level of the individual patient, client and customer.

In this chapter, we evaluated the predictive distribution for count data as they occur in a wide range of demographic application. Our focus was on using predictive count modelling to display the ability to correctly predict models that best describe the factors that affect children ever born. This approach was implemented to precisely identify the best model for predicting any data by comparing the performance of each model used. This chapter presents the method and techniques for analyzing the data are explained and the comparison results of the Poisson regression, Negative Binomial regression and Generalized Poisson regression followed by the conclusion.

6.1 Analysis of the data

One major issue in fitting a model is how well it performs when applied to new data. To solve this problem, to the data needs to be partitioned data into training, which is used to create the model validation, which is used to evaluate the model performance, and test which is used to access how well the algorithm was trained using the training dataset. Partitioning is performed randomly to protect against a biased partition according to the proportion specified by the user. In this case, using SAS version 9.4, a comparison of 60% training and 40% validation, 70% training and 30% validation, 80% training and 20% validation also 90% training and 10% validation was performed respectively to examine the three models behaviours (Poisson regression, Negative Binomial regression and Generalized Poisson regression). In addition, to examine the stability of the training parameters under each partition. Firstly, the model is fit on a training dataset, that is a set of examples used to fit the parameters of the model. Using a supervised learning method specifically, the model is trained on the training dataset. The training dataset usually consists of pairs of input vector (or scalar) and the corresponding output scalar (or vector) which is normally denoted as the target (or label). The training dataset is now run with the current model to produce a result that is compared with the target for each input vector in the training dataset. Regarding the result of the comparison and specific learning algorithm being used, the parameters of the models are adjusted, while variable selection and parameter estimation can be included in the model fitting (Brownlee, 2017).

Sequentially, in the validation dataset, the fitted model is used to predict the responses for the observations. While tuning the model’s hyperparameters, the validation dataset provides an unbiased evaluation of a model fit on the training dataset (James et al., 2013).

It is important to note that the underlying assumption, when presenting the root mean square error (RMSE), is that the errors are unbiased and follow a normal distribution. Therefore, using an RMSE helps to provide a complete picture of the error distribution (Chai and Draxler, 2014). Nevertheless, the mean absolute error (MAE) is also one of the metrics for assessing and summarizing the quality of a machine learning model, while mean squared error assesses the quality of a predictor or an estimator (Wackerly and Scheaffer, 2008), and the coefficient of determination (R^2) measures the closeness of the data to the fitted regression line.

In this work, mean absolute error (MAE), Mean squared error (MSE), root mean square error (RMSE) and coefficient of determination (R^2) are the performance evaluation metrics used. The formulas are presented below,

Root Mean Square Error (RMSE) is given as:

$$RSME = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$

Mean Absolute Error (MAE) is given as:

$$MAE = \frac{\sum_{i=1}^{n}|predicted_i - actual_i|}{N} = \frac{\sum_{i=1}^{n}|e_i|}{N}$$

Mean squared error (MSE) is given as:

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(predicted_i - actual_i)^2$$

Where N is the total number of observations.

Coefficient of determination (R^2):

$$R^2 = cor(actual_1, predicted_1)^2$$

Tables 6.1-6.4 contain the summarized results of the comparison of Poisson regression, Negative Binomial regression and Generalized Poisson regression using 60%:40%, 70%:30%, 80%:20% and 90%:10% respectively.

Table 6.1: Summary of Poisson, Negative Binomial and Generalized Poisson regression results for 60%:40%

		MAE	**MSE**	**RMSE**	**R^2**
Poisson	Training	1.613814	4.313774	2.076963	0.3624504
	Validation	1.600686	4.262919	2.064684	0.3604352
Negative Binomial	Training	1.613813	4.313784	2.076965	0.3624491
	Validation	1.600686	4.26293	2.064686	0.3604339
Generalized Poisson	Training	1.613700	4.315765	2.077442	0.3622371
	Validation	1.600644	4.264933	2.065171	0.3602402

Table 6.2: Summary of Poisson, Negative Binomial and Generalized Poisson regression results for 70%:30%

		MAE	MSE	RMSE	R^2
Poisson	Training	1.605426	4.273806	2.067319	0.3588404
	Validation	1.616214	4.344664	2.084386	0.366524
Negative Binomial	Training	1.605426	4.273815	2.067321	0.3588393
	Validation	1.616214	4.344675	2.084388	0.3665226
Generalized Poisson	Training	1.605347	4.276152	2.067886	0.3585859
	Validation	1.616177	4.347424	2.085048	0.3661882

Table 6.3: Summary of Poisson, Negative Binomial and Generalized Poisson regression results for 80%:20%

		MAE	MSE	RMSE	R^2
Poisson	Training	1.611256	4.305445	2.074957	0.3598963
	Validation	1.594595	4.218068	2.053794	0.3722824
Negative Binomial	Training	1.611255	4.305455	2.074959	0.359895
	Validation	1.594595	4.218074	2.053795	0.3722818
Generalized Poisson	Training	1.611217	4.307683	2.075496	0.3596542
	Validation	1.594352	4.218566	2.053915	0.3722467

Table 6.4: Summary of Poisson, Negative Binomial and Generalized Poisson regression results for 90%:10%

		MAE	MSE	RMSE	R^2
Poisson	Training	1.609629	4.299476	2.975103	0.3615511
	Validation	1.603307	4.213045	2.999453	0.3655643
Negative Binomial	Training	1.609629	4.299486	2.975114	0.3615499
	Validation	1.603307	4.213054	2.999464	0.3655631
Generalized Poisson	Training	1.609559	4.301766	2.977169	0.3613071
	Validation	1.603503	4.216727	3.001452	0.3650468

Based on the results of mean absolute error and root mean square error for Poisson, Negative Binomial and Generalized Poisson regression model, the performance evaluation for the training sample is higher than the validating sample, although with a slight difference (Aertsen et al., 2010; Onoro-Rubio and López-Sastre, 2016). The results as presented in Tables 6.1-6.4, and identified Poisson as the best predictive model as it gave the best performance for validating samples.

In conclusion, comparing the root mean square error, mean squared error, R-squared and mean absolute error for training and validating sample of each model, showed that all the three models had almost identical performance evaluation metrics (Ghanbari, 2019). The Poisson regression was chosen as the best because it is the simplest model, this being important because it balances the goodness of fit with simplicity and predicts the probability of the outcome. Complex models adapt their shape to fit the data but the additional parameter may not represent anything useful.

Chapter 7: Conclusion

Count data, known as a type of data that takes non-negative integers has a wide application in real life. The most common models for count data are Generalized Poisson regression, Poisson regression, and Negative Binomial regression. Count data models have been applied in most of the real-life happenings, it was applied in this study to identify the factors affecting total children ever born, which poses a concern to every society. This can be explained from the point that fertility is among the major determining factor of population growth and patterns.

Firstly, descriptive statistics were carried out to summarize the dataset in a useful and informative manner (Shmueli, 2010; Upadhyay, 2017). A comparison of three statistical methodologies on total children ever born revealed that, for Poisson regression, urban-rural status and age of household head had a significant effect on total children ever born. For urban-rural status, rural women had more children ever born than urban women. In terms of age of household head, a household head from age 42-67 had more children ever born than those at the age of 68years and above while a household head from age 41years and below had fewer children ever born than those of 68years and above.

From the interaction result, it is found that age at first marriage or cohabitation had a significant effect on total children ever with no more. Regardless of the age at first marriage or cohabitation, women whose fertility preferences choice is to have another birth are more powerful in the decision of how many children they will give birth to. Furthermore, in region, women from South South who desired to have another birth had more influence in child bearing decision making. In ever had pregnancy terminated via abortion, miscarriage or stillbirth, women in the group of have not and who are professional worker/others in terms of woman's occupation had upper hand in decision making. Consequently, for fertility preferences, women who are undecided/others and have single birth had more power in decision of child bearing. While in whether and when this pregnancy is wanted, women who wanted pregnancy then and have secondary/high educational level are most influential in child bearing decision making. With respect to child is alive, women with secondary/high educational level whose children are alive had more influence in decision making in the family.

Negative binomial was introduced to check for over-dispersion or under-dispersion, and it was found to be under-dispersed from the deviance result after fitting the data from Poisson regression.

Generalized Poisson regression seemed to be an appropriate model to detect factors affecting children ever born. Age of household head, age of respondent at the time of first birth, urban-rural status, and religion are significantly associated with total children ever born. Early marriage, religious belief and unawareness of women who dwell in rural areas should be checked to control total children ever born in Nigeria. This result follows the conclusion from Ozmen and Famoye, (2007) and Islam et al. (2013).

In the interaction of region with fertility preferences, women from South East who desire to have another birth, have more power in family decision making. However, for age at first marriage and fertility preferences (where women irrespective of their age who are undecided/others) have the highest decision-making power. Women who have not terminated pregnancy via abortion, miscarriage or stillbirth and are not currently working, have a better say in child bearing decision making. Regarding fertility preferences and child is twin, women who are undecided/others with single birth have dominant power for decision. For whether and when this pregnancy is wanted and education, women who wanted no more with primary education had more say in child bearing decision making. Finally, in child is alive and education, women in the group of no and have secondary/high educational level had more dominant power in family planning.

In the predictive modeling, all the three models showed almost identical performance evaluation metrics while the Poisson regression was chosen as the best as it is the simplest model. This is because the root mean square error, mean squared error and the mean absolute error of the three models showed almost identical performance metrics. From the results obtained, in the inferential modeling, the Generalized Poisson Model was found to be superior, while in the predictive modeling, all three models showed almost identical performance evaluation metrics, with the Poisson regression being choosen as the best due to it being the simplest model. These results provide important information on how the age of household head, age at first marriage or cohabitation, ever had pregnancy terminated via abortion, miscarriage or stillbirth, age of respondent at first birth, urban-rural status, region, religion woman's occupation, fertility preferences, whether and when this child's pregnancy is wanted, child is alive, child is twin and education are associated with total children ever born. Age of household head, age at first marriage or cohabitation, urban-rural status, region, religion, woman's occupation, fertility preferences, whether and when this child's pregnancy is wanted, child is alive, child is twin, and education were found to be the most important variables for predicting factors affecting total children ever born. In addition, uneducated women who are rural dwellers give birth more than their counterparts due to unawareness and lack of exposure to family planning programs. To authenticate the findings, larger study is needed to confirm these findings. Based on objective 2 of this study, Generalized Poisson regression served as alternative for handling count data associated with socio-economic and demographic factors affecting the total children ever born to respondents in Nigeria and, we believe that the findings to a large extent may be used to satisfy the fulfilment of the aim of the study.

There were some limitations to the research, specifically that secondary data were analyzed. Missing values, inaccurate information from the respondents and some important variables could not be investigated further. For example, currently pregnant, husband/partner lives in woman's household, ever been married and sex of household head were removed due to recording and coding error.

Future studies need to focus on more explanatory variables that might be available from other sources. It will be useful in future research to investigate the motive and reason for child bearing, and to establish the present trend, as the increase in total children ever born could have an adverse effect on the Nigerian economy and security.

Reference

Adhikari, R. (2010). Demographic, socio-economic, and cultural factors affecting fertility differentials in Nepal. *BMC pregnancy and childbirth* **10**(1): 19.

Adhikari, N. (2014). A study on number of child ever born: A case study of dayanagar VDC of Rupendehi Distric, Central Department of Economics, Tribhuwan University Kritipur, Kathmandu, Nepal.

Aertsen, W., Kint, V., Van Orshoven, J., and Ozkan, K. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological modelling* 221(8): 1119-1130.

Agresti, A. (2002). Categorical Data Analysis. John Wiley & Sons. Inc., Publication.

Alaba, O. O., Olubusoye, O. E., and Olaomi, J. O. (2017). Spatial patterns and determinants of fertility levels among women of childbearing age in Nigeria. *South African Family Practice* 59(4): 143-147.

Alkema, L., Raftery, A.E., Gerland, P., and Clark, S. J. (2012). Estimating trends in the total fertility rate with uncertainty using imperfect data: Examples from West Africa. *Demographic research* 26(15).

Archak, N., Ghose, A., and Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science* **57**(8): 1485-1509.

Bayarri, M., and DeGroot, M. (1987). Information in selection models. *Probability and Bayesian statistics, Springer***:** 39-51.

Bhende, A. A. and Kanitkar, T. (1982). Principles of population studies (demography). 2nd rev.

Brownlee, J. (2017). What is the Difference Between Test and Validation Datasets. Machine Learning Process,[https://machinelearningmastery. com/difference-testvalidation-datasets].

Bulatao, R. A., Lee, R. D., Hollerbach, P.E., and Bongarts, J. P. (1983). Determinants of fertility in developing countries.

Cameron, A. C., and Trivedi, P. K. (1996). 12 count data models for financial data. *Handbook of statistics* **14**: 363-391.

Cameron, A. C., and Trivedi, P. K (1998). Regression analysis of count data, Cambridge, United Kingdom: Cambridge University Press

Cameron, A. C., and Trivedi, P. K. (2013). Regression analysis of count data, Cambridge university press.

Chai, T., and Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)–Arguments against avoiding RMSE in the literature. *Geoscientific model development* **7**(3): 1247-1250.

Chatfield, C., Zidek, J., and Lindsey, J. (2010). An introduction to generalized linear models, Chapman and Hall/CRC.

Consul, P. C., and Jain, G. C. (1973). A generalization of the Poisson distribution. *Technometrics* **15**(4): 791-799.

Cook, J. D. (2009). Notes on the negative binomial distribution. Unknow, October **28**: 2009.

Crawley, M. J. (2012). The R book, John Wiley & Sons.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*: 278-292.

Dick, R., Agel, J., and Marshall, S. W. (2007). National Collegiate Athletic Association injury surveillance system commentaries: introduction and methods. *Journal of athletic training* **42**(2): 173.

Doty, J. H. (2017). Meta-Model Results For Polynomial Chaos Method. 55th AIAA Aerospace Sciences Meeting.

Du, J., Park, Y-T., Theera-Ampornpunt, N., McCullough, S. J., and Speedie, S. M. (2011). The use of count data models in biomedical informatics evaluation research. *Journal of the American Medical Informatics Association* **19**(1): 39-44.

Finlay, S. (2014). Predictive analytics, data mining and big data: Myths, misconceptions and methods, Springer.

Frome, E. L. (1983). The analysis of rates using Poisson regression models. *Biometrics*: 665-674.

Garenne, M. (2013). Situations of fertility stall in sub-Saharan Africa. *African Population Studies* **23**(2).

Ghanbari, A. (2019). Construction Demand Forecasting based on Conventional and Supervised Machine Learning Methods.

Grimmett, G. and Welsh, D. (2014). Probability: an introduction, Oxford University Press.

Gschlößl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical papers* **49**(3): 531.

Guikema, S. D. and Goffelt, J. P. (2008). A flexible count data regression model for risk analysis. Risk Analysis: *An International Journal* **28**(1): 213-223.

Hartnett, C. (2000). Principles and Recommendations for Population and Housing Censuses, Revision 1-United Nations, New York, 1998, 274 pages, ISBN 9211614023, UN Sales No. E. 98. XVII. 8, $35.00. *Journal of Government Information* **27**(2): 250-254.

Isgin, T., Bilgic, A., Forster, L., and Batte, M. T. (2008). Using count data models to determine the factors affecting farmers' quantity decisions of precision farming technology adoption. *Computers and electronics in agriculture* **62**(2): 231-242.

Islam, M. M and Alam, M. (2013). Predictors of the number of under-five malnourished children in Bangladesh: application of the generalized poisson regression model. *BMC public health* **13**(1): 11.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An introduction to statistical learning, Springer.

Jensen, J. L. W. V. (1902). Sur une identité d'Abel et sur d'autres formules analogues. *Acta mathematica* **26**(1): 307-318.

Kim, J. M and Jun, S. (2016). Zero-inflated poisson and negative binomial regressions for technology analysis. *Int. J. Softw. Eng. Appl*, 10, 431-448

Li, W. (1991). Testing model adequacy for some Markov regression models for time series. *Biometrika* **78**(1): 83-89.

Liu, Y., Huang, X., X., An, A., and Yu, X. (2008). Modeling and predicting the helpfulness of online reviews. Data mining, 2008. ICDM'08. Eighth IEEE international conference on, IEEE.

Long, S. J. (1997). SAGE Publications, Inc. Thousand Oaks, California. Regression models for categorical and limited dependent variables **7**.

Macro, I. and Commission, N. P. (2014). Nigeria demographic and health survey 2013.

Maxwell, O., Mayowa, B. A., Chinedu, I. U., and Peace, A. E. (2018). Modelling Count Data; A Generalized Linear Model Framework.

Melliana, A., Setyorini, Y., and Eko, H. (2013). Purhadi: The comparison of generalized Poisson regression and negative binomial regression methods in overcoming overdispersion. *IJSTR* **2**: 255-258.

Mohd, M. A. and Naing, N. N. (2012). Comparison between Negative Binomial and Poisson Death Rate Regression Analysis: AIDS Mortality Co-infection Patients.

Muluneh, E. K., Zewotir, T., and Yayeh, T. M. (2016). Fertility Status of Married Women and Its Determinants in Ethiopia.

Ntzoufras, I., Ntzoufras, I., Katsis, A., and Karlis, D. (2005). Bayesian assessment of the distribution of insurance claim counts using reversible jump MCMC. *North American Actuarial Journal* **9**(3): 90-108.

Onoro-Rubio, D. and López-Sastre, R. J. (2016). Towards perspective-free object counting with deep learning. European Conference on Computer Vision, Springer.

Osuji, G., Obubu, M., and Obina-Ilono, H. (2016). Uterine fibroid on women's fertility and pregnancy outcome in Delta State, Nigeria. *Journal of Natural Sciences Research* **6**(2): 27-33.

Ozmen, I. and Famoye, F. (2007). Count regression models with an application to zoological data containing structural zeros. *Journal of Data Science* **5**(4): 491-502.

Ramirez, O. A. and Shultz, S. D. (2000). Poisson count models to explain the adoption of agricultural and natural resource management technologies by small farmers in Central American countries. *Journal of Agricultural and Applied Economics* **32**(1): 21-33.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel* **83**: 902-926.

Shapiro, D. and Gebreselassie, T. (2013). Fertility transition in sub-Saharan Africa: falling and stalling. *African Population Studies* **23**(1).

Shmueli, G. (2010): To explain or predict? *Statistical science*, 25(3), 289-310.

Singh, K. P. and Famoye, F. (1993). Analysis of rates using a Generalized Poisson regression model. *Biometrical journal* **35**(8): 917-923.

Tüzen, M. F. and Erbaş, S. (2018). A comparison of count data models with an application to daily cigarette consumption of young persons. *Communications in Statistics-Theory and Methods* **47**(23): 5825-5844.

Upadhyay, H. P. and Bhandari, K. R. (2017). Factors Associated with Children Ever Born: A Case Study of Somadi Village Development Committee of Palpa District, Nepal. *Advanced Journal of Social Science* **1**(1): 15-29.

Wackerly, D. and Scheaffer, W. (2008). Mathematical Statistics with Applications . Belmont, CA, USA, Thomson Higher Education, ISBN 0-495-38508-5.

Winkelmann, R. (2008). Econometric analysis of count data, Springer Science & Business Media.

Zamani, H. and Ismail, N. (2012). Functional form for the Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* **41**(20): 3666-3675.

Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in R. *Journal of statistical software* **27**(8): 1-25.

Zhou, H., Seigel, P. Z., Barile, J., Njai, R. S., Thompson, W. W., Kent, C., and Liao, Y. (2014). Peer Reviewed: Models for Count Data with an Application to Healthy Days Measures: Are You Driving in Screws with a Hammer? Preventing chronic disease