University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Mathematics

Estimating Adolescents Fertility in Kenya Using Poisson and Negative Binomial Regression Models With 2014 DHS Data

Research Report in Mathematics, Number 32, 2018

Antony MUTUNGI

November 2018

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in Social Statistic

Estimating Adolescents Fertility in Kenya Using Poisson and Negative Binomial Regression Models With 2014 DHS Data

Research Report in Mathematics, Number 32, 2018

Antony MUTUNGI

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master of Science Project

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in Social Statistics

Prepared for The Director Graduate School
University of Nairobi
Monitored by School of Mathematics

Abstract

Due to non-functioning and incomplete vital registration systems in many developing countries, estimates of adolescents' fertility rate is by indirect methods which derive data from retrospective information obtained during national housing censuses which are conducted every ten years, or from specialized demographic sample surveys conducted quinquennially. Generally, most indirect estimation methods are characterized by: reliance on historical data which may suffer from recall errors, large volume of data requirements, low quality data, high data collection costs and application difficulties in the field. The indirect methods are developed not as a possible replacement of direct methods, but rather as temporary tools for measuring fertility aspects of the population where systems for registration of vital events do not exist or are too incomplete to be directly useful. Different methods have been advanced to calculate the ASFR metric, each with a different result.

This study seeks to estimate and compare adolescents' birth rate using the Poisson and negative binomial generalized linear regression models and to establish the reliability of the techniques in producing differential effects from socio-economic characteristics that influence teenage pregnancy. The study will use the birth histories of adolescents age 15-
19 years derived from the 2014 DHS data for Kenya.

By use of the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC), the results showed that the Poisson regression model; AIC (1737.542) and BIC (1878.325) presented a comparatively better fit in modeling the differential effects than the negative binomial regression model; AIC (1813.856) and BIC (1954.639). The estimated ASFR for the Poisson model was 88 births per 1000women while that of the negative binomial model was 86 births per 1000 women, in comparison to the national ASFR estimate of 96 births per 1000 women. For further research, it is recommended to compare the Poisson regression model with other models that estimate under-dispersion in count data as well as further detailed analysis through variable reduction.

Dissertations in Mathematics at the University of Nairobi, Kenya. ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this project report is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

______________________ ______________________

Signature Date

ANTONY MUTUNGI
Reg No. I56/89270/2016

In my capacity as a supervisor of the candidate, I certify that this report has my approval for submission.

______________________ ______________________

Signature Date

Prof. Ottieno Joseph A. M.
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: jottieno@uonbi.ac.ke

Dedication

This piece of research work is dedicated to my beloved departed young siblings Solace Munuve and Alfred Mutinda and to my dearest father Thomas Mutungi.

Contents

Acknowledgments

My sincere gratitude goes to Prof. Ottieno J.A.M. for the continuous support and guidance throughout the process. I would also like to applaud the contribution of all the lecturers in the School of Mathematics for imparting knowledge and skills requisite to undertake this study. Special thanks also go to my dear wife Irene and children Faith and Moses for their moral support, prayers and understanding during the many times I locked up myself in the room and or put in so many hours at night to work on the project.

The contribution of my job colleague Ezekiel Ngure is much appreciated especially during the initial stages of sieving through different ideas to come up with a research topic and also mentorship during the data analysis process.
May God bless you all abundantly.

Antony Mutungi

Nairobi, 2018.

CHAPTER 1: GENERAL INTRODUCTION

1.1 Background Information

Part of Kenya's population growth is tied to relatively young population. Today, 75% of Kenyans are below 30 years, while 45% of total population is below 15 years. It is estimated that in the next two decades, Kenyan youth will have reached 24 million. For these reasons part of the solution to socio-economic and sustainable environmental development is to drastically reduce the current level of population growth rate. Female higher education has emerged as one of the factors which could accelerate this process. Equally important are the late age at marriage, increased employment, improved health care services including increased use of modern contraceptive methods.

The 2009 Kenya Population and Housing Census (KPHC) shows that adolescents age 10-19 years represent about 24 % of the country's total population (9.2 million). In Kenya, as in other parts of Sub-Saharan Africa, adolescents face severe challenges to their lives and general well-being. They are vulnerable to early and unintended pregnancy, unsafe abortion, female genital mutilation (FGM), child marriages, sexual violence, malnutrition and reproductive tract infections including sexually transmitted infections (STIs) as well as HIV and AIDS. The 2014 KDHS data estimates that 15% of women age 15-19 have already given birth while 18 % have started childbearing i.e. had a live birth or are pregnant with their first child.

According to an article published on www.dsw.org/, adolescents are at higher risk than adults of illnesses and death from pregnancy related complications, HIV infections, or other sexually transmitted infections and gender based violence (GBV). In addition, unintended pregnancies and early marriages are often the main cause for girls leaving formal education, which hence affects their families, community, and the future development of the country. The underlying causes for adolescent pregnancy and resulting school dropout are complex and include issues such as parent-child communication, gender inequity, a lack of services and high-quality comprehensive sexuality education for young adolescents (age 10-14 years); With the majority of Kenya's population under 25 years, and two out of every five people under the age of 15, there is an urgent need to address these challenges.

Parametric and non-parametric models have been reported to have useful applications in demographic research. Apart from being useful when creating hypothetical rate schedules in forecasting and projection, they also serve to condense complex data into smaller indices (Schmertmann 2003; Peristera and Kostaki 2007). Several models have; therefore, been proposed to model fertility as the major determinant (of the three demographic variables namely, fertility, mortality, and migration,); of the size and structure of any population. These models have been commonly created for the developed countries of the world and usually fit excellently the population they are intended to model (Hoem et al. 1981).
It is pertinent, however, to mention that though there are many fertility models in the literature, few have been specifically generated to describe age-specific fertility patterns in Africa; despite the fact that most governments of the sub Saharan African countries are targeting lower total fertility rates to meet the Millennium Development Goals (MDGs) (United Nations 2000). To make reaching these targets possible, a better understanding of the current pattern of age specific fertility rate (ASFR) of African countries is required.

Mathematical models, when well-constructed, can aid in this understanding as they provide better insight into some characteristics of the distributional pattern of fertility in Africa. The goal of any modeling exercise is to extract as much information as possible from available data and to provide an accurate representation of both the known and unknown aspects of the phenomenon being studied (Salomon and Murray 2001). Modeling fertility in Africa has also become necessary to enable a meaningful comparison of fertility across the countries in the region in the face of the current fertility transition. Already, fertility can be compared using a wide variety of existing conventional measures, summary indices, or averages that are commonly reported for fertility data. These include total fertility rate (TFR), general fertility rate, and the crude birth rate. Few comparisons, however, are made based upon the detailed distribution of the age-specific fertility curve. Not all information in the curve can be conveyed by these summary indices. There is still much to be described in terms of the variance, skew, kurtosis, and symmetry of fertility distributions for individual countries on the continent.

In this study, I seek to use Poisson and negative binomial regression models to estimate Age Specific Fertility Rate (ASFR) in Kenya for adolescents age 15-19 years. Contrary to the basic fertility rate measurements, these models offer the flexibility to assess the differential effects of various variables associated with fertility rates.

1.2 Definitions, Notations and Terminologies

Adolescence: The World Health Organization (WHO) defines an adolescent as any person between the age of 10 and 19 years. Adolescence is a period marked by significant growth, remarkable development and changes in the life course for boys and girls, filled with vulnerabilities and risks, as well as incredible opportunities and potential. The experiences of adolescents shape the direction of their lives and that of their families.

Indirect estimation methods: Indirect estimation seeks to estimate a demographic parameter that is difficult to measure directly from some indicator that can be accurately recorded and is largely, but not exclusively, determined by the parameter of interest. The effects of confounding variables on the indicator are then allowed for, so that the parameter of interest can be estimated.

Age specific fertility rate: Annual number of births to women age 15-19 years per 1000 women in that age group. It is also referred to as the age-specific fertility rate for women age 15-19 years

Fertility: Is the actual number of live births a woman has had by the end of her reproductive life span (15 - 49 years).

Current fertility: Refers to the births that occurred to a woman in the 12 months before the census date.

Total Fertility Rate (TFR): The average number of children a woman would have assuming that current age specific birth rates remain constant throughout her childbearing years which are considered to be ages 15 to 49. TFRs are for the 36 month period prior to the census.

Fertility differentials: looks at the relationship between a woman's background characteristics and her fertility.

Cumulative fertility: – captures a woman's complete childbearing history.

Mean Age at First Birth: Refers to the average age at which women have their first born child.

Nuptiality: Refers to marriage as a population phenomenon, including the rate at which it occurs, the characteristics of persons united in marriage, and the dissolution of such unions through divorce, separation, widowhood and annulment (Haupt and Kane, 1998)

Generalized Linear Models: Are flexible generalizations of the Ordinary Least Squares (OLS) regression that relates to the distribution of the response variable (expected value) to the systematic portion of the experiment (linear predictors) though a function (canonical) called a link function.

Incident rate: Reflects the number of cases in a population that are observed to occur within a population at risk for some outcome.

Incidence Rate Ratios (IRR): Reflects the change in the incidence rate across levels of one or more predictors. For example, an IRR=1, means that the incidence rate for outcome A is neither increasing nor decreasing per one unit increase on predictor X. an IRR=2, means that the incidence rate for outcome A increases by a factor of 2 for every one unit increase on predictor X. an IRR=0.5 means that the incidence rate for outcome A decreases by one-half for every one unit increase on predictor X. or, you could say that the incidence rate for outcome b increases by 1/0.5=2 for every one unit increase on a predictor.

Over-dispersion: often (but not always) occurs when you have a large number of zeros (0's) in the observed data stemming from very low frequency events. It can also occur when there are large counts in the upper tail of the data distribution.

1.3 Research Problem

Due to non-functioning and incomplete vital registration systems in many developing countries, estimates of adolescents fertility is by indirect methods which derive data from retrospective information obtained during national housing censuses which are conducted every ten years, or from specialized demographic sample surveys conducted quinquennially. Moultrie et al. (2013) observes that the registration of births in developing countries such as Kenya is incomplete due to various contributing factors such as lack of registration incentives, non-capturing of neo-natal cases either as deaths or births and late registration by parents/ guardians. Additionally Moultrie et al. (2013) indicates that researchers rely on census data to estimate total and recent fertility rates and flags out several quality issues that affect census data. These include; underreporting associated with increasing mothers' ages especially in relation to cases of deceased children or those living independently and over-enumeration occasioned by lack of clarity on the census reference period or where respondents tend to peg recent births to the reference period. Fertility estimates can also be derived from Demographic Health Surveys (DHS) based on data collected on birth histories. DHS data also suffer from distortions related to shifting recent births to distant years to avoid additional questions related to the child's data (Cleland 1996). Underestimation of fertility rates may be caused by omission and displacement of births as corroborated by Schoumaker (2010, 2011).

Generally, most indirect estimation methods are characterized by: reliance on historical data which may suffer from recall errors, large volume of data requirements, limited sampling designs, low quality data, high data collection costs and application difficulties in the field. The indirect methods are developed not as a possible replacement of direct methods, but rather as temporary tools for measuring fertility aspects of the population where systems for registration of vital events do not exist or are too incomplete to be directly useful.

In Kenya, the DHS computes ASFR by summing the live births that occurred in the three-year period preceding the survey classified according to the age of the mother (in five-year age groups) at the time of the child's birth. This is used to derive the numerators of the rate. The denominators of the rates represent the number of woman-years lived by the survey respondents in each of the five-year age groups during the specified period.

The inadequacy to model fertility differentials, compounded by the complexities of having to make reference to the survey dates and by working in three and five year age periods of calendar time are plausible weaknesses of this approach

There are various studies conducted to estimate fertility using different methods such as Poisson and negative binomial regression but none seems to run a comparative analysis of results, leading to the same results.

This study seeks to estimate adolescents' fertility in Kenya using Poisson and negative binomial regression and to establish reliability of the techniques in producing differential effects from various socio-economic and proximate variables.

1.4 Study Objectives

a) Main objective

To compare different methods of estimating Adolescents' Age Specific Fertility Rate.

b) Specific objectives

i) To use Poisson regression model to estimate ASFR in the three years reporting period preceding the survey.

ii) To use negative binomial regression model to estimate ASFR in the three years reporting period preceding the survey.

iii) To compare the results obtained by Poisson and negative binomial regression models with the national ASFR estimate.

1.5 Methodology

The data to be used in this study are national secondary data on adolescent fertility derived from the 2014 KDHS. The Fifth National Sample Survey and Evaluation Program (NASSEP V) was used to generate the sampling frame.

The KDHS survey employed a two stage stratified sampling procedure. The first sampling stage entailed cluster selection. A total of 1616 clusters (617 urban and 995 rural) were realized. The second stage entailed the sampling unit (household) selection. 25 households were selected per cluster leading to a total sample size of 40,300 households. 98.5% (39,679 households) of the selected households were sampled. Respondents interviewed in each sampled household were women age 15-49 and men age 15-54 in half of the sampled households.

1.6 Literature review

The review of literature will give an overview of the various definitions given to the term ASFR, as well as a general introduction to the Poisson and negative binomial regression model. This will follow a critical review of literature on methods used to estimate Age-specific fertility rates based on published secondary sources, with a focus on Poisson and negative binomial regression analysis.

1.7 Significance of Study:

The Age Specific Fertility Rate (ASFR) for women in Kenya age 15 – 19 years is 96 births per 1,000 women and most recently it is estimated to have increased to 121 births per 1,000 women (Kenya ASRH Policy, 2015). Even though the ASFR for adolescent has declined over the years, the contribution of a large pool of young women to Total Fertility Rate (TFR) is significant and increasing because of early child bearing and marriage at a young age.

According to the KDHS 2014, 15% of young women between the ages of 15-19 are already mothers and 3% are pregnant with their first child.

Percentage of women age 15-19 who:			
Age	Have had a live birth	Are pregnant with first child	Percentage who have begun childbearing
15	1.7	1.6	3.2
16	5.9	2.0	8.0
17	10.3	4.7	15
18	21.5	4.4	25.9
19	35.3	4.6	39.9
Average			18.4%

Source: KDHS 2014

Teen pregnancy continues to be a problem for families, educators, health care professionals, the government and the general public at large. Sub-Saharan Africa has one of the highest levels of teenage pregnancies in the world. In spite of that, there is insufficient empirical research on causes of teenage pregnancies involving key influencers in African countries like Kenya.

This study will therefore seek to provide simple yet more powerful and informative methods which provide better estimates for policy makers, researchers and other stakeholders. This is critical to design targeted development strategies aimed at addressing the high rates of adolescent pregnancy, and ultimately contribute to social and economic pay-offs including reduction in school dropouts rates, improving neonatal health, better work force and overall poverty reduction.

CHAPTER 2: LITERATURE REVIEW

2.1. Introduction

Fertility is an important metric to government agencies and non-governmental institutions that deal with issues related to population. One fertility metric is the age fertility rate, which, according to United Nations Population Division (2009), refers to the measure of number of births to women within a given age group within a given year per 1000 women in that age-group. While many authors agree with this definition, a number of methods have emerged that are used to calculate this metric, each with a different result. The first section of this chapter reviews the various definitions given on the term ASFR, followed by a critical review of literature on methods used to estimate Age-specific fertility rates based on published secondary sources. The study firstly reviews literature that used Poisson, followed by literature that used the negative binomial regression analysis. The last part is a summary of the literature review. This chapter also presents the conceptual, operational, and analytical frameworks to the study.

2.1.1. Defining Age-Specific Fertility Rate (ASFR)

Different definitions and computation of ASFR exist leading to varying results.

The Johns Hopkins Bloomberg School of Public Health (2017) defines age-specific fertility rate as the number of births per 1000 women within a specific age group in a specific place at a specific time, usually a year. This definition does not have any restrictions about the range of age that should be considered and neither does it have any consideration about the size of geographical region to be used. This definition is derived from the one given by the UNPD (2009), and is usually compiled by considering only live births.

$$ASFR = \frac{Number\ of\ live\ births\ to\ women\ in\ a\ specified\ age\ group}{Number\ of\ women\ in\ the\ same\ age\ group}\ x\ 1000$$

For instance, suppose a country, P, has 310,000 women aged between 26 years and 30 years and the total number of live births recorded for this age group totals to 72,000 for the year 2018, then the ASFR can simply be calculated thus:

$$ASFR = \frac{72{,}000}{310{,}000} \text{ x } 1000$$

ASFR = 232.3

This implies that there are 232.3 births per 1000 women who are within the age of 26 year and 30 years in the given country. While the same can be calculated for each year, it is usually calculated on a five-year age group basis, covering seven age-groups, which comprise 15-19 years, 20-24 years, 25-29 years, 30-34 years, 35-39 years, 40-44 years and 45-49 years. The results can then be plotted on a graph for easier comparison as illustrated in figure 1, which shows the ASFR for Italy and Zambia in 2006.

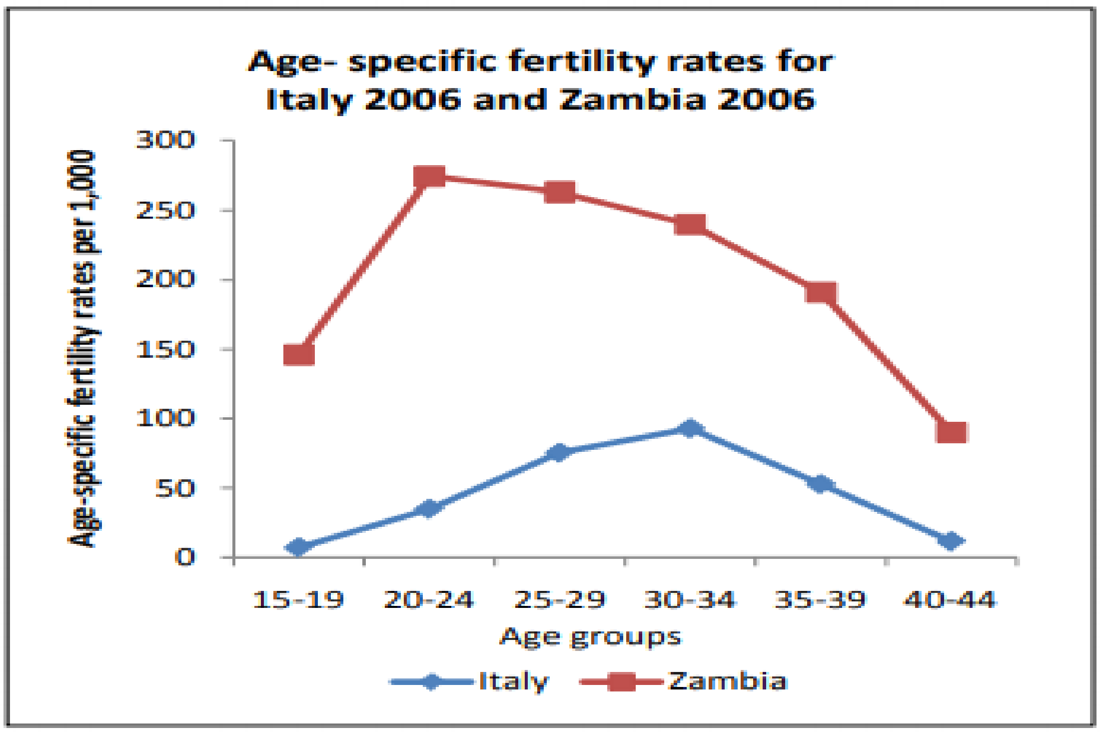

Figure 1: ASFR for Italy and Zambia in 2006

Source: Johns Hopkins Bloomberg School for Public Health (2017)

MEASURE Evaluation (2018), a non-government organization in the US, argues that ASFR can be calculated either on a single-year or five-years age groups. They define Age-specific fertility rates as the number of births occurring during a given year or reference period per 1000 women of reproductive age classified in single or five year age groups as shown in the formula below:

$$\mathrm{ASFR_a} = (\mathrm{B_a}/\mathrm{E_A}) * 1000$$

Where:

B_a = Number of births to women in age group (a) in a given year or reference period;

E_a = Number of person-years of exposure in age group (a) during the specified period.

2.1.2. Poisson regression and negative binomial regression

According to Gourieroux (2000), Poisson regression is a generalized linear model form of analysis, which is mostly used to model contingency tables and count data. Greene (2008) adds that Poisson regression has two main assumptions, the first of which is that it is possible to model the log of the expected value of a response variable through the use of unknown parameters. It also assumes that the response variable will have a Poisson distribution. Gourieroux (2000) gives the Poisson equation formula as follows:

$$log\big(E(Y/x)\big) = \beta_1 \mathrm{x}$$

Where x is a vector of the independent variables and concatenated to a vector of ones.

The Poisson distribution is represented as an equation as follows:

$$E(Y/x) = e^{\beta_1 \mathrm{x}}$$

Greene (2008) further points out that a negative binomial regression is a generalization of the Poisson regression. Its major strength is that it models the Poisson heterogeneity with a gamma distribution and this is because it allows the flexibility of the assumption made by Poisson distribution that the variance and the mean are equal (Jones et al., 2013).

Negative binomial distribution has the same sample space as the Poisson distribution, but also has an additional parameter that is used to model the variance. This parameter is referred to as the dispersion parameter (Gardner, Mulvey, & Shaw, 1995).

The image below provides a comparison between Poisson and negative binomial distributions (of different count outcomes), assuming that both have equal mean and variance (λ's), notice that the probabilities associated with the count outcomes differ between the distributions.

From figure 2 below, you notice that the probability distribution or relative frequencies for different values of the two distributions are different even though they look fairly similar. Around the center or mean of the distribution the relative frequencies for the Poisson distribution tend to be higher than the relative frequencies you'd observe for the negative binomial distribution.

As you move out into the upper tail the relative frequencies for the negative binomial counts tend to be higher than the relative frequencies that would be observed for the Poisson distribution.

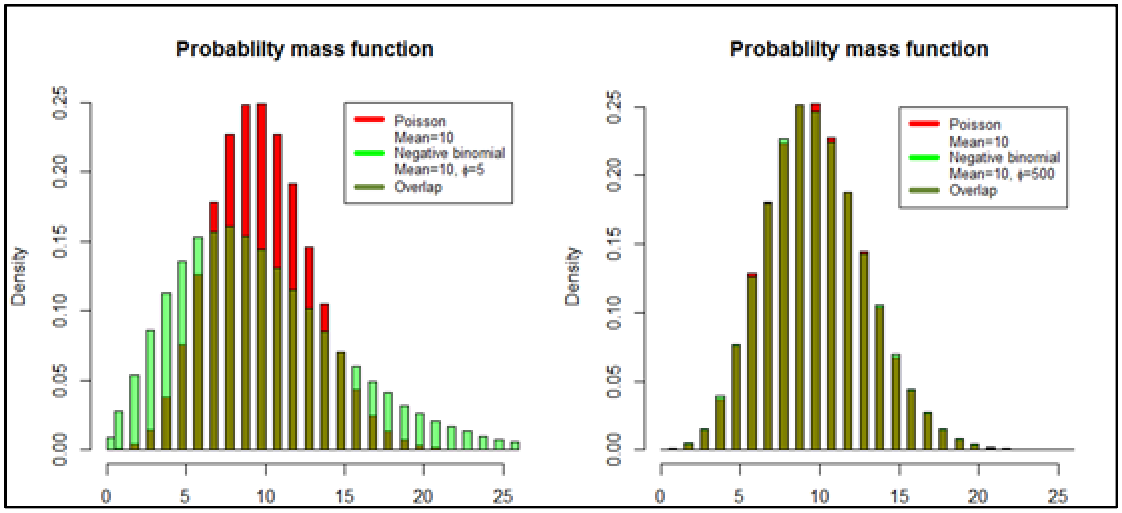

Figure 2: Comparison of Poisson and negative binomial distributions

Source: (doi:10.12746/swrccc2015.0310.135)

The aim of negative binomial regression is the same as that of Poisson regression: to model relationships between predictors and the likelihood of certain count outcomes. The probability distribution that is utilized – the negative binomial distribution – takes into account the issue of over dispersion in your data.

2.2 Estimation of fertility using Poisson and negative binomial regression models.

Various authors have used Poisson and negative regression to estimate count data, of which determination of fertility is one. Winkelmann and Zimmermann (1995) pointed out that Poisson regression model can be used to determine fertility of women by focusing on the number of children born at a particular period. Choi (2014) underscores two main advantages of using Poisson distribution. Firstly, it is able to capture non-negative discreet data and, secondly, it can be able to be used to determine probability of occurrence of an event. Li (2016) also studied Poisson distributions and noted one major problem. The authors argue that in Poisson model, the mean and variance of the dependent variable are curtained to be equal in theory.

However, in practice, the variance is often different from the mean, thus leading to inefficiency of the model estimates. Additionally, the parameters of regression analysis usually have errors and are mostly likely to bring too many null hypotheses (Winkelmann and Zimmermann, 1994). According to Choi (2014), because of its strengths, the Poisson regression model might be a good starting point in estimating fertility, but there is need to check the validity of the results afterwards. Li (2016) argues that one way this can be done is by running the negative binomial regression, especially if there is too much dispersion. However, Asiki et al. (2015) alludes that if there is equidispersion in the observed data, the negative binomial and the Poisson regression models should end up the same.

Based on the Poisson distribution model given above, which ensures that at all times, E(y/x)> 0

Thus, the number of children ever born to a woman can be equated as follows:

$$P(Y = y/x) = \frac{e^{-e(x'\beta)}e^{(x'\beta)^{y}}}{y!} \; Where\ y = 0,1,2, \dots N$$

To this end, the equation that represents estimated maximum fertility will be:

$$L(B) = L(\beta) = \sum_{1}^{i} \{y_i x_i \beta - e^{x_1 \beta}\}$$

The xi’s in the equation above represent the socioeconomic factors that can also influence the fertility of a woman as they significantly affect the physical and psychological characteristics.

$$\mu_i = e^{\beta_0 +} \sum_{j=1}^{k} \beta_j x_j$$

Where

μ_i = the expected number of children per woman i

e = the base of natural logarithms

β_0 = the intercept

β_j = regression coefficients

x_i*’s* = explanatory variables

Tsegaye (2010) studied the socio-economic factors that influence women’s fertility in Zambia in comparison to various other African countries, including Kenya, Lesotho, Rwanda, Uganda and Malawi, and used the Poisson model.

The study focused on total number of children ever born, and analyzed the data alongside 15 other variables that affect women, including age, education, marriage age, wealth index, rural or urban residence, province, occupation, contraceptives use and exposure to media like newspapers. They found that, in addition to age, all these factors also influence fertility although each has its different degree and direction of influence. They used Poisson results and multivariate analysis to confirm all their hypotheses, concluding that type of residence (rural versus urban) contraceptives, work, education and income are all negatively correlated with fertility decisions. When they used the same data to generate a negative binomial regression output, they found that the results were the same. This can be explained by considering that the variance and mean of the variables were equal.

As indicated in figure 3, Kenya has the second lowest Total fertility rates, just behind Lesotho, which, according to Tsegaye (2010), implies that they also are the most likely to have the lowest ASFR. Uganda has the highest followed by Zambia. This also implies that they are most likely the countries with the highest ASFR for all age groups listed.

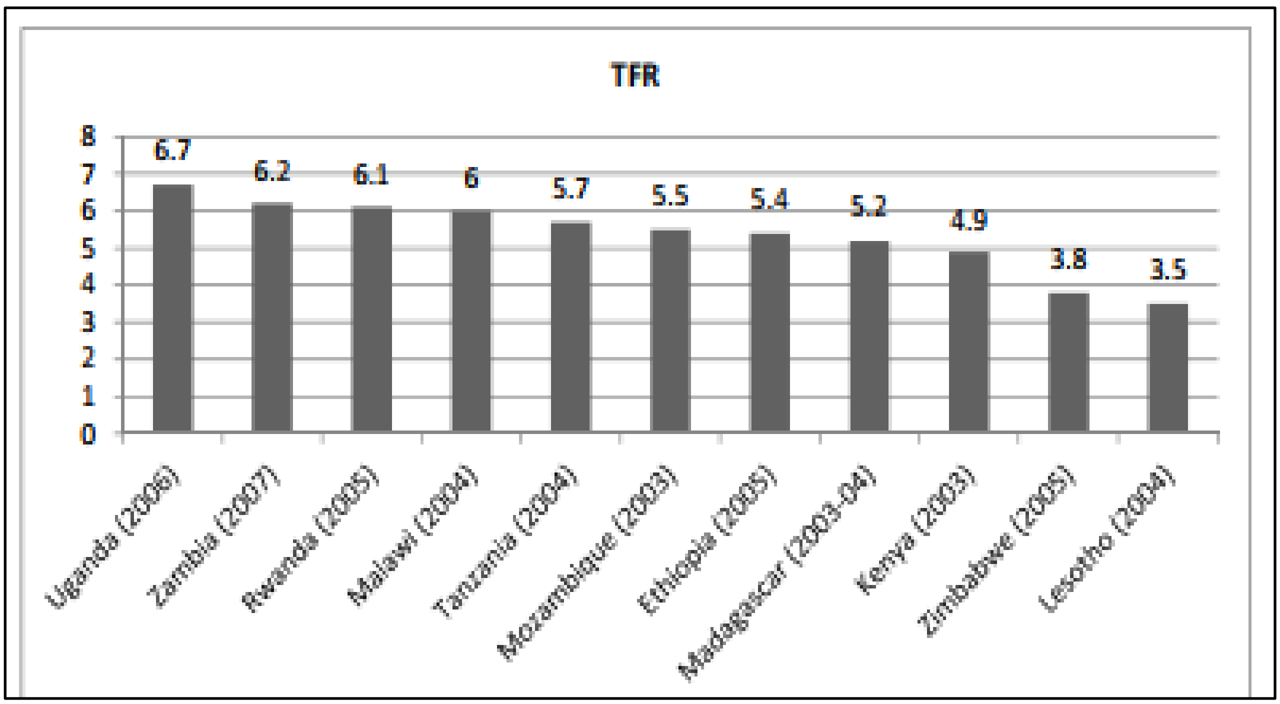

Figure 3: Total Fertility rates of various countries across Africa

Source: (Zambia Demographic and Health Survey, 2010)

Comparatively, Choi (2014) studied the fertility of women among Mexican immigrants in the US. The researcher considered data from the Mexican Life Survey, the 2006-2010 National Survey of Family and Growth (NSFG), and the 2002 NFSG. The first set of data comprised some 35,000 men and women and 8,400 households, and the 2002 NSFG data contained data on 7,600 women while the 2006-2010 NSFG data contained data regarding 12,200 women. The respondents ranged in age from 15 to 44 years and were grouped into five-year age brackets. The researcher as well considered their education, marital status, race and ethnicity, among other variables. Using Poisson regression, the ASFR was predicted and it was established that the fertility rates was 1.8 and 2.0 for whites and non-immigrants respectively, which is only slightly different from the official reports, which give the same data to be 2.0 and 2.2 in that order.

On the other hand, there was larger parity between those findings and the data given by official statistics, which said Mexican immigrants had a total fertility rate of 4.0 compared to the study's 2.8. There was over dispersion in the accounts of birth and this necessitated the use of negative binomial regressions. However, the model could not be used because it did not account well for marital status and other time-varying controls.

Li (2016) studied fertility and its determinants across three Asian countries, Cambodia, the Philippines, and Indonesia, which are three of the four countries that make up the ASEAN community. The author uses data from the Demographic Health Survey and considers the fertility of women between 15 years and 49 years of age. The data was analyzed using negative binomial regression analysis, which brought a number of conclusions to the fore. Firstly, the data showed that there was a negative correlation between education, as well as use of contraceptives. However, contrary to previous studies, the data showed that urbanization and husband's education level do not have a significant effect on the subject. The author determined that the women were influenced by education, use of contraceptives, as well as age at first marriage among other issues, to reduce the number of children they eventually have. However, the most influential variables were age at marriage and their propensity to use contraceptives.

Also contributing to the research is Ahmed and Ali (2016) also conducted a study that focused on modeling the factors influencing fertility in Sudan by using generalized linear models using Poisson regression models and negative binomial regression models.

This study was conducted in Sudan where women were used as the sampling frame. This research adopted a retrospective cross-sectional design and used the data collected by Sudan Household Health Survey (2010) (SHHS). From this, the researchers were able to identify the statistical advantages and suitability of the standard Poisson and negative binomial models for analyzing count data. The results indicated that both methods were viable to be used to predict the average number of children born by Sudanese women. The findings also revealed that a significant relationship existed between fertility and age at first marriage, gap between births, infant mortality, high level of education and wealth index. On the other hand, the negative binomial regression was better in prediction of the fertility of the Sudanese women compared to Poisson regression models. Despite this, the negative binomial regression method failed to detect the exact relationship that existed between women who are currently using family planning methods and the number of children born.

Opiyo (1987) studied the intercensal fertility estimation in Kenya.

The study considered the techniques used in estimating the intercensal fertility levels and accounted for a wide range of parameters such as gross rate of reproduction, net rate of reproduction, total fertility rate, crude birth rate and age at the year of data collection. The study used various methods to estimate the fertility rates, and the age-specific fertility rate was detained using regression method. The study found that the few studies on fertility had been carried out in the country, and the census data was used in order to compute the fertility rates. It was shown that different methods of estimating fertility rate yielded different results. Specifically, the study considered the Brass and Coale-Trussell ratio, the Coale-Demeny' ratio, the Gompertz model as well as the age-specific growth rate technique.

Contributing to this discussion is Alaba, Olubusoye and Olaomi (2017) who studied spatial patterns and determinants of fertility levels among women of childbearing age in Nigeria. The study used secondary sources from Nigerian Demographic Health Survey (NDHS) to analyze the determinants of fertility levels in Nigeria using the geo-additive model. Negative Binomial distribution was used to assess over dispersion of the dependent variable. In line with that, spatial effects were used to identify the key areas for high fertility levels. The researcher then drew their inferences using the Bayesian approach. The results of the study were presented within 95% credible interval (CI).

Using these methods, the results revealed that infertility among mothers was mostly associated with factors such as advanced education of the mother, family planning use, Yoruba ethnicity, higher wealth index, Christianity, and previous Caesarean birth.

Conclusion

The review of literature has focused on the ways in which age-specific fertility rate is measured using Poisson and negative binomial regression. A number of issues have been explored. It has been established that Poisson and negative binomial regression are closely related methods of analyzing data, with the main difference being that Poisson has strict assumptions where negative binomial regression relaxes the assumptions about variance and mean being equal. Additionally, it has also been shown that while the fertility rates can be calculated on a yearly basis, the most common method is to calculate it based on age groups of five years. Using both Poisson and negative binomial regression has also shown that the calculations take into account various variables, which include age, gender, ethnicity and awareness of the woman.

Notably, however, there are cases in which binomial regression was used instead of Poisson and the vice versa, and in each case there was good reason.

As such, the review concludes that the use of either Poisson or negative binomial regression depends on the need of the research as well as the underlying assumptions. Overall, the use of both binomial regression and Poisson distribution can help in estimating age-specific fertility rates among women in a reliable manner.

Conceptual/ Theoretical framework

Davis and Blake (1956) originally developed an analytical framework for fertility in which factors influencing fertility were categorized into two groups namely: background determinants and proximate determinants. The background determinants included: cultural, psychological, economic, social, health and environmental factors, which functioned through a framework of 11 proximate determinants.

Bongaarts (1978) restructured the work by Davis and Blake (1956) to develop a framework for the proximate determinants and a model for assessing their individual effect on fertility. The Bongaarts' new classification included eight proximate determinants of fertility namely: proportion married, contraception, induced abortion, lactational infecundability, frequency of intercourse, sterility, spontaneous intrauterine mortality and duration of the fertile period (Bongaarts, 1978).

Bongaarts and Potter (1983) analyzed data from 41 developed and developing countries and observed that, 96 per cent of variation in fertility among populations could be explained by four principal proximate determinants namely: proportion of women married, contraceptive use and effectiveness, induced abortion, and postpartum infecundability. In 1984, Bongaarts added a fifth variable, primary sterility to the proximate determinants model. This framework thus will form the basis for estimating age specific fertility rate in Kenya using Poisson and negative binomial regression.

Conceptual Framework (Bongaarts' Fertility Framework, 1984)

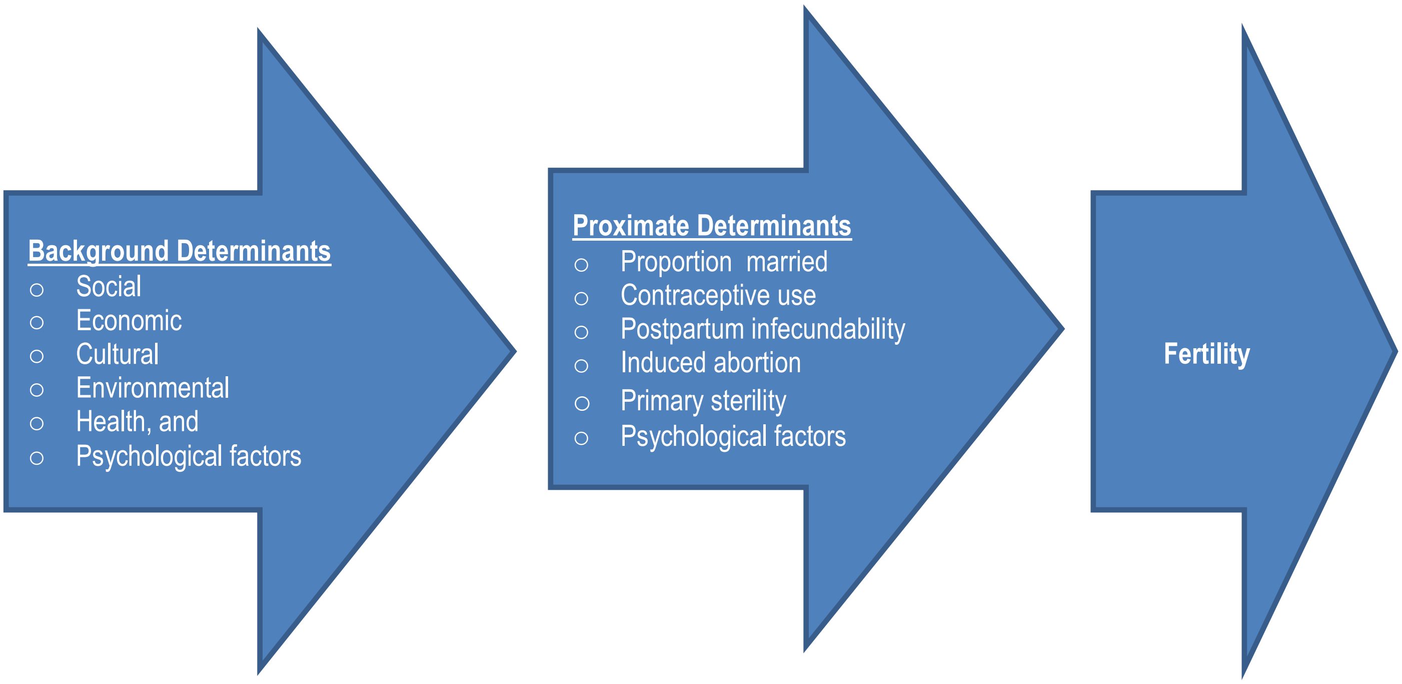

Figure 4: Bongaarts' Fertility Framework

Source: Bongaarts (1984)

Operational Framework

The study will operationalize the social and economic factors. These include educational level, contraceptive use, age at first sex, marital status, religion, frequency of listening to radio and wealth index. The operational framework is as shown in the figure below:

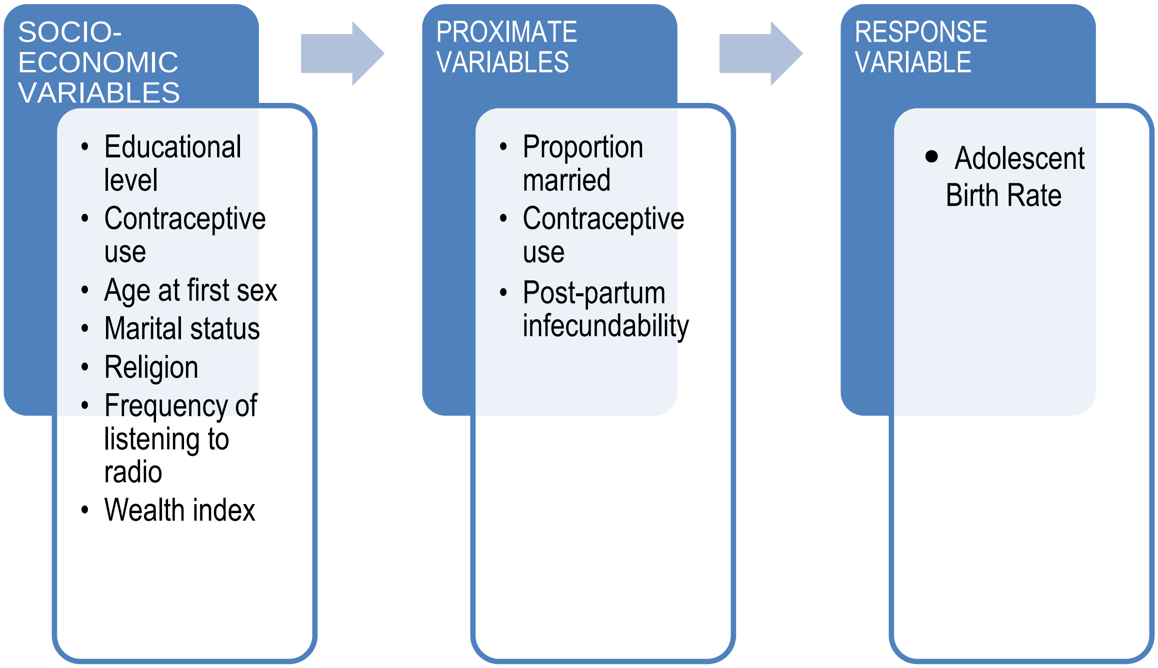

Figure 5: Operational Framework

Source: Adapted from John Bongaarts (1984)

CHAPTER III: Methodology

3.1 Introduction

This chapter presents a description of the source of data, its scope and limitations as well as the data analysis techniques. It also highlights the various statistical methods to be used and hypotheses to be evaluated.

3.2 Data Collection

The data to be used in this study are national secondary data on adolescent fertility derived from the 2014 KDHS. The Fifth National Sample Survey and Evaluation Program (NASSEP V) was used to generate the sampling frame.

The KDHS survey employed a two stage stratified sampling procedure. The first sampling stage entailed cluster selection. A total of 1616 clusters (617 urban and 995 rural) were realized. The second stage entailed the sampling unit (household) selection. 25 households were selected per cluster leading to a total sample size of 40,300 households. 98.5% (39,679 households) of the selected households were sampled. Women age 15-49 years were interviewed in each of the sampled households, while men age 15-54 years were interviewed in half of the sampled households. The number of eligible women for interview age 15-49 years was 31,172 out of which 31,079 were interviewed representing a 97% response rate.

Generally, the KDHS 2014 collected data on: family planning and fertility behaviour of the Kenya population, maternal and child health indicators, adult and maternal mortality levels, nuptiality, fertility preferences, awareness and use of family planning methods, use of maternal and child health services, levels and patterns of knowledge and behaviour related to prevention of AIDS and other sexually transmitted diseases.

Various questionnaires were used to collect data on the areas highlighted above: the full household questionnaire, short household questionnaire, full woman’s, short woman’s questionnaire and men’s questionnaires.

According to Moultrie et al. (2013), data gathered from the women's questionnaire is necessary to calculate the denominator of the fertility rates and the following key elements are required: the month and year of birth, month and year of interview, variables needed to adjust the data for sampling design and sample weights and covariates needed to model fertility differentials.

Additionally, the child's data are needed to calculate the numerator of the fertility rates and certain key elements are required such as the child's date of birth and essential details of the mother e.g. mother's date of birth.

For purposes of this study, I limit myself to data derived from the women's questionnaire and specifically on the birth histories of girls age 15 – 19 years. The analysis of the full dataset for cases of girls age 15 – 19 years yielded a sample size of 6078 cases (unweighted) and 5820 cases (weighted). The DHS 2014 used the number of weighted cases and a three year reference period to calculate the ASFR.

In this study Poisson and negative binomial models will be used to estimate the ASFR by operationalizing selected social and economic background variables through proximate variables. (Bongaarts, 1984). Just like the DHS, the estimates will be computed and compared using a three year reference period. (Moultrie et al., 2013). Reference periods of more than one year are preferred to compute ASFRs from DHS data, the justification is to reduce sampling variability associated with relatively small numbers of annual births occurring to women in single or five-year age groups and the distorting effects of reference period reporting errors. Various analyses of DHS fertility data use the three- or five-year period prior to the survey in calculating ASFRs (Arnold and Blanc, 1989; Lutz, 1990). When multiple years are used for computational purposes, average annual rates are normally presented.

3.2.1. Limitations of the DHS data

The DHS does not collect data for adolescent girls age 13 – 14 years. Therefore the central question is how to deal with the left censoring of under-15 exposure, which goes into the denominator of fertility rates, but is truncated because girls under age 15 are not included in the surveys.

Non-sampling and sampling errors have been highlighted as challenges affecting DHS data.

DeNavas, C. (1989) explains non sampling errors as the results of shortcomings in the implementation of data collection and data processing; such as failure to locate and interview the correct household, misunderstanding of questions on the part of either interviewer or respondent and data entry errors. Data quality assurance mechanisms were put into place for the KDHS 2014 to minimize these types of errors. Non-sampling errors are difficult to entirely avoid and to evaluate statistically. Estimated accuracy attributed to locating and interviewing the correct households is 95.4%.

Sampling errors on the other hand are a measure of the variability between all possible samples in a population. The degree of variability is not known exactly but can be estimated from the survey results. The KDHS 2014 sample was as a result of a multiple stage stratified design, which requires complex procedure to estimate sampling errors. Integrated System for Survey Analysis (ISSA) sampling error module software was used to calculate the sampling errors.

Often time's surveys are conducted over prolonged periods due to contextual dynamics and this makes it difficult to precisely locate the measure of fertility in time.

3.3 Data Analysis Techniques

a) Poisson Regression Model

Let Y be the number of events of a Poisson distribution.

The distribution of Y is given by:

$$P\ (Y = y) = \frac{e^{\lambda}\lambda^{y}}{y!}; y = 0,1,2,$$

λ is the rate parameter and gives the average number of events per unit period of time or space.

The mean and variance of Y are equal.

i) Using a simple model

Assume that the rate parameter is determined by a predictor.

The model is of the form:

$$Ln(\lambda) = \beta_0 + \beta_1 X$$

ln(λ) is the natural parameter of Poisson distribution.

Hence log function is the link function.

- **Interpretation of regression coefficient**

Taking the exponent of the model:

$$\lambda = e^{\beta_0 + \beta_1 X} = e^{\beta_0}\, e^{\beta_1 X}$$

- If a coefficient is positive, its transformed log value will be greater than one, meaning that the number of events of interest will increase.
- If a coefficient is negative, its transformed log value will be less than one, meaning that the number of events of interest will decrease.
- A coefficient of zero (0) has a transformed log value of 1.0, meaning that this coefficient does not change the outcome.

a) Continuous predictor

If the predictor variable is **continuous**:

Let X= *k*, (*k* is a constant); then the average number of events is:

$\lambda = e^{\beta_0} e^{\beta_1 X}$

Increasing X by one unit; i.e. X = *k + 1*, the average number of events is:

$\lambda = e^{\beta_0} e^{\beta_1 (k+1)} = e^{\beta_0} e^{\beta_1 k} e^{\beta_1}$

Thus to measure the change in response when we change the predictor by one unit; we compare the two rates using a ratio:

$$Rate\ ratio\ (RR) = \frac{\lambda\ when\ X = k+1}{\lambda\ when\ X = k} = \frac{e^{\beta_0} e^{\beta_1 k} e^{\beta_1}}{e^{\beta_0} e^{\beta_1 X}} = e^{\beta_1}$$

The transformed log value is a rate ratio.

b) Categorical predictor

If the predictor is a **categorical** variable with *m* levels:

Create *m -1* dummy variables

$$X_j = \begin{cases} 1, if\ in\ level\ j \\ 0, otherwise \end{cases}$$

When X_j = 0, then the average number of events is:

$\lambda = e^{\beta_0} e^{\beta_1 (X=0)} = e^{\beta_0}$

When Xj = 1, then the average number of events is

$\lambda = e^{\beta_0} e^{\beta_1 (X=0)} = e^{\beta_0} e^{\beta_1}$

- **Statistical inference on regression coefficients**

The 100(1- α) % C.I for β_1 is;

$\hat{\beta}_1 \pm Z_{\alpha/2} * s.e(\hat{\beta}_1)$

And the 100(1 - α) % C.I for e^{β_1} is:

$exp^{(}\hat{\beta}_1 \pm Z_{\alpha/2} * s.e(\hat{\beta}_1))$

- The hypothesis test for the significance of $\boldsymbol{\beta_1}$ is:

$H_0: \beta_1 = 0\ vs. H_1: \beta_1 \neq 0$

- The test statistic is:

$$Z = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} \sim N(0,1)$$

Alternatively test

$H_0: RR = 1 \; vs. H_1: RR \neq 1$

This test is carried out using 100(1-α)% confidence interval for e^{β_1} .

If the confidence interval includes the value 1, fail to reject H_0 and if it does not include 1 then reject H_0.

ii) <u>Multiple model</u>

The model is:

$$Ln(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \cdots + \beta_p X_p$$

Thus

$$\lambda = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \cdots + \beta_p X_p}$$

which can be used to estimate the average number of events.

- **Overall Significance of the model fit**

The test is carried out by comparing two models: null and fitted

H_0: null model is the better fit

H_1: fitted model is the better fit

Test statistic:

Likelihood ratio test statistic; its sampling distribution is chi- square distribution

Degrees of freedom: p; p is the no. of predictors

A small p-value indicates a significant fit.

- **Interpretation of partial regression coefficients**

Each regression coefficient predictor variable β_k is interpreted while holding all other predictors except X_k constant.

Let X_k be a continuous predictor variable:

When $X_k = m$;

$$\lambda = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_k m + \cdots + \beta_p X_p}$$

And when $X_k = m+1$;

$\lambda = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k m+\beta_k\ldots+\beta_p X_p}$

$$(R.R) = \frac{\lambda \; when \; X_k = m+1}{\lambda \; when \; X_k = m} = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k m+\beta_k\ldots+\beta_p X_p}}{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k m+\cdots+\beta_p X_p}} = e^{\beta_k}$$

Let X_k be a categorical predictor variable; a dummy variable is created for all levels of variable except the level chosen as reference level.

For reference level $X_k = 0$;

$\lambda = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k(0)+\cdots+\beta_p X_p}$

For other level $X_k = 1$;;

$\lambda = e^{\beta_0+\beta_1 X_1+\cdots+\beta_k(1)+\cdots+\beta_p X_p}$

$$(R.R) = \frac{\lambda \; when \; X_k = 1}{\lambda \; when \; X_k = 0} = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_k+\beta_k\ldots+\beta_p X_p}}{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}} = e^{\beta_k}$$

For the multiple Poisson regression model, the rate ratio is an **adjusted** rate ratio since we adjust (control) for all other predictors when assessing the effect of one predictor on the response variable.

- **Statistical inference on regression coefficients**

The 100(1- α) % C.I for β_k is;

$\hat{\beta}_k \pm Z_{\alpha/2} * s.e(\hat{\beta}_k)$

And the 100(1-α) % C.I for e^{β_k} is;

$[e^{\hat{\beta}_k - Z_{\frac{\alpha}{2}} * s.e(\hat{\beta}_k)} e^{\beta_k + Z_{\frac{\alpha}{2}} * s.e(\hat{\beta}_k)}]$

- The hypothesis test for the significance of β_k is

$H_0: \beta_k = 0 \; vs. H_1: \beta_k \neq 0$

- The test statistic is:

$$Z = \frac{\hat{\beta}_k}{s.e(\hat{\beta}_k)} \sim N(0,1)$$

Alternatively test:

$$\mathrm{H_0: e^{\beta_k} = 0 \; vs. H_1: e^{\beta_k} \neq 0}$$

This test is carried out using 100(1 - α) % confidence interval for e^{β_k} .

If the confidence interval includes the value 1, fail to reject H_0 and if it does not include 1 then reject H_0.

b) Negative Binomial Model

- A limitation of the Poisson distribution is the equality of its mean and variance.
- Quite often count data observed do not satisfy the condition of this equality.
- In particular, the conditional variance is larger than the conditional mean. This is termed as over-dispersion.
- Over dispersion renders the assumption of a Poisson distribution for the data untenable.
- A reasonable alternative under these circumstances is negative binomial regression. This model allows the variance to differ from the mean
- To use negative binomial regression model, the count data is assumed to follow negative binomial distribution.
- The negative binomial distribution is a two-parameter distribution. For positive integer ***y***, it is the distribution of the number of failures ***r*** that occurs in a sequence of trials before ***y*** successes have occurred, where the probability of success in each trial is ***p***.

The distribution is defined by:

$$f(y,r,p) = \binom{y+r-1}{y} p^y (1-p)^r$$

The mean and variance is given by:

$$E[Y] = \frac{pr}{1-p} \; and \; Var[Y] = \frac{pr}{(1-p)^2}$$

Let E[Y] = μ

$$\frac{pr}{1-p} = \mu \ \rightarrow p = \frac{\mu}{r+\mu}$$

And

$$Var[Y] = \frac{\mu}{r}\ (\mu + r)$$

Further if we let r = 1/α; then

$$E[Y] = \mu\ and\ Var[Y] = \mu(1 + \alpha\mu)$$

α is the dispersion parameter

$$f\left(y, \frac{1}{\alpha}, p\right) = \begin{pmatrix} y + \frac{1}{\alpha} - 1 \\ y \end{pmatrix} \left(\frac{\alpha\mu}{1+\alpha\mu}\right)^{y} \left(\frac{1}{1+\alpha\mu}\right)^{1/\alpha}$$

The coefficients have an additive effect in the log(y) scale and the IRR have a multiplicative effect in the y scale. The dispersion parameter alpha in negative binomial regression does not affect the expected counts, but it does effect the estimated variance of the expected counts.
The negative binomial distribution is a function of both mean (μ) and alpha (α); the dispersion parameter, as α → 0; the distribution becomes the Poisson distribution.
The form of the model equation for negative binomial regression is the same as that for Poisson regression.
The log of the outcome is predicted with a linear combination of predictors:

$$log(y) = \beta_0 + \beta_1 X_1 + \cdots \beta_k X_k + \cdots + \beta_p X_p$$

This implies:

$$y = e^{(\beta_0 + \beta_1 X_1 + \cdots \beta_k X_k + \cdots + \beta_p X_p)}$$

$$y = e^{\beta_0} * e^{\beta_1 X_1} * \ldots * e^{\beta_k X_k} * \ldots e^{\beta p X p}$$

- Negative binomial regression can be considered as a generalization of Poisson regression since it has the same mean structure as Poisson regression and it has an extra parameter to model the over-dispersion.

- The Poisson model is actually nested in the negative binomial model.
- To determine whether to use negative binomial regression, one can look at mean and variance values of response variables. For the negative binomial distribution, the variance is proportionally larger than the mean.
- Alternatively one uses a likelihood ratio test to compare the two models.
- The interpretation of regression coefficients is the same as that of Poisson regression models

3.3.1. Statistical methods

The study will use frequencies, descriptive statistics, scatter plots, skewness, kurtosis, cross-tabulation analysis and chi square tests, Omnibus test, normal Q-Q plots, Box plots, one sample Kolmogorov Smirnov Test and hypothesis testing to undertake both exploratory and confirmatory data analysis.

3.3.2. Hypotheses to test

The null hypotheses to be evaluated are:

There is no difference between Poisson and Negative Binomial regression models in the estimation of adolescent fertility rate.

Level of education, marital status, wealth index, frequency of listening to radio, contraceptive use, religion, type of residence and age at first sex do not influence adolescents' birth rate.

CHAPTER FOUR: DATA ANALYSIS AND RESULTS

4.1 Introduction

The data analysis process entails both exploratory data analysis and confirmatory data analysis. A comparative analysis of the two models is done to examine the goodness of fit, model effects and individual parameter estimates. Further the population parameter estimates are compared to the demographic estimates.

4.2. Exploratory Data Analysis (EDA)

This entailed performing basic analyses in SPSS through random manipulation to better understand the data and also generate some results. Data visualization was done by plotting bar graphs, scatter plots and contingency tables.

i) Case selection.

This was achieved by applying the weighting parameter then case selection based on the age cohort of interest (15-19 years). DHS data also includes: 20-24, 25-29, 30-34, 35-39, 40-44 and 45-49 age cohorts.

The result yielded a sample size (N) of 5820 weighted cases. The number of non-weighted cases was 6078. The full dataset contained a total of 31,079 cases for all the age cohorts.

ii) Types of data

a) Response variable:

The outcome variable is count data of the number of live births (804) by 5820 girls' age 15 - 19 years, three years preceding the survey. 12.6% had given birth to one child each, 1.2% to two children each and 0.1% to three children each. This information is as summarized in the table below:

Age in 5-year groups * Births in last three years Crosstabulation						
		Births in last three years				Total
		0	1	2	3	
15-19	Count	5016	732	67	5	5820
	% within Age in 5-year groups	86.2%	12.6%	1.2%	0.1%	100.0%
Total	Count	5016	732	67	5	5820
	% within Age in 5-year groups	86.2%	12.6%	1.2%	0.1%	100.0%

b) Predictors:

Eight predictors were considered for the full model fit and included seven categorical variables with different levels and one continuous variable. They include: Level of education, marital status, wealth index, frequency of listening to radio, contraceptive use, religion, type of residence and age at first sex; and were selected on the basis of theoretical and general understanding of the relationship being modeled. The categorical variables were treated as factors and the continuous variable as a covariate during data analysis.

c) Covariate:

The mean age (imputed) at sexual debut was 6.12 years with a standard deviation of 9.877 as shown in the table below:

Descriptive Statistics

	N	Mean	Std. Deviation
Age at first sex (imputed)	5814	6.12	9.877
Valid N (listwise)	5814		

iii) Categorical predictors:

The categorical variable information is summarized in the charts below:

a) Type of type of residence

68.1% and 31.9% of the respondents came from a rural and urban background respectively.

Age in 5-year groups * Type of type of residence Crosstabulation				
		Type of type of residence		Total
		Urban	Rural	
15-19	Count	1859	3961	5820
	% within Age in 5-year groups	31.9%	68.1%	100.0 %
Total	Count	1859	3961	5820
	% within Age in 5-year groups	31.9%	68.1%	100.0 %

b) Religion.

Protestants (69.9%), Catholics (20.9%) and Muslims (8.3%) constituted the majority of the respondents respectively.

Age in 5-year groups * Religion Crosstabulation							
		Religion					Total
		Roman Catholic	Protestant / Other Christian	Muslim	No religion	Other	
15-19	Count	1213	4062	485	47	6	5813
	% within Age in 5-year groups	20.9%	69.9%	8.3%	0.8%	0.1 %	100.0 %
Total	Count	1213	4062	485	47	6	5813
	% within Age in 5-year groups	20.9%	69.9%	8.3%	0.8%	0.1 %	100.0 %

c) Frequency of listening to radio

66.0% listened to radio at least once a week, 15.9% less than once a week and 18.1% not at all.

Age in 5-year groups * Frequency of listening to radio Crosstabulation

		Frequency of listening to radio			Total
		Not at all	Less than once a week	At least once a week	
15-19	Count	1053	922	3842	5817
	% within Age in 5-year groups	18.1%	15.9%	66.0%	100.0%
Total	Count	1053	922	3842	5817
	% within Age in 5-year groups	18.1%	15.9%	66.0%	100.0%

d) Contraceptive use and intention

The data shows that 10.1% were using contraceptives whereas 89.9% were non-users.

Age in 5-year groups * Contraceptive use and intention Crosstabulation

		Contraceptive use and intention				Total
		Using modern method	Using traditional method	Non-user - intends to use later	Does not intend to use	
15-19	Count	250	24	1728	715	2717
	% within Age in 5-year groups	9.2%	0.9%	63.6%	26.3%	100.0%
Total	Count	250	24	1728	715	2717
	% within Age in 5-year groups	9.2%	0.9%	63.6%	26.3%	100.0%

e) Marital status

The data indicates that 10.5% were married and 86.8% had never been in a union. This is summarized in table below:

Age in 5-year groups * Current marital status Crosstabulation								
		Current marital status						Total
		Never in union	Married	Living with partner	Widowed	Divorced	No longer living together/ separated	
15-19	Count	5052	610	85	4	20	49	5820
	% within Age in 5-year groups	86.8%	10.5%	1.5%	0.1%	0.3%	0.8%	100.0%
Total	Count	5052	610	85	4	20	49	5820
	% within Age in 5-year groups	86.8%	10.5%	1.5%	0.1%	0.3%	0.8%	100.0%

f) Wealth index

The distribution of wealth per quintiles shows that 17.9% fall under the poorest households' category, 21.0% poorer, 22.9% middle, 19.1% richer and 19.2% in the richest households' category.

Age in 5-year groups * Wealth index Crosstabulation							
		Wealth index					Total
		Poorest	Poorer	Middle	Richer	Richest	
15-19	Count	1040	1220	1331	1113	1116	5820
	% within Age in 5-year groups	17.9%	21.0%	22.9%	19.1%	19.2%	100.0%
Total	Count	1040	1220	1331	1113	1116	5820
	% within Age in 5-year groups	17.9%	21.0%	22.9%	19.1%	19.2%	100.0%

g) Highest educational level

It is expected that most of the girls age 15-19 are in primary and secondary level. The data shows that 2.3% had no education, 49.9% had primary education, 45.0% secondary education and 2.8% had higher education.

Age in 5-year groups * Highest educational level Crosstabulation						
		Highest educational level				Total
		No education	Primary	Secondary	Higher	
15-19	Count	133	2903	2619	165	5820
	% within Age in 5-year groups	2.3%	49.9%	45.0%	2.8%	100.0%
Total	Count	133	2903	2619	165	5820
	% within Age in 5-year groups	2.3%	49.9%	45.0%	2.8%	100.0%

4.3. Confirmatory Data Analysis (CDA)

a) Estimation of the Poisson regression model

i) One sample Kolmogorov Smirnov Test

The one sample K-S test is used to test whether the dependent variable – number of births follows a Poisson distribution. The computed Kolmogorov Smirnov Z value is 0.233 which is fairly close to zero. The mean (0.15) and variance (0.157) are highly similar hence a good indication that the dependent variable follows a Poisson distribution. The overall significance test (P-value = 1), means that the test is not statistically significant hence conclude that the data follows a Poisson distribution.

One-Sample Kolmogorov-Smirnov Test

		Births in last three years
N		5593
Poisson Parameter[a,b]	Mean	.15
Most Extreme Differences	Absolute	.003
	Positive	.003
	Negative	-.003
Kolmogorov-Smirnov Z		.233
Asymp. Sig. (2-tailed)		1.000
a. Test distribution is Poisson.		
b. Calculated from data.		

ii) Scatter plot

The mean predicted was plotted against the deviance residual to assess for cases that may have undue influence on both the Poisson and negative binomial regression results. The plot shows normality in the distribution is as shown below:

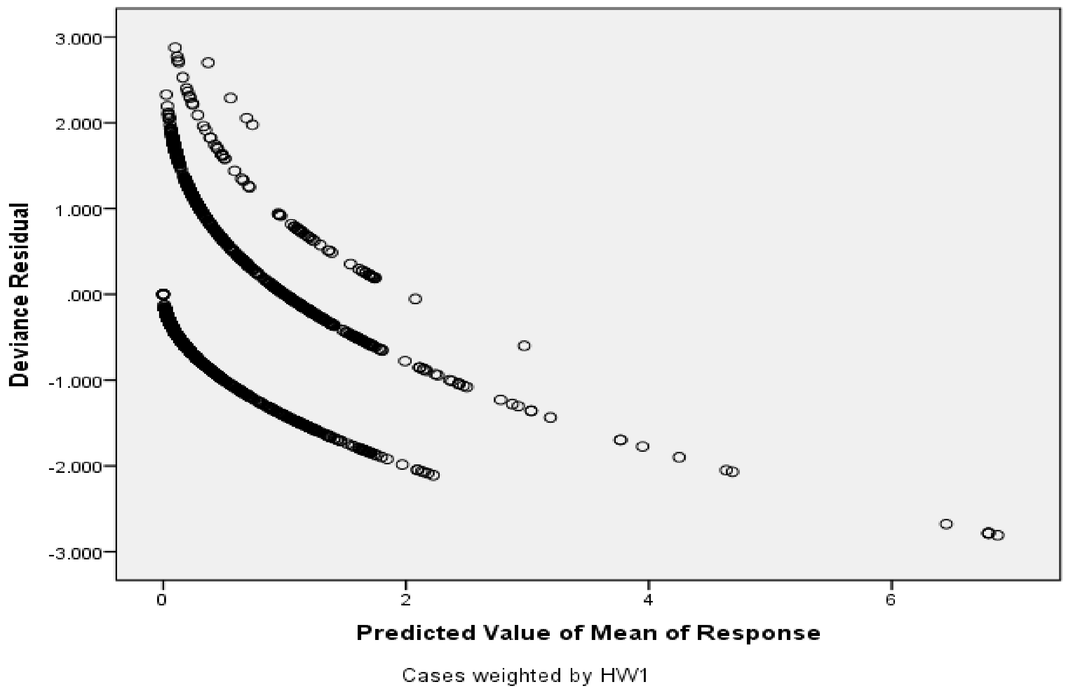

Figure 6: Scatter plot on predicted mean response and deviance residual

iii) Goodness of Fit

The computed Deviance Value/df = 0.373. This is an indication of under-dispersion in the data. An ideal value would be close to one and fall in the range of 0.9 – 1.1. The Akaike's Information Criteria (AIC) and the Bayesian Information criteria (BIC) values are 1737.542 and 1878.325 respectively.

Goodness of Fit[a]			
	Value	df	Value/df
Deviance	963.619	2583	.373
Scaled Deviance	963.619	2583	
Pearson Chi-Square	1864.174	2583	.722
Scaled Pearson Chi-Square	1864.174	2583	
Log Likelihood[b]	-844.771		
Akaike's Information Criterion (AIC)	1737.542		
Finite Sample Corrected AIC (AICC)	1738.007		
Bayesian Information Criterion (BIC)	1878.325		
Consistent AIC (CAIC)	1902.325		

iv) Omnibus Test

The hypotheses to test are:

H_0 : null model is the better fit

H_1 : fitted model is the better fit

The full predicted Poisson model represents a statistically significant (p=0.000) improvement in fit over a null model (intercept only model).

Omnibus Test[a]		
Likelihood Ratio Chi-Square	**df**	**Sig.**
611.264	23	.000

v) Statistical inference on regression coefficients

The Poisson regression analysis using the Likelihood Ratio Chi-Square test shows that the wealth index (p=0.009), contraceptive use (p=0.000), marital status (p=0.000) and age at first sex (p=0.000) were all significant predictors since their p-values are less than the stated alpha level of 0.05. On the other hand, education level (p=0.341), religion (p=0.062), frequency of listening to radio (p=0.846) and type of residence (p=0.500) were non-significant predictors. This is as summarized in the table below:

Tests of Model Effects			
Source	**Type III**		
	Likelihood Ratio Chi-Square	**df**	**Sig.**
(Intercept)	42.364	1	.000
Type of residence	0.454	1	0.500
Education Level	3.351	3	0.341
Religion	7.350	3	0.062
Frequency of listening to radio	0.333	2	0.846
Wealth index	13.458	4	0.009
Contraceptive use	24.800	3	0.000
Marital status	172.401	5	0.000
Age at first sex	25.807	1	0.000

vi) Interpretation of regression coefficients

The estimated Poisson regression coefficients for the dependent variables as well as their contribution in terms of the exponentiated data are interpreted to identify which individual predictor variables are uniquely accountable for the variation in the dependent variable. The categorical predictors are treated as factors and the dummy variables were ordered in a descending manner with the lowest coded category being treated as the reference group.

a) **Type of residence**

The dummy variable was coded into two levels representing the type of residence as either urban or rural. The urban category was treated as the reference variable.

- **Rural**

  Adjusting for education level, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in rural residence contributes to an average increase of 0.052 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in type of residence, the IRR increases by 5.4% for those residing in rural areas compared to those with no education, adjusting for level of education, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

b) **Highest level of education**

The dummy variable was coded into four levels to represent the various categories for the highest level of education variable with the ‘no education level’ being treated as the reference variable.

- **Higher education**

  Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in higher education contributes to an average decrease of -0.908 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in education level, the IRR decreases by 59.7% for those with higher education compared to those with no education, adjusting for religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Secondary education**

  Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in secondary education contributes to an average decrease of -0.346 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in education level, the IRR decreases by 29.2% for those with secondary education compared to those with no education, adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Primary education.**

  Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in primary education contributes to an average decrease of -0.235 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in education level, the IRR decreases by 20.9% for those with primary education compared to those with no education, adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

c) Religion

The dummy variable was coded into four levels to represent the various categories for religion with the 'Roman Catholic level' being treated as the reference variable.

- **No religion**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in no religion contributes to an average increase of 0.336 in the log counts of births. In terms of the exponentiated parameters, for every unit change in religious status, the IRR increases by 40.0% for those with no religious affiliation compared to those who are Roman Catholics, adjusting for education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Muslims**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in Muslims category contributes to an average decrease of -0.609 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in religious status, the IRR decreases by 45.6% for Muslims compared to those who are Roman Catholics, adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Protestants**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in Protestants category contributes to an average decrease of -0.008 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in religious status, the IRR decreases by 10.8% for Protestants compared to those who are Roman Catholics, adjusting for education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

d) Frequency of listening to radio

The dummy variable was coded into four levels to represent the various categories for the predictor variable with the ‘not at all level’ being treated as the reference variable.

- **Almost every day**

  Adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in the frequency of listening to radio almost every day contributes to an average decrease of -0.002 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in frequency of listening to radio, the IRR decreases by 0.2% for those who listened to radio almost every day compared to those who did not listen to radio at all, adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex.

- **At-least once a week**

Adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in frequency of listening to radio at-least once a week contributes to an average increase of 0.009 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in the frequency of listening to radio, the IRR increases by 0. 9% for those who listened to radio at-least once a week compared to those who did not listen to radio at all, adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex.

e) Wealth index

The dummy variable was coded into five levels to represent the various categories for wealth index variable with the 'poorest level' being treated as the reference variable.

- **Richest**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the richest quintile contributes to an average decrease of -0.836 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 56.7% for those in the richest quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Richer**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the richer quintile contributes to an average decrease of -0.038 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 31.6% for those in the richer quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Middle**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the middle quintile contributes to an average decrease of -0.358 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 30.1% for those in the middle quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Poorer**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the poorer quintile contributes to an average decrease of -0.219 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 19.7% for those in the poorer quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

f) Contraceptive use and intention

The dummy variable was coded into four levels to represent the various categories with the ‘modern method level’ being treated as the reference variable.

- **No intention to use**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those who had no intention to use contraceptives contributes to an average decrease of -1.248 in the log counts of number of births. In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 71.3% for those who had no intention to use contraceptives compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

- **Non user but intends to use**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those who were non-users but had intention to use contraceptives contributes to an average decrease of -0.898 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 59.3% for those who were non-users but had intention to use contraceptives compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

- **Traditional method**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those using traditional contraceptive methods contributes to an average decrease of -0.924 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 60.3% for respondents using traditional contraceptive methods compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

g) Marital status

The dummy variable was coded into six levels to represent the various categories with the ‘never in union level’ being treated as the reference variable.

- **Separated**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those separated contributes an average increase of 1.645 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 5.18 higher for those separated compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Divorced**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those divorced contributes to an average increase of 0.699 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 2.012 higher for those divorced compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Widowed**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those widowed contributes to an average increase of 1.837 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 6.275 higher for those widowed compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Living with partner**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those living together contributes to an average increase of 1.121 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 3.067 higher for those living with a partner compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Married**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those married contributes to an average increase of 1.703 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 5.492 higher for those married compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

h) Age at first sex

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and marital status, a predicted one year increase in age contributes to a 0.017 average increase in the log counts of number of births.

In terms of the exponentiated parameters, for every year increase in age at first sex, the IRR is 1.7% higher, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

Parameter Estimates										
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-1.237	.4156	-2.051	-.422	8.859	1	.003	.290	.129	.655
Type of residence										
[Rural=2]	.052	.1478	-.238	.342	.125	1	.724	1.054	.789	1.408
[Urban=1]	0[a]	.	.	.	.	.	.	1	.	.
Level of education										
[Higher=3]	-.908	.6145	-2.112	.297	2.181	1	.140	.403	.121	1.346
[Secondary=2]	-.346	.3582	-1.048	.356	.931	1	.334	.708	.351	1.428
[Primary=1]	-.235	.3379	-.897	.428	.482	1	.487	.791	.408	1.534
[No education=0]	0[a]	.	.	.	.	.	.	1	.	.
Religion										
[Other=96]	-25.869[b]	.	.	.	.	.	.	5.825E-012	.000	.000
[No religion=4]	.336	.3303	-.311	.984	1.037	1	.308	1.400	.733	2.674
[Muslim=3]	-.609	.2931	-1.184	-.035	4.322	1	.038	.544	.306	.966
[Protestant=2]	-.008	.1299	-.263	.247	.004	1	.951	.992	.769	1.280
[Roman Catholic=1]	0[a]	.	.	.	.	.	.	1	.	.
Frequency of listening to radio										
[At-least once-	-.002	.1386	-.274	.269	.000	1	.986	.998	.760	1.309

week=2]										
[Less than once-week=1]	.009	.1855	-.355	.372	.002	1	.963	1.009	.701	1.451
[Not at all=0]	0[a]	.	.	.	.	.	.	1	.	.
Marital status										
[Separated=5]	1.645	.2907	1.075	2.215	32.012	1	.000	5.180	2.930	9.158
[Divorced=4]	.699	.6090	-.494	1.893	1.318	1	.251	2.012	.610	6.638
[Widowed=3]	1.837	1.0627	-.246	3.919	2.987	1	.084	6.275	.782	50.366
[Living with partner=2]	1.121	.3114	.510	1.731	12.948	1	.000	3.067	1.666	5.646
[Married=1]	1.703	.1331	1.442	1.964	163.662	1	.000	5.492	4.231	7.130
[Never in union=0]	0[a]	.	.	.	.	.	.	1	.	.
Contraceptive use and intention										
[No intend to use=4]	-1.248	.1940	-1.629	-.868	41.387	1	.000	.287	.196	.420
[Non-user intends to use=3]	-.898	.1215	-1.136	-.660	54.649	1	.000	.407	.321	.517
[Traditional method=2]	-.924	.3480	-1.606	-.242	7.051	1	.008	.397	.201	.785
[Modern method=1]	0[a]	.	.	.	.	.	.	1	.	.
Wealth index										
[Richest=5]	-.836	.2397	-1.306	-.367	12.178	1	.000	.433	.271	.693
[Richer=4]	-.380	.1874	-.747	-.013	4.108	1	.043	.684	.474	.988
[Middle=3]	-.358	.1646	-.681	-.036	4.738	1	.030	.699	.506	.965
[Poorer=2]	-.219	.1592	-.531	.093	1.895	1	.169	.803	.588	1.097
[Poorest=1]	0[a]	.	.	.	.	.	.	1	.	.
Age at first sex										
Age at first sex	.017	.0030	.011	.023	31.813	1	.000	1.017	1.011	1.023

b) ESTIMATION OF THE NEGATIVE BINOMIAL REGRESSION MODEL

i) Goodness of Fit

The analysis was conducted based on an adjusted dispersion parameter of one. The Akaike's Information Criteria (AIC) and the Bayesian Information criteria (BIC) values are 1813.856 and 1886.191 respectively.

Goodness of Fit[a]			
	Value	df	Value/df
Deviance	753.019	2583	.292
Scaled Deviance	753.019	2583	
Pearson Chi-Square	1289.599	2583	.499
Scaled Pearson Chi-Square	1289.599	2583	
Log Likelihood[b]	-822.928		
Akaike's Information Criterion (AIC)	1813.856		
Finite Sample Corrected AIC (AICC)	1740.046		
Bayesian Information Criterion (BIC)	1886.191		
Consistent AIC (CAIC)	1911.191		

ii) Omnibus Test

The hypotheses to test are:

H_0 : null model is the better fit

H_1 : fitted model is the better fit

The full predicted negative binomial regression model represents a statistically significant (p=0.000) improvement in fit over a null model (intercept only model).

Omnibus Test[a]		
Likelihood Ratio Chi-Square	**df**	**Sig.**
539.373	23	.000

iii) Statistical inference on regression coefficients

The negative binomial regression analysis using the Likelihood Ratio Chi-Square test shows that the wealth index (p=0.000), contraceptive use (p=0.000), marital status (p=0.016) and age at first sex (p=0.000) were all significant predictors since their p-values are less than the stated alpha level of 0.05. On the other hand, education level (p=0.281), religion (p=0.176), frequency of listening to radio (p=0.856) and type of residence (p=0.613) are non-significant predictors. This is as summarized in the table below:

Tests of Model Effects			
Source	Type III		
	Likelihood Ratio Chi-Square	df	Sig.
(Intercept)	15.096	1	.000
Type of residence	.256	1	.613
Education Level	3.821	3	.281
Religion	4.949	3	.176
Frequency of listening to radio	.311	2	.856
Wealth index	95.690	5	.000
Contraceptive use	45.858	3	.000
Marital status	12.127	4	.016
Age at first sex	52.403	1	.000

iv) Interpretation of regression coefficients

The estimated negative binomial regression coefficients for the dependent variables as well as their contribution in terms of the exponentiated data are interpreted to identify which individual predictor variables are uniquely accountable for the variation in the dependent variable. The categorical predictors are treated as factors and the dummy variables were ordered in a descending manner with the lowest coded category being treated as the reference group.

a) Type of residence

The dummy variable was coded into two levels representing the type of residence as either urban or rural. The urban category was treated as the reference variable.

- **Rural**

Adjusting for education level, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in rural residence contributes to an average increase of 0.024 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in type of residence, the IRR increases by 2.4% for those residing in rural areas compared to those with no education, adjusting for level of education, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

b) Highest level of education

The dummy variable was coded into four levels to represent the various categories for the highest level of education variable with the ‘no education level’ being treated as the reference variable.

- **Higher education**

  Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in higher education contributes to an average decrease of -1.160 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in education level, the

IRR decreases by 69.7% for those with higher education compared to those with no education, adjusting for religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Secondary education**

Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in secondary education contributes to an average decrease of -0.465 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in education level, the IRR decreases by 27.2% for those with secondary education compared to those with no education, adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Primary education.**

Adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in primary education contributes to an average decrease of -0.338 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in education level, the IRR decreases by 28.6% for those with primary education compared to those with no education, adjusting for type of residence, religion, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

c) Religion

The dummy variable was coded into four levels to represent the various categories for religion with the ‘Roman Catholic level’ being treated as the reference variable.

- **No religion**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in no religion contributes to an average increase of 0.653 in the log counts of births.

  In terms of the exponentiated parameters, for every unit change in religious status, the IRR increases by 92.1% for those with no religious affiliation compared to those who are Roman Catholics, adjusting for education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Muslims**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in Muslims category contributes to an average decrease of -0.544 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in religious status, the IRR decreases by 41.9% for Muslims compared to those who are Roman Catholics, adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

- **Protestants**

  Adjusting for type of residence, level of education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in Protestants category contributes to an average decrease of -0.033 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in religious status, the IRR decreases by 3.3% for Protestants compared to those who are Roman Catholics, adjusting for education, frequency of listening to radio, wealth index, contraceptive use, marital status and age at first sex.

d) Frequency of listening to radio

The dummy variable was coded into four levels to represent the various categories for the predictor variable with the ‘not at all level’ being treated as the reference variable.

- **Almost every day**

  Adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in the frequency of listening to radio almost every day contributes to an average increase of 0.041 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in frequency of listening to radio, the IRR increases by 4.2% for those who listened to radio almost every day compared to those who did not listen to radio at all, adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex.

- **At-least once a week**

  Adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex, a predicted unit change in frequency of listening to radio at-least once a week contributes to an average increase of 0.010 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in the frequency of listening to radio, the IRR increases by 1.0% for those who listened to radio at-least once a week compared to those who did not listen to radio at all, adjusting for type of residence, level of education, religion, wealth index, contraceptive use, marital status and age at first sex.

e) Wealth index

The dummy variable was coded into five levels to represent the various categories for wealth index variable with the ‘poorest level’ being treated as the reference variable.

- **Richest**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the richest quintile contributes to an average decrease of -0.953 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 61.4% for those in the richest quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Richer**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the richer quintile contributes to an average decrease of -0.446 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 36% for those in the richer quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Middle**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the middle quintile contributes to an average decrease of -0.417 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 34.1% for those in the middle quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

- **Poorer**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex, a predicted unit change in income for those in the poorer quintile contributes to an average decrease of -0.238 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in wealth index, the IRR decreases by 21.2% for those in the poorer quintile compared to those in the poorest quintile, adjusting for type of residence, level of education, religion, frequency of listening to radio, contraceptive use, marital status and age at first sex.

f) Contraceptive use and intention

The dummy variable was coded into four levels to represent the various categories with the ‘modern method level’ being treated as the reference variable.

- **No intention to use**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those who had no intention to use contraceptives contributes to an average decrease of -1.283 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 72.3% for those who had no intention to use contraceptives compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

- **Non user but intends to use**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those who were non-users but had intention to use contraceptives contributes to an average decrease of -0.935 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 60.7% for those who were non-users but had intention to use contraceptives compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

- **Traditional method**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex, a predicted unit change in those using traditional contraceptive methods contributes to an average decrease of -0.988 in the log counts of number of births.

  In terms of the exponentiated parameters, for every unit change in contraceptive use, the IRR decreases by 62.8% for respondents using traditional contraceptive methods compared to those using modern contraceptive methods, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, marital status and age at first sex.

g) Marital status

The dummy variable was coded into six levels to represent the various categories with the ‘never in union level’ being treated as the reference variable.

- **Separated**

  Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those separated contributes an average increase of 1.454 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 4.282 higher for those separated compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Divorced**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those divorced contributes to an average increase of 0.540 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 1.716 higher for those divorced compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Widowed**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those widowed contributes to an average increase of 1.508 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 4.516 higher for those widowed compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Living with partner**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those living together contributes to an average increase of 0.810 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 2.248 higher for those living with a partner compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

- **Married**

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex, a predicted unit change in those married contributes to an average increase of 1.487 in the log counts of number of births.

In terms of the exponentiated parameters, for every unit change in marital status, the IRR is 4.424 higher for those married compared to those who have never been in a union, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

h) Age at first sex

Adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and marital status, a predicted one year increase in age contributes to a 0.050 average increase in the log counts of number of births.

In terms of the exponentiated parameters, for every year increase in age at first sex, the IRR is 5.1% higher, adjusting for type of residence, level of education, religion, frequency of listening to radio, wealth index, contraceptive use and age at first sex.

Negative binomial regression - Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-1.313	.5574	-2.405	-.220	5.545	1	.019	.269	.090	.802
Type of residence										
[Rural=2]	.024	.1812	-.332	.379	.017	1	.897	1.024	.718	1.460
[Urban=1]	0[a]	.	.	.	.	.	.	1	.	.

Level of education										
[Higher=3]	-1.160	.7139	-2.560	.239	2.642	1	.104	.313	.077	1.270
[Secondary=2]	-.465	.4802	-1.406	.476	.937	1	.333	.628	.245	1.610
[Primary=1]	-.338	.4598	-1.239	.564	.539	1	.463	.714	.290	1.757
[No education=0]	0^a	.	.	.	.	.	.	1	.	.
Religion										
[Other=96]	-25.741^b	.	.	.	.	.	.	6.622E-012	.000	.000
[No religion=4]	.653	.4854	-.298	1.604	1.809	1	.179	1.921	.742	4.975
[Muslim=3]	-.544	.3733	-1.275	.188	2.123	1	.145	.581	.279	1.206
[Protestant=2]	-.033	.1550	-.337	.270	.046	1	.830	.967	.714	1.310
[Roman Catholic=1]	0^a	.	.	.	.	.	.	1	.	.
Frequency of listening to radio										
[At-least once-week=2]	.041	.1727	-.298	.379	.056	1	.813	1.042	.743	1.461
[Less than once-week=1]	.010	.2247	-.430	.451	.002	1	.963	1.010	.651	1.570
[Not at all=0]	0^a	.	.	.	.	.	.	1	.	.
Marital status										
[Separated=5]	1.454	.3791	.711	2.197	14.713	1	.000	4.282	2.036	9.002
[Divorced=4]	.540	.7469	-.924	2.004	.523	1	.469	1.716	.397	7.420
[Widowed=3]	1.508	1.4947	-1.422	4.437	1.017	1	.313	4.516	.241	84.533
[Living with partner=2]	.810	.3786	.068	1.552	4.579	1	.032	2.248	1.070	4.721
[Married=1]	1.487	.1562	1.181	1.793	90.619	1	.000	4.424	3.258	6.010
[Never in union=0]	0^a	.	.	.	.	.	.	1	.	.
Contraceptive use and intention										
[No intend to use=4]	-1.283	.2282	-1.730	-.836	31.614	1	.000	.277	.177	.434
[Non-user intends to use=3]	-.935	.1543	-1.237	-.633	36.737	1	.000	.393	.290	.531
[Traditional method=2]	-.988	.5461	-2.059	.082	3.277	1	.070	.372	.128	1.085
[Modern method=1]	0^a	.	.	.	.	.	.	1	.	.

Wealth index										
[Richest=5]	-.953	.2875	-1.517	-.390	10.991	1	.001	.386	.219	.677
[Richer=4]	-.446	.2286	-.894	.002	3.811	1	.051	.640	.409	1.002
[Middle=3]	-.417	.2003	-.810	-.025	4.337	1	.037	.659	.445	.976
[Poorer=2]	-.238	.1954	-.621	.145	1.487	1	.223	.788	.537	1.156
[Poorest=1]	0[a]	.	.	.	.	.	.	1	.	.
Age at first sex										
V531	.050	.0089	.033	.067	31.511	1	.000	1.051	1.033	1.070
(Scale)	1[c]									
(Negative binomial)	1[c]									

CHAPTER V: DISCUSSION OF RESULTS

This chapter discusses the results by comparing the Poisson and negative binomial regression models for one and three year periods preceding the survey. Further the results will be compared with those obtained in the literature review.

5.1 The Mean and Variance Test

A descriptive analysis of the mean and variance of the dependent variable indicates that the mean (0.15) and variance (0.157) are highly similar but not exactly equal. Li (2016) argues that in Poisson model, the mean and variance of the dependent variable are curtained to be equal in theory. However, in practice, the variance is often different from the mean, thus leading to inefficiency of the model estimates.

5.2. Goodness of fit

The estimated Poisson regression model is a better fit compared to the negative binomial regression model since it has lower AIC (1737.542) and BIC (1878.325) values than the negative binomial regression, AIC (1813.856) and BIC (1954.639). According to Choi (2014), because of its strengths, the Poisson regression model might be a good starting point in estimating fertility, but there is need to check the validity of the results afterwards.

The deviance statistic shows under-dispersion in the data for both the Poisson (0.373) & the negative binomial regression models (0.292). However, Asiki et al. (2015) pointed out that if equidispersion holds in the data, the negative binomial regression and the Poisson regression end up the same.

5.3 Test of model effects

The Poisson regression analysis using the Likelihood Ratio Chi-Square test shows that the marital status (p=0.000), contraceptive use (p=0.000), wealth index (p=0.009) and age at first sex (p=0.000) were all significant predictors since their p-values are less than the stated alpha level of 0.05. On the other hand, type of residence (p=0.500), level of education (0.341), religion (0.062) and frequency of listening to radio (p=0.846) were non-significant predictors. The findings tend to differ with studies by Tsegaye (2010) who studied the socio-economic factors that influence women's fertility using the Poisson model. The study focused on total number of children ever born, and analyzed the data alongside 15 other variables that affect women, including age, education, marriage age, wealth index, rural or urban residence, province, occupation, contraceptives use and exposure to media like newspapers. They found that, in addition to age, all these factors also influence fertility although each has its different degree and direction of influence.

The negative binomial regression analysis shows very similar results to those of the Poisson regression model. The same dependent variables were found to be statistically significant with minute variations in the degree of influence: marital status (p=0.000), contraceptive use (p=0.000), wealth index (p=0.016) and age at first sex (p=0.000) were all significant predictors since their p-values are less than the stated alpha level of 0.05. On the other hand, type of residence (p=0.613), level of education (0.281), religion (0.176) and frequency of listening to radio (p=0.856) were non-significant predictors. This serves to validate the results obtained using the Poisson regression model (Tsegaye, 2010).

5.4 Comparative analysis of the regression coefficients (β's) and Incidence Rate Ratios (IRR)

In terms of the magnitude and direction of the parameter estimates for the dependent variables the results show very similar patterns for both Poisson and negative binomial regression models as summarized in the table below.

Predictors	Contribution to log counts **(β's)**		Contribution to log counts **IRR**	
Parameter	Poisson	Negative Binomial	Poisson	Negative Binomial
(Intercept)	-1.237	-1.313	0.290	0.269
Type of residence				
[Rural=2]	0.052	0.024	1.054	1.024
[Urban=1]	0a	0a	1	1
Level of education				
[Higher=3]	-0.908	-1.160	0.403	0.077
[Secondary=2]	-0.346	-0.465	0.708	0.245
[Primary=1]	-0.235	-0.338	0.791	0.290
[No education=0]	0a	0a	1	.
Religion				

[Other=96]	-25.869b	-25.741b	5.825E-012	6.622E-012
[No religion=4]	0.336	0.653	1.400	1.921
[Muslim=3]	-0.609	-0.544	0.544	0.581
[Protestant=2]	-0.008	-0.033	0.992	0.967
[Roman Catholic=1]	0a	0a	1	1
Frequency of listening to radio				
[At least once-week=2]	-0.002	0.041	0.998	1.042
[Less than once-week=1]	0.009	0.010	1.009	1.010
[Not at all=0]	0a	0a	1	1
Wealth index				
[Richest=5]	-0.836	-0.953	0.433	0.386
[Richer=4]	-0.380	-0.446	0.684	0.640
[Middle=3]	-0.358	-0.417	0.699	0.659
[Poorer=2]	-0.219	-0.238	0.803	0.788
[Poorest=1]	0a	0a	1	1
Contraceptive use and intent				
[No intend to use=4]	-1.248	-1.283	0.287	0.177
[Non-user intends to use=3]	-0.898	-0.935	0.407	0.290
[Traditional method=2]	-0.924	-0.988	0.397	0.128
[Modern method=1]	0a	0a	1	.
Marital status				
[Separated=5]	1.645	1.454	5.180	4.282
[Divorced=4]	0.699	0.540	2.012	1.716
[Widowed=3]	1.837	1.508	6.275	4.516
[Living with partner=2]	1.121	0.810	3.067	2.248
[Married=1]	1.703	1.487	5.492	4.424
[Never in union=0]	0a	0a	1	1
Age at first sex				
Age at first sex	0.017	0.050	1.017	1.051
		1c		

5.5 Estimation of the Age Specific Fertility Rate

According to the KDHS 2014, measures of current fertility were computed based on a three-year reference period (2011-2014) prior to the survey. The estimated national ASFR was 96 births per 1,000 women age 15 – 19 years. A three year reference period provides for collection of the most current fertility information as well as a statistically adequate sample to calculate the fertility rates.

Poisson and negative binomial regression were used to predict the ASFR for an adolescent girl age 15-19 years, using the average age at first sex (6.12). Other socio-economic background characteristics include: with primary level education, Catholic, listens to radio at-least once a week, from a poorer household, rural residence, non-user with intention to use contraceptives and was never in a union.

- The Poisson regression model is:

$$Ln(\lambda) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \cdots + \beta_p X_p$$

Thus

$$\lambda = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \cdots + \beta_p X_p}$$

$$\lambda = e^{\{-1.237 + -0.052 + 0 + -0.002 + -0.219 + 0.052 + -0.898 + 0 + (0.017 x 6.12)\}}$$

$$\lambda = e^{\{-2.435\}}$$

ASFR = (0.088*1000) = 88

- The negative binomial model is:

The form of the model equation for negative binomial regression is the same as that for Poisson regression. The log of the outcome is predicted with a linear combination of predictors:

$$log(y) = \beta_0 + \beta_1 X_1 + \cdots \beta_k X_k + \cdots + \beta_p X_p$$

This implies:

$$y = e^{(\beta 0 + \beta 1 X 1 + \cdots + \beta k X k + \cdots + \beta p X p)}$$

$$y = e^{\beta 0} * e^{\beta 1 X1} * \ldots * e^{\beta k X k} * \ldots e^{\beta p X p}$$

$$y = e^{\{-1.313+-0.338+0+0.041+-0.238+-0.935+0+(0.05*6.12)\}}$$

$$y = e^{\{-2.453\}}$$

ASFR = (0.086*1000) = 86

The estimated ASFR using the full Poisson regression model is 88 births while the estimated ASFR using the full negative binomial regression model is 86 births. Both estimates are nearly the same and adequately estimate the national estimated ASFR (96 births) based on the socio-economic variables at play. The advantage of using these models is that they offer the flexibility to include and exclude several covariates into the model as well as measuring their contribution to the ASFR and predicting future ASFR rates based on varying socio-economic characteristics.

CHAPTER VI: CONCLUSION

6.1 Summary

Part of Kenya's population growth is tied to relatively young population. Today, 75% of Kenyans are below 30 years, while 45% of total population is below 15 years. It is estimated that in the next two decades, Kenyan youth will have reached 24 million. For these reasons part of the solution to socio-economic and sustainable environmental development is to drastically reduce the current level of population growth rate. Female higher education has emerged as one of the factors which could accelerate this process. Equally important are the late age at marriage, rapid urbanization, increased employment, improved health care services including increased use of modern contraceptive methods. The government could therefore achieve this by encouraging the new devolved system of governance to implement these processes.

6.2. RECOMMENDATIONS

6.2.1. Policy

75% of Kenyans are below 30 years, while 45% of total population is below 15 years. It is estimated that in the next two decades, Kenyan youth will have reached 24 million. For these reasons part of the solution to socio-economic and sustainable environmental development is to drastically reduce the current level of population growth rate while accelerating efforts to harness the demographic dividend. Female higher education has emerged as one of the factors which could accelerate this process. Equally important are the late age at marriage, increased employment, improved health care services including increased use of modern contraceptive methods. To achieve significant strides, there is need to:

1. Track the implementation and monitoring of relevant policies and guidelines which promote access to comprehensive sexual and reproductive health information and services by young people especially adolescents. These include but not limited to the Adolescent Reproductive Health and Development Policy, National School Health Policy and the National Guidelines for Provision of Youth Friendly Services in Kenya (2005). The Kenyan government has achieved significant progress in addressing the adolescents sexual and

reproductive health rights and needs of young people but still a concerted effort and coordination is required. According to the 2008/09 KDHS, national estimates show that 47% of the births to young people age 15 to 19 and 40% of the births to young people age 20 to 24 years were not planned. Education programmes and the use of innovative approaches that reach out to young people with sexual and reproductive health information and services are critical to address young people's diverse needs. Even though ASFR is hardly used as an outcome measure in evaluating family planning programs due to the level of resources needed to collect the data, it is a variable that program administrators and policy makers track over time as a macro-level indicator of program effectiveness combined with non-program influences.

2. Strengthen investments in evidence generation by researchers and utilization in decision making by the policy makers both at the national and devolved levels of government. The ASFR should be of particular interest in countries or counties with adolescent reproductive health interventions designed to reduce unintended pregnancy. In Kenya with the devolved system of governance and the health ministry, the modeling of fertility differentials should be enhanced since Counties can budget for context specific surveys. Unlike the crude birth rate, the ASFR is unaffected by differences or changes in population age composition, and thus is more useful in comparing different populations or sub-groups and in measuring changes over time. This is useful to generate spatially-grained analyses based on context specific issues hence leading to better targeted programmes, legislations and policies.

6.2.2. Further Research

The following key issues should be explored further in regard to further research.

- The findings of this study point to the need for additional research on modeling ASFR as an intermediate computation in the derivation of Total Fertility Rates (TFR). Hence the application of Poisson and negative binomial regression analysis can be extended to estimation of Total Fertility Rates and trends over time using DHS or Census data.
- There is need for deeper analysis and comparison of the results obtained using other improved regression analysis models such as Quasi Poisson, tilded-Poisson and beta-binomial regression models etc. A critical analysis of data characteristics i.e. over-dispersed or under-dispersed as well as the contributing risk factors e.g. rising age of marriage should also inform the choice of estimation methods.

REFERENCES

1. Ahmed, H. M. M., & Ali, H. M. H. Using Count Regression Models To Determinate Factors Influencing Fertility Of Sudanese Women. Available at: https://www.researchgate.net/publication/309010918_using_count_regression_models_to_determinate_factors_influencing_fertility_of_sudanese_women
2. Alaba, O. O., Olubusoye, O. E., & 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.
3. Asiki, G. et al. (2015). Adverse pregnancy outcomes in rural Uganda (1996–2013): trends and associated factors from serial cross-sectional surveys, BMC Pregnancy and Childbirth (2015) 15:279
4. Brass W. 1964. Uses of census or survey data for the estimation of vital rates. Paper prepared for the African Seminar on Vital Statistics, Addis Ababa 14-19 December. Document No. E/CN.14/CAS.4/V57. New York: United Nations.
5. Brass, William. 1964. “Uses of Census or Survey Data for the Estimation of Vital Rates.” Paper presented to the African Seminar on Vital Statistics, Adis Ababa, December 1964.
6. Choi, K. (2014). Fertility in the context of Mexican migration to the United States: A case for incorporating the pre-migration fertility of immigrants. Demographic research, 130(24), pp. 703-737.
7. Christensen, Ronald (1997). Log-linear models and logistic regression. Springer Texts in Statistics (Second ed.). New York: Springer-Verlag.
8. Cleland J. 1996.” Demographic data collection in less developed countries”, Population Studies 50(3):433-450. Doi: 10.1080/0032472031000149556
9. Coale, Ansley, and Paul Demeny. 1968. Methods for Evaluating basic Demographic Measures from Limited and Defective Data. New York: United Nations.
10. DeNavas, C. (1989). The Hispanic population in the United States: March 1988. Washington, D.C.: U.S. Dept of Commerce, Bureau of the Census.

11. Gardner, W., Mulvey, E.P. and Shaw, E.C. (1995) Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological Bulletin, 118(3), 392-404.
12. Gayawan, E., Adebayo, S. B., Ipinyomi, R. A., & Oyejola, B. A. (2010). Modeling fertility curves in Africa. Demographic Research, 22, 211-236. doi:10.4054/demres.2010.22.10
13. Gouriéroux, Christian (2000). "The Econometrics of Discrete Positive Variables: the Poisson Model". Econometrics of Qualitative Dependent Variables. New York: Cambridge University Press. pp.270–83.
14. Greene, William H.(2008). "Models for Event Counts and Duration".Econometric Analysis(8th ed.). Upper Saddle River: Prentice Hall. pp. 906–944.
15. Haupt, A., & Kane, T. T. (1998). The Population handbook, 4th International Edition. Washington, D.C.: Population Reference Bureau.
16. Hilbe, J. M. (2007).Negative Binomial Regression. Cambridge University Press.
17. https://www.encyclopedia.com/social-sciences/encyclopedias-almanacs-transcripts-and-maps/estimation-methods-demographic
18. https://www.measureevaluation.org/prh/rh_indicators/family-planning/fertility/age-specific-fertility-rates
19. IYD 2014 - Improving sexual and reproductive health and rights for young adolescents in Kenya. (2016, October 26). Retrieved from http://www.dsw.org/en/2014/08/iyd-2014-sexual-reproductive-health-rights-young-adolescents-kenya/
20. Johns Hopkins Bloomberg school of public health (2017) Age-Specific Fertility Rates and the Total Fertility Rate. Available at: https://www.k4health.org/sites/default/files/Total%20Fertility%20Rate%20and%20Age-Specific%20Fertility%20Rate.pdf
21. Jones, Andrew M.;et al. (2013). "Models for count data".Applied Health Economics. London: Routledge. pp. 295– 341.
22. Li, L. (2016). Recent Fertility and Its Proximate Determinants In 3 ASEAN Countries. Thesis paper.
23. Mishra, R. et al. (2017) A Model for Age-Specific Fertility Rate Pattern of India Using Skew-Logistic Distribution Function. *American Journal of Theoretical and Applied Statistics.* Vol. 6, No. 1, 2017, pp. 32-37

24. Moultrie TA, RE Dorrington, AG Hill, K Hill, IM Timaeus and B Zaba (eds). 2013. Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population. Available at: http://demographicestimation.iussp.org/
25. Opiyo, P. (1987) Intercensal fertility estimation in Kenya. Thesis paper
26. Population Index – Volume 57 – Number 1. Retrieved from: http://popindex.princeton.edu/browse/v57/n1/f.html
27. Rutstein S and G Rojas. 2003. Guide to DHS Statistics. Calverton, MD:ORC Macro.
28. Schoumaker B. 2010. “Reconstructing fertility trends in sub-Saharan Africa by combining multiple surveys affected by data quality problems”. Paper presented at Population Association of America 2010 Annual meeting. Dallas, TX, April 15-17, 2010.
29. Schoumaker B. 2011. “Omissions of births in DHS birth histories in sub-Saharan Africa: Measurement and determinants”, Papre presented at Population Association of America 2011 Annual Meeting. Washington DC, March 31-April 2, 2011.
30. Schoumaker B. 2013. “A stat module for computing fertility rates and TFRs from birth histories: tfr2”, Demographic Research 28(Article 38): 1093-1144. Doi: 10.4054/DemRes.2013.28.38
31. Sloggett A, W Brass, SM Eldridge, IM Timaeus, P Ward and B Zaba (eds). 1994. Estimation of Demographic Parameters from Census Data. Tokyo: Statistical Institute for Asia and the Pacific.
32. The Zambia Demographic and Health Survey, (2007): key findings. Calverton, Maryland, USA: CSO and Macro International Inc.
33. Tsegaye T. (2010). Cross Sectional Study of Women Employment and Fertility in Ethiopia. Master’s Thesis, Department of Economic History, Lund University
34. UN Population Division. 1983. Manual X: Indirect Techniques for Demographic Estimation. New York: United Nations, department of Economic and Social Affairs, ST/ESA/SER.A/81.http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html
35. United Nations. 1983. Manual X: Indirect Techniques for Demographic Estimation. New York: United Nations.

36. Winkelmann, R., & Zimmermann, K. F. (1994). Count data models for demographic data∗. Mathematical Population Studies, 4(3), 205-221. doi:10.1080/08898489409525374

37. Winkelmann, R., & Zimmermann, K. F. (1995). Recent Developments In Count Data Modelling: Theory And Application. Journal of Economic Surveys, 9(1), 1-24. doi:10.1111/j.1467-6419.1995.tb00108.x *Texts, 2014*.