

DETERMINANTS OF LOW BIRTHWEIGHT IN GHANA

BY

APPIAH KUBI FELIX JUNIOR

(10638150)

**THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF
MASTER OF PHILOSOPHY STATISTICS DEGREE**

INTEGRI PROCEDAMUS
INTEGRI PROCEDAMUS

JULY 2019

DECLARATION

I, hereby declare that this submission is my own work towards the award of Master of Philosophy and that, to the best of my knowledge, it contains no materials previously published by another person nor material which has been accepted for the award of any other degree of the University, except where due acknowledgment has been made in the text.

Signature Date

Appiah Kubi Felix Junior
(Student)

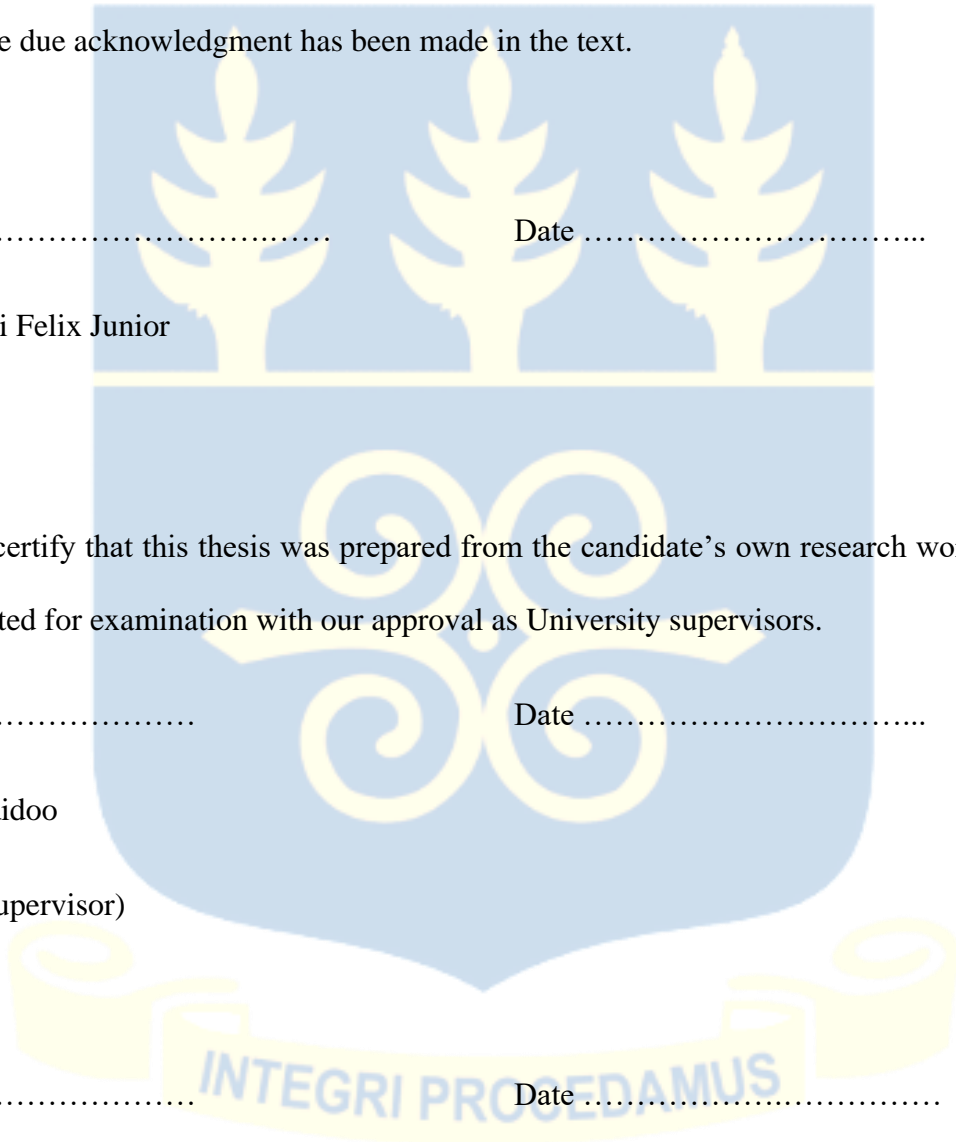
We hereby certify that this thesis was prepared from the candidate's own research work and has been submitted for examination with our approval as University supervisors.

Signature Date

Dr. Isaac Baidoo
(Principal Supervisor)

Signature Date

Dr. Felix O. Mettle
(Co-Supervisor)



ABSTRACT

The study objectives are to determine the key factors that may be accounting for birthweight of infants, estimate its distribution, prevalence and proposes a statistical model that can be useful in epidemiological studies in Ghana. Both estimation (Linear) and classification (Logistic) analysis was conducted using data from GDHS. Covariates found to statistically associated with newborn's weight are gender, size of child at birth, region, wealth index, birth order, total number children ever had, delivery place, preceding birth interval, and birth type. Size of child (Very Large) was the largest contributor to the explained variation in birthweight in the Multiple Linear regression. The Multivariable Logistic regression, for every one-unit change in mothers height, the log odds of low birthweight (against normal birthweight) of a child decreases by -0.107kg with an odds ratio of 0.898kg and for a unit decrease in birth order of the infant, the log odds of being low weight decreases by -0.714kg with an odds ratio of 0.490kg. Region depicted that (Brong Ahafo, Central, Greater Accra, Upper East and Volta) was found to be significantly related to newborn's weight, implying that a mother from any of these regions is more likely to have a NBW as compared to a mother from Ashanti. A male child was 0.397 times more likely to be NBW as compared to a female child. Single birth babies were 0.001 times more likely to be NBW as compared to multiple births. Only mothers from the poorest class was found to be significantly related to birthweight which implied that the newborns whose mothers are under the poorest class would 1.973 times more likely to be NBW as compared to mothers belonging to the middle class. Size of a child at birth was found to be significantly related to birthweight of a newborn. The study has contributed to the understanding of maternal determinants associated with infant birthweight at the population level. Findings have therefore, provided a starting point towards identifying risk factors and providing clues to health service providers on maternal determinants and birth outcomes factor to concentrate health promotion messages on.

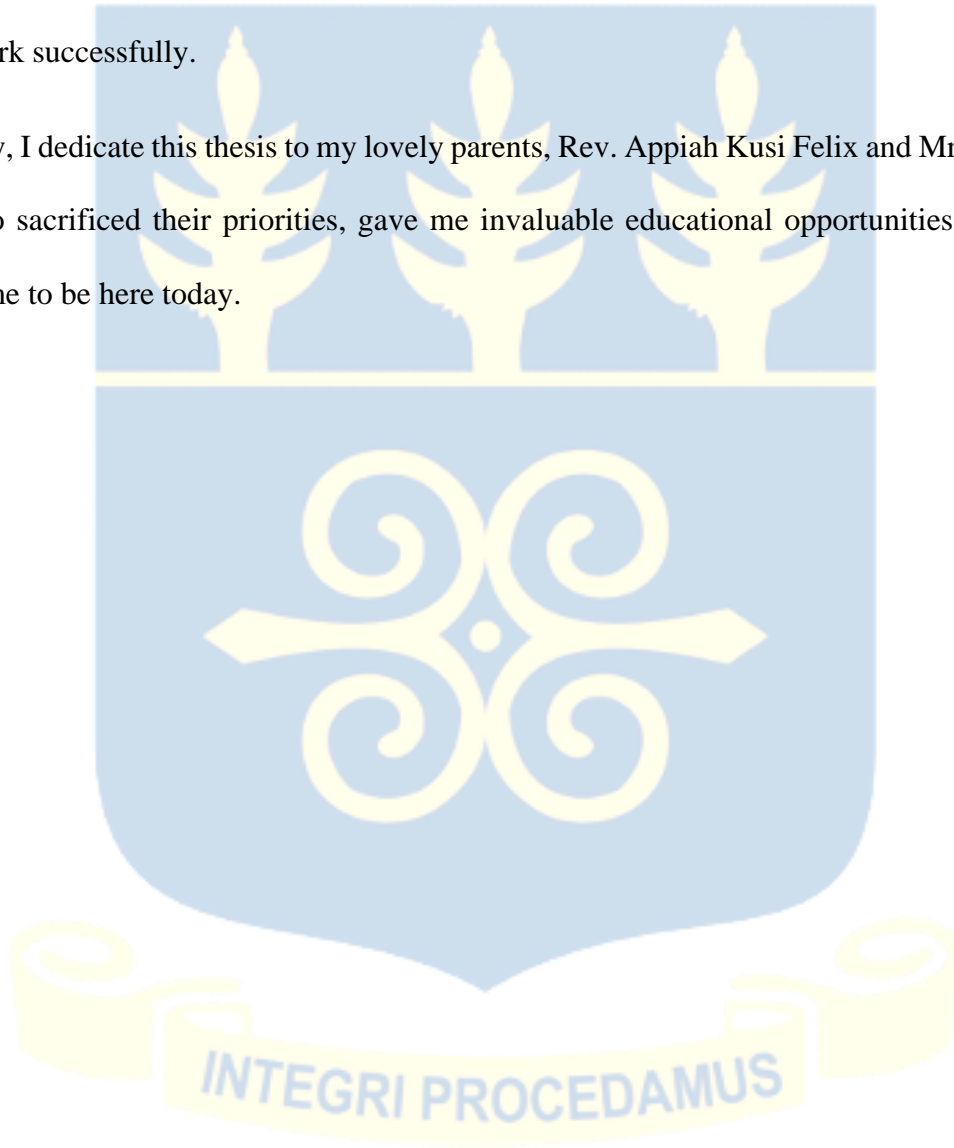
DEDICATION

I will give thanks to you, LORD, with all my heart; I will tell of all your wonderful deeds.

(Psalm 9:1)

This thesis is dedicated to the Almighty God, who led me through it all to complete my course and research work successfully.

Additionally, I dedicate this thesis to my lovely parents, Rev. Appiah Kusi Felix and Mrs. Dorothy Appiah who sacrificed their priorities, gave me invaluable educational opportunities and made efforts for me to be here today.



ACKNOWLEDGMENT

A study on the **Determinants of Low Birthweight in Ghana** was undertaken with an objective to analyze the factors affecting birthweight of a newborn. Many people helped me in every single imaginable approach to finish this exploration work to whom I wish to offer my thanks.

To start with, I might want to thank the all-powerful God for giving me the opportunity to pursue my graduate study at the Department of Statistics, University of Ghana.

I owe the most profound appreciation to Dr. Isaac Baidoo and Dr. Felix O. Mettle, my thesis advisors who assisted me throughout the dissertation process. I fumbled in many ways during the early stages but their rapid response and timely feedback, even during odd hours, was just extraordinary. I am truly amazed by the energy and the quality of the guidance they provided. Their stimulating support helped me in preparing my thesis within the timeframe.

My deepest thanks go to my parents, Rev. Appiah Kusi Felix and Mrs. Dorothy Appiah who has bolstered me right to fulfill my dream.

I also extend my gratitude to my friend Bright Antwi Boasiako, for playing a pivotal role in shaping my research methodology and data analysis. I am truly appreciative for liberally giving your opportunity to look at my research work, regardless of how bustling you were. Moreover, to Rachael Adalevo for her profitable, productive remarks and supportive gestures all through my study.

Finally, I extend my gratitude to Lecturers and Colleagues of the Department of Statistics and Actuarial Science, University of Ghana who contributed in many ways behind the scenes.

*Once again, to God the giver of life and the wherewithal to dream of and execute the program,
be the Glory.*

TABLE OF CONTENT

DECLARATION	ii
ABSTRACT	iii
DEDICATION	iv
ACKNOWLEDGMENT	v
TABLE OF CONTENT	vi
LIST OF TABLES	ix
LIST OF FIGURES	xii
LIST OF ABBREVIATIONS	xiii
CHAPTER ONE	1
INTRODUCTION	1
1.0 Introduction	1
1.1 Background of Study	1
1.2 Statement of Problem	2
1.3 Research Objectives	3
1.3.1 General Objective	3
1.3.2 Specific Objectives	3
1.4 Research Questions	3
1.5 Scope of the Study	4
1.6 Rationale for the current study	5
1.7 Study significance	5
1.8 Organization of the thesis	6
CHAPTER TWO	8
LITERATURE REVIEW	8
2.0 Introduction	8
2.1 Overview of Birthweight	8
2.2 Classification of birthweight	8
2.3 Consequences of Low birthweight	8
2.4 Determinants of Birthweight	9
2.4.1 Normal factors	10
2.4.2 Intermediate factors	10
2.4.3 Non-Normal factors	14

2.5 Related Literature Reviews outside Ghana	17
2.6 Related Literature Reviews in Ghana	18
2.8 Conclusion	20
CHAPTER THREE	21
METHODOLOGY	21
3.0 Introduction.....	21
3.1 Data description	21
3.1.1 Type and Source of Data	21
3.1.2 Sampling and Sampling Procedures	21
3.1.3 Instrumentation and Operationalization	22
3.1.4 Our Study assumptions	23
3.2 Application of Statistical Methods.....	24
3.2.1 Descriptive Analysis.....	24
3.2.2 Cross-tabulation for birthweight and birth size (2 x 2 Contingency table)	25
3.2.3 Kappa Statistics and Fleiss.....	27
3.2.4 The Chi-square χ^2 test.....	28
3.3 Statistical Model Building	28
3.3.1 Our study model building.....	29
3.3.2 Linear Regression Model (Multiple)	29
3.3.3 Logistic Regression Model.....	36
3.3.4 Variable Selection for Model Building.....	43
3.3.5 Developing the Screening tool.....	44
3.4 Conceptual framework.....	44
CHAPTER FOUR	46
DATA ANALYSIS AND DISCUSSION OF RESULTS	46
4.0 Introduction.....	46
4.1 Birthweight Data Analysis.....	46
4.1.1 Birthweight Distribution	47
4.1.2 Classification of Birthweight.....	48
4.2 Prevalence rate of Low birthweight (LBW)	48
4.2.1 WHO's Recommendation.....	48
4.2.2 Researcher's Adjusted Prevalence	48
4.3 Mother's Reporting of Birth Size Reliability Analysis	49

4.3.1 How accurate is a mother at reporting child's birthweight?	49
4.3.2 Assessment of Birthweight verse Size of an infant at birth	50
4.3.3 Calculation of Kappa Statistics.....	51
4.3.4 Reporting of Birthweight: (Mother's Memory Recall or Health Card)	52
4.4.1 Factor One: Outcome of Pregnancy	53
4.4.2 Factor Two: Socio-Economic and Demographic Factors.....	60
4.4.3 Factor Three: Maternal Anthropometry	74
4.4.4 Factor Four: Maternal Reproductive Factors.....	76
4.5 Multivariate Analysis of Predictors and Birthweight.....	89
4.5.1 Towards Building a Multiple regression model	89
4.5.2 Towards Building a Logistic Regression Model.....	104
CHAPTER FIVE	122
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	122
5.0 Introduction.....	122
5.1 Summary.....	122
5.1.1 Key Observations from Study Objectives.....	123
5.1.2 Final Comments on Multiple Linear Regression fit.....	124
5.1.3 Final Comments on Logistic Regression fit	126
5.1.4 Comparison of Two Sets of Fits.....	127
5.2 Conclusions.....	127
5.3 Recommendations	128
REFERENCES.....	130
APPENDICES	135

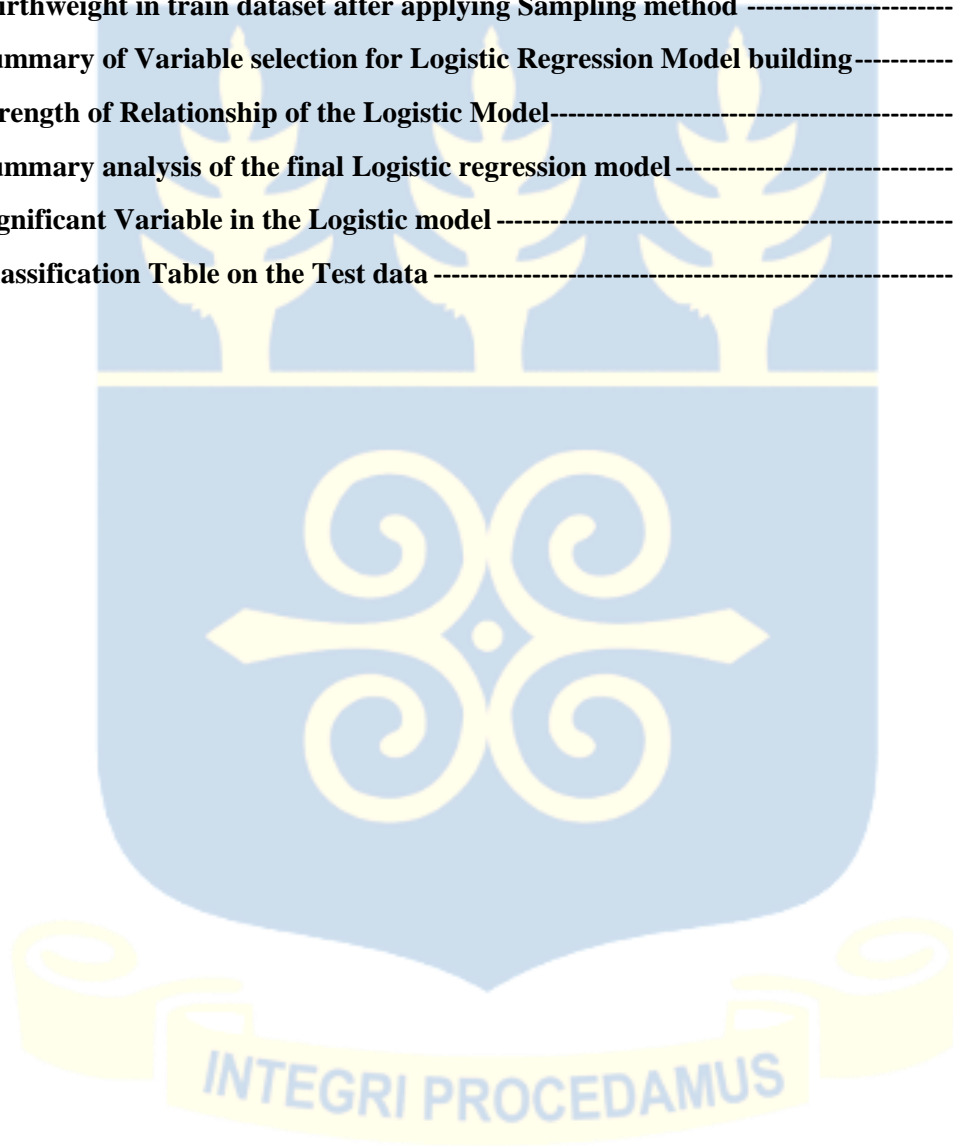


LIST OF TABLES

Table 1: Variables with the level of Measurement	23
Table 2: Standard 2×2 Table contingency Table	25
Table 3: Interpretation of Cohen's kappa & Fleiss Agreement	27
Table 4: ANOVA Table for Linear Regression (Multiple)	33
Table 5: Logit transformation	40
Table 6: Descriptive Statistics for Birthweight	47
Table 7: Descriptive Statistics of birthweight-based classification	48
Table 8: Birthweight groups against Size of child Crosstabulation	49
Table 9: Kappa Statistics for the contingency table	52
Table 10: Tabulation of Gender of child	53
Table 11: Descriptive Analysis for Gender of a child across birthweight classifications	54
Table 12: Chi-Square Tests for Gender of a child against birthweight classification	55
Table 13: Symmetric Measures for Gender of child and birthweight classification	56
Table 14: Tabulation of type of delivery by Caesarean	56
Table 15: Tabulation of type of delivery by Caesarean	57
Table 16: Chi-square tabulation for Type of delivery by cesarean and birthweight	57
Table 17: Descriptive analysis of Size of the child	58
Table 18: Chi-Square Tests for Size of child and birthweight	59
Table 19: Symmetric Measures for Size of child and birthweight classification	59
Table 20: Descriptive Statistics of Mother Age	60
Table 21: Descriptive statistics of Mothers Age by group and Birthweight	61
Table 22: Descriptive and Crosstabulation for Mothers Age and Birthweight classification	61
Table 23: Chi-Square Tests for Mothers age and birthweight	62
Table 24: Mean birthweight across Maternal Educational level	63
Table 25: Crosstab for mother education and birthweight	63
Table 26: Chi-Square tests between Mother education and birthweight	64
Table 27: Mean birthweight across various Regions	65
Table 28: Region and Birthweight Crosstabulation	65
Table 29: Chi-Square Tests across regions and birthweight	67
Table 30: Symmetric Measures for regions and birthweight	67

Table 31: Tabulation of Type of residence -----	68
Table 32: Crosstab of Type of residence and birthweight -----	68
Table 33: Chi-Square Tests type of residence and birthweight -----	69
Table 34: Religion and Birthweight Crosstabulation -----	70
Table 35: Chi-Square Tests for mother recode religion and birthweight -----	71
Table 36: Symmetric Measures of Religion and birthweight -----	71
Table 37: Mean weight of infants across the Wealth Index -----	72
Table 38: Crosstabulation Wealth index and Birthweight group -----	73
Table 39: Chi-Square Tests for Wealth index and Birthweight -----	74
Table 40: Symmetric Measures for Wealth index and birthweight -----	74
Table 41: Descriptive Statistics for Mothers Height -----	75
Table 42: Descriptive Statistics for Mother weight -----	76
Table 43: Crosstab for Birth order (Parity) and birthweight -----	77
Table 44: Chi-Square Tests for Birth order (Parity) and birthweight -----	77
Table 45: Symmetric Measures for Birth order (Parity) and birthweight -----	78
Table 46: Tabulation of Total children ever had against Birthweight classification -----	78
Table 47: Crosstabulation of Total Children ever born and Birthweight group -----	80
Table 48: Chi-Square Tests for total children ever born and birthweight -----	80
Table 49: Symmetric Measures for total children ever had/born and birthweight -----	81
Table 50: Mean Tabulation of delivery place -----	81
Table 51: Delivery place and Birthweight Crosstabulation -----	82
Table 52: Chi-Square Tests for Delivery place and Birthweight -----	83
Table 53: Symmetric Measures for delivery place and birthweight -----	83
Table 54: Descriptive Statistics for Preceding birth interval (Months) -----	84
Table 55: Tabulation of Preceding birth interval and birthweight -----	85
Table 56: Crosstab of Preceding birth interval and Birthweight -----	85
Table 57: Chi-Square Tests for Preceding birth interval and birthweight -----	86
Table 58: Symmetric Measures for Preceding birth interval and birthweight -----	86
Table 59: Descriptive statistics for Birth type and Birthweight -----	87
Table 60: Birth type and Birthweight group Crosstabulation -----	87
Table 61: Chi-Square Tests for Birth type and birthweight -----	88
Table 62: Symmetric Measures for Birth type and Birthweight -----	88

Table 63: Variable selection for Multiple Regression Model building -----	93
Table 64: Summary of Model: OLS Backward Selection -----	94
Table 65: Summary after resolving the assumptions behind our fitted model -----	101
Table 66: ANOVA Table for Multiple Linear Regression -----	103
Table 67: Classification Table on full low weight data -----	105
Table 68: Birthweight in train dataset after applying Sampling method -----	106
Table 69: Summary of Variable selection for Logistic Regression Model building -----	107
Table 70: Strength of Relationship of the Logistic Model -----	113
Table 71: Summary analysis of the final Logistic regression model -----	114
Table 72: Significant Variable in the Logistic model -----	117
Table 73: Classification Table on the Test data -----	121



LIST OF FIGURES

Figure 1: Data Cleaning Procedure -----	22
Figure 2: Conceptual framework for studying determinants of Birthweight -----	45
Figure 3: Histogram of Birthweight distribution (kg) -----	47
Figure 4: Birthweight classification Bar plot according to gender -----	54
Figure 5: Birthweight histogram across gender -----	55
Figure 6: Age distribution of mothers -----	60
Figure 7: Birthweight distribution across various Regions -----	66
Figure 8: Distribution of birthweight across region and type of residence -----	70
Figure 9: Simple Scatter with Fit Line of Birthweight by Mothers height -----	75
Figure 10: Simple Scatter with Fit Line of Birthweight by Mothers weight -----	76
Figure 11: Bar Plot of Total Children ever born/had -----	79
Figure 12: Boxplot of Birth and Total children ever born -----	79
Figure 13: Preceding Birth Interval in Months -----	84
Figure 14: QQ-Plot for Longitudinal birthweight (kg) -----	90
Figure 15: Fitted Verse Residuals for the Multiple Linear Model fit -----	96
Figure 16: ROC Curve for our Logistic Model -----	111
Figure 17: Cooks distance for the Logistic Model -----	112

LIST OF ABBREVIATIONS

ANC	Antenatal Care
ANOVA	Analysis of Variance
AOR	Adjusted Odds Ratio
CEB	Children Ever Born
CI	Confidence Interval
CM	Centimeter
GDHS	Ghana Demographic and Health Survey
GMHS	Ghana Maternal Health Survey
GM	Gram
KG	Kilogram
LRT	Likelihood Ratio Test
MICS	Multiple Indicator Cluster Survey
NBW	Normal Birthweight
OLS	Ordinary Least Squares
OR	Odds Ratio
REG	Region
Std.	Standard Deviation
SOC	Size of Child
UNICEF	United Nations International Children's Emergency Fund
WHO	World Health Organization



CHAPTER ONE

INTRODUCTION

1.0 Introduction

At the point when a newborn enters the world, the question of survival of life comes to mind. When one thinks about the survival of human life, then health becomes the first indicator. Numerous efforts, plans, and interventions are arranged by the world to save the life of a child. In recent years, the World Health Organization (WHO) has adapted to the slogan “Children’s health is tomorrow’s wealth”.

1.1 Background of Study

Birthweight is a significant determinant of a newborn’s exposure to the risk of diseases and his/her odds of survival. The World Summit has selected the incidence of birthweight as an essential indicator for observing and monitoring major health goals for Children (UN, 2002). Birthweight is a projected (plan) characteristic of future development, improvement and a reassessment (review) determinant of maternal nutrition and health status. Globally, birthweight is a good summary measurement for the heterogeneous public health difficulty that reflects the undernourishment of mothers and poor pregnancy healthcare.

Birthweight, therefore, is a noteworthy indicator of mortality, morbidity and infancy and childhood disability (WHO, 2010). In addition, to be an integral endpoint, birthweight is affiliated with several unfavorable health outcomes in childhood and adulthood. The challenge of birthweight is multidimensional, and it needs a coordinated methodology to incorporate socio-economic, medical and educational measures to address the issues. However, to recognize the factors influencing birthweight has been the paramount challenge in the public health sector.

1.2 Statement of Problem

The World Health Organization characterizes low weight of infant as weight less than 2.5kg (2500g) at birth irrespective of community, region, culture or gestational age with measurement following birth, preferably between the first hour of life before critical postnatal weight reduction happens. This cut off is utilized worldwide and depends on epidemiological observations of newborn babies. About 20.5 million babies, an estimated 14.6 percent of all newborns all around the world that year, suffered from low birthweight. (UNICEF, 2015).

As indicated by the World Bank collection of development indicators, low birthweight babies in Ghana was 10.70% as of 2011. Its highest value over the past 18 years was 16.10% in 2003, while its lowest value was 9.10% in 2006. Notwithstanding, the latest recent WHO data published in 2017, low birthweight deaths in Ghana was 8224, approximately (3.91%) of total deaths. Low birthweight is positioned seventh among the best 20 causes for mortality in Ghana and 29th on the world (WHO, 2017).

The low weight of neonates is a demanding public health menace in Ghana. Its criticality originates because infant morbidity is an extreme determinant that contributes generously to the comprehensive handicap of childhood mortality. Low weight neonates are at a high risk of serious medical issues, enduring disabilities, and even death. It is, hence, a significant indicator that has a persevering impact on health results in adult life. Being undernourished in the uterine accelerates the possibility of death in the early months and infant's years ahead. The survived is prone to suffer from an impaired immune disability and a high likelihood of contracting diseases; likely to remain malnourished, minimized muscle strength and conceptual abilities during their lives. As adults, they are subjected to an excessive rate of diabetes and heart-related diseases.

In spite of the fact that indications of advancement in infant medical care have significantly diminished the number of deaths affected by low weight during childbirth, hearing misfortune and

vision, learning issues, cerebral paralysis and a little level of survivors creates mental impediment. It is viewed as the single most critical indicator of mortality particularly of death inside the inside month of baby's life. However, the present study is “To find out the prevalence of low birthweight and assess the different factors affecting the weight of a newborn child in Ghana”.

1.3 Research Objectives

1.3.1 General Objective

To identify a method for the explanation of birthweight in terms of cognizable factors so that using this method a possible procedure is defined to assess the weight of the newborn infant beforehand.

1.3.2 Specific Objectives

- ☉ To calculate the prevalence estimates of birthweight (low and normal).
- ☉ To examine the association between the key risk factors affecting birthweight.
- ☉ To assess the type of birthweight reporting system in Ghana.
- ☉ To investigate the use of the Mother's perception of child size at birth (variable) as a proxy to birthweight of an infant if weigh data is missing.
- ☉ To develop statistical prediction models for the indicators of birthweight using mothers' characteristics and birth outcomes.

1.4 Research Questions

Different investigations have been done with various factors influencing birthweight with various settings. The review of literature unveils a major public health burden associated with birthweight. As evident from the available literature, very few studies attempted at utilizing the available large-scale quality data like Ghana Demographic and Health Survey (GDHS) to carry out systematic statistical analysis of existing burden of low birthweight and related epidemiological models. In

view of these perspectives, the present examination has been done to answer a portion of the pertinent questions like:

- (i) What are the prevalence rates of birthweight (low and normal)?
- (ii) What is the type of birthweight reporting system in Ghana?
- (iii) Can we investigate the use of the Mother's perception of child size at birth (variable) as a proxy to birthweight of an infant if weigh data is missing?
- (iv) Which factors contribute to and influences the birthweight of newborns in Ghana?
- (v) If possible, can we hypothesize a statistical model in predicting the likelihood of a child's birthweight belonging to a low or normal class using chosen factors such as maternal factors, outcomes of birth and maternal anthropometric variables?

Such studies from time to time will substantiate the epidemiological evidence on low birthweight and provide insight for planning appropriate interventions. Epidemiological model on low birthweight for the country as well as for specific region may give extra insights for the policy planners.

1.5 Scope of the Study

Indeed, even today, the study of risk factors contributing to birthweight is famous among researchers. Since the research carried out in various geographical environment (settings) indicated diverse outcomes, the causal relationship between critical risk factors and birthweight establishment is hard. This study brings together evidence from past and current researches with an objective to address a significant public health challenge in relation to what already persists in Ghana. Moreover, the emphasis of the study is to establish the association between key risk indicators and infants' weight from a population-based dataset and to explore how these relationships were mediated by some covariates. Women aged between 15 and 49 years with a

history of live delivery were included in the research. External validity is a critical challenge in research when results from a sample of data are inferred into the whole population. The fact that GDHS data were used and approximate very closely to a census, the results may be generalizable.

1.6 Rationale for the current study

As a statistician attempting to utilize data modeling skills in giving a reasonable clarification to birthweight, which is overwhelmingly an issue related to the medical profession, may pose many challenges. From the available literature on this problem, it may be observed that there is not a unique method to estimate birthweight, which suits every environment. As the competing variables are many in number and they vary in their characteristics due to variations in demographic, behavioral and prenatal care of mothers, a unified methodology to estimate birthweight seems to be a difficult task. This might be an explanation behind the continuous enthusiasm for this issue. In this context, this study is aimed at providing an explanation of birthweight using all available related maternal characteristics, demographic, behavioral, pre-natal care profiles. The main issue being addressed is the question of identifying a procedure that could be applied to a large class of situations. Many statistical tools are available to evaluate the characteristics of variables under consideration and generate a model that may provide a stable procedure to estimate birthweight. The study involves data obtained from Ghana Demography and Health Survey 2014.

1.7 Study significance

Few statistical models have been developed yet to address the issue. This study proposes to analyze these elements and build up an exploratory model of these determinants of birthweight that might be valuable in epidemiological investigations. Moreover, the study is significant because it uses the descriptive information from the Ghana Demography Health Survey 2014 report to draw conclusions about the general population of Ghanaian children with respect to the information

obtained from the sample. In this regard, the research findings will be a contribution to knowledge and would be a basis for further research into birthweight in Ghana and the world. The research findings will empower Stakeholders, Obstetricians, Pediatricians, Public Health Physicians, and Health Policymakers to have much insight into the significant contributing factors of birthweight. Additionally, to comprehend the epidemiology of low birthweight, decrease its prevalence and associated public health consequences in order to decrease the scope of intervention policies to accelerate the success of achieving the Millennium Development Goal 4. It is also amazing to note that, this study will serve as a guide in detailing, formulation, and implementation of child survival policies in the country.

1.8 Organization of the thesis

The present investigation under the entitled “Determinants of Low birthweight in Ghana” is organized into five chapters.

Chapter one gives a concise introduction and details the background for our study. It likewise discusses the problem statement under study, the study objectives (general and specific), the research questions and the study significance.

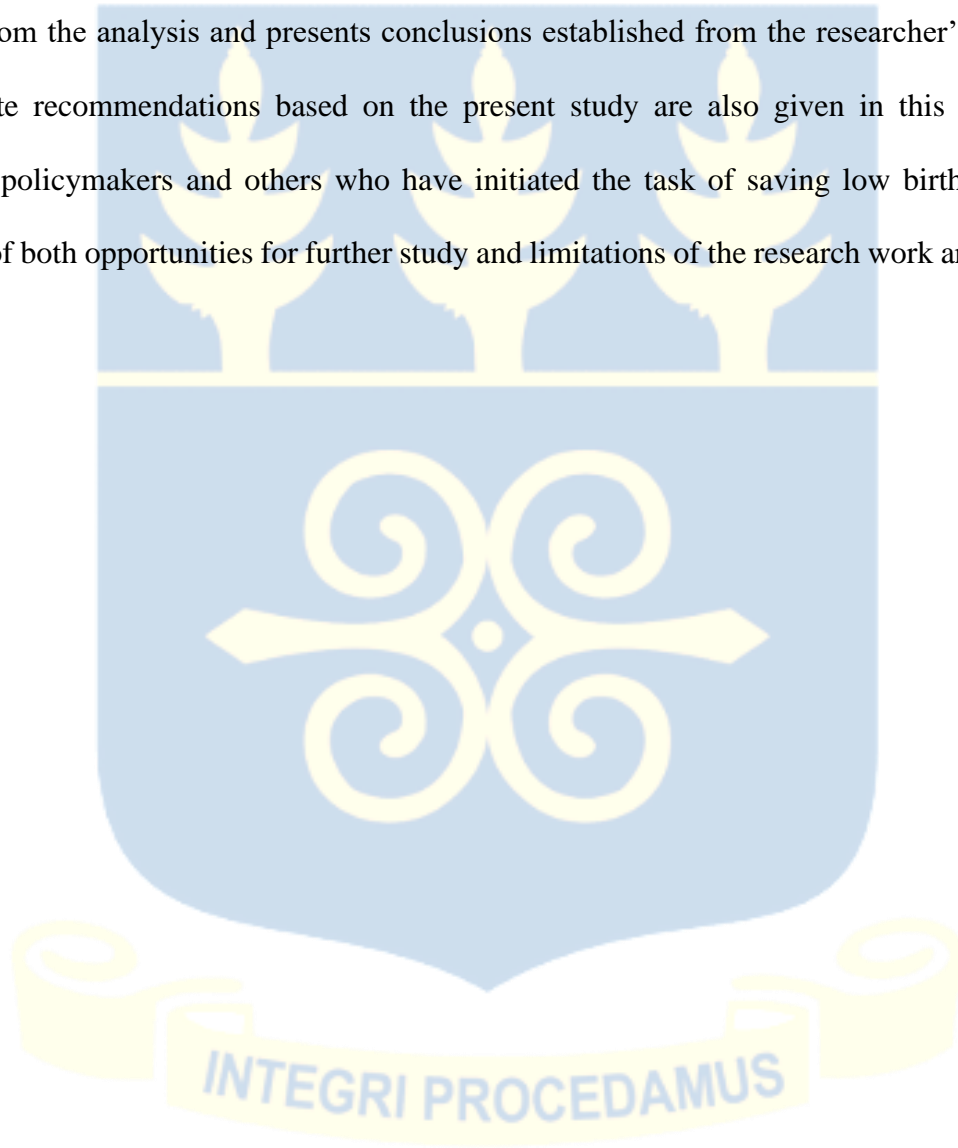
Chapter two deals with different studies done in parts of the world. It deals with the review of developed countries, developing countries, Africa and Ghanaian studies. It additionally manages the hypothetical foundation on birthweight by different researchers with a comprehensive review to establish relationships between birthweight and different factors. It gives us a frame of reference on the key risk determinants significant to the study.

The methodology adopted for the study is described and detailed in chapter three.

In chapter four, the data is explored, analyzed and discussed using R Software packages and SPSS 25 to produce results. Different tools such as frequency, percentage analysis, Analysis of Variance (ANOVA), Crosstabs and Chi-square test were employed.

Chapter five presents the discussion (overall findings to the objectives understudy) based on results produced from the analysis and presents conclusions established from the researcher's findings.

The concrete recommendations based on the present study are also given in this chapter to authorities, policymakers and others who have initiated the task of saving low birthweight. A discussion of both opportunities for further study and limitations of the research work are likewise noted.



CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

A theoretical foundation in our birthweight study is formulated in this chapter. Its primary concern is to review pieces of literature from epidemiological setting the significant indicators whose influence has been reported and discussed.

2.1 Overview of Birthweight

Birthweight is the first weight of the newborn baby measured following birth, preferably between the first hour of life before critical postnatal weight reduction happens. In all countries (developing and developed), birthweight is likely the most significant single factor that influences neonatal mortality, notwithstanding be a significant determinant of post neonatal mortality and Child mortality. It is said to literally follow a person from the cradle to the grave as it is related not just with morbidity and mortality of infants, but with outcomes happening later in life, including adult mortality (Basso, 2008). Thus, birthweight has been considered as a sensitive index (indicator) of a country's wellbeing and viability of a newborn infant.

2.2 Classification of birthweight

For past centuries, birthweight has been treated as dichotomous. 'Low birthweight' (LBW) is the class of infants weighing under 2500 grams (up to and including 2499g) during childbirth, and 'Normal birthweight' is all the rest (that is greater than or equal to 2500 gram).

2.3 Consequences of Low birthweight

The effect of birthweight seems to expand well beyond infancy. Low birthweight infants are vulnerable to early growth retardation, exposed to infectious disease and considered death during infancy and childhood (WHO, 2015). As indicated by the fetal origin's hypothesis (Barker, 2012), infant undernutrition for which low birthweight is a marker, may for all time program the body by

lessening the number of cells in specific organs, changing the circulation of cell types and affecting metabolic processes. These customized changes are related to an assortment of chronic infection results during the adult stage and old age, for example, cardiovascular ailment, diabetes, and hypertension.

Moreover, existing literature suggests that infants of low weight are more likely than normal birthweight infants to have physiological, neurodevelopmental complications and congenital abnormalities extending from early stages through adolescence and into adulthood. Studies have discovered a significant relationship among birthweight and school-age inabilities (Avchen, Scott and Mason, 2001), behavior issues (Sommerfelt, Ellertsen and Markestad, 1993), cognitive function during young adulthood (Richards, Hardy, & Kuh, 2001), preterm birth (Porter, Fraser, Hunter, Ward & Varner, 1997) and gestational diabetes (Innes, 2002). It is also postulated that the most common cause of morbidity and mortality are associated with low birthweight babies (Maheswari, 2014). It is commonly perceived that being brought into the world with low weight is a hindrance for the infant.

2.4 Determinants of Birthweight

By going through the available literature on research relating to birthweight, it is observed that a vast majority of researchers have reported the relationship between wide spectrums of factors. Considering such factors, the present study classified them into three namely; (1) Normal, (2) Intermediate and (3) Non-normal. The studies highlight the importance of these variables sequentially below.

2.4.1 Normal factors

These are indicators that are an element of all pregnancy and where the related increment or decrement in fetal development rate does not of itself impact the prognosis for the child or the mother future pregnancies.

2.4.1.1 Sex of infant

Of a similar gestation, male newborn children appear to be heavier than female babies. This discovering shows up dependably in a wide scope of population studies from Australia (Kettle, 1960) to Canada (Love and Kinch, 1965) to Aberdeen (Thomson, 1968) and to Malta (Camilleri and Cremona, 1970). Although differences do happen (Adams and Niswander, 1968), in general, the reported variation is somewhere in the range of 120g and 150g.

2.4.2 Intermediate factors

These are factors present in all pregnancy but where the consequences for development rate may themselves impact prognosis or where the impacts are bewildered with the impacts of ‘non-normal’ indicators.

2.4.2.1 Parity (Birth order)

Parity is the number of deliveries a woman has born after a pregnancy duration of at least 28 completed weeks. It is one of the central factors fundamentally associated with birthweight, obviously, an element of all pregnancy. It was settled that first newborn is lighter than later newborn at all ages of gestation, that is birthweight increases with birth order (Seidman et al., 1988). Greenwood R, et al. (1994) compared singleton survivors born to multigravida and showed that higher the gravidity, lower is the risk of poor pregnancy outcome for current pregnancy. Data from the Indian population shows similar results. For example, on investigating 331 Bengalee

mother-infant sets, (Bisai 2006) saw that the difference of mean birthweight among first and second pregnancy was 145 grams, while that among second and third pregnancy was 72 grams. While (Thomson et al., 1968) study reported no further increment after second pregnancy, Camilleri and Cremona (1970) in Malta found that weight of newborns increases up to the ninth or tenth pregnancy. Additional researches have demonstrated that Mothers who had 3-4 children were 24% more likely to have low weight infant as compared with women who had brought forth at least five children (Muula, Siziya, and Rudatsikira, 2011).

2.4.2.2 Maternal Size

2.4.2.2.1 Maternal height

The maternal height provides a measure of past nutrition of the mother and therefore an indicator of long-term nutritional status. The positive correlation between birthweight and maternal height all-inclusive finding. Thomson et al., (1968) demonstrated a difference of 200g to 300g between the shortest women in their study (under 5ft 1in) and the tallest (greater than 5ft 4in). A further study in 1971 by Thomson demonstrated that when a newborn's weight was plotted against maternal height for several populations with various mean birthweights, the slope of the graphs was approximately parallel. This finding would contend that maternal stature of itself comparably affects birthweight in different population.

Additionally, a study conducted to examine the independent effect of maternal height on birthweight demonstrated that increasing maternal height was significant and positively related to newborn child weight (Abrams B., Selvin S., 2000). However, various attempts have been made to identify cutoffs of height for estimating the risk of low birthweight. For instance, Phaneendra Rao et al. (2001) noted that mothers who were shorter than 145cm gave birth to children who weighed almost 600g less than mothers with height more than 160cm did. According to Eltahir

Elshibly and Gerd Schmalisch (2008), the maternal height of less than 156cm was found to increase the relative risk for low birthweight about 52% in Sudanese women.

Studies done on Indian mothers provide a range of short heights from 145 cm to 156 cm as risk cutoffs associated with low birthweight (Mohanty C. et al., 2006; Nahar S. et al., 2007). Further, Anitha C. J. (2009) showed that an increase in height of 1cm contributed a 13g increase in birthweight. As maternal height does not change during the short duration of pregnancy, it can be used for prediction of risk of low birthweight at the time of registration.

2.4.2.2.2 Maternal weight

Weight is the simplest anthropometric measurement recorded in field studies and is an indicator of current nutritional status. Pre-pregnancy maternal weight is one of the significant indicators of birth outcome. Serial measurements of weight during gestation are useful for estimating weight gain that ensures the progress of fetal growth and therefore is an influential parameter associated with birth outcomes. Using weight, Miller et al., (1978) have demonstrated that being maternal underweight is a critical risk characteristic related to infants small for gestational age. Luke and Petrie (1980) discovered that, at the other extreme, among obese women, increasing weight detrimentally affected birthweight. Anderson (1989) recommended that weight in from the early pregnancy, up to the thirteenth week could be utilized as a surrogate of pre-pregnancy weight.

While Pune Maternal Nutrition Study (PMNS) reports average pre-pregnancy to be 41.7 kg \pm 5.1 kg in (Rao S. et al. 2001), it was 43.7 kg \pm 6.6 in Karnataka. Phaneendra Rao et al., (2001) and Agarwal K. N. et al., (2001) reported it to be 42.5 \pm 3.9kg in rural Uttar Pradesh, whereas mothers from Bangalore had an average weight of 51.2 (46.2–57.5) kg (Muthayya S., 2005). Among all indicators of low infant weight, Nahar S., et al. (2007) observed the maternal weight to be the most indicator and every 1 kg increment in maternal weight was demonstrated to be related with an

increase in birthweight approximately 260g with statistically significant correlation ($r = 0.4$, $p < 0.001$).

2.4.2.3 Maternal age

While mother's age is a factor normal to every pregnancy, the age at which mothers begin their childbearing are firmly identified with socio-economic status (Thompson, 1982) and due to this close relationship between parity and mother's age, it is hard to evaluate the independent effect of mother's age. Leppert et al., (1986) conducted their study among adolescents and older mothers in New York and reported maternal age as a significant predictor of birthweight. Viegas et al., (1989), based on a study conducted in Singapore validated a quadratic relationship between birthweight and maternal age. Fraser et al. (1995) found that a younger maternal age conferred an increased risk for low birthweight. Gage et al. (2009) reported that a mother's age at birth significantly influences the weight distribution in a study from eight populations in New York State.

Selvin and Janerich (1971) and Miller et. al., (1978) contend that adolescent pregnancy is related to small infants. Meanwhile, some latter researchers suggested that mothers beyond age 35 possess a higher risk of giving birth to children small for gestational age. Low birthweight disparities by maternal age are complexly related to socioeconomic disadvantage and current social and behavioral factors. Adolescents or teenage mothers (less than 20 years) frequently have reproductive conditions, worse socioeconomic and perinatal outcomes when compared to other age classes, for example, those between 20-29 years.

2.4.2.4 Ethnic origin

Differences in, for instance, lifestyle and nutrition among various ethnic classes represent significant confounding features. Additionally, the intricate impacts of various ethnic blends in the various guardians are hard to evaluate. In Ghana, where our populace is of moderately

homogeneous ethnic roots, it is absurd to address the issue of ethnicity in either research or clinical standards. Even when data is accessible, the standard development explicit to ethnicity for clinical purposes might be in fact doable; regardless of whether it is appropriate, could possibly rely upon numerous different factors. Additionally, while it might be useful for analysts to know about the dispersion of various ethnicity in the populace since the difference in child weight-related with differences in ethnicity may confound different investigations.

2.4.3 Non-Normal factors

These are factors that might happen and where there is at first sight evidence to assume that the related increment or decrement in the fetus affect prognosis for the child or mother future pregnancies.

2.4.3.1 Socio-economic factors

Socioeconomic status is an incorporated measurement of individual work experience, or family's socioeconomic position in relation to another, based on wealth, educational status and occupation. The prominent factors of socioeconomic status as viewed by various researchers with respect to low birthweight are discussed below.

2.4.3.2.1 Educational level

Maternal literacy is one of the important factors that are known to be associated with infant mortality rate and nutritional status of a child. The educational level of individuals in the family has a huge influence on the social welfare of members of the family. Therefore, higher levels of education have relatively larger and increasing benefits (Rolleston, 2011). Less educated mothers are known to have low birthweight infants (Chiavarini, Bartolucci, Gili, Pieroni, & Minelli, 2012). Infants of women with low or intermediate education have significantly higher odds of low birthweight than those of higher education (Gisselmann, 2005).

Additionally, Kleinman and Madans (2006) found that women with less than 12 years of education were two times higher odds ratio of low birthweight compared to mothers with greater years of education. Conversely, Subramanyam M. A. et al., (2010) detailed that in India, newborn children were more likely to be brought into the world with low weight, on the off chance that they had mothers with more than 12 years of training compared with 1 to 5 years of education with a relative risk (RR) of 0.79. Some studies do not show an association of low birthweight with maternal education. Maternal education may have direct or indirect effects on pregnancy outcome, its significance may be affected by other social variables when considered simultaneously.

2.4.3.2.2 Residential status

In Ghana, most inhabitants live in a rural residence. Residence determines the availability of social amenities like housing, health care, education, and it is shown that rural-urban difference poses these inequalities (Sahn & Stifel, 2003). Living in rural areas in Sub-Saharan Africa means living in a deprived community in terms of social amenities, infrastructures and job opportunities that convey an increased risk of low birthweight. It has also been shown in a study in Ghana that, being a rural dweller increased the probability of having a low weight infant (Kayode et al., 2014). Hillemeier et al., (2007) detailed rural areas further as large rural city-focused areas and more rural areas compared to urban areas and low weight risk is related with some however not a wide range of rural settings when compared with the urban setting. Auger et al., (2009) concluded that rural relative to urban area as well as low socioeconomic status (represented by maternal education) as having an association with low birthweight.

2.4.3.3 Mode of delivery

Spontaneous vaginal delivery is the commonest mode of delivery. A study in China has shown that the cesarean section rate of low birthweight infants increased with the increasing of gestational

weeks (Chen et al., 2015). It indicated that the cesarean section rate of low birthweight infants was 61.14%, which was higher than that of normal birthweight infants (52.947%).

2.4.3.4 Number of antenatal care (ANC) visits

This is the totality a pregnant woman visits the clinic to receive antenatal care until delivery. As per GDHS (2017) report, almost all (98%) women age 15 to 49 with live birth or stillbirth received antenatal care from skilled providers during pregnancy. While 64% of mothers had their visit during the first trimester of pregnancy, 89% followed the WHO recommendation of at least four visits during pregnancy.

An inadequate number of ANC visits, laboratory studies and exams has appeared to have a higher risk of low weight newborns. A study in Nepal to find the factors of low birthweight noted that mothers who do not attend antenatal care, have increased odds of having a low weight infant by twice more (Odds Ratio=2.3) (Khanal, Zhao, & Sauer, 2014). Further studies have found that even after adjusting for other differences like socioeconomic status and maternal age, low birthweight was highest 36.8% among the mothers who had no antenatal checkup, but it was 15.9% among those who had check-up more than 7 times (Nahar N. et al., 1998).

Negi K. et al., (2006) also noted that mothers with one antenatal care visit had almost six times increased risk of having a low birthweight baby as compared with mothers who received more than four antenatal visits (Odds Ratio = 5.71). Raatikainen et al. (2007) concluded that none or under-attendance for antenatal care is related to an elevated risk of low birthweight. Tayie and Lartey (2008) revealed that early antenatal care is crucial to favorable pregnancy outcomes including birthweight.

2.5 Related Literature Reviews outside Ghana

Tema (2006) conducted a cross-sectional descriptive study to determine the prevalence and key indicators of low birthweight infants in Jimma zone, South West Ethiopia from September 1, 2002, to March 30, 2003. A sample size of 645 newborn mother pairs who delivered in the four health centers, one specialized referral hospital, home deliveries and those who received care in the above health facilities within 24 hours of delivery during the study period were considered. Among the 645 live births included in this study, 145 were low birthweight indicating a prevalence rate of 22.5%. Those mothers residing in urban areas had a high proportion of delivering newborn babies with low weight compared to rural mothers with differences statistically significant ($p = 0.00$). The other factors related to the socioeconomic status of the mothers like age, religion, ethnicity and marital status revealed no statistically significant association ($p > 0.05$) with low birthweight. Low birthweight in relation to maternal obstetric history showed that mothers who delivered before 37 weeks of gestation and those mothers who had multiple pregnancies had a higher proportion of low birthweight babies with the differences statistically significant ($p = 0.01, 0.00$ respectively). Mothers who had weight loss and those who did not receive additional nutritional supplement during pregnancy had an increased risk of delivering low birthweight babies with the differences statistically significant ($p = 0.00, 0.00$) respectively. The other determinants such as the number of pregnancies, history of STI, hypertension, anemia, engaging in light work during pregnancy, family size and history of chronic illness had no association with LBW delivery.

Mehri Rejali et al., (2017), the study evaluated assessed elements connected with low birthweight and deployed a decision curve analysis (DCA) to define a scale to predict the likelihood of having a low birthweight newborn baby. As a hospital-based case-control, the study included 470 mothers with normal birthweight neonates and 470 mothers with low birthweight neonates. Factors found

to be significantly associated with low birthweight were former low weight infants (OR = 2.99 [1.510–5.932]), last trimester of pregnancy bleeding (OR = 2.58 [1.018–6.583]), hypertension in pregnancy (OR = 2.39 [1.429–4.019]), premature membrane rupture (odds ratio [OR] = 3.18 [1.882–5.384]), mother age >30 (OR = 2.17 [1.350–3.498]) and premature pain (OR = 2.70 [1.659–4.415]). However, with decision curve analysis, the prediction model made on these 15 variables had a net benefit (NB) of 0.3110.

2.6 Related Literature Reviews in Ghana

Fosu M. Ofori et al., (2013), the study demonstrated the incidence of low weight among newborns and its associated maternal risk indicators in Manhyia District Hospital, Kumasi-Ghana. The study was a facility-based cross-sectional analysis from the maternity ward of the hospital. 1,200 women were sampled within the reproductive age (15 – 49 years) between 2010 and 2012 from a total delivery of 24,025. Multiple logistic regression was employed to determine the relationship between maternal risk factors and low birthweight. The estimated low birthweight prevalence was reported as 21.1%. The leading factors reported to be significantly associated with low birthweight included maternal age (p-value=0.0160), Residence (p-value =0.0000), Hemoglobin level (p-value =0.0020), Fetal infection (p-value<0.0000) and Antenatal Care (p-value =0.0040). All other variables considered not significant (p-values > 0.05) were height, weight, gestational age and baby's sex.

Mensah Y. Evelyn (2015), the study comprised data obtained through personal interviews from mothers in their postpartum at Ridge Hospital, Kumasi and a secondary data generated from the mother's antenatal book to identify risk factors associated with low birthweight (LBW) and Low Apgar score (LAS). The study encompassed 330 women who delivered at the hospital between February and March 2015. The prevalence of LBW and LAS at Ridge Hospital were 18.8% and

15.2% respectively. Using logistic regression, the significant risk factors associated with LBW were found to be Retroviral (HIV) status of the mother, Gestational Age, Daily Hours Rested, Frequency of Eating and Type of Cooking Fuel used. The factors that significantly influenced LAS were Retroviral (HIV) status of the mother, Gestational Age, and Daily Expenditure. The results showed that HIV-Positive Mothers are more likely to give birth to a newborn with LBW and LAS. Retroviral (HIV) status of the mother was found to be the most important determinant for both LBW and LAS.

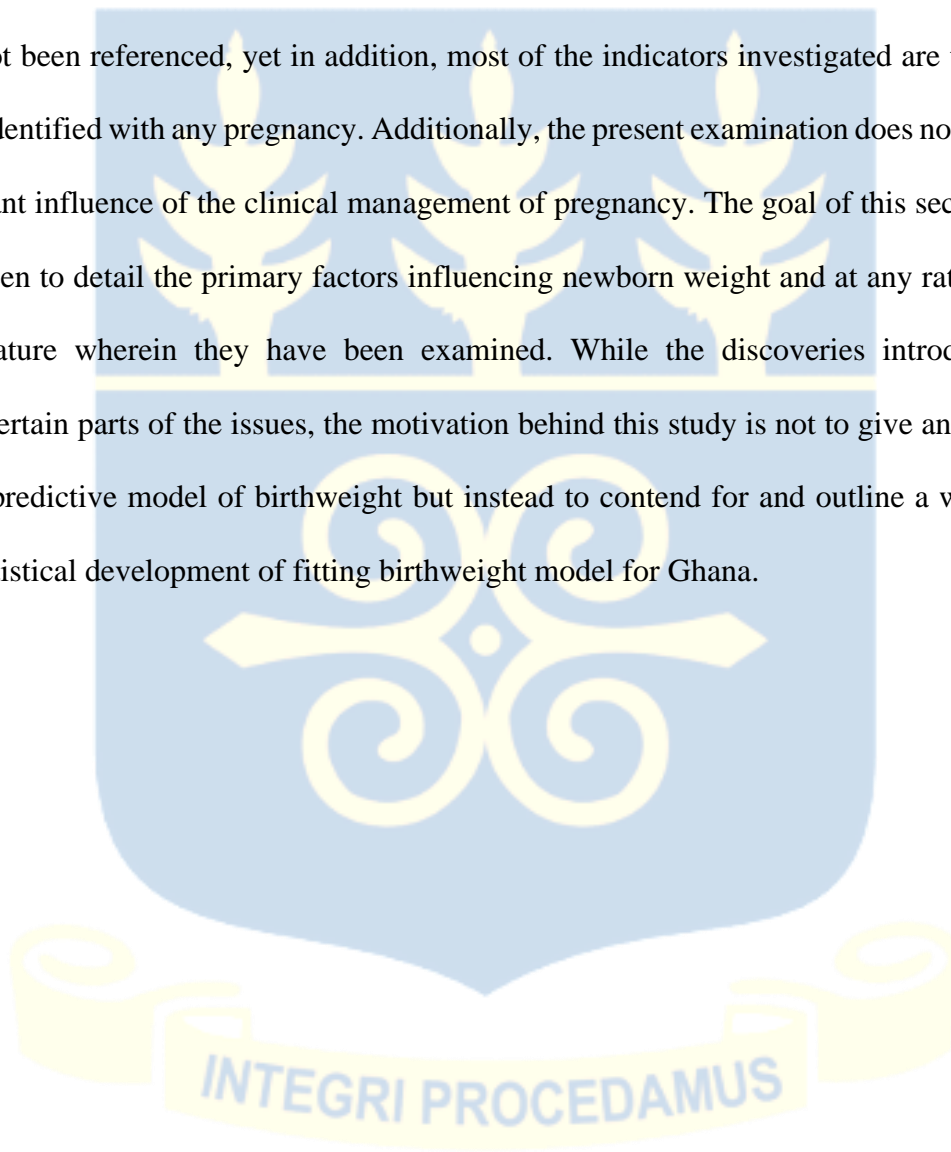
Atinuke O. Adebajani and Puurbalanta R. (2015), in the study, a logistic regression model was utilized to identify the determining variables in predicting LBW babies based on the birth records of 500 mothers of singleton neonate's resident in the Tamale metropolitan area of the Northern Region of Ghana from November 2010 to January 2011. The significant model coefficients were Gestation (p-value = 0.0008), Household size (p-value = 0.0160), Maternal food intake (p-value = 0.0002), Maternal health (p-value = 0.0000), Passive smoking (p-value = 0.0003) and Type of fuel used for cooking (p-value = 0.0418). A test of predictive ability of the model showed correct classifications of 93% for normal birthweight infants and 76.8% for low birthweight infants. The likelihood ratio and Nagelkerk R square tests showed a positive correlation between the predictors and LBW.

Tampah-Naah M. Anthony et al., (2016), population-based study design and a cross-sectional study using data from the Ghana Multiple Indicator Cluster Survey 2011 on some selected maternal factors. A binary logistic regression model was generated to assess factors associated with low birthweight among mothers. Mothers with no education (Odds ratio = 0.566, 95% CI. = 0.349 – 0.919) were less likely to have children with low birthweight, and those not in union (OR = 1.698, 95% C.I. = 0.993 – 2.905) had a higher likelihood of giving birth to children with low

birthweight. Maternal factors such as educational status and marital status showed to influence the birthweight of a child.

2.8 Conclusion

The literature on birthweight is enormous and ever-expanding, not exclusively are there variables that have not been referenced, yet in addition, most of the indicators investigated are themselves intricately identified with any pregnancy. Additionally, the present examination does not talk about the significant influence of the clinical management of pregnancy. The goal of this section at this point has been to detail the primary factors influencing newborn weight and at any rate to allude to the literature wherein they have been examined. While the discoveries introduced may illuminate certain parts of the issues, the motivation behind this study is not to give an all-around acceptable predictive model of birthweight but instead to contend for and outline a way to deal with the statistical development of fitting birthweight model for Ghana.



CHAPTER THREE

METHODOLOGY

3.0 Introduction

The analytical methods and their mathematical aspects are detailed in the chapter. Moreover, it represents a step forward for the analysis for the next chapter.

3.1 Data description

3.1.1 Type and Source of Data

The study utilized secondary data from the 2014 Ghana Maternal Health Survey (GMHS), a nationally representative household survey to gather comprehensive information on 9,396 women age 15 to 49 in the country.

3.1.2 Sampling and Sampling Procedures

Data for birthweight was recorded using a measurement scale (in grams). During data collection, birthweight was confirmed by way of a documented evidence on the child health card or verbally from the mother. The DHS data were weighted prior to analysis to reflect population representativeness. Figure 1 shows the stages of how the final dataset for analysis was arrived at. Of the 26,003 original participants in the dataset, birthweight was not documented on 20,122 of the infants. 543 participants did not know about birthweight, while data from 1,971 of the births were excluded because they were not weighed at birth. In my study, 3,361 records were available for analysis.

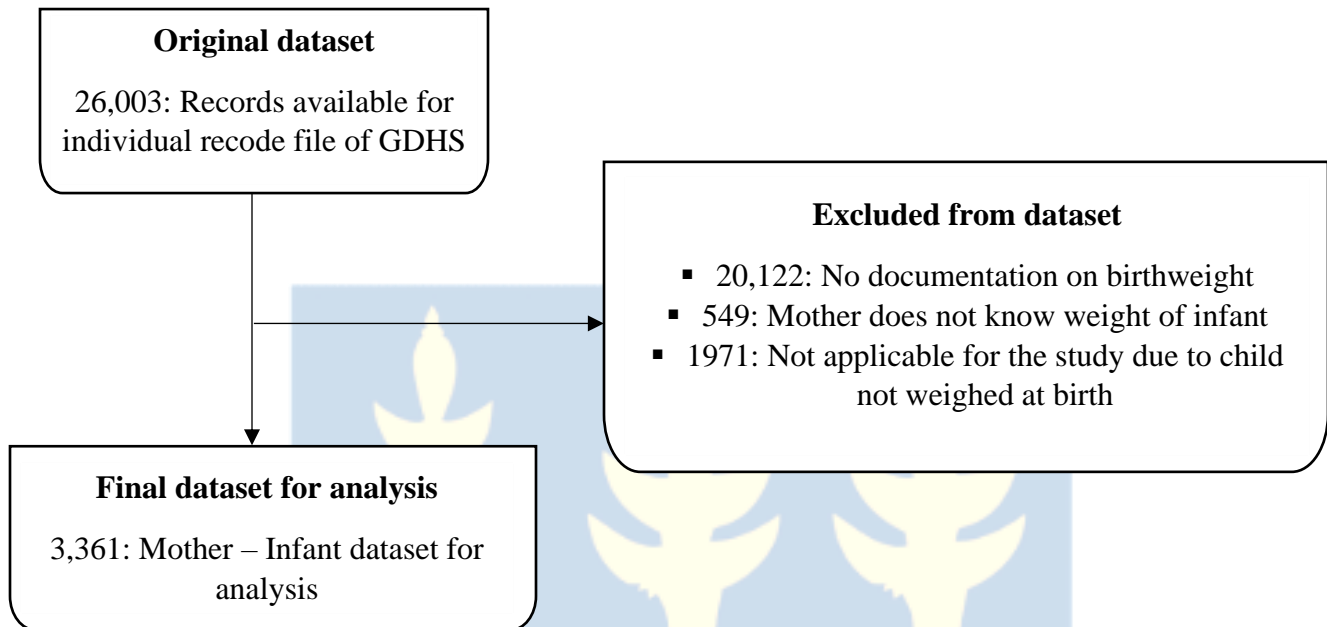


Figure 1: Data Cleaning Procedure

3.1.3 Instrumentation and Operationalization

Validated questionnaires per the protocol for undertaking a GDHS were used. Data generated using the women's questionnaire were recoded into an individual record file, which was then used to select specific variables for the study. Of significant to note is that the women's questionnaire contains both maternal and infant data. Understanding the measurement level on each of the variables helps to select the most appropriate statistical method to use when analyzing the data. The scales of measurement of data were nominal, ordinal, and continuous. Variables that are measured on an ordinal scale take on intrinsic ordering whereas on a nominal scale the variables are categorical and mutually exclusive. Data measured on a continuous scale allow for the use of advanced statistical analysis. (See Table 1)

Table 1: Variables with the level of Measurement

Variable	Level of measurement
Total Number of children ever had Wealth index Education level Birth order (Parity) Size of the child	Ordinal
Marital status Place of residence Region Ethnicity Fuel type for cooking Toilet facility type Gender of child Place of delivery Birth type	Nominal
Birthweight Mothers weight Mothers height Maternal age	Continuous

3.1.4 Our Study assumptions

Four key assumptions underpinned this study. These assumptions are necessary considering that data collection was cross-sectional.

- First, the sensitivity of instruments used to weigh infants at birth was assumed to be the same across all health facilities in the country.
- Second, as individuals from households either agreed, declined, or were unavailable at the household during the time of the survey, it was assumed that characteristics of all individuals, hence birth outcomes regardless of participation in the study were randomly distributed.
- Third, the socio-economic situation of the mother as presented during data collection was assumed to be the same as at the entire duration of the pregnancy through to delivery. For

example, if at the time of the survey the mother was using firewood or LPG as fuel for cooking in the household, the study assumed that the same form of energy was also used during the lifetime of the pregnancy.

- Finally, the underlying influence of covariates was additive, each additional covariate contributing the same additional risk were assumed to be independent.

3.2 Application of Statistical Methods

3.2.1 Descriptive Analysis

The study data collected needs to be analyzed systematically so that patterns of association can be identified. Here, we are concerned with describing the numerical distributions that characterize birthweight variables. The study considered dataset containing 18 variables with 3361 observations each. Three phases of analysis were done including descriptive (univariate) analysis and inferential (bivariate and multivariable) statistical approaches.

- *Univariate analysis*

Two approaches were used in the univariate analysis depending on the type of the variable being analyzed. Firstly, for continuous variables including age and weight measures of central tendency (mean and median) were calculated along with measures of distribution (standard deviation and range) to determine the distribution of the variables. These descriptive statistics were summarized and presented. Secondly, for each categorical variable, the univariate analysis was presented as a table of frequency distribution containing both the frequency and corresponding percentage.

- *Bivariate analysis*

The bivariate analysis involved cross tabulating each independent variable against birthweight classifications and comparing the proportion of mothers in the different level of each

independent factor who had low birthweight delivery. The chi-square test is performed to test for significant associations between risk characteristics and birthweight classification. An alpha cut-off level of 0.05 would be used to determine statistically significant associations.

- *Multivariable analysis*

The multivariable analysis was conducted using regression models (Multiple Linear and Logistic) with birthweight delivery as the dependent variable.

3.2.2 Cross-tabulation for birthweight and birth size (2 x 2 Contingency table)

This 2 × 2 contingency table is a useful and simple technique in the identification of statistical significance of the association (relationship) of two binary (dichotomous) characteristics usually coded as 1 for “Yes” or “Success” and 0 for “No” or “Failure.” With the help of this contingency table, we can easily understand the various steps of the analysis concepts of birthweight (per health card) and birth size (per mother’s recall).

Table 2: Standard 2 × 2 Table contingency Table

Gold Standard	Birthweight reported in terms of birth size			
	Low (Small size)	Normal (Large Size)	Total	
Birthweight reported	Low birthweight (<2500g)	tp (True Positive)	fp (False Positive)	tp+fp
	Normal birthweight (>2500g)	fn (False Negative)	tn (True Negative)	fn+tn
	Total	tp+fn	fp+tn	N=tp+fp+fn+tn

From Table (2) above, the test can lead to the conclusion that:

- **True Positive (TP):** Proportion of children having a birthweight below 2500 grams and reported as of small size. It is denoted by ‘tp’.
- **False Positive (FP):** Proportion of children having a birthweight below 2500 grams but reported as of large size. It is denoted by ‘fp’.

- **False Negative (FN):** Proportion of children having a birthweight above 2500 grams but reported as of small size. It is denoted by ‘fn’.
- **True Negative (TN):** Proportion of children having a birthweight above 2500 grams and reported as of large size. It is denoted by ‘tn’.

- **Sensitivity** is the ability of the test to correctly identify those who have the disease (tp) from all individuals with the disease (tp+fn). That is the strength of birth size to identify correctly children with low birthweight (LBW-Small size) as true under-recorded birthweight. It is calculated as:

$$Sensitivity = \left\{ \frac{tp}{tp + fn} \right\} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- **Specificity** is the ability of the test to correctly identify those who do not have the disease (tn) from all individuals free from the disease (fp+tn). The strength of birth size to identify correctly children with normal birthweight as true under the recorded birthweight. It is calculated as:

$$Specificity = \left\{ \frac{tn}{fp + tn} \right\} = \frac{True\ Negative}{False\ Positive + True\ Negative}$$

- **Positive predictive value (PPV):** The proportion of children of small size who remained low birthweight as per recorded birthweight. It is calculated as:

$$PPV = \left\{ \frac{tp}{tp + fp} \right\} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- **Negative predictive value (NPV):** The proportion of children of normal size who remained normal as per recorded birthweight. It is calculated as:

$$NPV = \left\{ \frac{tn}{fn + tn} \right\} = \frac{True\ Negative}{False\ Negative + True\ Negative}$$

3.2.3 Kappa Statistics and Fleiss

An agreement between two is required when a diagnostic test is being compared with a Gold standard. What amount is the agreement among negative and positive results of the two tests must be answered? It is basic to evaluate the reproducibility of the test by many observers. At least two observers ought to independently assess the test results without accessing the data. The Kappa lie between -1 to $+1$ as most relationship insights do.

Mathematically, kappa statistic is defined and calculated as

$$Kappa = k = \frac{p_o - p_e}{1 - p_e} = \frac{\text{Observed Agreement (O)} - \text{Expected Agreement (E)}}{1 - \text{Expected Agreement (E)}}$$

Where; Relative Observed Agreement = $\left(\frac{tp + tn}{n}\right)$ and

Total number = $n = tp + fp + fn + tn$

Therefore, Proportion of Chance Agreement = $E = \left[\left(\frac{tp+fp}{n}\right)\left(\frac{tp+fn}{n}\right) + \left(\frac{fp+tn}{n}\right)\left(\frac{fn+tn}{n}\right)\right]$

Table 3: Interpretation of Cohen's kappa & Fleiss Agreement

Kappa statistics	Agreement level	Fleiss	Agreement level
0–0.20	Slight	<0.4	Poor
0.21 –0.39	Fair	0.4-0.75	Fair to Good
0.40 –0.59	Moderate	>0.75	Excellent
0.60 –0.79	Substantial		
Above .80	Almost Perfect		

Fleiss' kappa (after Joseph L. Fleiss, 1981) statistic has been used for a similar measure of agreement in categorical rating when there are more than two raters (Avijit Hazra, 2013).

3.2.4 The Chi-square χ^2 test

The Chi-Square test is frequently utilized for testing relationships between categorical variable. The null hypothesis states that no relationship exists on the categorical factors in the observation; they are independent. Here, the values of χ^2 are also calculated to check the association between birthweight and different predictors. The test statistic for the Chi-Square Test of Independence is computed as:

$$\chi^2 = \sum_{i=1}^n \left[\frac{(O_i - e_i)^2}{e_i} \right]$$

Where: O_i = Observed frequencies and e_i = Expected frequencies

3.3 Statistical Model Building

In general, by statistical models, we mean models of real systems obtained through empirical methods. Statistical models summarize the relationship between a dependent and an independent factor. Mathematically, the way response variable depends upon the values of certain explanatory variables may be explained using a general form,

$$\text{Response variable} = \text{Systematic component} + \text{Random component}$$

Where the systematic component summarizes how variability in a response variable is accounted for by explanatory variables and the random component summarizes deviation of the response values from the systematic component. Suppose there is a set of 'n' observed values of a response variable, then Y can be written as,

$$y_i = \eta_i + \xi_i, \quad i = 1, 2, \dots, n$$

Where η_i is the systematic component and ξ_i is the random component.

It is possible that η_i may be linear in unknown parameters, β , for example,

$$\eta_i = \beta_0 + \beta_1 x_i, \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_i^2, \text{ and } \eta_i = \beta_0 + \beta_1 x_1 + \beta_2 x_{2i} \text{ or}$$

$$\text{Nonlinear as in; } \eta_i = \beta_0 \exp(\beta_1 x_i), \eta_i = \frac{x_i}{\{\beta_0 + \beta_1 x_i\}} \text{ and } \eta_i = \beta_0 \exp(\beta_1 x_{1i} + \beta_1 x_{2i})$$

3.3.1 Our study model building

All statistical related learning problems are built to minimize expected loss. Mathematically, the issue of determining birthweight is that of investigating a given set of elements, the one that predicts the infant weight in the most ideal manner. We minimize our risk factor in a circumstance where the joint distribution of our variables (dependent and independents) are unknown so to choose the best accessible indicators.

Two main types of problems considered in this study:

- Regression estimation
- Classification

In the regression estimation, our problem is to minimize the risk function with the squared error loss function, while the classification problem, we find for an indicator function that minimizes our misclassification error.

3.3.2 Linear Regression Model (Multiple)

Linear models are those statistical models in which a series of parameters are organized as a linear combination. That is, inside the model, no parameter shows up as either a multiplier, divisor or exponent to some other parameter. The linear model is used to serve the following goals: (1) Modeling the relationship between variables, (2) Prediction of the target variable (forecasting) and (3) Hypothesis testing. The effect of some factor on a dependent variable may be affected by the presence of other factors because of redundancies or effect interactions (modifications).

Consequently, to give a comprehensive study examination, it might be desirable to consider as many factors as possible and sort out which ones are most closely associated with the response variable.

Now, we write our multiple regression model as

$$Y = \beta_0 + \sum_{j=1}^n \beta_j X_{ji} + \epsilon_i \quad \text{or}$$

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i, \quad i = 1, 2, 3, \dots, n$$

Where Y_i represent our dependent variable, the X 's are our independent variables, and ϵ is the error term. We have a dependent factor and k th independent factors excluding the intercept term. The inferential technique of a regression intensely relies upon the following assumptions:

1. Linearity and additivity of the relationship between the dependent and independent variables:
 - a) The expected value of the dependent variable is a straight-line function of each independent variable, holding the others fixed.
 - b) The slope of that line does not depend on the values of the other variables.
 - c) The effects of different independent variables on the expected value of the dependent variable are additive.
2. The X 's are non-stochastic variables whose values are fixed.
3. The error has zero expected value: $E(\epsilon) = 0$
4. Homoscedasticity (constant variance) of the errors for all observations, that is

$$E(\epsilon^2) = \sigma^2, \quad i = 1, 2, \dots, n$$

5. Statistical independence of the errors. Thus, $E(\epsilon_i \epsilon_j) = 0$, for all $i \neq j$.
6. Normality of the error distribution.

Now, we can present our multiple regression model in matrix form as: $Y = X\beta + \epsilon$

Where;

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_k \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,k} \end{bmatrix}$$

The ordinary least square estimate of $k + 1$ unknown parameters are obtained by minimizing the sum of squares errors $\sum_{i=1}^n e_i^2 = \epsilon' \epsilon = (Y - X\beta)'(Y - X\beta)$ yielding $\hat{\beta} = (X'X)^{-1}X'Y$.

Moreover, we obtain $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$. For this model, the residuals then become:

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki}; \quad i = 1, 2, \dots, n$$

However, the Matrix $W = X(X'X)^{-1}X'$ is known as a Leverage Matrix. An unbiased and consistent estimate of σ^2 is

$$S^2 = \sum_{i=1}^n \frac{\hat{\epsilon}_i^2}{(n - k - 1)}$$

The estimated standard error of

$$\hat{\beta}_j \text{ is } s_{\hat{\beta}_j} = \sqrt{S^2 V_j} \text{ where } V_j \text{ is the } j - \text{th diagonal element of } (X'X)^{-1}$$

When the errors are normally distributed, then $\frac{\hat{\beta}_j - \beta_j}{s_{\hat{\beta}_j}} \sim t_{n-k-1}$

3.3.2.1 Goodness of fit check in multiple regression

We can use the R^2 statistic as a measure of goodness of fit for the multiple regression model but the difficulty is that it does not account for the number of degrees of freedom. Although, we know that

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

A natural solution is to use variances, not variations that help to define a corrected (adjusted) R^2 as

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \frac{n-1}{n-k-1} = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

A formal technique to test the goodness of fit by the multiple regression line can be developed using the *ANOVA of Y*. At first, we set the null hypothesis that the overall regression is not significant in a sense that the k explanatory variables considered in this model are not able to explain the response variable in a satisfactory way. We then compute the ratio $\frac{RMS}{EMS}$ that follows a F distribution with k and $n - k - 1$ degrees of freedom. If the calculated value of this ratio is greater than $F_{k,n-k-1,0.05}$, we reject the null hypothesis and conclude that the overall regression is significant at the 5% level of significance.

Hence, for a multiple regression model, we have

$$TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = Y'Y - n\bar{Y}^2$$

$$TSS = RSS + ESS \quad \text{where } RSS = \hat{\beta}'X'Y - n\bar{Y}^2$$

Table 4: ANOVA Table for Linear Regression (Multiple)

Component	Sum of Squares	Degree of Freedom	Mean Sum of Squares	F Statistics
Regression	RSS	k	$RMS = \frac{RSS}{k}$	$\frac{RSS}{ESS} \sim F_{k,n-k-1}$
Error	ESS	$n - k - 1$	$EMS = \frac{ESS}{n - k - 1}$	
Total	TSS	$n - 1$		

3.3.2.2 Tests for Normality

Statistical methods such as correlation, experimental design, and regression are altogether founded on one basic assumption, that the data sets follow normal (Gaussian) distribution. As such, there is an expectation that samples accumulated from the population are normally distributed. At the point when the data sets are not distributed normally, the associated chi-square tests are inaccurate and subsequently, the t and F tests are not generally valid in finite samples.

3.3.2.2.1 Normal Q-Q Plot

This graphical tool empowers us to evaluate if our data set plausibly originate from some theoretical distribution, for example, a Normal or exponential. Two sets of quantiles are plotted against each other in a scatterplot form, if both sets of quantiles originate from similar distribution, we should see our data points forming a line that looks roughly straight.

3.3.2.2.2 Shapiro-Wilk Test

The Shapiro-Wilk (1965) test is based on the correlation of true observations and the expectation of normalized order statistics. The form of the test statistic is:

$$W = \frac{\{\sum a_i \epsilon_{(i)}\}^2}{\sum (\epsilon_i - \bar{\epsilon})^2}$$

Where $\epsilon_{(i)}$ is the $i - th$ order statistics and a_i is the $i - th$ expected value of normalized order statistics.

3.3.2.3 Unusual Observations in Linear Regression

3.3.2.3.1 Outlier in regression

An outlier is a data point that deviates from the linear relationship determined from the other points, or possibly from the greater part of those points.

3.3.2.3.2 High Leverage Points

According to Hocking and Pendleton (1983), ‘high leverage points are those for which the input vector x_i , in some sense, far from the rest of the data have an enormous impact on the distance from the center. To identify outliers, we should consider first looking at the residual plot of e_i versus Y_i . Recall the property of residuals:

$$e_i = Y_i - \hat{Y}_i \sim N(0, \sigma^2(1 - h_{ii}))$$

Where

$$h_{ii} = \frac{1}{n} + \frac{(X_i - \bar{X})^2}{S_{xx}}, \text{ which is called the Leverage of the } i\text{th case}$$

3.3.2.4 Detecting of Outliers

- Deleted Studentized (Externally Studentized or R-Student)

To account for the different variances among residuals, we consider “studentizing” the residuals (that is dividing by an estimate of their standard deviation). It is defined as:

$$r_i = \frac{e_i}{\sqrt{MSE(1 - h_{ii})}} ; i = 1, 2, \dots, n$$

Where $E(r_i) = 0$ and $Var(r_i) = 1$, so that the studentized residuals have a constant variance regardless of the location of the X's. An absolute residual value exceeding three can be considered an outlier.

3.3.2.5 Detecting of Influential Observations

- **Cook's Distance:**

Cook (1977) proposed the use of the distance measure known in statistical literature as Cook's distance. We define the $i - th$ Cook's distance as:

$$CD = \frac{(\hat{\beta} - \hat{\beta}^{-i})^T (X^T X) (\hat{\beta} - \hat{\beta}^{-i})}{(k + 1) \hat{\sigma}^2} ; \quad i = 1, 2, \dots, n$$

Where $\hat{\beta}^{(-i)}$ is the estimated parameter of β with the $i - th$ observation deleted.

3.3.2.6 Multicollinearity (Variance Inflation Factor and Tolerance)

Linear regression assumes that there is little or no multicollinearity in the data. Multicollinearity occurs when the independent variables are too highly correlated with each other. Tolerance and Variance Inflation Factor (VIF) are computed in evaluating both our pairwise and multiple variable collinearities. A tolerance estimate nearing zero indicates that the variable is highly collinear with the other independent variables. Inversely, the variance inflation factor (VIF) is associated with the tolerance estimate: $VIF = \frac{1}{TOLERANCE}$. Larger variance inflation factor estimates (a usual threshold of 10.0 that corresponds to a 0.10 tolerance) indicate a high degree of multicollinearity or collinearity among our predictor variables.

3.3.2.5.1 Corrections for Multicollinearity

To remedy multicollinearity, the following steps can be adopted:

- More data collection
- Drop off insignificant variables
- Use Ridge regression
- Use Principal component regression

3.3.3 Logistic Regression Model

Logistic regression is helpful for situations in which we need to anticipate (predict) the likelihood (for example, yes or no) of a result dependent on estimations of a lot of predictor factors. Its strategies utilize any of these three types of categorical dependent variables: binary, ordinary and nominal.

Now, considering a simple k variable regression model $Y_i = \beta_0 + \beta_1 X_1 + \beta_p X_p + \dots + \beta_j X_{ji}$ where $j = p + 1$ can easily be generalized and expressed as:

$$\lambda(X) = \frac{1}{1 + \exp[-(\beta_0 + \sum_{j=1}^k \beta_j X_{ji})]} \quad i = 1, 2, \dots, n$$

Or, equivalently,

$$\ln \frac{\lambda(X)}{1 - \lambda(X)} = \beta_0 + \sum_{j=1}^k \beta_j X_{ji}$$

This leads to the likelihood function

$$L(\lambda(X)) = \prod_{i=1}^n \frac{[\exp(\beta_0 + \sum_{j=1}^k \beta_j X_{ji})]^{y_i}}{1 + \exp(\beta_0 + \sum_{j=1}^k \beta_j X_{ji})} \quad y = 0, 1$$

Application of computer-packaged programs such as R, STATA, and SPSS can iteratively estimate our parameters.

We would logically let,

$$y_i = \begin{cases} 0, & \text{if the } i - \text{th unit does not have the characteristic} \\ 1, & \text{if the } i - \text{th unit does possess the characteristic} \end{cases}$$

Here, the logistic regression model that specifies the probability of a birthweight depending on a set of n -explanatory variables X . Where $\lambda(X)$ represent the conditional probability that the low birthweight is present. $P(Y = 1|X)$ and the θ 's are the parameters representing the effects of the X_n 's on the risk (probability) of low birthweight.

3.3.3.1 Testing hypotheses of Multiple Logistic Regression

When we have fit a multiple logistic regression model and our estimates are obtained for the various parameters of interest, we need to respond to questions concerning the contributions of different factors to the prediction of the binary response factor.

3.3.3.1.1 Overall Regression Tests

An overall test for a model containing k factors, say,

$$\lambda(X) = \frac{1}{1 + \exp[-(\beta_0 + \sum_{j=1}^k \beta_j X_{ji})]} \quad i = 1, 2, \dots, n$$

Therefore, our null hypothesis is stated as: “All k independent variables considered together do not explain the variation in the responses.” In other words,

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

To test our null hypothesis, two likelihood-based statistics can be utilized with each following an asymptotic chi-square distribution with k degrees of freedom under H_0 .

1. Likelihood Ratio Test

$$\chi^2_{LR} = 2[\ln L(\hat{\beta}) - \ln(0)]$$

2. Score Test

$$\chi_S^2 = \left[\frac{\delta \ln(0)}{\delta \beta} \right]^T \left[\frac{\delta^2 \ln(0)}{\delta \beta^2} \right]^{-1} \left[\frac{\delta \ln(0)}{\delta \beta} \right]$$

3.3.3.1.2 Tests for a Single Variable

The null hypothesis can be stated as; ‘‘Factor X_i does not have any value added to the prediction of the response given that other factors are already included in the model.’’ In other words,

$$H_0: \beta_i = 0$$

To test such a null hypothesis, we can perform a likelihood ratio chi-square test, with 1 degree of freedom:

$$\chi_{LR}^2 = 2[\ln L(\hat{\beta}; \text{all } X\text{'s}) - \ln L(\hat{\beta}; \text{all other } X\text{'s with } X_i \text{ deleted})]$$

A much easier alternative method is using $z_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$. Where $\hat{\beta}_i$ is the corresponding estimated regression coefficient and $SE(\hat{\beta}_i)$ is the estimate of the standard error of $\hat{\beta}_i$, both of which are printed by standard computer-packaged programs.

3.3.3.2 Wald Test

Wald statistic is purposed in evaluating the estimate of individual regressors when different regressors are in the model. It is defined as:

$$W = \frac{\hat{\beta}_i}{s.e(\hat{\beta}_i)} \sim \chi_1^2 \text{ or another version of it as } W = \frac{\hat{\beta}_i^2}{V(\hat{\beta}_i)} \sim \chi_1^2$$

It is based on large sample sizes, that is, the large-sample normality of parameter estimates. The Wald test can be used to determine a $100(1 - \alpha)\%$ confidence interval for $\hat{\beta}_i$ to show that the true

parameter lies in the interval with boundaries; $\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}}(Se(\hat{\beta}_i))$. Where $Z_{1-\frac{\alpha}{2}}$ is the critical value for the two-sided normal distribution of size α .

$$\exp\left\{\hat{\beta}_i \pm Z_{1-\frac{\alpha}{2}}(Se(\hat{\beta}_i))\right\}$$

$$\text{Lower Limit} = \exp\left\{\hat{\beta}_i - Z_{1-\frac{\alpha}{2}}(Se(\hat{\beta}_i))\right\} \quad \text{and} \quad \text{Upper Limit} = \exp\left\{\hat{\beta}_i + Z_{1-\frac{\alpha}{2}}(Se(\hat{\beta}_i))\right\}$$

In terms of the odds ratio, confidence intervals are formed by finding exponents of the boundaries.

3.3.3.3 Logit Transformation

For proper interpretation of the estimated coefficients, the expression given below

$$\lambda(X) = \frac{1}{1 + \exp[-(\beta_0 + \sum_{j=1}^k \beta_j X_{ji})]} \quad i = 1, 2, \dots, n$$

may be transformed as:

$$g(X) = \ln\left[\frac{\lambda(X)}{1 + \lambda(X)}\right] = (\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \dots + \theta_p X_p)$$

The equation above is termed as logit transformation. The importance of this transformation is that $g(X)$ possesses many of the desirable properties of a linear regression model. The logit, $g(X)$, is linear in its parameters and may also range from $-\infty$ to $+\infty$, depending upon the range of values of X .

3.3.3.4 Interpretation of Coefficients

Like in the case of linear regression, the estimated regression coefficients represent the slope or rate of change in log logit function of the dependent variable because of per unit change in the independent variable. In the logistic model, θ represents a change in the logit for a one-unit change in the covariates, X , that is, $g(X + 1) - g(X)$; where $g(X)$ is the logit transformation defined by the expression above. In case of consideration of only one dichotomous independent variable that

is coded as 0 or 1, the values of the regression model may be expressed as described below in table 5.

Table 5: Logit transformation

Outcome Variable (Y)	Independent Variable (X)	
	X = 1	X = 0
Y = 1	$\lambda(1) = \frac{\exp(\theta_0 + \theta_1)}{1 + \exp(\theta_0 + \theta_1)}$	$\lambda(1) = \frac{\exp(\theta_0)}{1 + \exp(\theta_0)}$
Y = 0	$1 - \lambda(1) = \frac{1}{1 + \exp(\theta_0 + \theta_1)}$	$1 - \lambda(1) = \frac{1}{1 + \exp(\theta_0)}$
Total	1.0	1.0

Odds of the outcome (Low birthweight) being present among individuals with $X = 1$ is defined as $\frac{\lambda(1)}{1-\lambda(1)}$. On the similar pattern odds of the outcome being present among individuals with $X=0$ is defined as $\frac{\lambda(0)}{1-\lambda(0)}$. Therefore, the log of the odds, called logit, is defined as:

$$g(1) = \ln \left[\frac{\lambda(1)}{1 - \lambda(1)} \right] \text{ and } g(0) = \ln \left[\frac{\lambda(0)}{1 - \lambda(0)} \right]$$

The odds ratio denoted by φ is defined as the ratio of the odds for $X = 1$ to the odds for $X = 0$:

$$\varphi = \frac{\left[\frac{\lambda(1)}{1 - \lambda(1)} \right]}{\left[\frac{\lambda(0)}{1 - \lambda(0)} \right]}$$

The log of the odds ratio, termed as a log-odds ratio or log-odds is given by the equation:

$$\ln \varphi = \ln \frac{\left[\frac{\lambda(1)}{1 - \lambda(1)} \right]}{\left[\frac{\lambda(0)}{1 - \lambda(0)} \right]} = g(1) - g(0)$$

Now, using the expressions for the logistic regression model as shown in **Table 5**, the odds ratio is:

$$\varphi = \frac{\left(\frac{e^{\theta_0+\theta_1}}{1+e^{\theta_0+\theta_1}}\right)\left(\frac{1}{1+e^{\theta_0}}\right)}{\left(\frac{e^{\theta_0}}{1+e^{\theta_0}}\right)\left(\frac{1}{e^{\theta_0+\theta_1}}\right)} = \frac{e^{\theta_0+\theta_1}}{e^{\theta_0}} = e^{\theta_1}$$

Hence, under the logistic regression analysis with a dichotomous independent variable, $\varphi = e^{\theta_1}$ and the logit difference, or log-odds, is $\ln \varphi = \ln e^{\theta_1} = \theta_1$. Therefore, the odds ratio approximates how much more likely (or unlikely) it is for the low birthweight to be present among those categories with $X = 1$ than those among with $X = 0$.

3.3.3.5 Goodness-of-Fit test in Logistic Regression

We employ a test for Goodness of fit to determine how well the proposed model fits the data. A model is poorly fit if either the model's residual variation is large or it does not follow the variability postulated by the model (Hallet, 1999).

3.3.3.5.1 R^2 in Logistic Regression

The standard regression theory tells us that:

$$\frac{2}{l^n} = \frac{1}{1-R^2} \quad \text{where } l = \frac{L(\hat{\beta})}{L(0)} \text{ is the Likelihood Ratio Statistics}$$

We can rewrite the above expression as:

$$R^2 = 1 - \left[\frac{L(0)}{L(\hat{\beta})}\right]^{-\frac{2}{n}}$$

Since the likelihood function $L(\hat{\beta})$ is a product of probabilities, it follows that the value of the function must be less than or equal to 1. Thus, the maximum possible value for R^2 is given by

$$\text{Max}(R^2) = 1 - [L(0)]^{\frac{2}{n}}$$

In linear regression $\hat{Y} = \bar{Y}$ (for the null model). Similarly, in logistic regression, we would have $\hat{\pi} = \gamma$ for the null model, with γ denoting the percentage of 1's in data set. It follows that

$$L(0) = \prod_{i=1}^n y_i^{y_i} (1 - y_i)^{1 - y_i} = \gamma^{n\gamma} (1 - \gamma)^{n - n\gamma}$$

Here, we can rewrite

$$Max(R^2) = 1 - [\gamma^{n\gamma} (1 - \gamma)^{n - n\gamma}]^{\frac{2}{n}}$$

In other words, the Cox-Snell R^2 is a *Pseudo - R²* statistic and the ratio of the likelihoods reflect the improvement of the full model over the intercept-only model with a smaller ratio reflecting greater improvement. It is given by:

$$Cox - Snell R^2 = 1 - \left[\frac{L(R)}{L(F)} \right]^{\frac{2}{n}}$$

Where,

$L(R)$ = Likelihood of intercept only model

$L(F)$ = Likelihood of specified model

n = number of observations

The maximum possible value will be close to zero when our data is quite sparse. Nagelkerke

(1991) therefore suggests that \bar{R}^2 be used, with $\bar{R}^2 = \frac{R^2}{Max(R^2)} = \frac{1 - \left[\frac{L(R)}{L(F)} \right]^{\frac{2}{n}}}{1 - [L(R)]^{\frac{2}{n}}}$ which is known as adjusted R^2 .

3.3.3.5.2 Hosmer – Lemeshow Test

Hosmer-Lemeshow goodness-of-fit is used for grouped data and it is defined as:

$$HLT = \sum_{i=1}^g \frac{(O_i - N_i \bar{\pi}_i)^2}{N_i \bar{\pi}_i (1 - \bar{\pi}_i)}$$

where g is the number of groups, where N_i is the total frequency of subjects in the i th group, O_i is the total frequency of event outcomes in the i th group, and $\bar{\hat{\pi}}_i$ is the average estimated predicted probability of an event outcome for the i th group.

3.3.4 Variable Selection for Model Building

Our goal is to identify from numerous available indicators a small subset of factors that significantly relates to our response. In the process of identification, our wish is to avoid a large type I error (false positive). Such a goal can be accomplished by utilizing a methodology that enters or removes from a regression model one factor at a time according to a certain order of relative significance.

a) Backward Elimination

1. Start with all the predictors in the model
2. Remove the predictor with the highest p-value greater than the alpha value
3. Refit the model and go to 2
4. Stop when all p-values are less than the alpha value

b) Forward Selection

1. Start with no variables.
2. For all predictors not in the model, check their p-value if they are added to the model. Choosing the one with the lowest p-value less than the alpha value
3. Continue until no new predictors can be added.

3.3.5.2 Criterion-Based Procedures

$$\text{Akaike's Information Criterion (AIC)} = n \ln(\text{SSE}) - n \ln n + 2p$$

$$\text{Bayesian Information Criterion (BIC)} = n \ln(\text{SSE}) - n \ln n + p \ln n$$

$$\text{Amemiya's Prediction Criterion (APC)} = \frac{n+p}{n(n-p)} \text{SSE}$$

Where 'n' is sample size and 'p' is the number of regression coefficients in the model being evaluated including the intercept.

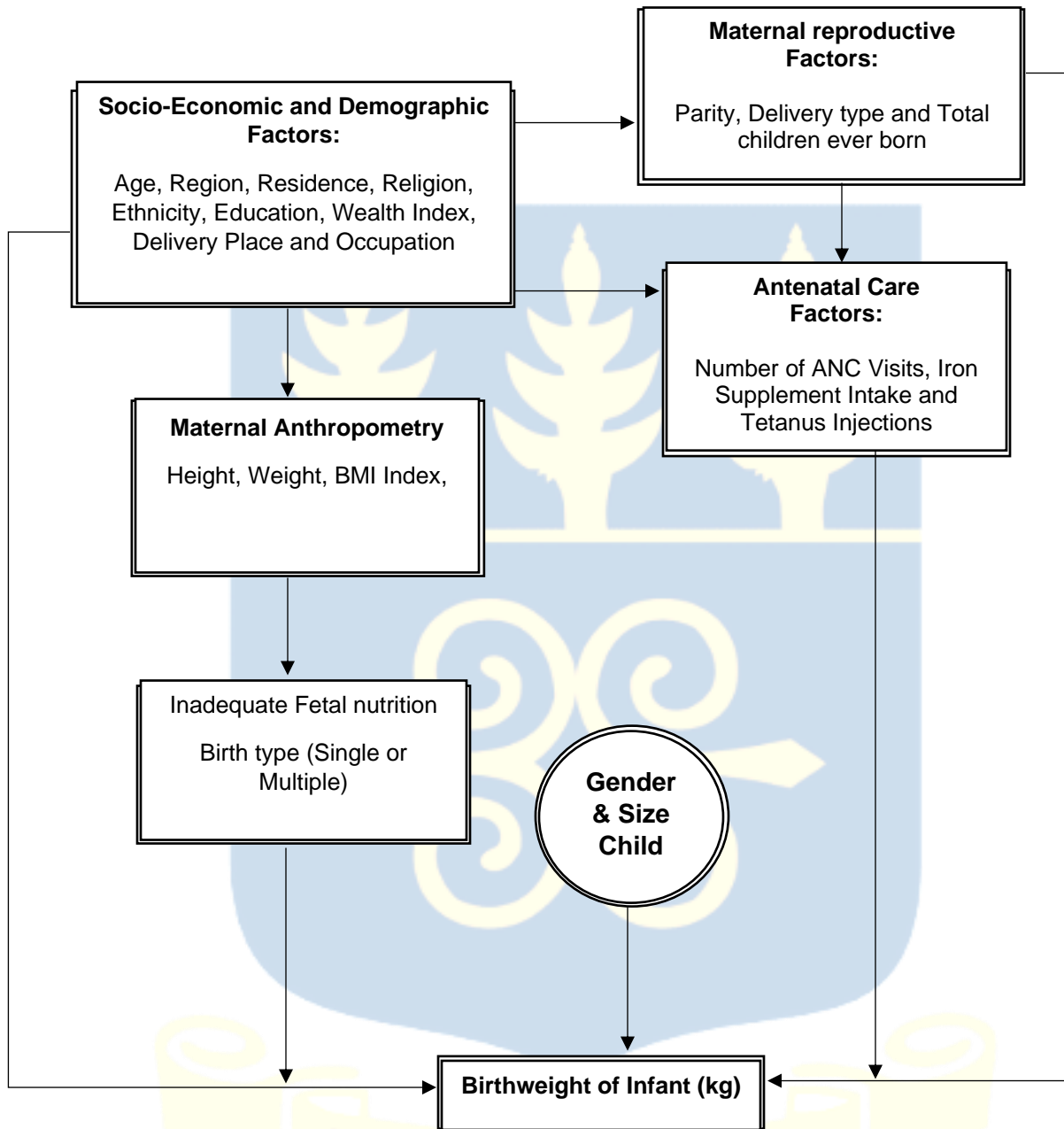
3.3.5 Developing the Screening tool

The discrimination of a screening tool is a measure of how well a model categorizes individuals by order of risk, such that individuals are correctly identified as being high-risk or low risk. Usually, discrimination is measured by plotting the receiver operating characteristic (ROC) curve with the calculation of the Area Under the ROC curve (AUROC) or c-statistic. Using the proportion of positive observations that are correctly considered as positive and the proportion of negative observations that are erroneously considered as positive, a tradeoff graphic is plotted to demonstrate the rate at which you can accurately predict something with the rate of inaccurately predicting something. In the y-axis, true positive rate (TPR) or sensitivity is displayed and x-axis false positive rate (FPR) or 1 – specificity is displayed, the curve approaching 1 represents the best performance of the model.

3.4 Conceptual framework

So as to decide the relationship between various factors and birthweight in Ghana, this conceptual framework was adopted. We hypothesized that birthweight of infants was likely going to be contributed by the following classes of factors, namely; social, demographics, mothers' reproductive behavior, maternal healthcare and health status including nutritional status (mother and newborn) factors. These factors may influence neonate's birthweight either directly or indirectly. Several factors which do not show direct associations with unfavorable birth outcomes contribute to these outcomes indirectly through intermediate factors. Based on the reviewed literature, the inter-relationship among various variables is shown in the flow chart below.

Figure 2: Conceptual framework for studying determinants of Birthweight



Source: Author (2019)

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.0 Introduction

We report our findings in the study in order to ascertain which variables and to what degree are responsible for birthweight prediction. The chapter intends to fulfill the objectives of the study and focuses on findings by analyzing our prospective data gathered in the survey. Analyses in the univariate sense, bivariate sense, and multivariate sense were conducted to find out the proportion of birthweight (low or normal) including influential factors associated with birthweight.

Data employed in the study are related to birthweight of infants incorporated in the pregnancy-related section of the individual questionnaire. As the study focus at determining significant risk factors with birthweight association, the covariates have been exclusively selected for consistency where possible. Further consideration was given to variables reported in works of literature to be associated with birthweight. With these two restraints, a limited number of covariates entered our model building process.

4.1 Birthweight Data Analysis

Initially, we consider birthweight dataset containing 18 variables each with 3361 observations except for mother's weight and height with missing values. The analysis was conducted in three phases incorporating the usage of descriptive (univariate) statistics and inferential (bivariate and multivariable) statistical methods. Details of analytical approaches used in the three phases are presented in chapter three.

4.1.1 Birthweight Distribution

Using the data, the first variable to be considered is the birthweight.

Table 6: Descriptive Statistics for Birthweight

	Obs.	Min	Max	Mean	Median	Std. Error	Std. Deviation
Birthweight	3361	.800	5.500	3.137	3.133	0.01	.5928
Valid N	3361						

In table 6, the mean birthweight for the study is 3.137kg with a standard error of 0.01kg and standard deviation of 0.593kg. The minimum birthweight recorded is 0.8kg while the maximum is 5.5kg. The median birthweight is 3.130kg which is approximately equal to the mean birthweight suggesting a possible nearing symmetric distribution.

Looking at the overall data, there seems to be evidence for normality, vide Histogram (Figure 3), but there may be some extreme cases as indicated by the Boxplot (See Appendix 3).

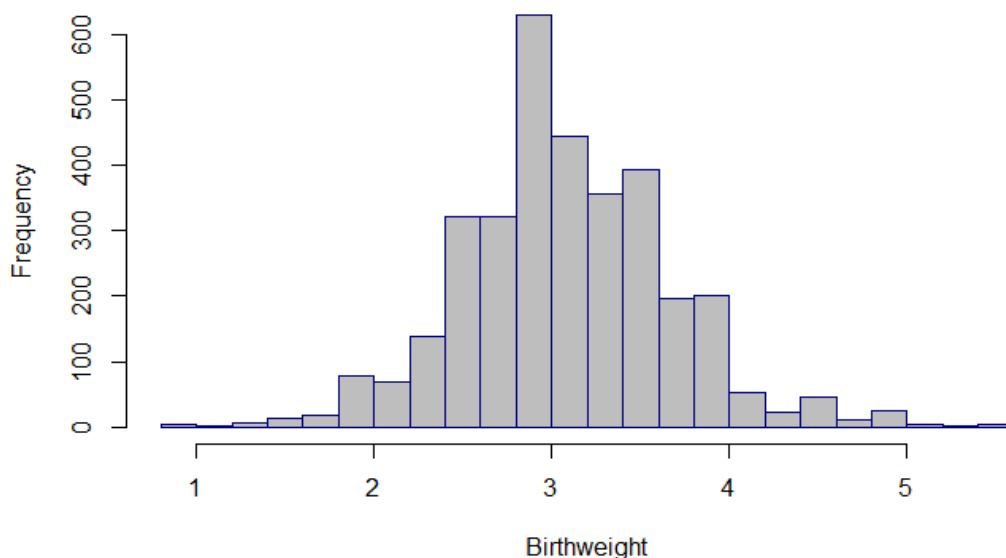


Figure 3: Histogram of Birthweight distribution (kg)

From figure 3 above, a large cluster of data points seems to be found in the 2.50kg to 4.0kg range of birthweight.

4.1.2 Classification of Birthweight

In the study, newborn classification based on birthweight was in two parts: Low birthweight (LBW) as birthweight less 2.5 kilograms (less 2,500g) and Normal birthweight (NBW) considered as birthweight above 2.499 kilograms.

Table 7: Descriptive Statistics of birthweight-based classification

Count	Mean	Standard Deviation	Minimum	Maximum	95.0% Lower CL for Count	95.0% Upper CL for Count
335	2.118	.293	.80	2.499	2.086	2.149

a. Low birthweight neonates

Count	Mean	Standard Deviation	Minimum	Maximum	95.0% Lower CL for Count	95.0% Upper CL for Count
3026	3.25	.503	2.50	5.50	3.232	3.27

b. Normal birthweight neonates

4.2 Prevalence rate of Low birthweight (LBW)

4.2.1 WHO's Recommendation

The official definition of low birthweight is birthweight of less than 2500g (World Health Organization, 2004a). With the study's aim to explore the prevalence statistics on low birthweight, of 3361 women, 9.9% delivered low birthweight infants.

4.2.2 Researcher's Adjusted Prevalence

It was observed from the DHS data there was birthweight heaping. With this kind of problem, the researcher employed a new birthweight prevalence method calculation. From the data, there were 194 infants who were reported to weigh exactly 2500g.

Many of these infants may actually have a birthweight of less than 2500g and therefore should be classified as having low birthweight. The exclusion of these infants will have an influence over the estimated proportion of infants with low birthweight. Per the researcher's analysis, of 3361

infants, 529 births had low birthweight, which reports an adjusted low infant weight at 15.7%, a 5.8% increase in the WHO’s low birthweight recommendation.

4.3 Mother’s Reporting of Birth Size Reliability Analysis

4.3.1 How accurate is a mother at reporting child’s birthweight?

One interesting question asked on the DHS is “the mother is asked for her assessment of the size of her baby at birth”. Size at the birth of a child has been utilized in certain studies as a proxy for birthweight (Ghosh, 2006; Magadi et al., 2007). However, the exact relationship existing between birthweight and perception of infant size is not yet established in Ghana. The point is to investigate whether a mother’s perception of the infant’s size gathered in a retrospective survey like Demographic and Health Survey is reliable and valid to be considered as a proxy for birthweight when missing.

The sizes of a child at birth are recorded as ‘very small’, ‘smaller than average’, ‘average’, ‘large than average’ and ‘very large’. Recategorization of very small and smaller than average was recoded as “Small size at birth” and likewise average, larger than average and very large as “Large birth size”. These two new classifications formulated was compared to the birthweight reported to assess the perception of the mother on a child’s weight.

Table 8: Birthweight groups against Size of child Crosstabulation

		Size of child		Total
		Low (Small size)	Normal (Large size)	
Low birthweight	Count	178	157	335
	% within groups	53.1%	46.9%	10.0%
Normal birthweight	Count	389	2637	3026
	% within groups	12.9%	87.1%	90.0%
Total	Count	567	2794	3361
	% within groups	16.9%	83.1%	100.0%

Pearson $\chi^2(1) = 348.9247$ Pr = 0.000

Chi-square test was utilized to assess if there exists an association between the size of child at birth and birthweight recorded, with this association observed at 5% alpha level. From the result, since our p-value (0.000) calculated is less than the significance level of (0.05), we must reject our null hypothesis and make the conclusion that the variables are associated. Hence, this result is evidence that birth size of neonates as reported by the mother might be a decent proxy of birthweight. (See Table 8)

4.3.2 Assessment of Birthweight verse Size of an infant at birth

To ascertain our accuracy of a mother's assessment of the size of a child, it was important to pass judgment on the size during childbirth against recorded birthweight. The aim is to discover whether a child who is declared as being normal size by the mother is indeed normal with reference to newborn child actual birthweight. Likewise, it is additionally imperative to perceive what number of individuals who are classified as being small would likewise be named being of low birthweight. If there is close agreement between low weight and small size of child assessment, at that point this may enable size during childbirth to be utilized in the calculation of birthweight estimates in Ghana.

The following analyses were reported from the contingency table:

- **True Positive (TP):** Proportion of children having a birthweight below 2.50 kg and reported as of small size is 53.1%
- **False Positive (FP):** Proportion of children having a birthweight below 2.50 kg but reported as of large size is 46.9%
- **False Negative (FN):** Proportion of children having a birthweight above 2.50 kg but reported as of small size is 12.9%

- **True Negative (TN):** Proportion of children having a birthweight above 2.50 kg and reported as of large size is 87.1%
- **Sensitivity:** The strength of birth size to identify correctly children with low birthweight (LBW- small size) as true under-recorded birthweight. It is calculated as:

$$\text{Sensitivity} = \frac{178}{178 + 389} = 0.314 \approx 31.4\%$$

- **Specificity:** The strength of birth size to identify correctly children with normal birthweight as true under-recorded birthweight. It is calculated as:

$$\text{Specificity} = \frac{2637}{157 + 2637} = 0.944 \approx 94.4\%$$

- **Positive predictive value (PPV),** that is our proportion of children of small size who remained low birthweight as per recorded birthweight. It is calculated as:

$$\text{PPV} = \frac{178}{178 + 157} = 0.531 \approx 53.1\%$$

- **Negative predictive value (NPV),** that is our proportion of children of normal size who remained normal as per recorded birthweight. It is calculated as:

$$\text{NPV} = \frac{2637}{389 + 2637} = 0.871 \approx 87.1\%$$

4.3.3 Calculation of Kappa Statistics

From Table 9 presented below, the kappa statistics tells us that if each mother had made her determination randomly, we would expect two mothers to agree on 75.53% of their infants' weight. In fact, they agreed on 83.75% of the child's weight or 30.80% of the way between random agreement and perfect agreement. Moreover, the amount of agreement ($\text{Prob} > Z = 0.000$) which is less than our alpha value (0.05) indicates that the hypothesis of mothers making their

determinations randomly can be rejected. Therefore, we conclude that mother’s perception of their neonate’s size is a true reflection of the child’s actual birthweight.

Table 9: Kappa Statistics for the contingency table

Agreement	Expected Agreement	Kappa	Standard Err.	Z	Prob>Z
83.75%	76.53%	0.3080	0.0165	18.68	0.000

4.3.4 Reporting of Birthweight: (Mother’s Memory Recall or Health Card)

For neonates with a recorded birthweight in our study, their reporting method was noted. These two reporting techniques are ‘mother’s memory recall’ or ‘read directly from health card’. In Ghana, not long after childbirth when a newborn is weighed, a health card is issued to the mother containing all significant data with respect to the newborn including birthweight. The distribution of the birthweights related to the reporting method, either via card or from memory can be investigated to distinguish whether there is any distinction among the distribution of birthweight by the two techniques of reporting.

The sample size, mean birthweights, their corresponding standard deviations and standard errors for each reporting technique. The study showed a higher mean birthweight for memory recall (3.2304kg ± 0.7292kg) as compared to mean weight on card reports (3.1114kg ± 0.5472kg). (See Appendix 1)

The p-value (0.000) associated with the two-sample t-test utilized to test if there exists a difference between the mean birthweights by reporting type. The analysis showed a significant difference among the mean birthweights reported from health card against the mean birthweights from mothers recall at a significant difference of 5% level. Based on the result, it might be hypothesized

that birthweights extracted from a health card should be more accurate than those birthweights extracted from memory recall. (See Appendix 2)

4.4 Bivariate analysis of study factors associated with birthweight

The researcher starts by looking at the epidemiological factors affecting birthweight in the socio-demographic sense, physical characteristics of the mother, antenatal care availability, obstetric history as well as morbidities during pregnancy. At this point, before we utilize Multiple Linear and Logistic regression to make conclusions about the data, each of these characteristics is evaluated to establish relationships between the maternal indicators and birthweight. A crosstab will put in effect a preliminary explanatory analysis as significant covariates are categorical. Then after, to get an idea about the association between different explanatory variables and dependent variable, Chi-square test performed.

4.4.1 Factor One: Outcome of Pregnancy

The outcome of the current pregnancy is studied to see the factors that may be influencing birthweight. The important variables are the sex of the newborn baby, birth size of neonates and the mode of delivery (either by cesarean). The distribution of birthweight regarding outcomes of pregnancy is presented below.

4.4.1.1 Gender of child

4.4.1.1.1 Analysis of Gender of a child

Gender of a child is a binary variable, either male or female, defined by the biological sex of the child.

Table 10: Tabulation of Gender of child

Gender of child	Obs.	Mean	Standard Dev.	Min	Max
Male	1,749	3.195	0.596	0.8	5.5
Female	1,612	3.074	0.583	1.0	5.5

Looking at Table 10, birthweight of male children is slightly higher compared to female children with respectively mean birthweight of $3.195\text{kg} \pm 0.596\text{kg}$ and $3.074\text{kg} \pm 0.583\text{kg}$.

Table 11: Descriptive Analysis for Gender of a child across birthweight classifications

		Low birthweight				Normal birthweight			
		Obs.	Mean	Stand. dev.	Birthweight %	Obs.	Mean	Stand. dev.	Birthweight %
Gender of child	Female	187	2.13	.28	5.6%	1425	3.20	.49	42.4%
	Male	148	2.10	.31	4.4%	1601	3.30	.51	47.6%

As already shown in Table 11, male infants constitute 52.0% (1749 infants) of the sample whereas female infants reported 48.0% (1612 infants). This means there are more male newborn children than female babies in the study. The minimum birthweight recorded for male infants is 0.8kg whereas the maximum is 5.5kg. Likewise, the minimum and maximum birthweights recorded for female infants are 1.0kg and 5.5kg respectively.

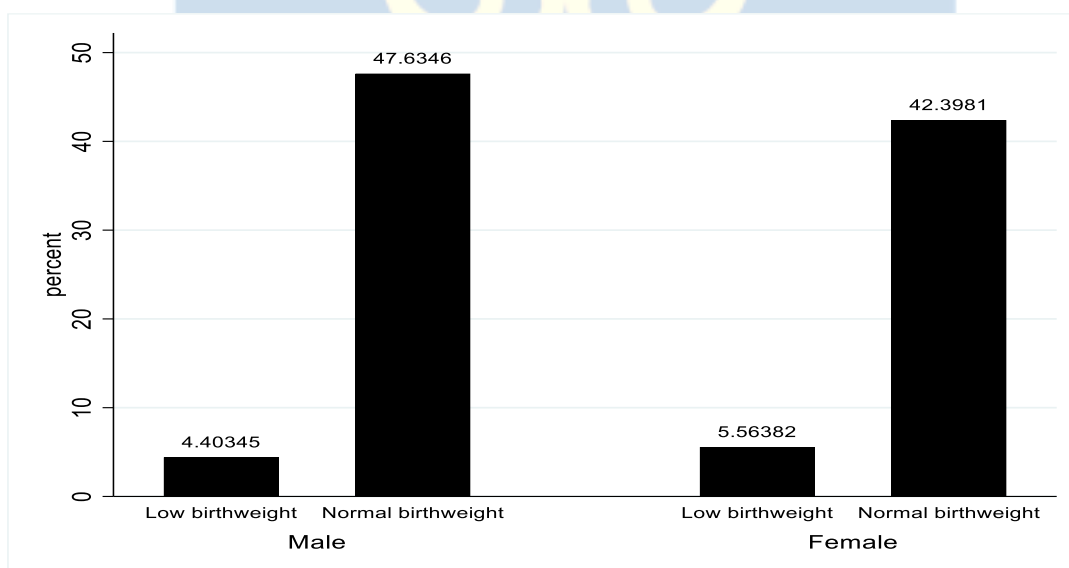


Figure 4: Birthweight classification Bar plot according to gender

Assessment of the group bar chart presented in figure 4 shows the category of low birthweight infants who (weighed less than 2.5kg at birth) is 335 (9.97%) while those with normal birthweight who (weighed above 2.499kg) is 2981 (90.03%) of the total sample size.

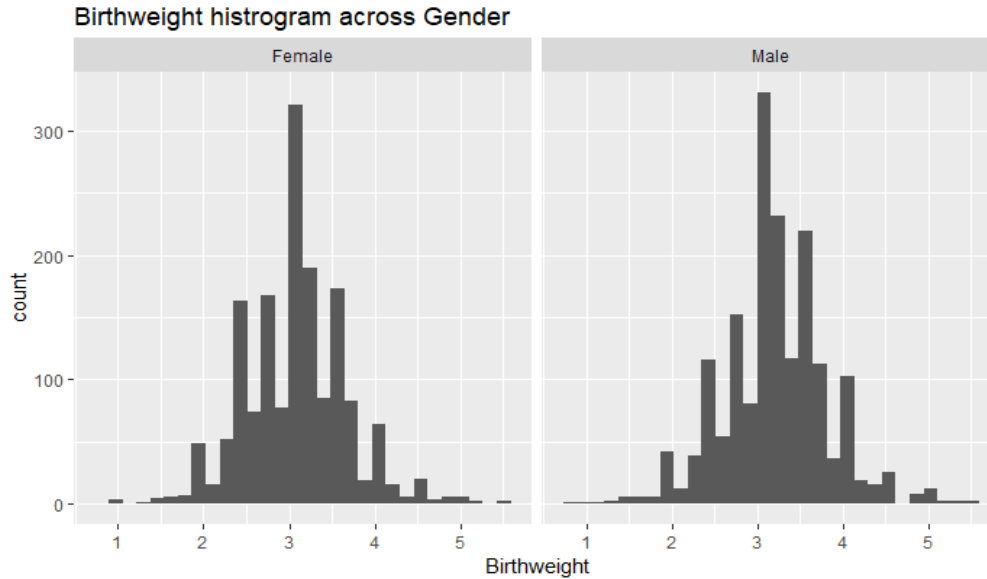


Figure 5: Birthweight histogram across gender

From the histogram plot above in Figure 5, we can see clearly that the levels in males are on average higher than the levels of female infants.

4.4.1.1.2 Association between birthweight classification and gender of the child

H₁₀: There exists no significant association between the gender of the child and newborn birthweight

H₁₁: There exists a significant association between Gender of child and newborn birthweight.

Table 12: Chi-Square Tests for Gender of a child against birthweight classification

	Value	Df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.208 ^a	1	.002		
Continuity Correction ^b	8.861	1	.003		
Likelihood Ratio	9.207	1	.002		
Fisher's Exact Test				.003	.001
Number of Valid cases	3361				

The result of the Chi-square test as presented in Table 12 was performed at 5% level of significance. The Pearson Chi-square significant value is reported as 0.002 with 1 degree of freedom. Therefore, the null hypothesis faces rejection and hence, a significant association between the gender of the child and newborn birthweight exist. It might be concluded that the gender of a child and newborn birthweight are not independent of each other.

Table 13: Symmetric Measures for Gender of child and birthweight classification

		Value	Approximate Significance
Nominal by Nominal	Phi	.107	.002
	Cramer's V	.107	.002
Number of Valid cases		3361	

In Table 13, measures of indexes of the agreement are utilized to evaluate the strength of our association. The value of this measure as per the Chi-square table is significant with 0.002 and the degree of association between these two variables is 10.7%.

4.4.1.2 Type of delivery by cesarean

4.4.1.2.1 Analysis of the type of delivery by cesarean

In here, the “YES” group includes those deliveries conducted through cesarean section (delivery with the surgical procedure). It is seen that the average birthweight of infants delivered by cesarean section was higher than those infants who were delivered by normal mode (delivery outside of cesarean section) with respective mean birthweights of 3.225 kg ± 0.710kg and 3.122kg ± 0.569kg (See Table 14).

Table 14: Tabulation of type of delivery by Caesarean

		Obs.	Mean	Standard Deviation
Type of delivery by cesarean	No	2867	3.122	.569
	Yes	494	3.225	.710

In Table 15, the distribution of the study about the type of delivery by cesarean is illustrated. There were 279 (8.3%) and 56 (1.7%) of low birthweight incidences compared to 2,588 (77.0%) and 438 (13.0%) of normal birthweight cases that gave births through cesarean section and normal modes of delivery respectively.

Table 15: Tabulation of type of delivery by Caesarean

			Low birthweight	Normal birthweight	
Type of delivery by cesarean	No	Obs.	279	2588	2867
		% of Total	8.3%	77.0%	85.3%
	Yes	Obs.	56	438	494
		% of Total	1.7%	13.0%	14.7%
Total		Obs.	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

4.4.1.2.2 Association between birthweight and type of delivery by cesarean

The hypothesis for the data shown in Table 15 for the Chi-square test is:

H_{2_0} : There exists no significant association between type of delivery by cesarean and newborn birthweight.

H_{2_1} : There exists a significant association between type of delivery by cesarean and newborn birthweight.

Table 16: Chi-square tabulation for Type of delivery by cesarean and birthweight

	Value	Df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.209 ^a	1	.272		
Continuity Correction ^b	1.037	1	.309		
Likelihood Ratio	1.171	1	.279		
Fisher's Exact Test				.290	.154
Number of Valid cases	3361				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 49.24.

b. Computed only for a 2x2 table

In Table 16, the Chi-square test performed at a 5% level of significance. Here, the Pearson Chi-square significant value is 0.272 with 1 degree of freedom. Therefore, the null hypothesis fails rejection and hence, no significant association between type of delivery by cesarean and newborn birthweight exist. It might be concluded that the type of delivery by cesarean and newborn birthweight are independent of each other. Since there exists no association between type of delivery by cesarean and birthweight, the measure of the strength of association was not computed.

4.4.1.3 Size of the child at birth

4.4.1.3.1 Analysis of the size of the child at birth

Size of an infant at birth is a nominal variable which took the form of the mother placing the child into one of five size categorizations: very small, smaller than average, average, larger than average or very large. This information was asked of newborn babies born in the last five years preceding the survey in order to minimize recall bias.

Table 17: Descriptive analysis of Size of the child

		Obs.	Mean	Standard Deviation
Size of Child	Very small	91	2.447	.640
	Smaller than average	303	2.693	.515
	Average	749	3.010	.471
	Larger than average	819	3.256	.510
	Very large	393	3.631	.543

In table 17, with each perception of size classification, the mean birthweight was computed. The mean birthweight in each classification of size follows the expected trend within each classification. The mean birthweight in the very large classification is the heaviest recording of 3.631kg \pm 0.543kg, with the mean in each subsequent classification being lighter than the preceding. Additionally, birthweight in the very small classification recorded the lowest with mean weight and standard deviation as 2.447kg \pm 0.640kg.

4.4.1.3.2 Association between the type of delivery by Size of child and birthweight

H_{3_0} : There exists no significant association between the size of newborn and newborn birthweight.

H_{3_1} : There exists a significant association between the size of newborn and newborn birthweight.

Table 18: Chi-Square Tests for Size of child and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	450.737 ^a	4	.000
Likelihood Ratio	367.062	4	.000
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 12.66.

The results of the Chi-square test as presented in Table 18 was performed at 5% level of significance. The Pearson Chi-square significant value is 0.000 with 4 degrees of freedom. Therefore, the null hypothesis is rejected and hence, a significant association between Size of child and newborn birthweight exist. It might be concluded that the size of a child and newborn birthweight are not independent of each other.

Table 19: Symmetric Measures for Size of child and birthweight classification

		Value	Approximate Significance
Nominal by Nominal	Phi	.366	.000
	Cramer's V	.366	.000
Number of Valid cases		3361	

In Table 19, measures of indexes of the agreement are utilized to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.000 and the degree of association between the two variables is 36.6%.

4.4.2 Factor Two: Socio-Economic and Demographic Factors

4.4.2.1 Mothers Age

4.4.2.1.1 Analysis of Mothers Age

Mother's age is a continuous measure of the age in years at the time of interview. It was calculated from a century month code (CMC) using the birth date and the interview date. In order to facilitate the analysis, it was later recorded from its earlier continuous state to a categorical variable using a five years interval of age as 15 to 19 years, 20 to 24 years, 25 to 29 years, 30 to 34 years, 35 to 39 years, 40 to 44 years and ≥ 45 years.

Table 20: Descriptive Statistics of Mother Age

	N	Minimum	Maximum	Mean	Median	Std. Deviation
Mother Age	3361	15	49	30.36	30.00	6.551
Valid N (listwise)	3361					

From Table 20, the mean age of women whose infants were studied is 30.36 (approximately 30) years ranging from 15 to 49 years with a median age of 30 years.

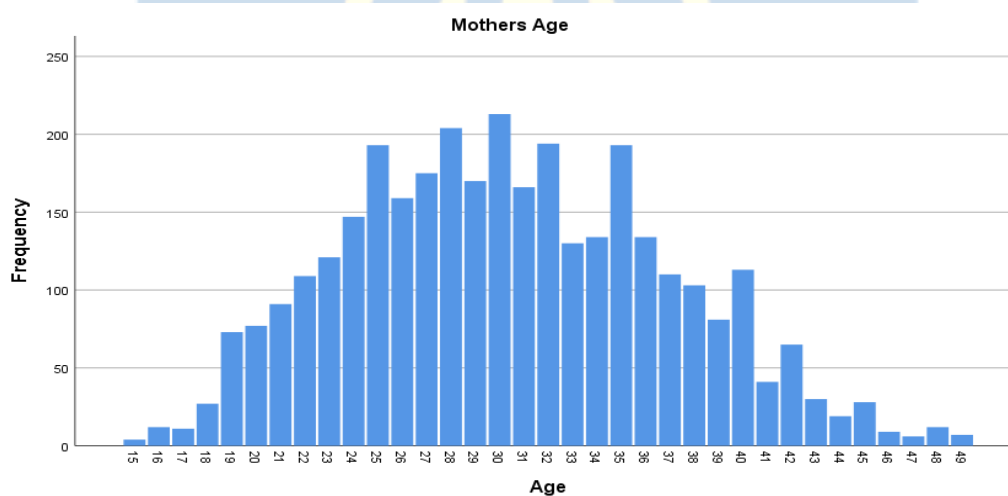


Figure 6: Age distribution of mothers

Figure 6 above exhibit that most of the births occurred between the age of 18 years to 40 years.

Table 21: Descriptive statistics of Mothers Age by group and Birthweight

		Obs.	Mean	Standard Deviation
Mother Age	15-19	127	2.978	.596
	20-24	545	3.051	.593
	25-29	901	3.123	.592
	30-34	837	3.159	.591
	35-39	621	3.204	.588
	40-44	268	3.200	.554
	45-49	62	3.180	.691

Table 21 above depicts the mean birthweight of infants across the various categories of age intervals. It is shown that the age interval of 15-19 years recorded the lowest mean weight of 2.978kg while the class with the highest birthweight recorded was the group of 35-39 years accounting for 3.204kg.

Table 22: Descriptive and Crosstabulation for Mothers Age and Birthweight classification

			Low birthweight	Normal birthweight	
Mother Age Group	15-19	Obs.	19	108	127
		% of Total	0.6%	3.2%	3.8%
	20-24	Obs.	67	478	545
		% of Total	2.0%	14.2%	16.2%
	25-29	Obs.	87	814	901
		% of Total	2.6%	24.2%	26.8%
	30-34	Obs.	84	753	837
		% of Total	2.5%	22.4%	24.9%
	35-39	Obs.	53	568	621
		% of Total	1.6%	16.9%	18.5%
	40-44	Obs.	20	248	268
		% of Total	0.6%	7.4%	8.0%
	45-49	Obs.	5	57	62
		% of Total	0.1%	1.7%	1.8%
Total		Observation	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

The majority 901(26.8%) of the mother's range between the ages of 25-29 years. Also, there were 127(3.8%), 545(16.2%), 837(24.9%), 627(18.5%), 268(8.0%) and 62(1.8%) of mothers in the year

groups 15-19, 20-24, 30-34, 35-39, 40-44 and ≥ 45 respectively (Table 24). The mean age of the study was observed at 30.36 ± 6.55 years old. Of the total observation of low birthweight infants recorded, 19(5.7%) of the observation were delivered by mothers in the age category 15-19 years, 67 (20.0%) were delivered by mothers in the age category 20-24 years, 87(26.0%) of low birthweight infants were delivered by women in the age category between 25-29 years, 84(25.1%) were delivered by mothers from the age category 30-34 years, 53 (15.8%) were delivered by mothers from the age category 35-39 years, 20 (6.0%) were delivered by mothers in the age category 40-44 years and 5(1.5%) low birthweight infants were delivered by women of age above 44 years. (See Table 22)

4.4.2.1.2 Association between newborn birthweight and maternal Age

The hypothesis for the data shown in Table 22 for the Chi-square test is:

H_{4_0} :: There exists no significant association between mothers age and newborn birthweight.

H_{4_1} : There exists a significant association between mothers age and newborn birthweight.

Table 23: Chi-Square Tests for Mothers age and birthweight

	Value	Df	Asymptotic Significance (2-sided)
Pearson Chi-Square	10.461 ^a	6	.107
Likelihood Ratio	10.079	6	.121
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 6.18.

In Table 23, the Pearson Chi-square test was performed at a 5% level of significance shows a significant value of 0.107 with 6 degrees of freedom. Therefore, the null hypothesis fails rejection and hence, no significant association among mother age and newborn birthweight exist. It might be concluded that maternal age and newborn birthweight are independent of each other.

4.4.2.2 Maternal Education

4.4.2.2.1 Analysis of maternal education

Mother’s education is an ordinal categorical variable defining the highest educational level.

Categories include primary, secondary, higher and no education.

Table 24: Mean birthweight across Maternal Educational level

		Birthweight		
		Obs.	Mean	Standard Deviation
Educational Level	No education	853	3.124	.563
	Primary	622	3.132	.603
	Secondary	1672	3.142	.605
	Higher	214	3.165	.587

The mean weight descriptive analysis in table 24 shows that as the mother’s educational level goes up, so as the weight of the infant becomes higher. The proportion of the study population by maternal educational level showed that the majority 1,679 (49.7%) of the mothers had secondary level education. Whiles, there were 18.5%, and 6.4% of mothers that accomplished the educational level up to Primary and Tertiary respectively and 25.4% of the respondent has no education. (See below in Table 25)

Table 25: Crosstab for mother education and birthweight

			Low birthweight	Normal birthweight	
Educational Level	No education	Obs.	81	772	853
		% of Total	2.4%	23.0%	25.4%
	Primary	Obs.	59	563	622
		% of Total	1.8%	16.8%	18.5%
	Secondary	Obs.	174	1498	1672
		% of Total	5.2%	44.6%	49.7%
	Higher	Obs.	21	193	214
		% of Total	0.6%	5.7%	6.4%
Total		Observation	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

In Table 25, of the total number of low birthweight infants reported, 81 (24.2%) of low weight infants were delivered by mothers with no education, 59 (17.6%) of low birthweight were delivered by mothers with primary level education. The majority, 174 (51.9%) of low birthweight infants were delivered by women who have attained education to the secondary level while 21(6.3%) low birthweight infants were delivered by women with tertiary attainment of education.

4.4.2.2.2 Association between newborn birthweight and mother education

The hypothesis for the data shown in Table 26 for the Chi-square test is:

H_{8_0} : There exists no significant association between maternal education and newborn birthweight.

H_{8_1} :: There exists a significant association between maternal education and newborn birthweight.

Table 26: Chi-Square tests between Mother education and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.738 ^a	3	.864
Likelihood Ratio	.738	3	.864
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 21.33.

The results of the Chi-square test as presented in Table 26 was performed at 5% level of significance and obtained a calculated significant value of 0.864 with 3 degrees of freedom. Therefore, the null hypothesis fails rejection and hence, no significant association existed between mother education and newborn birthweight. It might be concluded that mother education and newborn birthweight are independent of each other.

4.4.2.3 Maternal Region

4.4.2.3.1 Analysis of Region of respondent

The region is a categorical variable with 10 categories which aligns with the administrative regions of Ghana.

Table 27: Mean birthweight across various Regions

		Obs.	Mean	Standard Deviation
Region	Ashanti	436	3.158	.586
	Brong Ahafo	406	3.123	.555
	Central	325	3.123	.607
	Eastern	261	3.153	.716
	Greater Accra	358	3.213	.594
	Northern	274	3.012	.572
	Upper East	387	3.064	.546
	Upper West	297	2.951	.463
	Volta	302	3.325	.598
	Western	315	3.231	.618

From Table 27, the region with the highest mean infants' weight was Volta with 3.231kg \pm 0.618kg while the region that recorded the lowest birthweight was Upper West accounting for 2.951kg \pm 0.463kg. Brong Ahafo and Central region had the same mean weight of 3.123kg but a varied standard deviation of 0.555kg and 0.607kg respectively.

Table 28: Region and Birthweight Crosstabulation

			Low birthweight	Normal birthweight	
Region	Ashanti	Obs.	44	392	436
		% of Total	1.3%	11.7%	13.0%
	Brong Ahafo	Obs.	44	362	406
		% of Total	1.3%	10.8%	12.1%
	Central	Obs.	29	296	325
		% of Total	0.9%	8.8%	9.7%
	Eastern	Obs.	34	227	261
		% of Total	1.0%	6.8%	7.8%

Greater Accra	Obs.	24	334	358
	% of Total	0.7%	9.9%	10.7%
Northern	Obs.	37	237	274
	% of Total	1.1%	7.1%	8.2%
Upper East	Obs.	43	344	387
	% of Total	1.3%	10.2%	11.5%
Upper West	Obs.	36	261	297
	% of Total	1.1%	7.8%	8.8%
Volta	Obs.	19	283	302
	% of Total	0.6%	8.4%	9.0%
Western	Obs.	25	290	315
	% of Total	0.7%	8.6%	9.4%
Total	Obs.	335	3026	3361
	% of Total	10.0%	90.0%	100.0%

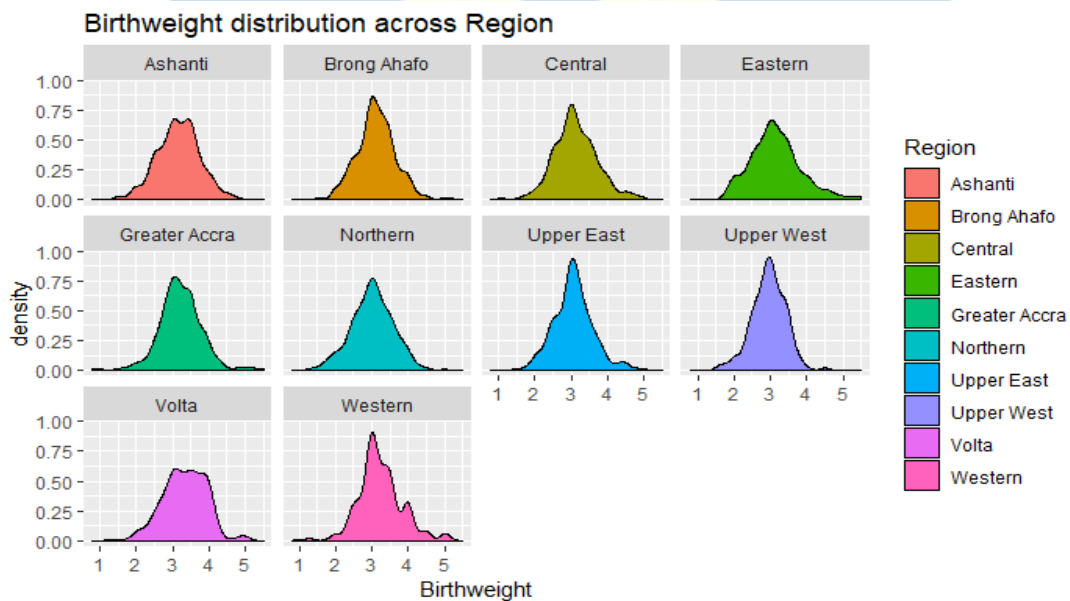


Figure 7: Birthweight distribution across various Regions

To graphically illustrate the birthweight across various regions, a normal distribution was performed. From figure 7 above, we could see that except for Western (having multimodal distribution), the distribution of birthweight across various regions seems to be similar.

4.4.2.3.2 Association between Region and newborn birthweight

The hypotheses for the data shown in table 28 for the Chi-square test is:

H_{5_0} : There exists no significant association between region and newborn birthweight.

H_{5_1} : There exists a significant association between region and newborn birthweight.

Table 29: Chi-Square Tests across regions and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	19.629 ^a	9	.020
Likelihood Ratio	20.183	9	.017
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 26.01.

The results of the Chi-square test as presented in Table 29 was performed at 5% level of significance. It obtained a significant value of 0.020 with 9 degrees of freedom. Therefore, the null hypothesis is rejected, indicating a significant association between mother region and newborn birthweight exist. It might be concluded that region and newborn birthweight are not independent of each other.

Table 30: Symmetric Measures for regions and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.076	.020
	Cramer's V	.076	.020
Number of Valid cases		3361	

From table 30, measures of indexes of the agreement are utilized to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.020 and the degree of association between two variables is 7.60%.

4.4.2.4 Type of Residence

4.4.2.4.1 Analysis of Type of residence

Type of residence is a binary variable that defines the residential place as either rural or urban setting. It is seen from the table that the average birthweight of infants delivered at urban cities was higher than those infants who were delivered at the rural areas with respective mean weights of $3.145\text{kg} \pm 0.595\text{kg}$ and $3.128\text{kg} \pm 0.590\text{kg}$.

Table 31: Tabulation of Type of residence

		Obs.	Mean	Standard Deviation
Type of Residence	Rural	1634	3.128	.590
	Urban	1727	3.145	.595

The distribution of subjects by residence indicates that 1634 (48.6%) of respondents were living in the rural whiles 1727 (51.4%) in the urban. For low birthweight and normal birthweight, 4.9% against 5.0% and 43.7% against 51.4% were living in rural and urban cities respectively (Table 31).

Table 32: Crosstab of Type of residence and birthweight

			Low birthweight	Normal birthweight	
Type of Residence	Rural	Obs.	166	1468	1634
		% of Total	4.9%	43.7%	48.6%
	Urban	Obs.	169	1558	1727
		% of Total	5.0%	46.4%	51.4%
Total		Obs.	335	3026	3361
		% of Total	9.9%	90.1%	100.0%

Table 32 depicts that, there is a 4.9% and 5.0% prevalence rate of low birthweight according to the type of maternal residence.

4.4.2.4.2 Association between newborn birthweight and type of residence

The hypotheses for the data shown in table 32 for the Chi-square test is:

$H7_0$: There exists no significant association between the type of residence and newborn birthweight.

$H7_1$: There exists a significant association between type of residence and newborn birthweight.

Table 33: Chi-Square Tests type of residence and birthweight

	Value	Df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.130 ^a	1	.718		
Continuity Correction	.092	1	.761		
Likelihood Ratio	.130	1	.718		
Fisher's Exact Test				.730	.381
Number of Valid cases	3361				

The Chi-square test was performed at a 5% level of significance. Our Pearson Chi-square significant value is 0.718 with 1 degree of freedom. Therefore, the null hypothesis fails rejection and thus indicating that there is no significant impact of the distribution of maternal residence on fetal birthweight. It might be concluded that mother residence and newborn birthweight are independent of each other. (See table 33)

4.4.2.4.3 Finding distribution of Birthweight across Region and Type of Residence

In trying to establish the birthweight distribution across regions and the type of residence, Figure 8 below depicts several distributions of which some regions have higher means of birthweight. Furthermore, it shows that the mean birthweight across regions with respect to the respondents living in rural areas have higher birthweight means as compared to the people living in urban cities.

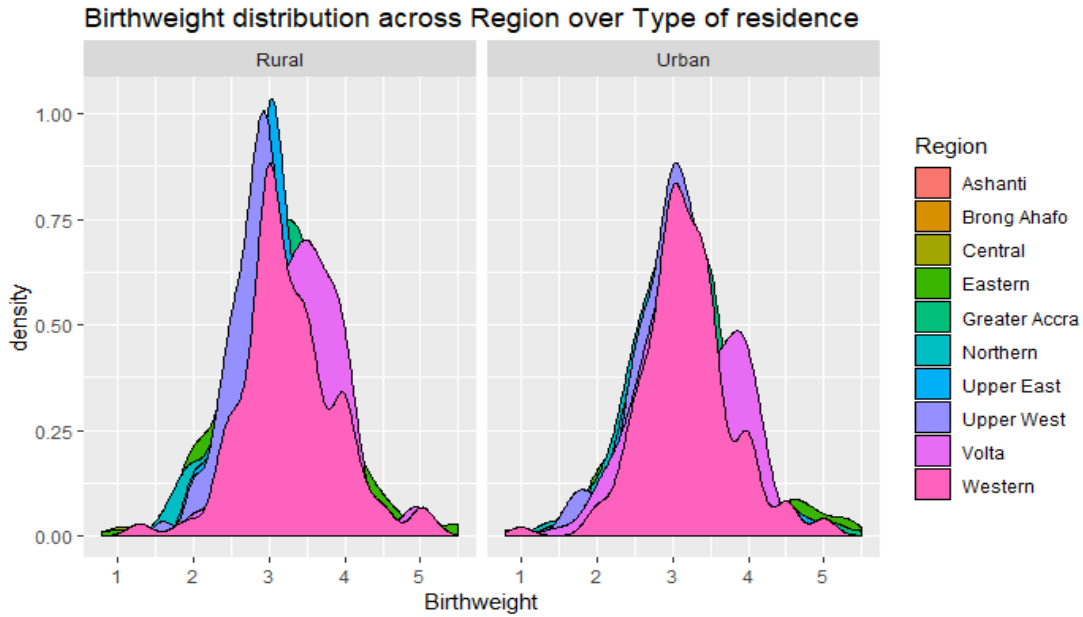


Figure 8: Distribution of birthweight across region and type of residence

4.4.2.5 Mother Religion

4.4.2.5.1 Analysis for Mothers Religion

Religion is a categorical nominal variable with three (3) divisions. Regarding religion, majority of the study are Christians constituting 2494(74.2%), followed by Islam, Other with respectively frequencies and percentages as 717(21.3%) and 150(4.5%) as displayed in Table 34 below.

Table 34: Religion and Birthweight Crosstabulation

			Low birthweight	Normal birthweight	
Religion	Christianity	Obs.	234	2260	2494
		% of Total	7.0%	67.2%	74.2%
	Islam	Obs.	87	630	717
		% of Total	2.6%	18.7%	21.3%
	Other	Obs.	14	136	150
		% of Total	0.4%	4.0%	4.5%
Total		Obs.	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

Further, table 37 indicates that women in Ghana who are Christians experience 7.0% likelihood of given birth to a low weight infant in comparison to the 2.6% Islam and 0.4% Other religion women.

4.4.2.5.2 Association between mother religion and birthweight

The hypotheses for the data shown in Table 34 for the Chi-square test is:

H_{9_0} : There exists no significant association between mother recode religion and birthweight.

H_{9_1} : There exists a significant association between mother recode religion and birthweight.

Table 35: Chi-Square Tests for mother recode religion and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	20.645	2	.008
Likelihood Ratio	22.091	2	.005
Number of Valid cases	3361		

Our Chi-square test was performed at a 5% level of significance as presented in Table 35. The Pearson Chi-square significant value is 0.008 with 2 degrees of freedom. Therefore, the null hypothesis is rejected. Hence, a significant association between mother religion and newborn birthweight exist. It might be concluded that mother religion and newborn birthweight are not independent of each other.

Table 36: Symmetric Measures of Religion and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.078	.008
	Cramer's V	.078	.008
Number of Valid cases		3361	

In Table 36, measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.008 and the degree of association between two variables is 7.80%.

4.4.2.6 Maternal Wealth Index

4.4.2.6.1 Analysis of Wealth index

Wealth quintile is presented as five ordinal categories. Regarding Demography and Health Survey, wealth is calculated as “a composite measure of a household's cumulative living standard” regarding possession of some selected assets including “televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities.” This data is used to determine a wealth index score, which is then broken into quintiles and presented in a separate variable.

Table 37: Mean weight of infants across the Wealth Index

		Birthweight		
		Obs.	Mean	Standard Deviation
Wealth index	Poorest (Lowest)	783	3.042	.573
	Poorer (Second)	593	3.145	.562
	Middle	686	3.199	.615
	Richer (Fourth)	662	3.142	.606
	Richest (Highest)	637	3.174	.594

The distributions in Table 37 depict the degree to which mean birthweight is distributed (evenly or unevenly) by wealth index. From the study, the group with the lowest mean infants' weights are women who are classified as poorest ($3.042\text{kg} \pm 0.573\text{kg}$) while the Middle-income earners of women had the highest mean birthweight of $3.199\text{kg} \pm 0.615\text{kg}$. This indicates a 4.91% increase in mean weight of the poorest against the middle-class women of the wealth index distribution. Respectively, the following mean infants' weight was recorded for poorer, richer and richest ($3.145\text{kg} \pm 0.562\text{kg}$, $3.142\text{kg} \pm 0.606\text{kg}$ and $3.174\text{kg} \pm 0.594\text{kg}$).

Table 38: Crosstabulation Wealth index and Birthweight group

			Low birthweight	Normal birthweight	
Wealth index	Poorest (Lowest)	Obs.	99	684	783
		% of Total	2.9%	20.4%	23.3%
	Poorer (Second)	Obs.	50	543	593
		% of Total	1.5%	16.2%	17.6%
	Middle	Obs.	60	626	686
		% of Total	1.8%	18.6%	20.4%
	Richer (Fourth)	Obs.	69	593	662
		% of Total	2.1%	17.6%	19.7%
	Richest (Highest)	Obs.	57	580	637
		% of Total	1.7%	17.3%	19.0%
Total	Observation		335	3026	3361
	% of Total		10.0%	90.0%	100.0%

With reference to Table 38, the proportion of the combined wealth index for the poorer and the poorest was 40.9% (n = 1,376) while the proportion of the combined wealth index of the middle, richer and richest constitute 61.9% (n = 1985) of the population. This illustrates that 40.9% of the population is below the middle wealth quintile. The prevalence of low birthweight among children born to mothers in the poorest (lowest) and fourth household quintiles of wealth (each above 20%) was higher than for those in the second quintile (14.9%), middle quintile (17.9%) and highest quintile (17.0%).

4.4.2.6.2 Association between newborn birthweight and household wealth index

The hypotheses for the data shown in Table 38 for the Chi-square test is;

H_{10_0} : There exists no significant association between wealth index and newborn birthweight.

H_{10_1} : There exists a significant association between wealth index and newborn birthweight.

Table 39: Chi-Square Tests for Wealth index and Birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	9.838 ^a	4	.043
Likelihood Ratio	9.542	4	.049
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 59.11.

Our Chi-square test was performed at a 5% level of significance and presented in Table 39. The Pearson Chi-square significant value is 0.043 with 4 degrees of freedom. Therefore, the null hypothesis is rejected. Hence, a significant association exists between wealth index and newborn birthweight. It might be concluded that mother wealth index and newborn birthweight are not independent of each other.

Table 40: Symmetric Measures for Wealth index and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.054	.043
	Cramer's V	.054	.043
Number of Valid cases		3361	

In Table 40, measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.043 and the degree of association between these two variables is 5.40%.

4.4.3 Factor Three: Maternal Anthropometry

To get an initial feel for the relationships between the anthropometry variables - and, in particular, between mother height or weight and birthweight, it was interesting to observe the scatterplots produced by plotting each variable against all others, as well as the specific distribution of birthweight values within each level of mother anthropometry characteristics.

4.4.3.1 Mothers Height

Mother height was taken subjectively from the mother as a continuous variable. Obtained data was recorded in the questionnaire as centimeters.

Table 41: Descriptive Statistics for Mothers Height

	Obs.	Minimum	Maximum	Mean	Std.
Mothers height	3361	139.20	184.60	159.30	4.16
Valid N (listwise)	3361				

From the data, the minimum height was 139.2 centimeters while the maximum height was 184.6 centimeters with a mean height of 159.30 centimeters \pm 4.16 centimeters. (See Table 41)

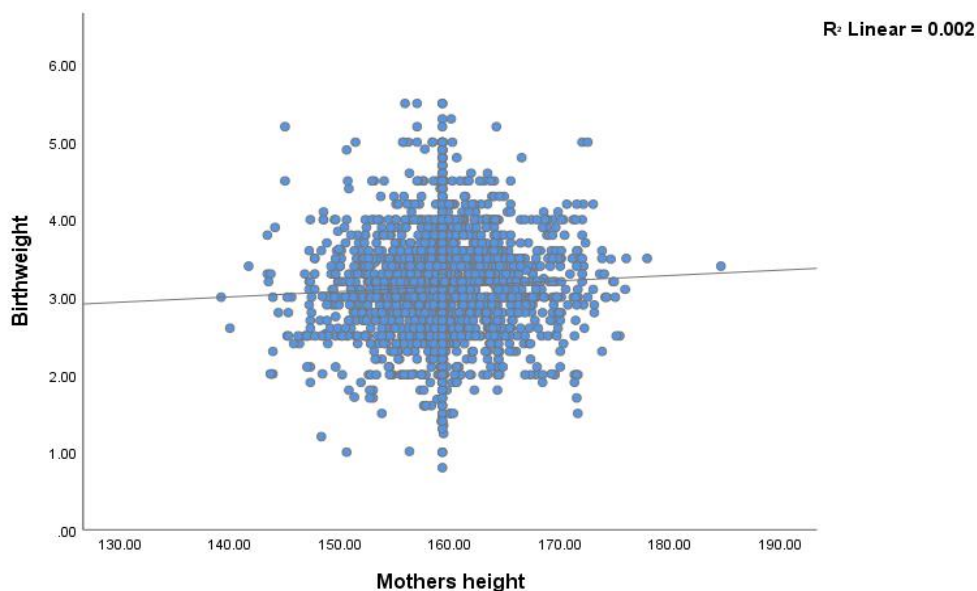


Figure 9: Simple Scatter with Fit Line of Birthweight by Mothers height

It is clear from the scatterplot above in Figure 9 that the birthweight does not depend on mother height very strongly, however, there seems to be a slight association between the two.

4.4.3.2 Mothers weight

Mother weight was likewise taken subjectively from the mother as a continuous variable. Obtained data was recorded in the questionnaire as kilograms.

Table 42: Descriptive Statistics for Mother weight

	Obs.	Minimum	Maximum	Mean	Standard dev.
Mothers weight	3361	38.90	159.40	63.60	9.70
Valid N (listwise)	3361				

From the data, the minimum weight was 38.90kg while the maximum weight was reported as 159.40kg with a mean weight of 63.60kg \pm 9.70kg. (See Table 42)

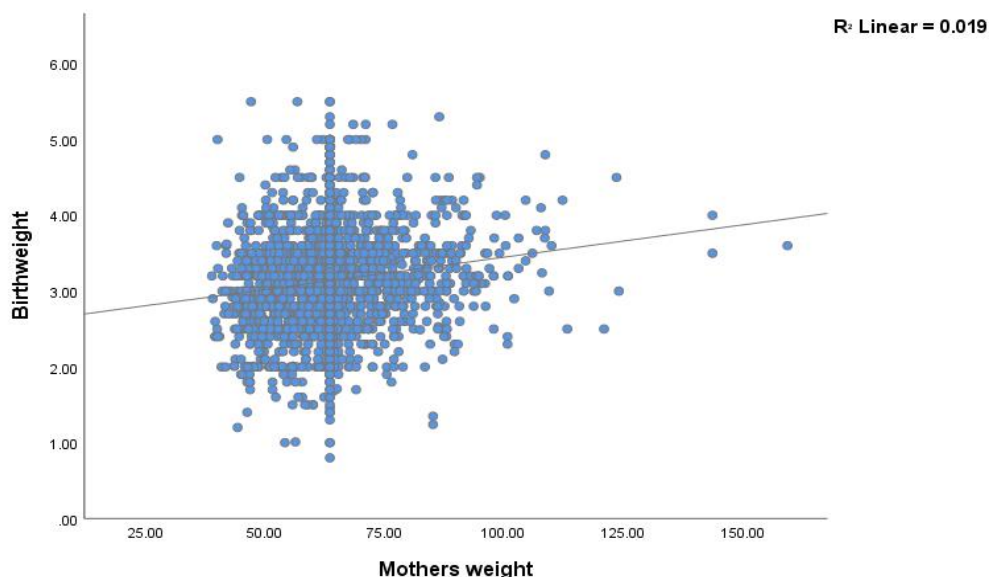


Figure 10: Simple Scatter with Fit Line of Birthweight by Mothers weight

Likewise, it is clear from the scatterplot above in Figure 10 that the birthweight does not depend on mother weight very strongly, however, there seems to be a slight association between the two.

4.4.4 Factor Four: Maternal Reproductive Factors

4.4.4.1 Birth order (Parity)

4.4.4.1.1 Analysis of Birth order

Birth order is a discrete-continuous variable corresponding to the numerical order in that the infant of interest was born. In order to facilitate our analysis, mothers were divided as primipara (single

pregnancy) and multipara (2 or more pregnancy) such as 2 to 4, 5 to 8 birth order and above 8 birth order.

Regarding Table 43 below, the proportion of low birthweight occurring within the first birth order is 3.5% (n = 118) while the proportion of the combined birth order above first child is reported as 6.4% (n = 217) of the population.

Table 43: Crosstab for Birth order (Parity) and birthweight

			Low birthweight	Normal birthweight	
Birth order	1	Obs.	118	777	895
		% of Total	3.5%	23.1%	26.6%
	2-4	Obs.	161	1627	1788
		% of Total	4.8%	48.4%	53.2%
	5-8	Obs.	52	581	633
		% of Total	1.5%	17.3%	18.8%
	Above 8	Obs.	4	41	45
		% of Total	0.1%	1.2%	1.3%
Total		Obs.	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

4.4.4.1.2 Association between newborn birthweight and birth order

The hypotheses for the data shown in Table 43 for the Chi-square test is:

H_{11_0} : There exists no significant association between birth order (parity) and newborn birthweight.

H_{11_1} : There exists a significant association between birth order (parity) and newborn birthweight.

Table 44: Chi-Square Tests for Birth order (Parity) and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	14.394 ^a	3	.002
Likelihood Ratio	13.724	3	.003
Number of Valid cases	3361		

The results of the Chi-square test as presented in Table 44 with a 5% level of significance. Our Pearson Chi-square significant value is 0.002 with 1 degree of freedom. Therefore, the null hypothesis is rejected. Hence, a significant association exists between birth order of infant and newborn birthweight. It might be concluded that birth order and newborn birthweight are not independent of each other.

Table 45: Symmetric Measures for Birth order (Parity) and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.065	.002
	Cramer's V	.065	.002
Number of Valid cases		3361	

In Table 45, measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.002 and the degree of association between two variables is 6.50%.

4.4.4.2 Total number of children ever had or born

4.4.4.2.1 Analysis of Total Children ever had

The total children ever born is a measure of the children born alive to a woman which is an attribute of the counting of unit 'person'. In this study, the variable was applied to women 15 years of age and over but excluded mothers with no child (none). However, it used a standard single-level classification with thirteen categories.

Table 46: Tabulation of Total children ever had against Birthweight classification

	N	Minimum	Maximum	Mean	Std. Deviation
Total Children ever born	3361	1	13	3.23	1.943
Valid N (listwise)	3361				

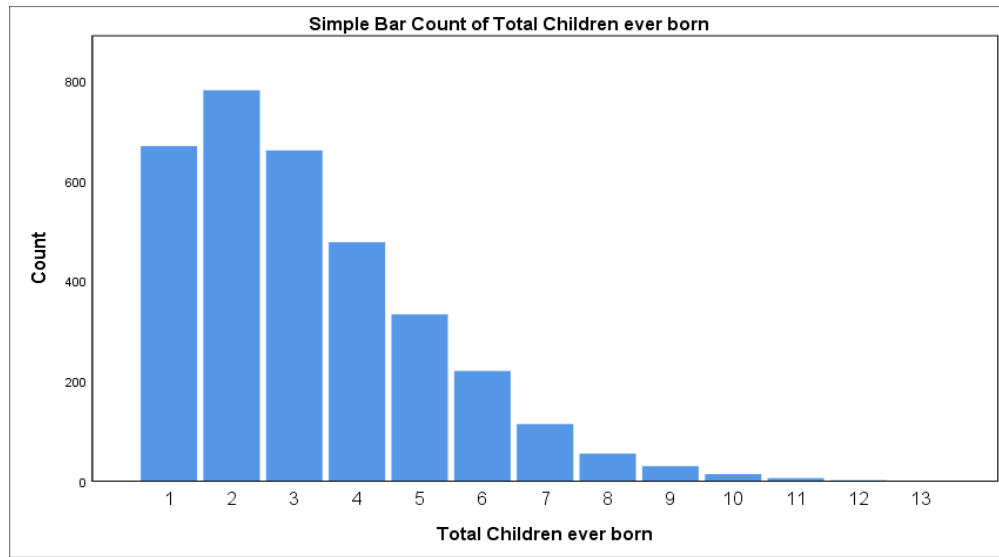


Figure 11: Bar Plot of Total Children ever born/had

Because of huge disparity among the counts of the different total children ever born, the above bar plot may not be very informative. Looking at Figure 11 above, birthweights across various total children ever had shown that at categories above three (3) children, the frequencies seem to decrease since fewer women give birth to the extent. Hence, the study adopted different levels of granularity for total children ever born. The following categories were considered: Small size (1 to 2 children), Moderate size (3 to 5 children) and Large size (children above 5).

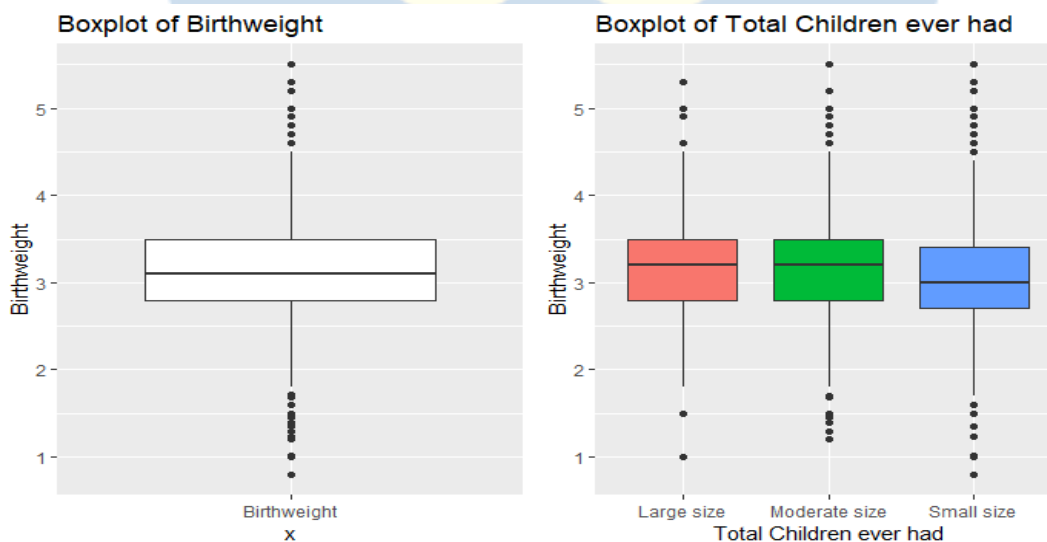


Figure 12: Boxplot of Birth and Total children ever born

The new stratification followed by boxplots lets us see the distribution of each group now has a similar distribution but with slightly different medians. Moreover, we can see that there are points on both the high and the low ends of the variable that appears to be outliers. (See Figure 12)

Table 47: Crosstabulation of Total Children ever born and Birthweight group

			Low birthweight	Normal birthweight	
Total Children ever born	Large	Obs.	42	400	442
		% of Total	1.2%	11.9%	13.2%
	Moderate	Obs.	128	1342	1470
		% of Total	3.8%	39.9%	43.7%
	Small	Obs.	165	1284	1449
		% of Total	4.9%	38.2%	43.1%
Total		Obs.	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

4.4.4.2.2 Association between newborn birthweight and total children ever had or born

The hypotheses for the data shown in Table 47 for the Chi-square test is:

H_{12_0} : There exists no significant association between children ever born and newborn birthweight.

H_{12_1} : There exists a significant association between children ever born and newborn birthweight.

Table 48: Chi-Square Tests for total children ever born and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	5.962 ^a	2	.039
Likelihood Ratio	5.939	2	.039
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 44.06.

The result of the Chi-square test was performed at a 5% level of significance as presented in Table 48. The Pearson chi-square significant value is 0.039 with 2 degrees of freedom. Therefore, the

null hypothesis is rejected. Hence, a significant association between children ever born and newborn birthweight exist. It might be concluded that children ever born and newborn birthweight are not independent of each other.

Table 49: Symmetric Measures for total children ever had/born and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.065	.039
	Cramer's V	.065	.039
Number of Valid cases		3361	

Measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.039 and the degree of association between two variables is 6.50%. (See Table 49)

4.4.4.3 Delivery place

4.4.4.3.1 Analysis for Delivery place

Choice of delivery place might be influenced for by several reasons based on respondent such geographical destination (region and type of residence), socio-economic status (such as wealth index) of the respondent and many more.

Table 50: Mean Tabulation of delivery place

		Obs.	Mean	Standard Deviation
Delivery place	Government facility	2795	3.126	.586
	Respondent Home	274	3.159	.645
	Maternity home	48	3.215	.532
	Other	9	2.944	.611
	Private facility	235	3.237	.609

In Table 50, the facility with the highest mean birthweight was recorded at the private facility with $3.237\text{kg} \pm 0.609\text{kg}$ while the facility that recorded the lowest mean birthweight of infants were delivery that happened at “Other” facilities accounting for $2.944\text{kg} \pm 0.532\text{kg}$.

Table 51: Delivery place and Birthweight Crosstabulation

			Low birthweight	Normal birthweight		
Delivery place	Government facility	Obs.	281	2514	2795	
		% of Total	8.4%	74.8%	83.2%	
	Home	Obs.	30	244	274	
		% of Total	0.9%	7.3%	8.2%	
	Maternity home	Obs.	2	46	48	
		% of Total	0.1%	1.4%	1.4%	
	Other	Obs.	1	8	9	
		% of Total	0.0%	0.2%	0.3%	
	Private facility	Obs.	21	214	235	
		% of Total	0.6%	6.4%	7.0%	
	Total		Obs.	335	3026	3361
			% of Total	10.0%	90.0%	100.0%

Table 51 tells us that, majority of births occurred in the government facility accounting for 2795 (83.2%) of the study, followed by 244 (8.2%) of births happening at respondent home.

4.4.4.3.2 Association between Delivery place and Birthweight

The hypotheses for the data shown in Table 51 for the Chi-square test is:

H_{13_0} There exist no significant association between delivery place and newborn birthweight.

H_{13_1} : There exists a significant association between delivery place and newborn birthweight.

Table 52: Chi-Square Tests for Delivery place and Birthweight

	Value	Df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.409 ^a	4	.011
Likelihood Ratio	2.865	4	.011
Number of Valid cases	3361		

a. 2 cells (20.0%) have expected count less than 5. The minimum expected count is .90.

Our Chi-square test was performed at a 5% level of significance with the results presented in Table 52. The Pearson Chi-square significant value is 0.011 with 4 degrees of freedom. Therefore, the null hypothesis is rejected. Hence, there is a significant association existing between the delivery place and newborn birthweight. It might be concluded that the delivery place and newborn birthweight are dependent on each other.

Table 53: Symmetric Measures for delivery place and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.065	.039
	Cramer's V	.065	.039
Number of Valid cases		3361	

Measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.011 and the degree of association between two variables is 6.50%. (See Table 53)

4.4.4.4 Preceding birth interval

4.4.4.4.1 Analysis of Preceding birth interval

Here, preceding birth interval (inter-pregnancy interval) is considered as the interval between two consecutive pregnancy.

Table 54: Descriptive Statistics for Preceding birth interval (Months)

	N	Min.	Max.	Mean	Median	Std. Deviation
Preceding birth interval (Months)	2451	12	217	49.82	42.00	28.437
Valid N	2451					
Invalid N (Missing data)	901					

Within the context of this study, the corresponding mean and median birth intervals are 49.82 months (approximately 50 months or 4 years and 2 months) and 42.00 months (approximately 3 years and 6 months) respectively. The minimum birth interval was 12 months (exactly a year) while its highest birth interval went as far as 217 months (more than 18 years). (See Table 54)

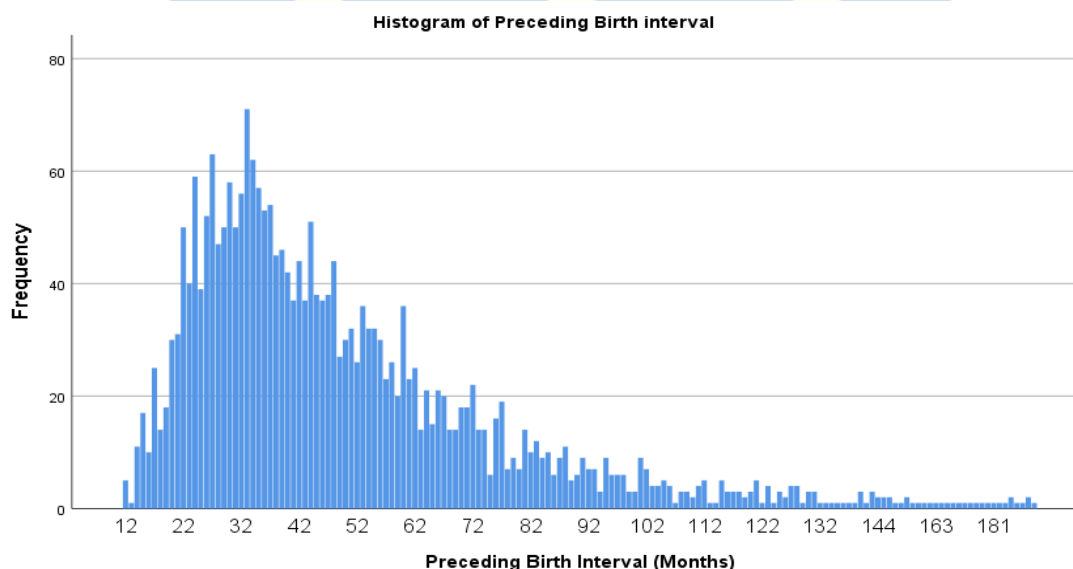


Figure 13: Preceding Birth Interval in Months

Figure 13 above gives a description of a fluctuated (move or sway in a rising and falling or wavelike pattern) trend of birth interval. In order to facilitate the analysis, the preceding pregnancy interval is divided into 4 groups; less than 24 months, 24 – 72 months, 73 – 120 months and interval above than 120 months. Mothers with the first parity group are considered as not applicable (included as part of missing data).

Table 55: Tabulation of Preceding birth interval and birthweight

		Obs.	Mean	Std.
Preceding birth interval	***	910	3.035	.599
	Less 24 Months	252	3.190	.570
	24-72 Months	1805	3.177	.587
	73-120 Months	315	3.150	.605
	Above 120 Months	79	3.163	.559

*** Missing data

Table 55, in comparing the mean birthweight across the various categories of preceding pregnancy interval, the highest birthweight was recorded at an inter-pregnancy interval less than 24 months and the lowest birthweight recorded at an inter-pregnancy interval between 73-120 months with respective birthweight means of 3.190 ± 0.570 and 3.150 ± 0.605 .

Table 56: Crosstab of Preceding birth interval and Birthweight

			LBW	NBW	
Preceding birth interval	***	Obs.	127	783	910
		% of Total	3.8%	23.3%	27.1%
	Less 24 Months	Obs.	18	234	252
		% of Total	0.5%	7.0%	7.5%
	24-72 Months	Obs.	149	1656	1805
		% of Total	4.4%	49.3%	53.7%
	73-120 Months	Obs.	37	278	315
		% of Total	1.1%	8.3%	9.4%
	Above 120 Months	Obs.	4	75	79
		% of Total	0.1%	2.2%	2.4%
Total		Observation	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

*** Missing data

In Table 56, majority of the respondents 1805(53.7%) gave birth between 24-72 months while less than 24 months birth interval, 73-120 months and above 120 months accounted for the following respectively percentages [252(7.5%), 315(9.4%) and 79(2.4%)]. Of 53.7% of birth occurring between 24-72 months, the study results showed that 4.4% of infants born had low birthweight. In other to easily facilitate the analysis, the mean preceding birth interval (months) is 49.8 which is

approximated as 50 months were replaced by the missing value which likewise falls within the highest distribution of birth the preceding months.

4.4.4.4.2 Association between newborn birthweight and preceding birth interval

The hypotheses for the data shown in Table 56 for the Chi-square test is;

$H14_0$: There exists no significant association between the preceding birth interval and newborn birthweight.

$H14_1$: There exists a significant association between Preceding birth interval and newborn birthweight.

Table 57: Chi-Square Tests for Preceding birth interval and birthweight

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	27.500 ^a	4	.000
Likelihood Ratio	26.817	4	.000
Number of Valid cases	3361		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 7.87.

The Pearson Chi-square test was performed at a 5% level of significance as presented in Table 57. Its significant value is 0.000 with 4 degrees of freedom. Therefore, our null hypothesis is rejected. Hence, a significant association between the preceding birth interval and newborn birthweight exist. It might be concluding that preceding birth interval and newborn birthweight are dependent on each other.

Table 58: Symmetric Measures for Preceding birth interval and birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.090	.000
	Cramer's V	.090	.000
Number of Valid cases		3361	

Measures of indexes of the agreement are employed to evaluate the strength of our association. The value of this measure is as per Chi-square table is significant with 0.000 and the degree of association between two variables is 9.0%. (See Table 58)

4.4.4.5 Birth Type

4.4.4.5.1 Analysis of Birth type

The birth type was categorized into two, namely single birth and multiple births (the delivery of more than one infant in a single birth event). Table 59 below shows that infants from single births are much less likely than those from multiple births to have low birthweight. The mean birthweight of a singleton child is 0.6kg more than multiple birth infant.

Table 59: Descriptive statistics for Birth type and Birthweight

		Obs.	Mean	Standard Deviation
Birth type	Multiple birth	167	2.521	.679
	Single birth	3194	3.169	.570

In the study, (86)2.6% of 167(5.0%) of infants from multiple births had low birthweight as compared to 249(7.4%) of 3194(95%) singleton births. Moreover, the likelihood of having low birthweight when an infant is from multiple births is 51.50% as compared to normal birthweight of 48.50%. (See Table 60 below)

Table 60: Birth type and Birthweight group Crosstabulation

			Low birthweight	Normal birthweight	
Birth type	Multiple Births	Obs.	86	81	167
		% of Total	2.6%	2.4%	5.0%
	Single birth	Obs.	249	2945	3194
		% of Total	7.4%	87.6%	95.0%
Total		Observation	335	3026	3361
		% of Total	10.0%	90.0%	100.0%

4.4.4.5.2 Association between newborn birthweight and birth type

The hypotheses for the data shown in Table 60 for the Chi-square test is;

H_{15_0} : There exists no significant association between birth type and newborn birthweight.

H_{15_1} : There exists a significant association between birth type and newborn birthweight.

Table 61: Chi-Square Tests for Birth type and birthweight

	Value	Df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	337.747 ^a	1	.000		
Continuity Correction ^b	332.895	1	.000		
Likelihood Ratio	200.260	1	.000		
Fisher's Exact Test				.000	.000
Number of Valid cases	3361				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 16.65.

b. Computed only for a 2x2 table

Our Pearson Chi-square test was performed at a 5% level of significance as presented in Table 61. It obtained a significant value of 0.000 with 1 degree of freedom. Therefore, the null hypothesis is rejected. Hence, a significant association between birth type and newborn birthweight exist. It might be concluding that birth type and newborn birthweight are dependent on each other.

Table 62: Symmetric Measures for Birth type and Birthweight

		Value	Approximate Significance
Nominal by Nominal	Phi	.317	.000
	Cramer's V	.317	.000
Number of Valid cases		3361	

In Table 62, Cramer V, one of the measures of indexes of the agreement is employed to evaluate the strength of association. The value of this measure is as per Chi-square table is significant with 0.000 and the degree of association between two variables is 31.7%.

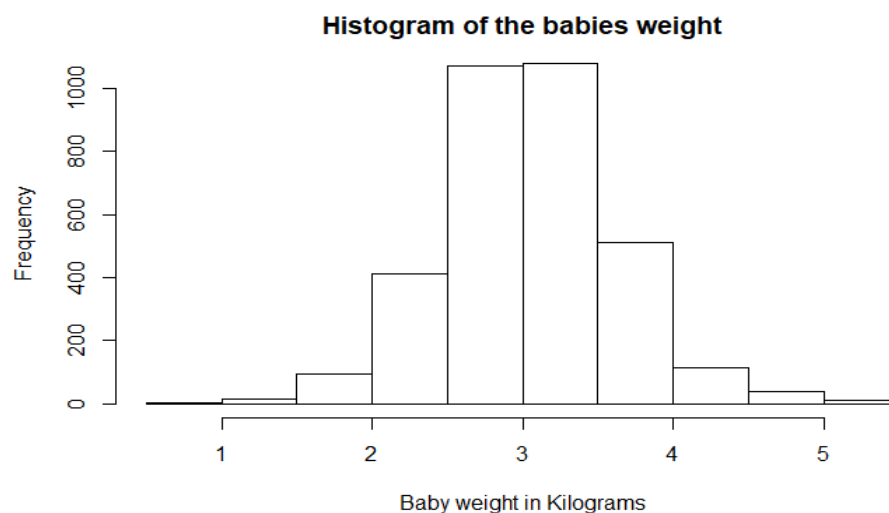
4.5 Multivariate Analysis of Predictors and Birthweight

The present study has various strengths compared with past researches conducted on the roles of maternal characteristics and outcome of a newborn on birthweight. This is a population-based that utilizes 2014 GDHS survey data that included important information on women who uses both traditional and modern healthcare. The study is expected to identify best indicators related to birthweight than past researches that depended on clinical data.

4.5.1 Towards Building a Multiple regression model

4.5.1.1 Step Zero: Understanding the Problem

Understanding the factors that can influence the baby weight is an important question. To answer such a broad question, we will begin by looking at the shape of the distribution of the baby's weights. It will allow us to argue whether a parametric or a non-parametric model is the best for this dataset.



Given this histogram it is quite hard to estimate if the data is truly normally distributed, thus we can draw the corresponding Q-Q plot.

Here, we can notice that the observations are quite aligned on the line which means that the sample quantiles correspond to the quantile of a theoretical normal distribution (see figure).

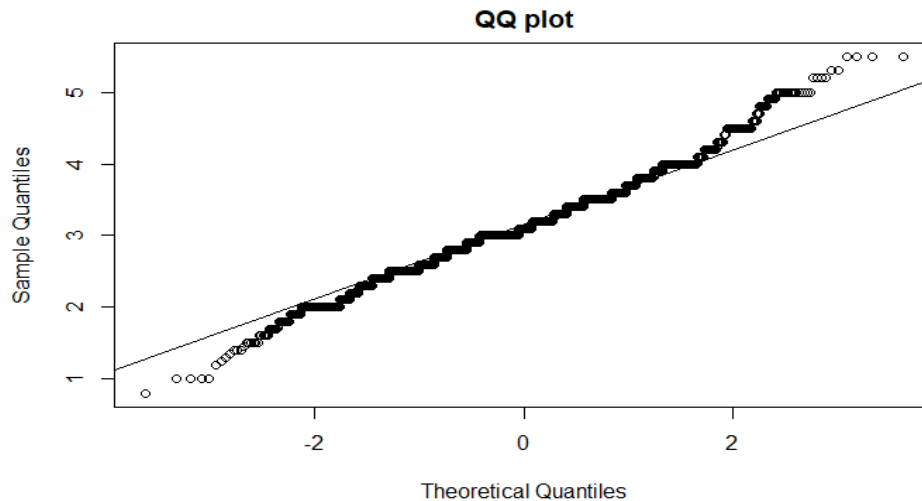


Figure 14: QQ-Plot for Longitudinal birthweight (kg)

4.5.1.2 Step One: Data Splitting

After, we will first split the data into a train and test set to be able to compare our models on the test error after building our model with the train set. A major objective of this procedure is to identify an algorithm $f(x)$ that most precisely predicts future values (y) in view of a set of inputs (x). As it were, we need an algorithm that fits well to our past data as well as more critically, one that predicts a future result precisely and accurately. This is known as the generalizability of our algorithm.

- **Training set:** these data are utilized to prepare our algorithms.
- **Test set:** these data are utilized to estimate our prediction error (generalization error) after chosen our final model.

To provide an accurate understanding of the generalizability of our final optimal model, the study split our longitudinal dataset into 70% training and 30% test using a randomized sampling from the caTool package in R software.

4.5.1.3 Step Two: Variables Selection Approaches

In multiple linear regression modeling, there are several methods of selection that determine how many independent variables can enter inside the analysis. By applying different selection methods, several regression models from the same set of indicators are formulated. We would like to try a different linear model to make a prediction on the weight of a newborn. In order to decide which predictors to choose and assess the relative importance of independent variables, three variable selection method was employed to yield the most appropriate regression equation with birthweight as the dependent variable.

- The Ordinary Least Squares (Backward selection)
- The Ridge regression
- The Lasso regression

The point of this selection is to decrease our set of predictor variables to those that are vital and represent about as a significant part of the variance by the total set. This will assist us in determining the degree of significance of every predictor variable.

Step Three: Multiple Regression model development

(1) RegModel 0: Baseline model

The baseline model uses the mean birthweight of an infant in its prediction. It is the solution to the problem without applying any machine learning techniques. This means that any model we build must improve upon our solution. Since we are predicting birthweight (which is a quantitative variable), the study considered the average of all birthweights in our dataset. The prediction for

child weight according to the baseline model is 3.137kg. Additionally, this resulted in an RMSE of 0.5873. So, any model we build should have an RMSE lower than 0.5873.

(2) Advanced Models

We evaluated by fitting train data to a multiple regression model. We considered both Ordinary Least Squares (OLS) and Regularized regression approaches: Ridge and Least Absolute Shrinkage and Selection Operator (LASSO).

RegModel 1: Ordinary Least Squares Model (Backward Selection)

First, a model of birthweight regressed on all other variables was fit; then, an iterative pruning procedure based on removing terms from the model and evaluating the impact on the AIC (Akaike's Information Criterion) was employed. We applied the backward selection method on the training set and compared the indicators of the AIC's. Finally, the AIC indicator converged in the idea that we should only consider 10 predictors in the model.

Now, we measure the quality of a linear regression using the predictors selected by the backward selection method. Based on the summary, we note that our linear regression model is decent as almost all predictors have less than 0.05 significance values. However, our R-Square is small with a value of 0.36, that is only 36% of the variance of the data is explained by our model. After, we applied our model on the test set in order to test its performance. It was noted that this linear model has a low accuracy on our test set with an RMSE of 0.4772.

Application of Shrinkage methods (Regularized Regression)

We further tried two shrinkage methods to verify the selected predictors. In both cases (Ridge and Lasso), we tested 100 different tuning parameter lambdas to find the one that minimizes the error on 10 folds cross-validation.

RegModel 2: Ridge Regression

We applied the ridge regression on the training set to select our predictors using their coefficients. The coefficient indicator converges in the idea that we should only consider 10 predictors in our analysis. Moreover, for the ridge regression, we minimize our RMSE for a tuning parameter of 0.0321. We then perform the ridge regression on the full training set to compute the optimal coefficients. Finally tested our model on the test dataset and obtained an RMSE of 0.47726.

RegModel 3: Lasso Regression

A Lasso regression on the training set was applied and compared the coefficients of the predictors. The coefficient indicator likewise converged in the idea that we should only consider 10 of the 18 predictors in our analysis. Moreover, the Lasso regression minimized our RMSE for a tuning parameter of 0.02924. Thereafter performed the Lasso regression on the full training set to compute the optimal coefficients. Finally tested our model on the test dataset and obtained an RMSE of 0.4837441.

Step Four: Model Selection Evaluation

All the above models were evaluated using a standard regression error metric called the Root Mean Square Error (RMSE).

Table 63: Variable selection for Multiple Regression Model building

	Model 0 Baseline Model	Model 1 OLS Regression (Backward Selection)	Model 2 Ridge Regression	Model 3 Lasso Regression
Number of variables Entered	Null	10	10	10
Variables Selected		<ul style="list-style-type: none"> • Region • Gender of child • Delivery place • Children ever born • Birth type • Birth order • Mothers weight 	<ul style="list-style-type: none"> • Region • Gender of child • Delivery place • Children ever born • Birth type • Birth order • Mothers weight 	<ul style="list-style-type: none"> • Region • Gender of child • Delivery place • Children ever born • Birth type • Birth order • Mothers weight

		<ul style="list-style-type: none"> Mothers height Wealth index Size of child 	<ul style="list-style-type: none"> Mothers height Wealth index Size of child 	<ul style="list-style-type: none"> Mothers height Wealth index Size of child
RMSE	0.5873	0.47720	0.47726	0.48374

From Table 63, it can be inferred that all models (OLS, Ridge, and Lasso) have smaller Root Mean Squared Errors than the baseline model. It could also be seen that these Root Mean Squared Errors don't look very different from other variable selections, which means it's probably better to choose a more interpretable one. However, it appears the "Best Model" is the OLS backward selection technique. Considering this, the study adopted the OLS backward selection predictors in the modeling building.

Table 64: Summary of Model: OLS Backward Selection

		Model Likelihood Ratio Test		Discrimination Indexes	
Obs.	2355	LR chi2	1066.72	R-Square	0.364
Sigma0	4474	Df	29	R-Square Adj.	0.356
Df	2325	Pr (> chi2)	0.0000	G	0.403
Residuals Analysis					
Minimum	1 st Quartile	Median	3 rd Quartile	Maximum	
-1.60931	-0.29305	-0.03294	0.25982	2.17720	

Coefficients of the Multiple Regression Model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.353231	0.394974	3.426	0.000623 ***
RegionBrong Ahafo	0.027643	0.041144	0.672	0.501739 (NS)
RegionCentral	0.026571	0.043117	0.616	0.537791 (NS)
RegionEastern	0.106046	0.046011	2.305	0.021266 *
RegionGreater Accra	0.188214	0.042338	4.446	0.0001 ***
RegionNorthern	-0.085306	0.046320	-1.842	0.065653 (NS)
RegionUpper East	0.122830	0.044967	2.732	0.006352 **
RegionUpper West	-0.048152	0.046670	-1.032	0.302299 (NS)

RegionVolta	0.289806	0.044773	6.473	<0.0001 ***
RegionWestern	0.067039	0.043183	1.552	0.120691 (NS)
Gender_of_childMale	0.069121	0.019835	3.485	0.000502 ***
Delivery_placeGovernment hospital	0.027651	0.025106	1.101	0.270840 (NS)
Delivery_placeMaternity home	0.151639	0.082981	1.827	0.067769 (NS)
Delivery_placeOther	0.069177	0.100287	0.690	0.490394 (NS)
Delivery_placePrivate hospital, clinic	0.120785	0.044236	2.730	0.006372 **
Delivery_placeRespondent's home	0.023122	0.039968	0.579	0.562977 (NS)
Delivery_typeYes	0.098489	0.030092	3.273	0.001080 **
Children_ever_bornModerate	-0.038881	0.031816	-1.222	0.221808 (NS)
Children_ever_bornSmall	-0.145925	0.032875	-4.439	<0.0001 ***
Birth_typeSingle birth	0.576971	0.045826	12.590	<0.0001 ***
Mothers_weight	0.002764	0.001129	2.448	0.014425 *
Mothers_height	0.005810	0.002553	2.276	0.022953 *
Wealth_indexPoorer	-0.077610	0.033197	-2.338	0.019481 *
Wealth_indexPoorest	-0.047433	0.035224	-1.347	0.178241 (NS)
Wealth_indexRicher	-0.047822	0.031996	-1.495	0.135149 (NS)
Wealth_indexRichest	-0.089551	0.034818	-2.572	0.010174 *
Size_of_childLarger than average	0.234356	0.024583	9.533	<0.0001 ***
Size_of_childSmaller than average	-0.288376	0.032863	-8.775	<0.0001 ***
Size_of_childVery large	0.584341	0.031019	18.838	<0.0001 ***
Size_of_childVery small	-0.548516	0.053502	-10.252	<0.0001 ***

*NS=Not significant *p-value <0.05 **p-value <0.01 ***p-value <0.001*

Step Five: Checking Model Assumptions

a. Checking the linearity and Constant Variance Assumption

1. Fitted versus Residuals Plot

Our most helpful instrument will be a Fitted versus Residuals Plot to validate both the linearity and constant variance assumption. In here, we search for two things in the plot. At any fitted value, the mean of the residuals ought to be approximately 0. If so, the linearity assumption is valid and

at each fitted estimate, the spread of the residuals ought to be generally equivalent. If so, the constant variance assumption is valid.

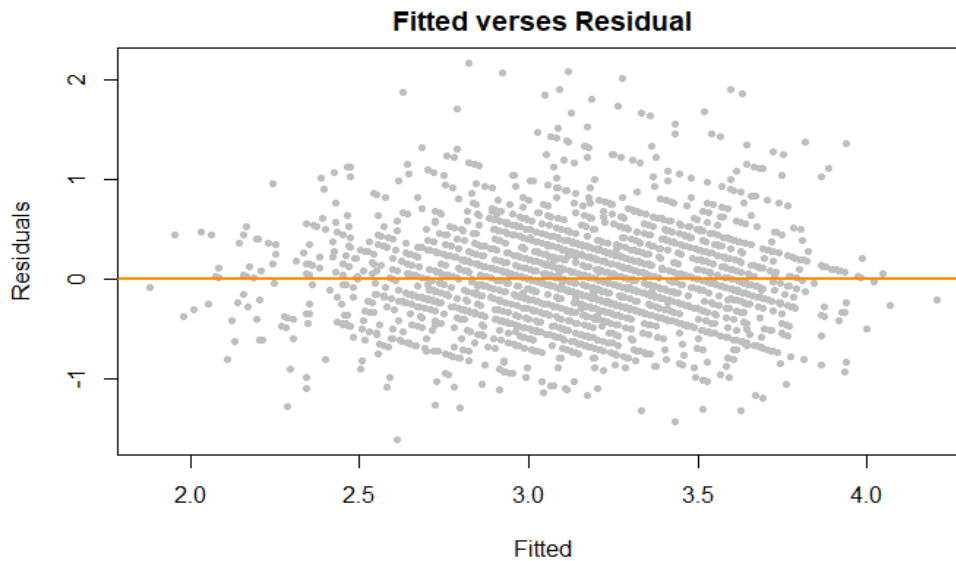


Figure 15: Fitted Verse Residuals for the Multiple Linear Model fit

The plot showed in figure 15 demonstrates the scatterplot of standardized residuals against the fitted. Finding uncovered from our graphical representation shows no obvious difficulties. For any fitted value, the residuals appear to be roughly centered around 0. A random pattern is obvious in the plot, this is a good sign of the linearity assumption not violated. Notwithstanding, we likewise observe clearly, that for larger fitted estimates, the spread (dispersion) of the residuals is larger which is terrible portraying a violation in constant variance assumption.

2. Breusch-Pagan Test

While the fitted versus residuals plot gave us a thought regarding homoscedasticity, sometimes a more formal test is preferable. There are numerous constant variance tests, however, the study employed the Breusch-Pagan Test with our null and alternative considered as:

- H_0 : Homoscedasticity (About the true model, the errors have constant variance)
- H_1 : Heteroscedasticity (About the true model, the errors have non-constant variance)

Studentized Breusch-Pagan test			
Data	Breusch-Pagan	df	p-value
Model	82.533	29	<0.00001

For our model, we see a small p-value, so we reject the null of homoscedasticity indicating that a violation of constant variance assumption. This matches our finding with a fitted versus residuals plot.

b. Test for Independence Assumption

Durbin-Watson test for autocorrelation was calculated. If our residuals are autocorrelated, the regression results can be inaccurate and unreliable. The Durbin-Watson test estimates lie between 0 to 4.

1. 2 means no autocorrelation
2. >2 to 4 is negative autocorrelation
3. 0 to <2 is positive autocorrelation

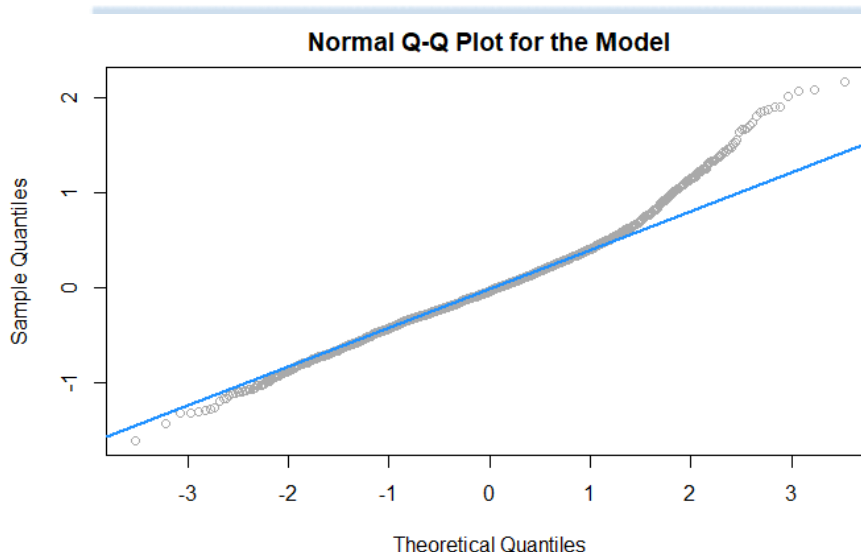
In our Durbin-Watson test, our null hypothesis (H_0) is that there is no first-order autocorrelation with an alternate hypothesis (H_1) stating the existence of first-order correlation.

Lag	Autocorrelation	D-W Statistic	p-value
1	-0.0371087	2.074037	0.084
Alternative hypothesis: $\rho \neq 0$			

In the study case, $p > 0.05$ (0.084), hence failed to reject our null hypothesis. There is no indication of the presence of autocorrelation in the residuals and our independence assumption is satisfied.

c. Normality assumption using QQ-Plot

In order to evaluate our normality assumption, the normal Q-Q plot of regression standardized residuals is derived. Be that as it may, if it shows up roughly normal, then we will accept the errors could, in fact, be normal.



Looking at the figure above, it displays a speculated Q-Q plot. We would most likely not accept the errors comes from a normal distribution. Therefore, more formal testing known as the Shapiro–Wilk test was computed.

Shapiro-Wilk normality test		
Data	W	p-value
Model residuals	0.97436	<0.0001

The table above provides us an estimation of the test statistic of 0.97436 with its p-value (<0.0001). The null hypothesis assumes the study data were sampled from a normal distribution; thus, a small p-value in here demonstrates that just a small likelihood of our study data could have been sampled from a normal distribution. Both the Q-Q plot and our Shapiro-Wilk test suggests normality assumption violation.

d. Multicollinearity of our predictor's checking***Multicollinearity assumption check***

		Tolerance	VIF
Predictor	Region	.977	1.024
	Gender of child	.998	1.002
	Delivery Place	.997	1.003
	Delivery type	.935	1.070
	Children ever born	.955	1.047
	Birth type	.969	1.032
	Wealth index	.937	1.067
	Mothers weight	.850	1.177
	Mothers height	.876	1.142
	Size of child	.991	1.009

None of our variance inflation factor value is greater than 10.0, and our tolerance values indicate that collinearity does not explain greater than 10 percent of any predictor variable's variance. Our birthweight model problem indicates no evidence of significant collinearity.

Step Six: Unusual Observations

Notwithstanding checking multiple linear regression assumptions, the study additionally examines any unusual data points in our trained dataset. Frequently few data points could have an extreme effect on our regression model, occasionally to such an extent that there is a violation in our regression model as a result of these points.

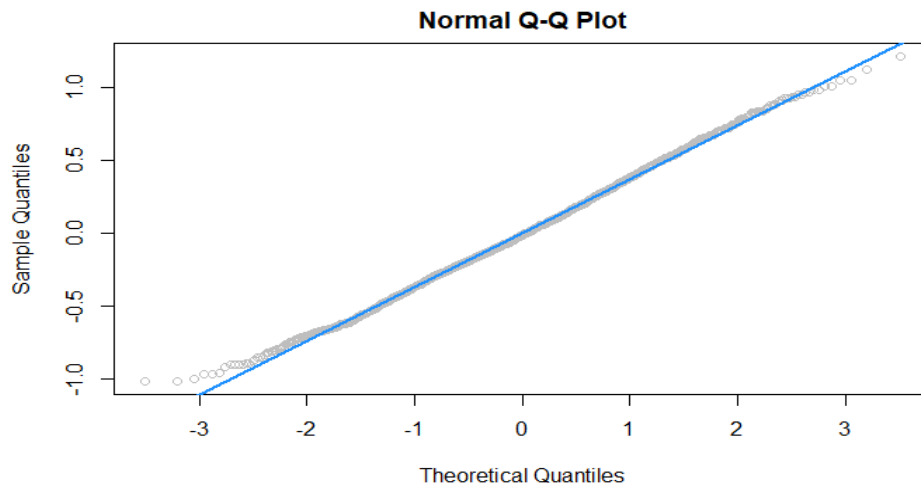
1. Leverage

An observation with high leverage is a data point that could have a large influence during our model fitting. In doing our leverage analysis, data points 2145 and 3066 were found.

2. Outliers

To identify outliers, the study looked for observations with large residuals. With an added point, the study then calculates each of the residuals and standardized residuals. Since our sample size

(n=2353) is large, the standardized residuals greater than 2 in magnitude was used to identify “large” residuals. In the analysis of the standardized residuals, we realized that 128 added points have large standardized residual. Hence, we considered to remove these points and recheck our model for improvement.



The figure above indicated that after knocking out these influential data points (leverages and outliers), findings display a much better normal Q-Q plot and now Shapiro-Wilk rejects for a high p-value.

Likewise, the removal of the influential points, the homoscedasticity assumption using the Breusch-Pagan test was rechecked. See table below

Shapiro-Wilk normality test		
Data	W	p-value
Model residuals	0.99782	0.08

For the fixed model (after influential data points removed), we see a large p-value indicating a failure to reject our null hypothesis of homoscedasticity, hence our constant variance is assumed (not violated).

Studentized Breusch-Pagan test			
Data	Breusch-Pagan	df	p-value
Model after fix	22.591	24	0.544

Summary after resolving our assumptions behind our fitted model

After these diagnostic tests, the study results about to interpreted is reliable.

Table 65: Summary after resolving the assumptions behind our fitted model

		Model Likelihood Ratio Test		Discrimination Indexes	
Obs.	2209	LR chi2	1373.25	R-Square	0.463
Sigma0	3760	Df	24	R-Squ. Adj.	0.457
Df	2184	Pr (> chi2)	0.0000	G	0.386
Residuals Analysis					
Min	1 st Quartiles	Median	3 rd Quartiles	Max	
-1.04266	-0.24925	-0.01725	0.24779	1.24011	

Coefficient of the Fixed Model

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.1805372	0.3249786	3.633	0.000287 ***
RegionBrong Ahafo	0.0260511	0.0331938	0.785	0.432644 (NS)
RegionCentral	0.0165153	0.0349598	0.472	0.636682 (NS)
RegionEastern	0.0160550	0.0383110	0.419	0.675206 (NS)
RegionGreater Accra	0.1854645	0.0344450	5.384	<0.0001 ***
RegionNorthern	-0.0536515	0.0371116	-1.446	0.148410 (NS)
RegionUpper East	0.1108757	0.0361808	3.064	0.002207 **
RegionUpper West	-0.0298572	0.0373381	-0.800	0.424004 (NS)
RegionVolta	0.2783656	0.0360307	7.726	< 0.0001 ***
RegionWestern	0.0538071	0.0350977	1.533	0.125404 (NS)
Gender_of_childMale	0.0708896	0.0160964	4.404	< 0.0001 ***
Delivery_typeYes	0.0606169	0.0245646	2.468	0.013676 *
Children_ever_bornModerate	-0.0376362	0.0260730	-1.443	0.149026 (NS)
Children_ever_bornSmall	-0.1238582	0.0268070	-4.620	< 0.0001 ***
Birth_typeSingle birth	0.6049093	0.0392759	15.402	< 0.0001 ***

Mothers_weight	0.0023190	0.0009271	2.501	0.012445 *
Mothers_height	0.0068369	0.0021008	3.254	0.001154 **
Wealth_indexPoorer	-0.0720377	0.0268109	-2.687	0.007267 **
Wealth_indexPoorest	-0.0472767	0.0280248	-1.687	0.091754 (NS)
Wealth_indexRicher	-0.0138321	0.0261286	-0.529	0.596592 (NS)
Wealth_indexRichest	-0.0227756	0.0280711	-0.811	0.417250 (NS)
Size_of_childLarger than average	0.2087758	0.0198437	10.521	< 0.0001 ***
Size_of_childSmaller than average	-0.3053589	0.0266587	-11.454	< 0.0001 ***
Size_of_childVery large	0.5621383	0.0252864	22.231	< 0.0001 ***
Size_of_childVery small	-0.5746811	0.0469995	-12.227	< 0.0001 ***

*NS=Not significant *p-value <0.05 **p-value <0.01 ***p-value <0.001*

Note: Delivery place was removed after model fixing

Step Seven: Results Interpretation of Model

Before one makes use of a multiple regression equation (MRE), it is desirable to determine first whether it is, in fact, worth using. In this regard, the coefficient of multiple determination (R-Square) is utilized together with our assessment of its significance. An adjusted R-Square estimate is necessary in order to make the coefficient comparable.

In the present analysis, only four of the seven maternal variables (mother's weight, height, region and wealth index), two of three outcome of birth (gender and size of child) and three of seven obstetric variables (delivery type, children ever born and birth type) met the criteria (probability of $p < 0.050$) to enter into the models for developing multiple regression equation (MRE) (Table 8.1).

In our final model, the calculated R-Square is 0.463, stating a significant portion (above 46%) of the variation in newborn birthweights is explained by these nine predictors and R (multiple correlation coefficient) was 0.68, which is not high and not very close to the ideal value '1'. It indicates that MRE might not be of much use for the estimation of birthweight of newborn for the given value of the mother's weight in the segment of the population under study. Our adjusted

squared multiple correlation coefficient reported as (0.457) is not very different from the unadjusted value of (0.463) because the number of independent variables ($k=24$) is much less than the number of observations ($n=2209$).

Table 66: ANOVA Table for Multiple Linear Regression

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	266.137	24	11.089	78.65	.000 ^b
	Residual	308.737	2184	0.141		
	Total	574.874	2208			

a. Dependent Variable: Birthweight

b. Predictors: (Constant), Region, Mothers height, Mothers weight, Gender of a child, Size of a child, Delivery type, Wealth index, Birth type, Children ever born

In Table 66, we then examine our ANOVA table associated with our multiple linear regression. Our sum of squares due to the regression model is reported as 266.137 and the sum of squares due to error (residual sum of squares) is 308.737. The residual mean square is an estimate of the variance and is equal to 0.141. The value of the F statistic is equal to 78.65 with our corresponding p-value of 0.000 providing very strong evidence of the utility of the model.

Presently, the study dissects the part of the findings providing the estimates of the regression parameters: According to our model, the estimated regression line of birthweights on our nine predictors is:

$$\begin{aligned} \mu\{BWT\} = & 1.181(\text{Intercept}) + 0.026(\text{Reg} = B. Ahafo) + 0.017(\text{Reg} = Central) + \\ & 0.016(\text{Reg} = Eastern) + 0.185(\text{Reg} = G. Accra) - 0.054(\text{Reg} = Northern) + \\ & 0.111(\text{Reg} = U. East) - 0.030(\text{Reg} = U. West) + 0.278(\text{Reg} = Volta) + 0.054(\text{Reg} = \\ & Western) + 0.071(\text{GoC} = Male) + 0.061(\text{Del. type} = Yes) - 0.038(\text{CeB} = Moderate) - \\ & 0.124(\text{CeB} = Small) + 0.605(\text{B. type} = Single) + 0.002(\text{MWeight}) + 0.007(\text{MHeight}) - \\ & 0.072(\text{W. index} = Poorer) - 0.047(\text{W. index} = Poorest) - 0.014(\text{W. index} = Richer) - \end{aligned}$$

$0.023(W.index = Richest) + 0.209(SoC = Larger\ than\ average) - 0.305(SoC = Smaller\ than\ average) + 0.562(SoC = Very\ large) - 0.575(Soc = Very\ small)$

4.5.2 Towards Building a Logistic Regression Model

The analyses in our study also utilized logistic regression modeling due to its conversion into a binary dependent variable (birthweight), indicating low or normal weight. This modeling approach will try to use the predictors available in the dataset to predict the weight of the newborn baby. The multiple regression models demonstrated not many quality results with respect to its R-Square, that might suggest that these predictors are not enough to explain all the variance observed in the birthweight response variable. However, the major goal of the study is to identify our risk factors accounting for a low-weight infant at birth. For that, we will reformulate our modeling problem as a classification problem, testing for threshold in the dataset, which will split healthy infants from infants at risk, and fitting logistic regression on this binary outcome — *no* for normal infants and *yes* for infants with low weight.

4.5.2.1 Fitting Logistic Regression

The study began by trying to model the probability of birthweight based on low weight. The birthweight dataset contains information on 3,361 neonates. Birthweight classification is the dependent variable in the data, where 1 denotes presence of low weight and 0 denotes the absence of low weight (normal weight). There are 90.1% of normal weight infants and 9.9% of infants are low weight in this data. This dataset is used to develop logistic regression in order to predict our likelihood of having low birthweight.

Step Zero: Creating a Classification table with full data

First, a “cut-off” value c (0.5) is chosen. For each subject in the low birthweight dataset, the study “predicts” the baby’s birthweight status as 0 coded to be normal if their fitted probability of being

normal birthweight is greater than c, otherwise, we predict it as 1 coded to low). Then after, a table is constructed to show how many of the observations our study has correctly predicted.

Table 67: Classification Table on full low weight data

Observed		Predicted		
		Birthweight group		Percentage Correct
		LB	NB	
Birthweight	NBW	3026	0	100.0
	LBW	335	0	0
Overall Percentage				90.0

There are 3026 normal birthweights and 335 low birthweights. Thus, our odds of low birthweight are equal to $335/3026=0.1107$. Our odds are confirmed in the output for the model with a constant only.

Variables in the Equation

	Est	Standard Error	Wald Statistics	Degree of freedom	Sig. value	Exp(B)
Constant	-2.201	.058	1440.072	1	.000	.111

Step One: Data Splitting for the Logistic model building

Likewise, we load the cleaned data and split it into train and test subsets. We do this because we will use the train set for model selection and use the test set for backtesting the performance of the chosen model. In providing an accurate understanding of the generalizability of our final optimal model, a 70% train and 30% test split of our data is done using the same randomized sampling from the caTool package in R software.

Step Two: Apply Sample methods to Imbalanced Data

As the study has less low birthweight (less 10%), we apply sample methods to balance the data. The study employed the use of Over, Upper, Mixed (Both) and ROSE sampling methods using ROSE package and SMOTE sampling method using DMwR package in R.

- **Oversampling**, as Birthweight (1) is having less occurrence, so this oversampling method will increase the birthweight records until matches good records 3361. Here $N = 2352 * 2 = 4704$.
- **Under-sampling**, as Birthweight (1) is having less occurrence, so this under-sampling method will decrease the normal birthweight records until matches birthweight records, but the limitation to this is that we can lose significant information from the sample.
- **Mixed Sampling** apply both undersampling and oversampling on this imbalanced data.
- **ROSE Sampling**, this helps us to generate data synthetically. It generates artificial data instead of duplicate data.
- **SMOTE (Synthetic Minority Over-sampling Technique)** is a technique based on nearest neighbors judged by Euclidean Distance between data points in feature space. There is a percentage of Over-Sampling which indicates the number of synthetic samples to be created and this percentage parameter of Over-sampling is always a multiple of 100.

Table 68: Birthweight in train dataset after applying Sampling method

Sampling Method	Birthweight Classification		Area Under Curve (ROC) on test data
	LBW	NBW	
Train dataset	234	2118	0.652
Over Sampling	2586	2118	0.736
Under Sampling	234	766	0.721
Mixed Sampling	1707	1654	0.752
ROSE Sampling	1173	1179	0.813
SMOTE Sampling	468	468	0.793

Now, we have five different types of inputs which are balanced and ready for prediction. We then apply the logistic classifier to all these five datasets and calculate the prediction performance of each on the test set. The highest data accuracy is obtained using the ROSE method. However, there is not much variation in these sampling methods. The present data employed the ROSE Sampling technique to balance data for the logistic model fitting. (See Table 68)

Step Three: Variables selection for Logistic Model building

LogModel 1: Stepwise variable selection using AIC (Backward Selection)

Next, we applied the backward selection method on the training set and compared indicators using the AIC. The AIC indicator converges in the idea that we should only consider nine predictors in our analysis. We selected the predictors with p-value less 0.05.

LogModel 2: Variables significance using Chi-square test

In LogModel 2, a chi-square test of significance of variables is conducted to validate the selection on the training set made by the stepwise backward selection. The p-value indicators converged in the idea that we should also consider nine variables.

Table 69: Summary of Variable selection for Logistic Regression Model building

	LogModel 0: Full Model	LogModel 1: Stepwise variable selection using AIC (Backwards Selection)	LogModel 2: Chi-square test for significance of variables
Number of variables Entered	17	9	9
Variables Entered	All predictors	<ul style="list-style-type: none"> • Region • Gender of child • Delivery place • Birth type • Birth order • Mothers height • Preceding birth interval • Wealth index • Size of child 	<ul style="list-style-type: none"> • Region • Gender of child • Delivery place • Birth type • Birth order • Mothers height • Preceding birth interval • Wealth index • Size of child

Since each variable selection method yielded the same nine predictors, the study concluded by using these significant predictors in the logistic model building and knocked out eight variables considered as insignificant. The significant predictors selected were reran in order to build the final logistic regression model.

Step Four: Model Evaluation

Our logistic regression model has been formulated and their corresponding coefficients will later be analyzed. In any case, few critical questions need to be answered. Is our model any good? How well does our model fit the training data? Are our predictions accurate? Which independent variables are most significant?

(1) Goodness of Fit test on the model

- **Likelihood Ratio Test (Drop-in-Deviance)**

Our model is said to produce the best fit to our dataset if it depicts an improvement over a reduced model (fewer covariates). Given that our null hypothesis (H_0) holds that the reduced model is true, a p-value less than (0.05) for our overall model fit statistic would compel us to reject our null hypothesis. This would provide enough evidence against the reduced model in favor of the full model.

Likelihood ratio test

Model	#df	Likelihood Ratio test	Difference	Chisq	Pr(>Chisq)
LogModel 0	44	-1019.44			
LogModel 1	30	-992.55	-14	53.777	0.1083

In the likelihood ratio test table above, value for the reduced model is higher than the full model - $992.55 > -1019.44$, the p-value (0.1083) is greater than the alpha value of (0.05). This provides evidence for the null hypothesis that our null hypothesis is true.

- **Hosmer-Lemeshow Test**

To validate our goodness of model fit, the Hosmer-Lemeshow statistic was also computed. This investigates whether our observed proportions of events are like the predicted probabilities of occurrence in subgroups of the dataset using a Pearson chi-square test. The study expresses our null and alternative hypotheses as follows:

H_0 : The model is a good fit for the data against H_1 : The model does not fit for the data well

Hosmer and Lemeshow Test

Step	X-squared	Chi-square	Df	Sig.
1	2352	8.110	8	0.43

From the Hosmer and Lemeshow test table, a Chi-square of 8.110 with 8 degrees of freedom and the corresponding p-value is computed as 0.43. P-value greater than 0.05 tells us that the study fails to reject our null hypothesis that there is no difference between the observed and model-predicted values. It suggests that our model provides an adequate fit to the train data.

(2) Measures of the Proportion of Variation explained by our logistic model

After assessing our model fit, it was necessary to establish the strength of our association among the significant predictors with birthweight. Few statistics have been proposed on account of logistic modeling that can be viewed as generally equal in interpretation to the coefficient. Nonetheless, the study utilized the application of the Pseudo R-Square.

- **McFadden's R-square**

The measurement lies between 0 to just under 1, with estimates nearer to zero indicating that our model has no power of prediction, values between above 0.2 generally being considered as satisfactory. If McFadden's R-square is less than 0.2, then the model does a poor job in explaining the target variable.

- **Nagelkerke's R-Square and Cox and Snell's R-Square**

This is based on the computation of our relative change in the log-likelihood for our intercept-only-model to our full model.

Strength of Relationship of the Logistic Model

	Pseudo R-squared
McFadden	0.3911766
Cox and Snell (ML)	0.4185809
Nagelkerke (Cragg and Uhler)	0.5581091

The interpretation in table above is that our model (region, gender of child, delivery place, birth type, birth order, mother’s height, preceding birth interval, wealth index and size of child) as our predictors explains about 39% of the variation in our model using the McFadden R-Square, 42% and 56% respectively in Cox and Snell and Nagelkerke .

(3) Statistical Tests for Individual Predictors

- **Wald Test**

The Wald test was utilized to examine each of our coefficient’s statistical significance in our model. The idea is to test the hypothesis that our coefficient of predictors is not significantly different from zero. If our test fails to reject the null hypothesis, the suggestion is that removing the variable would not substantially harm our fitted model.

Variable	Degree of freedom	F-Statistics	P-value
Region	9	7.966189	0.0000*
Gender of child	1	21.64232	0.0000*
Delivery place	5	3.431877	0.0043*
Birth type	1	210.4018	0.0000*
Birth order	1	24.3509	0.0000*
Mothers height	1	15.58437	0.0000*
Wealth index	4	3.316142	0.0102*
Size of child	4	114.0508	0.0000*

** Significant at alpha 0.05*

Since our p-value for the Wald statistics is less 0.05, it ought to force us to reject our null hypothesis and accept that our variable ought to be incorporated into the fitted model. Notwithstanding, a p-value exceeding 0.05 proposes that such covariates can be excluded from the model.

(4) Validation of Predicted Values using ROC Curve (AUROC)

In the y-axis, true positive rate (TPR) or sensitivity is displayed and x-axis false positive rate (FPR) or $1 - \text{specificity}$ is displayed, the curve approaching 1 represents the best performance of the model. The Area under Curve (AUC) values greater than 70% is considered a model with high predictive accuracy. In Figure 19, illustrating how our logistic regression model, Area under curve (AUC) value is around 81%, which is considered an accurate model. (See figure below).

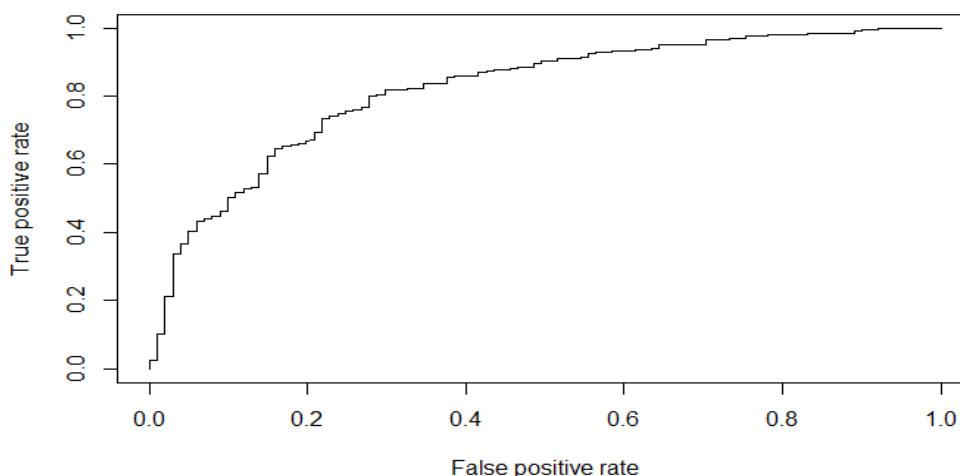


Figure 16: ROC Curve for our Logistic Model

Step Five: Diagnostic Test of the Model

1. Variance Inflation Factor for Multicollinearity

Variance inflation factor was calculated for each predictor in our model. While there is disagreement on the appropriate cutoff for identifying whether the model suffers from multicollinearity, the following rules provide a rough guideline.

- If $VIF > 1$ and $VIF < 2.5$, then those explanatory variables are moderately correlated.
- If $VIF > 2.5$, then those variables are highly correlated.

	GVIF	Df	GVIF^{1/(2*Df)}
Region	2.185337	9	1.058560
Gender of child	1.050815	1	1.025093
Delivery place	1.568413	5	1.046035
Birth type	1.219042	1	1.104102
Birth order	1.216167	1	1.102800
Mothers height	1.089146	1	1.043622
Wealth index	2.069814	4	1.125221
Size of child	1.398831	4	1.042847

Looking at the rule of thumb above, one could conclude that there is no existence of multicollinearity in our predictor variables.

2. Influential Observations

After our model development, we also want to examine whether certain observations in the dataset that had a significant influence on the estimates.

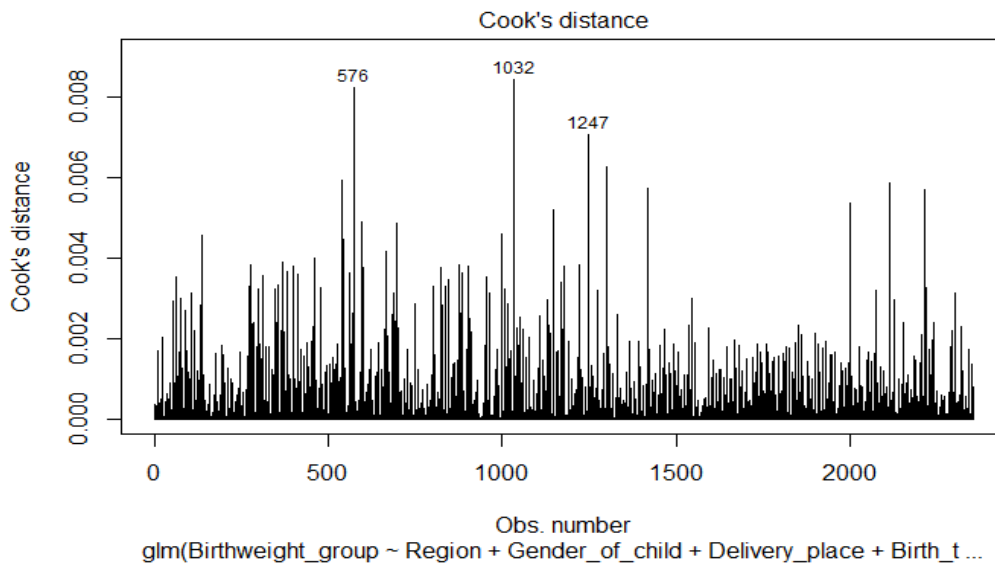


Figure 17: Cooks distance for the Logistic Model

During the analysis, 18 data points were significantly different from others, which in one way or the other might had to affect our parameter estimates. Hence, a decision was made to examine these influential data points. The influential data points were removed, and our model was rerun, it could be seen from the table below an improvement on our model performance.

Table 70: Strength of Relationship of the Logistic Model

	Old Pseudo R-Squared	New Pseudo R-Squared
McFadden	0.3911766	0.595460
Cox and Snell (ML)	0.4185809	0.561877
Nagelkerke (Cragg and Uhler)	0.5581091	0.749265

Step Six: Model Summary and Interpretation

The variables in our Parameter Estimates in Table 71 has several significant components. The Wald statistic and its associated probabilities provide us an index of each predictor importance in our equation. In our Table below, the second column the value of ‘B’ is our estimated coefficient, with standard error (S.E.) (third column of the same table). The ratio of ‘B’ to Standard error, squared, equals to Wald statistic. The simplest way to assess Wald statistic is to take the significance values (as shown in the fifth column of the same table) and if it is less than 0.05, then reject the null hypothesis as the variable does make a significant contribution. Further, it is to be concluded that the parameter is useful to the model.

“Exp (B)” column in the table below represents the extent to which raising the corresponding measure by one unit influences the odds ratio. In other words, “Exp (B)” or odds ratio is the predicted change for a unit change in odds for a unit increase or decrease in the predictor. The “exp” refers to the exponential value of B. When Exp (B) is less than 1; the increasing value of the variable corresponds to the decreasing odds of the event’s occurrence. When Exp (B) is greater than 1, increasing the value of the variable correspond to the increasing odds of the event’s

occurrence. In other words, if the value of Exp (B) exceeds 1, the odds of an outcome occurring increase; if the value is less than 1, any increase in the predictor leads to a drop in the odds of the outcome occurring.

Table 71: Summary analysis of the final Logistic regression model

		Model Likelihood Ratio Test		Pseudo R-Squared		Rank Discrimination Indexes	
Obs.	2352	LR chi2	1256.51	McFadden	0.595460	C	0.885
NB	1179	df	29	Cox and Snell	0.561877		
LB	1173	Pr (> chi2)	<0.0001	Nagelkerke	0.749265		

Where;

- Obs. is the number of observations in our model fit
- LR is model likelihood ratio χ^2
- df is the degree of freedom
- Pr (> chi2) is the P-value
- c index is the area under ROC curve
- The McFadden R-Square, Cox and Snell and Nagelkerke R-Square



Parameter Estimates								
	Estimate (B)	Std. Error	Wald Test	z value	Pr (> z)	Odds ratio = Exp (B)	CI'S for the Odds Ratios	
							Lower	Upper
(Intercept)	12.014*	2.090	33.043	5.749	<0.0001	164999.800	2856.718	10365548.000
Base(RegionAshanti)								
RegionBrong Ahafo	-1.375*	0.350	15.439	-3.929	<0.00001	0.253	0.127	0.500
RegionCentral	-1.975*	0.380	27.009	-5.197	<0.00001	0.139	0.065	0.289
RegionEastern	0.373	0.354	1.112	1.055	0.292	1.452	0.726	2.913
RegionGreater Accra	-2.274*	0.370	37.826	-6.150	<0.00001	0.103	0.049	0.211
RegionNorthern	-0.318	0.430	0.547	-0.740	0.460	0.727	0.313	1.692
RegionUpper East	-1.109*	0.362	9.406	-3.067	0.002	0.330	0.162	0.668
RegionUpper West	0.265	0.350	0.573	0.757	0.449	1.303	0.656	2.590
RegionVolta	-2.216*	0.410	29.167	-5.401	<0.00001	0.109	0.048	0.240
RegionWestern	-0.271	0.344	0.618	-0.786	0.432	0.763	0.387	1.494
Base(Gender_of_childFeMale)								
Gender_of_childMale	-0.925*	0.155	35.710	-5.976	<0.00001	0.397	0.292	0.536
Base(Delivery_place_Gov. clinic)								
Delivery_placeGovernment hospital	0.221	0.198	1.246	1.116	0.264	1.248	0.847	1.843
Delivery_placeMaternity home	-1.200	0.664	3.266	-1.806	0.709	0.301	0.072	1.019
Delivery_placeOther	1.074	0.863	1.549	1.245	0.213	2.927	0.581	16.176
Delivery_placePrivate hospital, clinic	0.463	0.353	1.717	1.310	0.190	1.588	0.797	3.185
Delivery_placeRespondent's home	1.516	0.304	24.937	4.994	0.124	4.553	2.526	8.314
Base(Birth_typeMultiple birth)								

Birth_typeSingle birth	-6.848*	0.647	111.986	-10.582	<0.00001	0.001	0.000	0.003
Birth_order	-0.714*	0.161	19.687	-4.437	<0.00001	0.490	0.356	0.669
Mothers_height	-0.107*	0.019	32.086	-5.664	<0.00001	0.898	0.865	0.932
Base(Wealth_indexAverage)								
Wealth_indexPoorer	0.447	0.282	2.513	1.585	0.113	1.564	0.900	2.722
Wealth_indexPoorest	0.680*	0.305	4.964	2.228	0.026	1.973	1.086	3.596
Wealth_indexRicher	0.291	0.270	1.156	1.075	0.282	1.337	0.787	2.274
Wealth_indexRichest	-0.179	0.323	0.307	-0.554	0.580	0.836	0.444	1.575
Base(Size_of_childAverage)								
Size_of_childVery small	2.277*	0.224	103.331	10.189	<0.0001	9.752	6.371	15.332
Size_of_childSmaller than average	1.540*	0.149	106.824	10.322	<0.0001	4.666	3.494	6.274
Size_of_childLarger than average	-1.486*	0.151	96.846	-9.844	<0.0001	0.226	0.168	0.303
Size_of_childVery large	-3.404*	0.403	71.346	-8.443	<0.0001	0.033	0.014	0.069
Note: Base is the reference category						* means $p < 0.05$		



4.5.2.3 Summary of our Logistic Model Fit

Considering low weight (< 2.5kg) and normal weight (≥ 2.5 kg), results as depicted in Table 71 shows that variables found to be significantly related to birthweight of a newborn are:

Table 72: Significant Variable in the Logistic model

Variable Entered	Estimated (B)	Odds ratio	Sig.	Overall
Intercept	12.014	33.043	<0.0001	
Region=Brong Ahafo	-1.375	0.253	<0.0001	2.0484976
Region=Central	-1.975	0.139	<0.0001	3.5554126
Region=Greater Accra	2.274	0.103	<0.0001	4.01998
Region=Upper East	-1.109	0.330	0.0022	1.6830045
Region=Volta	-2.216	0.109	<0.0001	3.3761294
Gender of child=Male	-0.925	0.397	<0.0001	4.5074859
Birth type=Single birth	-6.848	0.001	<0.0001	14.130663
Birth order	-0.714	0.490	<0.0001	5.6868232
Mothers height	-0.107	0.898	<0.0001	4.0588195
Wealth index=Poorest	0.680	1.973	0.0258	3.60798
Size of child=Very small	2.277	9.752	<0.0001	10.1894
Size of child=Smaller than average	1.540	4.666	<0.0001	10.322
Size of child=Larger than average	-1.486	0.226	<0.0001	9.84358
Size of child=Very large	-3.404	0.033	<0.0001	8.44257

***Overall = Variable Importance**

Variables (Entered and Removed) Interpretation

For every one-unit change in mothers height, our log odds of low birthweight (against normal birthweight) of a child decreases by -0.107kg with an odds ratio of 0.898kg. Likewise, for a unit decrease in the birth order of the infant, the log odds of being low weight decreases by -0.714kg with an odds ratio of 0.490kg.

The indicator variables for Region depicts that Brong Ahafo, Central, Greater Accra, Upper East, Volta are found to be significantly related to birthweight of a newborn with respective estimated (B) as [-1.375, -1.975, 2.274, -1.109, -2.216]. It implies that the newborn whose mother comes from Brong Ahafo would 0.253 times more likely to be normal weight as compared to a mother from the Ashanti. Similarly, the newborn whose mother's region of residence is at Central would 0.139 times more likely to be NBW than whose mother is from the Ashanti. The newborn whose mother is from Volta would 0.109 times more likely to be normal weight than whose mother is from the Ashanti. Infants of mothers from Greater Accra would 0.103 times more likely to be NBW as compared to mothers from Ashanti. Finally, the newborn whose mother's region of residence is from Upper East would 0.330 times more likely to be NBW than whose mother is from the Ashanti. In terms of odds, the result can be concluded as follows.

For the mothers' region, we would conclude that newborn whose mothers' region of residence is in Brong Ahafo, Central, Greater Accra, Upper East, Volta are less likely to be LBW, compared to mothers' region of residence is in Ashanti (the reference category). The odds of newborn being LBW whose mother's Region is in Brong Ahafo are $[0.253 - 1] * 100\% = -74.7\%$, i.e. 74.5% lower than the odds of being LBW whose mother's region is in Ashanti.

The indicator variables for the gender of the child indicated that male is found to be significantly related to birthweight of a newborn with an estimated (B) value of as -0.925. It implies that the newborn whose gender is male would 0.397 times more likely to be NBW as compared to a female child. The odds of newborn being LBW whose gender is male are $[0.397 - 1] * 100\% = -60.3\%$, i.e. 60.3% lower than the odds of being LBW whose gender is female.

The indicator variables for birth type showed that Single birth is found to be significantly related to birthweight of a newborn with an estimated (B) value of as -6.848. It implies that the newborn who came as single born was 0.001 times more likely to be NBW as compared to multiple births (as birth with more than one child).

The indicator variables for maternal wealth index depicted that only mothers from the poorest class are found to be significantly related to birthweight of a newborn with an estimated (B) value as 0.680. It implies that the newborn whose mothers are classified under the poorest class would 1.973 times more likely to be NBW as compared to mothers who belong to the middle class of wealth index.

Finally, the indicator variables for the size of a child depicted that sizes of child [Very Small, Smaller than average, Larger than average and Very Large] are found to be significantly related to birthweight of a newborn with respective estimated (B) as [2.277, 1.540, -1.486, -3.404]. It implies that the newborn whose size at birthweight is Very Small would 9.752 times more likely to be NBW as compared to a child whose size at birth is average. Similarly, the newborn whose size at birth is Smaller than average would 4.666 times more likely to be NBW than whose child whose size at birth is average. The newborn whose size at birth is Larger than average would 0.226 times more likely to be NBW than whose birth size is average. Finally, the newborn whose birth size is very large would 0.033 times more likely to be NBW than infants with average birth size.

4.5.2.4 Constructing the Logistic Model fit line

Each predictor in our logistic regression model fit provides us a coefficient 'B' which measures its individual contribution of variations on our dependent variable. The study noted that our dependent variable can take only on one of the two values: 0 or 1, therefor our logistic regression model is as follows.

$$\ln \left[\frac{\pi}{1-\pi} \right] = 1.181(\text{Intercept}) + 0.026(\text{Reg} = \text{B.Ahafo}) + 0.017(\text{Reg} = \text{Central}) + 0.016(\text{Reg} = \text{Eastern}) + 0.185(\text{Reg} = \text{G.Accra}) - 0.054(\text{Reg} = \text{Northern}) + 0.111(\text{Reg} = \text{U.East}) - 0.030(\text{Reg} = \text{U.West}) + 0.278(\text{Reg} = \text{Volta}) + 0.054(\text{Reg} = \text{Western}) + 0.071(\text{GoC} = \text{Male}) + 0.061(\text{Del.type} = \text{Yes}) - 0.038(\text{CeB} = \text{Moderate}) - 0.124(\text{CeB} = \text{Small}) + 0.605(\text{B.type} = \text{Single}) + 0.002(\text{MWeight}) + 0.007(\text{MHeight}) - 0.072(\text{W.index} = \text{Poorer}) - 0.047(\text{W.index} = \text{Poorest}) - 0.014(\text{W.index} = \text{Richer}) - 0.023(\text{W.index} = \text{Richest}) + 0.209(\text{SoC} = \text{Larger than average}) - 0.305(\text{SoC} = \text{Smaller than average}) + 0.562(\text{SoC} = \text{Very large}) - 0.575(\text{SoC} = \text{Very small})$$

Where π is probability of low birthweight, Reg = Region, Del.type= Delivery typ, B.type=Birth type, GoC=Gender of child, CeB=Children ever born, MWeigh=Mothers weight, MHeight=Mothers height, W.index=Wealth index and Soc = Size of child.

4.5.2.4 Classification Rate of our Logistic Model

How good the classification model (after including all the independent variables) is? The answer to this question is given in a classification table shown in Table 67. One way to assess how well our model fits the observed data is to obtain a classification table on our test dataset. This simple technique would indicate how good our model is at predicting our outcome variable.

First, a “cut-off” value $c=0.5$ will be chosen. For each subject in the sample, newborns are “predicted” birthweight status as 0 (coded as normal weight) if their fitted probability of being normal birthweight is greater than c , otherwise predicted it as 1 (coded as low weight). The table constructed below shows how many of our test data had a correct prediction.

Table 73: Classification Table on the Test data

Observed			Predicted		
			Birthweight group		Percentage Correct
			NB	LB	
Final Step	Birthweight group	NB	822	62	93.0
		LB	28	97	77.6
	Overall Percentage				91.08

a. The cut value is .500

We should note that seven explanatory variables were used in the fitted model above for our test data. In the study, we noticed that our model percentage of correct predictions reported in Table 73 is 91.08% of the observations in the testing set, indicating an improvement on the full data reported earlier in table 69.

Table 73 shows that out of 884 cases predicted to be the NBW group, only 822 cases are observed to be in the NBW group, while 62 in the LBW group. Similarly, out of 125 cases predicted to be LBW group, only 97 cases are correctly classified as LBW, while 28 in the NBW group. So out of 1009 cases, only 919 (822 + 97) are correctly classified and 90 (62+28) cases are misclassified. From this, it can be said that $(1009 - 90) / 1009$, or 91.08% of the cases are correctly classified with this model. The researcher can predict with 91.08% accuracy on the test data. Our proportional by chance accuracy rate was calculated by computing the proportion of cases for each group based on the number of cases in each group in the classification table.

The proportion in the "LBW" classification is $97/1009 = 0.0961$. The proportion in the "NBW" classification is $822/1009 = 0.815$. Then, we square and sum the proportion of cases in each group ($0.0961^2 + 0.815^2 = 0.673$). That is 67.3% is proportional by chance accuracy rate. Our accuracy rate was reported as 91.08% which is greater than or equal to the proportional by chance accuracy criteria of 84.13% ($1.25 * 67.3 = 84.13\%$). Hence, the criterion for classification accuracy is satisfied.

CHAPTER FIVE

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.0 Introduction

Results of both the longitudinal and low birthweight data are summarized in relation to other findings, regarding limitations and strength of the study. The conclusions, recommendations, and future research are also discussed.

5.1 Summary

There are social and economic costs attached to managing infants born with a birthweight of less than 2.5kg. The 10%-15% prevalence of low birthweight in the Ghanaian population is a cause for alarm. However, establishing maternal level characteristics associated with infants of low weight provides a starting point to identify other key risk factors and modify to improve birth outcomes of neonates. The purpose of this research was to establish indicators of birthweight in an estimation and a classifier way using data from mother-child data from GDHS.

A very wide spectrum of factors including physical, demographic, social, economic as well as previous obstetric factors has been considered to identify their relationship with the birthweight of the child. The study started with the explanation of the different type of factors considered and the exploration of their place in the determination of birthweight. The problem is considered in its complete form by considering all possible type of factors having a bearing on the birthweight.

Research Questions addressed in the present investigation are (i) What are the prevalence rates of birthweight (low and normal)? (ii) What is the type of birthweight reporting system in Ghana? (iii) Can we investigate the use of the Mother's perception of child size at birth (variable) as a proxy to birthweight of an infant if weigh data is missing? (iv) Which factors contribute to and influences the birthweight of newborns in Ghana? (v) If possible, can we hypothesize a statistical model in

predicting the likelihood of a child's birthweight belonging to a low or normal class using chosen factors such as maternal factors, outcomes of birth and maternal anthropometric variables?

5.1.1 Key Observations from Study Objectives

1. The mean birthweight for the study was 3.137kg, median birthweight was 3.130kg with a standard error of 0.01kg and standard deviation of 0.593kg. The minimum birthweight recorded is 0.8kg while the maximum is 5.5kg. The male infant recorded a higher average birthweight (3.195kg) compared to female infants (3.074kg).
2. Towards answering the first research question on the prevalence rate of birthweight, the statistics on low birthweight was recorded as 9.9% using the WHO's definition of birthweight classification while the Researcher's Adjusted Prevalence reported an adjusted low infant weight at 15.7%, a 5.8% increase in the WHO's low birthweight recommendation.
3. The second research question revealed that almost two-fifth births in Ghana were reported at birth through mother memory recall. This finding suggests that the picture of reporting of birth weight in Ghana is still challenging. There is a need to have a formal procedure or recording of birth weight even birth outside a health facility. These findings corroborate with the study by Channon et al., which suggested that health systems in poor countries should initiate efforts to systematically monitor the recording of birth weight data ensuring for both quality and comparability at the international levels (Channon et al., 2011). Based on the result from the type of birthweight reporting system, it might be hypothesized that birthweights extracted from a health card were more accurate than those acquired from memory recall.
4. Moreover, in developing countries majority of birth occurs outside health facilities. The estimates of birth weight are prone to biases in the form of being inaccurate in measurement;

imprecise in methods of reporting and varying background characteristics that influences the reporting on part of mother. These measurement issues can substantially distort the actual prevalence of LBW and hence the intervention formulated on this empirical information remains ineffective in practical sense. Therefore, accurate reporting of prevalence of LBW is important for monitoring the health of a population. It was reported that there were portions of proof that the size at birth of the neonates as disclosed by the mother may count as a possibility of a good proxy of birthweight in case, we have missing information of birth weight or unrecorded birth weight, even or deliveries in the institutional settings.

5. On account of the research in attempt to answer the forth research question stating which factors contribute to and influences the birthweight of newborns in Ghana, the study identified numerous risk factors for LBW, covariates found to statistically associated with newborn's weight are Gender (0.02), Size at birth (0.000), Region (0.02), Wealth index (0.043), Birth order (0.02), Children ever born (0.039), Delivery place (0.011), Preceding birth interval (0.000), and Birth type (0.000). For parsimony, the significance level was established at $p < 0.05$, a 95% confidence interval (CI).

5.1.2 Final Comments on Multiple Linear Regression fit

The objective of the research was to build a model that allows us to estimate the separate effects of some maternal characteristics and birth outcomes on an infant's birthweight. The following seventeen independent variables were considered: Gender of child, Type of delivery by caesarean, Size of child at birth, Mothers age, Children ever born, Birth type, Maternal anthropometry (Weight and height), Maternal education, Mother's region of residence, maternal religion, type of residence, Birth order, Delivery Place, Preceding birth interval and Wealth index.

We examined the relationship between neonate's birthweight and the seventeen independent variables in the multiple linear regression modeling. Thereafter, a variable selection was conducted and found to fish out significant predictors for the model building. It converged in the idea that, of the seventeen variables, nine had p-values less 0.05.

We found that over 46 % of the variation in newborn birthweight is clarified by these nine covariates. The value of the determination coefficient (0.46) indicates that the model can be improved by adding new significant variables. The above model demonstrated a very strong effect of size of a child at birth on infant birthweight subsequent to representing both maternal and birth outcome factors. The region, Mothers height, Mothers weight, Gender of a child, Delivery type, Wealth index, Children ever born, and Birth type is likewise important contributors. However, the remaining eight variables: Mothers age, Maternal education, Type of residence, Maternal Religion, Birth order, Delivery Place and Preceding birth interval were found to be nonsignificant contributors.

The comparison of our nine predictor variables by means of individual t-statistics depicts the relative magnitude of the unique contribution of each variable to the overall variability in birthweight. According to the above output, the Size of child=Very Large is the largest contributor to the explained variation in birthweight. The regression coefficient associated with the Size of child=Very Large is 0.5621383 with a corresponding t-ratio of 22.231, indicating a very strong effect of infants' weight after accounting for the effect of maternal and birth outcome variables. Birth type=Single (15.402), Size of child=Larger than average (10.521) and Region=Volta (7.726) are the next three most important contributors.

The linear fit was backed by a normality test. Our model was free from multicollinearity. Few unusual data points that affected our initial fit were treated and hence we can be dependent on the inference made from the linear fit.

5.1.3 Final Comments on Logistic Regression fit

Likewise, the study examines impacts of the seventeen predictors (maternal characteristics, the outcome of child and obstetric factors) on infant and low birthweight in Ghana using chi-square analysis, binary logistic, likelihood, and odds ratio tests corresponding with the level of significance. Out of these seventeen predictors, only seven were significantly related to low birthweight. We found that between 56% to 75% of the statistical variation in infant low birthweights is explained by seven covariates using the three metric strength test score in table 73 above. The value of the determination coefficients above indicates that the model can still be improved by adding new significant variables.

The findings suggest that birth type of child status has been discovered as the most significant obstetric predictor of newborn's weight determination. That means the birthweight rate decreases with multiple birth type. Further, some maternal characteristics variables have a considerable influence on newborn birthweight. Among these variables include the region of residence and maternal classification of wealth were found to be having contributions to a child's weight. That is, children whose mother is from Greater Accra, Volta, Brong Ahafo, Upper East, Central and from the richer home have a lower risk of low weight at birth. Factors like the mother's height and birth order of child are likewise found to be significant factors to determine infants' weight.

In conclusion, the findings of the research demonstrate that different factors such as Region, Mothers height, Gender of a child, Wealth index, size of a child, Children ever born, and Birth type have statistically significant impacts to determine infants' weight in Ghana. The logistic fit

was backed by its assumptions. Our model was multicollinearity free. Few unusual data points that affected our initial fit were treated, henceforth we can depend on the inferences produced using the logistic fit.

5.1.4 Comparison of Two Sets of Fits

Nearly similar factors appear to be significant with its direction of influence perfectly coordinating. Among longitudinal and low birthweight dataset, a few factors were common. Among the explanatory factors: Size at birth, Region, Mothers height, Gender of child, Delivery type, Wealth index, Children ever born and Birth type emerge as significant, while Mothers age, Maternal education, Type of residence, Maternal Religion, Birth order, Delivery Place and Preceding birth interval were not statistically significant. The study discovered only one opposite insignificance of the variable mother's weigh, the longitudinal data, mothers weight seems to be significant while mothers weight turned out to be insignificant in the low birthweight data.

5.2 Conclusions

In the Ghanaian context, this study has contributed to the understanding of maternal determinants associated with infant birthweight at the population level. Findings from this study have, therefore, provided a starting point towards identifying risk factors and providing clues to health service providers on maternal determinants and birth outcomes factor to concentrate health promotion messages on. Almost all identified risk factors are for low birthweight in the study is modifiable and preventable. The problem is multidimensional and hence need an integrated approach incorporating medical, social, economic, and educational measures to address this issue. In conclusion, a comprehensive approach which institute a combination of interventions to improve the overall health of women are needed. Such approaches are likely to be most effective in reducing the low birthweight of infants in Ghana.

5.3 Recommendations

The study closed by pointing out some significant limitations to the extent of our study, and proposed areas for further studies.

- Better pregnancy outcomes can be expected by providing adequate antenatal care and nutrition, effective management of complications and the provision of proper family planning services for proper spacing and family size.
- Regional and community educational programmes should be organized to educate women on the importance of prenatal care.
- Access to quality antenatal and health facilities should be accessible in all regions.
- More micro level analyses (both Qualitative and Quantitative) in establishing the associations between maternal health and their socio-economic, cultural and programmatic antecedents would go a long way in tailoring services in a locally relevant way.
- While we have focused on three key influences of birthweight distribution (maternal factors, the birth outcome of child and obstetric history), there are without a doubt different indicator that impact birthweight. The future researcher has a scope to include more psychological and genetic factors to study the effect on birthweight.
- The present study looked at data for the 2014 GDHS only. To generate trends and have a better understanding of the nature of determinants across time periods, a similar analysis needs to be done with previous and future GDHS data.
- Given the small numbers of low birthweight, no attempt was made to explore data on whether an associated existed between regional categories and low birthweight. Future studies to consider large sample sizes so that urban and rural dynamics can be explored in greater detail and depth.

- Finally, our results center around the status of the newborn child during childbirth, or not long after birth. Along these lines, we can't make any decisions about the connection between low birthweight (or birthweight) and longer-run outcomes, for example, such as cognitive development, educational attainment, and adult health. A more straightforward investigation is by all accounts a valuable bearing for future studies.



REFERENCES

- Abrams, B., Hoggatt, K., Kang, M., & Selvin, S. (2008). History of weight cycling and weight changes during and after pregnancy. *Paediatric and Perinatal Epidemiology*, 15(4), A1-A1. doi:10.1111/j.1365-3016.2001.381-1.x
- Antwi – Boasiako Ishmael (2011). Assessing the Risk Factors Associated with Low Birthweight (LWB) And Mean Actual Birthweight of Neonates: A Case Study of St. Martin's Hospital, Agroyesum. Master of Science. Kwame Nkrumah University of Science and Technology, Kumasi
- Atinuke O. Adebajji, A., & Puurbalanta R, A. (2015). Determinants of Low Birth Weight Neonates: A Case Study of Tamale Metropolis in Ghana. *Journal for Studies in Management and Planning*, 90-102..
- Avchen, R. N. (2001). Birth Weight and School-age Disabilities: A Population-based Study. *American Journal of Epidemiology*, 154(10), 895-901. doi:10.1093/aje/154.10.895
- Babaei, Z., Rejali, M., Mansourian, M., & Eshrati, B. (2017). Prediction of low birth weight delivery by maternal status and its validation: Decision curve analysis. *International Journal of Preventive Medicine*, 8(1), 53. doi:10.4103/ijpvm.ijpvm_146_16
- Balakrishnan, N., Barnett, V., & Lewis, T. (1995). Outliers in Statistical Data. *Biometrics*, 51(1), 381. doi:10.2307/2533352
- Barker, D. (2002). EDITORIAL: The developmental origins of adult disease. *European Journal of Epidemiology*, 18(8), 733-736. doi:10.1023/a:1025388901248
- Barker, D. J. (2011). Fetal Origins of Adult Disease. *Fetal and Neonatal Physiology*, 192-197. doi:10.1016/b978-1-4160-3479-7.10018-7

- Basso, O., Olsen, J., Johansen, A. M., & Christensen, K. (1997). Change in social status and risk of low birth weight in Denmark: Population-based cohort study. *BMJ*, *315*(7121), 1498-1502. doi:10.1136/bmj.315.7121.1498
- Carr-Hill, R., & Pritchard, C. (1985). The Development and Exploitation of Empirical Birthweight Standards. doi:10.1007/978-1-349-07434-1
- Chatterjee, S., & Hadi, A. S. (1986). Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, *1*(3), 379-393. doi:10.1214/ss/1177013622
- David W. Hosmer, J., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Hoboken, NJ: John Wiley & Sons.
- Evelyn Yaa Dufie Mensah (2015). Statistical Analysis of Retroviral (HIV) Status and Other Maternal Risk Factors Associated with Low Birthweight and Low Apgar Score of Infants: Evidence from the Greater Accra Regional Hospital. Master of Philosophy. University of Ghana
- Fonseca, W., Kirkwood, B. R., Barros, A. J., Misago, C., Correia, L. L., Flores, J. A., ... Victora, C. G. (1996). Attendance at day care centers increases the risk of childhood pneumonia among the urban poor in Fortaleza, Brazil. *Cadernos de Saúde Pública*, *12*(2), 133-140. doi:10.1590/s0102-311x1996000200002
- Hilbe, J. M. (2016). Practical Guide to Logistic Regression. doi:10.1201/b18678
- Kanade, A., Rao, S., Yajnik, C., Margetts, B., & Fall, C. (2005). Rapid assessment of maternal activity among rural Indian mothers (Pune Maternal Nutrition Study). *Public Health Nutrition*, *8*(6), 588-595. doi:10.1079/phn2004714

- Kayode, G. A., Amoakoh-Coleman, M., Agyepong, I. A., Ansah, E., Grobbee, D. E., & Klipstein-Grobusch, K. (2014). Contextual Risk Factors for Low Birth Weight: A Multilevel Analysis. *PLoS ONE*, *9*(10), e109333. doi:10.1371/journal.pone.0109333
- Kramer, M. S. (1987). Determinants of Low Birthweight: Methodological Assessment and Meta-Analysis. *Bull World Health Organization*, *65*, 663–737
- Leppert, P. C., Namerow, P. B., & Barker, D. (1986). Pregnancy outcomes among adolescent and older women receiving comprehensive prenatal care. *Journal of Adolescent Health Care*, *7*(2), 112-117. doi:10.1016/s0197-0070(86)80006-7
- Logan, M. (2010). Biostatistical Design and Analysis Using R. doi:10.1002/9781444319620
- MacLeod, S., & Kiely, J. (1988). The effects of maternal age and parity on birthweight: a population-based study in New York City. *International Journal of Gynecology & Obstetrics*, *26*(1), 11-19. doi:10.1016/0020-7292(88)90191-9
- Malhotra, N., & Dash, S. (2013). Future of research in marketing in emerging economies. *Marketing Intelligence & Planning*, *31*(2). doi:10.1108/mip.2013.02031baa.001
- Murray, Barbara A. (1999) *A Statistical Analysis of Low Birthweight in Glasgow*. Ph.D. Thesis.
- Noora Nidhal Saleh (2016). *Nature and Causes of Low Birthweight of Babies: A Statistical Analysis*. Master of Science. Ball State University Muncie, Indiana.
- Paul Nesara (2018). Determinants of Low Birthweight in A Population-Based Sample of Zimbabwe. (Doctoral Studies). Walden University, USA
- Porter, T., Fraser, A., Hunter, C., Ward, R., & Varner, M. (1997). The risk of preterm birth across generations. *Obstetrics & Gynecology*, *90*(1), 63-67. doi:10.1016/s0029-7844(97)00215-9

- Sareer Badshah (2007). Exploratory Analysis of Low Birthweight Data from a Survey of Births Delivered During 2003 at Four Main Public-Hospitals in Peshawar.
- Shea Oscar Rutstein (2006). Guide to DHS Statistic, Demographic and Health Surveys Methodology, ORC Macro Calverton, Maryland
- Sommerfelt, K., Troland, K., Ellertsen, B., & Markestad, T. (2008). Behavioral Problems in Low-Birthweight Preschoolers. *Developmental Medicine & Child Neurology*, 38(10), 927-940. doi:10.1111/j.1469-8749.1996.tb15049.x
- Tampah-Naah, A., Anzagra, L., & Yendaw, E. (2016). Factors Correlate with Low Birth Weight in Ghana. *British Journal of Medicine and Medical Research*, 16(4), 1-8. doi:10.9734/bjmmr/2016/24881
- Tema, T. (2006). Prevalence and determinants of low birth weight in Jimma zone, Southwest Ethiopia. *East African Medical Journal*, 83(7). doi:10.4314/eamj.v83i7.9448
- Wardlaw, T. M. (2004). *Low Birthweight: Country, Regional and Global Estimates*. UNICEF. UNICEF & WHO: Reduction of Low Birthweight: A South Asia Priority, 2002. United Nations Children's Fund - Regional Office for South Asia.
- UNICEF (2008) Global Database on Low Birthweight. Low Birthweight Incidence by Country (1999-2006)
- UNICEF-WHO (2004) United Nations Children's Fund and World Health Organization, (2004) Low Birthweight; Country and Global Estimates. UNICEF, New York
- Wilcox, A. J. (2001). On the importance—and the unimportance— of birthweight. *International Journal of Epidemiology*, 30(6), 1233-1241. doi:10.1093/ije/30.6.1233

Wilcox, A. J., & Skjaerven, R. (1992). Birth weight and perinatal mortality: the effect of gestational age. *American Journal of Public Health*, 82(3), 378-382.

doi:10.2105/ajph.82.3.378

Zupan, J., Åhman, E., & World Health Organization. (2006). *Neonatal and Perinatal Mortality: Country, Regional and Global Estimates*. WHO.



APPENDICES

Appendix 1: Group Statistics for Birthweight Reporting Type

	Reporting Type	N	Mean	Std. Deviation	Std. Error Mean
Birthweight	From mother's recall	719	3.2304	.7292	.0272
	From written card	2642	3.1114	.5472	.0107

Appendix 2: Independent Samples Test for Birthweight Reporting type

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% CI of the Difference	
									Lower	Upper
Birthweight	Equal variances assumed	86.810	.000	4.790	3359	.000	.1190	.0249	.0703	.1678
	Equal variances not assumed			4.076	948.861	.000	.1190	.0292	.0617	.1764



Appendix 3: Boxplot of Birthweight data

