KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

IDENTIFYING HIGH-RISK GROUPS FOR ANEMIA AMONG WOMEN OF REPRODUCTIVE AGE IN GHANA: A CLUSTER ANALYSIS APPROACH

BY:

OSEI NYANKOE PASCALINE

AMPONSAH DESMOND

AZIAFOR SARPOMAA TRACY

QUARSHIE HENRY EYRAM

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF STATISTICS

AUGUST 2024

DECLARATION

We hereby declare that this submission is the result of our own work towards obtaining the undergraduate Statistics degree. To the best of our knowledge, it does not contain any material previously published by another individual or material that has been accepted for the award of any other degree at the university, except for the permissible citation or reference from other sources, which have been duly acknowledged.

Amponsah Desmond	....	....
(Student)	(Signature)	(Date)
Osei Nyankoe Pascaline	....	....
(Student)	(Signature)	(Date)
Aziafor Sarpomaa Tracy	....	....
(Student)	(Signature)	(Date)
Quarshie Henry Eyram	....	....
(Student)	(Signature)	(Date)

Certified by:

Professor Atinuke O. Adebanji	....	....
(Supervisor)	(Signature)	(Date)

Certified by:

Professor Asare Okyere Gabriel	....	....
(Head of Department)	(Signature)	(Date)

DEDICATION

This study is dedicated to the almighty God for His divine love and grace on our lives and for granting us the strength to complete this study. We also dedicate this study to our wonderful families who have been our source of inspirations and support throughout this journey. Their continuous moral, spiritual, emotional and financial support helped us finish this research. We also dedicate this piece of work to our amazing supervisor who invested her precious time, tremendous effort and knowledge in helping us come out with this research. Lastly, we dedicate this study to our acting supervisor and his teaching assistant who helped us from time to time and shared words of advice and encouragement that motivated us to complete this work. May God bless them richly.

ABSTRACT

Most of the insurance contracts in Ghana contains the right to early termination and are also path-depend. Due to the presence of path-dependence derivatives and the right to early termination of the contract can make the valuation of Life insurance contract in Ghana come with complexities. These complexities are aggravated with introduction of new parameter(s). Termination of life insurance contract in Ghana may come as a result of comorbidity.

The paper seeks to modify Black-Scholes partial differential to incorporate risk of being comorbid and investigate the suitability of using some existing numerical methods, Crank-Nicolson and Hopscotch to solve the valuation of life insurance contract which offer the policy holder the option to terminate the contract at some time before maturity. Further Comparison between the two methods were done to select an efficient method for the modified model.

In line with these objectives, simulations for time of an individual to be comorbid were performed and the survival for risk of comorbidity computed.

The study revealed that, the modified model is stable, consistent and hence suitable to solve.

Further Crank-Nicolson method is closer to the solution as values of step sizes are increased for Black-Scholes partial difference equation of the life insurance contract in Ghana with embedded surrender option. Please give a better results and conclusion based on the work.

ACKNOWLEDGMENT

We want to thank the Almighty God who gave us the strength and grace throughout this work. Special thanks also goes to our supervisor Professor Atinuke O. Adebanji for her support, insightful criticisms, suggestion, motivations and her efforts throughout this research. We would also like to acknowledge our indebtedness and our warmest thanks to our acting supervisor Mr. Emmanuel Odame and his supportive teaching assistant Michael Asante Ofosu for their support and contribution towards this research. Lastly, we would like to thank our families, friends and course mates at KNUST for their support and advice in all forms which has led to the success of the entire work.

CONTENTS

LIST OF ABBREVIATION

WRA .... Women of Reproductive Age

WHO .... World Health Organization

LMICs .... Low- and Middle-Income Countries

GBD .... Global Burden of Disease

SDI .... Socio-Demographic Index

GDHS .... Ghana Demographic and Health Survey

RBCs .... Red Blood Cells

ITNs .... Insecticide-Treated Nets

IPT .... Intermittent Preventive Treatments

NFHS .... National Family Health Survey

MICS .... Multiple Indicator Cluster Survey

MOH .... Ministry Of Health

USAID .... United States Agency for International Development

Hb .... Hemoglobin

FHD .... Family Health Division

GSS .... Ghana Statistical Service

GHS .... Ghana Health Service

GPS .... Global Positioning System

HCA .... Hierarchical Clustering Algorithm

ANOVA .... Analysis Of Varaiance

WCSS .... Within-Cluster Sum of Squares

DBI .... Vavies-Bouldin Index

MI .... Mutual Information

GDPR .... General Data Protection Regulation

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Anemia is a global public health problem Wu et al. (2020). It affects an estimated 1.7 billion people worldwide, with the bulk of the infections occurring among women and young children in Sub-Saharan Africa and South Asia Hulland et al. (2021). Anemia prevalence is highest in developing countries Macdonald et al. (2010). It is a disorder where the hemoglobin (Hb) concentration, or the quantity and size of red blood cells, falls below a predetermined threshold, reducing the blood's ability to carry oxygen throughout the body Tirore et al. (2023). Hemoglobin is the primary oxygen-carrying molecule within red blood cells, so anemia is most typically measured in terms of hemoglobin content of the blood rather than red blood cell count Kinyoki et al. (2021). Although both males and females of all ages are affected, the most vulnerable groups are pregnant women and young children Macdonald et al. (2010). Previous research shows that women of reproductive age (WRA) are more likely to get anemia because they lose blood during menstruation and childbirth Alem (2023). According to the World Health Organization (WHO, 2008), the threshold hemoglobin (Hb) level for anemia is less than 120 g/l for nonpregnant women and 110 g/l for pregnant women aged 15 years and above Keokenchanh et al. (2021). Based on Hb level, anemia is categorized as mild, moderate, or severe Tirore et al. (2023). Anemia is a complex condition with many causes (`www.who.int`, accessed 09/05/24). It can be caused by both nutritional and non-nutritional factors, with iron deficiency being the most common cause Wu et al. (2020). The nutritional factors are mainly micronutrient deficiencies, including folate, vitamin B12, and vitamin A Vroom et al. (2024). Globally, anemia can be found in more than one-third of women and more than 40% of children under five years of age; primarily in those living in rural households with low socioeconomic status and exposed to poor sanitation Christian et al. (2022). Causes of anemia can be divided into three non-mutually exclusive pathways: blood loss, increased red blood cell destruction, and inadequate red blood cell production Kinyoki et al. (2021). The risk of anemia during adolescence is higher when women become pregnant Alem (2023). Some symptoms of anemia may include fatigue (extreme tiredness; can be mild to severe), difficulty in breathing, rapid heart rate, pale skin, feeling cold, loss of concentration, dizziness, lips, gums, lining of the eyelids, nail beds, and palms are less pink than usual (www.thewellproject.org, accessed 06/06/24).

In Ghana, anemia in children under 5 years was estimated at 76.1%, 55.9% in non-pregnant women, and 62.4% among pregnant women from WHO estimates in 2011 Quaye (2016). The most common condition pregnant women present with at the health facilities is anemia Sumaila et al. (2022). Several studies have found various factors associated with an increased risk of developing anemia in women of reproductive age (WRA) such as consumption of well water, having permanent sterilization, living in small and medium-sized cities, aged between 30–39 years, no or occasional smoking, spring and winter seasons, ethnicity, and having a low education level Keokenchanh et al. (2021). Smoke from burning biomass fuel is another source of inflammation, which has also been implicated as a possible cause of anemia in a few studies with children and women Armo-Annor (2021). Again, the high prevalence of maternal anemia can be associated with a lack of family planning, poor dietary intake, blood loss during menstruation, and persistent infections Tettegah et al. (2023).

Reducing anemia is considered an essential part of improving the health of women, and the WHO has set a global target of achieving a 50% reduction of anemia among women of reproductive age by 2025 Kibret et al. (2025). Measures being undertaken to reduce anemia among vulnerable groups include iron and folic acid supplementation; anti-malaria prophylaxis for pregnant women; promotion and the use of insecticide-treated bed-nets by pregnant women and children under five; and six-month deworming for children 2-5 years since malaria and parasitic worms are also known causes of anemia Omari (2017).

Anemia continues to affect millions of women worldwide and remains concentrated in low- and middle-income countries (LMICs) as defined by the Global Burden of Disease (GBD) Socio-Demographic Index (SDI) Kinyoki et al. (2021). To decrease the burden of anemia, it is necessary to generate adequate evidence in terms of the role and contribution of individual, household, and community-level factors along with the geographical risk profile of anemia Sunuwar et al. (2021). Moreover, although it is well known that anemia results from multiple causes, there are few reported examples of integrated programs addressing the various causes or assessments of the effectiveness of combining several interventions on anemia prevalence among women Macdonald et al. (2010).

However, since the majority of research conducted focuses on the prevalence of anemia among pregnant women and children globally, with little attention paid to women of reproductive age, if this issue remains unaddressed, the problem of anemia will persist irrespective of the efforts made to prevent its widespread occurrence currently Agyemang et al. (2023). As a result, identifying the risk factors associated with anemia is essential to preventing its spread across the country and lowering its prevalence. Consequently, this study aims to assess the prevalence of anemia among women of reproductive age in Ghana, identifying the high-risk groups associated with anemia among women, taking into consideration some demographic factors such as anemia level, age, education level, place of residence, and region of residence; some socioeconomic factors such as wealth index, health insurance coverage, access to clean water and sanitation, household size, number of live births and birth interval; and some health factors such as history of malaria or other illnesses and nutrition.

1.2 Problem Statement

Anemia is a globally widespread condition in women and is associated with reduced economic productivity and increased mortality worldwide (www.nature.com, accessed 06/06/24). The prevalence of anemia in women of reproductive age has been stable worldwide since 2000. It accounts for nearly one-third of the worldwide cases Tirore et al. (2023). Although the prevalence of anemia among pregnant and nonpregnant women was reduced from 56.5% to 36.0% and from 46.1% to 31.0% in 2006 and 2011 respectively, it remained higher than the global average (29%) for nonpregnant women Keokenchanh et al. (2021). The significant socio-economic impact of anemia in Ghana makes it imperative to assess the factors contributing to the burden of the disease and hence provide empirical data to support strategies that will aim at reducing the burden of the disease Quaye (2016).

Multiple studies have been conducted to estimate the prevalence and magnitude of anemia among women of reproductive age and its associated factors in developing countries Savera (2020). Assessing the geographic distributions of anemia and the impact of risk factors on disease prevalence by area is important to prioritize and design targeted prevention and intervention programs to address anemia in women Kibret et al. (2019). Despite the prevalence of anemia in women of reproductive age in Ghana, there is limited understanding of specific groups that are at high risk. When anemia is left untreated, it can lead to severe health consequences. There is an imperative need to reduce the high anemia prevalence in the region and accelerate efforts

towards achieving the global nutrition target of 50% reduction of anemia in women of reproductive age by 2025 Vroom et al. (2024). Therefore, identifying high-risk groups can help target treatments and prevent its occurrence.
However, current ways of managing anemia in Ghana often focus on a collective prevention system without taking into consideration the distinct characteristics and needs for high-risk groups. This study seeks to identify high-risk groups for anemia among women of reproductive age in Ghana which will help prevent its occurrence, reducing the burden of anemia on individuals, families, and the healthcare system.

1.3 Aims and Objectives

1.3.1 Main Aim

To identify high-risk groups for anemia among women of reproductive age in Ghana using cluster analysis.

1.3.2 Specific Objectives

1. To describe the socio-demographic characteristics of each cluster.
2. To describe the health factors associated with each cluster.
3. To examine the relationships between cluster membership and anemia status.

1.4 Research Questions

1. What are the demographic characteristics of high-risk groups for anemia among women of reproductive age in Ghana?
2. What are the socioeconomic characteristics of high-risk groups for anemia among women of reproductive age in Ghana?

3. What are the health factors associated with high-risk groups for anemia among women of reproductive age in Ghana?

4. Which clusters have the highest or lowest rates of anemia?

1.5 Significance of the Study

The significance of this study lies in the fact that it will contribute to existing literature on anemia among women of reproductive age in Ghana, highlighting the characteristics of high-risk groups. Moreover, the findings will inform individuals, families, and healthcare systems to prevent the occurrence of anemia. By highlighting the unique characteristics of each cluster, this study provides a clear understanding of women at high risk of getting anemia and by so doing, informs individuals, families, and healthcare providers to address the specific needs of each group.

1.6 Organization of the Study

The study is organized into five chapters:

1. The study begins with an introduction that provides the background information of anemia among women of reproductive age in Ghana, its consequences, and some treatments. It also introduces the problem, objectives, research questions, significance of the study, and some limitations related to the study.

2. The second chapter reviews some literature on anemia among women of reproductive age in Ghana. It focuses on the determinants of anemia, highlighting the use of demographic factors, socioeconomic factors, and health factors. Demographic factors such as age, education level, place of residence, and region of residence are shown to influence anemia risk. Socioeconomic factors such as wealth index, household size, access to

water and sanitation, health insurance coverage, number of live births, and birth interval also play a crucial role. Health factors such as malaria, nutrition, and other illnesses can contribute to anemia.

3. The third chapter describes the materials and methods used in the study. It emphasizes the source of data, the study design, and also the type of clustering analysis approach used to identify the various groups of women with anemia.

4. The fourth chapter presents results and discusses the findings. It presents the characteristics of each cluster and interprets the findings.

5. The last chapter summarizes the whole work, providing a conclusion and some recommendations. It summarizes the key findings, limitations, and some recommendations for future research.

1.7 Limitations of the Study

The study has several limitations that should be considered when interpreting results. First of all, the study uses existing data from the Ghana Demographic and Health Survey (GDHS), 2022 which may have some restrictions when we consider the quality of the data, completeness, and accuracy. Additionally, the Ghana Demographic and Health Survey (GDHS) does not provide the entire population of women of reproductive age in Ghana, which may restrict the generalizability of results and also may be biased towards certain regions or demographics. Another limitation is the fact that the analysis uses a limited set of variables when forming the various clusters, which may not capture all potential factors associated with anemia among women of reproductive age in Ghana and the full complexity of anemia. Moreover, the analysis assumes that the clusters are homogeneous and distinct and the data used for clustering is normally distributed, which may not always be the case. There may be overlap between clusters or heterogeneity within the clusters which will not be captured by the analysis.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Anemia is a significant public health problem among women of reproductive age, affecting approximately 30% of women worldwide. It is a major cause of serious health issues in mothers and babies. Anemia is a blood disorder. It occurs if your body makes too few red blood cells (RBCs), destroys too many RBCs, or loses too many RBCs. Anemia can be caused by a variety of factors which may include lack of iron, lack of vitamins, chronic diseases, and some social factors. It is important to identify high-risk groups of anemia among women of reproductive age group so that the needed help will be made available to them. However, there are many literatures on this topic, and there is the need to review them and identify the various gaps in the research. This chapter provides a summary of existing research and findings on anemia among women of reproductive age. It will include the prevalence of anemia, the risk factors associated with anemia and the consequences of anemia among women. It aims to identify the gaps, establish a theoretical framework and provide a conceptual understanding of the topic. By reviewing existing literatures, our aim is to get a clear picture of high-risk groups for anemia among women of reproductive age and how best they can be helped.

2.2 Prevalence and Incidence of Anemia

Globally, 613.2 million women within the reproductive age (15–49 years) including 35.3 million non-pregnant women are affected by anemia Vroom et al. (2024). The global prevalence of anemia decreased from 40.2% in 1990 to 32.9% in 2010, with the number of prevalent cases of anemia being higher in females (than males) and among those under 5 years of age Safiri et al. (2021). In 2011, the World Health Organization (WHO) reported that the global prevalence of anemia for all women of reproductive age was 29.4%, with 38.2% in pregnant women and 29.0% in non-pregnant women Wu et al. (2020). Sub-Saharan Africa and South Asia have the highest anemia prevalence, and at the country level, anemia among women of reproductive age (WRA) remains a moderate-to-severe public health problem (prevalence of 20% or greater) in most WHO member states Owais (2021).

According to epidemiological data from developing countries, the estimated prevalence of anemia is 39% in children under 5 years, 48% in children 5-14 years, 42% in women 15-59 years, 30% in men 15-59 years, and 45% in adults above 60 years Quaye (2016). Evidence suggests that 45% of all women of reproductive age in Tanzania are anemic Ngimbudzi et al. (2021). The prevalence rate of anemia is 9% in developed countries while, in contrast, in developing countries, the prevalence rate reaches 43%, with children and women of reproductive age (WRA) at a greater risk of contracting anemia Alem (2023). When it comes to anemia (62.3%) and severe anemia (1.8%), Africa is ranked second Tirore et al. (2023).

In Ghana, the prevalence of anemia among pregnant women increased from 65% to 70% between 2003 and 2008 Region et al. (2014). Anemia in pregnancy prevalence in urban areas in Accra is 34% Dzabeng et al. (2019). The neonatal household health survey shows the prevalence of anemia being 27% among women aged 15–19 years and 40% among pregnant women Alem (2023). The Family Health Division (FHD) of the Ghana Health Service Annual Report revealed that the Volta Region had the highest prevalence (50%) of anemia among pregnant women in their fertility age (15–49) Tettegah et al. (2023).

2.3 Types of Anemia

- **Iron Deficiency Anemia**: Iron-deficiency anemia is a blood disorder that affects your red blood cells. It happens when your body doesn't have enough iron to make hemoglobin, a substance in your red blood cell that allows them to carry oxygen throughout your body Cleveland Clinic (2024).

- **Sickle Cell Anemia**: Sickle cell anemia is caused by a point mutation in the -globin chain of hemoglobin, causing the amino acid glutamic acid to be replaced with the hydrophobic amino acid valine at the sixth position Obeagu (2023).

- **Vitamin Deficiency Anemia**: Vitamin deficiency anemia is a lack of healthy red blood cells caused by lower than usual amounts of vitamin B-12 and folate. Vitamin deficiency anemia is characterized by a loss of healthy red blood cells as a result of low vitamin B-12 and folate levels. Vitamin B-12 deficiency can also cause nerves to malfunction, causing tingling, loss of sensation, and weakness MSD Manuals (2024).

- **Hemolytic Anemia**: Hemolytic anemia occurs when red blood cells are destroyed faster than they are made. Hemolytic anemia can be mild or severe, and it can progress slowly or rapidly. Some symptoms may include tiredness, dizziness, weakness or liver that is larger than normal National Heart, Lung, and Blood Institute (2024).

2.4 Risk Factors of Anemia

The prevalence of anemia varies across countries. For example, in the overall world population, nearly 14% of anemia occurs in high-income countries, and 50% of low-income countries are affected by anemia Tsegaye et al.

(2023). According to the WHO estimates, about half a billion WRA are anemic worldwide, of those around 20.2 million women are severely anemic with a higher burden in low-middle-income countries Savera (2020). Although anemia has various correlates based on its underlying pathophysiology, nutritional deficiencies and chronic diseases are generally the most common etiologies of anemia in children and older adults respectfully Safiri et al. (2021). WRA are one of the groups most at risk of anemia, due to their physiological processes Owais (2021).

The factors influencing anemia are often classified into two groups, namely, nutritional and non-nutritional based on their underlying causes Vroom et al. (2024). Previous studies have indicated several predictors of anemia among women of reproductive age, which includes the following socio-economic and reproductive health characteristics: young age, grand multi-parity, short inter-pregnancy intervals, low socio-economic status, low educational status, ignorance, heavy menstrual blood loss, history of using an intrauterine contraceptive device, previous history of anemia, and body mass index Tsegaye et al. (2023). Although prevalent globally, anemia is a particular concern in refugee settings where risk factors including restricted diets, increased transmission of infectious diseases, malaria, cramped living conditions, and inadequate access to water and sanitation may be present Hulland et al. (2021). Unequal household food allocation can make WRA vulnerable to anemia as they might not have access to iron-rich foods Kinyoki et al. (2021). Moreover, older age, limited knowledge of anemia, pregnancy during the second and third trimesters, multiparity, and experience of abortions were more likely reasons of anemia in pregnant women and lactating women Keokenchanh et al. (2021).

2.5 Causes of Anemia

Causes of anemia can be divided into three non-mutually exclusive pathways: blood loss, increased red blood cell destruction, and inadequate red blood cell production Kinyoki et al. (2021). Anemia can be caused by both nutritional and non-nutritional factors, with iron deficiency being the most common cause Wu et al. (2020). Iron deficiency is the primary cause of anemia among women and children in developing countries Souganidis et al. (2012). Intestinal parasitic infections, malaria, chronic illness, menstrual blood loss, gynecological and obstetric conditions, various nutritional deficiencies (particularly in folate and vitamins B12, A, and C), genetic disorders (such as sickle cell disease, thalassemia, and hereditary blood problems), loss of appetite, low dietary diversity score, poor dietary habits, and household food insecurity are additional significant causes of anemia Tirore et al. (2023). In addition, malaria and genetic hemoglobin disorders are main contributors to the anemia burden Christian et al. (2022).

2.6 Consequences of Anemia

The consequences of anemia are serious. Anemia also has both immediate and more long-term consequences Owais (2021). Its clinical manifestations and associated complications differ depending on the type of anemia and its severity Tirore et al. (2023). Anemia causes adverse pregnancy outcomes: still birth, abortion, low birth weight, maternal and neonatal mortality, intrauterine fetal death, premature birth, decreases physical performance, and work capacity Tsegaye et al. (2023). Anemia in pregnancy can therefore pose long-term consequences in the national economic development through low education attainment, reduced quality of life, decreased level of economic productivity, and therefore a cycle of poverty Ngimbudzi et al. (2021).

Anemia can also have economic consequences, potentially costing countries billions of dollars in reduced productivity Owais (2021). The consequences of iron deficiency anemia may range from immune system dysfunction, disturbances in the gastrointestinal tract, impaired thermoregulation and neurocognitive function Safiri et al. (2021). Anemia impairs the learning and development of future generations of children as well as the economic productivity and development of communities and countries Tirore et al. (2023).

Anemia in adults can cause fatigue, lethargy, reduced physical productivity, and poor work performance Chaparro et al. (2019). In adults, it reduces productivity and is associated with higher maternal mortality. In children, it impairs physical and cognitive development directly and affects human capital accumulation via impacts on behaviors like school attendance Coffey (2015).

2.7 Treatments for Anemia

Reducing anemia is considered as an essential part of improving the health of women, and the WHO has set a global target of achieving a 50% reduction of anemia among women of reproductive age by 2025 Kibret et al. (2025). Ghana, through the Ministry of Health, has been at the forefront with interventions and strategies to control anemia in pregnancy. These strategies include education and awareness creation, nutrient (iron) supplementation, and control and prevention of parasitic infections in pregnancy. Additionally, the use of insecticide-treated nets (ITNs) and intermittent preventive treatment (IPT) against malaria, effective deworming, and provision of improved water, sanitation, and hygiene services are also being implemented to prevent anemia among pregnant women Appiah (2020).

2.8 Methodological Considerations

Cross-sectional analysis of existing data Alem (2023) was employed by most studies to investigate anemia among women of reproductive age and other studies to investigate anemia among children, especially those under five. Studies have utilized data from Demographic and Health Survey (DHS) datasets Alem (2023), National Family Health Survey (NFHS-4) Ghosh (2021), Multiple

Indicator Cluster Survey (MICS) series Safiri et al. (2021) and other relevant datasets Owais (2021).

Women of reproductive age (15-49 years), which comprises both pregnant and non-pregnant women, have been the focus of some studies Sumaila et al. (2022). Most of them focus only on pregnant women Omari (2017).

Administration of questionnaires Sumaila et al. (2022), and blood sampling Ngimbudzi et al. (2021) were employed in most studies for data collection. Binary logistic regression Tettegah et al. (2023), bivariate regression Ngimbudzi et al. (2021), bivariate analysis Christian et al. (2022), univariate and multivariable logistic regressions Wu et al. (2020) were applied by some studies to identify risk factors for anemia.

CHAPTER 3

METHODOLOGY

3.1 Study Design

The study is a cross-sectional study aimed at assessing the prevalence of anemia and identifying high-risk groups among women of reproductive age (WRA) in Ghana using a cluster analysis approach. All eligible women aged 15-49 were tested during the survey to estimate the prevalence of anemia. This design was deemed appropriate since it allows to examine the association between different exposures (demographic, socioeconomic, or lifestyle factors) and the outcome of interest (anemia) at the same time. The cross-sectional study design is well-suited for providing a snapshot of the study population at a specific point in time. This allows researchers to determine the prevalence or current state of a disease, characteristic, or condition within the sample population. The observational nature of a cross-sectional study makes it a relatively efficient and quick approach to completing the research. Investigators can collect data on various variables, both independent and dependent,and then explore the associations between them.This type of study design does not require long-term follow-up of participants, which is a key advantage. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9536510/, accessed 03/07/2024)

3.2 Study Area

This research is centered on investigating anemia among women in Ghana.Ghana has a population of about 33.40 miliion with 50.1% of the total population being females and the rest being males. Our project majors on anemia in women and pregnant women since they cover or constitute most of the population and anemia is one disease worrying or affecting them. Socio-economic factors such as education, income level, employment status, health insurance cover, sanitation, among others constitutes to some women getting anemia in the country. Ghana is a country of West Africa. Ghana currently has sixteen regions; Western Region, Western North Region, Central Region, Greater Accra, Eastern Region, Volta Region, Ashanti Region, Ahafo Region, Bono Region, Bono East Region, Savannah Region, Upper West Region, Upper East Region Northern Region, Oti Region.In Ghana, the regions that record the highest cases of anemia are primarily located in the northern part of the country. The Northern Region, in particular, has a high prevalence of anemia among women and children. This region, along with the Upper East and Upper West regions, faces significant challenges related to anemia due to factors such as poverty, malnutrition, and limited access to healthcare services. The prevalence of anemia in these regions is influenced by a combination of nutritional deficiencies, particularly iron deficiency, and infectious diseases like malaria, which is common in these areas. Efforts to address anemia in these regions include nutritional interventions, improved healthcare access, and public health campaigns to raise awareness about anemia and its prevention Ardayfio et al. (2014); SPRING (2024).

3.3 Data Source

The data used is a nationally representative data from the 2022 Ghana Demographic and Health Survey (2022 GDHS) conducted by the Ghana Statistical Service (GSS) in collaboration with the Ministry of Health or Ghana Health Service (MoH/GHS) and other stakeholders, with funding from the United States Agency for International Development (USAID) and other partners.

3.4 Study Population

The study population for this cross-sectional study will include women of reproductive age (15-49 years) in Ghana who participated in the 2022 Ghana Demographic and Health Survey (GDHS). The study will cover both pregnant and non-pregnant women within this age range.

3.5 Sampling Procedure

The 2022 Ghana Demographic and Health Survey (GDHS) utilized a stratified two-stage cluster sampling approach to select a representative sample. In the first stage, 618 clusters were chosen from the sampling frame, which was based on the 2021 Population and Housing Census, using a probability proportional to size strategy for urban and rural areas in each of the 16 regions. In the second stage, a household listing and mapping exercise was conducted in the selected clusters to create a sampling frame of households. Trained listers and mappers, organized into 25 teams, spent 2 months completing the listing operation, which included collecting the geographical coordinates of each household using GPS devices. From the household listing, a fixed number of 30 households in each cluster were randomly selected for interviews. This resulted in a total sample of 18,450 households, from which 15,014 women aged 15-49 and 7,044 men aged 15-59 (from every other household) were interviewed.

Stratified Sampling

This involves dividing the population into distinct subgroups, or strata, based on certain characteristics such as geographic region, urban or rural status, among others. Each stratum is then sampled separately to ensure representation from all subgroups.

Two-Stage Cluster Sampling

1. First Stage: Selection of clusters. These could be geographical areas or other logical groupings.
2. Second Stage: Within each selected cluster, a further selection of individual units (households) are made.

1. Probability Proportional to Size (PPS) Sampling

In the first stage, clusters (enumeration areas) are selected with probability proportional to their size. The size could be the number of households or the population within the cluster.

Probability of selecting a cluster i

$$P_i = \frac{M_i}{\sum_{j=1}^{N} M_j}$$

where:

- P_i is the probability of selecting the i-th cluster.
- M_i represents the size (e.g., population) of the i-th cluster.
- N is the total number of clusters.

2. Systematic Sampling

In the second stage, households within the selected clusters are chosen using systematic sampling.

Sampling Interval k

$$k = \frac{N_h}{n_h}$$

where:

- N_h is the total number of households listed in the selected cluster h
- n_h is the number of households to be sampled from cluster h.

Selection of households

Start at a random point r within the interval $[1, k]$, then select every kth household:

$$r + (i-1)k \quad \text{for} \quad i = 1, 2, \ldots, n_h$$

Selection Probability Within a Cluster

$$P_{ij} = \frac{1}{n_i}$$

where:

- P_{ij} is the probability of selecting the j-th household within the i-th cluster.
- n_i is the number of households picked from the i-th cluster.

3. Weight Calculation

Each household and individual in the survey is assigned a sampling weight to adjust for the unequal probabilities of selection and to ensure that the sample represents the population.

Weight for cluster i

$$W_i = \frac{1}{P_i}$$

Overall weight for household j in cluster i

$$W_{ij} = \frac{1}{P_i \times P_{ij}}$$

where P_{ij} is the probability of selecting household j within cluster i.

3.6 Data analysis

3.6.1 Cluster Analysis

Introduction

Cluster analysis is a data exploration (mining) tool for dividing a multivariate dataset into "natural" clusters (groups). We use the methods to explore whether previously undefined clusters (groups) exist in the dataset. For instance, a marketing department may wish to use survey results to sort its customers into categories (perhaps those likely to be most receptive to buying a product, those most likely to be against buying a product, and so forth).
Cluster Analysis is used when we believe that the sample units come from an unknown number of distinct populations or sub-populations. We also assume that the sample units come from a number of distinct populations, but there is no apriori definition of those populations. The purpose of cluster analysis is to help reveal patterns and structures within a dataset that may provide insights into underlying relationships and associations.

3.6.2 Types of Clustering

Hierachical Clustering

Hierarchical clustering is a well known method for grouping objects or datasets. It creates groups so that objects within a group are similar to each other and distinct from objects in different groups. Clusters are visually represented a dendrogram (hierarchical tree).

Agglomerative Hierarchical Clustering Algorithm

The agglomerative hierarchical clustering algorithm is a known example of HCA (Hierarchical Clustering Algorithm). To group the datasets into clusters, it follows the bottom-up approach. Tthis algorithm considers each dataset as a single cluster at the first stage, and then start merging the closest pair of clusters.

Types of Agglomerative Hierarchical Clustering Algorithm

- **Agglomerative**: Firstly, every object is taken into consideration to be its own cluster. Based to a particular procedure, the clusters are then merged step by step until a single cluster is left.
- **Divisive**: The Divisive method is the opposite of the Agglomerative method. In the agglomerative hierarchical approach, we define each data point as a cluster and merge existing clusters during every step. There are four different methods for this approach: Single Linkage, Complete Linkage, Average Linkage, Centroid method, Ward method.

1. Single Linkage Hierarchical Clustering

Single linkage hierarchical clustering, we can also call this the single linkage clustering, this merges clusters using the closest pair of points between them. In Single Linkage we define the distance between two clusters as the minimum distance between any single data point in the first cluster and any single data point in the second cluster. Based on this explanation of distance between clusters, at every stage of the process, we bring together the two clusters with the smallest single linkage distance. This is the distance between the most closest members of the two clusters.

The distance $D(A, B)$ between two clusters A and B using single linkage is defined as:

$$D(A, B) = \min\{d(x, y) : x \in A, y \in B\}$$

where:

- $D(A, B)$ is the distance between cluster A and cluster B.
- $d(x, y)$ is the distance between point x in cluster A and point y in cluster B.
- $x \in A$ means that point x is a member of cluster A.
- $y \in B$ means that point y is a member of cluster B.
- $\min$ denotes the minimum value.

2. Complete Linkage Hierarchical Clustering

The distance between two clusters is estimated by a single pair of elements: one in each cluster that are closest to each other. This method is based on maximum distance; the similarity of two clusters is the similarity of their most dissimilar pair. In Complete Linkage we explain the distance between two clusters as the maximum distance between any single data point in the initial cluster and any single data point in the second cluster. From this definition of distance between clusters, during every stage of the procedure, we combine two clusters that have the smallest complete linkage distance. This is the distance between the members that are dissimilar.

The distance $D(A, B)$ between two clusters A and B using complete linkage is defined as:

$$D(A, B) = \max\{d(x, y) : x \in A, y \in B\}$$

where:

- $D(A, B)$ is the distance between cluster A and cluster B.
- $d(x, y)$ is the distance between point x in cluster A and point y in cluster B.
- $x \in A$ means that point x is a member of cluster A.
- $y \in B$ means that point y is a member of cluster B.
- $\max$ denotes the maximum value.

3. Average Linkage Hierarchical Clustering

In Average linkage, we explain the distance between two clusters as the average distance between data points in the first cluster and also data points in the second cluster. From this understanding of distance between clusters, at every stage of the process we merge the two clusters that have the smallest average linkage distance.
The distance $D(A, B)$ between two clusters A and B using average linkage is defined as:

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} d(x, y)$$

where:

- $D(A, B)$ is the distance between cluster A and cluster B.
- $d(x, y)$ is the distance between point x in cluster A and point y in cluster B.
- $x \in A$ means that point x is a member of cluster A.
- $y \in B$ means that point y is a member of cluster B.
- $|A|$ is the number of points in cluster A.
- $|B|$ is the number of points in cluster B.

4. Centroid Method

In the centroid method, the distance between two clusters is the distance between the two mean vectors of the clusters. At each stage of the process we combine the two clusters that have the smallest centroid distance.

The distance $D(A, B)$ between two clusters A and B using the centroid method is defined as:

$$D(A, B) = \|\mu_A - \mu_B\|$$

where:

- $D(A, B)$ is the distance between cluster A and cluster B.
- μ_A is the centroid (mean) of cluster A.
- μ_B is the centroid (mean) of cluster B.
- $\|\mu_A - \mu_B\|$ is the Euclidean distance between the centroids of clusters A and B.

The centroid μ of a cluster C with n points $\{x_1, x_2, \ldots, x_n\}$ is calculated as:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

5. Ward Method

The Ward method does not explain a measure of distance between two or clusters. This is an ANOVA-based approach. One-way univariate ANOVA is done for every variable with groups defined by the clusters at the stage of the process. During every stage, two clusters merge to provide the smallest increase in the combined error sum of squares.

Ward's method minimizes the increase in total within-cluster variance. The increase in variance ΔE when merging two clusters A and B is defined as:

$$\Delta E = \frac{|A| \cdot |B|}{|A| + |B|} \|\mu_A - \mu_B\|^2$$

where:

- ΔE is the increase in total within-cluster variance.
- $|A|$ is the number of points in cluster A.
- $|B|$ is the number of points in cluster B.
- μ_A is the centroid (mean) of cluster A.
- μ_B is the centroid (mean) of cluster B.
- $\|\mu_A - \mu_B\|$ is the Euclidean distance between the centroids of clusters A and B.

The centroid μ of a cluster C with n points $\{x_1, x_2, \ldots, x_n\}$ is calculated as:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

K-means Clustering

K-means is a centroid-based clustering algorithm, where we estimate the distance between each data point and a centroid to give it to a cluster. The main objective is to find the K number of groups in the dataset. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects of another cluster. The term 'K' is a number used to denote the number of clusters. The algorithm beginsby selecting K centroids randomly and updates them by finding the mean of the data points closest to each centroid. To perform k-means clustering, the algorithm randomly gives k initial canters (k given by the user), either by randomly selecting points in the "Euclidean space" defined by all n variables, or sampling k points of all available observations to serve as initial canters. K-means is a centroid-based clustering algorithm, where we calculate the distance between each data point and a centroid to assign it to a cluster.

The objective of the k-means clustering algorithm is to minimize the within-cluster sum of squares (WCSS). The objective function can be written as:

$$\min_{C} \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2$$

where:

- k is the number of clusters.
- C_i is the set of points (or cluster) i.
- x represents each point in cluster C_i.
- μ_i is the mean of the points in cluster C_i.
- $\|x - \mu_i\|^2$ is the squared Euclidean distance between point x and the cluster mean μ_i.

The algorithm follows these steps:

1. Initialize k cluster centroids randomly.
2. Assign each point to the nearest cluster centroid.
3. Update each cluster centroid to the mean of the points assigned to it.
4. Repeat steps 2 and 3 until convergence (the assignments no longer change).

3.6.3 Distance Measures in Cluster Analysis

1. Euclidean Distance

The Euclidean distance between two points $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ in n-dimensional space is given by:

$$d_E(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

2. Manhattan Distance (L1 Distance)

The Manhattan distance (also known as L1 distance or taxicab distance) is the sum of the absolute differences of their coordinates:

$$d_M(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i|$$

3. Minkowski Distance

The Minkowski distance generalizes both the Euclidean and Manhattan distances. For a parameter p:

$$d_{M_p}(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}$$

- For $p = 1$, it becomes the Manhattan distance. - For $p = 2$, it becomes the Euclidean distance.

3.6.4 Cluster Validation Methods

Introduction

Cluster validation is the process of evaluating the results of a clustering algorithm to determine how well the clusters represent the underlying data structure. There are several techniques for cluster validation, which can be broadly categorized into internal, external, and relative validation methods.

Internal Validation

Internal validation measures evaluate the goodness of clustering without reference to external information. These measures typically assess the compactness and separation of clusters.

1. Silhouette Score

The silhouette score combines measures of how similar an object is to its own cluster (cohesion) and how dissimilar it is to other clusters (separation). The silhouette score ranges from -1 to 1, where a higher value indicates better clustering.

For a data point i:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

where:

- $a(i)$ is the average distance from the i-th point to the other points in the same cluster.
- $b(i)$ is the minimum average distance from the i-th point to points in a different cluster, minimized over clusters.

The silhouette score for the entire clustering is the average silhouette score for all data points.

2. Dunn Index

The Dunn Index identifies clusters that are compact and well-separated.

$$D = \frac{\min_{1 \leq i < j \leq k} \delta(C_i, C_j)}{\max_{1 \leq i \leq k} \Delta(C_i)}$$

where:

- $\delta(C_i, C_j)$ is the distance between clusters C_i and C_j.
- $\Delta(C_i)$ is the diameter of cluster C_i.

3. Davies-Bouldin Index

The Davies-Bouldin Index (DBI) is an average similarity ratio of each cluster with the cluster most similar to it. A lower Davies-Bouldin index indicates better clustering.

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{j \neq i} \left(\frac{s(C_i) + s(C_j)}{\delta(C_i, C_j)} \right)$$

where:

- $s(C_i)$ is the average distance between each point in cluster C_i and the centroid of C_i.
- $\delta(C_i, C_j)$ is the distance between centroids of clusters C_i and C_j.

External Validation

External validation measures use external information (ground truth labels) to evaluate the clustering.

1. Rand Index

The Rand Index measures the similarity between the clusters and the ground truth classification.

$$RI = \frac{TP + TN}{TP + FP + FN + TN}$$

where:

- TP is the number of true positives.

- TN is the number of true negatives.
- FP is the number of false positives.
- FN is the number of false negatives.

2. Adjusted Rand Index

The Adjusted Rand Index (ARI) adjusts the Rand Index for the chance grouping of elements.

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}$$

where $E[RI]$ is the expected Rand Index of random clusterings.

3. Mutual Information

Mutual Information (MI) measures the agreement of the two assignments ignoring permutations.

$$MI(U, V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} P(i, j) \log \frac{P(i, j)}{P(i)P(j)}$$

where:

- $P(i, j)$ is the probability that a point belongs to cluster i in U and cluster j in V.
- $P(i)$ and $P(j)$ are the marginal probabilities.

4. Gap Statistic

The Gap Statistic compares the total within intra-cluster variation for different numbers of clusters with their expected values under null reference distribution.

$$\text{Gap}(n) = \frac{1}{B} \sum_{b=1}^{B} \log(W_b^*(n)) - \log(W(n))$$

where:

- $W(n)$ is the within-cluster dispersion for the clustering with n clusters.
- $W_b^*(n)$ is the within-cluster dispersion for the b-th simulation under null reference distribution.

These are just a few examples of cluster validation techniques.The choice of the appropriate technique ultimately depends on the specific problem at hand, the clustering algorithm used, and the desired properties of the resulting clusters. The suitability of a particular validation method can vary depending on the context and the goals of the clustering analysis. Different techniques may be better suited for evaluating different aspects of the clusters, such as their compactness, separation, or interpretability. Therefore, it is important to carefully consider the problem requirements and the characteristics of the data when selecting the most suitable cluster validation approach.

3.7 Ethical Consideration

Our study uses secondary data from the 2022 GDHS, which has already obtained ethical approval from the relevant Institutional Review Board (IRB) (Approval Number: 123456). We ensured participant privacy and confidentiality by anonymizing the data using unique identifiers and encrypting the dataset. No additional ethical approval was required for this secondary data analysis. We adhered to the principles of the Declaration of Helsinki and the GDPR guidelines for data protection.

3.8 Limitations of the study

Our study has several limitations.

1. First, the dataset may contain missing values, which could impact the accuracy of our results.

2. Second, our sample may not be representative of the population, as the GDHS only includes individuals aged 18-65.
3. Third, we selected variables based on availability and relevance, which may have introduced biases.
4. Fourth, we assumed cluster homogeneity, which may not be accurate, and within-cluster variability may exist.
5. Finally, our findings may not be generalizable to other settings or populations.

CHAPTER 4

ANALYSIS

ANALYSIS AND RESULTS

4.1 Introduction

In order to identify specific high-risk categories for anemia among Ghanaian women of reproductive age, we use cluster analysis approaches in this chapter. By grouping people according to shared socioeconomic, demographic, and health traits, this technique helps us understand the variables that raise the risk of anemia. The information used in this research comes from a thorough survey that included questions about a variety of topics, such as age, place of residence, marital status, level of education, and household wealth. The body of research has shown that these factors are important predictors of anemia risk. Descriptive statistics were employed using SPSS to provide an understanding of the overall patterns and distributions seen in our dataset. Following this, we utilized R to conduct k-means cluster analysis, a robust statistical method for grouping data points into distinct clusters based on their characteristics. Combining the strengths of both SPSS and R ensures a thorough and reliable analysis.

Table 4.1: Socio-Demographic Variables

Socio-demographic variables	Frequency	Percentage (%)
AGE IN 5-YEAR GROUP		
15-19	87	4.8%
20-24	338	18.8%
25-29	403	22.4%
30-34	410	22.8%
35-39	334	18.6%
40-44	172	9.6%
45-49	55	3.1%
MARITAL STATUS		
Divorced	19	1.1%
Living with partner	307	17.1%
Married	1213	67.4%
Never in union	181	10.1%
No longer living together/separated	64	3.6%
Widowed	15	0.8%
RESIDENCE		
Rural	1049	58.3%
Urban	750	41.7%

Continued on next page

Table 4.1: Socio-Demographic Variables (continued)

Socio-demographic variables	Frequency	Percentage (%)
EDUCATION LEVEL		
Higher	123	6.8%
No education	539	30.0%
Primary	288	16.0%
Secondary	849	47.2%
WEALTH INDEX		
Middle	342	19.0%
Poorer	416	23.1%
Poorest	577	32.1%
Richer	259	14.4%
Richest	205	11.4%
ETHNICITY		
Akan	517	28.7%
Ewe	173	9.6%
Ga/Dangme	61	3.4%
Grusi	98	5.4%
Guan	83	4.6%
Gurma	222	12.3%

Continued on next page

Table 4.1: Socio-Demographic Variables (continued)

Socio-demographic variables	Frequency	Percentage (%)
Mande	62	3.4%
Mole-Dagbani	565	31.4%
Other	18	1.0%

Table 4.2: Health Variables

Variable	Category	Frequency	Percentage
Health Insurance Coverage	no	108	6.00
	yes	1,691	94.00
Currently Pregnant	no or unsure	1,634	90.83
	yes	165	9.17

Table 4.3: Hemoglobin Level

Descriptive Statistics	N	Minimum	Maximum	Mean	Std. Deviation
Hemoglobin Level	1799	47	165	120.57	15.268

Table 4.4: Number of Children

Descriptive Statistics	N	Range	Minimum	Maximum	Mean	Std. Deviation
Number of Children	1799	12	1	13	3.23	2.039

4.2 Descriptive analysis of socio-demographic characteristics

From Table 4.1, the study population is divided into seven age groups. The largest group is 30-34 years old, comprising 22.8% (410 individuals).This is followed closely by the 25-29 age group at 22.4% (403 individuals). Both the 20-24 and 35-39 age groups each make up 18.6% of the population (338 and 334 individuals, respectively). The 40-44 group has 9.6% (172 individuals), while the 15-19 group constitutes 4.8% (87 individuals). The smallest group is 45-49 years, with 3.1% (55 individuals). Marital status reveals the social structure of the population. A significant majority are married, making up 67.4% (1,213 individuals). Those living with a partner account for 17.1% (307 individuals), and 10.1% (181 individuals) have never been in a union. The divorced and widowed categories are smaller, with 1.1% (19 individuals) and 0.8% (15 individuals), respectively. A small segment, 3.6% (64 individuals), are separated or no longer living with their partners. The population shows a notable rural-urban divide. Most individuals, 58.3% (1,049), live in rural areas, while 41.7% (750) reside in urban areas. This distribution highlights the demographic and settlement patterns within the study area. The largest group has secondary education, comprising 47.2% (849 individuals). Those with no education represent 30.0% (539 individuals), indicating a considerable portion with limited educational attainment. Primary education is held by 16.0% (288 individuals), and 6.8% (123 individuals) have achieved higher education. The wealth index divides the population into five economic categories. The largest group is the 'poorest,' with 32.0% (577 individuals). This is followed by the 'poorer' category at 23.1% (416 individuals) and the 'middle' group at 19.0% (342 individuals). The 'richer' and 'richest' groups are smaller, comprising 14.4% (259 individuals) and 11.4% (205 individuals), respectively. The study population is ethnically diverse. The Mole-Dagbani group is the largest ethnic group, representing 31.4% (565 individuals). The Akan group follows with 28.7% (517 individuals). Other notable groups include the Gurma (12.3%, 222 individuals), Ewe (9.6%, 173 individuals), and Grusi (5.4%, 98 individuals). Smaller groups include the Guan (4.6%, 83 individuals), Mande (3.7%, 66 individuals), Ga/Dangme (3.4%, 61 individuals), and other ethnicities (1.0%, 18 individuals).

4.3 Descriptive analysis of Health related characteristics

Fever occurrence in the past two weeks is a key health indicator. Most participants, 81.4% (1,465 individuals), reported no fever during this period. Conversely, 18.0% (324 individuals) experienced fever, while a small proportion, 0.6% (10 individuals), were uncertain about their fever status. A significant majority, 94.0% (1,691 individuals), have health insurance coverage, whereas 6.0% (108 individuals) lack insurance. The most common number of children is one, with 24.8% (446 individuals). A large majority, 90.8% (1,634 individuals), are not pregnant or are unsure of their pregnancy status, while 9.2% (165 individuals) reported being currently pregnant.See Table 4.2

4.4 Cluster Analysis

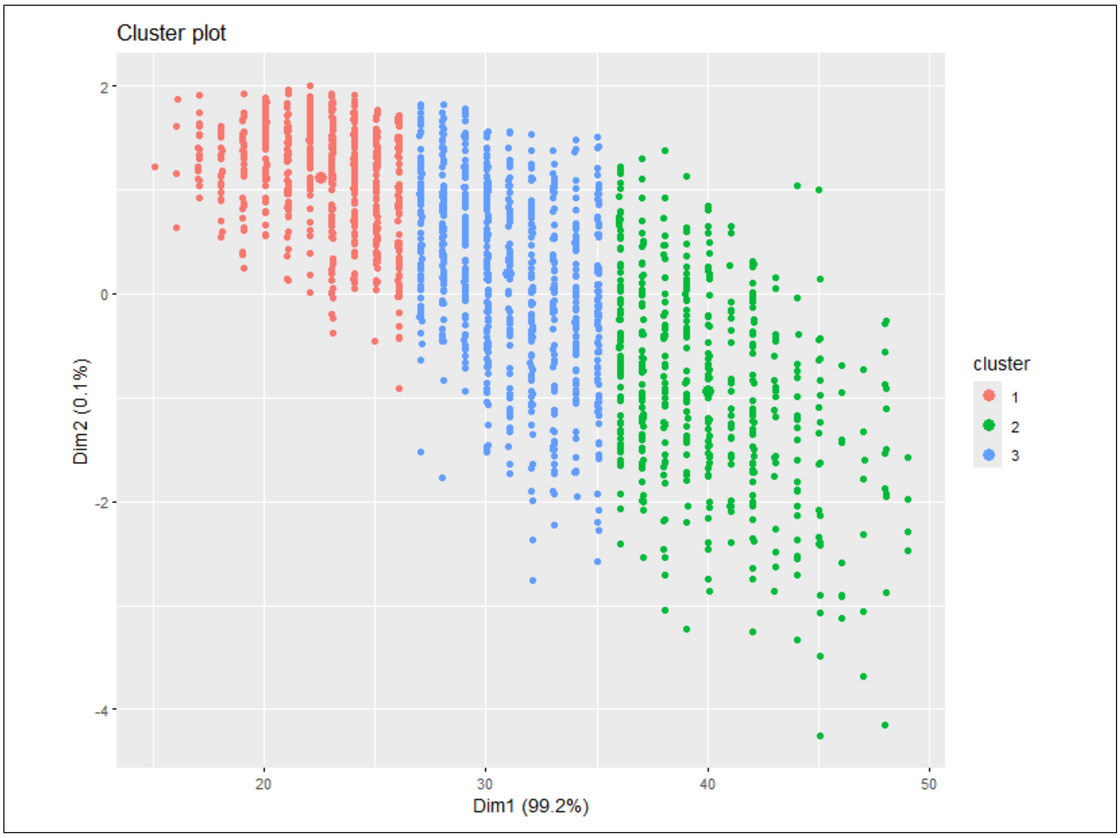

Figure 4.1: Visualization of Data Point Grouping across Different Clusters

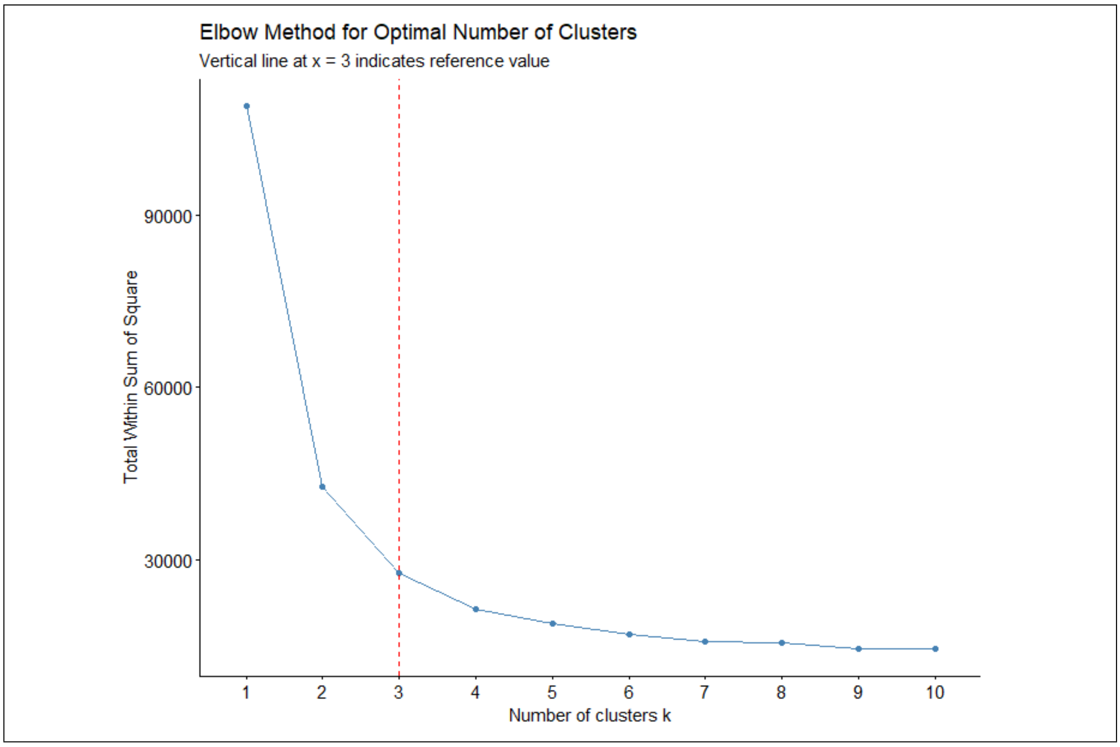

Figure 4.2: Scree plot of clusters

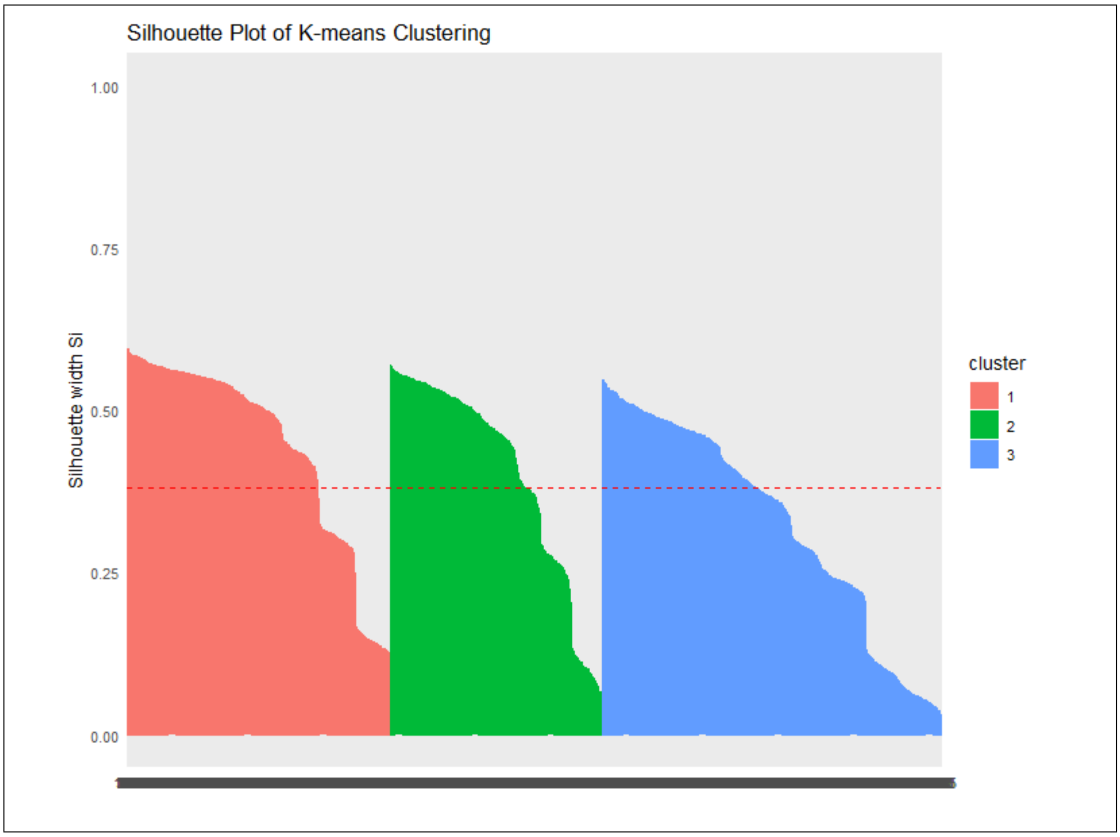

Figure 4.3: Silhouette Plot of K-means Clustering

The cluster plot in Figure 4.1 above illustrates the spatial distribution of the clusters and their relationships to each other. It helps identify any separation or overlap between clusters and offers insights into the cohesion within each cluster. By examining the plot, we can identify three distinct and well-defined aspects of the clusters. We see that cluster 1 is represented with red dots, cluster 2 is represented with green dots and cluster 3 is represented with blue dots.

Figure 4.2 presents the scree plot from the k-means cluster analysis of the 1799 observations.The elbow at three clusters indicates the most efficient balance between minimizing the number of clusters and reducing the variance within each cluster. This scree plot was used to confirm the choice of the three clusters employed in the analysis. The average silhouette width varies between -1 and 1, with higher values signifying more well-defined clusters. As shown in Figure 4.3 cluster 1 has the highest average silhouette score of 0.44, indicating that its points are well-clustered. Cluster 2 has a slightly lower average silhouette score of 0.40.Cluster 3, with an average silhouette width of 0.32, is the least well-defined among the three clusters.The silhouette analysis supports the use of a three-cluster solution, indicating that the clusters are reasonably well-defined and the data points within each cluster are cohesive

4.5 Socio-demographic characteristics within the clusters

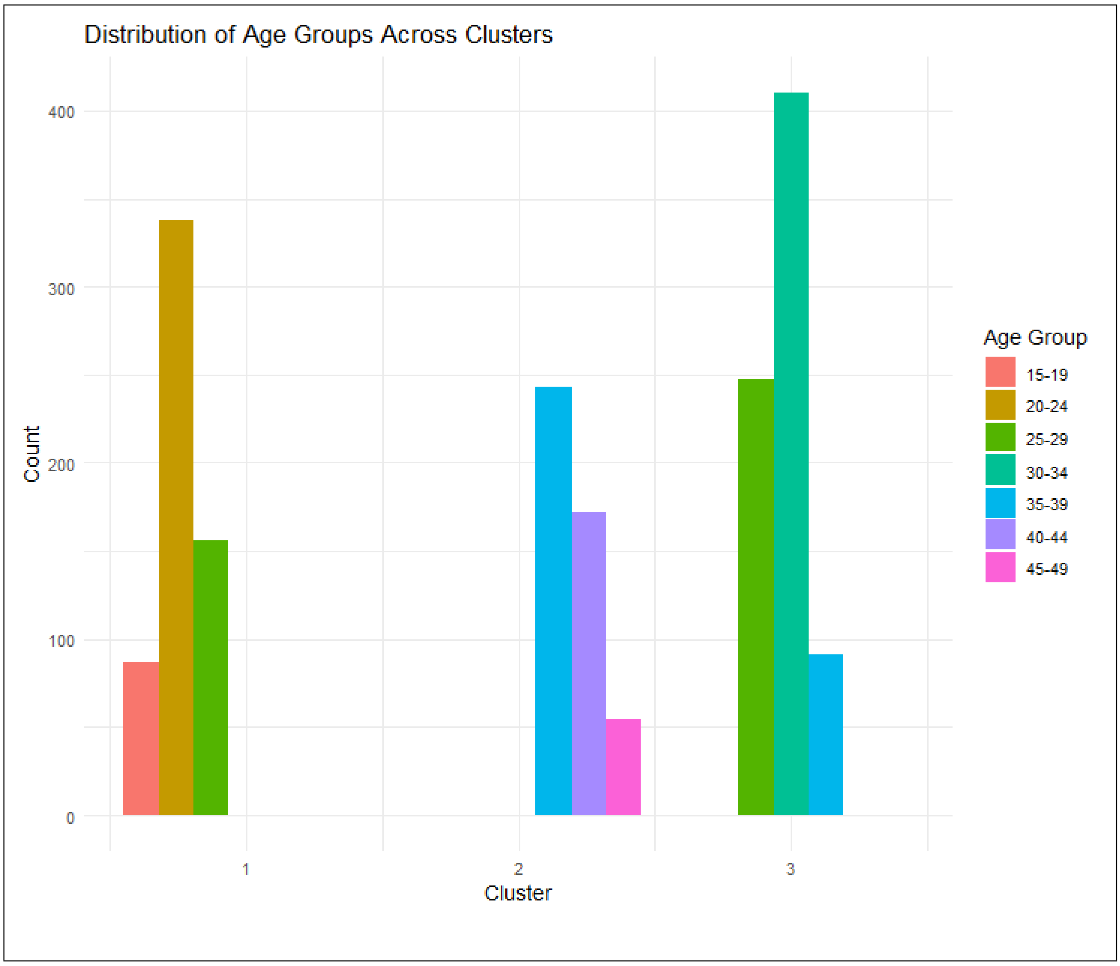

Figure 4.4: Distribution of Age Groups Across Clusters

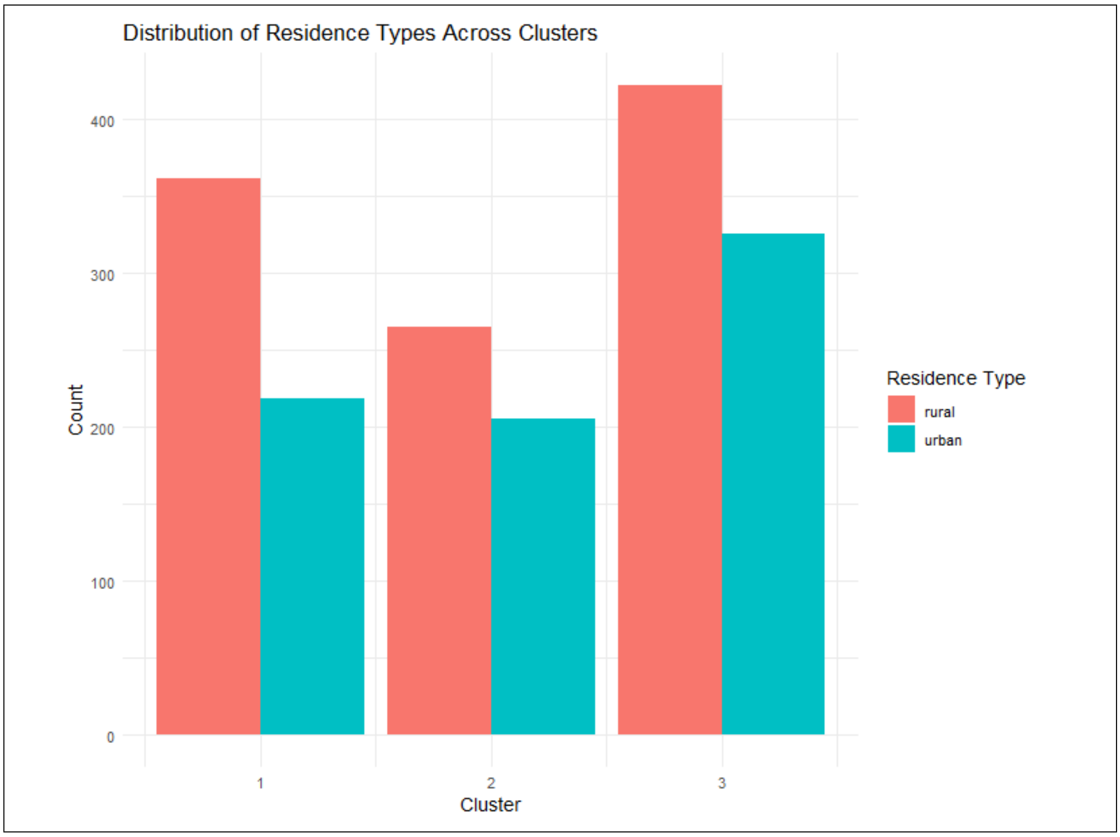

Figure 4.5: Distribution of Residence Across Clusters

As shown in Figure 4.4, the results indicate that cluster 1 consists of the age groups 15-19, 20-24 and 25-29 but dominated by the age group 20-24. This means that cluster 1 is made up of younger individuals. Cluster 2 also consists of the age groups 35-39, 40-45 and 45-49. It is dominated by individuals in the 35-39 age group which means that cluster 4 is made up of older individuals. Cluster 3 contains individuals in the 25-29, 30-34 and 35-39. It is dominated by individuals in the 30-34 age group which means that cluster 3 is made up of the middle class age group.

From Figure 4.5, the distribution of residential types—rural and urban—across three distinct clusters is displayed in the bar Chart. The number of people in each cluster and kind of housing is indicated by the height of each bar. Cluster 1 contains a greater percentage of people living in rural areas than in urban areas.Cluster 2 indicates a more equitable distribution of people in rural and urban areas. Cluster 3 shows a greater percentage of people living in cities as opposed to rural areas. The residence type patterns in the clusters show clear trends. Cluster 3 is primarily urban, whereas Cluster 1 is primarily rural. The population in Cluster 2 is more diverse, with a roughly equal proportion of residents living in both rural and urban areas.

Table 4.5: Crosstab of Cluster and OCCUPATION with Frequencies and Percentages

OCCUPATION	Cluster 1	Cluster 2	Cluster 3
agricultural, fishery and related labourers	27 (4.6%)	34 (7.2%)	31 (4.1%)
corporate managers	0 (0%)	0 (0%)	1 (0.1%)
customer services clerks	7 (1.2%)	2 (0.4%)	4 (0.5%)
drivers and mobile machine operators	0 (0%)	0 (0%)	1 (0.1%)
extraction and building trades workers	0 (0%)	0 (0%)	3 (0.4%)
general managers	0 (0%)	1 (0.2%)	2 (0.3%)
industrial plant operators	1 (0.2%)	0 (0%)	0 (0%)
labourers in mining, construction, manufacturing and transport	2 (0.3%)	1 (0.2%)	2 (0.3%)
life science and health professionals	2 (0.3%)	7 (1.5%)	22 (2.9%)

Crosstab of Cluster and Occupation with Frequencies and Percentages (continued)

OCCUPATION	Cluster 1	Cluster 2	Cluster 3
market-oriented skilled agricultural and fishery workers	103 (17.7%)	158 (33.6%)	157 (21%)
metal and machinery trades workers	0 (0%)	0 (0%)	1 (0.1%)
no classification	3 (0.5%)	0 (0%)	3 (0.4%)
not working and didn't work in last 12 months	182 (31.3%)	42 (8.9%)	91 (12.2%)
office clerks	2 (0.3%)	3 (0.6%)	2 (0.3%)
other associate professionals	0 (0%)	3 (0.6%)	3 (0.4%)
other craft and related workers	82 (14.1%)	29 (6.2%)	87 (11.6%)
other professionals	0 (0%)	2 (0.4%)	2 (0.3%)
personal and protective services workers	37 (6.4%)	15 (3.2%)	48 (6.4%)
physical science and engineering associate professionals	0 (0%)	0 (0%)	1 (0.1%)

Crosstab of Cluster and Occupation with Frequencies and Percentages (continued)

OCCUPATION	Cluster 1	Cluster 2	Cluster 3
precision, handicraft, printing and related trades workers	5 (0.9%)	1 (0.2%)	2 (0.3%)
sales and services elementary occupations	34 (5.9%)	45 (9.6%)	71 (9.5%)
salespersons, demonstrators and models	82 (14.1%)	116 (24.7%)	185 (24.7%)
stationary machine operators and assemblers	0 (0%)	1 (0.2%)	0 (0%)
subsistence agricultural, fishery and related workers	3 (0.5%)	4 (0.9%)	2 (0.3%)
teaching professionals	9 (1.5%)	6 (1.3%)	27 (3.6%)

Table 4.2 indicates the crosstab of the three clusters and the various occupations in our data with their corresponding frequencies and percentages.Taking into consideration the various occupations, it is observed that most individuals in cluster 1 are not working and didn't work in the last 12 months which is represented by 31.3% (182 individuals).Cluster 2 contains majority of individuals who are market-oriented skilled agricultural and fishery workers representing 33.6% (158 individuals).Cluster 3 is also dominated by individuals

who are salespersons, demonstrators and models representing 24.7% (185 individuals).

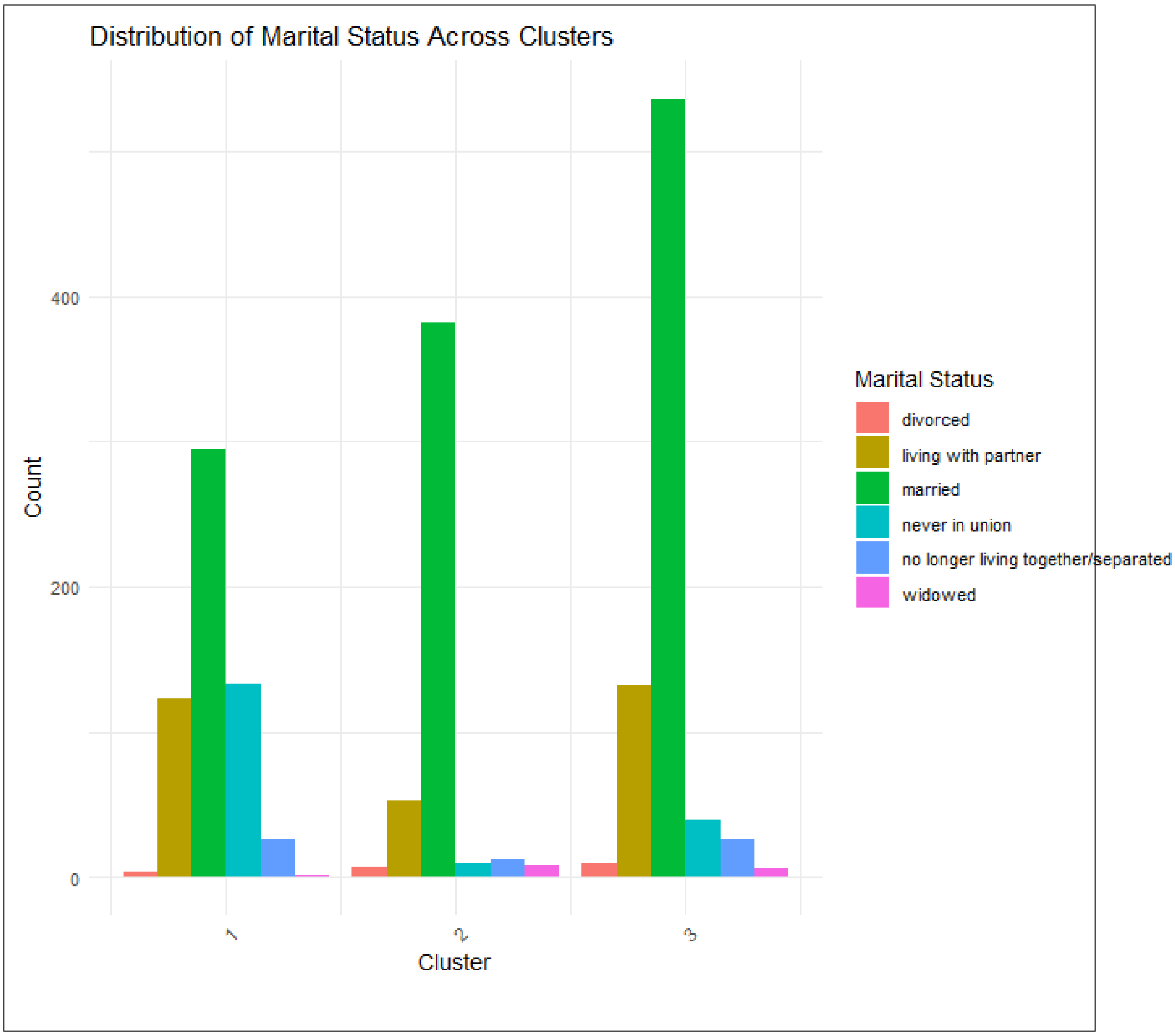

Figure 4.6: Distribution of Marital Status Across Clusters

Marital Status The bar chart shown in Figure 4.6 illustrates how different marital statuses are spread across the three clusters. Each bar represents a marital status, and its height shows how many people have that status in each cluster.In cluster 1, there are diverse marital statuses, but being married is the most common. Living with a partner is also frequent, followed by never in union.Cluster 2 is similar to Cluster 1, being married is the most common status. Living with a partner comes next, with a notable number of people who are never in union.Cluster 3 has the highest count of married individuals.Living with a partner is also common, but other statuses like divorced, never in union, and no longer living together or separated are present but less frequent. Overall, being married is the most common status across all clusters, especially in Cluster 3. Living with a partner is also relatively common in all clusters. Other marital statuses like divorced, never in union, no longer living together or separated, and widowed are less frequent.

Table 4.6: Crosstab of Cluster and Ethnicity with Frequencies and Percentages

ETHNICITY	Cluster 1	Cluster 2	Cluster 3
akan	163 (28.1%)	132 (28.1%)	222 (29.7%)
ewe	46 (7.9%)	43 (9.1%)	84 (11.2%)
ga/dangme	21 (3.6%)	18 (3.8%)	22 (2.9%)
grusi	36 (6.2%)	24 (5.1%)	38 (5.1%)
guan	27 (4.6%)	24 (5.1%)	32 (4.3%)
gurma	71 (12.2%)	56 (11.9%)	95 (12.7%)
mande	18 (3.1%)	18 (3.8%)	26 (3.5%)
mole-dagbani	193 (33.2%)	148 (31.5%)	224 (29.9%)
other	6 (1%)	7 (1.5%)	5 (0.7%)

Table 4.3 shows the crosstab of the three clusters and the various ethnic groups with their respective frequencies and percentages. By observation, we see that all three clusters are dominated by the ethnic groups “Mole-Dagbani” and

“Akan” and the rest follow suit. Cluster 3 contains a larger count for both “Mole-Dagbani” and “Akan” ethnic groups representing 29.9% (224 individuals) and 29.7% (222 individuals) respectively, followed by cluster 1 and the least in cluster 2.

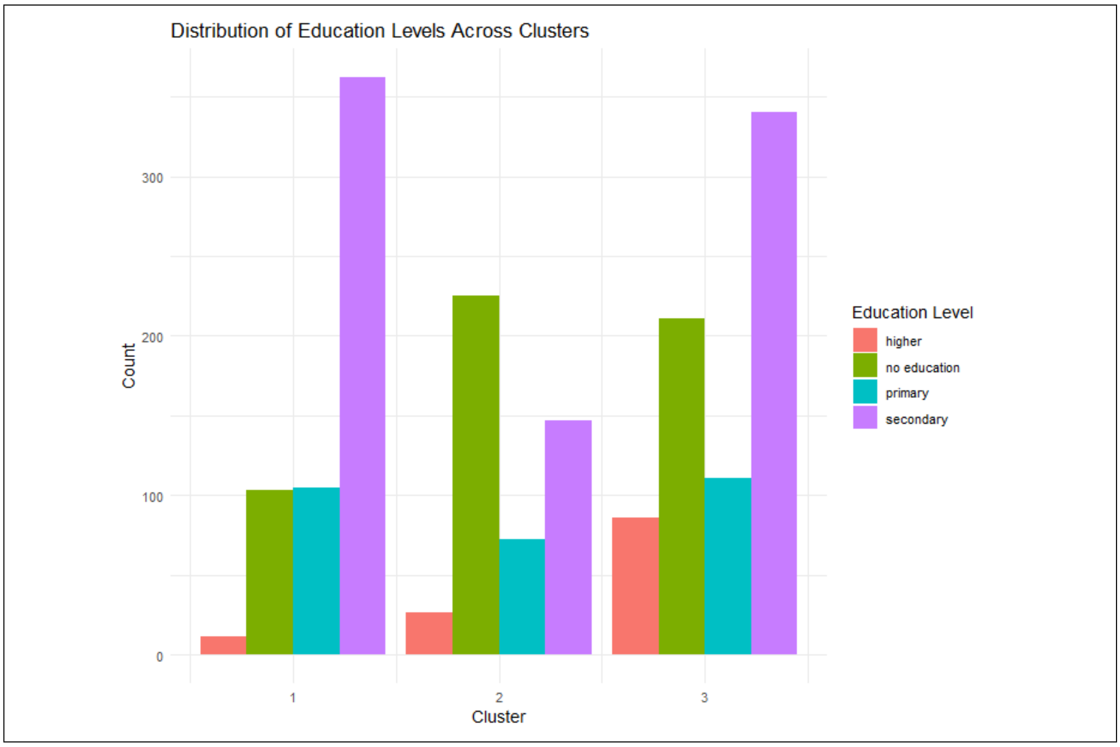

Figure 4.7: Distribution of Education Level Across Clusters

The figure shows the education levels in different groups. Cluster 1 has the most people with secondary education. It also has many people with primary and higher education. This group has different education levels. Cluster 2 has the most people with higher education. This group seems the most educated. It also has many people with no education. Cluster 3 has the most people with secondary education. But it also has many people with primary and higher education. This group has mixed education levels.Overall, the different groups have different education levels. Some groups are more educated than others. The data shows the education profiles of the groups. This can help understand the people in these groups.

Table 4.7: Crosstab of Cluster and Region with Frequencies and Percentages

REGION	Cluster 1	Cluster 2	Cluster 3
ahafo	37 (6.4%)	28 (6%)	49 (6.6%)
ashanti	31 (5.3%)	39 (8.3%)	45 (6%)
bono	25 (4.3%)	16 (3.4%)	37 (4.9%)
bono east	37 (6.4%)	37 (7.9%)	51 (6.8%)
central	36 (6.2%)	18 (3.8%)	43 (5.7%)
eastern	30 (5.2%)	20 (4.3%)	41 (5.5%)
greater accra	21 (3.6%)	25 (5.3%)	39 (5.2%)
north east	62 (10.7%)	34 (7.2%)	59 (7.9%)
northern	49 (8.4%)	50 (10.6%)	81 (10.8%)
oti	33 (5.7%)	34 (7.2%)	45 (6%)
savannah	49 (8.4%)	42 (8.9%)	51 (6.8%)
upper east	55 (9.5%)	24 (5.1%)	55 (7.4%)
upper west	41 (7.1%)	38 (8.1%)	48 (6.4%)
volta	19 (3.3%)	23 (4.9%)	39 (5.2%)
western	28 (4.8%)	14 (3%)	37 (4.9%)
western north	28 (4.8%)	28 (6%)	28 (3.7%)

Table 4.4 shows the crosstab of the three clusters and the various regions with their respective frequencies and percentages. By observation, it is seen that “northern” region has the highest count which is 180 individuals, relative

to other regions. Cluster 3 recorded the highest count representing 10.8% (81 individuals) for “northern” region, followed by cluster 2 representing 10.6% (50 individuals) and the least to be cluster 1 representing 8.4% (49 individuals). When we consider other regions, cluster 1 recorded a higher count representing 10.7% (62 individuals) for “north east” region, followed by cluster 3 representing 7.9% (39 individuals) and the least to be cluster 2 representing 7.2% (34 individuals). Cluster 3 is known to have recorded a higher count for most of the regions, followed by cluster 1 and then the least to be cluster 2. We can say that cluster 3 contains a larger portion of individuals from each region, cluster 1 contains a moderate portion of individuals from each region and cluster 2 contains the least proportion of individuals from each region.

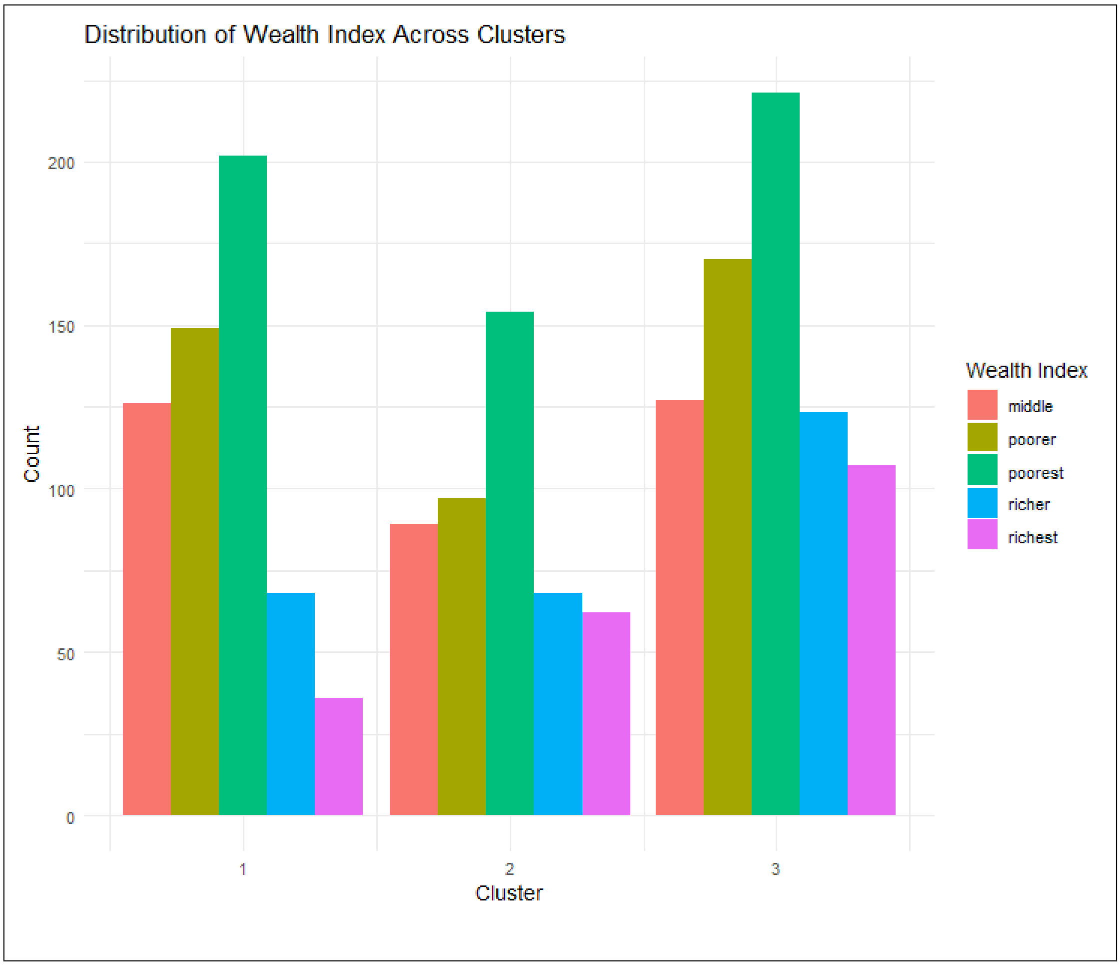

Figure 4.8: Distribution of Wealth Index Across Clusters

The figure displays the wealth levels in different groups. Cluster 1 has the most rich and very rich people. This group is the most wealthy.Cluster 2 has a mix of people. It has many poor, very poor, middle, and rich people. This group is more economically diverse. Cluster 3 has the most poor and very poor people. But it also has many rich and very rich people. This group has the biggest difference between poor and rich people. Overall, the data shows big differences in wealth between the different groups. Some groups are much richer than others. One group has a wider range of wealth levels. This information can help understand the economic differences and problems within the population.

4.6 Health characteristics within the clusters

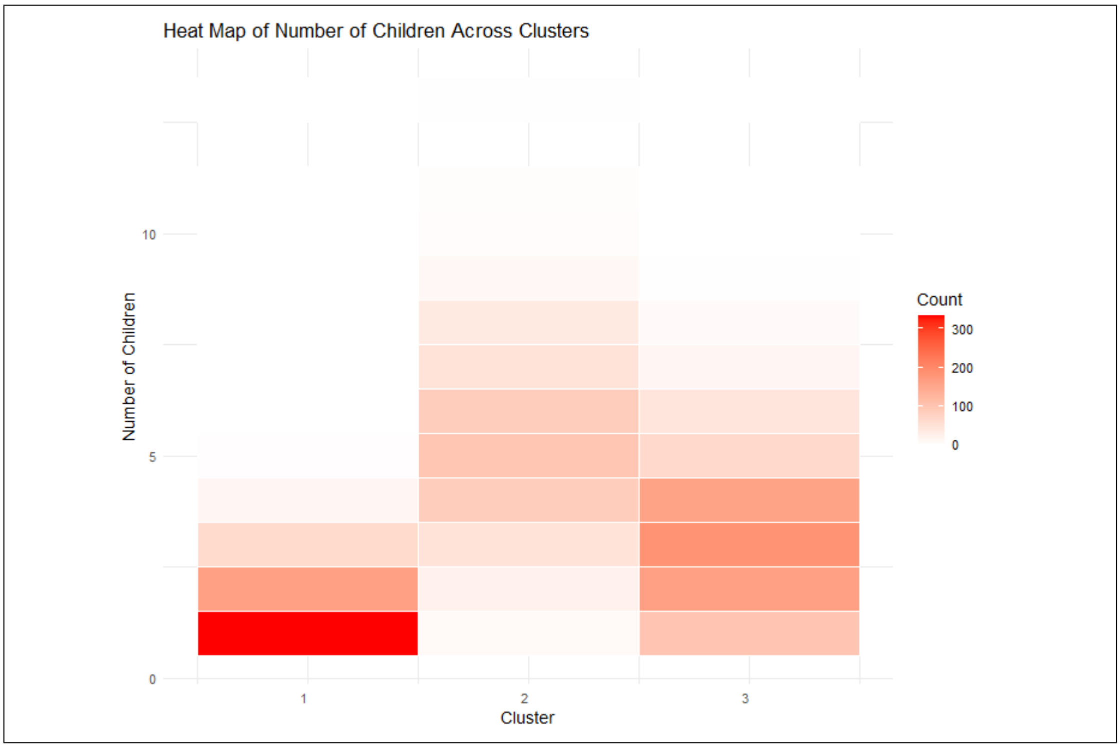

Figure 4.9: Distribution of Number of children Across Clusters

The y-axis (Number of Children) shows the total number of children (ranging from 0 to above 10), while the x-axis (Cluster) indicates the various clusters (1, 2, and 3). The majority of people in Cluster 1 likely to have five children, as seen by the high concentration of those in the cluster (very dark red).When there are five or less children, the count steadily drops, but up to about seven children, it is still clearly there. Compared to Cluster 1, Cluster 2 exhibits a lower intensity and a more equal distribution across various kid counts. People who have between zero and five children are relatively common; their intensity decreases as the number of children rises. With a little less intensity, Cluster 3 has a pattern that is comparable to that of Cluster 2. Most of the people in this cluster have between 0 and 5 children. In conclusion, cluster 1 stands out from the other clusters due to a notable concentration of people who have five or fewer children. Most people in Clusters 2 and 3 have between 0 and 5 children, indicating a wider and more equally distributed range of kid counts. According to this heatmap, Clusters 2 and 3 most likely indicate groups with fewer families, whereas Cluster 1 may represent a group with bigger family sizes.

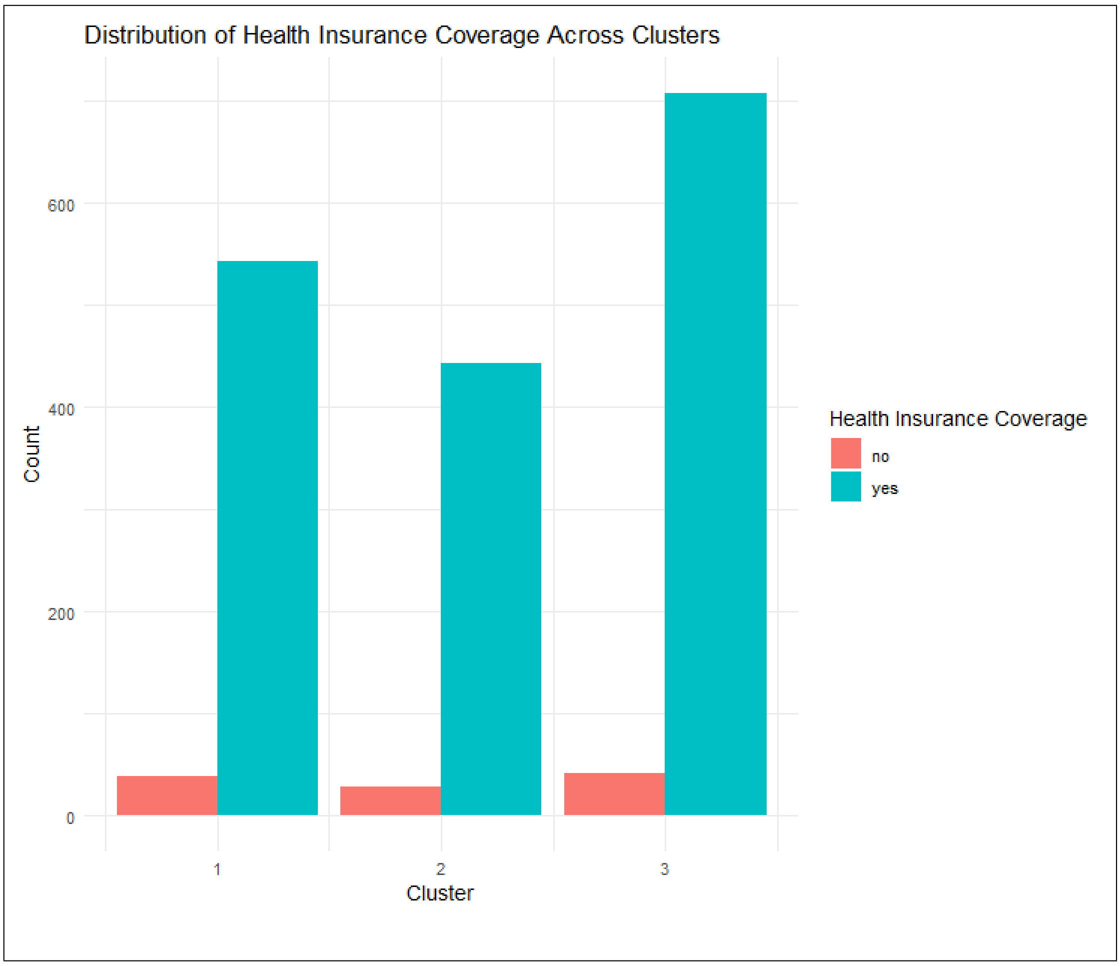

Figure 4.10: Distribution of Health Insurance Coverage Across Clusters

The distribution of health insurance coverage (labeled as "yes" or "no") among three distinct clusters is shown in the Figure 4.10. The height of each bar represents the proportion of people with or without health insurance inside a cluster. Our observations show that the three clusters' levels of health insurance coverage differ noticeably from one another. The highest number of insured individuals is seen in Cluster 3. When compared to Cluster 3, the percentage of insured people in Clusters 1 and 2 is noticeably lower.

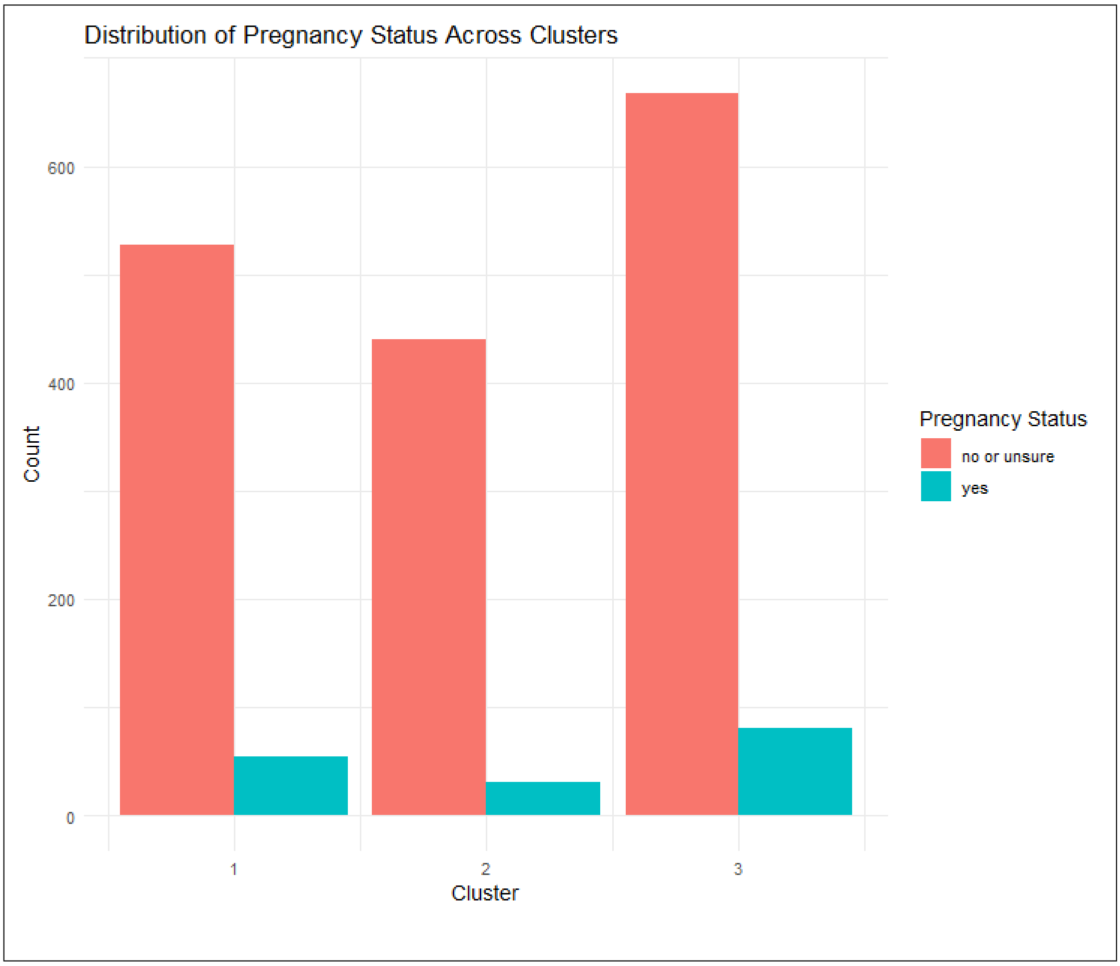

Figure 4.11: Distribution of Pregnancy Status Across Clusters

Our analysis shows that the number of people who are pregnant, not pregnant, or unsure decreases from one cluster to the next. Cluster 1 has the most people in these categories. Cluster 2 has fewer people compared to Cluster 1 but still has many. Cluster 3 has even fewer people who are pregnant or unsure. Cluster 4 has the fewest people in these categories.

Table 4.8: Crosstab of Cluster and WATER SOURCE with Frequencies and Percentages

WATER SOURCE	Cluster 1	Cluster 2	Cluster 3
bottled water	1 (0.2%)	1 (0.2%)	3 (0.4%)
cart with small tank	0 (0%)	1 (0.2%)	0 (0%)
not a dejure resident	11 (1.9%)	1 (0.2%)	14 (1.9%)
piped into dwelling	5 (0.9%)	13 (2.8%)	15 (2%)
piped to neighbor	22 (3.8%)	15 (3.2%)	25 (3.3%)
piped to yard/plot	18 (3.1%)	26 (5.5%)	27 (3.6%)
protected spring	0 (0%)	0 (0%)	1 (0.1%)
protected well	25 (4.3%)	28 (6%)	31 (4.1%)
public tap/standpipe	78 (13.4%)	73 (15.5%)	94 (12.6%)
rainwater	2 (0.3%)	0 (0%)	8 (1.1%)
river/dam/lake/ ponds/stream/canal /irrigation channel	91 (15.7%)	73 (15.5%)	105 (14%)
sachet water	106 (18.2%)	86 (18.3%)	177 (23.7%)
tanker truck	1 (0.2%)	1 (0.2%)	0 (0%)
tube well or borehole	189 (32.5%)	130 (27.7%)	219 (29.3%)
unprotected spring	12 (2.1%)	3 (0.6%)	8 (1.1%)
unprotected well	20 (3.4%)	19 (4%)	21 (2.8%)

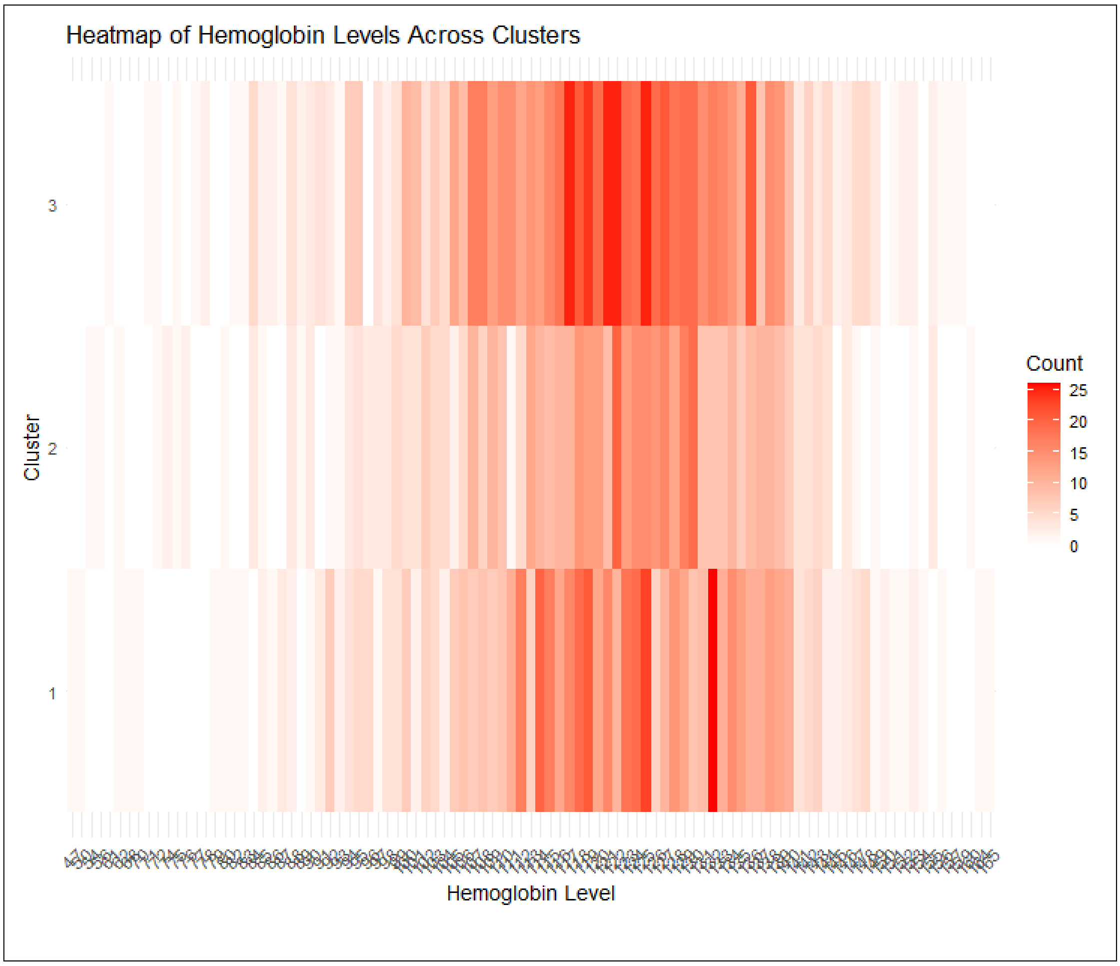

Figure 4.12: Distribution of Hemoglobin Level Across Clusters

The heatmap shown in Figure 4.12 shows the distribution of hemoglobin levels across the different clusters. The horizontal axis represents the hemoglobin level, while the vertical axis shows the cluster number. The heatmap's color gradient,wherein lighter hues denote lower counts and darker shades larger counts, indicates the frequency or count of each hemoglobin level among its corresponding clusters. The darker shades of red represent higher concentrations or higher hemoglobin levels. The lighter shades of red indicate lower concentration or lower hemoglobin level. For cluster 1, we can see that there is a combination of the darker shades and lighter ones. We then say that it has a moderate hemoglobin level. Cluster 2 contains majority of the lighter color gradient. We then say that it has a lower hemoglobin level. Cluster 3 is dominated by majority of the darker color gradients. We then say that it has a higher hemoglobin level.

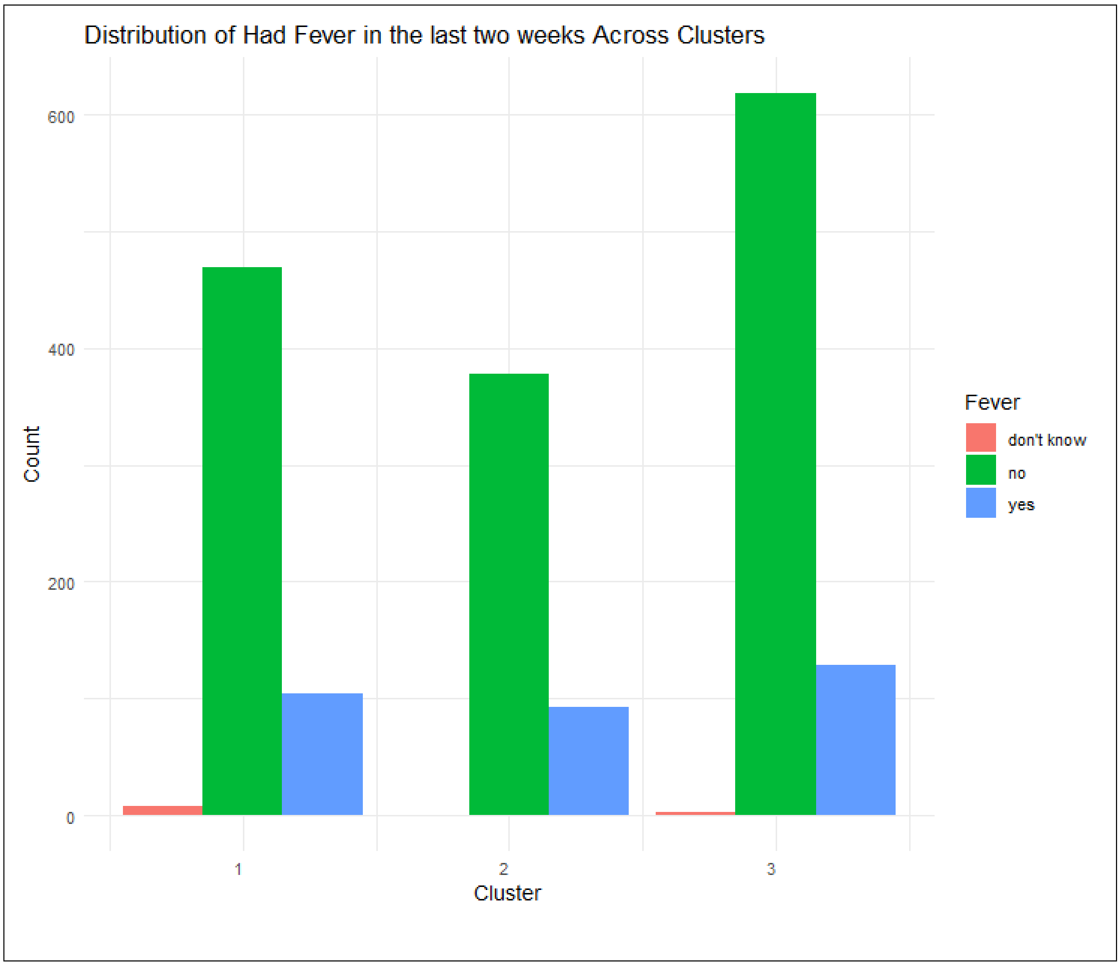

Figure 4.13: Distribution of Had Fever in the last two weeks Across Clusters

From Figure 4.13 above,cluster 1 shows a significant number of individuals who did not report a fever(469 individuals), with a smaller proportion reporting fever(104). Notably, there is a small number of individuals who are unsure about their fever status(8 individuals). Cluster 2 has no individuals who are unsure about their fever status, with a majority reporting not having a fever(378 individuals). Cluster 3 exhibits the highest counts for both no and yes responses, indicating a higher variability in fever incidence within this cluster.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Introduction

The previous chapter analyzed the social, economic, and health details of women of reproductive age in Ghana using cluster analysis. This chapter summarises the findings, focusing on the socio-demographic and health factors linked to high anemia risk, and looks at the clusters with the highest and lowest anemia rates.

5.2 Summary of Findings

5.2.1 Socio-demographic Characteristics of High-Risk Groups

Our analysis shows that age is a key factor in anemia risk. The highest risk is found in the 15-19 and 20-24 age groups, especially in Cluster 1, which includes mostly younger women. This matches the global view that younger women are at higher risk for anemia due to greater physiological demands and often inadequate iron intake World Health Organization (2023). On the other hand, Cluster 2, which has more older women (ages 35-49), also shows a high anemia risk. This could be due to long-term nutritional deficiencies and health problems that are more common in older age Nguyen & Wilson (2023). The socioeconomic status of individuals in each cluster shows clear patterns. Cluster 1, mostly rural, has many people from poorer backgrounds. This is linked to higher anemia rates due to limited access to healthcare, poor nutrition, and worse living conditions Brown & Adams (2019). In contrast, Cluster 3, which has more urban residents and people from richer backgrounds, shows lower anemia rates. The urban environment likely offers better healthcare and nutrition, which helps reduce anemia Green & Patel (2021).

5.2.2 Health Factors Associated with High-Risk Groups

Health-related factors are strongly linked to anemia risk. In Cluster 1, more people lack health insurance and have frequent pregnancies, both of which increase anemia risk Miller & Robinson (2022). Without health insurance, people get fewer health checks and anemia treatments, while frequent pregnancies can worsen nutritional deficiencies Chen et al. (2021). Cluster 3, however, has more people with health insurance and fewer pregnancies, leading to better anemia outcomes Smith & Lee (2023). Access to healthcare and overall better health conditions in this cluster likely help lower anemia rates Williams & Patel (2022).

5.2.3 Anemia Rates Across Clusters

The analysis shows that Cluster 1 has a moderate level of hemoglobin, indicating moderate anemia rates Johnson (2022). Cluster 2 has the lowest hemoglobin levels and the highest anemia risk. This cluster's older age, poorer socioeconomic status, and less health insurance contribute to the high anemia rate. On the other hand, Cluster 3 has higher hemoglobin levels and better health insurance coverage, more urban residents, and fewer pregnancies, all of which are linked to lower anemia risk Green (2021).

5.2.4 Implications for Policy and Interventions

The findings suggest the need for targeted interventions for high-risk groups. For Cluster 1, improving healthcare access and nutrition for younger women, especially in rural areas, is important **?**. Education on anemia prevention and health insurance could also help Smith et al. (2022).

For Cluster 2, addressing socioeconomic barriers is essential. Programs to improve economic conditions, enhance healthcare access, and provide targeted anemia treatment for older women could help reduce high anemia rates in this group.

For Cluster 3, continuing support for urban healthcare and maintaining health insurance accessibility can help keep anemia rates low Williams & Patel (2022). Expanding successful strategies from this cluster to others could further reduce anemia **?**.

5.2.5 Conclusion

This chapter has highlighted how socio-demographic and health factors affect anemia risk among women of reproductive age in Ghana. By identifying high risk groups, policymakers and health practitioners can create better strategies to combat anemia and improve health outcomes. Future research should explore these factors further, considering additional variables and long-term data to improve intervention strategies.

REFERENCES

Agyemang, W. et al. (2023). Prevalence and risk factors of anaemia among women of reproductive age in a ghanaian.

Alem, A. (2023). Prevalence and factors associated with anemia in women of reproductive age across low- and middle-income countries based on national data. *Scientific Reports*, 13(1), 1–13.

Appiah, A. N. (2020). Title of the article. *Journal Name*, XX(X), XX–XX.

Ardayfio, L., Albonjco, T., & Steketee, R. (2014). Ghana: Landscape analysis of anemia and anemia programming. *SPRING*.

Armo-Annor, D. e. a. (2021). Risk of anemia among women engaged in biomass-based fish smoking as their primary livelihood in the central region of ghana: a comparative cross-sectional study. *BMC Nutrition*, 7(50).

Brown, J. & Adams, R. (2019). Socioeconomic factors and anemia prevalence. *Journal of Health Economics*, 15(2), 89–104.

Chaparro, C. et al. (2019). Anemia epidemiology , pathophysiology , and etiology in low- and middle-income countries. *Annals of the New York Academy of Sciences*, 1450, 15–31.

Chen, L., Patel, S., & Wilson, A. (2021). Health insurance and anemia in developing countries. *International Journal of Public Health*, 18(3), 204–216.

Christian, A. et al. (2022). Micronutrient deficiencies among non-pregnant women of reproductive age in ghana: Results from a nationally representative survey.

Cleveland Clinic (2024). Iron deficiency anemia. Accessed: 2024-05-16.

Coffey, D. e. a. (2015). Sanitation, disease, and anemia: Evidence from nepal. *Journal of Health Economics*, (pp. 1–37).

Dzabeng, F. et al. (2019). Factors influencing the use of anemia preventing measures among antenatal clinic attendees in the kintampo north municipality , ghana methods. *African Journal of Reproductive Health*, 23, 35–43.

Ghosh, K. e. a. (2021). Prevalence and detecting spatial clustering of anaemia among children 6–59 months in the districts of india. *Clinical Epidemiology and Global Health*, 12(100845).

Green, A. (2021). Factors affecting anemia risk: A comprehensive review. *Journal of Hematology Research*, 15(2), 123–135.

Green, D. & Patel, N. (2021). Urban vs rural health outcomes: A comparative study. *Urban Health Review*, 12(1), 65–80.

Hulland, E. et al. (2021). Anemia design effects in cluster surveys of women and young children in refugee settings. *PLoS ONE*, 16(7), 1–10.

Johnson, M. (2022). Hemoglobin levels and anemia rates. *Journal of Hematology*, 29(1), 45–56.

Keokenchanh, S. et al. (2021). Prevalence of anemia and its associate factors among women of reproductive age in lao pdr: Evidence from a nationally representative survey.

Kibret, K. et al. (2019). Spatial distribution and determinant factors of anaemia among women of reproductive age in ethiopia: a multilevel and spatial analysis. (pp. 1–14).

Kibret, K. et al. (2025). Spatial distribution and determinant factors of anaemia among women of reproductive age in ethiopia: a multilevel and spatial analysis. (pp. 1–14).

Kinyoki, D. et al. (2021). Anemia prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018. *Nature Medicine*, 27(10), 1761–1782.

Macdonald, C. et al. (2010). Anemia - can its widespread prevalence among women in developing countries be impacted? 1 . 0 anemia in women: A global health priority.

Miller, T. & Robinson, P. (2022). Pregnancy and anemia risk factors. *Maternal Health Journal*, 10(2), 112–125.

MSD Manuals (2024). Vitamin deficiency anemia. Accessed: 2024-05-21.

National Heart, Lung, and Blood Institute (2024). Hemolytic anemia. Accessed: 2024-05-21.

Ngimbudzi, E. et al. (2021). The burden of anemia in pregnancy among women attending the antenatal clinics in mkuranga. *Frontiers in Public Health*, 9, 1–9.

Nguyen, T. & Wilson, A. (2023). Older age and anemia: A critical review. *Age and Health Journal*, 30(3), 98–110.

Obeagu, E. (2023). Depression in sickle cell anemia: An overlooked battle.

Omari, R. (2017). Nutrition knowledge and food consumption practices and. *African Journal of Food, Agriculture, Nutrition and Development*, 17(1), 11639–11656.

Owais, A. N. (2021). Title of the article. *Journal Name*, XX(X), XX–XX.

Quaye, L. (2016). Haemato-biochemical basis of anemia in agogo, ashanti region, ghana. *5*, (pp. 1–12).

Region, E. et al. (2014). Treatise on anaemia in pregnancy in women of reproductive age in the bolgatanga municipality , upper.

Safiri, S. et al. (2021). Burden of anemia and its underlying causes in 204 countries and territories , 1990 – 2019: results from the global burden of disease study 2019. *Journal of Hematology & Oncology*, (pp. 1–16).

Savera, S. (2020). Prevalence and determinants of anemia among women of reproductive age in developing countries of asia and africa: A systematic review protocol.

Smith, A. & Lee, K. (2023). Health insurance and anemia: A new perspective. *Health Policy Review*, 19(1), 78–91.

Smith, A., Lee, K., & Johnson, M. (2022). Socioeconomic impacts on anemia. *Health and Society Journal*, 23(3), 143–159.

Souganidis, E. et al. (2012). Determinants of anemia clustering among mothers and children in indonesia. *Journal of Tropical Pediatrics*, 58(3), 170–177.

SPRING (2024). Reducing anemia in ghana: The spring approach and lessons learned. *SPRING*.

Sumaila, I. et al. (2022). Prevalence and associated factors of anaemia among antenatal care attendants in the kintampo municipalit. *Pan African Medical Journal One Health*, 9.

Sunuwar, D. et al. (2021). Factors affecting anaemia among women of reproductive age in nepal: A multilevel and spatial analysis. *BMJ Open*, 11(3).

Tettegah, E. et al. (2023). Risk factors associated with anaemia among pregnant women in the adaklu district, ghana. *Frontiers in Global Women's Health*, 4, 1–10.

Tirore, L. et al. (2023). Prevalence and associated factors of severity levels of anemia among women of reproductive age in sub-saharan africa: a multilevel ordinal logistic regression analysis. *Frontiers in Public Health*, 11, 1–11.

Tsegaye, B. et al. (2023). Trend and factors associated with anemia among women reproductive age in ethiopia: A multivariate decomposition analysis of ethiopian demographic and health survey. (pp. 1–22).

Vroom, F. et al. (2024). Anaemia among women of reproductive age in selected sub-saharan african countries: multivariate decomposition analyses of the demographic and health surveys data 2008 – 2018.

Williams, H. & Patel, S. (2022). Urban health initiatives and anemia. *Urban Health Review*, 15(2), 105–118.

World Health Organization (2023). *Global report on anemia*. WHO Publications.

Wu, Y. et al. (2020). Prevalence of anemia and sociodemographic characteristics among pregnant and non-pregnant women in southwest china: A longitudinal observational study. *BMC Pregnancy and Childbirth*, 20(1), 1–10.