Determinants and Spatial Distribution of HIV/AIDS in Jimma Zone, Southwest Ethiopia

By:

Nemso Geda Bedaso

A Thesis Submitted to Department of Statistics, College of Natural Sciences, Jimma University as a Partial Fulfillment for the Requirements of the Degree of Master of Science (MSc) in Biostatistics

September 2020

Jimma, Ethioipa

Determinants and Spatial Distribution of HIV/AIDS in Jimma Zone, Southwest Ethiopia

By:

Nemso Geda Bedaso

Advisor: **Prof. Legesse Kassa Debusho (PhD)**

Co Advisor: **Geremew Muleta (Ass.Prof, PhD Scholar)**

September 2020

Jimma, Ethioipa

STATEMENT OF THE AUTHOR

As author of this research study, I declare that the thesis is a result of my work, support of my supervisors and help hands of other individuals. Thus, all those had who participated in the study and sources of materials used for writing this thesis have been duly acknowledged. I have submitted this thesis to Jimma University as a partial fulfillment for the requirements of Degree of Master of Science in Biostatistics. The library directorate of Jimma University can deposit the copy of the thesis in the university library so that students and researchers can refer it.

Nemso Geda Bedaso

Date: ———

Signature: ———

August 2020

Jimma, Ethioipa

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

MSc THESIS APPROVAL SHEET

We, the undersigned, member of the Board of Examiners of the final open defense by **Namso Geda** have read and evaluated his/her thesis entitled "**Determinants and Spatial Distribution of HIV/AIDS in Jimma Zone, Southwest Ethiopia**" and examined the candidate. This is therefore to certify that the thesis has been accepted in partial fulfillment of the requirements for the degree Master of Science in **Biostatistics** (____________________________).

Akalu Banbeta		03.09.2020
Name of the Chairperson	Signature	Date
Prof Legesse Kassa Debusho		04/09/2020
Name of Major Advisor	Signature	Date
Akalu Banbeta		03-09-2020
Name of the Internal Examiner	Signature	Date
Zeytu Gashaw Asfaw (PhD)		03.09.2020
Name of the External Examiner	Signature	Date

ACKNOWLEDGEMENTS

Above all, I would like to thank the almighty God for the gift of health, wisdom and strength throughout my entire life.

I would like to express the most sincere thanks and profound sense of gratitude to my advisor Prof. Legesse Kassa Debusho (PhD) from University of South Africa, South Africa, for his professional and deep comments, advice and guidance at all stages of this thesis. I appreciated his friendly communication, his motivational advice in academic related matter and his speedy feedbacks for the work that I have been submitting related to this research.

I sincerely appreciate my co-advisor Geremew Muleta (Assistant Prof, PhD Scholar) for his valuable suggestions and comments for the successful realization of this thesis. It is a great pleasure to express his commitment to share his experience, knowledge and his smooth contacts is also appreciable.

My thanks also to Jimma Zone Health Bureau and all health center of Jimma Woredas for providing me with the data for this thesis. Grateful acknowledgment is also expressed to the Department of Statistics, Jimma university for kind assistance in many ways.

Furthermore, I want to thank all my friends, my family members and others who have provided me a great help and moral support directly and indirectly contributed towards the completion of this study.

Contents

List of Tables

List of Figures

ACRONYMS

AIDS	Acquired Immunodeficiency Syndrome
AIC	Akaike Information Criterion
ART	Antiretroviral therapy
BIC	Bayesian Information Criterion
CDC	Centers for Disease Control
DPC	Disease prevention and control department
EDHS	Ethiopian Demographic Health Survey
EPHIA	Ethiopian Population-based HIV Impact Assessment
ENRM	Ethiopia National Road Map
FMoH	Federal Ministry of Health
GLM	Generalized linear models
HIV	Human Immunodeficiency Virus
MLE	Maximum Likelihood Estimation
PLWAH	Peoples Living with HIV AIDS
RMLE	Restricted Maximum Likelihood Estimation
SSA	Sub-Saharan Africa
UNAIDS	United Nation Acquired Immunodeficiency Syndrome
WHO	World Health Organization

ABSTRACT

Background: *HIV is a major cause of illness and death in large parts of the developing countries, especially in Africa. Accurate estimates of HIV distribution are required for planning, implementation and evaluation of HIV control programs. Studies that shows the comparison of distribution of HIV over place to place in Ethiopia specially at Jimma Zone with at Woreda level is very limited. Hence, this study did investigate the spatial distribution of HIV and associated factors at Wereda level in Jimma Zone.*

Objectives: *The main objective of this study was to examine spatial patterns and risk factors of HIV distribution in Jimma zone, Oromia Region, Ethiopia.*

Methodology: *The study was conducted in Jimma zone of entire districts and the data is secondary which is obtained from Jimma zone health center. Spatial distribution of HIV were identified using global and local measures of spatial auto correlation. Binary regression model was used to analyze covariates related to HIV distribution using R software and to analyze spatial point pattern data relevant software GeoDa and ArcGIS were used.*

Results: *The results of this study indicated that, the presence of significant global spatial autocorrelation of HIV Distribution in Jimma zone. Based on the p-values of Moran's* $I(0.002 < 0.05)$ *and Geary's* $C(0.0009 < 0.05)$ *coefficients, we reject the null hypothesis of no spatial autocorrelation. The results of binary regression model including spatial dependence show that, there is statistically significant relationship between HIV status and Age, Marital status,religion,and place of residence.*

Conclusions: *There is evidence of significant HIV clustering in Jimma zone, southwest Ethiopia. Significant hot spots clusters were identified in eight districts and cold spots of HIV clusters were identified in ten districts. Clustering of dissimilar values identified in four Woreda. Model based data analysis showed that there is significant relationship between HIV distribution and Age,Gender,Marital status,Education level,religion,residence and condom use. The study recommends that interventions should be facilitated in highly clustered HIV distribution areas by giving special attention in targeting intervention and health services to the highly risk exposed districts and neighboring districts.*

Key words: *Binary regression model, Geary's coefficient, Global Moran's coefficient, HIV/AIDS, Local Moran's coefficient, Spatial distribution*

CHAPTER ONE

1 INTRODUCTION

1.1 Background of the Study

Human Immunodeficiency Virus (HIV) is the most dangerous virus and attacks the body's immune system [1]. Over a period of time, the disease can progress to a stage where the infected person has a collection of symptoms and a variety of infections, this condition is then known as Acquired Immunodeciency Syndrome (AIDS), which means if HIV is not treated it can lead to AIDS [2].

Human Immunodeficiency Virus is transmitted through direct contact of a mucous membrane or the blood stream with a bodily fluid containing HIV, such as blood, semen, vaginal fluid, pre-seminal fluid, and breast milk. These transmission can come in the form of anal, vaginal or oral sex, blood transfusion, contaminated hypodermic needles and HIV positive mother can transmit HIV to her baby in during pregnancy, childbirth (also called labor and delivery), breastfeeding, or other exposure to one of the above bodily fluids [3]. With increasing access to effective HIV prevention, diagnosis, treatment and care, HIV infection has become a manageable chronic health condition, enabling people living with HIV to lead long and healthy lives. However, if medical advise is not followed HIV continues to be a major global public health issue.

According to 2018 global HIV statistics an estimated 37.9 million people were living with HIV (including 1.7 million children). South Africa has the highest number of people living with HIV (7.7%) in the world, whereas Swaziland has the highest prevalence in the world, 27.3% [4]. Mexico, Malaysia, and Afghanistan have the lowest rates of HIV/AIDS infection in the world. According to the HIV Fact Sheet 2019 statistics, there were 1.7 million new infection cases in 2019 worldwide. East and Southern Africa is the region most affected by HIV in the world and is home to the largest number of people living with HIV. Despite availability of improved HIV testing services, still two out of every ten people living with HIV are unaware of their status.

The number of Peoples Living with HIV/AIDS (PLHIV) in this region is continues to increase, but access to antiretroviral treatment is increasing as well. Although laws and cultural traditions vary between East and Southern African countries, there are a number of ingrained cultural, structural and legal barriers to HIV prevention [4]. Ethiopia's first cases of HIV were reported in 1986 and the disease rapidly spread [5]. Ethiopia faces an epidemic among sub-populations and geographic areas, with an estimated overall HIV prevalence rate of 1.4 percent. Ethiopia has large and very vulnerable population, with an estimated 15% of the population living below the poverty line. HIV/AIDS is one of the key challenges for the overall development of Ethiopia, as it has led to seven year decrease in life expectancy and greatly reduced workforce [4].

In 2016, according to (UNAIDS) Program on HIV/AIDS in Ethiopia, there were 718,500 people who live with HIV among which 65,088 were children from 0-14 years of age. In 2017, UNAIDS reported that 127,619 HIV positive people and the epidemic is primarily associated with areas of urban and major transport corridors. The HIV epidemic in Ethiopia is heterogeneous by sex, geographic areas and population groups. Among women and men combined, HIV prevalence is seven times higher in urban areas than in rural areas (2.9% versus 0.4%) [6]. HIV prevalence is 3.6% among women in urban areas compared with 0.6% among women in rural areas. Seven out of nine regional states and two city administration have HIV prevalence above 1% [6]. According to HIV related estimates and projections for Ethiopia there are 610,335 people living with HIV (PLHIV) with 0.96% estimated adult HIV prevalence [7].

The 2016 Ethiopian demographic and health survey (EDHS) report shows that Gambella region (4.8%) had the highest HIV prevalence rates followed by Addis Ababa (3.4%) while Somali ($< 0.1\%$) and Southern Nations, Nationalities and peoples regions had less than 0.1% and 0.4% HIV prevalence rates, respectively. The report also shows that the adult (15-59 years old) HIV prevalence in Ethiopia was 0.9%. Knowing the spatial distribution of HIV can be helpful to put in place HIV control and prevention programs as well as for generating etiologic hypothesis. Spatial auto correlation is useful for [56]luster mapping of Woreda's health care problems ˙

Cluster mapping helps to clarify issues such as the spatial aspects of both internal and external correlations of leading health care events which in turn help planners to assess spatial risk factors and identify the most plausible types of health care policies for planning and implementation of health care services.

Southwest district of Ethiopia involved different ethnic population and refugee camp, and adds lots of HIV patients attending ART clinic in Jimma town public health facilities. Also, Jimma town is nearby Gambella state which is a small and sparsely populated region that has the highest regional HIV prevalence. Still the rise of the new HIV infection epidemic in Ethiopia is broadly documented [8]. The prevalence of HIV infection among infants born to HIV infected mothers in Jimma Zone, southern west Ethiopia is 5.3% [10]. The main goal of this study is to examine spatial patterns or clusters of HIV distribution and associated factors in Jimma Zone. It seeks to identify HIV "hotspot" and "coldspot" Woreda by producing map of clustering observation and fit appropriate spatial models for HIV distribution in Jimma Zone.

1.2 Statements of the Problems

The epidemic of HIV is spread unevenly in the world, with high burden in Sub-Saharan Africa including Ethiopia. Some regions of the World are highly affected by this pandemic, and in other areas it is negligible. The rapid expansion of HIV in sub-Saharan African countries has a profound impact on the health sector as well as the socio-economic development of the region in general. In Ethiopia HIV is the leading causes of morbidity and mortality by accounting million HIV cases and thousands of deaths annually. In Jimma zone HIV is one of the public health problems affecting many people each year.

Different researches have been done to see the progress of disease from time to time [1] & [7]. However, information that shows the comparison of HIV distribution over place to place in Ethiopia specially at Jimma Zone with in Woreda level is very limited. A study conducted on HIV positive sero-status disclosure and its determinants among people living with HIV/AIDS following ART clinic in Jimma University Specialized Hospital indicates that age, sex, educational status and marital status had significant association with HIV sero-status disclosure [1]. Since study was done only in one hospital, generalizing of these findings for entire population of Jimma Zone is not possible and also, comparing the spatial distribution of HIV/AIDS from Woreda to Woreda is difficult.

But in this study, the author fill such gap adding more health center facilities rather than one hospital by clustering Jimma Zone at Woreda level. In spatial dependence the Districts that are in closer proximity are expected to have similar characteristics. Measures of Spatial autocorrelation helps to assess the degree of dependency among observations in a geographic space. It requires measuring a spatial weights matrix that reflects the intensity of the geographic relationship between observations in a neighborhood. Controlling HIV at Woreda (districts) level needs identifications of spatial pattern of HIV distribution among the districts of Jimma zone. We had measured spatial dependence using global and local measures of spatial autocorrelation to find spatial pattern of HIV in districts of Jimma zone, southwest Ethiopia.

kibret et al. (2019) and Adal. (2019) have investigated the trends and spatial distributions of HIV cases in Ethiopia and both reported that Gamebela has the highest prevalence (4.8%) in the country [9] and [7].These studies focused only on HIV prevalence at regional level, however conducting such analysis at a lower levels, for example Zones level or district level can produce more information that could be used to develop public health policy, which helps to minimize the HIV/AIDS transmission rate. This study did investigate the spatial distribution of HIV/AIDS and modeled the HIV status using the generalized linear model (GLM) for data collected from Jimma Zone.

To the author knowledge there is no study (no published research) that shows the spatial patterns and associated factors for spatial distribution of HIV cases at woreda or district level in Jimma Zone until the completion of this thesis. Therefore, it is necessary to explore the spatial pattern of HIV, which is very important for the government to formulate appropriate strategies, to give more attention for woredas with high HIV infection rate.

The following research questions are considered in this study:

- Does the spatial distribution of HIV spatially random or clustered?
- Which woredas are under hot spot and cold spot?
- How to check spatial dependence in the model?
- Which predictor variables significantly affect the distributions of HIV?

1.3 Objectives of the Study

1.3.1 General Objectives of the Study

The general objective of this study is to examine spatial patterns and risk factors of HIV distribution in Jimma zone, Oromia Region, Ethiopia.

1.3.2 Specific Objectives of the Study

- To identify the spatial distribution of HIV in Jimma zone.
- To identify woredas under hot and cold spot.
- To check spatial dependency in the model.
- To identify variable that associated with the HIV distribution.

1.4 Significance of the Study

The result of this study will help in the health sector to formulate appropriate strategies and interventions. The study is may help the organization as well as individuals who work in this area to get a clue on to what extent HIV distribution is serious across the Woredas of Jimma zone. The other basic significance of the study is that it will further assist other researchers interested in this area to use the results obtained as a benchmark for their future works in identifying spatial distribution of disease either in this area or other areas. Furthermore, the study helps practitioners how to do similar analysis at lower level geographical areas.

1.5 Limitation of the Study

In this study one of the basic limitation is the factor which related to health area (clinical diagnostic related variables) that associated with HIV positive status is not recorded in health center facilities or this study was limited to few variables recorded at the health office. Therefore, the researcher used only the socio-demographic Characteristics and some Behavioural factors which are documented in health center.

1.6 Organization of the Thesis

This thesis has five chapter and organized as follows. The first chapter describes brief introduction about the study, statement of the problem, Objectives of the study, Significance of the study and limitation of the study. The second chapter provides overview of HIV/AIDS, model based and explanatory spatial data analysis, HIV/AIDS in Ethiopia and determinants of HIV/AIDS distribution. The third chapter describes the study area, source of data, variables in the study and the methodology that used for analysis. The fourth chapter provides outputs of explanatory spatial data analysis, outputs of model based data analysis and discussion of each output. Chapter five provides Conclusion and Recommendation based on the findings of the study.

CHAPTER TWO

2 LITERATURE

In this chapter we present the literature review relevant to the study conducted including application of statistical methods used in this study.

2.1 Overview of HIV/AIDS

HIV is positioned among the most important public health problems in Sub-Saharan Africa (SSA) for being a major cause of chronic infections and premature deaths [11]. The disease causing agent of HIV and AIDS is a virus from the retrovirus family. There are two different viruses both from the same family which causes HIV. They have been identified as HIV-I and HIV-II. The type of virus which is mostly confined to the West African setting is the HIV-I [12]. The HIV/AIDS epidemic is no longer merely a health problem but one of the greatest development challenges the world has ever faced [30]. HIV/AIDS has decreased life expectancy and the quality of life of human kind and is a serious impediment to economic development.

The Demographic and Health Surveys (DHS) conducted in the Eastern Africa region reported the prevalence ranging from 5.7-6.4%, and varying greatly across regions [14]. It is also estimated that in 2014, 36.9 million people were living with HIV/AIDS and 0.8% of which are within the active labor force age of 15 to 49 years worldwide [15]. Majority of the studies that report on HIV issues such as Demographic and Health Surveys (DHS), UNAIDS or USAID reports, have generalized that the HIV sero-prevalence according to regions and yet due to variations, differences of high and low prevalence exist within any given region [11]. According to the end of 2019 UNAIDS Report on the Global AIDS epidemic, approximately 24.5 million people are living with HIV worldwide. Sub-Saharan Africa remains among the hardest hit regions by the pandemic, with nearly one in every 25 adults (4.2%) living with HIV, accounting for nearly two-thirds of the global total HIV cases. Senegal and Uganda are still the only two SSA countries that have successfully controlled HIV/AIDS through vigorous and sustained prevention campaigns, by successfully mobilizing all sectors of society in the fight against HIV/AIDS [15].

2.2 Model Based and Explanatory Spatial Data analysis

Zhang et al.(2015) made a population based study on AIDS epidemic and economic input impact factors in Chongqing, China [16]. The main aim of the study was to analyse the spatial-temporal clustering of the HIV epidemic in Chongqing and to explore its association with the economic indices of AIDS prevention and treatment. Spatial clustering, temporal spatial clustering and spatial regression were used to conduct statistical analyses. The result of the study indicates that the HIV/AIDS epidemic showed a non-random spatial distribution (Moran's I ≥ 0.310).

Heimeret et al. (2008) used Moran's I and nearest neighbour analysis to study the spatial distribution of HIV prevalence and incidence among injection drug users in St Petersburg.[17] [?] adopted two spatial statistical methods to explore the clustering of HIV infection in the rural population in KwaZulu-Natal, South Africa, while Jia et al (2011) used the spatial analysis model to investigate the spatial distribution of HIV/AIDS in China from to 2009.

The application of spatial analysis and Geographical Information Systems (GIS) in identifying and visualising clusters and spatial distribution of diseases are methods that are being used in public health and epidemiological research to manage and analyse health data. GIS techniques generate maps which usually provide a more comprehensive display of the pattern and magnitude of disease from local to global scales [21].

Khademi et al. (2016) aimed to investigate the distribution pattern of HIV using the spatial statistics and spatial clustering tests (Average Nearest Neighbor, Moran's I index and Getis-Ord general) in Kermanshah Metropolis, western Iran [19]. It is believed that the application of GIS towards studying the distribution of HIV, could supply the health care policy makers with insights into this disease in Iran. The results indicate that HIV prevalence is expanding across the metropolis of Kermanshah and HIV prevalence was clustered (not accidental) and followed a regular basis over the period under study.

Tsai et al. (2009) employed spatial auto correlation methodologies, including Global Moran's I and Local Getis-Ord statistics [22] . The objective of the study was to describe and map spatial clusters, and areas in which these are situated, for 20 leading causes of death in Taiwan. The results indicate that Cluster mapping helps to illuminate issues such as the spatial aspects of correlations for leading health care events.

Seguy et al. (2006) targeted to assess HIV infection clustering and trends in Libya [20]. The results of the study indicated that HIV cases steadily increased within the Libyan population, particularly among those aged less than 27 years. Spatio temporal analysis showed marked geographic and temporal variation of HIV infection, particularly during 2005–2012. The risk factors varied from one region to another, and the contribution of injection drug use to infection increased with time.

Daw et al. (2018) use spatial analysis of HIV/AIDS clusters in order to identify epidemic and trends in the country [20] . The prevalence of HIV infection is evaluated according to sex, age, and altitude of participating households. Kulldorff's spatial scan statistic are used to test HIV clustering in the study area. The results of Kulldorff's spatial scan statistic detected that, one marginally significant HIV-positive and one significant HIV-negative cluster in the study area [63].

2.3 HIV/AIDS in Ethiopia

According to World heath organization 2005 HIV testing and counseling began in Ethiopia in the late 1980s and expanded during the 1990s . Since then, the epidemic has spread to the general population in both urban and rural areas. Even though Ethiopian health institute has a different structure of working and introduced sectors to prevent communicable disease, but still the issue of controlling the communicable disease remains to be not well addressed. Thus, according to the FMoH report, to do further on this problem the country has formed Health Sector Transformation Plan (HSTP) during the second growth and transformation plan of 2015 by giving more emphasis for the disease like malaria, HIV and TB [23] .

The FDRE has committed in reducing new adult HIV infections by 50 percent by 2020 and to ending AIDS as a public health threat by 2030 [6]. HIV/AIDS is one of the key challenges for overall national development in Ethiopia. It has led to a seven-year loss in life expectancy, close to a million orphans and a loss of productivity and income at the workplace with severe effects at the household and community levels. The high rates of morbidity and mortality associated with HIV/AIDS have strongly affected the health sector and are among the major impediments to delivering quality care to its full capacity. The 25 countries with the highest numbers of new HIV infections were selected for the Global HIV Prevention Coalition, of these, 17 are African countries, including Ethiopia [6].

In of 2003, Ethiopia had an estimated 950 000 to 2.3 million people living with HIV, among the highest in the world. An estimated 120 000 adults and children died from HIV in of 2003, and 720 000 children younger than 17 years had been orphaned by HIV at the end of of 2003 [24]. HIV transmission in Ethiopia occurs mainly through heterosexual contact. Some transmission also occurs from mother to child and through transfusion of infected blood and unsafe medical practices. With 45% of Ethiopia's population under 15 years of age, young people are especially vulnerable. Other vulnerable population groups include unemployed people, long-distance truck drivers, migrant workers and internally displaced populations ([24]; [25]).

The FMoH reported that the national HIV prevalence was between 3.5% and 5% among blood donors in 2005, where the prevalence for those blood donors in the age group 15–19 years was 2.9% but the highest prevalence occurred among donors in the age group of 30–39 years. The age range at which people become infected was 15 to 24 years for females and 25 to 34 years for males. By 1988, high rates of HIV prevalence (17%) were detected among commercial sex workers residing along the main trading roads and long distance truck drivers (13%) [23] (FMoH, 2002). The number of HIV infections among adult Ethiopians was estimated at 722248 in 2017, increasing by 3748 infections from 2016. The highest prevalence rates are in the age group 15–24 years, and women in this age group are especially vulnerable. The highest estimated prevalence's among adults were in

Addis Ababa (5%) and Gambella (4%). The two administrative regions, Gambella and Addis Ababa persist with high load of HIV cases for long time so far [26]. The study conducted on spatial dstribution of HIV transmission in Ethiopia and characteristics of HIV clusters using the 2005, 2011 and 2016 DHS data shows that there were 11,383, 29,812, and 26,753 individuals with HIV in years 2005, 2011, and 2016, respectively [9]. Four HIV clusters were identified consistently over the three time points, with the clusters representing 17% of the total population and 47% of all HIV cases. Looking at prevalence by region, HIV is highest in Gambella (4.8%) followed by Addis Ababa (3.4%), Dire Dawa (2.5%) and harari (2.4%) [14].

Abebe et al. (2018) aimed to examine health-related quality of life and associated factors among HIV positive women receiving in health facilities of Jimma town [27]. A cross-sectional study was conducted, and consecutive sampling technique was employed to select 377 HIV positive women who were on antiretroviral therapy. Descriptive statistics, bivariate, and multivariable logistic regression analyses were performed. The study demonstrated high proportion of HIV positive women on ART had poor health-related quality of life which was afected by wealth index, social support, and duration on ART. The high incidence of sexually transmitted infections, the prevalence of multiple sexual partners and harmful traditional practices such as female genital mutilation and body piercing have also contributed to the spread of the epidemic [28].

2.4 Determinants of HIV Disease

The risk of incidence and prevalence of HIV infection can be reduced or spread out widely depending on the culture, demographic and socioeconomic factors that determine prevention and the mitigating abilities of a given society ([29] [30] . The studies on the prevalence and correlates of HIV infections in Sub-Saharan Africa has shown large differentials in the prevalence of HIV by age, sex, place of residence and geographical region within and between countries ([11] [31]). There are also other factors, such as, educational attainment, occupation and exposure to the media, that can influence risk taking behavior's ([21] , [32] [22]) and also lead to increased risk of HIV infection.

2.4.1 Gender

The world health organization report of 2017 quantified that the number of male HIV infected is greater than that of the female. The study done in South Africa showed that women were more likely to disclose their HIV positive sero - status as compared to their male counterparts [33]. In research conducted among tissue donors in the United States, the incidence of HIV infection was estimated to be 40,000 cases per year, with high cases in males (70%) and low cases in females(30%). The 2010 USAID report on the HIV health profile in East Africa indicated that compared to males, female adolescents in Ethiopia were 1.4 times at risk of HIV infections. In Kenya and Tanzania, the prevalence of HIV in young women was 4 times higher than in young men [11]. According to EDHS 2011 report, HIV prevalence was significantly higher among adult females (1.8%) compared to adult males (1.0%). The Study conducted on spatial distribution and risk factors for HIV infection in the Kenyan fishing communities of Lake Victoria indicates significant differences between men (29%) and women (38%).

2.4.2 Age

Age of the people is one of the key factors that is assumed to have an influence on HIV distribution over the world. The WHO (2017) report that out of 40,000 cases per year the patients were younger than 30 years,(18%), 30 to 49 years (71%) and 50 years and above (11%).

According to the end of 2018 UNAIDS Report on the Global AIDS epidemic approximately 37.9 million people globally were living with HIV virus, 36.2 million adults and 1.7 million children ($<$ 15 years). Amornkul et al. (2009) Reported that in Kenya different HIV prevalence among older adolescents (15-19) females (3.5%) and males (1.8%), respectively [34]. The study conducted in Libya was reported that, Compared with the younger age group (15–19 years), adults aged 30–34 years were 6.71 times more likely to be HIV-positive. The estimated HIV-positive population among women was 1.43 times larger than among men[20].

2.4.3 Occupation

Munoz-Laboy et al. (2014) showed the relationship between occupation and sexual behaviour in that those with higher incomes were more likely to engage in extra-relational sexual encounters [35]. Working in the manual labour (daily worker) and hospitality industries were associated with higher sexual risk behaviours than working in the sales, retail and professional industries [36]. Since the economic recession was declared in December 2007, the rate of unemployed citizens has continued to rise, leading to a loss of over 7 million jobs. Emerging and young adults may be particularly vulnerable to the effects of unemployment and poverty. It was also found that when divided into regular and irregular workers, group differences emerged between the genders showing that irregular workers potentially have greater sexual appetite and sexual risk [36]. The results highlight important relationships that must be further explored to better understand how emerging and young adults are affected by poverty and unemployment.

2.4.4 Religion

Varas-Diaz et al.(2010) explore the role of religion in HIV stigma manifested by Puerto Rican health professionals in practice and in training [37]. The results of the study was report that religion plays high role in conceptualizations of health and illness among participants. Furthermore, the importance placed on religion and participation in such activities was related to higher levels of HIV stigma. Some studies have found that religious beliefs are related to the idea that infection is a punishment from God, and that PWHA are to blame because they did not follow established moral or religious codes [34].

2.4.5 Place of residence

Most of studies report a patient residing in urban had a higher prevalence than rural residents. From that, according to the EPHIA Ethiopian Population-based HIV Impact Assessment report for period October 2017 to April 2018 that, the annual incidence of HIV among adults ages 15-64 years in urban Ethiopia is 0.06%, which corresponded to approximately 7,000 new cases of HIV annually among adults ages 15-64 years living in urban Ethiopia. Prevalence of HIV among adults ages 15-64 years in urban Ethiopia is 3.0%. This corresponds to approximately 380,000 people living with HIV ages 15-64 years in urban Ethiopia as of April 2018. Prevalence of HIV among children ages 0-14 years in urban Ethiopia is 0.3%, the same among both females and males. The 2011 EDHS reported that adult HIV prevalence was 4.2% in urban areas and 0.6% in rural areas.

2.4.6 Marital status

Marital status is the factor that was found to be associated with increased risk of HIV prevalence. Many authors report that people who married are more effected by HIV ([8]; [38]; [39]). The study entitled Factors Associated with HIV/AIDS in Sudan with the main objective to examine the main factors associated with HIV in Sudan is reported that among the 870 participants single people are 339 (39.0%), married people are 434 (49.9%) and Divorced people are 97 (11.1%) means that people who married are highly effected.The study on the Influence of marital status on HIV infection in south Africa (2000-2017) indicate that, the risk of HIV infection was approximately two times in those who were never married while those who widowed presented as much as twice the risk of HIV infection.

Lakew et al. (2015) was conduct the study to identifies social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia [38]. The result of this study was reported that, Never married 0.4%, currently married 1.4% and formerly married 6.9% therefore, married is more risked for HIV. The study conducted on Seroprevalence of HIV and Associated Factors among Pregnant Women was report that, from a total of 400 pregnant women attending health centers enrolled in the study married women are 378. So, married women were highly affected by HIV [39].

2.4.7 Educational level

Education is a strong factor in improving population health by building in individuals the capacity to process and understand risks related to the HIV/AIDS pandemic. Since HIV/AIDS first emerged globally, the role of behaviour change has been recognized as critical to the control of the pandemic and the sentence "Education is the only vaccine against AIDS" was commonly aired to control it [40]. Adversely, poor information hinders individuals from analyzing their behavioural choices by masking potential health risks [41].

The 2016 Ethiopian Demographic and Health Surveys reported that 76% of the respondents were is with secondary or higher levels of education and the HIV prevalence among these groups were the highest among the respondents. Those with any education are at significantly higher risk of HIV compared to those with no education. Thus a university graduate is half as likely to be infected as someone with no education.

The study conducted by Aregash and Yasmin (2018) in Jimma Zone show that educational status was significantly associated with ART adherence status where majority in the study reported primary or read and write education level.

Kassahun et al. (2018) explore HIV disclosure and associated factors among reproductive age women in health facilities of Jimma Town. The study included explanatory variables to determine the most significant factors that influence HIV disclosure status among HIV reproductive women attending ART services in Jimma town public health institution. From those, education level was partitioned as illiterate, read and write only, Primary School and Secondary school. As the result of this study report, about (40.7%) of respondents had attended primary education [8].

CHAPTER THREE

3 DATA AND METHODOLOGY

3.1 Study Area

Jimma is one of the zones in the Oromia regional state of Ethiopia and is named for the former kingdom of Jimma, which was absorbed into the former province of Kaffa in 1932. The capital town of the zone is Jimma which is the largest city in south-west Ethiopia. The zone has a latitude and longitude of $7^{0}40$'N $36^{0}50$'E/7.66^{0} N 36.833^{0} E and the temperature at Jimma is in a comfortable range, with the daily mean staying from 20 to 25 degree Celsius. Recently the zone ,includes around 22 districts including Jimma city as one distict. Based on the 2007 census conducted by the Central Statistical Agency of Ethiopia, this zone has a total population of 2,486,155 and with an area of 15,568.58 square kilometers. Jimma has a population density of 159.69 per km^2.

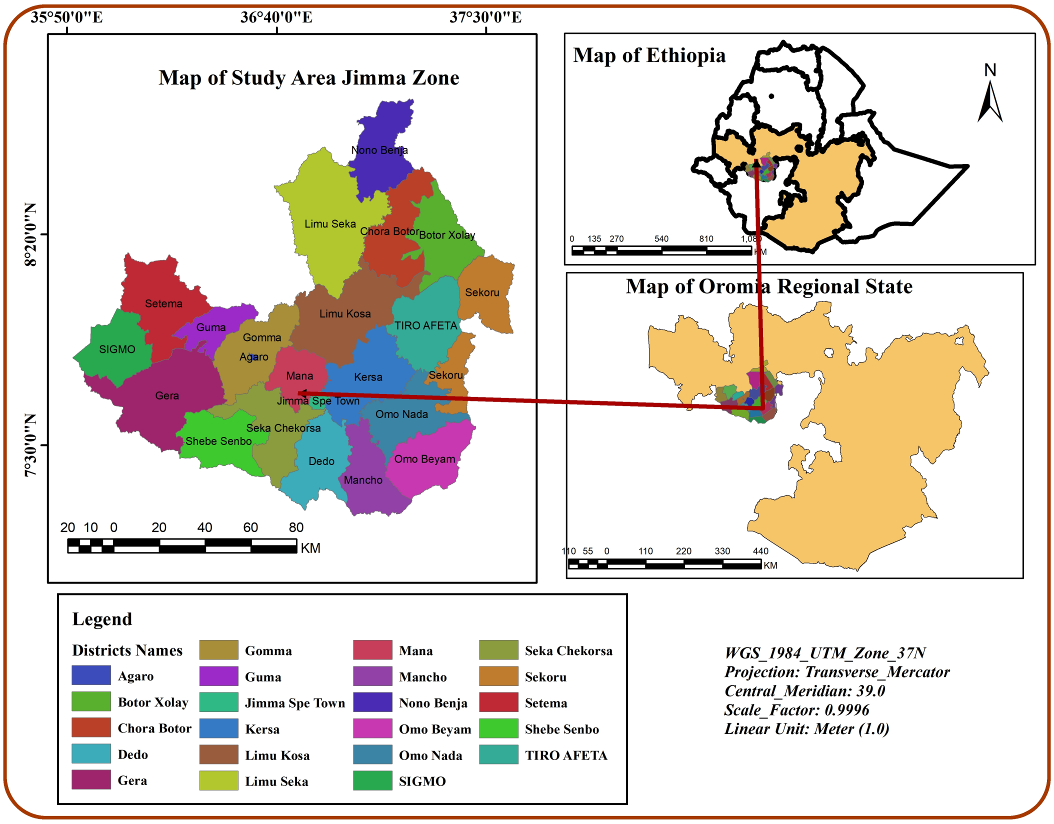

Figure 3.1: Map of Jimma, Oromia, Ethiopia (source: ArcMap 10.5 version)

3.2 Source of Data

The data for this study was secondary data. It was obtained from Jimma Zone districts health center office. All data on any forms of HIV cases and identified covariate's are gathered from Jimma Zone districts health center offices. The data were registered from September 2018 to August 2019, which is recorded for a year.

3.3 Variables under Study

3.3.1 Response Variable

The dependent variable of this study is HIV status (Negative=0 or Positive=1) in each district or woreda of Jimma zone recorded under the health center office .

3.3.2 Explanatory Variables

Based on literature and data available in Jimma Zone districts health center offices, the explanatory variables considered for this study are;

- Gender (Female, Male)
- Age (< 15, 15-19, 20-24, 25-49, > 50)
- Marital status (Single, Married, Divorced and Widowed)
- Educational Level (No education, Primary, Secondary and Superior)
- Condom use (No, Yes)
- Religion (Protestant, Muslim and Orthodox)
- Occupation (No job, Daily worker, Farmer, Merchant and Government employee)
- Place of residence (Rural,Urban) and
- In addition to these variable there are other factors like Additional drugs (Alcohol or Chat), Number of life time sexual partners or Multiple sex, Average monthly income, CD4 count, HIV awareness and behavior and many others factors which affects the HIV distribution. But unable to consider in this study due to the fact that secondary data limitation.

3.4 Statistical Methodology

This study was use spatial analytical techniques which had advantages over standard statistical techniques to identify geographical variations of HIV distribution in Jimma Zone. The exploratory spatial data analysis Global Moran's I, Geary's C and local measures of spatial autocorrelation such as local Moran I and moran scatter plot were used to test for significance of spatial clustering and to know the distribution of events. To identify factors related to HIV distribution, binary or Logistic regression model were used. Chi-square test was used to assess whether there is a significant association between the dependent variable and each explanatory variables.

In this study binary regression model was applied to estimate the effect of covariates on the probability of being HIV positive. A geographical epidemiology is the description of spatial patterns of HIV/AIDS, as part of descriptive epidemiological studies. Once the distribution is identified, another resolution enhancement in spatial data offers an opportunity to model HIV cases, because it also demands on a corresponding enhancement in identifying factors that affect HIV distributions. This is mainly used to identify the relation between HIV Status and explanatory variables. The Generalized linear spatial model under classical approach are used to model spatial data. Where the spatial effects are used to account spatial dependence. The objective of the analysis was to determine if the point pattern of an event is either random or clustered. In order to know the distribution of HIV, exploratory spatial data analysis Moran's I, Geary's C and LISA mainly Moran scatter plot were used.

3.4.1 Spatial Autocorrelation

Spatial auto correlation measures and analyzes the degree of dependency among observations in a geographic space. It needs measuring a spatial weight matrix that reflects the intensity of the geographic relationship between observations in a neighborhood. Spatial auto correlation statistics such as global Moran's I and Geary's C estimate the overall degree of spatial dependence of HIV status overall the entire woreda. Local spatial auto correlation statistics mainly the local Moran's I identified from the Moran scatter plot (Anselin, 1995) is useful in identifying local patterns or hot spots.

To analyze spatial point pattern data relevant software such as GeoDa, ArcGIS and R program can be used for practical modeling and analysis. Spatial autocorrelation analysis is a technique used to detect disease patterns and measures the extent to which the occurrence of an event in a real unit contains or makes more probable to the occurrence of an event in neighboring areal unit. It uses a measure known as spatial autocorrelation coefficient to measure and test how clustered or dispersed points are in space with respect to their attribute values. A positive spatial autocorrelation refers to a map pattern where geographic features of similar value tend to cluster on a map, whereas a negative spatial autocorrelation indicates a map pattern in which geographic units of similar values scatter throughout the map. When no statistically significant spatial autocorrelation exists, the pattern of spatial distribution is considered to be random. The most widely used measures for the proximity of locations and the similarity of the characteristics of these locations are Moran's I and Geary's statistics.

3.5 Method of Testing Spatial Randomness

Testing for complete spatial randomness is the first step in the analysis of spatial point pattern data. Basically, the main question here is: are locations randomly distributed through the study area or do the locations indicate some structure? There are several methods and algorithms that are used to answer the scientific question of spatial randomness or clustering of cases. However, quadrant count and nearest neighbor methods are commonly used to test the spatial randomness or clustering of events.

3.5.1 Quadrant Count Method

The basic idea of quadrant method is to divide the region D into subsets often rectangular shape and then counting the number of events in each of the subsets. The use of quadrant counts can be used to access whether there is any spatial pattern in the data. If clustering is present in the data, then one would expect quadrants with higher counts to be located near each other or if the quadrant counts are spread out over the region, then there is evidence of uniformity. The quadrant count method can be described simply as partitioning of the data set into n equal sized sub regions; these sub regions are called quadrants. In each quadrant the number of events that occur will be counted and it is the distribution of quadrant counts that will serve as an indicator of pattern. The choice of the quadrant size can greatly affect the analysis, where large quadrants produce a coarse description of the pattern. If the quadrant size is too small then many quadrants may contain only one event or they might not contain any events at all.

3.5.2 Nearest Neighbor Method

This method is commonly applied for irregular areal data which are divided into different polygons (woredas in this study). Identification of polygons which are nearest to each other is a primary concern in exploratory spatial data analysis. In this approach, the nearest neighbor distance method is used to define spatial weight matrix that helps to develop the statistical method used in testing randomness. Nearest neighbor method gives a flexible weight matrix that represents spatial dependence based on a decay relationship and the number of neighbors. The measurements of areal units that are nearer to each other tend to be similar. When the measurements are independent, then no spatial pattern is expected. The first stage to implement a spatial pattern analysis is the construction and estimation of the weight matrix, given the spatial arrangement of the observations [42].

3.5.3 Weight Matrix

A general spatial weight matrix can be defined as a symmetric binary contiguity matrix, which can be generated from topological information based on either adjacency or distance criteria. The spatial linkages or proximity of the observations are measured by defining a spatial weight matrix. The spatial weight matrix represents the strength of the potential interaction between locations. However, it has to be noted that the determination of the proper specification for the elements of a spatial weight matrix is one of the difficult and controversial methodological issues in spatial data analysis [**?**]. There are two methods that are used in computing spatial weight matrix, namely the Euclidean distance method and the proximity method. The most common method is to consider two or more regions as neighbors if they share a common border or vertex. According to the adjacency criteria, the element of spatial weight matrix is **1** if location **i** is adjacent to location j and 0, otherwise. The following examples show how weight matrix is constructed from a regular/irregular study area. Neighborhood relations are defined as either Rooks, Bishop's, or Queen's (King's) case.

- **Rook's case** considers contiguity by a neighborhood of four locations adjacent to each cell (locations which share a common border are considered as neighbors).

- **Bishop's case** only considers the diagonals (common edge) to define a neighborhood.

- **Queen's case** considers all neighborhoods (common boundary and or edge) to define neighborhood. This is the most commonly used method.

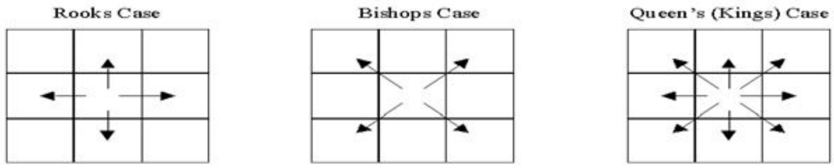

Figure 3.2: Contiguity Case of Representation of Spatial Weight Matrix

3.5.4 Global Measures of Spatial Clustering

This is investigated by use of the local and global spatial auto-correlation using the Geary (**G**) and Moran (**M**) indices respectively. These statistics are computed to test the null hypothesis of no significant clustering of HIV distribution in jimma zone. Moran's I global measures summarize spatial association with respect to the whole Districts. Spatial auto correlation index measures spatial association in the data considering simultaneously both locational and attribute information. The Moran I on the other hand, described the overall spatial dependence of HIV distribution over the study area its values range from -1 to +1 and the general formula for computing Moran's I is given by:

$$I= \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^{n}(y_i-\bar{y})^2}, \qquad [3.1]$$

where n is the number of polygons (woredas in this study) indexed by i and j; y is the variable of interest; $\bar{y}$ is the mean of y; and W_{ij} is the element in the spatial weight matrix. W_{ij} is 1 if the geographic areas associated to y_i and y_j are neighbors and 0 otherwise.

The index has a positive value in case of positive spatial auto correlation, when the pairs of deviations from the mean for contiguous locations having the same sign are prevalent. In contrast, when the pairs of deviations from the mean have prevalent opposite sign the index has a negative value, therefore showing negative spatial auto correlation. The observed value of I can be compared to its distribution under the null hypothesis of no spatial auto correlation or no clustering, i.e when the values of y_i are independent of the values $y_j (i \neq j)$ at neighboring locations. This is equivalent to say that under the reference null distribution, data are randomly distributed over locations. Therefore, inference can be based on the standardized version of I, namely

$$\mathrm{Z}(I)= \frac{I-E(I)}{\sqrt{var(I)}}.$$

The expected value of Moran's I under the null hypothesis of no spatial auto correlation is $\mathrm{E(I)}=\frac{-1}{n-1}$ and Its variance equals $\mathrm{Var}(I)=E(I^2)-(E(I))^2=\frac{n^2(n-1)S_1-n(n-1)S_2-2S_o^2}{(n+1)(n-1)S_o^2}$,

where $S_o=\sum_{i\neq j}^{n} W_{ij}$, $\quad S_1=\sum_{i\neq j}(W_{ij}-W_{ji})^2$, $\quad S_2=\sum_{k=1}^{n}(\sum_{j=1}^{n} W_{kj}-\sum_{i=1}^{n} W_{ik})^2$.

Moran index is used for identification of spatial distribution as dispersion, random or cluster patterns. Indices close to zero indicate the presence of random pattern. Indices close to **+1** indicate a tendency toward clustering [44]. Besides the fact that Moran's I takes the usual form of autocorrelation, its distribution is well studied so that it can be used for testing the significance of spatial autocorrelation in neighboring plots or counties in a study area. A statistically significant estimate of I indicates that neighboring region have a similar prevalence rate and that the cases are likely to cluster at the regional level (woreda in this study) [43].

Geary's C

Geary's C is a measure of spatial autocorrelation or an attempt to determine if adjacent observations of the same phenomenon are correlated. Geary's C interactions are not the cross product of the deviations from the mean, but the deviations in intensities of each observation location with one another. In this study it is useful in identifying local patterns of HIV distribution. Geary's C is given by:

$$\mathrm{C}=\frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(y_i-y_j)^2}{2\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\sum_{i=1}^{n}(y_i-\bar{y})^2}, \qquad [3.2]$$

where the notations are the same as expression (3.1).Usually the values of C range between 0 and 2. Values of C between 1 and 2 indicate presence of negative spatial autocorrelation while values between 0 and 1 indicate presence of positive spatial autocorrelation [45].

Moran's I gives a more global indicator, whereas the Geary's coefficient is more sensitive to differences in small neighborhoods. Testing the significance is done by using the standardized version of C, namely

$$Z(C) = \frac{C-E(C)}{\sqrt{var(C)}},\ E(C) = 1 \quad \text{and} \quad var(c) = \frac{((2s_1+s_2)(n-1)-4s_o^2)}{2(n+1)s_o},$$

where the notations are the same as expression (3.2) and E(C) and var(C) are the expectation and variance of C coefficients respectively. The interpretation of Geary's is analogous to that of Moran's I. The only difference is that when C lies in the interval (1,2) indicates the presence of negative spatial auto correlation (HIV distribution clustering of dissimilar values), if it lies in the interval (0,1) indicates positive spatial auto correlation representing the HIV distribution clustering of similar values. Smaller p-values correspond to stronger autocorrelation for both I and C statistics. Based on the preceding remarks, we have positive spatial autocorrelation when $ZI > 0$ or $ZC < 0$ and we have negative autocorrelation when $ZI < 0$ or $ZC > 0$ [45].

3.5.5 Local Indicators of Spatial Autocorrelation

While the strength of Moran's I lies in its simplicity, its major limitations is that it tends to average local variations in the strength of spatial autocorrelation. The local indicators of spatial auto correlation category of tools examines the local level of spatial autocorrelation in order to identify areas where values of the variable are both extreme and geographically homogeneous. This leads to identification of hot spots, cold spots and clustering of dissimilar values.

Anselin (1995) defines a LISA statistic satisfying the following two conditions:

- The LISA for each observation measures the extent of sign, when positive spatial clustering of similar values around the observation and when negative spatial clustering of dissimilar values around observations.
- The sum of LISAs for all observations is proportional to a corresponding global indicator of spatial autocorrelation

The local value of a LISA is computed as:

$$I_i = \frac{\sum_{j=1}^{n} W_{ij}(z_i - \bar{z})(z_j - \bar{z})}{(z_i - \bar{z})^2}. \qquad [3.3]$$

From the proportionality condition, $\sum_{i=1}^{n} I_i = \gamma I$. Where I_i is the LISA statistic for each observation, γ is a scale factor and I, is a corresponding global spatial autocorrelation measure. The Moran scatter plot is a useful visual tool for exploratory spatial analysis. It enables us to assess how similar an observed value is to its neighboring observations.

The Moran scatter plot provides a visual representation of spatial association (dependence) in the neighborhood around each observation. Depending on their position in the plot, the Moran scatter plot data points express the level of spatial association of each observation with its neighboring ones. The Moran scatter plot can be divided into four quadrants as it is indicated in Figure 3 below: the top right and the bottom left quadrants contain observations showing positive spatial autocorrelation respectively with high-high and low-low data values indicating presence of clusters. The top left quadrant contains low values in a neighborhood of high values (low high), while the bottom right quadrant contains high values in a neighborhood of low values (high low). In both cases they are showing values of dissimilar clustering [45].

Local spatial correlation analysis focuses on the relationship between a spatial unit and its surrounding units. The most widely used tool is the Moran scatter diagram. It uses a two-dimensional graph to express the relationship between a space unit and its surrounding space units. If Z is used to represent a space unit and its surrounding space unit is *W-Z*, then Z and *W-Z* constitute four quadrants as below [47].

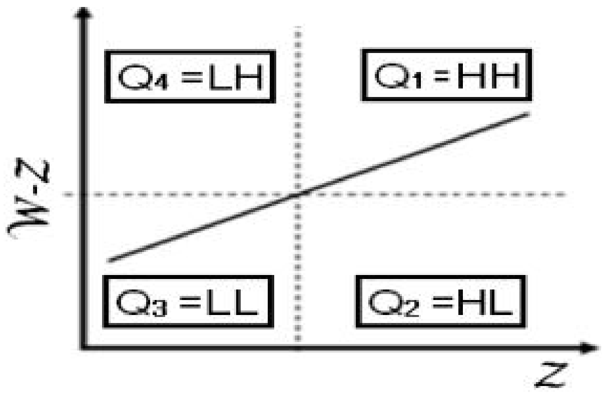

Figure 3.3: Moran Scatter plot (Source: Anselin. L, 1995)

3.6 Generalized Linear Model

3.6.1 Exponential Family Distributions

Consider a single random variable Y whose probability distribution depends on a parameter θ and ϕ. A distribution is belongs to the exponential family of distributions if its probability density function, or probability mass function, can be written as

$$f_{\theta}(y) = exp[\{y^{\theta} - b(\theta)\}/\mathrm{a}(\phi)+c(y, \phi)], \qquad [3.4]$$

where b, a and c are arbitrary functions, ϕ an arbitrary 'scale' parameter, and θ is known as the 'canonical parameter' of the distribution (in the GLM context, θ will completely depend on the model parameters β [48]). It is possible to obtain general expressions for the the mean and variance of exponential family distributions, in terms of a, b and ϕ. The mean of any exponential family random variable, is given by the first derivative of b with respect to θ.

$$\mathrm{E(y)}=\mu=\mathrm{b}^{'}(\theta)$$

and

$$\mathrm{var(y)}=b^{''}(\theta)\mathrm{a}(\phi).$$

a could in principle be any function of ϕ, and when working with GLMs there is no difficulty in handling any form of a, if ϕ is known. For example for Bernoulli distribution a(ϕ)=1, b(θ)=(log[1+exp(θ)]) and c(y;ϕ)=1 and for binomial distribution with n number of trials a(ϕ)=1/n, b(θ)=(log[1+exp(θ)]) and c(y;ϕ)=log[n!/{(ny!)(n-ny)!}], where y is the number of successes in n trial.

3.6.2 Generalized Linear Model

Generalized linear models were formulated by John Nelder and Robert Wedderburn in 1972 as a way of unifying various other statistical models, including linear regression, logistic regression and Poisson regression. Generalized linear models is an extension of the linear modeling process that allows models to be fitted to data that follow probability distributions other than the Normal distribution and allow for a degree of non-linearity in the model structure [49]

GLM has three components these are, Random Component which is refers to the probability distribution of the response variable (Y), Systematic Component which specifies the explanatory variables $(X_1, X_2, ...X_k)$ in the model and Link Function g(μ) which specifies the link between random and systematic components. This model is defined in terms of a set of independent random variables Y_1,... ,Y_N each with a distribution from the exponential family. A Generalized linear model has the basic structure

$$\mathbf{g}(\boldsymbol{\mu}_\mathbf{i}) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p = \mathbf{x}_\mathbf{i}^\mathbf{T} \boldsymbol{\beta}, \qquad [3.5]$$

where $\mu_i = E(y_i)$, g is a smooth monotonic differentiable function called link function, $\mathbf{X_i}$= (1, $x_1, ...x_p)'$ is the $\mathbf{i}^{th}$ row of a model matrix $\mathbf{X}$, and $\boldsymbol{\beta} = (\beta_0, \beta_1....\beta_p)$ is a vector of unknown parameters and it is estimated by maximum likelihood estimation method.

3.6.3 Model Validation in GLM

Generalized linear models does not assume a linear relationship between the dependent variable and the independent variables, but it does assume linear relationship between the transformed response in terms of the link function and the explanatory variables; e.g., for binary logistic regression logit(π) = $\beta_0 + \beta$x. Model checking is important in Generalized linear models. Residuals, particularly Pearson residuals and Deviance residuals are used to asses model diagnostics. The pearson residual has the form

$$p_i = \frac{y_i - \hat{\mu_i}}{\sqrt{var(y_i)}}$$

The result is called the Pearson residual because the square of p_i is the contribution of the i^{th} observation to Pearson's chi-squared statistic. A measure of discrepancy between observed and fitted values is the deviance statistic, and it is given by

$$D = 2\sum y_i log(\frac{y_i}{\hat{\mu_i}}) + (n_i - y_i)log(\frac{n_i - y_i}{n_i - \hat{\mu_i}}). \quad [3.6]$$

where y_i is the observed and $\hat{\mu_i}$ is the fitted value for the i^{th} observation. Alternatively The binomial deviance has the form

$$D = 2\sum o_i \log(\frac{o_i}{e_i}),$$

where o_i denotes observed, e_i denotes expected frequencies (under the model of interest) and the sum is over both "successes" and "failures" for each i frequencies. With grouped data, the distribution of the deviance statistic as the group sizes $n_i \longrightarrow$ inf for all i, converges to a chi-squared distribution with n-p degrees of freedom (d.f) and parameter p. An alternative or other diagnostic tools for measure of goodness of fit is Pearson's chi-squared statistic and can be written as

$$X_P^2 = \sum \frac{n_i(y_i - \hat{\mu_i})^2}{\hat{\mu_i}(n_i - \hat{\mu_i})}. \quad [3.7]$$

Here each term in the sum is the squared difference between observed and fitted values y_i and $\hat{\mu_i}$, divided by the variance of y_i, which is $\mu_i(n_i - \mu_i)/n_i$ estimated using $\hat{\mu_i}$ for μ_i. With grouped data, Pearson's statistic has approximately a chi-squared distribution with n-p d.f for large samples and is asymptotically equivalent to the deviance or likelihood-ratio (LR) chi-squared statistic [48].

The likelihood ratio test is used to test the null hypothesis that any subset of the β's is equal to 0. The LR test is performed by estimating two models and comparing the fit of one model to the fit of the other. Removing predictor variables from a model will almost always make the model fit less well (i.e., a model will have a lower log likelihood), but it is necessary to test whether the observed difference in model fit is statistically significant. The LR test does this by comparing the log likelihoods of the two models, if this difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model.
The formula for the LR test statistic is:

$$LR = -2ln\left(\frac{L(m_1)}{L(m_2)}\right) = 2(loglik(m_2) - loglik(m_1)), \quad [3.8]$$

where $L(m_i)$ i=1,2 denotes the likelihood of the respective model (either Model 1 or Model 2), and $loglik(m_i)$ the natural log of the model's final likelihood (i.e., the log likelihood).

Where m_1 is the more restrictive model, and m_2 is the less restrictive model. The resulting test statistic is distributed chi-squared, with degrees of freedom equal to the number of parameters that are constrained [50]. In GLM the homogeneity of variance does not need to be satisfied. In fact, it is not even possible in many cases given the model structure, and overdispersion (when the observed variance is larger than what the model assumes) may be present [49].

3.6.4 Generalized Linear Spatial Model

A logistic regression model predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables, that can be continuous or categorical. Binary data is modelled by a GLM, with the canonical logit link. This model is known as the logistic regression model and is the most popular for binary data. Let $y_i, i = 1....n$ be a binary dependent variable and x_i be the corresponding p x 1 vector of explanatory variables or covariates. A binary response is represented as a variable taking either 0 or 1 modeled by Bernoulli or binomial distribution with number of trial is equal to 1. For example a response variable with y=0 or y=1, a binary response wether a people is HIV negative and HIV positive. The main goal here is to explain y in terms of explanatory variable contained in x. when π the probability that $y_i = 1$ then y has Bernoulli distribution. The density function of Bernoulli distribution is given by

$$\text{f}(\text{y};\pi)= \pi^y\ (1\text{-}\pi)^{(1-y)}. \qquad [3.9]$$

By taking the logarithm then exponentiating the above equation the density function can be expressed as

$$\text{f}(\text{y};\pi)=\exp\{\log(\pi^y(1\text{-}\pi)^{(1-y)})\}$$

$$=\exp\{\text{y}\log\pi+\ (1\text{-y})\log(1\text{-}\pi)\}$$

$$=\exp[\text{y}\log(\tfrac{\pi}{1-\pi})\ +\log(1\text{-}\pi)].$$

Now let $\theta=\log\ (\frac{\pi}{1-\pi})$, so that $e^\theta= \frac{\pi}{1-\pi}$ and therefore, $1+ e^\theta= \frac{1}{1-\pi}$
$\implies$ $-\log(1+ e^\theta)= \log(1-\pi)$ $\implies$ $b(\theta)= \log(1+ e^\theta)$.

Hence,

$$f(y;\pi)=\exp(y\theta - b(\theta)),$$

which is exponential form with $a(\phi)= \phi=1$ and $c(y, \phi) \equiv 0$. A canonical link is one which when applied to the mean, μ, gives the canonical parameter, so for the Bernoulli it is clearly: $g(\mu)= g(\pi) = \log(\frac{\pi}{1-\pi})$, where $g(\pi)$ a special case of the logit link.

3.6.5 Link Function for Binary Regression

Each distribution that is a member of the exponential family has compatible link functions meant to be used under different situations. The binomial-logit family consisting of the Bernoulli/binomial distributions are used to model discrete or proportional responses. This family can be used to model number of successes out of a number of trials. The links that are commonly used with this family are logit, probit, log-log, complementary log-log. The logit link is equivalent to logistic regression where log-odds are modeled while the probit link is used to model data in terms of normal-based probabilities. The complementary log-log defines a sigmoid curve where the upper part is more stretched out than the logit or probit, and the log-log defines a sigmoid curve where the lower part is more stretched out than the logit or probit. The log link produces estimates of the log risk ratio, the log-complement estimates log health ratio, and the identity link yields estimates of the risk difference [51]. In GLM, the logistic link function is most widely used and the model with this type of link function is called logistic regression model. The link functions and mean function for the Generalized Linear Models under logistic regression are:

$$\mathbf{X}\boldsymbol{\beta} = \boldsymbol{ln}(\tfrac{\mu}{1-\mu}),$$

which implies that $\boldsymbol{\mu}_i$ is the exponential function of independent variables $\mathbf{X}_i$ with parameter $\boldsymbol{\beta}$ and can be expressed as:

$$\mu_i = \frac{\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k}}{1+\exp^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k}} \qquad \text{or} \qquad \boldsymbol{\mu}_i = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1+\exp(\mathbf{x}'\boldsymbol{\beta})}.$$

The left-hand-side is the familiar probability scale, the right-handside is a non-linear function of the predictors. The probability μ_i is assumed to fall between 0 and 1 over a finite range of x values. In this study also the response variable is binary which are HIV status whether Negative or Positive.

3.6.6 Estimation of $\boldsymbol{\beta}$

The method of maximum likelihood (ML) can be used to estimate the parameters in the linear predictor ($\boldsymbol{\mu_i}$). In logistic regression the occurrence of the event y_i given $\mathbf{x}_i$ is modelled via probability as

$$\text{p}(y_i = 1|\mathbf{x}_i) = \frac{e^{x'_i\beta}}{1+e^{x'_i\beta}}, \qquad i = 1, \ldots, n$$

where p x 1 vector $\boldsymbol{\beta}$ is the parameter of interest. The likelihood function for the response vector $\mathbf{y} = (y_1, \ldots y_n)$ can be expressed as

$$\mathbf{L}(\boldsymbol{\beta},\text{y}) = \prod_{i=1}^{n}\left(\frac{e^{x'_i\boldsymbol{\beta}}}{1+e^{x'_i\beta}}\right)$$

and hence the log likelihood function for the response vector $\mathbf{y}$ is given by

$$l = \ln(\mathbf{L}(\boldsymbol{\beta},\text{y})) = \boldsymbol{\beta}' \sum_{i=1}^{n} \mathbf{x}_i - \sum_{i=1}^{n} \mathbf{ln}(1 + e^{x'_i\boldsymbol{\beta}}).$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is obtained by maximizing the above function with respect to $\boldsymbol{\beta}$. Particularly, differentiating l with respect to $\boldsymbol{\beta}$ yields the score estimating equations

$$\frac{\partial \boldsymbol{l}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{1}{1+e^{x'_i\boldsymbol{\beta}}}\, x_i \text{-} \sum_{i=1}^{n} \mathbf{x}_i = \mathbf{0}.$$

In practice this equation is solved numerically by iteration, e.g using the Newton Raphson or Fisher scoring method. The information matrix for parameter $\boldsymbol{\beta}$ is given by

$$I_{\boldsymbol{\beta}} = \text{-E}\left(\frac{\partial^2 l}{\partial\boldsymbol{\beta}\ \partial\boldsymbol{\beta}'}\right) = \sum_{i=1}^{n} \frac{e^{x'_i\boldsymbol{\beta}}}{(1+e^{x'_i\boldsymbol{\beta}})^2}\ \boldsymbol{x_i x'_i}.$$

Recall that either the Newton Raphson or Fisher scoring method can be used to maximize the log likelihood function in above expression. Both methods have the following generic form for the MLE of $\boldsymbol{\beta}$ **:**

$$\hat{\boldsymbol{\beta}}_{k+1} = \hat{\boldsymbol{\beta}}_k + \lambda_k \boldsymbol{I}_{\boldsymbol{\beta}}^{-\mathbf{1}}\left(\frac{\partial\, \boldsymbol{l}}{\partial\, \boldsymbol{\beta}}\right)\Big|_{\boldsymbol{\beta}=\boldsymbol{\beta_k}},$$

where $0 < \lambda_k \leq 1$. For the Newton Raphson method $I_{\boldsymbol{\beta}}$ is negative Hessian matrix whereas for the Fisher scoring method $I_{\boldsymbol{\beta}}$ is the expected Hessian matrix. At the final iteration $\boldsymbol{I}_{\boldsymbol{\beta}}^{-\mathbf{1}}$ gives the asymptotic variance covariance matrix ofthe MLE of $\boldsymbol{\beta}$.

3.6.7 Test for Goodness of Fit

The goodness of fit of statistical model describes how well it fits a set of observations. The logit link function and the binary dependent variable of interest make the logistic regression model distinct from linear regression model. The goal of logistic regression model is to model the probability of the occurrence of an event depending on the value of covariates x. A model is said to fit poorly if the model's residual variation is large and systematic [52]. Influential observations and outliers can also lead to a poor fit.

Many methods on assessing the goodness-of-fit for logistic regression models have been developed [53]. The current methods are based on covariate patterns, which include Pearson's Chi-square X^2 test (Pearson 1900), Deviance, Osius and Rojek's normal approximation test [54]. The residual is based on the deviance or likelihood ratio chi squared statistic. For a measure of goodness of fit, analogous to the residual sum of squares for normal models, two such measures, the generalized Pearson X^2 statistic and the log likelihood-ratio statistic, called the deviance are commonly used statistics in GLMs. For a binary response we count the number of successes for which y_i is the observed count while $n_i p_i$ is the expected count case and failures for which $n_i - y_i$ and $n_i(1 - \hat{p}_i)$ which results in:

$$X^2 = \textstyle\sum_{i=1}^{n} \frac{(y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i(1-\hat{p}_i)}.$$

and for any other model with p<n parameters

$$\text{D=2} \textstyle\sum_{i=1}^{n}\{y_i log(\frac{y_i}{\hat{p_i}}) + (n_i - p_i)log(\frac{n_i - y_i}{n_i - \hat{p_i}})\}.$$

The deviance has a general advantage as a measure of discrepancy in that it is additive for nested sets of models, leading to likelihood-ratio tests. Furthermore, the X^2 approximation is usually quite accurate for the differences of deviances even though it could be inaccurate for the deviances themselves. Another advantage of the deviance over the X^2 is that it leads to the best normalizing residuals [55].

3.7 Methods of Model Selections

AIC (Akaike information criterion) and BIC (Bayesian information criterion, also known as the Schwarz criterion) can be used as methods of model selection.

3.7.1 Akaike Information Criterion (AIC)

The AIC is a measure of fit that can be used to assess models. This measure also uses the Log likelihood, but adds a penalizing term associated with a number of variables. It is well known that by adding variables, one can improve the fit of models. Thus, the AIC tries to balance the goodness of fit versus the inclusion of variables in the model. The AIC is given by AIC = -2lnL + 2p where p is the number of unknown parameters included in the model and lnL is the log Likelihood. The model which has smaller value of AIC is the better model.

3.7.2 Bayesian Information Criterion (BIC)

Similar to AIC, BIC also employs a penalty term associated with the number of parameters p and the sample size n. This measure is also known as the Schwarz Information Criterion and computed as BIC = - 2 lnL + pln n . Again, model which has smaller value of BIC has better fit.

CHAPTER FOUR

4 RESULTS AND DISCUSSION

In this study, generalized linear models are used to see the significance and type of relationship that exists between the dependent and independent variables after distribution pattern of HIV cases are identified using Global Moran's I and Geary's C statistics under both randomization and normality assumption. In addition to that, Moran scatters plots and local Moran's I are used to identify local spatial clustering mainly to identify clustering of high values and low values. The explanatory variables included in this study will expected to have significant effect on the HIV distribution in this chapter. In section 4.1 Descriptive Data Analysis, 4.2 spatial distribution of HIV case by woreda was presented. In section 4.3, Tests of spatial autocorrelation such as global and local measures of spatial autocorrelation were used to identify spatial pattern of HIV. Under global measures of spatial autocorrelation global moran's I and Geary's C, were used. Under local measures of spatial autocorrelation local moran's I and moran scatter plots were used to identify hotspot, cold spot and clustering of dissimilar values. In section 4.4 model based data analysis was presented. In Sections 4.5 and 4.6 results and discussion of Model Diagnosis checking and Model Selection are presented.

4.1 Descriptive Data Analysis

The following Table 4.1 presents the counts and percentages of HIV positive cases in each districts of Jimma-zone by considering Sex and Age. It is indicated that, without considering the effect of sex and ages, Sokoru district had the lowest prevalence, 65 (1.60%) HIV cases, whereas Jimma city recorded the highest cases 535 (13.24%). The overall numbers of male with HIV cases 2134 (53%) were greater than those of females 1907 (47%). The Table 4.1 show that the youngest age group < 15 years of age, had the lowest per cent (3%) of cases and the eldest age group, $>= 50$ years of age had the second lowest per cent of positive cases (14%). Whereas the three middle age groups, 15-19, 20-24 and 25-49 years of age had 24%, 25% and 34% of the total HIV cases, respectively implying that people in the age group 25-49 years of age were highly affected by HIV in Jimma zone. Districts like Limu Seka, Omo Beyam, Limu Kosa, Sokoru, Kersa, Mana, Goma, Sigmo and Guma has no HIV cases in the age group younger than (<15) years of age.

Table 4.1: Summary of positive HIV cases by Age group, Gender and District of Jimma Zone

Districts	**Case by Gender%**		**HIV case by Age(%)**					**Per each**
	Female	**Male**	$<$**15**	**15-19**	**20-24**	**25-49**	$\geq$**50**	**Districts**
Chora Botor	37(37)	63(63)	4(4)	20 (20)	29(29)	35(35)	12(12)	100(2.47)
Botor xolay	24(23)	81(77)	2(2)	14(13)	17(16)	61(58)	11(11)	105(2.59)
Mancho	44(49)	46(51)	3(3)	22(25)	19(21)	38(42)	8(9)	90(2.22)
Dedo	53(66)	27(34)	4(5)	15(19)	15(19)	35(43)	11(14)	80(1.97)
Nono benja	46(61)	29(39)	1(1)	16(21)	12(16)	26(35)	20(27)	75(1.85)
Limu seka	69(46)	81(54)	0	9(6)	44(29)	73(49)	24(16)	150(3.71)
Omo nada	44(52)	40(48)	2(2)	35(42)	19(23)	25(18)	13(15)	84(2.07)
Omo beyam	33(47)	37(53)	0	28(40)	10(14)	15(21)	17(24)	70(1.73)
Limu kosa	177(51)	173(49)	0	105(30)	49(14)	114(33)	82(23)	350(8.66)
Sokoru	36(52)	29(48)	0	12(18)	9(14)	22(34)	22(34)	65(1.60)
Tiro afeta	31(43)	41(57)	1(1)	23(32)	11(15)	22(31)	15(21)	72(1.78)
Kersa	78(52)	72(48)	0	48(32)	17(11)	57(38)	28(19)	150(3.71)
Mana	131(49)	139(51)	0	66(27)	70(26)	94(35)	40(15)	270(6.68)
Goma	185(53)	165(47)	1(0)	86(25)	112(32)	101(29)	50(14)	350(8.66)
Gera	92(46)	108(54)	2(1)	63(31)	56(28)	59(30)	20(10)	200(4.94)
Seka chokorsa	119(48)	131(52)	6(2)	47(9)	75(30)	82(33)	40(16)	250(6.18)
Sigmo	87(54)	73(46)	1(0)	30(19)	60(38)	48(30)	21(13)	160(3.95)
Setama	64(51)	61(49)	5(4)	26(21)	47(38)	35(28)	12(10)	125(3.09)
Shabe sonbo	39(39)	61(61)	7(7)	26(26)	28(28)	31(31)	8(8)	100(2.47)
Guma	107(49)	113(51)	0	45(20)	55(25)	100(45)	20(9)	220(5.44)
Agaro	215(49)	225(51)	14(4)	98(22)	125(28)	150(34)	53(12)	440(10.88)
Jimma city	196(37)	339(63)	73(14)	129(24)	135(25)	143(27)	55(10)	535(13.24)
Total case	1907(47)	2134(53)	126(3)	963(24)	1014(25)	1356(34)	582(14)	4041(100)

The summary of positive HIV cases for Marital status, Educational level,Religion, Place of Residence and Condom Use are presented in Appendex Table 5.2 and 5.3

4.2 Spatial Distribution of HIV AIDS by Districts

The total number of positive HIV cases in Jimma zone was 4041 in the year 2011 Ethiopian calendar. The study data was collected from 22 woredas (districts) of this zone. Figure 4.1 presents the percentages of HIV cases in each districts of Jimma zone. From total HIV cases of 4041 Setema, Shabe Sombo, Kersa and Tiro Afeta woredas are accounted for 3.09, 2.47, 3.71 and 1.78 per cent of HIV cases, respectively. Whereas Sigmo, Dedo, Mancho, Omo Beyam, Omo Nada, Sokoru, Botor Tolay, Chora Botor, Limu Seqa, and Nono Benja woredas accounted for 3.95, 1.97, 2.22, 1.73, 2.07, 1.60, 2.59, 2.47, 3.71 and 1.85 per cent of HIV cases respectively. On the other hand Gera, Guma, Seka Chokorsa, Jimma city, Mana, Goma, Agaro and Limu kosa woredas accounted for 4.94, 5.44, 6.18, 13.23, 6.68, 8.66, 10.88 and 8.66 per cent of HIV cases respectively. In general, a higher HIV cases were observed in the central part of the study area while the eastern part of study area has low HIV cases.

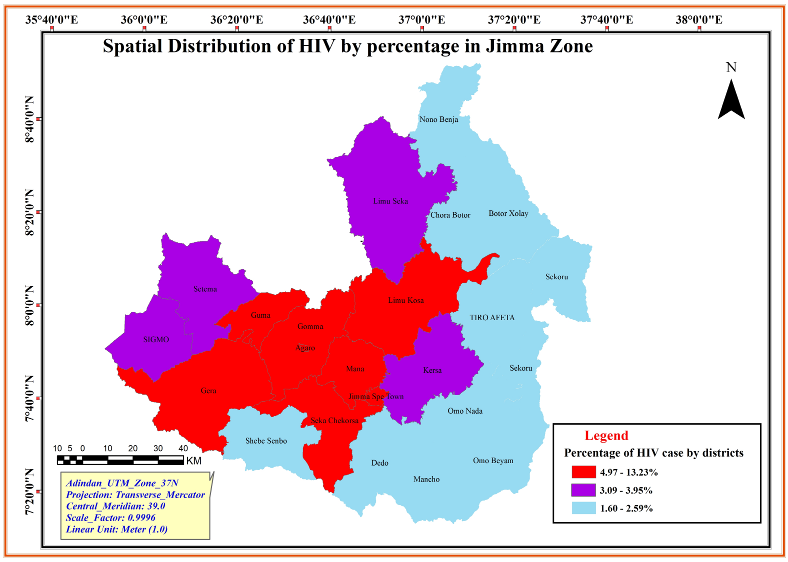

Figure 4.1: Spatial Distribution of HIV Cases among Districts in Jimma Zone from 2018 to 2019

The above figure is showed that, the Spatial Distribution of HIV case by per cent in study area. From this we see that, the red part of the above map is bounded between 4.97 and 13.23 per cent, the blue color is bounded between 3.09 and 3.95 per cent and the Ultra blue color is bounded between 1.60 and 2.65.

4.3 Testing for Spatial Autocorrelation.

In this section, our focus is on the Moran's I and Geary's C coefficients for Spatial Autocorrelation. The contiguity spatial weight matrix, in Jimma Zone is presented in Appendix Table 8. The element of the Contiguity weight matrix was calculated by using Queen's method because Queen's method considers common edge and/or vertex in defining the spatial dependence. Spatial autocrrelation analysis includes tests and visualization of both global test (Moran's I and Geary's C) and local test for clustering (local Moran's I) statistics. The global test is visualized by means of Moran scatter plot, in which the slope of the regression line corresponds to Moran's I. Local analysis is based on the local Moran's I. First, the global Moran's I and Geary's C test statistics were computed to test the null hypothesis $H_o : \rho = 0$ of no significant clustering of HIV case in the entire study area (α=0.05).

4.3.1 Moran's I and Geary's C Tests for Global Spatial Autocorrelation

The main objective of estimating spatial autocorrelation coefficient (global and local) is to measure the strength of spatial autocorrelation amongst neighboring woredas' HIV cases to seek for spatial pattern. The variance of global moran's I and Geary's C vary under both normality and randomization assumptions as shown in Tables 4.2 and 4.3. The null hypothesis states spatial independence (uncorrelated error terms) for the data under consideration and the alternative hypothesis states spatial dependence (correlated error terms). The test results indicate the presence of significant global spatial autocorrelation of HIV in Jimma zone at 5% level of significance, see Tables 4.2 and 4.3. The test results are also shown in Moran's I scatter plot Figure 4.2. These global results in the distribution of HIV need to be further explored using local spatial statistics.

Table 4.2: Results of Global Moran's I and Geary's C Statistics under Normality assumption

Assumption	**Coefficient**	**Observed**	**Expected**	**Dev Std**	**Z**	**P**
Normality	Moran's I	0.45710880	-0.04761905	0.17713	2.8495	0.00219
Normality	Geary's C	0.40473545	1.00000000	0.19190	3.1019	0.0009613

Table 4.3: Results of Global Moran's I and Geary's C Statistics under Randomization assumption

Assumption	**Coefficient**	**Observed**	**Expected**	**Dev Std**	**Z**	**P**
Randomization	Moran's I	0.45710880	-0.04761905	0.17241	2.9275	0.001709
Randomization	Geary's C	0.40473545	1.00000000	0.199375	2.9856	0.001415

Hypothesis test under the global measure of spatial autocorrelation is H_0 : no spatial autocorrelation ($H_0 : \rho = 0$) versus $H_1 : \rho \neq 0$ (There is spatial autocorrelation (spatial dependence). The variance of Moran's I and Geary's C varies under the assumptions of normality and randomization.

Based on the p-values of Moran's I and Geary's C coefficients, we reject the null hypothesis of no spatial autocorrelation at 5% level of significance. Furthermore, the computed Z- statistic for Moran's I is positive under both Normality and Randomization assumptions and Geary's C is negative under both Normality and Randomization assumptions indicating the existence of significant positive spatial autocorrelation (HIV case clustering of similar value).

In order to visualize global spatial autocorrelation, we use Moran's scatter plot under the assumption of normality Figure 4.2. It shows HIV case can be assumed to occur with unequal distribution at all cluster (woredas). The Moran scatter plot is a useful visual tool for exploratory spatial analysis because it enables us to assess how similar an observed value is to its neighboring observations. Moran Scatter Plot for HIV case in Jimma Zone is showed below and the values of Moran's I is 0.457109.

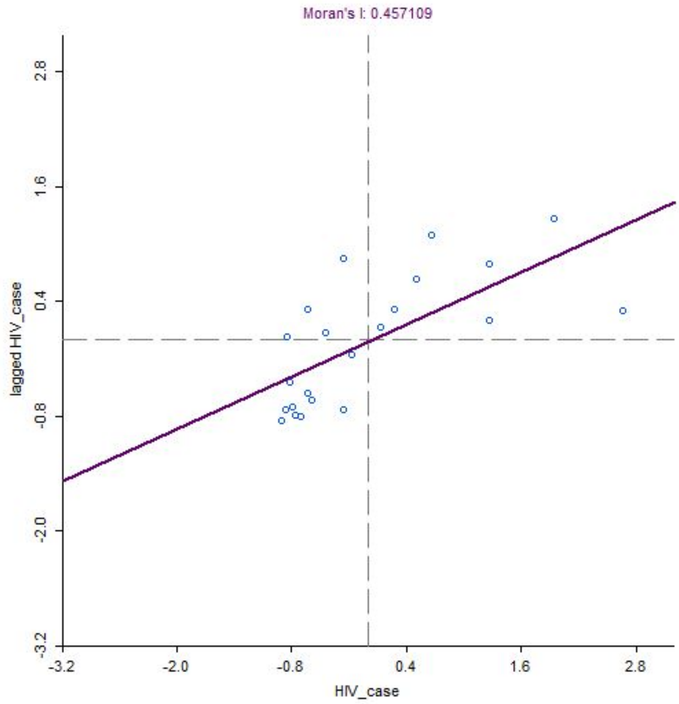

Figure 4.2: Moran Scatter plot for HIV case in Jimma Zone

From Figure 4.2 we conclude that the distribution of HIV cases in Jimma zone in each woreda is spatially correlated with the distribution of HIV cases in the neighboring. From the first quadrant (upper right) of Moran scatter plot we understand that in eight woredas the distribution of HIV cases are highly clustered. This result indicates that in these eight woredas, there is high HIV cases clustering of similar values (hot spots).

From the 3rd quadrant (lower left) we see that the distribution of HIV cases in ten woredas is less clustered. This indicates that in these ten woredas the distribution of HIV cases are cold spots (low low). On the other hand, as it is seen from the fourth quadrant of the Moran scatter plot there is HIV cases clustering of dissimilar values in four woredas (low high value). However, identification of woredas for the presence of significant HIV cases clustering is done based on local measures of spatial autocorrelation and depicted in Figure 4.3 as High-High, Low-High and Low –Low.

Local measures of spatial autocorrelation is used to show the significancy of HIV cases clustering in Jimma Zone and showed as below.

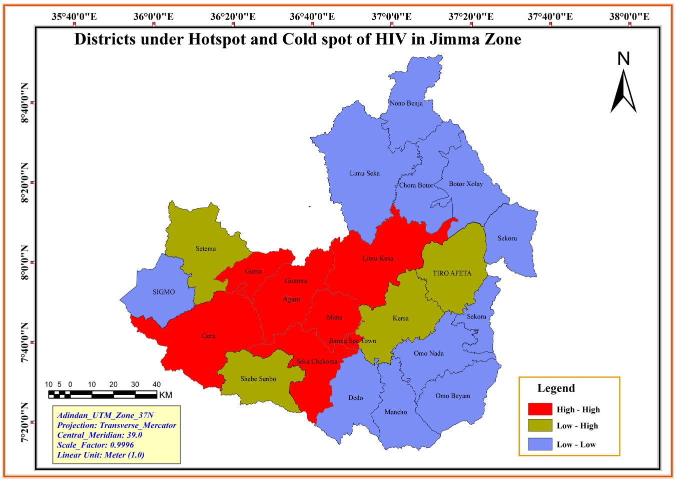

Figure 4.3: Significant HIV Cases Clustering

In Figure 4.3, the red color indicates the presence of hot spots (high-high HIV case clustering) in eight districts as indicated in Moran scatter plot, while the Ultra blue color indicates the presence of cold spots (low-low HIV case clustering) in the ten districts (woredas). The Olivenite green or Avocado color indicates clustering of dissimilar value.

The map reveals that in Agaro town, Gera, Guma, Seka chokorsa, Jimma city, Mana, Goma, and Limu Kosa woredas there is HIV cases clustering of high high values (hotspots). In Sigmo, Dedo, Mancho, Omo Beyam, Omo Nada, Sokoru, Botor Tolay, Chora Botor, Limu Seqa, and Nono Benja woredas there is HIV cases clustering of low-low values (cold spots) is observed. On the other hand in Setema, Shabe Sombo, Kersa and Tiro Afeta woreda there is HIV cases clustering dissimilar values (low-high values) or they have low values, but surrounded by high values. Generally measures of spatial autocorrelations show that the spatial distribution of HIV cases among Woredas in Jimma zone is spatially correlated with neighboring Woredas, where high-high (hot spot) value grouping occurred and another one at the opposite extreme with low-low (cold spot) values occurred. The results of Local Moran's I Test for all District in Jimma Zone are presented in Table 4.4.

Table 4.4: Results of Local Moran's I Test

ID	Woreda	Ii	E.Ii	Var.Ii	Z.Ii	Pr(z > 0)
1	Chora Botor	1.11459089	-0.14285714	2.3883087	0.81366346	0.2079188876
2	Botor Tolay	0.39674210	-0.04761905	0.8704121	0.47629237	0.3169330515
3	Mancho	1.78969751	-0.14285714	2.3883087	1.25050822	0.1055569776
4	Dedo	0.58527599	-0.04761905	0.8704121	0.67837406	0.2487672764
5	Nono Benja	0.76858737	-0.09523810	1.6665150	0.66914677	0.2517009274
6	Limu Seqa	0.39041033	-0.09523810	1.6665150	0.37619877	0.3533845659
7	Omo Nada	1.24552183	-0.09523810	1.6665150	1.03859543	0.1494964671
8	Omo Beyam	1.32455076	-0.09523810	1.6665150	1.09981376	0.1357066374
9	Limu Kosa	1.07505787	-0.19047619	3.0357933	0.72633628	0.2338163213
10	Sokoru	0.79867384	-0.04761905	0.8704121	0.90710640	0.1821752663
11	Tiro Afeta	-0.09390765	-0.14285714	2.3883087	0.03167401	0.4873660127
12	Kersa	-1.13368443	-0.23809524	3.6089687	-0.47143008	0.6813331790
13	Mena	3.72714570	-0.23809524	3.6089687	2.08726708	0.0184319995
14	Goma	6.29273280	-0.28571429	4.1078349	3.24576367	0.0005856802
15	Gera	0.10118783	-0.28571429	4.1078349	0.19089503	0.4243039175
16	Seqa chokorsa	2.00967065	-0.28571429	4.1078349	1.13252823	0.1287062118
17	Sigmo	0.06045205	-0.09523810	1.6665150	0.12060256	0.4520029210
18	Setama	-0.10238188	-0.14285714	2.3883087	0.02619054	0.4895526811
19	Shabe sombo	-0.41668234	-0.09523810	1.6665150	-0.24900097	0.5983199856
20	Guma	0.27126224	-0.14285714	2.3883087	0.26796639	0.3943625948
21	Agaro	2.56875056	-0.04761905	0.8704121	2.80437856	0.0025206837
22	Jimma city	2.51816660	-0.14285714	2.3883087	1.72188251	0.0425454016

Observe from Table 4.4 that Tiro Afeta, Kersa, Setama and Shabe Sombo woredas spatial correlations are negative (Ii < 0). These indicate that in these Woredas high value is surrounded by low values or low value is surrounded by high values of neighboring Woredas. The rest of the Woredas exhibit positive spatial correlations (Ii> 0) indicating grouping of similar values or Low value is surrounded by low values and (high,high) values.

4.4 Model Based Data Analysis

Most of the explanatory variable are categorical and to see the effect of each item on positive HIV case, first we convert all categorical into a factor to indicate that they should be treated as a categorical variable. The independent variable like Gender, place of residence and condom use are dummy variables and the results of fitted model are displayed in Table 4.5.

Table 4.5: Parameter Estimates of Binary Regression using MLE

Coefficients	Categories	Estimate	Std. Error	z value	$Pr(>\|z\|)$
Intercept		-0.85290	0.17645	-4.834	0.000
Age	15-19	1.83469	0.16080	11.410	0.000
	20-24	0.97478	0.15782	6.177	0.000
	25-49	1.21823	0.15470	7.875	0.000
	>50	1.21999	0.17103	7.133	0.000
Gender	Male	0.15520	0.06747	2.300	0.021
Marital status	Married	0.61969	0.09504	6.520	0.000
	Divorced	0.49980	0.10718	4.663	0.000
	Widowed	0.50976	0.12406	4.109	0.000
Education level	Primary	0.44029	0.08169	5.390	0.000
	Secondary	0.22248	0.09616	2.314	0.020
	Superior	0.22020	0.14434	1.526	0.127
Religion	Muslim	0.35337	0.07494	4.716	0.000
	Orthodox	0.26724	0.09625	2.777	0.005
Occupation	Daily worker	0.21711	0.09617	2.257	0.023
	Farmer	0.26482	0.10610	2.496	0.012
	Merchant	0.26543	0.11487	2.311	0.020
	Government employee	0.27193	0.16702	1.628	0.103
Place of Residence	Urban	0.18338	0.06771	2.708	0.006
Condom use	Yes	-4.18179	0.07268	-57.534	0.000

The above output shows the coefficients, standard errors, Z-statistics (wald Z-statistics) and the associated p-values. The last column or $\Pr(> |z|)$ shows the two-tailed p-value testing the null hypothesis that the coefficient is equal to zero (i.e. no significant effect on HIV status). The results shows at 5% level of significance, except the coefficients for superior category of education level and government employee category of occupation all other coefficients are statistically significant because their respective p-values are less than 0.05. Residual deviance for the fitted binomial regression was given as 6089.7 on 8903 df.

An odds ratio (OR) is a statistic that quantifies the strength of the association between two events, which are positive and negative in this study. The odds ratio is defined as the ratio of the odds of positive in the presence of negative and the odds of negative in the absence of positive. Relative risk ratios allow an easier interpretation of the logit coefficients. They are the exponentiated value of the logit coefficients and presented with Confidence interval as Table 4.6.

The indicator variables have a slightly different interpretation. For every one unit of change in Age the log odds of HIV status positive versus Negative increase. From this result as the user of condom is increased the HIV positive patient is decreased. Having grouped age category of 15-19, versus category of <15, changes the log odds of HIV positive by 6.26. The odds of male versus female, changes the log odds of HIV positive by 1.16. From the marital status married terms are highly affected by HIV. Having marital status being Married, versus single, changes the log odds of HIV positive by 1.85. Muslim religion followers patients are highly affected by HIV.Muslim versus protestant, changes the log odds of HIV positive by 1.42. Most of subjects included in the study are take the primary school education. Having Primary Educational level versus no Education, changes the log odds of HIV positive by 1.55.

Table 4.6: Odds Ratio

coefficient	Categories	OR	2.5%	97.5
(Intercept)		0.42617924	0.30130307	0.60194160
Age	15-19	6.26321158	4.57124079	8.58870272
	20-24	2.65059139	1.94412641	3.61038049
	25-49	3.38118420	2.49552328	4.57804166
	>=50	3.38713944	2.42224454	4.73709757
Gender	Male	1.16788873	1.02326983	1.33315455
Marital status	Married	1.85834773	1.54269172	2.23928445
	Divorced	1.64839320	1.33632499	2.03425568
	Midowed	1.66488375	1.30612818	2.12426286
Education level	Primary	1.55315741	1.32341202	1.82302499
	Secondary	1.24917589	1.03476108	1.50861538
	Superior	1.24632207	0.93975269	1.65464013
Religion	Muslim	1.42385748	1.22949145	1.64939602
	Orthodox	1.30635007	1.08217138	1.57824981
Occupation	Daily worker	1.24247610	1.02883273	1.50001502
	Farmer	1.30319526	1.05853309	1.60456049
	Merchant	1.30398999	1.04130776	1.63368999
	Government employee	1.31249320	0.94718538	1.82261098
Residence	urban	1.20126600	1.05198295	1.37183762
Condom use	yes	0.01527115	0.01322281	0.01758271

4.4.1 Bivariate Association between Dependent and Independent variables

The test for an overall effect of predictors variables can be tested using the wald.test function. Means that, categorical independent variables test of association was carried out using the Pearson Chi-Square. The order in which the coefficients are given in the table of coefficients is the same as the order of the terms in the model. This is important because the wald.test function refers to the coefficients by their order in the model.

Table 4.7: Bivariate Association between HIV status and its Determinants .

Independent variables	Chi-squared test	df	P($> X^2$)
Age	144.2	4	0.000
Gender	5.3	1	0.021
Maritalstatus	43.4	3	0.000
Educationlevel	29.4	3	0.000
Religion	23.0	2	0.000
Occupation	8.2	4	0.086
Residence	7.3	1	0.006
Condom use	3310.1	1	0.000

The Wald Chi-square test statistic of 43.4 with 3 degrees of freedom and associated p-value of 2e-09 indicates that the overall effect of Marital status is statistically significant, with p-value of 1e-05 indicating that the overall effect of Religion is statistically significant. Furthermore, the p-value of Age, Gender, Educational level, Religion, Residential area and Condom use are less than 0.05. These show that the overall effects of each of these variables on individual HIV status was significant at 5% level of significance. However, the p-value for type of Occupation was 0.086, hence its overall effect was statistically non-significant at 5% level.

4.5 Model Diagnosis Checking

Diagnostics for regression models are tools that assess a model's compliance to its assumptions and investigate if there is a single observation or group of observations that are not well represented by the model. The default residual for generalized linear model is Pearson residual. Figures 4.4 and 4.5 display plot of Pearson residual versus linear predictors and plots for diagnosis of outliers and influential observations, respectively.

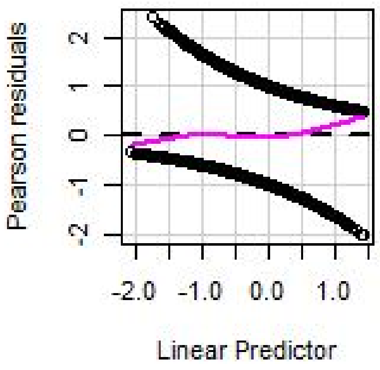

Figure 4.4: Pearson Residual versus Linear Predictors Plot.

4.5.1 Leverage: Hat Values

Observations that are relatively far from the center of the regressor space, taking account of the correlational pattern among the regressors, have potentially greater influence on the least-squares regression coefficients; such points are said to have high leverage. The most common measures of leverage are the h_i, or hat-values. The hatvalues function works for both linear models and GLMs. One way of examining the hat-values and other individual-observation diagnostic statistics is to construct index plots, graphing the statistics against the corresponding observation indices. Figure 4.5 hat-values (on bottom) plot shows the observations 87 and 8370 are stand out from the rest observations in the plot of hat-values, which indicating that their regressor values are unusual relative to the other observations.

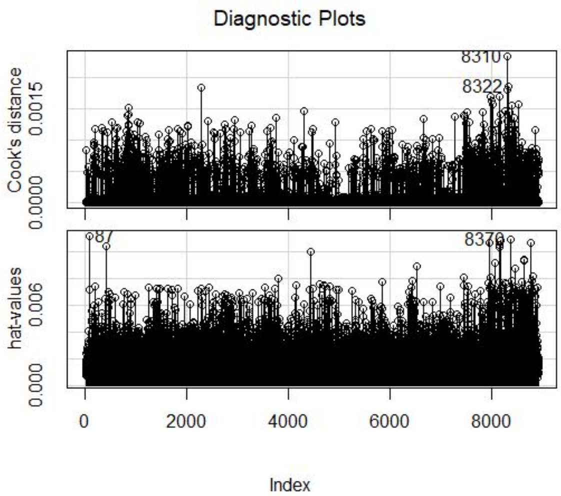

Figure 4.5: Cook's Distance (on top) and Hat-Values (on bottom) Plots.

Observing figure 4.5 cook's distance which is a measure of the influence of each observation on the regression coefficients. The Cook's distance statistic is a measure, for each observation in turn, of the extent of change in model estimates when that particular observation is omitted. Any observation for which the Cook's distance is close to 1 or more, or that is substantially larger than other Cook's distances (highly influential data points). Observations 8310 and 8322 has larger Cook's distance than other data points in Cook's distance plot(on top) plot. Which is the most influential according to Cook's distance.

4.6 Method of Model Selection

To compare the model,the best way for classical approach is AIC. The AIC is a measure of fit that can be used to assess models. It is well known that by adding variables, one can improve the fit of models. Smaller values of AIC indicate the best models to be selected. Two models were considered in this thesis to illustrate model selection, the first model contains only socio-demographic factors (called Model 1) and the second model contains all socio-demographic factors and behavioural factors (Model 2).

Model 1: HIV status ~ Age + Gender + Marital status + Education level + Religion + Occupation + Residence

Model 2: HIV status ~ Age + Gender + Marital status + Education level + Religion + Occupation + Residence + condom use

---

The two models were fitted using the method that was discussed in the previous sections. Their AIC, BIC and Log Likelihood Ratio test values are displayed in Table 4.7.

Table 4.8: AIC, BIC and log likelihood ratio test statistics values of Model 1 and 2

Candidate model	**Methods of comparisons**		
	AIC	**BIC**	**Log Likelihood Ratio test**
Model:1	11760.146	11894.977	-5,861.073
Model:2	6129.666	6271.594	-3,044.833

Recall that the high value of log-likelihood statistic and lower values of AIC and BIC suggest the model is better fit the data. The second model has low value of AIC and BIC and highvalue of log-likelihood, therefore the second model is fit good.

4.6.1 Accounting Spatial Dependence to the Model

The above fitted model does not take in to account a spatial dependence between observations and used to fit fixed effects only. However, from explanatory spatial data analyses we have observed that there is a positive spatial autocorrelation in the distribution of HIV cases among Woredas in Jimma zone. Bayesx function is used to include spatial dependency. Fitting logistic regression model by ignoring spatial dependence under estimate the standard errors of regression coefficients due to this reason to account spatial dependence we extend application of GLM to Bayesian GLM by adding random effect and correlation structure with default prior information. In Bayesian GLM the researcher used Restricted Maximum likelihood method of estimation.

Under spatial dependency two regression model was fitted. These are, Structured and Unstructured Spatial Effects of Regression Model. With regard to spatial effects, the geographical pattern of the Jimma Zone is presented in figure 5.1 and 5.2 (Appendex), which depicts maps of Jimma Zone for HIV cases showing structured spatial effects and unstructured spatial effect on the HIV cases in Jimma Zone respectively. Obviously there exists a district-specific geographical variation of Jimma Zone.

A spatially structured effect is typically included in a spatial model as a spatially structured error term, i.e. in order to account for any spatial autocorrelation unexplained by covariates in the model. Unstructured Spatial Effects of Regression Model is include only determinants of HIV distribution where as, the structured Spatial Effects of Regression Model is include all determinants of HIV distribution and random effects considering the boundary of all Districts in Jimma Zone. Under the Parameter Estimation of Binary Regression, using Maximum Likelihood Estimation most of explanatory variable has positive relationship between dependent Variable.

But based on structured Spatial Effects of Regression Model (presented below in Table 4.9) is indicating the presence of positive relationship between dependent and some explanatory variables.

Table 4.9: Parameter Estimation by Accounting Spatial Dependence using RMLE

coefficients	Estimate	Std. Error	t value	Pr(> $\|t\|$)	95% Confidence Interval
(Intercept)	-0.6960	0.0808	-8.6154	0.000	(-0.8544 -0.5376)
Age15-19	-0.9487	0.1194	-7.9484	0.000	(-1.1826 -0.7147)
Age20-24	0.7791	0.0806	9.6630	0.000	(0.6210 0.9371)
Age25-49	-0.0995	0.0676	-1.4725	0.140	(-0.2319 0.0329)
Age >50	0.1441	0.0628	2.2951	0.021	(0.0210 0.2672)
Gender1	-0.0856	0.0339	-2.5237	0.011	(-0.1521 -0.0191)
Maritalstatus1	-0.4335	0.0698	-6.2088	0.000	(-0.5703 -0.2966)
Maritalstatus2	0.2235	0.0542	4.1273	0.000	(0.1173 0.3296)
Maritalstatus3	0.1011	0.0627	1.6112	0.107	(-0.0219 0.2240)
Educationlevel1	-0.2274	0.0622	-3.6551	0.0003	(-0.3493 -0.1054)
Educationlevel2	0.2149	0.0565	3.8018	0.0001	(0.1040 0.3256)
Educationlevel3	-0.0062	0.0661	-0.0932	0.925	(-0.1357 0.1234)
Religion1	-0.2057	0.0486	-4.2302	0.000	(-0.3009 -0.1103)
Religion2	0.1521	0.0467	3.2585	0.001	(0.0605 0.2435)
Occupation1	-0.1924	0.0753	-2.5560	0.010	(-0.3400 -0.0448)
Occupation2	0.0272	0.0599	0.4537	0.650	(-0.0902 0.1446)
Occupation3	0.0497	0.0691	0.7201	0.471	(-0.0856 0.1851)
Occupation4	0.0607	0.0764	0.7948	0.426	(-0.1585 -0.0253)
Residence1	-0.0920	0.0340	-2.7066	0.006	(-0.1585 -0.0253)
usecondom1	2.1156	0.0369	57.2980	0.000	(2.0432 2.1880)

An alternative method for analyzing the relationship between predictor variables and response variable is by accounting spatial dependence to binary regression model. The table is indicating the presence of positive relationship between dependent and some explanatory variables. The results shows at 5% level of significance, except the coefficients for

Age category of 25-49, widowed marital status, superior education level and farmer, merchant, government employee category of occupation all other coefficients are statistically significant because their respective p-values are less than 0.05. There is difference between parameter estimation with accounting spatial dependence and without accounting spatial dependency. Without accounting spatial dependence, the results shows at 5% level of significance, only the coefficients for superior category of education level and government employee was insignificant. But in accounting spatial dependency; Age category of 25-49 all occupation category expect daily worker was statistically insignificant because of their respective p-values are higher than 0.05.

4.7 Discussion

The descriptive results of the study indicated that the total number of HIV/AIDS cases at study area was 4041. The study indicated that the number of males with HIV positive status (52.8%) was greater than the number of females with the same status. These results were in line with world health organization (WHO) report of 2017 quantified that the number of male HIV infected is greater than that of the female worldwide. Similarly, in research conducted among tissue donors in the United States, the prevalence of HIV infection was estimated to be 40,000 cases per year, with high cases in males (70%) and low cases in females(30%).

From the category of marital status, married people with HIV case (40.9%) was higher than the other marital status. These result is supported and related to the study conducted by Lakew et al. (2015) and Yibeltal (2018) which indicated that category of married has highly iffected than those who were never married, divorced and those who widowed and the study explored by [8] is also report the majority (49%) of the respondents were married ([38], [39] and[8]).

The results of this study also show that the number of HIV positive cases more in Urban area (57.8%) than rural area. This result was in line with the study conducted by Ethiopian Population-based HIV Impact Assessment (EPHIA, 2018) which reported that the prevalence of HIV in urban Ethiopia is higher than that of rural.

The number of people with primary Educational level (48.3%) is high rather than other education level.This result is in line with the Kassahun et al. (2018) studies which reports primary or read and write education level is high. About (40.7%) of respondents had attended primary education [8]. In this study patient residing in urban had a higher prevalence than rural residents (57.8%). The results is supported by Ethiopian Population-based HIV Impact Assessment report for period October 2017 to April 2018 that, the annual incidence of HIV in urban Ethiopia is high [23].

Based on the p-values of Moran's I and Geary's C coefficients, we reject the null hypothesis of no spatial autocorrelation and accept alternative hypothesis of there is spatial autocorrelation at 5% level of significance. Furthermore, the computed Z- statistic for Moran's I is positive under both Normality and Randomization assumption is indicating the existence of significant positive spatial autocorrelation (HIV cases clustering of similar values). As the identification of local patterns of spatial association is an important concern, the result of local measure of spatial auto correlation in the study was showed that HIV cases in Jimma zone of each districts (cluster) is spatially correlated with neighboring districts (clusters).

The result of local moran'I shows positive spatial correlation in eighteen (18) districts which means their observed values are greater than zero and negative spatial autocorrelation in four(4) districts (woredas) because their observed values are less than zero. From the local indicators of spatial autocorrelation, the moran scatter plot was divided the districts into four quadrants.

The first quadrant (upper right) of Moran scatter plot is showed that the distribution of HIV cases are highly clustered in Eight(8) woredas (clusters). Those woredas are Gera, Guma, Seka Chokorsa, Jimma city, Mana, Goma, Agaro and Limu kosa. This result indicates that in these Eight woredas, there are high HIV cases clustering of similar values (hot spots). In these quadrants all districts have high HIV cases and surrounded by woredas which have high HIV cases(high high).

Quadrant two (lower right) is classified for woredas which have high values of HIV cases and surrounded by woredas with low HIV cases but there is no woreda included in this quadrant. The 3rd quadrant (lower left) of Moran scatter plot showed that the distribution of HIV cases in Ten woredas are less clustered and those woredas are Sigmo, Dedo, Mancho, Omo Beyam, Omo Nada, Sokoru, Botor Tolay, Chora Botor, Limu Seqa, and Nono Benja. In these woredas the HIV cases clustering was low-low values (cold spots). On the other hand quadrant four (upper left) of moran scatter plot showed that the distribution of HIV cases in four woredas of Jimma zone were clustering of dissimilar values and those woredas are Setema, Shabe Sombo, Kersa and Tiro Afeta woredas. Those woredas have low values of HIV cases and surrounded by woredas which have high values of HIV cases.

Generally measures of spatial autocorrelations shows distribution of HIV cases in Jimma zone in each woreda(cluster) is spatially correlated with neighboring woreda (cluster), where high-high (hot spot) value grouping occurred and another one at the opposite extreme with low-low (cold spot) values occurred. The Previous studies are also consistent with this result that HIV is spatially correlated with neighboring districts (cluster),where high (hot spot) value grouping occurred and another one at the opposite extreme with low (cold spot) values grouping occurred[19]; [17]; [20] and [21].

The study also considered the model based data analysis, i.e. generalized linear model (GLM). Fitting binary regression model, all terms for Age, Gender, Marital status, Religion, Residence and condom use are statistically significantly affect the HIV distribution. By extending application of generalized linear model to Bayesian generalized linear model spatial dependence is accounted using function bayesx. Previous studies are also consistent with this result that HIV distribution is statistically significant associated with Age, Educational level, marital status and place of residence ([20],[38], [40] and [39]).

CHAPTER FIVE

5 CONCLUSIONS AND RECOMMENDATIONS

This study describes the spatial pattern of HIV/AIDS in Jimma Zone, southwest Ethiopia using the documented data and It is a cross sectional study where the detailed information on HIV status with associated covariates were collected at health care facilities.

5.1 Conclusions

The results of this study showed that HIV cases in Jimma Zone is exhibits a spatial pattern which is dependent on socio- demographic and some Behavioral factors. HIV case in the study area is significantly clustered with high levels in the central part of the zone and in the Northern,eastern and half part of southern study area have low levels of HIV cases. clustering of dissimilar value is occurred in western part of this zone. Geographical clusters of HIV cases were identified through exploratory spatial data analysis, using Global Moran's I, Geary's C and also local indicators. The results obtained reveal that the distribution of HIV case in Jimma Zone is clustered.

The Moran scatter plot also depicts the presence of strong local spatial clustering of high values in Eight(8) woredas (hots spot). Those woredas are Gera, Guma, Seka Chokorsa, Jimma city, Mana, Goma, Agaro and Limu kosa woredas. Significant HIV case clustering of low values were observed in ten woredas those are Sigmo, Dedo, Mancho, Omo Beyam, Omo Nada, Sokoru, Botor Tolay, Chora Botor, Limu Seqa, and Nono Benja woredas from among the woredas shown in Moran scatter plot.

On the other hand, in Setema, Shabe Sombo, Kersa woreda and Tiro Afeta woredas there were clustering of disimilar HIV status or they has low values, but surrounded by high values. Based on binary regression model with spatial dependence Age, Marital status,religion,education level and residence are effects to increase HIV distribution.

5.2 Recommendation

Based on the findings of this study, the researcher recommended the following points.

- Based on the results obtained, the study recommends that interventions should be facilitated in highly clustered HIV distribution areas by giving special attention in targeting intervention and health services to the highly risk exposed districts and neighboring districts.

- This thesis was limited to few variables recorded at the health office. Thus, it will be worthy to extend this study clinical diagnostic related variables.

- Add new health facilities and strengthen old health facilities especially in districts identified as hot spots and neighboring districts .

- The government and non-governmental Organizations should consider these results when planning HIV control measures and more education to increase people awareness, about the HIV distribution.

5.3 Future Works

In this study the researcher found spatial distribution of HIV in Jimma zone and model HIV distribution using few available Socio-demographic Characteristics variables in classical approach. Other researchers should extend further on spatial distribution and modeling HIV status by using large dataset.

References

[1] Tesfaye, T., Darega, J., Belachew, T., & Abera, A. (2018). HIV positive sero-status disclosure and its determinants among people living with HIV/AIDS following ART clinic in Jimma University Specialized Hospital, Southwest Ethiopia: a facility-based cross-sectional study. Archives of Public Health, 76(1), 1.

[2] Marx, J. (1982). New disease baffles medical community. Science, 217(4560), 618-621.

[3] Centers for Disease Control and Prevention (CDC. (2001). HIV and AIDS–United States, 1981-2000. MMWR. Morbidity and mortality weekly report, 50(21), 430.

[4] Global HIV/AIDS Epidemic. Fact sheet. 2017. www.unaids.org/en/resources/documents/2017/2017 data book.

[5] Lester, F., Ayehunie, S., & Zewdie, D. (1988). Acquired immunodeficiency syndrome: seven cases in an Addis Ababa hospital. Ethiopian medical journal, 26(3), 139.

[6] Federal, H. I. V. Control Office 2018: HIV Prevention in Ethiopia National Road Map 2018–2020.

[7] Adal, M. (2019). Systematic review on HIV situation in Addis Ababa, Ethiopia. BMC public health, 19(1), 1544.

[8] Kassahun G, Tenaw Z, Belachew T, Sinaga M (2018) Determinants and Status of HIV Disclosure among Reproductive Age Women on Antiretroviral herapy at Three Health Facilities in Jimma Town, Ethiopia, 2017. Health Sci J. Vol. 12 No. 2: 558.

[9] Kibret, G., Ferede, A., Leshargie, C., Wagnew, F., Ketema, D., & Alebel, A. (2019). Trends and spatial distributions of HIV prevalence in Ethiopia. Infectious diseases of poverty, 8(1), 90.

[10] Asfaw Mesfin, Temesgen Kabeta and Daniel Emana (2017) Diagnosis of Early Infant HIV Infection among Sero-Positive Mother in Jimma Zone, Southern West Ethiopia, Jimma

[11] UNAIDS.(2006). Overview of the global aids epidemic: Report on the global aids epidemic: washington: Usa. Tech. rep.,

[12] Haji, O. (2008). Tuberculosis and HIV Coinfection in Two Districts in Somaliland. Tampere University Press.

[13] Unaids, J. (2017). Fact sheet—latest global and regional statistics on the status of the AIDS epidemic. Geneva: UNAIDS.

[14] Demographic, C. (2016). Health Survey-2011. Central Statistical Agency Addis Ababa. Ethiopia ICF International Calverton, Maryland, USA. 2012.

[15] Ateka, G. (2001). Factors in HIV/AIDS transmission in sub-Saharan Africa.(Perspectives). Bulletin of the World Health Organization, 79(12), 1168-1169.

[16] Su, L., Zhang, Y., von Briesen, H., Xing, H., Köstler, J., & Wagner, R. (2000). Characterization of a virtually full-length human immunodeficiency virus type 1 genome of a prevalent intersubtype (C/B9) recombinant strain in China. Journal of virology, 74(23), 11367-11376.

[17] Heimer, R., Barbour, R., Shaboltas, A., Hoffman, I., & Kozlov, A.(2008). Spatial distribution of HIV prevalence and incidence among injection drugs users in St Petersburg: implications for HIV transmission. AIDS (London, England), 22(1), 123.

[18] Tanser , F., Bärnighausen, T., Cooke, G. S., & Newell, M. L. (2009). Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. International journal of epidemiology, 38(4), 1008-1016.

[19] Khademi, N., Reshadat, S., Zanganeh, A., Saeidi, S., Ghasemi, S., & Zakiei, A. (2016). Identifying HIV distribution pattern based on clustering test using GIS software, Kermanshah, Iran. HIV & AIDS Review, 15(4), 147-152.

[20] Daw, M., Daw, A., Sifennasr, N. , Draha, A., Daw, A., Daw, A. & Adam, S. (2019). Spatiotemporal analysis and epidemiological characterization of the human immunodeficiency virus (HIV) in Libya within a twenty five year period: 1993–2017. AIDS research and therapy, 16(1), 14.

[21] Zulu, L., Kalipeni, E., & Johannes, E. (2014). Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010. BMC infectious diseases, 14(1), 285.

[22] Tsai, P. , Lin, M., Chu, C., & Perng, C. (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. BMC Public Health, 9(1), 464.

[23] Ethiopian Federal Ministry of Health. (2015). Health Sector Transformation Plan (2015/16-2019/20), 20(May), 1–118.

[24] Word Health Organization 2005

[25] Niens, L., & Lowery, D. (2009). Gendered Epidemiology: Sexual Equality and the Prevalence of HIV/AIDS in Sub-Saharan Africa. Social Science Quarterly, 90(5), 1134-1144.

[26] Word Health Organization 2017

[27] Abebe Weldsilase,Y., Likka, M., Wakayo, T., & Gerbaba, M. (2018). Health-related quality of life and associated factors among women on antiretroviral therapy in health facilities of Jimma Town, Southwest Ethiopia. Advances in Public Health, 2018.

[28] FMOH, 2016. HIV epidemic estimates by regional states and Ethiopia. FMOH(Federal Ministry of Health), HIV/AIDS Prevention and Control Office

[29] Hargreaves, J., & Glynn, J.(2002). Educational attainment and HIV-1 infection in developing countries: a systematic review. Tropical Medicine & International Health, 7(6), 489-498.

[30] Unaids, J. (2017). Fact sheet—latest global and regional statistics on the status of the AIDS epidemic. Geneva: UNAIDS.

[31] Tarantola, D., & Schwartländer, B. (1997). HIV/AIDS epidemics in sub-Saharan Africa: dynamism, diversity and discrete declines?. AIDS (London, England), 11, S5.

[32] Sreenivasulu, T., & Jahnavi, K. (2018). A cross sectional study on factors affecting treatment outcome among TB patients. International Journal of Advances in Medicine, 5(1), 175.

[33] Thupayagale-Tshweneagae, G., & Mgutshini, T. (2015). Disclosure of HIV status to sexual partners by people living with HIV.

[34] Zou, J., Yamanaka, Y., John, M., Watt, M., Ostermann, J., & Thielman, N. (2009). Religion and HIV in Tanzania: influence of religious beliefs on HIV stigma, disclosure, and treatment attitudes. BMC public health, 9(1), 75.

[35] Munoz-Laboy, M., Severson, N., & Bannan, S. (2014). Occupations, social vulnerability and HIV/STI risk: The case of bisexual Latino men in the New York City metropolitan area. Global public health, 9(10), 1167-1183.

[36] Mark, K., Janssen, E., & Milhausen, R. (2011). Infidelity in heterosexual couples: Demographic, interpersonal, and personality-related predictors of extradyadic sex. Archives of sexual behavior, 40(5), 971-982.

[37] Varas-Díaz, N., Neilands, T., Malavé Rivera, S., & Betancourt, E. (2010). Religion and HIV/AIDS stigma: Implications for health professionals in Pu

[38] Lakew, Y., Benedict, S., & Haile, D. (2015). Social determinants of HIV infection, hotspot areas and subpopulation groups in Ethiopia: evidence from the National Demographic and Health Survey in 2011. BMJ open, 5(11), e008669.

[39] Yibeltal Getaneh. (2018). Seroprevalence of HBV, HIV and Associated Factors among Pregnant Women Attending Selected Health Centers at Yeka Sub City, Addis Ababa, Ethiopia

[40] Liskin L, Church CA, Piotrow PT, Harris JA. AIDS Education: a Beginning. (1989). Baltimore Centre for Communication Programs, Johns Hopkins School of Public Health, W.DC, USA

[41] Population Research Institute Pop.psu.edu. 2017-08-10.Retrieved 2017-10-04.

[42] Anselin, L. and Hudak, J. (1998). Spatial econometrics in practice and reviews of soft ware options. Regional Science and Urban Economics 22, 509-536.

[43] Osei, P. (2010). Statistical Methods for Disease Mapping (Doctoral dissertation, University of Plymouth, UK).

[44] Cliff, A.D. and Ord, J.K. (1973) Spatial Autocorrelation, Pion: London

[45] Anselin, L. (1995). Local indicators of spatial association - LISA. Geographical Analysis 27 (2), 93–115.

[46] Anselin, L. (1996). "The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association." In Spatial Analytical Perspectives on GIS, edited by M. Fischer, H. Scholten, and D. Unwin, 111–125. London: Taylor & Francis.

[47] yu Chen et al.(2020) The Spatial Spillover Effect in Hi-Tech Industries: Empirical Evidence from China

[48] Dobson, Annette J., (1945) An introduction to generalized linear models; Second edition p. cm.—(Chapman & Hall/CRC texts in statistical science series)

[49] Nelder, J. & Wedderburn, R. (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), 370-384.

[50] McCullagh, P., & Nelder, J. (1989). Generalized linear models., 2nd edn.(Chapman and Hall: London). Standard book on generalized linear models.

[51] Deroche, C. B.(2015). Diagnostics and Model Selection for Generalized Linear Models and Generalized Estimating Equations. (Doctoral dissertation). Retrieved from https://scholarcommons.sc.edu/etd/3059

[52] Hosmer, D., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. Statistics in medicine, 16(9), 965-980.

[53] Evans, S., & Li, (2005). A comparison of goodness of fit tests for the logistic GEE model. Statistics in medicine, 24(8), 1245-1261.

[54] Osius, G., & Rojek, D. (1992). Normal goodness-of-fit tests for multinomial models with large degrees of freedom. Journal of the American Statistical Association, 87(420), 1145-1152.

[55] Pierce, D.A. and Schafer, D.W. (1986). Residuals in Generalized Linear Models.Journal of the American Statistical Association, 81, 977-986.

[56] CSA, Calverton M, USA. (2016). Ethiopian demographic and health surveys. Ethiopia and Calverton, Maryland, USA: Central Statistics Agency and ORC Macro.

[57] Porto, A. (1996). Use of spatial statistics to identify and test significance in geographic disease patterns. Preventive Veterinary Medicine 11, 205-224.

[58] Cressie, N. (1993). Statistics for Spatial Data. New York: John Willey & Sons. Inc.[Google Scholar].

[59] Anselin, L. and Hudak, J. (1998). Spatial econometrics in practice and reviews of soft ware options. Regional Science and Urban Economics 22, 509-536.

[60] Bradley, H., Bedada, A., Brahmbhatt, H., Kidanu, A., Gillespie, D., & Tsui, A. (2007). Educational attainment and HIV status among Ethiopian voluntary counseling and testing clients. AIDS and Behavior, 11(5), 736-742.

[61] Collett, D. (1991). Models for binary and binomial data. In Modelling binary data (pp. 43-91). Springer US.

[62] HAPCO. HIV prevention in Ethiopia National Road map; 2018.https:// ethiopia.unfpa.org/en/.../hiv-prevention-ethiopia-national-road-map. Accessed

[63] Kulldorff, M. (1997). A spatial scan statistic. Communications in Statistics-Theory and methods, 26(6), 1481-1496.

[64] UNAIDS. (2016). Fact sheet-latest statistics on the status of the AIDS epidemic; 2017. https://www.unaids.org/en/resources/fact-sheet. Accessed 4 Aug 2019

Appendex

Geo-statistics Geostatistics is a class of statistics used to analyze and predict the values associated with spatial or spatiotemporal phenomena. It incorporates the spatial coordinates of the data within the analyses and is a spatial process indexed over a continuous space. It is began with mining type applications and the prefix "geo" indicates that geostatistics dealt initially with statistics pertaining to the earth.

Lattice Data are spatial data indexed over a lattice data. Lattice type data provide the closest analogue to time series data. In time series data sets, observations are typically obtained at equally spaced time points.

Point Patterns Pertains to the location of events in the area of interest [57]. Consider a region D and in this region we are interested in locations of certain "events." We can understand whether events of interest are occurring randomly throughout the region, or if the events tend to cluster together. Data analysis of point patterns corresponds to studies where the interest lies in where events of interest occur.

Spatial association Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable. Finding the degree of spatial association (correlation) among data representing related location is fundamental to the statistical analysis of dependence and heterogeneity in spatial patterns. Analysis of the distribution patterns of two phenomena is done by map overlay. If the distributions are similar, then the spatial association is strong, and vice versa.

Spatial Dependence A basic property of spatially located data in a set of values (x_i) is likely to be related over space. Many authors in various disciplines discuss presence of dependence among observations related on diseases in space. It is believed that spatial dependence is present in every direction and gets weaker the more the dispersion in the data localization increases [58].

Woreda (Districts)

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	1	1	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1	0	0
16	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0

Table 5.1: Spatial Weight Matrix

Remark: Identification numbers in Table 5.1, 5.2 and 5.3 for woredas are given as: 1=Chora Botor,2=Botor Tolay, 3= Manch, 4=Dedo, 5=Nono Benja, 6=Limu Seqa, 7=Omo Nada, 8=Omo Beyam, 9=Limu Kosa, 10=Sokoru, 11=Tiro Afeta, 12=Kersa, 13=Mena, 14=Goma, 15=Gera, 16=Seqa chokorsa, 17=Sigmo, 18=Setama, 19=shabe sombo, 20=Guma, 21=Agaro, 22=Jimma city

Table 5.2: Summary of positive HIV cases by Marital status, Educational Level and District of Jimma Zone

Districts	**Marital status**				**Educational Level**				**Cases**
	0	**1**	**2**	**3**	**0**	**1**	**2**	**3**	
Chora Botor	1	48	29	22	0	62	24	14	100(2.47)
Botor Tolay	6	43	33	23	11	57	26	11	105(2.59)
Manch	20	37	20	13	31	28	23	8	90(2.22)
Dedo	28	27	13	12	22	44	10	4	80(1.97)
Nono Benja	21	19	14	21	21	31	13	10	75(1.85)
Limu Seqa	21	66	31	32	47	67	32	4	150(3.71)
Omo Nada	16	35	22	11	20	29	24	11	84(2.07)
Omo Beyam	26	25	12	7	17	40	13	0	70(1.73)
Limu Kosa	110	112	51	77	86	179	68	17	350(8.66)
Sokoru	22	15	15	13	1	43	15	6	65(1.60)
Tiro Afeta	23	20	15	14	2	49	16	5	72(1.78)
Kersa	68	45	18	19	29	64	46	11	150(3.71)
Mena	90	88	48	44	69	123	54	24	270(6.68)
Goma	94	125	92	39	78	148	103	21	350(8.66)
Gera	22	106	51	21	30	125	36	9	200(4.94)
Seqa chokorsa	18	122	72	38	52	135	52	11	250(6.18)
Sigmo	5	98	34	23	39	64	42	15	160(3.95)
Setama 18	3	40	73	9	17	61	37	10	125(3.09)
Shabe sombo	1	48	39	12	25	52	22	1	100(2.47)
Guma	65	91	48	16	82	94	38	6	220(5.44)
Agaro	125	125	135	55	135	197	84	24	440(10.88)
Jimma city	74	317	88	56	151	260	86	38	535(13.24)
Total	859	1652	953	577	965	1952	864	260	4041(100)

Marital status (Single=0, Married=1, Divorced=2 and Widowed=3), Educational Level (No education=0, Primary=1, Secondary=3 and Superior=4)

Table 5.3: Summary of positive HIV cases by Condom use, Religion, Residence and District of Jimma Zone

Districts	**Condom use**		**Religion**			**Residence**		**Cases**
	0	**1**	**0**	**1**	**2**	**0**	**1**	
Chora Botor	93	7	18	62	20	5	95	100(2.47)
Botor Tolay	98	7	15	73	17	17	88	105(2.59)
Mancho	77	13	37	37	16	12	78	90(2.22)
Dedo	67	13	21	48	11	38	42	80(1.97)
Nono Benja	56	19	26	33	16	37	38	75(1.85)
Limu Seqa	139	11	63	76	11	64	86	150(3.71)
Omo Nada	83	1	31	39	14	32	52	4(2.07)
Omo Beyam	46	24	23	31	16	47	23	70(1.73)
Limu Kosa	295	55	123	173	54	138	212	350(8.66)
Sokoru	64	1	22	38	5	25	40	65(1.60)
Tiro Afeta	71	1	30	28	14	32	40	72(1.78)
Kersa	140	10	54	75	21	69	81	150(3.71)
Mena	257	13	96	110	64	134	136	270(6.68)
Goma	329	21	122	132	96	172	178	350(8.66)
Gera	197	3	62	96	42	110	90	200(4.94)
Seqa Chokorsa	245	5	79	116	55	82	168	250(6.18)
Sigmo	149	11	51	83	26	75	85	160(3.95)
Setama	114	11	49	57	19	50	75	125(3.09)
Shabe Sombo	95	5	26	50	24	48	52	100(2.47)
Guma	209	11	57	117	46	120	100	220(5.44)
Agaro	423	17	150	207	83	178	262	440(10.88)
Jimma city	473	62	151	286	98	218	317	535(13.24)
Total	3720	321	1306	1967	768	1703	2338	4041(100)

Condom use (No=0, Yes=1) , Religion (Protestant=0, Muslim=1 and Orthodox=2) and Place of residence(Rural=0, Urban=1)

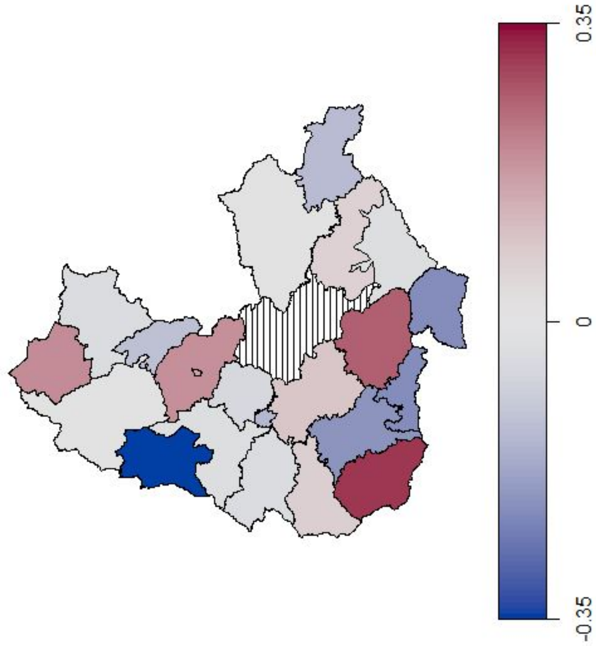

Figure 5.1: Maps of Jimma Zone for HIV cases showing unstructured spatial effects

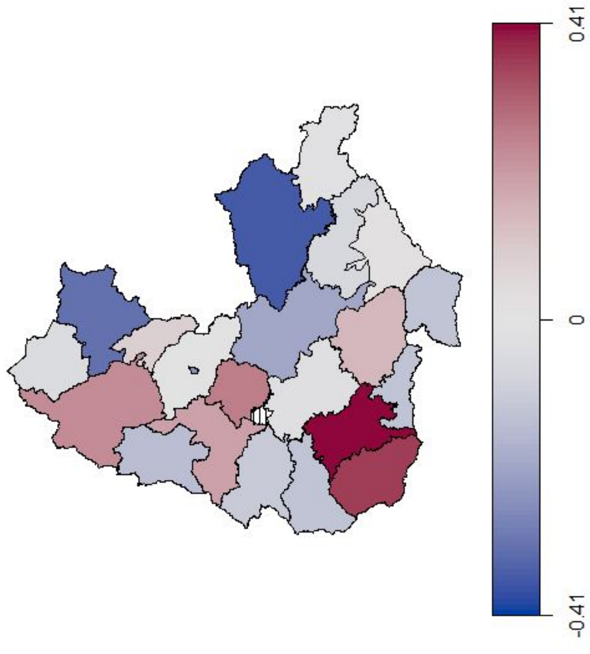

Figure 5.2: Maps of Jimma Zone for HIV cases showing structured spatial effects

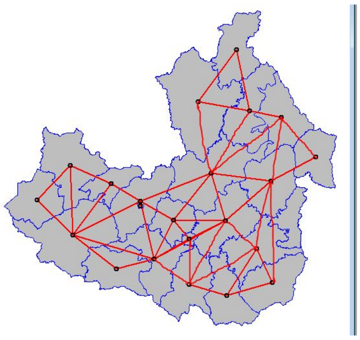

Figure 5.3: Neighborhoods Spatial Weight Matrix

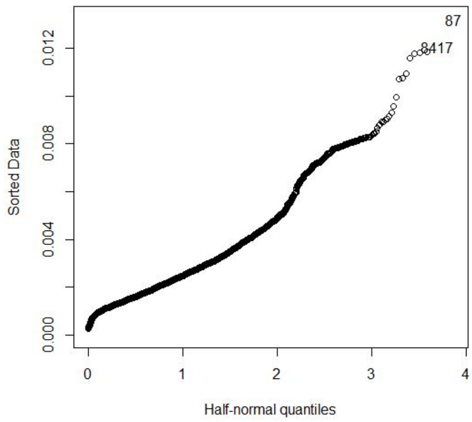

Figure 5.4: Half Normal Quantiles Plot

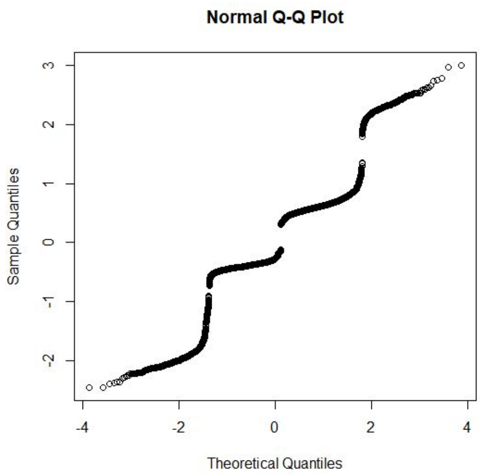

Figure 5.5: Normal QQ Plot