KWAME KRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF SCIENCE

FACULTY OF PHYSICAL SCIENCES

SURVIVAL ANALYSIS OF DROPOUT RATE AT THE BASIC SCHOOL LEVEL

(CASE STUDY: NORTHERN GHANA)

BY

VIDA AFOSAA

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS IN

PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF PHILOSOPHY IN MATHEMATICS

SEPTEMBER, 2009

DECLARATION

It is hereby declared that this thesis is a true account of the candidate's own research work except for references to other people's work, which have been duly acknowledged

SIGNATURE:…………………………. DATE: ....

VIDA AFOSAA (PG1542007)

(STUDENT)

SIGNATURE:………………………… DATE:…………………………….

DR. F. T. ODURO

(SUPERVISOR)

SIGNATURE:………………………….. DATE:……………………………..

DR. S. A. AMPONSAH

(HEAD OF DEPARTMENT)

ACKNOWLEDGEMENTS

I would like to begin by thanking my heavenly Father for the blessings and abilities that made this accomplishment possible. I would like to thank Him for his help and guidance in enabling me to accomplish this life long dream.

I would next like to express me heartfelt thanks to my supervisor Dr. F. T. Oduro a wonderful man and great teacher for his support, his help, and his guidance in completing this work. I am very grateful to God for your life, I say thank you and may God richly bless you.

I would also like to thank all the lectures at the Mathematics Department of KNUST, especially, Dr. S.A. Opoku, Dr. E. Prempeh, Dr. S.K. Amponsah (Action Man), Mr. W. Obeng –Denteh, Mr. Dedu K. Vincent, Nana K. Frimpong and others for their valuable input and suggestions.

I would also like to give this thanks to Professor F. K. Allotey (Director, Institute of Mathematical Sciences, and Accra) for his fatherly love, contribution, support and encouragement in completing this work.

I would now like to thank my family, who without all of their love and support none of this would have been possible. I would like to thank my parents; Mr. K. Amankwaa Asante and Madam Ernestina Yeboah who gave me the necessary love and support to help me get to where I am today. I would like to thank Mr. Daniel B. Senyah for checking in from time to time to see how my progress was going. I would next like to thank all my friends; Mr Kingsely Kodom, Grace Mirekuah-Asante and others who

helped me in diverse ways to make this dream a reality. God bless you all and replace whatever you have lost because of me.

Lastly, I would like to dedicate this work to the memory of Mr. D. Nsiah a wonderful man and a great teacher who was taken from us much too early.

ABSTRACT

Development and growth cannot be experienced in any part of the world without good education. Basic education is one of the effective investments in improving economies, reducing illiteracy, creating self-reliant and healthy society. In view of this, government and NGOs have constituted various policies and intervention programmes to make sure children of school going age enroll, attend and remain in school till completion (at least the basic level). Despite these efforts, studies have shown that dropout rates remain high in rural areas and in the three northern regions of Ghana. The purpose of the research is to estimate the probability of a child dropping out of school at some point within the basic level in Northern Ghana. Data was obtained from the Ministry of Education Youth and Sports on annual enrollment levels for each district for the years 2000 to 2007. Survival Analysis which is a time to event analysis was used to investigate the effect of gender and region on survival time (i.e. time before dropout). Particularly, the Kaplan-Meier method was utilized for plotting the survival and hazard functions. Log rank test was used to compare the survival curves. The Cox Proportional Hazards Model was used to fit the data. Findings indicated that the region of a child has significant effect on the potential of his/her dropout of school. On the other hand, gender had no significant influence on a pupil's dropout rate. The analysis also indicated that the Upper West region has the lowest hazard of dropout and Northern Region has the highest Hazard of dropout of school. It was also discovered that on the average, the potential of a child dropping out of school occurred mostly at primary 2, 5 and 6. It is suggested that interventions to reduce dropout rates should not be focused only on female pupils but on male pupils as well.

TABLE OF CONTENTS

CHAPTER THREE: REVIEW ON SURVIVAL ANALYSIS

CHAPTER FOUR: DATA ANALYSIS

CHAPTER FIVE: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND TO THE STUDY

The purpose of this project is to study the probability of a child dropping out of school at the basic level after some time in the three northern regions of Ghana.

Northern Ghana presents an interesting case of the limitations of the conventional school system in reaching underserved and deprived populations with basic education. Due to the peculiar nature of its demographic characteristics and the socio-economic challenges that confront this area of Ghana, conventional school system is unable to thrive and make an impact in remote areas. Many of these communities are sparsely populated and scattered making distance a hindrance to school attendance. The greatest barrier to access and participation is also the direct or indirect costs to families. Direct costs arises from schooling accessories such as uniforms, books and writing materials whilst the indirect costs are largely in the form of income lost from the child's potential employment or contribution to household income through direct labor. Yet another obstacle is the official school calendar which usually conflicts with the families' economic activities to which the child is a crucial contributor. [20]

Education is, nevertheless, the main key to unlocking the full potentials of a nation's human resources for development. It creates situations and events and facilitates the infusion of innovations and new or improved productivity and overall production in all sections of the national or local economy [2]. In view of this fact, many educational plans have been established since independence by various governments, Non Governmental Organizations and other donor organizations; to make education available and accessible by every child of school going age.

1.1.1 EDUCATIONAL POLICIES IN GHANA

The development of education since independence has been and continuous to be guided by various Education Acts and programmes, the most fundamental being the Education Act 1961. It is the principal legislation on the right to education and its states in section 2(1):

"Every child who has attained the school going age as determined by the Minister shall attend a course of instruction as laid down by the Minister in a school recognized for the purpose by the Minister." [14].

The 1992 Constitution gives further impetus to the provision of education as a basic right for all Ghanaians. Article 38, sub-section 2 states:

"The Government shall within two years after parliament first meets after coming into force of this constitution draw up a programme for the implementation within the following ten years for the provision of a free, compulsory universal basic education."[27]

In 1996, the Free and Compulsory Universal Basic Education (FCUBE) programme was launched. This is a 10-year programme (1995-2005) designed to establish the policy frame work, strategies and activities to achieve free and compulsory basic education for all children of school going age.[27]

The establishment of the Girl's Education Unit in 1997 marked a major step in the country's commitment to ensuring the respect for the general principle of securing a non-discriminatory environment and the reduction of gender disparities in the education sector. [27]

The 1951 Accelerated plan declared the first cycle of education to be free and compulsory; some minimal fees were introduced in the 1980s to meet textbook costs.

There is also a policy to achieve a Universal Basic Completion (UBC) rate by 2015 and every child in the relevant age group also to complete second cycle education in Ghana by 2020.

The Education Strategic Plan (ESP) (2003-2015) is very gender sensitive with its allocation of funds for the education of girls.

The Millennium Development Goal (MDG) on education enjoins all countries to work towards achieving universal education by the year 2015. Specifically, MDG2 is to ensure that all boys and girls complete a full course of primary schooling (World Education Forum Dakar, 2000). [24]

The government has also decided to pay capitation grants and implement schooling feeding programmes to cover fees and cost of feeding respectively for public schools. The purpose is to remove user-fees, and feeding cost which has become a barrier to many poor families and communities.

The Northern Scholarship scheme was set up by Ghana's first president, Dr. Kwame Nkrumah, exclusively for the people of north, to correct in the instance of the need for equity in national resource allocation, the injustices of the colonial era which deprived the people of northern decent education opportunities. [8]

1.1.2 PROFILES OF NORTHERN GHANA

Geography and climate [1][16]

There are three regions in the north of Ghana presently: Upper East, Upper West and Northern Regions. These three regions are in the Guinea Sudan and Sahel climate zones with population of approximately 2.4million and the youth constituting about 64 percent. The area has a mean annual rainfall values ranging from 600-1,100 mm. The rainy season begins in May and ends in October. The dry season starts in November and ends in March/April. The Long periods of droughts are followed by episodic torrential rains which sometimes result in floods in some areas.[16]

Studies have noted that Northern Ghana falls short by almost all indicators, one hindrance is geography. The three northern regions are far from the ports, roads, railways, markets, industrial centers and fertile farming areas that help stimulate greater economic and human development in southern Ghana.[1]

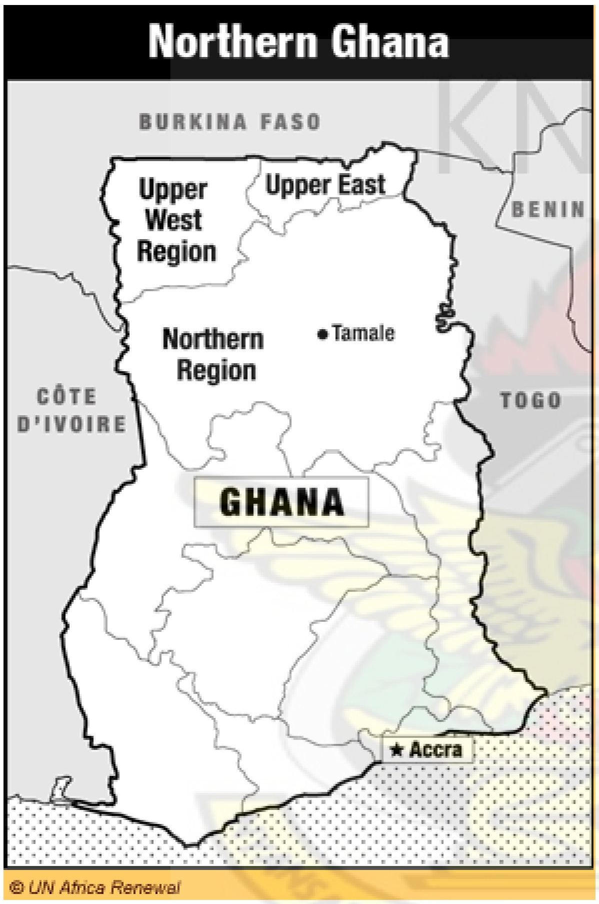

Figure 1.1: The map of Ghana showing the location of Northern Ghana.[16]

Religious affiliation [14]

Three main religious grouping are found in the three Northern Regions of Ghana, namely: Christianity, Islam and the Traditional. Considering them on regional bases we have;

(i) Northern Region: Islam is the dominant religion, of the region(56.1%), Traditional religion is the next(21.3%) and Christians represent 19.3%

(ii) Upper East Region: Traditional religion is the most common form of worship in the region (46.4%), followed by Christianity (28.3%) and Islam (22.6%).

(iii) Upper West Region: Christianity is the most common form of worship in the region (35.5%), followed by Islam (32.2%) and Traditional religion (29.3%).

Economic characteristics

Peasant farming is therefore the main occupation and livelihood activities for the inhabitants of the North. The crops they grow are maize, yam, millet, cassava, Soya bean, rice. Other livelihood activities include fetching of firewood, Shea nut picking and butter extraction, charcoal burning, and the rearing of animals such as sheep, goats and cattle. They also take care of birds like the hen and guinea fowls in very small numbers. Incidentally, Food crop farmers are among the very poor in the country.[16] The incidence of poverty in the Northern Region declined only slightly over the same period, from 63 per cent to 52 per cent. In the Upper West Region it remained static, at 88 per cent, while in the Upper East Region it actually increased, from 67 per cent to 70 per cent. [1]

The nature of their occupation, (peasant farming) degenerated by unfavourable weather conditions and the fast depleting soil's fertility, makes them unable to produce enough for their families throughout the year. By the fourth month after harvesting (March/April), more than two thirds of households begin to eat for just once or at best twice a day**.** Poverty has therefore raised its ugly heard all over the north making it the most poverty stricken areas among the 10 regions of Ghana. [16]

Educational Attainment and Interventions

One consequence of the north's limited development is that few resources can be generated from within the region for essential social services.[1]. Education has been far lagging behind that of the southern Ghana by over 5 decades.[16] While nearly 70 per cent of all school-age children are enrolled in primary schools nationally, in the Northern Region the rate is just 50 per cent and in the Upper West and Upper East regions only 51 per cent and 56 per cent, respectively. The three northern regions can claim only about half the national secondary-school enrolment rate. [1] In Ghana Statistical Service's Core Welfare Indicators Questionnaire Survey of 2003 for example, access to secondary education was recorded at 7.9 percent in the Upper East Region as against 63.4 and 56.1 percentage in Greater Accra and Ashanti Regions respectively.[16]

Though Education is said to be "FREE" it has not been easy to access by the northerners even within their own territories let alone in tertiary institutions elsewhere (outside the north) due to several factors including facility fee payments, books, accommodation, etc.[16]

Relying mainly on the human resource of the youth and children as farm hands or shepherds for livelihood, parents traditionally are confronted with the challenge of enrolling their wards in schools as a difficult option. School drop-out rates are consequently high at the northern regions of Ghana and inevitably street hawking and arms-begging by strong youth have gradually become common.[16]

Regrettably, farms culturally handed down to the youth yield just a little**:** thus accounting for their involvement in other unfruitful augments, petty fights, political disturbances, tools for chieftaincy disturbances and sometimes high street rubbery and home wrecking. Often, the very few young men and women who are able to find their way into institutions do that on their own with no or very little assistance from their parents or the government.[16]

Efforts to Get Children Especially Girls into School [1][3][16]

As elsewhere in Ghana, special efforts have been made over the past decade to get more children into school. Poor families receive small grants when they enroll children, and school feeding programmes provide an additional incentive. In areas where the school feeding programme has been introduced, notes Mr. Dordunoo, “enrolment has so increased that infrastructure, school buildings, facilities are not able to cope with the increase.” [1].

Ghana School Feeding Programme and Capitation Grants are implemented in Basic Schools with the aim of increasing enrolment, attendance, and retention rates. Enrolment rates in World Food Programme(WFP)-assisted schools in the Bolgatanga and Bongo

Districts increased by 14 and 10 percent respectively.[3] The take-home ration component has also had a tremendous impact on girls´ education in Ghana´s three northern regions. Girls´ enrolment in assisted schools grew from 9,000 to 42,000 at the peak of the programme, whilst retention rates doubled to 99 percent. In order to qualify for take-home rations, girls have to attend school for a minimum of 80 percent of the month. This has led to two interesting developments. First of all, parents allow their daughters to attend school more willingly and regularly, because the take-home rations are considered as compensation for the loss of the economic activity which their daughters would have provided if they had remained at home. Secondly, regular school attendance has resulted in better academic performance, enabling more girls to qualify into high schools.[3]

Other Non Governmental bodies and Christian Aid like the Catholic Relief Services(CRS) are also helping immensely get more children of school going age into school. In fact, according to source some of these bodies started providing food, school uniforms, teaching and learning materials, etc to some of the pupils, especially girls in the rural areas of Northern Ghana before the implementation of the government interventions. Now, it is noted that the CRS are concentrated at the Upper West Region, where there are more Catholic schools. Bonaboto Educational Fund(BEAF) was also launched in 2001 to assist very poor and brilliant school children in some areas in the Upper East Region.

‘Dehumanizing cultural practices’

Across most social and economic indicators — from school enrolment to health to access to land — northern women and girls fare far worse than their male counterparts. That is true across Ghana (as in most countries worldwide), but the north in particular is influenced by the persistence of “dehumanizing cultural practices,” says the *Ghana Human Development Report*.

” Female genital cutting remains widespread. “Those practices are still there, even though they are declared illegal.”

Other practices are also common. Many young girls are abducted or forced into marriage at an early age. In the Northern Region only 2 per cent of landholdings are held by women, and in Upper West just 4 per cent. Widows have few rights to inheritance.[1]

1.2 PROBLEM STATEMENT

School enrolment in the northern part of Ghana has been hampered by the involvement of children in farm and domestic labour. Drop out from school is a common phenomenon in the Northern Regions of Ghana. For example, in the West Mamprusi District especially in the Yagaba Circuit (overseas area), the annual drop out rate is 3.6%. In addition, there is a high incidence of seasonal drop out. During the farming season, not less than 40% of the pupils in the district stop attending school. They help their parents in their farms. [14] The high drop-out rate of boys and girls in school is the most significant phenomenon in Northern Ghana. Some of the reasons are:

- As pupils grow up, they are needed in the house to take part in the tedious tasks of fetching potable water and fire wood tasks considered to be for children.
- Parental poverty issues which compel them to force girls into early marriages and the boys to help on the farm and herding of cattle or child labor at early ages.
- Lack of good quality of teaching and learning materials.
- Parents unwilling to allow their children especially, young ones walk very long distances to school for safety reasons.
- The presence of anti girl-child education and socio-cultural practices pulling teenage boys and girls out of school.
- Lack of access to school. [24] [14]

1.3 OBJECTIVES OF THE STUDY

The main objectives of this research are to:

(i) Estimate the probability that a child will drop out of school after some time (years).

(ii) Determine if there is a significant difference between the drop out rate of boys and girls in Northern Ghana at the basic level; particularly, in the aftermath of the implementation of the school feeding programme and the payment of the capitation grant.

(iii) Compare the basic school dropout rates among the three Northern Regions of Ghana.

1.4 METHODOLOGY

The research was limited to secondary data obtained from Ministry of Education Youth and Sports on annual enrolment levels of the Basic Schools in each district at the three Northern Regions for the years 2001 to 2007. Survival Analysis considers subjects individually; so here, the various districts were considered as individual. Since the data is not continuous, a threshold of 35% was set whereby a percentage difference in enrolment between two adjacent years exceeding the threshold was considered to indicative of a dropout situation.

Kaplan-Meier Survival and Hazard curves were plotted for the various factors. Log-Rank (standard nonparametric method) and Breslow tests were used to compare the Kaplan-Meier Survival Curves. The Cox Proportional Hazards Model was used to fit the data.

The Statistical Product for Social Scientists (SPSS) was used as a computational tool for the analysis.

1.5 LIMITATIONS

There are some limitations to this project which includes:

The nature of the data for the analysis which is a secondary data and since the data collection was not done directly, the reliability / authenticity of the data cannot be 100% guaranteed.

Also, all the regions in the country could not be studied due to the limited time available for the study and also due to a lack of funding for this project.

Finally, the records of pupils appeared not properly kept so that data on annual enrolment levels for some particular years and posed difficulties of retrieval for the study.

1.6 STRUCTURE OF THE THESIS

Chapter one contains the introduction. Basic concepts in Set Theory, Probability Theory and test of hypothesis are in chapter two. Also, we have in chapter three, the review on principles and application of the survival analysis. The Data Analysis of the study can be found in chapter four. However, chapter five contains the Discussion, Conclusion and Recommendation of the study. Finally, we have the list of references, tables and Appendix.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

A dropout is considered, a student who for any reason other than death leaves school before graduation without transferring to another school. Dropping out of school is a well documented social problem and often present daunting circumstances for adolescents. Dropping out is also associated with delinquency, and low school achievements. Dropouts also cost the nation billions of dollars in lost tax revenues and in welfare, unemployment, and crime prevention programs [5]. Studies have shown that one of the major reasons for children being kept out-of-school was the lack of education of parents.

Research describes dropout and repetition as educational wastage of inputs since the output is not there as planned (UNESCO, 1998). On the other hand, educational wastage is defined as the input/output ratio, as an issue of cost-efficiency. Dropout as a measure of wastage or efficiency of school or education system is measured by a method of cohort analysis. Cohort is a group of pupils who enter the first cycle of a school in the same year (UNESCO, 1998, p14). The commonly used way of cohort analysis is that the enrollment in one grade in a given year is compared with enrollment of the consecutive grade during the following years.[28]

The determination of school dropout deals with concept of enrollment which is defined as the total count of pupils who were enrolled on the first day in the academic year, or at any time in the course of the school year, and used as the population figure against which dropout is subsequently counted.

In this study, dropout is indicated by a percentage difference in enrollment exceeding a percentage margin of 35%. School dropout is a world wide problem, and it affects the futures of our students every day.

2.1 ISSUES AND CAUSES OF SCHOOL DROPOUT IN AMERICA

High School Dropout Students dropping out of high school is a major problem facing America today. Millions of young people are dropouts without a high school diploma (Schwartz). Nearly half a million students are dropping out each year (Schwartz). The dropout rate is declining a little each year, yet it is still a severe problem facing America. A large portion of dropouts happen before the tenth grade (Schwartz). The main causes of school dropouts are personal factors, home and school stability, school experiences, social behavior, and rebellion. Personal problems affecting students seem to be the main cause for students to drop out of high school. Children seem to be the main personal problem facing dropout students, especially in women. Close to half of the dropouts students, both male and female, have children or are expecting one (Schwartz). Marriage is another great personal factor to the dropout rate. Marriage is a very stressful factor to any student. This stress could and does send many students to the point where they have too much to deal with. School becomes a second priority and is often discarded to lighten the load married couples deal with. Most of dropouts are married. Jobs also increase the percentage of students dropping out of school. Some students may and do have to take on a job to support themselves or their family. The job may interfere with school hours, school homework, and/or school activities. Drug problems are very serious and have major side effects students. This serious problem causes many students to drop out of school. They do this to either to help their drug addiction or to get a handle on their problem. Students in broken homes are more than twice as likely to drop out of school than those with families intact (Schwartz). This is so because of the fact that this is another stressful matter these young minds must also deal with. Home and school stability is another cause to for students dropping out of high school. More than half of dropouts have moved within their four years of being in high school (Schwartz). If a student does not have a stable home or a stable school life, then they are more likely to drop out of school. Stableness allows the student to feel comfortable enough to try to work at school. If they have a stable home and school life, then that is one less worry for them. This allows them to concentrate on staying in school instead. The more stable a situation is, the more comfortable the student becomes with the surroundings, the better they get along with teachers and students, and the easier it is for the to fit in and work hard at school. Bad school experiences are also a large contributor for the school dropout rate. A large majority of dropout students were only taking the bare minimum general high school requirements. This is because no one pushed them to try harder. This made students feel that school was not important enough to try hard at. These students then do not even care because they do not think school is important. These students also said they did not have much attention given to them when dealing with their schoolwork. This also emphasized that school was not important. A large majority of dropout students were held back a grade at least once in school life. This made those students feel as if they were not as smart as the other students, so why even bother. Social behavior is another cause for students dropping out of high school. Most students who dropped out of school did not like school to begin with. These students were failing as it was. They could not keep up with their schoolwork. They did not get along with their teachers and/or other students. They may have had disciplinary problems. A great portion of dropout students were suspended at one time or another. Frequently absent students also make them more likely to drop out. A good deal of dropout students had even been previously arrested. This is because high school dropouts tended to believe they have no control over their own lives. Some students did not feel like they fitted in or they may have felt unsafe. At this time in a students life school is the most awkward place to be. Students will stay away from it if they can help it. They will use any tactic including dropping out. There is also the rebellion factor. Rebellion is a very big cause of students dropping out of school. To some students, school is a place where their parents force them to go every weekday. These students feel like the teachers make the sit down and listen to lecture after lecture. They also feel like they do not learn to think but rather to only listen and repeat. This is where the rebellion factor comes into play. Those students do not wish to be asked to repeat something, but rather listen, think, and say their thoughts feelings and views. These students do not want to be told when they can and cannot have a personal opinion on a topic. These students rebel because they feel that they do not have a choice. If they speak up they think they will be considered a problem child. If they quietly sit in the back they feel they are going against their own selves. Those students drop out of school to find the world. They leave school to be open-minded. Not all students are like this but there are a few who are. If only those students could be redirected to something useful in the school then they might not drop out. The high school dropout rate is reducing. It is still, however, a major concern that should worked on. The home and school stability, bad school experiences, social behavior, and rebellion all have one main effect: increasing of the nation dropout rate of high school students in the United States. With half a million people dropping out of school each year, these key causes need to be looked at and examined closely. If these causes are worked at, maybe the effect will not be has high as it is now. [29]

2.1.1UNDERLYING CAUSES AND CONSEQENCES OF SCHOOL DROPOUT IN THE U.S HIGH SCHOOLS[4]

What Are the Characteristics of Students Who Drop Out?

Socio-economic Background National data show that students from low-income families are 2.4 times more likely to drop out of school than are children from middle-income families, and 10.5 times more likely than students from high-income families. **Disabilities** Students with disabilities are also more likely to drop out. The National Transition Study estimates that as many as 36.4% of disabled youth drop out of school before completing a diploma or certificate.

Race-ethnicity Hispanics and African Americans are at greater risk of dropping out than whites, with Hispanics at a greater risk of dropping out than either white or African American students. Nearly 40% of Hispanics who drop out do so before the eighth grade.

Academic Factors National research also indicates that academic factors are clearly related to dropping out. Students who receive poor grades, who repeat a grade or who are over-age for their class, are more likely to drop out.

Absenteeism Students who have poor attendance for reasons other than illness are also more likely to drop out. Clearly, students who miss school fall behind their peers in the classroom. This, in turn, leads to low self-esteem and increases the likelihood that at-risk students will drop out of school.

Occupational Aspirations Young people's perceptions of the economic opportunities available to them also play a role in their decision to drop out or stay in school. Dropouts often have lower occupational aspirations than their peers.

Six Predictive Factors The following individual-level factors are all strongly predictive of dropping out of high school:

- Grade retention (being held back to repeat a grade)
- Poor academic performance
- Moves during high school
- High absenteeism
- Misbehavior
- The student's feeling that no adult in the school cares about his or her welfare. What Reasons Do Young People Give for Dropping Out?

According to a National Longitudinal Study conducted by the U.S. Department of Education Statistics, here is a summary of the key reasons why 8^{th} to 10^{th} grade students dropped out:

School related:

- Did not like school (51%)
- Could not get along with teachers (35.0%)
- Was failing school (39.9%)

Job related:

- Couldn’t work and go to school at the same time (14.1%)
- Had to get a job (15.3%)
- Found a job (15.3%)

Family related:

- Was pregnant (51.0%)
- Became parent (13.6%)
- Got married (13.1%)

What are the Consequences of Dropping Out of School?

In recent years, advances in technology have fueled the demand for a highly skilled labor force, transforming a high school education into a minimum requirement for entry into the labor market.

Because high school completion has become a basic prerequisite for many entry-level jobs, as well as higher education, the economic consequences of leaving high school without a diploma are severe.

Earnings Potential On average, dropouts are more likely to be unemployed than high school graduates and to earn less money when they eventually secure work. Employed dropouts in a variety of studies reported working at unskilled jobs or at low-paying service occupations offering little opportunity for upward mobility. Dropping out and severely impairing a young person's job prospects and earnings potential, in turn, causes other secondary, indirect problems:

- **Public Assistance** High school dropouts are also more likely to receive public assistance than high school graduates who do not go on to college. In fact, one national study noted that dropouts comprise nearly half of the heads of households on welfare.

- **Single Parents** This increased reliance on public assistance is likely due, at least in part, to the fact that young women who drop out of school are more likely to have children at younger ages and more likely to be single parents than high school graduates.

- **Prisons.** The individual stresses and frustrations associated with dropping out have social implications as well: dropouts make up a disproportionate percentage of the nation's prisons and death row inmates. One research study pointed out that 82% of America's prisoners are high school dropouts

2.2 CAUSES AND ISSUES OF SCHOOL DROPOUT IN INDIA [25]

Issue of school dropouts in Mumbai By S. Govande (2008)

The education department of Mumbai runs 1,177 primary schools and 49 secondary schools that reach out to more than 500,000 children from the low-income groups. Rough estimates state that over 53 per cent of the children from municipal schools drop out at about 10 –16 years of age.

The reasons for dropping out are many. Onset of puberty resulting into engagement and marriage, household chores mainly looking after children, financial crisis at home are very common reasons for high dropout rate among girls. For boys, the main reasons for dropping out are financial crisis, inability to give good results in the school examinations and the need to be productive and contribute to the family income.

Another very important cause for dropping out is the belief that education is unnecessary and of no use. The quality of education (both content and methodology) is very poor. The curriculums do not equip children with various skills that they require to enter the world of responsibilities. Practical learning is missing from the education system.

Keeping in mind the needs of the low-income groups, the education department needs to rework on the curriculum content and methodology. Work based practical learning that provide children with skills, which is necessary to start working by the age of 16 - 18 years will ensure a low dropout rate. Currently, such work-based training is limited to the ITI (Industrial Training Institutes) and ITCs (Industrial Training

Centres). Some private vocational training institutes also provide such work based learning opportunities.

Many education-focused organisations have been working toward building child friendly curriculums that include life skills development in the children. Unfortunately, implementation of these curriculums remains limited to the schools that these organizations reach out. Many organizations have also invested in infrastructure development, provision of teaching aids and toy and book library in schools.

Schools in Mumbai also lack the much-required infrastructure. Dilapidated buildings, broken furniture, poorly maintained toilets, poor teaching aids and lack of library resources together add to the problem of high dropout.

2.3 GENERAL CAUSES OF DROPOUTS FROM SCHOOL

There many factors related to the school environments that affect school completion and dropout rates. Some of these factors are relatively unchanging from one county to another, namely school size, location, the percentage of English Language Learners (ELL), and the demographic make-up of the school. There are other factors that schools have greater control over; these are teacher-quality, class size, and school safety and students discipline. When combined, both sets of factors lay the groundwork for understanding the problem of students' dropout and how to address it.

Finn (1989) explained that either low participation in school activities or early school failure leads to low self-esteem, problem behaviours and then alienation from school. In a later study (1993), he added that: “the likelihood that a youngster will successfully complete 12 years of schooling is maximized if he or she maintains multiple, expanding forms of participation in school-relevant activities”. Building on this idea, schools can reduce dropout by encouraging multiple types of extracurricular opportunities for students and ensuring that all students can participate (e.g. avoiding exorbitant fees that would preclude participation by students from lower income backgrounds).

Other factors related to the school setting include the student not liking school, being unable to get along with teachers or peers, having difficulty with the materials being taught, or even having safety concerns while at school(Gonzalez,2003). In addition students who receive disciplinary measures (detentions, suspension and expulsions) run a higher risk of dropping out as well as for retention between grades.

Finally, grade retention has been correlated with very high rates has of school dropout: retention for one year leads to a 50% likelihood of dropout while retention for a second year a dropout rate of 90%. As a result of this, programmes that aim to address school completion and dropout must place a considerable emphasis on improving students’ academic achievement while also being cautious in the area of grade retention (Baker et al, 2001).

There are numerous behaviours that describe students who are at-risk for dropping out of school; various behaviours have a symptom that is visible from within the school environment. These are as shown in Table 2.1.

Table 2.1: Behaviours leading to Dropout in School

Behaviour	**Symptom**
Poor school attitude	Disinterest in school work, often off-task
Low ability level	Poor grade
Attendancy/truancy	Ongoing absences, trouble getting caught up
Behaviour/discipline	
Pregnancy	Detentions, suspensions, expulsion
Drug abuse	Trouble connecting with school work
Poor peer relationships	Deviant friendships; illegal behaviour on campus
Non-participation	Isolates self from peers
Friends have dropped out	Anger, aggression, competing interests(like employment or sports)
Illness/disability	Poor relationship with peers, possible cognitive difficulties with assignments

(Adapted from NDPC/N, 2004)

The common element between many of these behaviour is what Finn (1993) termed as a student's "disengagement" from school. Considering the sources of each of these symptoms, teachers and administrators can help students to deal positively with their life's challenges and be successful in their school setting.

2.4 PREVIOUS STUDIES ON DROPOUT FROM SCHOOL

Johnson and Kyle(2001), examined the determinants of school enrollment in Ghana, considering historical and social information to formulate an econometric model of school enrollment patterns for households. Data came from a 1989 survey of households in Ghana. The survey collected basic information about community characteristics, health and school facilities, and living conditions. This 1989 survey was the second in a series of surveys in Ghana. It included a sample of children age 6-20 years. Analysis of the data indicated that gender of the child and school attendance of the child's mother was the most significant predictors of school enrollment status. Boys were more likely to attend schools than girls, and girls were more likely to drop out of school than boys. Uneducated mothers were three times more likely to have children who did not attend school. Girls of mothers who did not attend school were 1.8 times more likely to drop out and half as likely to attend school than girls of mothers who attended school. The mean cost of schooling had no measurable effect on school enrollment status.[17]

Braimah and Oduro-Ofori (2005), assessed the trend of basic school dropout in Amansie West, a predominantly rural district in Ghana and to further determine the main causes and policy implications of the phenomenon in the district. Analysis of the data revealed a downward trend in the dropout rates. At the primary school level, the dropout rate reduced from 5.4 percent in 1998/99 to 4.5 percent in the year 2000/01. At the Junior Secondary School (JSS) level, the dropout rate also reduced from 9.7 percent in 1998/99 to 6.7 percent in 2000/01. This trend was attributed to the diversification of the income sources of parents, which enabled them to earn more income to take care of their wards in school. Further analysis of data gathered revealed that about 45.4% of the parents of school dropouts in the district were extremely poor with annual incomes less than GH¢60.00. The views of all stakeholders of education in the district confirmed that the causes of basic school dropout were mainly poverty related. In view of the strong inverse relationship between rates of school dropout and income levels it is recommended that pro-poor programmes be initiated and implemented in order to increase enrolment and retention of children in school for the ultimate benefits of public investment in education to be derived.[9]

Lavado and Gallegos[1] (2005), tried to capture the systematic characteristics, by gender and urbanity, which induce dropouts. The framework used is based on survival and duration models. Finally, a brief simulation of a cash direct transfer programme is performed to quantify the expenditure needed in order to replicate international successful experiences.

Shiyuan and Wallace (2008), applied discrete times survival analysis techniques to analyze education duration in Jamaica. Based on the Jamaica Survey of Living Conditions 2002, we are able to estimate the effects of household, individual, and other related covariates on the risks of students dropping out. We compare the discrete time

Cox model and discrete lime logit model and determined that the two estimations are consistent. The estimation results measure the effects of the covariates and can be used to predict the dropout risks of particular students in each grade, which could provide useful implications for the formation of policy to improve education in Jamaica. [23]

Sottie (2008), examined factors responsible for student disaffection and dropout at the basic level of education in Ghana, looks at official policies and programmes put in place to address the problem and identify ways in which gaps in those policies and programmes could be addressed. Specific areas of interest include, pre-dropout difficulties related to individual, school, home and socio-cultural factors. Children who have dropped out of school, children in school, educators, parents, and welfare workers form the main respondents for the study. It is expected that an outcome of this study will be the development of practical strategies for providing the support children require to complete the statutory years of schooling. [2]

Ampiah and Adu-Yeboah(2009), examined the issue of school dropout in six communities in the Savelugu-Nanton District in the Northern Region of Ghana. The study focused on 89 children (64 boys and 25 girls) aged 7-16 years, who had dropped out of school. A snowballing sampling method was employed to recruit participants to the study. Two researchers interviewed the children using semi-structured interview schedules over a period of three weeks. School dropouts were asked to tell their own stories about their schooling experiences and the factors which led to them leaving school. From their accounts dropping out of school appears to be the result of a series of events involving a range of interrelated factors, rather than a single factor. The complex nature of the processes leading to dropout demands input from various actors (i.e. teachers, head teachers, parent-teacher associations, school management committees and community members) to detect and address at-risk factors early in order to reduce the likelihood of dropout. [18]

Imoro (2009), investigated the various dimensions of basic school dropouts in rural Ghana using the Asutifi district as a case study. The analysis of data (both quantitative and qualitative) gathered from several stakeholders of basic education in the district, revealed that the causes of school dropout were rather complex. Poor educational outcomes in terms of performance of candidates in the final examinations of the basic level as a result of the poor quality of teaching and learning in the rural environment was directly linked to the high rate of drop-out. Although some stakeholders claimed that poverty was the main cause of school dropout, the significance of the loss of confidence in the educational system cannot be overemphasized. The policy implication is that quality consideration in the basic education delivery should now be the priority in order to regain the confidence of parents and their wards in the educational system in general so that enrolments and retention of children in school could be enhanced.[16]

Rolleston (2009), examined access to and exclusion from basic education in Ghana over the period 1991 to 2006, using data derived from the Ghana Living Standards Surveys. It uses the CREATE zones of exclusion model to explore schooling access outcomes within the framework of the household production function. Empirical findings indicate that the period was marked by large-scale quantitative access gains in Ghana. However, rates of progress through the system, as well as rates of dropout, showed no such improvements. Progress towards completion of the basic phase of education was found to be the preserve of the relatively privileged, raising questions of equity in relation to both the supply and demand for schooling. While Ghana may be one of few countries in Africa to achieve universal initial access to education, considerable challenges lie ahead in terms of improving rates of retention and completion.[22]

The factors affecting children's educational attainment in recent research can be summarized as including house hold and societal characteristics. Some of these characteristics are parent's educational level, family size, gender, geographical location, state of education expenditure, age of the child and score at school, child's expectation/self-esteem, religious activities and distance to school. Children may be withheld from school to tend to younger siblings, to earn wages, to do household chores or farm work. These costs may be as important as or more important the enrolment fees. Actually it is not what the child must pay that is the problem; it is what they give up in the time involved in schooling. Although, other methods were used, Survival Analysis techniques have become more and more popular in most recent literature.

CHAPTER THREE

REVIEW ON SURVIVAL ANALYSIS

3.1 What Is Survival Analysis? [19]

Generally, survival analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs.

By time, we mean years, months, weeks, or days from the beginning of follow-up of an individual until an event occurs; alternatively, time can refer to the age of an individual when an event occurs.

By event, we mean death, disease incidence, relapse from remission, recovery (e.g., return to work) or any designated experience of interest that may happen to an individual.

In a survival analysis, we usually refer to the time variable as survival time, because it gives the time that an individual has "survived" over some follow up period.
We also typically refer to the event as a failure, because the event of interest usually is death, disease incidence, or some other negative individual experience.
However, survival time may be "time to return to work after an elective surgical procedure," in which case failure is a positive event.

Survival analysis attempts to answer questions such as: what is the fraction of a population which will survive past a certain time? Of those that survive, at what rate will they die or fail? Can multiple causes of death or failure be taken into account? How do particular circumstances or characteristics increase or decrease the odds of survival? [19]

The uses in the survival analysis of today vary quite a bit. Applications now include time until onset of disease, time until stock market crash, time until equipment failure, time until earthquake, and so on.

3.2 CENSORED DATA [19]

Censoring occurs when we have some information about individual survival time, but we don’t know the survival time exactly.

There are generally three reasons why censoring may occur:

(1) A Subject does not experience the event before the study ends;

(2) A Subject is lost to follow-up during the study period;

(3) A Subject withdraws from the study because of death (if death is not the event of interest) or some other reason (e.g., adverse drug reaction or other competing risk)

These situations are graphically illustrated below. The graph describes the experience of several persons followed over time. An X denotes a person who got the event.

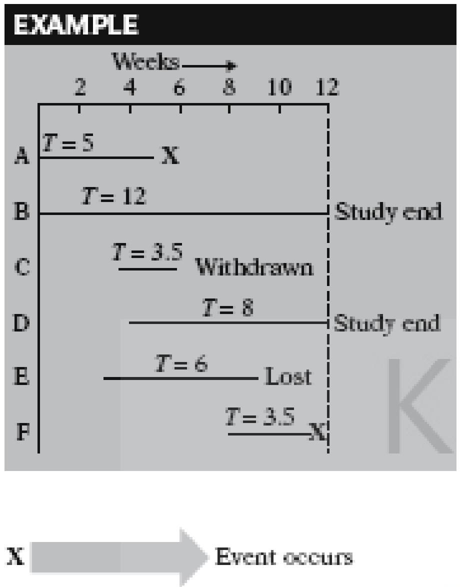

Figure3.1: Graphical representation of censored data

Person A, for example, is followed from the start of the study until getting the event at week 5; his survival time is 5 weeks and is not censored.

Person B also is observed from the start of the study but is followed to the end of the 12-week study period without getting the event; the survival time here is censored because we can say only that it is at least 12 weeks.

Person C enters the study between the second and third week and is followed until he withdraws from the study at 6 weeks; this person's survival time is censored after 3.5 weeks.

Person D enters at week 4 and is followed for the remainder of the study without getting the event; this Person's censored time is 8 weeks.

Person E enters the study at week 3 and is followed until week 9, when he is lost to follow-up; his censored time is 6 weeks. Person F enters at week 8 and is followed until getting the event at week 11.5. As with person A, there is no censoring here; the survival time is 3.5 weeks.

In short, of the six persons observed, two got the event (persons A and F) and four are censored (persons B, C, D, E)

3.2.1 Three Kinds of Censoring Commonly Encountered [19]

Right censoring

Left censoring

Interval censoring

3.2.1.1 Right censoring [19]

It occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. For example, we consider patients in a clinical trial to study the effect of treatments on stroke occurrence. The study ends after 5 years. Those patients who have had no strokes by the end of the year are censored. If the patient leaves the study at time t_e; then the event occurs in (t_e, ∞).

3.2.1.2 Left Censoring [19]

Left censoring is when the event of interest has already occurred before enrolment. This is very rarely encountered. For example, if we are following persons until they become HIV positive, we may record a failure when a subject first tests positive for the virus. However, we may not know exactly the time of first exposure to the virus, and therefore do not know exactly when the failure occurred. Thus, the survival time is censored on the left side since the true survival time, which ends at exposure, is shorter than the follow-up time, which ends when the subject tests positive.

3.2.1.3 Interval Censoring [19]

It occurs if the life time is know to be greater than some lower limit ‘L’ and less than some upper limit ‘U’.

We should note that Left Censoring is different from Truncation

3.3 Definition: Truncation [8]

Truncation is the exclusion of individuals from the data, where the investigator is unaware of their existence. Truncation is deliberate and due to study design.

Truncation comes in two forms, namely; Right truncation and Left truncation.

3.3.1 Right truncation: occurs when the entire study population has already experienced the event of interest (for example: a historical survey of patients on a cancer registry).

3.3.2 Left truncation: occurs when the subjects have been at risk before entering the study (for example: a life insurance policy holder where the study starts on a fixed date, event of interest is age at death).

3.4 SURVIVAL MODELS [8]

Definition 3.4.1: A Survival Model is a probabilistic model of a random variable that represents the waiting time until the occurrence of an unpredictable event. Survival Models are classified in two groups, namely; parametric and non-parameter models.

Survival Models can be applied in the modeling of:

- The time until surrender of a policy holder's life insurance policy.
- The time until the full payment of a loan.
- The time until the death of a patient
- The waiting time until a woman gets married.
- The time until an insurance claim is paid or settled.
- The time until a claim is made on an automobile insurance policy.

3.5 NOTATION AND TERMINOLOGY ([8],[19])

Let T denote the waiting time (survival time) till the occurrence of an unpredicted event as measured from some specified starting point also a non negative random variable with a probability density function *f(t)*.

Next, we denote by a small letter t any specific value of interest for the random variable capital T. For example, if we are interested in evaluating whether a person survives for more than 5 years after undergoing cancer therapy, small t equals 5; we then ask whether capital T exceeds 5. Thus, $T > t = 5$

Finally, we let the Greek letter delta (δ) denote a (0, 1) random variable indicating either failure or censorship. That is, δ = 1 for failure if the event occurs during the study period, or δ = 0 if the survival time is censored by the end of the study period.

The distribution function of T is therefore given by:

$$F(t) = P(T < t)$$

$$F(t) = \int_{-\infty}^{t} f(u)du$$

We next introduce and describe two quantitative terms considered in any survival analysis. These are the survivor function, denoted by S(t), and the hazard function, denoted by h(t).

3.5.1 Survival Function ([8],[19])

The survivor function S(t) gives the probability that the survival time T, is longer than some specified time t:

That is:

$$S(t) = P(T \geq t)$$
$$S(t) = 1 - F(t)$$

The survival function is also called the survivor function or survivorship function in problems of biological survival, and the reliability function in mechanical survival problems.[19]

The survivor function is fundamental to a survival analysis, because obtaining survival probabilities for different values of t provides crucial summary information from survival data.

Theoretically, as t ranges from 0 up to infinity the survivor function can be graphed as a smooth curve. As illustrated by the graph below, where t identifies the X-axis.

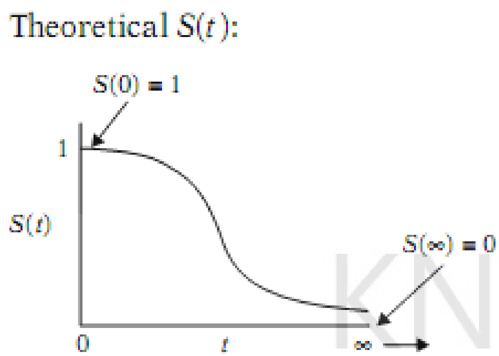

Figure 3.2: Theoretical graph of survivor function

3.5.2 Properties of Survival Function [19]

All survivor functions have the following characteristics:

(1) They are nonincreasing; that is, they head downward as t increases;

(2) at time t = 0, S(t) = S(0) = 1; that is, at the start of the study, since no one has gotten the event yet, the probability of surviving past time 0 is one;

(3) at time t =∞, S(t) = S(∞) = 0; that is, theoretically, if the study period increased without limit, eventually nobody would survive, so the survivor curve must eventually fall to zero.

Note that these are theoretical properties of survivor curves.

In practice, when using actual data, we usually obtain graphs that are step functions, as illustrated below, rather than smooth curves. Because the study period is never infinite in length, it is possible that not everyone studied gets the event. The estimated survivor function, denoted by a caret over the S in the graph, thus may not go all the way down to zero at the end of the study.

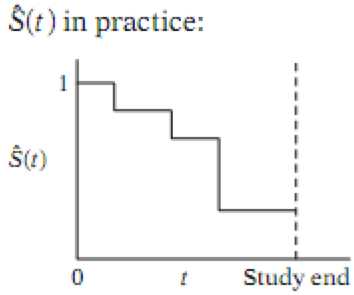

Figure 3.3: Practical graph of Survivor function

3.6 Lifetime distribution function and event density ([8],[25])

Related quantities are defined in terms of the survival function. The lifetime distribution function, usually denoted by *F*, is defined as the complement of the survival function,

$$F(t) = \Pr(T \leq t) = 1 - S(t)$$

and the derivative of *F* (i.e., the density function of the lifetime distribution) is denoted by *f*,

$$f(t) = \frac{d}{dt}F(t)$$

f is sometimes called the **event density**; it is the rate of death or failure events per unit time

3.7 Hazard and Cumulative Hazard Functions ([8],[22],[19])

The hazard function, denoted by $h(t)$ is widely used to express the risk or hazard of death at some time t, and is obtained from the probability that an individual dies at time t conditional on he/she having survived to that time.

For a formal definition of the hazard function, let us consider the probability that a random variable associated with an individual's life time T, lies between 't' and 't+dt', conditional on T being greater than or equal to 't' written:

$$P(t \leq T < t + d\ /T \geq t).$$

This conditional probability is then expressed as a probability per unit time by dividing the time interval dt, to give a rate.

The hazard function $h(t)$ is then the limiting value of this quantity, as dt tends to zero. That is:

$$h(t) = \lim_{dt \to 0} \left\{ \frac{p\left(t \leq T \leq t + dt/T \geq t\right)}{dt} \right\}$$

The hazard function h(t) gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t. Note that, in contrast to the survivor function, which focuses on not failing, the hazard function focuses on failing, that is, on the event occurring. Thus, in some sense, the hazard function can be considered as giving the opposite side of the information given by the survivor function.

As with a survivor function, the hazard function h(t) can be graphed as t ranges over various values. The graph below illustrates three different hazards. In contrast to a survivor function, the graph of h(t) does not have to start at 1 and go down to zero, but rather can start anywhere and go up and down in any direction over time.

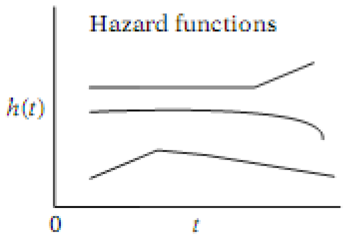

Figure 3.4: Graph of different Hazards

For a specified value of t, the hazard function h(t) has the following characteristics:

(1) it is always nonnegative, that is, equal to or greater than zero: $h(t) \geq 0$;

(2) it has no upper bound.

Considering, S(t) and h(t), the survivor function is more naturally appealing for analysis of survival data, simply because S(t) directly describes the survival experience of a study cohort. However, the hazard function is also of interest for the following reasons:

(1) it is a measure of instantaneous potential. Whereas a survival curve is a cumulative measure over time;

(2) it may be used to identify a specific model form, such as an exponential, a Weibull, or a lognormal curve that fits one's data;

(3) it is the vehicle by which mathematical modeling of survival data is carried out; that is, the survival model is usually written in terms of the hazard function.

3.7.1 The Cumulative Hazard Function [19]

$$H(t) = \int_0^t h(u)\,du$$

3.8 The relationship of h(t) to S(t) [8][19]

If you know one you can determine the other.

By the concept of conditional probability;

$$h(t) = \lim_{dt \to 0} \left\{ \frac{P(t \le T < t + dt)}{dt \bullet P(T \ge t)} \right\}$$

$$h(t) = \lim_{dt \to 0} \left\{ \frac{F(t+dt) - F(t)}{S(t) \bullet dt} \right\}$$

Where F(t) is the distribution function of T. This implies F(t) = 1- S(t)

But $F'(t) = \lim_{dt \to 0} \left\{ \frac{F(t+dt) - F(t)}{dt} \right\}$

$$\Rightarrow h(t) = \frac{F'(t)}{S(t)}; \quad F'(t) = f(t) = -S'(t)$$

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

Where *f(t)* is the density function of T.

$$\Rightarrow h(t) = -\frac{d}{dt}\left[\log S(t)\right]$$

$$S(t) = \exp\left[-H(t)\right]$$

$$H(t) = \int_0^t h(u)\,du$$

Where $H(t)$ is the integrated or cumulative hazard function.

$$H(t) = -\log S(t)$$

More generally, the relationship between S(t) and h(t) can be expressed equivalently in either of two calculus formulae given below.

$$S(t) = \exp\left[-\int_0^t h(u)du\right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)}\right]$$

3.9 Goals of Survival Analysis [19]

We now state the basic goals of survival analysis.

Goal 1: To estimate and interpret survivor and/or hazard functions from survival data.

Goal 2: To compare survivor and/or hazard functions.

Goal 3: To assess the relationship of explanatory variables to survival time.

3.10 ESTIMATION OF SURVIVAL FUNCTIONS ([19],[25],[8])

3.10.1Non Parametric Estimation in Survival Function

Non-parametric approach is that in which one is not restricted to fit a known distribution to the data. The Kaplan-Meier estimate is a mechanical process that will allow us to estimate the survival function from an investigation where censoring occurs. If censoring does not occur, we use the empirical estimate.

3.10.1.1 Empirical Estimation of the Survival Function [19]

Suppose we have a sample of survival times, where none of the observations are censored. Then the empirical survival function S(t) is

$$S(t) = \frac{\text{Number of individuals with survival times} \geq t}{\text{Number of individuals in the data survival time}}$$

The estimated survival function $\hat{S}(t)$ is assumed to be constant between two adjacent death times and so a plot of S(t) against 't' is a step function. The function decreases immediately after each observed survival time.

Example on **3.10.1.1** [8]

If the survival or waiting times in months of 11 claimants to be paid their benefits are as below:

11, 13, 13, 13, 13, 13, 14, 14, 15, 15, 17

Estimate the values of the survivor function at time 11, 13, 14, 15, and 17 months.

Table 3.1.: Solution: Survival Time and their functions

SURVIVAL TIME (T)	S(t)
11	$^{11}\!/_{11} = 1$
13	$^{10}\!/_{11} = 0.90$
14	$^{5}\!/_{11} = 0.45$
15	$^{3}\!/_{11} = 0.27$
17	$^{1}\!/_{11} = 0.09$

3.10.1.2 MAXIMUM LIKELIHOOD ESTIMATION [8]

We shall begin with the definition of the Likelihood Estimate Functions and then consider maximum likelihood function estimate of censored and truncated data and how it can lead us to compute the survival functions.

Definition: A likelihood function can be thought of as being the probability of getting the data that have been observed.

Suppose in the investigation of ‘n’ lifes, T_j is the complete future time of life j, j = 1, 2, …, n and tj denote the observed future life time of life j. assuming that future life times are independent continuous random variables then the likelihood function L is given by

$$L = \prod_{j=1}^{n} f_{T_j}(t_j)$$

Where $f_{T_J}(.)$ is the probability density function of the complete future life time of random variable T_j.

For independent and discrete random variables:

$$L = \prod_{j=1}^{n} \mathrm{P}\left(T_j = t_j\right)$$

The method of maximum likelihood is applied if we assume T_j to be identically distributed and belongs to a particular family.

Example: A lifetime random variable has a Gamma $(40, \theta)$ distribution. A random sample of 10 lifetimes resulted in a mean of 75.3.

Calculate the maximum likelihood estimate of θ based on these observations.

Solution

Let T_j denote the j^{th} life time random variable.

The probability density function of T_j is

$$f_{T_j}\left(t_j\right) = \frac{\left(t_j \big/ \theta\right)^{40} e^{-t_j/\theta}}{t_j \Gamma(40)}, t_j > 0$$ where, c = constant.

Taking the 'ln' on both sides,

$$\ln L = \ln c - 400 \ln \theta - \frac{1}{\theta}\sum_{j=1}^{10} T_j$$

Differentiating ln L with respect to θ ,

$$\frac{d}{d\theta}(\ln L) = \frac{-400}{\theta} + \frac{1}{\theta^2}\sum_{j=1}^{10} T_j$$

$$\frac{d}{d\theta}(\ln L) = \frac{1}{\theta}\left[-400 + \frac{1}{\theta}\sum_{j=1}^{10} T_j\right]$$

Setting the above equal to zero we obtain the maximum likelihood estimator

$$\hat{\theta} = \frac{1}{400}\sum_{J=1}^{10} T_J$$

But $\sum \hat{\theta} = 753 \times 1$

$$\sum \hat{\theta} = 753$$

$$\Rightarrow \quad \hat{\theta} = \frac{753}{400} = 1.88$$

Therefore, maximum likelihood estimator of θ is $\hat{\theta} = 1.882$

3.10.1.3 Theorem: Properties of Maximum Likelihood Estimators [8]

(i) All maximum likelihood estimators are asymptotically normally distributed, that is, as the sample size increases the distribution of the estimator approaches that of a normal distribution.

(ii) All maximum likelihood estimators are unbiased. That is, $E\left(\hat{\theta}\right) = \theta$

(iii) The asymptotic variance of a maximum likelihood estimator is given by the Cramer-Rao lower bound (C.R.L.B).

$$\text{C.R.L.B} = \frac{1}{-E\left[\dfrac{d^2 \ln L}{d\theta^2}\right]} = \frac{1}{E\left\{\left[\dfrac{d \ln L}{d\theta}\right]^2\right\}}$$

(iv) They are consistent. An estimator θ_n for a parameter θ is said to be consistent if for all $\delta > 0$, $\lim_{n\to\infty} P\left(\left|\theta_n - \theta\right| > \delta\right) = 0$

NOTE: Parameter Estimation Using Incomplete Data

Incomplete data may occur either through censoring or truncation.

3.10.1.4 Construction of the Likelihood Function for Censored Data ([8],[9],[22])

Suppose we want to carry out mortality study involving lives with the assumption that lifetime and censored lives are independent, then the likelihood function L is the product of the individual lifetime probability density functions.

$$L = \prod_{j=1}^{n} f_{T_j}\left(t_j\right)$$

Where T_j denotes the lifetime random variable for life j and t_j is the observed value of T_j, j = 1, 2, …, n.

Now suppose that some of the survival times are right censored and that the j^{th} life to be censored is censored at age c_j.

Our L is now:

$$L = \prod_{i \in D} f_{T_i}\left(t_i\right) \prod_{j \in c} \Pr\left(T_j > C_j\right)$$

$$L = \prod_{i \in D} f_{T_i}\left(t_i\right) \prod_{j = c} S_j\left(C_j\right)$$

Where: $S_j(.)$ = survival function for life j

D = set of lives that experienced the event during the study.

C = set of lives that were right censored.

Similarly, when the life times of the j^{th} life censored is now known exactly but is known to lie in the interval (C_j, d_j), then the likelihood function L becomes:

$$L = \prod_{L \in D} f_{T_i}\left(t_i\right) \prod_{j \in C} \mathrm{P}\left(\alpha_j < T_j < d_j\right)$$

$$L = \prod_{i \in D} f_{T_i}\left(t_i\right) \prod_{j \in C} \left[S_j\left(c_j\right) - S_j\left(d_j\right)\right] \qquad \text{(3a)}$$

Example 3.1 [8]

In a mortality investigation, lifetimes are right censored at times 10. The following set of observations was recorded from 7 independent lives: 4.35, 8.9, 10, 5.66, 10, 10, and 1.25. Assuming that the lifetime random variable of each of these lives follow an exponential distribution with parameter θ.

(i) Write down the likelihood function.

(ii) Calculate the maximum likelihood estimator of θ.

(iii) Obtain survival function of the lifetime random variable.

<u>Solution</u>

Let T_i denote the lifetime random variable for i^{th} life to die and T_j denote the lifetime random variable for the j^{th} life to be censored. The likelihood function L is:

$$L = \prod_{i\in D} f_{T_i}(t_i) \prod_{j\in C} P(T_j > 10)$$

$$L = \prod_{i\in D} \frac{1}{\theta} e^{-t/\theta} \prod_{j\in C} e^{-10/\theta}$$

$$L = \frac{1}{\theta^4} \exp\left[-\sum_{i\in D} t_j/\theta\right]\left(e^{-10/\theta}\right)^3$$

$$L = \frac{1}{\theta^4} e^{-20.16/\theta} e^{-30/\theta}$$

$$L = \frac{1}{\theta^4} e^{-50.16/\theta}$$

(ii) $\log L = -4\log\theta - \frac{50.10}{\theta}$

Differentiating with respect to θ.

$$\frac{d\log L}{d\theta} = \frac{-4}{\theta} + \frac{50.10}{\theta^2}$$

Setting $\frac{d\log L}{d\theta} = 0.$

$$\hat{\theta} = \frac{50.26}{4}$$

$$\hat{\theta} = 12.54$$

Thus the maximum likelihood estimate of θ is 12.54

(iii) $S(t) = \Pr(T > t)$

$$S(t) = 1 - \left[1 - e^{-t/12.54}\right]$$

$$S(t) = e^{-t/12.54};\ t > 0$$

3.10.1.4 Construction of the Likelihood Function for Truncated Data ([8],[9],[22])

We now consider the form of the likelihood function when left truncation is present.The likelihood function is the form:

$$L = \prod_i f_{T_i/U_i<T<V_i}(t_i)$$

$$L = \prod_i \frac{f_{T_i}(t_i)}{\Pr(U_i < T_i < V_i)}$$

If $V_i = \infty$, there is no right truncation V_i.

3.10.1.5 Kaplan-Meier Estimation of the Survivor Function ([8],[9],[22])

The **Kaplan-Meier estimator** (also known as the **product limit estimator**) estimates the survival function from life-time data. An important advantage of the Kaplan-Meier curve is that the method can take into account "censored" data — a loss from the sample before the final outcome is observed. When no truncation or censoring occurs, the Kaplan-Meier curve is equivalent to the empirical distribution.

A plot of the Kaplan-Meier estimate of the survival function a step function in which the estimated probabilities are constants.

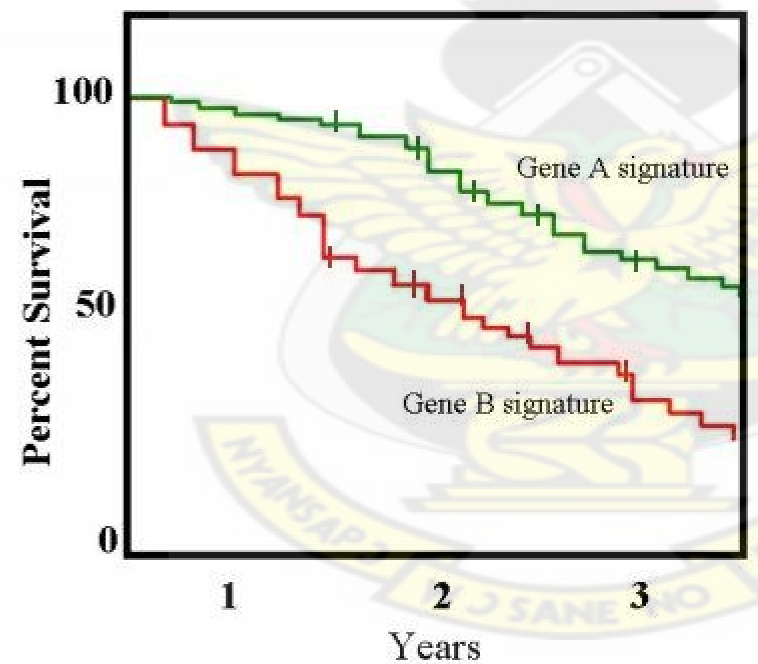

Figure 3.5: Kaplan-Meier plot for two conditions Gene A and B

An example of a Kaplan-Meier plot for two conditions associated with patient survival

On the plot, small vertical tick-marks indicate losses, where patient data has been censored.

In medical statistics, a typical application might involve grouping patients into categories, for instance, those with Gene A profile and those with Gene B profile. In the graph, patients with Gene B die much more quickly than those with gene A. After two years about 80% of the Gene A patients still survive, but less than half of patients with Gene B.

3.10.1.7 Formulation ([8],[9],[22])

The Kaplan-Meier estimator is the nonparametric maximum likelihood estimate of the survival function, *S*(*t*). It is a product of the form

$$\hat{S}(t) = \prod_{j=1}^{r} \hat{P}_j \quad \text{for } j = 1, 2, 3, \ldots, r$$

Where $\hat{P}_j = \frac{n_j - d_j}{n_j}$ is the estimated probability that an individual (or subject) survives through the interval t_j to t_{j+1}

Where j = 1, 2, …, r

r = total number of events in the whole period of observation.

n_j = number of subjects who survives just before time t_j (i.e. When there is no censoring). With censoring, n_j is the number of survivors less the number of losses (censored cases).

d_j = number of individuals who experienced the event at time t_j.

And S(t) = 1 if t < t_1 and S(t) = 0 if t > t_r, where t_r is the largest observation event time which is uncensored.

3.10.1.8 Variance and Standard Error of Kaplan-Meier Estimator ([8],[9])

The variance of Kaplan-Meier estimate of the survivor function is

$$V\left(\hat{S}(t)\right) = [S(t)]^2 \sum_{j=1}^{k} \frac{d_j}{n_j\left(n_j - d_j\right)}$$

<u>Proof</u>

The number of individuals who survive through the interval beginning at t_j can be assumed to have a binomial distribution with parameter *nj* and p_j, where p_j is the true probability of survival through the interval. The observed number who survive is n_j-d_j, and using the result that the variance of a binomial random variable with parameters n, p is np(1-p) the variance of $(n_j$-$d_j)$ is given by

$$\text{var}\left(n_j - d_j\right) = n_j p_j \left(1 - p_j\right)$$

Since

$$\hat{P} = \frac{\left(n_j - d_j\right)}{n_j}$$

The variance of $\hat{P}_j$ is $\text{var}\left(n_j - d_j\right) / n_j^2$ that is

$$\frac{\hat{P}\left(1 - \hat{p}_j\right)}{n_j} \qquad 3.2$$

The variance of $\hat{P}_j$ may then be estimated by

$$\frac{\hat{P}\left(1-\hat{p}_j\right)}{n_j}$$

Since

$$\operatorname{var}\left(g(x)\right) \approx \left\{\frac{d\,g(x)}{dx}\right\}^2 \operatorname{var}(x) \qquad 3.3$$

Using the above equation, the approximate variance of $\log \hat{P}_j$ is $\left(1-\hat{P}_j\right) \big/ \left(n_j \hat{P}_j\right)$

Which on substitution for $\hat{P}_j$ reduces to $\frac{d_j}{n_j\left(n_j-d_j\right)}$

But:

$$\operatorname{var}\left(\log \hat{S}(t)\right) = \sum_{j=1}^{k} \operatorname{var}\left(\log \hat{P}_j\right)$$

$$\Rightarrow \operatorname{var}\left(\log \hat{S}(t)\right) \approx \sum_{j=1}^{k} d_j \big/ n_j\left(n_j-d_j\right)$$

Further application of equation 3.2 yields

$$\operatorname{var}\left(\log \hat{S}(t)\right) = \frac{1}{\left[\hat{S}(t)\right]^2} \operatorname{var}\left(\hat{S}(t)\right)$$

$$\Rightarrow \operatorname{var}\left(\hat{S}(t)\right) \approx \left[\hat{S}(t)\right]^2 \sum_{j=1}^{k} \frac{d_j}{n_j\left(n_j-d_j\right)}$$

This result is known as Greenwood’s formula.

The standard error of the Kaplan-Meier estimate of the survival function is the square root of the estimator variance of the estimate, which is given by:

$$Se\left[\hat{S}(t)\right] \approx \hat{S}(t)\left\{\sum_{j=1}^{k}\frac{d_j}{n_j\left(n_j - d_j\right)}\right\}^{\frac{1}{2}}$$

For $t_{(k)} \le t \le t_{k+1}$

3.10.1.9 Pointwise Confidence Interval for the Survival Function [8]

We can use the Kaplan-Meier estimator and its standard error to provide a linear confidence interval *I(t)* for the survival function $\hat{S}(t)$ on the assumption that *S(t)* is normally distributed. That is:

$$I(t) = \left\{\hat{S}(t) - Z_{(1-\alpha/2)}Se\left[\hat{S}(t)\right], \hat{S}(t) + Z_{(1-\alpha/2)}Se\left[\hat{S}(t)\right]\right\}$$

The above confidence interval *I(t)* gives 100(1-α)% confidence of enclosing *S(t)* at a single point t, where α is the level of significance.

3.10.1.10 Estimate of the Hazard Function, $h(t)$ [8]

A natural way of estimating the hazard function of a survival data is to take the ratio of the number of deaths at a given death time to the number of individuals at risk at that time.

If there are d_j deaths at the j^{th} death time $t_{(j)}$, j = 1, 2, …, r and the n_j at risk at time t_j then the hazard function in the interval from t_j to t_{j+1} can be estimated as $h(t) = \frac{d_j}{n_j \tau j}$ for

$t_j < t < t_{j+1}$ and $\tau_j = t_{j+1} - t_j$

3.10.1.11 Estimate of the Cumulative Hazard Function, $H(t)$ [8]

The cumulative hazard at time t, $H(t) = -\log\left[S(t)\right]$ and so if *S(t)* is the Kaplan-Meier estimate of the survival function, then $\hat{H}(t) = -\log\left[S(t)\right]$ is an approximate estimate of the cumulative hazard to time t.

$\Rightarrow \hat{H}(t) = -\sum_{j=1}^{k} \log\left(\frac{(n_j - d_j)}{n_j}\right)$ for $t_k \le t \le t_{k+1}, k = 1,2, . \; n$ and $t_{(1)}, t_{(2)}, t_{(3)}, \ldots, t_{(r)}$ are the r-ordered death times with $t_{(r+1)} \le \infty$.

Example 3.3 ([8],[9])

The table below shows the data given for 18 women , all of whom were aged between 18 and 35 years and who had experience two previous pregnancies and therefore commences the use of a contraceptive bur some discontinue the use of the IUD for health complications. Discontinuation times that are censored are labeled with an asterisk.

The times in weeks till discontinuation of the use of an intrauterine device (IUD) are as follows:

10, 13*, 18*, 19, 30, 36, 38*, 54*, 56*, 59, 75, 93, 97, 104*, 107, 107*, 107*.

Construct a table showing S(t), $\lambda(t)$ and 95% confidence interval for S(t).

Table 3.2: A table showing S(t), $h(t)$ and 95% CI for S(t)

Time interval	τ_j	n_j	d_j	$\frac{n_j - d_j}{n_j}$	S(t)	$h(t)$	$Se(\hat{S}(t))$	95% Confidence Interval
[0,10)	10	18	0	1	1.0000	0.0000	0.0000	-
[10,19)	9	18	1	0.9444	0.9444	0.0062	0.0540	(0.839,1.000)
[19,30)	11	15	1	0.9333	0.8815	0.0061	0.0790	(0.727,1.000)
[30,36)	6	13	1	0.9231	0.8137	0.0128	0.0908	(0.622,1.000)
[36,59)	23	12	1	0.9167	0.7459	0.0036	0.1107	(0.529,0.963)
[59,75)	15	8	1	0.8750	0.6526	0.0078	0.1303	(0.397,0.908)
[75,93)	18	7	1	0.8571	0.5594	0.0079	0.1412	(0.283,0.836)
[93,97)	4	6	1	0.8333	0.4662	0.0417	0.1452	(0.182,0.751)
[97,107)	10	5	1	0.8000	0.3729	0.0200	0.1430	(0.093,0.653)
107	0	3	1	0.6667	0.2486	-	0.1392	(0.000,0.522)

In some cases, one may wish to compare different Kaplan-Meier curves. This may be done by several methods including:

- The Log rank Test
- The Wilcoxon Test or Breslow Test.

3.10.1.12 Log Rank Test [8]

The **logrank test** (sometimes called the **Mantel-Cox test**) is a hypothesis test to compare the survival distributions of two samples. It is a nonparametric test and appropriate to use when the data are right censored.

The log-rank test is constructed by considering separately each death time in two groups that is Group I and Group II. Suppose there are ‘r’ distinct death times $t_{(1)} < t_{(2)} < \ldots < t_{(r)}$

across the two groups and that at time t_{1j}, d_{1j} individuals in Group I and d_{2j} individuals in Group II die for j = 1, 2, …, r.

Unless two or more individuals in a group have the same recorded death time, the values of d_{1j} and d_{2j} will either be zero or unit. Consequently, at time t_{ij} there are $d_j = d_{1j} + d_{2j}$ deaths in total out of $n_j = n_{1j}-n_{2j}$ individuals at risk.

The test statistic for the log rank test is $\frac{U_L^2}{V_L}$, which has a chi-square distribution with one degree of freedom.

Where:

$$U_L = \sum_{j=1}^{r} \left(d_{1j} - e_{1j} \right)$$

$$e_{1j} = E\left(d_{1j} \right)$$

$$V_L = Var\left(U_L \right) = \sum_{j=1}^{r} V_{1j}$$

$$V_{1j} = \frac{n_{1j} n_{2j} \left(n_j - d_j \right)}{n_j^2 \left(n_j - 1 \right)}$$

3.10.1.13 Wilcoxon Test / Bresclow Test [8]

This test just like the log rank test is used to test the null hypothesis that there is no difference in the survival function for two groups of survival data.

The test statistics is $W_w = \frac{U_W^2}{V_W}$ which has chi-square distribution with one degree of freedom when the hypothesis is true.

Where:

$$U_W = \sum_{j=1}^{r} n_j \left(d_{1j} - e_{1j}\right)$$

$$V_W = \sum_{j=1}^{r} n_j^2 V_{1j}$$

$$V_{1j} = \frac{n_{1j} n_{2j} d_j \left(n_j - d_j\right)}{n_j^2 \left(n_j - 1\right)}$$

3.10.2 Some Survival Distributions [25]

Parametric survival models are constructed by choosing a specific **probability distribution** for the survival function. It is relatively easy to substitute one distribution for another, in order to study the consequences of different choices.

The choice of survival distribution expresses some particular information about the relation of time and any exogenous variables to survival. It is natural to choose a statistical distribution which has non-negative support since survival times are non-negative. There are several distributions commonly used in survival analysis, which are listed in the table below.

Table 3.3: Some Common Distributions Used in Survival Analysis

Distribution	**Survival function S(t)**
Exponential (special case of Weibull)	$e^{-\lambda t}$
Weibull	$e^{-(\lambda t)^{\gamma}}$
Gompertz	$e^{\lambda/\theta\left(1-e^{\theta t}\right)}$
Log-normal	$1-\Phi\left(\frac{\ln(t)-\mu}{\sigma}\right)$
Log-logistic	$\left[1+\left(t/\alpha\right)^{\beta}\right]^{-1}$

Where Φ is the cumulative distribution function of the standard normal distribution

3.11 Cox Proportional Hazards Model [19]

The model allows us to predict the hazard rates for individuals with different risk characteristics, such as sex smoking habits, marital status etc.

The Formula for the Cox PH Model:

$$\boxed{h(t,\mathbf{X}) = h_0(t)\exp\left[\sum_{i=1}^{p}\beta_i X_i\right]}$$

Where:

β_i is the coefficient of covariates

$h_0(t)$ is called the baseline hazard function.

X Denotes a collection of p explanatory variables

The model is semi parametric because $h_0(t)$ is unspecified.

Survival curves can be derived from the Cox PH model.

The larger the value of the hazard function the greater the potential for the event to happen.

Cox model survival function is given by the formula:

$$S(t,X)=\left[S_0(t)\right]^{\exp\sum_{i=1}^{p}(\beta_i X_i)}$$

Where:

β_i is the coefficient of covariates

$S_0(t)$ is called the baseline survival function.

X denotes a collection of p explanatory variables

$i=1,2,3,\ldots,p$

3.11.1 The Hazard Ratio (HR) [19]

In general, a hazard ratio (HR) is defined as the hazard for one individual divided by the hazard for a different individual. Thus,

$$\widehat{HR}=\frac{\hat{h}(t,\mathbf{X}^*)}{\hat{h}(t,\mathbf{X})}$$

where

$$\mathbf{X}^*=(X_1^*,X_2^*,\cdots,X_p^*)$$

and

$$\mathbf{X}=(X_1,X_2,\cdots,X_p)$$

denote the set of X's for two individuals

We now obtain an expression for the HR formula in terms of the regression coefficients by substituting the Cox model formula into the numerator and denominator of the hazard ratio expression:

$$\widehat{HR} = \frac{\hat{h}(t, \mathbf{X}^*)}{\hat{h}(t, \mathbf{X})} = \frac{\cancel{\hat{h}_0(t)}\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i^*}}{\cancel{\hat{h}_0(t)}\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i}}$$

$$\widehat{HR} = \frac{\hat{h}_0(t)\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i^*}}{\hat{h}_0(t)\, e^{\sum_{i=1}^{p} \hat{\beta}_i X_i}} = e^{\sum_{i=1}^{p} \hat{\beta}_i (X_i^* - X_i)}$$

Therefore, Hazard Ratio is written below:

$$\boxed{\widehat{HR} = \exp\left[\sum_{i=1}^{p} \beta_i (X_i^* - X_i)\right]} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(*)$$

3.11.2 Assumptions of the Cox Proportional Hazards Model [41]

The Cox Proportional Hazards Model is only appropriate if:

(i) Hazard curves (plots) do not cross.

(ii) Hazard ratio is independent of time: thus from equation (*),

(i) Baseline hazard function not involved in the HR formula.

(ii) Hazard ratio for two X’s are proportional:

$$h(t, \mathbf{X}^*) = \theta\, h(t, \mathbf{X})$$

Example on 3.11 ([28],[19])

A hazard rate model is dependent on two covariates, sex and alcohol consumption. The covariates take values

$$\begin{cases} 1 & \text{if life is a male} \\ 0 & \text{if life is a female} \end{cases}$$

X_{i2} = average number of units of alcohol consumed each week by life 'i'.

The parameters of the model are: $b_1 = 0.7$

$b_2 = 0.11$

The baseline hazard rate at age t is $h_0(t) = 0.00001 \times 1.1^t$

The model is therefore

$$h_i(t) = h_0(t)\exp\left[0.7X_{i1} + 0.11X_{i2}\right]$$

Calculate the hazard rate using the above model for

(a) A male life aged 55 who consume an average of 2 units of alcohol per week.

(b) A female life aged 50 who consume 6 units of alcohol each week.

<u>Solution</u>

(a) $h_1(55) = 0.00001 \times 1.1^{55}\exp\left[0.7 \times 1 + 0.11 \times 2\right]$

$$h_1(55) = 0.00001\, e^{0.7} e^{0.22}\, 189$$

$$h_1(55) = 0.00001 \times 2.0138 \times 1.2461 \times 189$$

$$h_1(55) = 0.00474$$

(b) Female 50.6 units of alcohol

$$h_2(5\ \ )=0.0\ \ 0\times 1.1^{5}1e^{0}\ \exp[0.71\times 0+0.1\ \times 16]$$

$$h_2(5\ \ )=0.0\ \ 0\ \times 1e^{0}1e^{0.6}\ 76$$

$$h_2(5\ \ )=0.0\ \ 0\ \times 11\times 1.97\ \ 3\ \ \ 5$$

$$h_2(5\ \ )=0.0\ \ 0\ \ 2\ \ 2\ \ 6$$

3.11.2 Obtaining the Partial and Estimating Parameters ([8],[9])

If there are no ties in the data and there are p parameters, then we will have n-ordered death times of $y_1, y_2, \ldots, y_n$.

Let 'j' be the life that dies at times y_j and R_j the number of people exposed to risk at time y_j. Then the contribution to the partial likelihood from the death occurring at time y_j is:

$$\frac{K_j}{\sum_{i\in R_j} K_i}$$

Where $K_j = \exp\left[b_1x_{i1} + b_2x_{i2} + \ldots + b_px_{ip}\right]$

The partial likelihood function LL is

$$LL = \sum_{j=1}^{n} \ln(K_j) - \sum_{j=1}^{n} \ln\left[\sum_{i\in R_j} K_i\right]$$

$$LL = \sum_{j=1}^{n}\sum_{k=1}^{p} b_k x_{jk} - \sum_{j=1}^{n} \ln\left(\sum_{i\in R_j} K_i\right)$$

Taking partial derivatives of the log-likelihood, and setting to zero, forms the simultaneous equations. The solutions to the equations give the maximum likelihood estimate of b_k.

3.12 Test of Parameter Significance ([8],[9])

Global test of all p parameters test:

$$H_0 : b_1 = b_2 = \ldots = b_p = 0$$

Global test involve testing whether one or more parameter equal zero,

$$H_0 = b_k = 0$$

3.12.1 Likelihood Ratio Test [8]

The Test Statistic for the Likelihood Ratio Test is as shown below:

$$x^2 = 2\left[LL\left(\hat{b}\right) - LL\left(\hat{b}^{\bullet}\right)\right]$$

Where $LL\left(\hat{b}\right)$ is the log-likelihood according to a parameter set $\hat{b}$. Parameter set $\hat{b}$ includes 's' particular parameters fitted to the data, whilst in set $\hat{b}^{\bullet}$ only u of these parameters have been fitted to the data *(u < s)*, $\hat{b}^{\bullet}$ will include the other *'s-u'* parameters but at a predetermined value as defined by H_0.

X^2 should therefore have a $\chi^2_{[s-u]}$ distribution, according to H_0 and is tested accordingly.

3.13 Parameter Selection for the Model [8]

STEPS

Perform global test, if not significant the model will include no covariates.

If significant then the following stages:

(1) Start with no covariate and test significance of including just one additional covariate.

(2) Repeat previous step for each potential covariate, add the most significant covariate to the model.

(3) Continue to add covariates to the model always choosing the most significant until no further significance parameters are found.

CHAPTER FOUR

ANALYSIS OF DATA

4.0 INTRODUCTION

In this section we will use a non parametric model called Kaplan-Meier method to plot the survival and the Hazard functions. Log Rank and Breslow tests would be used to test the equality of the survival curves. Then the Cox PH Model (semi parametric) will be considered for the parametric analysis to select among the covariates, those that have a higher explanatory power on the Hazard rate of dropout from school.

4.1 DESCRIPTION OF DATA: The data analysis consists of two covariates; the region of the pupil and gender of pupil. Dropout of school is indicated by a drop in enrolment; thus, a percentage difference in enrolment between two adjacent years which exceeds the threshold of 35%. The threshold of 10%, 20% causes a drop in enrollment in almost all the districts and 30% causes a drop in most of the districts and thus making the result trivial. Refer to Table A2 in the appendix for the percentage differences in annual enrollment levels.

For this study, data was collected on Basic Schools in 24 districts from the Three Northern Regions. The data spans seven years from 2000/2001 to 2006/2007 academic year. This data is a secondary data obtained from the Ministry of Education, National Headquarters-Accra.

4.2 CLASSIFICATION OF VARIABLES

The predictor variables selected to have the potential effect on the waiting time till a pupil drops out of school are as follows:

Table 4.1 Characteristics of Pupils

COVARIATES	LEVELS	NO. OF LEVELS
REGION	NORTHERN, UPPER EAST, AND UPPER WEST	3
GENDER	MALE, FEMALE	2

4.3 Definition (Dropout): Dropout of school simply means one who quits school. In general terms dropout means one who has withdrawn from a given social group or environment.

4.4 NON-PARAMETRIC DATA ANALYSIS

4.3.1 Kaplan-Meier (Product Limit) Estimator

We will look at the Kaplan-Meier curves for all categorical predictors. This will provide insight into the shape of the survival and the Hazard Functions. Then, we will evaluate whether or not K-M curves for two or more groups are statistically equivalent by using log-rank and Breslow tests.

4.4.1 Kaplan-Meier curves for the three Northern Regions

Case Processing Summary

Table4.2: Case Processing Summary for the Regions

Region	Total No Of Districts	No of Events	Censored	
			N	Percent
Northern Region	26	16	10	38.5%
Upper East Region	14	8	6	42.9%
Upper West Region	8	0	8	100.0%
Overall	48	24	24	50.0%

NOTE: *We have 24 districts in all but each district is considered twice; for male and female pupils. This resulted in the overall for Total N to be 48.*

From the table, column 1 from the left shows the different regions. Column 2 shows the number of districts in each region. Column 3 shows the number of districts in each region who experienced a drop in the enrolment of pupils exceeding the threshold from one stage to another at the basic school; which indicates dropout from school. Column 4 and 5 show the number of districts in each region whose drop in enrolment of pupils fell below the threshod.sk

The Northern Region has the highest dropout rate of 61.5% followed by Upper East with 57.1% and the Upper West with 0%.

4.4.1.1 Kaplan-Meier Survival Curves of Regions

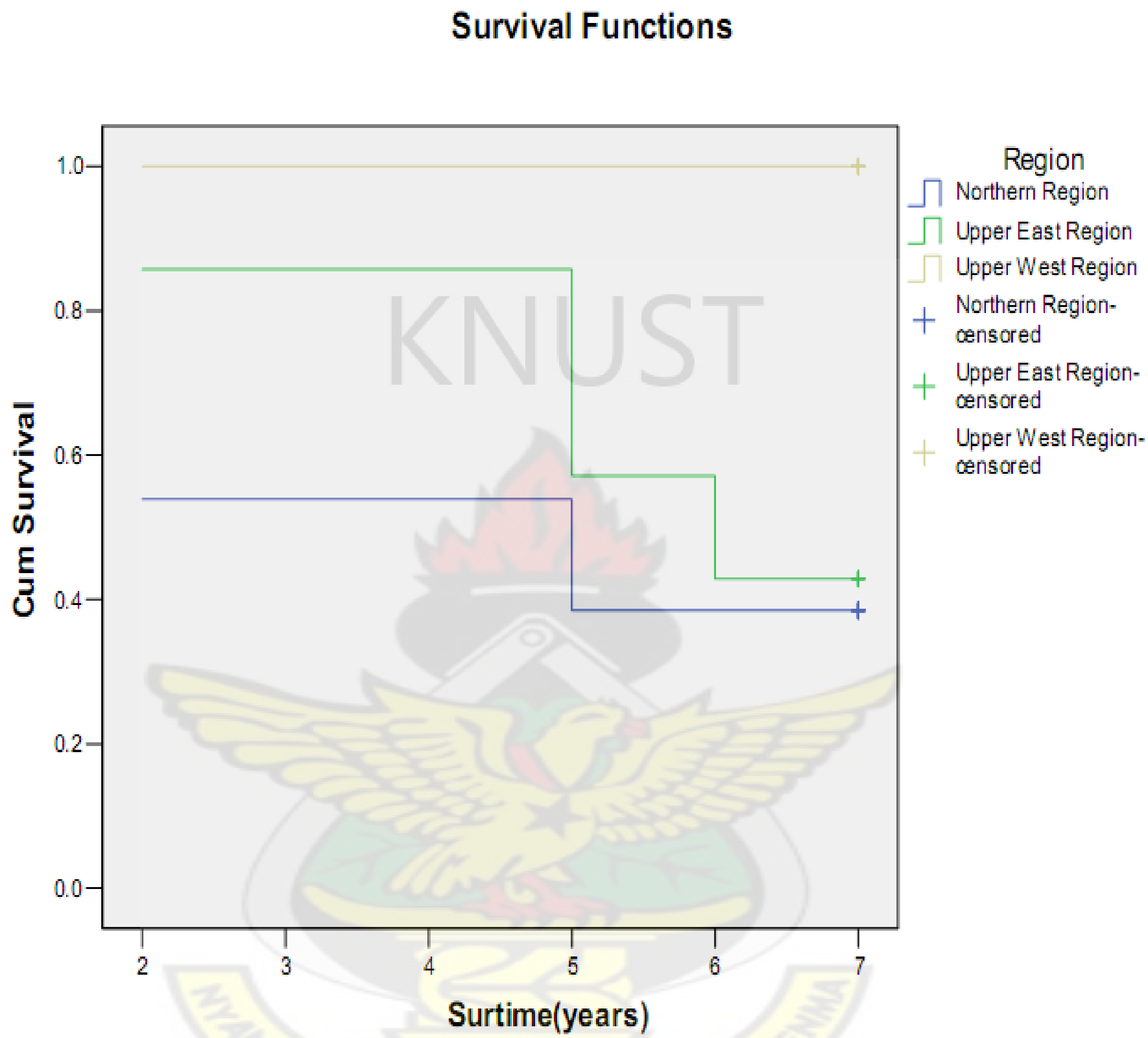

Fig.4.1. Kaplan-Meier survival plot of the three Northern Regions

With the Kaplan –Meier Method we have been able to get a graphical representation of probability of survival against time in years.

In the survival plot, drops in the survival curves occur whenever there is an indication of dropout from school (i.e. a drop in enrolment which exceeds the threshold).

However, Upper East and Northern Regions had a drop in enrolment levels between successive years exceeding the threshold at year two (primary two); indicating dropout of school. Also, between year two(2) and year five (5) we have the survival curves to be parallel to the time axis with a constant probability of staying in school (i.e. survival or a drop in enrolment levels not exceeding the threshold) equal 0.55 for the Northern Region, 0.85 for Upper East Region. There is another drop in survival at year five (primary 5) for the Northern and the Upper East Regions with probability of staying in school at approximately 0.4 and 0.6 respectively. Upper East had another drop in population at year six (primary 6).

It is observed from the graph that Upper West Region never had a drop in enrolment levels exceeding the threshold through out the study period (thus, indicating survival).

Comparing the three Regions from Fig.4.1, Northern Region has the lowest survival value of 0.4 followed by Upper East (0.45) and finally Upper West survival value at 1.0 at the end of the study.

In summary, the survival plots suggest that there is a high indication of pupils dropping out of school at year 5 for the Northern Region, and Upper East Region curve gives indication of dropout at years 5-6 and 6- 7.

4.4.1.2 K-M Hazard Curves of Regions

Figure 4.2 below, shows the hazard function which gives instantaneous potential of a child dropping out of school, given he/she survived (stayed in school) up to some time, t years (or has not yet dropped out from school).

The horizontal axis shows time in years for pupil to dropout of school, while the vertical shows the rate of a child dropping out of school; which is indicated by a drop in enrolment levels between adjacent stages exceeding the threshold.

A rise in step of the plot shows that there is a drop in population of pupils indicating dropout of school.

From Figure 4.2, it is observed that the hazard curve for Upper West is parallel to the time axis indicating a drop in population of pupils from one stage to another during the study period did not exceed the threshold. Hence, there is no dropout of school. Northern Region shows the highest rate of dropout of 0.95 at year 5 as compared to the Upper East region with dropout rate of 0.58 at the same year. Primary 6(year 6) recorded dropout rates of 0.95, 0.85 and 0.00 for the northern, Upper East and the Upper West Regions respectively. Comparing the three Hazard plots, the plot for the Northern Region consistently lies on the rest, followed by Upper East and finally Upper West Region. This indicates that the Northern Region has the highest hazard rate (potential) of pupils dropping out from school and Upper West Region experienced the lowest hazard rate of 0.0 dropouts from school.

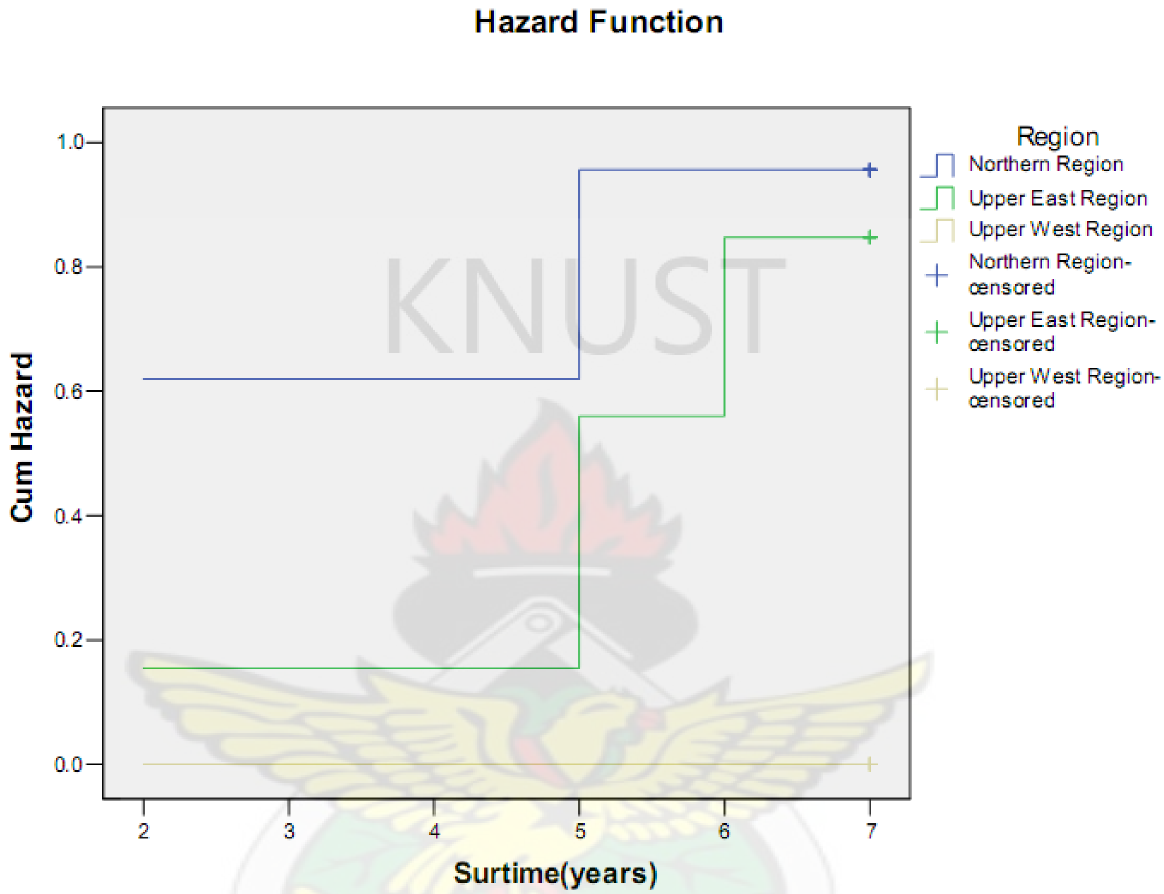

Figure 4. 2: K-M plot of Hazard function of the Regions

4.4.2 Overall Comparisons of the three Northern Regions

Table 4.3: Test of equality of survival distributions for the different levels of Region.

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	8.549	2	0.014
Breslow (Generalized Wilcoxon)	9.226	2	0.010

Tarone-Ware	8.956	2	0.011

Note: The null hypothesis is that there is no difference between the survival curves.

The table above gives a significant level of 0.014 for the Log Rank (Mantel-Cox) Test which indicates a significant difference between the regional survival curves. Breslow Generalized Wilcoxon) and Tarone-Ware also give significant levels of 0.010 and 0.11 respectively, which are less than 5% (0.05) ; and hence supporting the Log Rank Test conclusion.

4.4.3 Kaplan-Meier Survival and Hazard Curves For Gender

Case Processing Summary

Table 4. 4: Case Processing Summary for gender

Gender	Total N	No. of events	Censored	
			N	Percent
Boy	24	14	10	41.7%
Girl	24	10	14	58.3%
Overall	48	24	24	50.0%

NOTE: *We have 24 districts in all but each district is considered twice; for male and female pupils. This resulted in the overall for Total N to be 48.*

The table shows that of the districts in the three Northern Regions, 7districts and 5 districts experienced a drop in enrolment exceeding the threshold for boys and girls respectively; indicating dropout of school. Again, 12 districts out of the 24 districts in the entire three Northern Regions had a drop in the enrolment of pupils between adjacent years falling below the threshold; indicating no dropout of school.

The last two columns show the number and percentage of districts in the three Northern Regions whose male and female pupils are in school at the time of observation. Surprisingly, the males have a dropout rate of 58.3% and the females have a lower dropout rate of 41.7% over the entire regions.

4.3.3.1 K-M Survival plots of Gender

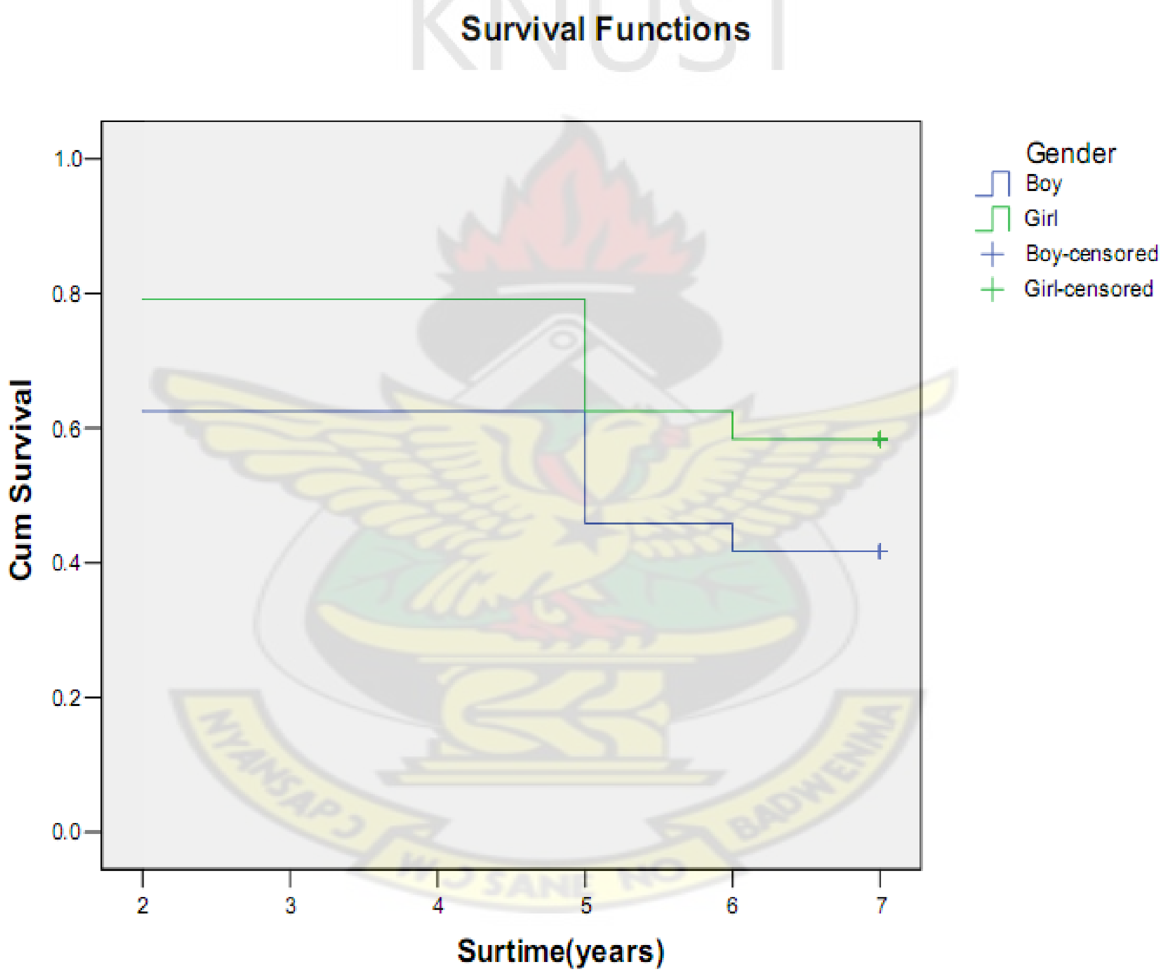

Figure 4. 3 K-M Survival curves for Gender.

From Figure 4.3 above, the survival curve for the girls lies consistently above that of the boys throughout the study period. This indicates that the female pupils' probability of staying in school these days is higher than that of males. This could be due to the retention fees paid at the end of each month to parent(s) of female pupils in the three Northern Regions. A female pupil qualifies for this fee if she gets 80% of the whole month's attendance.

However, it is also observed that there is an indication of dropout at years 5 -6 and 6-7 for both boys and girls

4.4.3.2 K-M Hazard plots of Gender

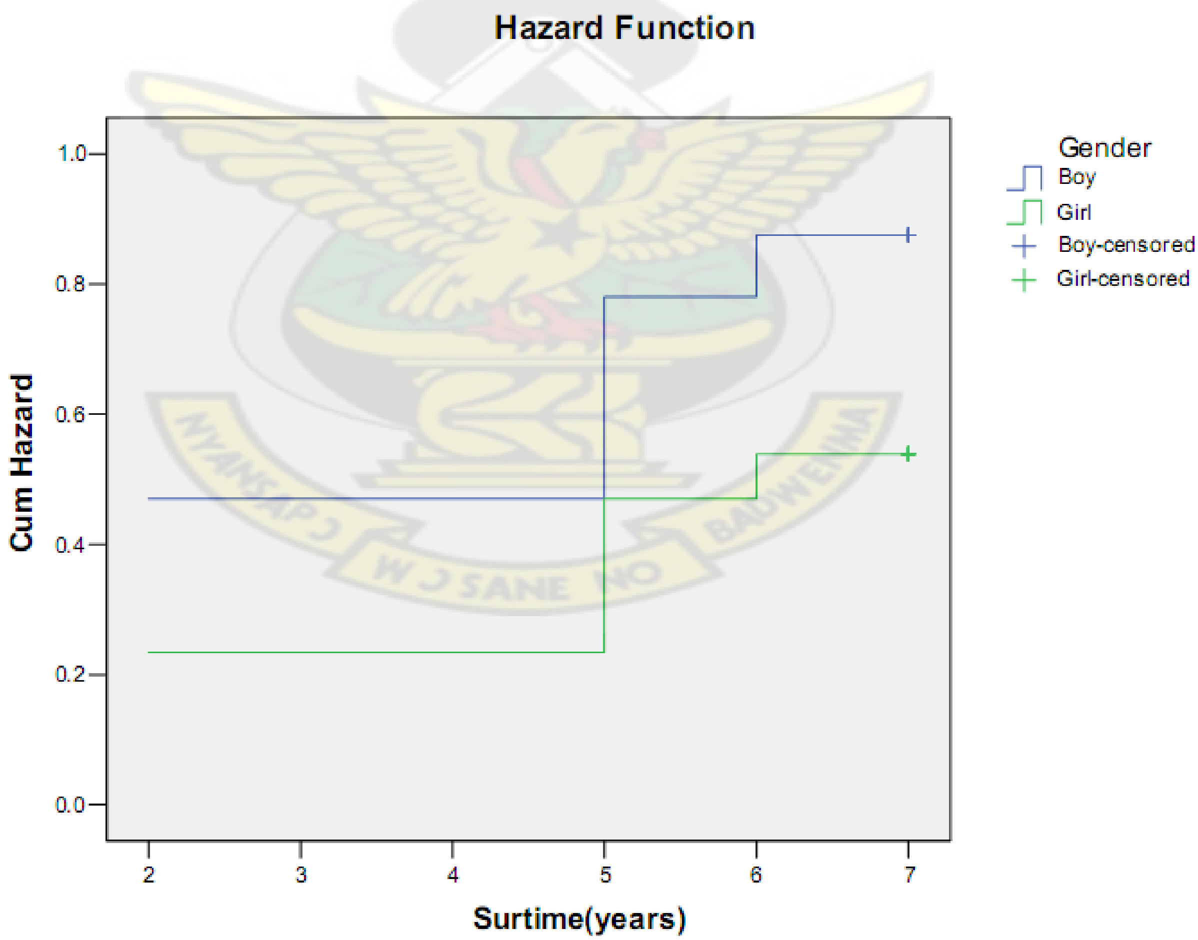

Figure 4.4 K-M Hazard Curves for Gender

This plot shows the hazard function which gives the potential of a male or female pupil dropping out from school (i.e. a significant drop in enrolment between adjacent stages) for gender in the districts of Northern Ghana.

The blue curve (male) which is above the green curve (female) suggests that the males have a higher hazard rate of dropping out of school than that of the females. At year five both boys and girls have the same potential of dropping out from school at an approximate rate of 0.5 as shown in Figure 4.4 above.

To determine whether these pictorial differences are not due to chance, we take a look at the comparison table below where we have three different and independent tests (methods) of statistical comparison scales.

4.4.3.3Overall Comparisons of Gender

Table 4.5: Test of equality of survival distributions for the different levels of Gender.

	Chi-Square	Df	Sig.
Log Rank (Mantel-Cox)	1.512	1	.219
Breslow (Generalized Wilcoxon)	1.693	1	.193
Tarone-Ware	1.620	1	.203

The table above provides overall tests of the equality of the survival times of the male and female pupils at the basic school in the three Northern Regions.

Since the significant values of the tests are all more than 5%, there is **not** enough evidence against the null hypothesis that the two survival times are the same.

Therefore, we conclude that there is no significant difference between the survival times of male and female pupils at the basic school level in Northern Ghana.

4.4.3.4 Means for Survival Time on Gender bases

Table 4. 6 Mean Survival Time for Gender

Gender	Mean			
	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Boy	4.750	.457	3.854	5.646
Girl	5.583	.404	4.792	6.374
Overall	5.167	.311	4.557	5.776

(a) Estimation is limited to the largest survival time if it is censored.

From Table 4.6, the mean waiting time for a male and a female to drop out of school are 4.750 and 5.583 years respectively, and this gives a numerical comparison of the male and female pupils' average waiting time to dropout from school at the basic level.

Since there is much proportion of overlap in the Confidence Intervals of the mean survival times, it indicates that there is approximately no difference in the average survival times of the two groups.

4.5 PARAMETRIC DATA ANALYSIS AND MODELLING

4.5.1 Statistical Identification of the Model.

In this section, we assume a semi-parametric model called Cox Proportional Hazards Model which is sometimes known as the Cox Regression. This model is widely used because the results from using the Cox Model will closely approximate the results for the correct parametric model.

For instance if correct model is Weibull; Cox Model will approximate Weibull. Also, if correct model is exponential; Cox Model will approximate exponential.

Thus, when in doubt, the Cox Model is a safe choice of model and one does not need to worry about whether the wrong parametric model is chosen. As described previously, the formula of the Cox Model is given as:

$$h(t,x) = h_0(t)\, Exp \sum_{i=1}^{p} \beta_i x_i$$

Where $h_0(t)$ = the baseline hazard function.

$X = (X_1, X_2, ..., X_p)$ explanatory variables /covariates.

i = 1, 2,…., p (number of covariates)

β_i = coefficient of covariates

The Cox PH Model selects among the covariates, those that have a higher influence on hazard rate of dropout from school.

In fitting the different levels of covariates to the model, we begin by estimating the parameters in the model; by using the Partial Maximum Likelihood Estimate method followed by determine the predictor variables in the model and finally, plots of Survival and Hazard curves for the mean of covariates.

4.4.2 Case Processing Summary of Data

Table 4.7 Case Processing Summary of Data

		N	Percent
Cases available in analysis	Event(a)	24	50.0%
	Censored	24	50.0%
	Total	48	100.0%
Cases dropped	Cases with missing values	0	0.0%
	Cases with negative time	0	0.0%
	Censored cases before the earliest event in a stratum	0	0.0%
	Total	0	0.0%
Total		48	100.0%

Table 4.7 above gives the total number of districts under study. Event cases are the number of districts which experienced a drop in enrolment levels exceeding the threshold; which indicates a dropout of school at the basic school level. Censored cases are the districts that did not experienced a drop in population exceeding the threshold. Hence, there is no enough indication of dropout from school.

Censored cases are not used in the computation of the regression coefficient but are used in the computation of the baseline hazard function.

4.4.2 Categorical Variable Coding

The Categorical Variable Coding is a useful reference in the interpretation of the regression coefficients for categorical covariates.

Table 4.8: Categorical Variable Coding

COVARIATES		MANUAL CODE	FREQUENCY	SPSS PROG. CODE
GENDER	MALE	0	24	1
	FEMALE	1	24	2
REGION	NORTHERN	1	26	1
	UPPER EAST	2	14	2
	UPPER WEST	3	8	3

4.5.3 Parameter Estimation and Model Selection

These are done together by the computer Software package; SPSS

The Omnibus test is a measure of how well the model performs. The chi-square change from previous step is the difference between the -2log likelihood model at the previous step and the current step.

4.4.3.1 Omnibus Tests of Model Coefficients

Table 4.9: Omnibus Test measures how well the model performs

Step	-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
		Chi-square	Df	Sig.	Chi-square	Df	Sig.	Chi-square	df	Sig.
1(a)	171.064	5.920	1	0.015	1.559	1	0.212	1.559	1	0.212

(a) Variable Removed at Step Number 1: Gender

(b) Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 177.848

(c) Beginning Block Number 2. Method = Forward Stepwise (Conditional LR)

The variable selection system includes a variable when the significance of the change is less than 5% and excludes or removes a variable when the significant value of the change is more than 10%.

4. 5.3.2 Variables in the Equation

Table 4.10: Explanatory Variable(s) in the model

		B	SE	Wald	Df	Sig.	Exp(B)	95.0% CI for Exp(B)	
								Lower	Upper
Step 1	Region	-0.814	0.350	5.393	1	0.020	0.443	0.223	0.881

After series of variable selection processes the model selection included only Region in the Cox PH Model with parameter coefficient of -0.814. The Hazard ratio, Exp(B) = 0.443 with its 95% CI as shown in Table 4.10.

From table 4.10, Region is selected as the only covariate because it has a significant value of 0.020 which is less than 5%.

4. 5.3.3 Variables not in the Equation

Table 4.11: Variables not in the Model

		Score	Df	Sig.
Step 1	Gender	1.565	1	0.211

(a) Residual Chi-Square = 1.565 with 1 df Sig. = 0.211

From table 4.11, gender is not included in the model because it has a significant value of 0.211 which is more than 10% (0.10). This indicates that, gender do not have significant influence in pupils dropout rate. In other words, it implies that according to this study, gender has not enough effect on the rate (potential) of a child dropping out at the basic school level.

4.5.3.4 Survival Table at the mean of covariates

Table 4.12 Survival Table

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
2	1.179	0.730	0.053	0.314
5	2.191	0.558	0.065	0.584

6	2.499	0.514	0.067	0.666

Table 4.12 shows the baseline Cumulative hazard at each year representing the rate of dropout with respect to the various times in school when no covariates are considered. For instance, the rate of a child dropping out of school at year two is 1.179.

However, it is observed that the Baseline Cumulative Hazard and the Cumulative Hazard values are directly proportional to time and the Survival values from the table are inversely proportional. This implies, as the years increase, the probability of a child staying in school decreases and the dropout rate increases. Thus, there is a high dropout rate of pupils as they move up the educational ladder.

The Cox PH survival and hazard graphs are at the appendix.

CHAPTER FIVE

5.0 DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 DISCUSSION

The following deductions are made from the research:

i) From the Kaplan-Meier plots, there is indication of dropout of school at year two (2) and year five(5) for the Northern and Upper East Regions. The Upper West Region never had a drop in population based on the threshold. However, drop out of school is approximately minimal for the first five years of the basic school education. Comparing the three regions, Northern region has the highest dropout rate. Hence we can conclude that most of the people at the Northern Ghana have some basic education up to primary five.

ii) Different levels of regions: There is a significant difference in the potential of a child dropping out of school among the three Northern Regions. In the test of equality of survival distributions for the different levels of Region, the log rank test gave a significant value of 0.014 which is less than 5% = 0.05 indicating a significant difference in the survival plots of the various regions. The region of a child has effective influence on his/ her drop out rate with the parameter coefficient of -0.814. From the K-M Hazard Plots, dropout rates at year five(primary 5) for the Northern, Upper East and Upper West Regions are 0.95, 0.85 and 0.00 respectively.

iii) Gender: The survival probabilities for both male and female pupils are not statistically different. Thus, the Log- Rank and Breslow tests gave significant values of 0.219 and 0.193 respectively. Again, the graphs from the Non-parametric analysis gave the probability of not dropping out of school as 0.45 and 0.65 for male and female respectively at year five(5) – primary 5. The rate of dropout was approximately 0.80 and 0.49 for male and female respectively at year five (primary 5).

However, according to this study the gender of a pupil has not enough influence on his/her rate of drop out of school. Thus, in the model selection process the significant value for gender was 0.212 > 10% =0.01, implying the effect of gender on the model is statistically insignificant.

5.2 CONCLUSION

From the above observations and discussion of results we can make the following conclusions:

Drop out of school is approximately minimal for the first five years of the basic school education at the three Northern Regions in Ghana. Hence, we can conclude that most of the people at the three northern regions have some basic education up to primary five.

The region of a pupil has influence on his/her potential of dropout of school with the Northern Region having the highest rate and the Upper West with the lowest rate.

There is statistically no difference in the dropout rate of male and female pupils at the basic school in the three Northern Regions of Ghana.

This research has brought to the notice of the public that the usual trend of "males having higher survival rate than females", is changing with interventions to reduce female dropout rate which are focused on rural areas in the three Northern Regions of Ghana

5.3 RECOMMENDATIONS

In view of the crucial role education plays in the advancement of society, every effort must be put to ensure that every child of school going age completes at least the nine – year Basic Education period.

The following recommendations are hereby made:

(i) Since it is observed that there is no significant difference between the dropout rate of boys and that of girls, interventions by Educational Policy Makers , the Government, NGOs ,etc to reduce dropout rate should not be focused on only female pupils but also on male pupils as well.

(ii) It was observed that there were indications of dropout mostly at primary two and primary five. This problem is recommended for further studies to identify the main cause(s).

(iii) Records on pupils must be properly kept in Schools, Districts, Regional and National Education Offices to be available and accessible; for researches to come out with findings that could assist Educational Policy Makers in making good decisions concerning education, and in the implementation of Educational Policies.

(iv) Incentive packages must be given be given to teachers at the three Northern Regions, especially those at the rural areas to motivate them give up their best. Sometimes a pupils performance or a teacher‘s attitude can drive the child out of school.

REFERENCES:

[1] Africa Renewal. http://www.org/ecosocdev/geninfo/afrec/Vol22no3/223-closing-ghanas-poverty-gap.hmtl. visited 2/07/09.

[2] Akorfa Sottie C. (2008); School Dropout in Ghana: Experiences, Perspectives, and Policy Direction. (Ghana Government) .© Queen's University Belfast 2009, University Road Belfast, BT7 1NN, Northern Ireland, UK

[3] America Chronicles. http://www.americanchronicles.com/articles/view/103751, Fighting hunger in Ghana with school meals. Visited 29/08/09.

[4] Article on underlying causes of High School Dropout. www.gafcp.org/pubs/rep/causeshsdropout.doc. Visited 13/12/09

[5] Andrew Hahn and J. Danzberger(1987), Dropouts in America: Enough Is Known for Action, Institute for Educational Leadership, 1987, p. 6.)

[6] B Imoro (2009); Dimension Of Basic School Dropouts In Ghana: The Case Of Asutifi District. Journal of Science and Technology (Ghana) >Vol 29, No 3

[7] Bennett S. and Whitehead, J.R. (1981) Fitting logistics and Log-logistic regression models to censored survival data using GLIM. GLIM Newsletter, 4, 12-9. Collection 5, 3.

[8] Boateng A.S (2008). Predictors of the Hazard Rate of surrender of a Universal Life Insurance Policy (Cox Proportional Hazards Approach), M.Sc. Thesis. unpublished

[9] Braimah I. and Oduro-Ofori E. (2005), Basic school dropout in Ghana: a case study of the Amansie West district. Journal Of Science And Technology(Ghana) Vol. 25 (1) 2005: 67-76

[10] Debate on Capitation Grant http://www.isodec.org.gh/Papers/debateoncapitationgrant.html., 30/10/08, 6:30 pm)

[11] Distributive function. http://en.wikipedia.org/wiki/ distributive_ function. (Tuesday 4th Nov. 2008)

[12] Finn, J.D. (1989), "Withdrawing from school." Review of Educatoinal Research, 59(2):117-142

[13] Finn, J.D. (1993), School engagement and students at risk. Washington, DC: National Center for Education Statistics.

[14] Ghana Districts Repository. http://www.ghanadistricts.gov.gh, visited 14/03/09

http://www.indepth.network.org, 30/10/08, 6:30 PM)

[15] (2004), http:/www.dropoutprevention.org/stats/whos_risk/sit_riskhtm?source=arise, Quick facts: Situations that put youth at risk. National Dropout Prevention Center (NDP/C)….Visited 14/03/09

[16] Human Rights and Poverty Seminar. Friday 20 August 2004 10:11. Accessed on 23/05/09

[17] J. Rebecca and S. Kyle, (2001); The Determinants of Girls' Educational Enrollment in Ghana. Publications, Department of Applied Economics and Management, Warren Hall, Cornell University, Ithaca, NY 14853-7801.

[18] J.G. Ampiah [a] and C. Adu-Yeboah [b] (2009); Mapping the incidence of school dropouts: a case study of communities in Northern Ghana. Comparative Education, Volume 45, pages 219 -232

[19] Kleinbaum D. and Klein M. (2005). Survival Analysis: A Self-Learning Text, 2nd ed. Springer Science+Business Media, Inc., New York, USA, 1,4-7.

[20] Kwame Akyeampong, CICE Hiroshima University, Journal of International Cooperation in Education, Vol .7, No. 1, (2004) pp.44. .52 (23rd October, 2008, 11:30am)

[21] Pablo Lavado and José Gallegos1(2005), "Schooling dropout in Peru: a framework using duration models". Centro de Investigación de la Universidad del Pacífico](www.test.aup.edu/lacea 2005/system/step2_php/papers/lavado_plav.pdf)

[22] Rolleston C.(2009); The determination of exclusion: evidence from the Ghana Living Standards Surveys 19912006. Comparative Education, Vol. 45, No. 2. (2009), pp. 197-218. http://www.aem.cornell.edu/research/workpaper3.html.

[23] Shiyuan C. and Wallace S. (2008), Determinants of Education Duration in Jamaica. International Studies Programme Working Paper 08-03. Georgia State Universty.

[24] Statement by National Education Collation Campaign http://www.liberationafrique.org/spip.php?page=article_pdf&id_article=866

[25] Survival Analysis. http://en.wikipedia.org/wiki/ survival_ analysis. (Tuesday 4th Nov. 2008)

[26] S. Govande (2008). Article on issues of dropout , India Wed, Jan 23, 2008 16:18:45 IST

[27] The development of Education National Report of Ghana by the Basic Education Division Ghana Education Service; June 2007. (ghana schs.pdf ,30/10/08, 6:30 pm)

[28] Unesco (1998), Wasted opportunities: when schools fail. Repletion and dropout in primary schools Paris.

[29] Schwartz, Wendy (1995), <u>“School Dropout: New Information About an Old Problem.”</u> National Center For Education Statistics. 1995. (*http://eric-web.tc.columbia.edu/digests/dig109.htm)

APPENDIX

Table A1: Model if Term Removed

Term Removed		Loss Chi-square	df	Sig.
Step 1	Region	6.783	1	.009

Table A2: Covariate Means

	Mean
Region	1.625
Gender	1.500

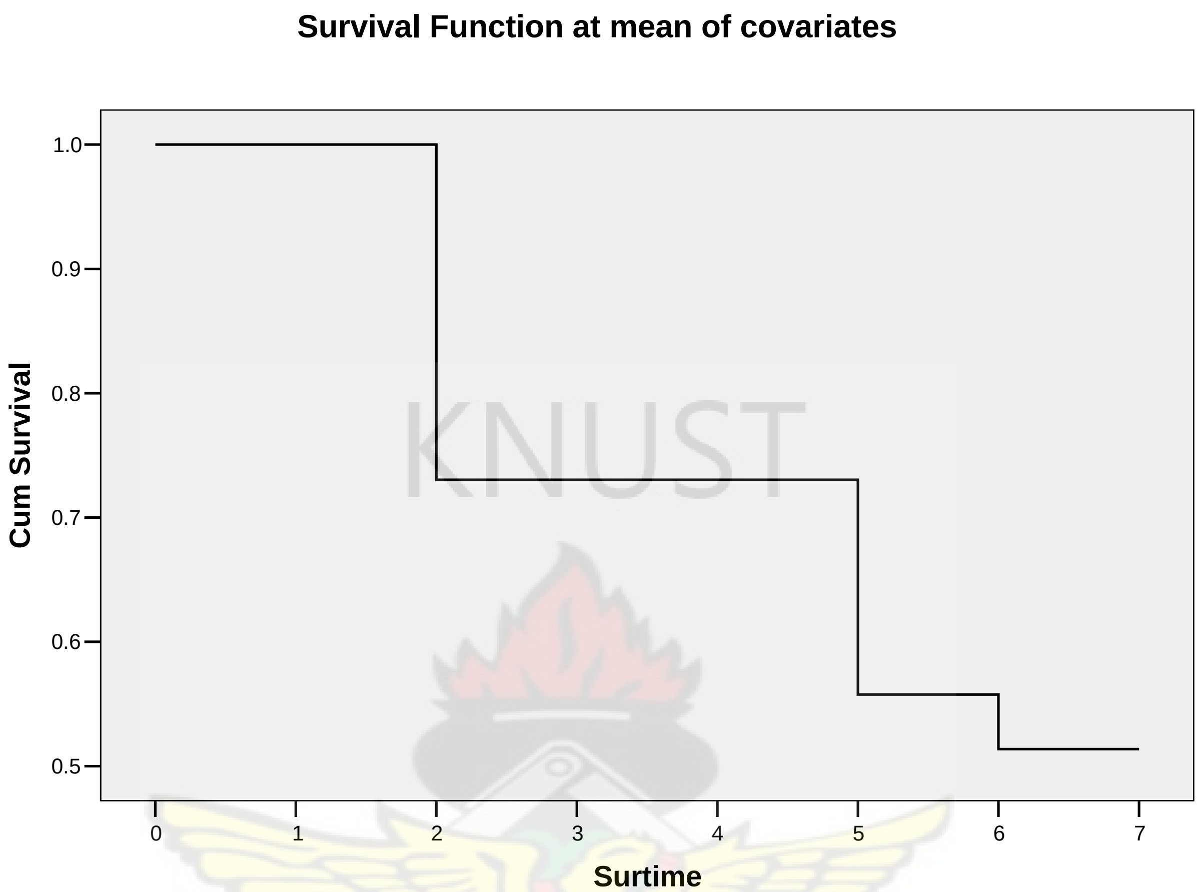

Figure A1: Plot of Survival Function at mean of Covariate

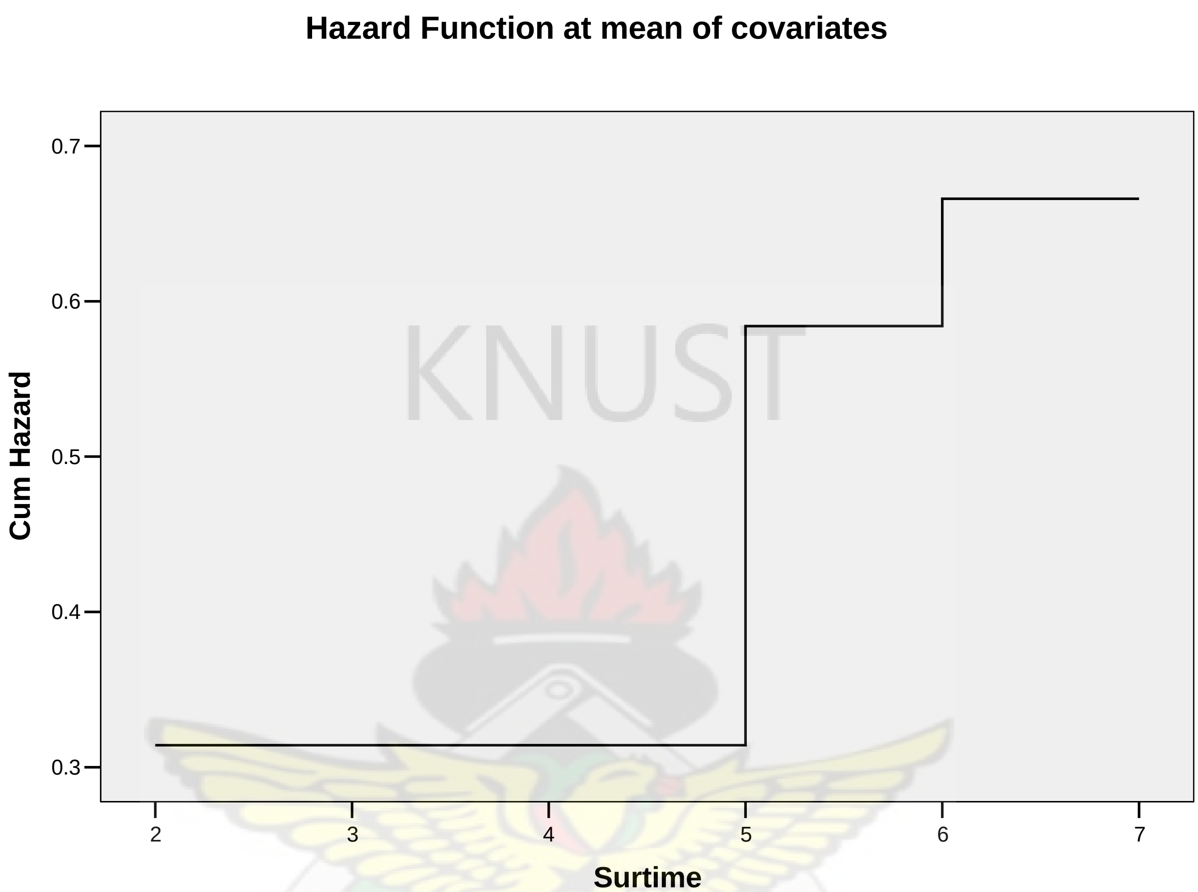

Figure A2: Plot of Hazard function at mean of covariates

Table A

Region		Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
				Estimate	Std. Error		
Northern Region	1	2.000	Dropped	.	.	1	25
	2	2.000	Dropped	.	.	2	24
	3	2.000	Dropped	.	.	3	23
	4	2.000	Dropped	.	.	4	22
	5	2.000	Dropped	.	.	5	21
	6	2.000	Dropped	.	.	6	20
	7	2.000	Dropped	.	.	7	19
	8	2.000	Dropped	.	.	8	18
	9	2.000	Dropped	.	.	9	17
	10	2.000	Dropped	.	.	10	16
	11	2.000	Dropped	.	.	11	15
	12	2.000	Dropped	.538	.098	12	14
	13	5.000	Dropped	.	.	13	13
	14	5.000	Dropped	.	.	14	12
	15	5.000	Dropped	.	.	15	11
	16	5.000	Dropped	.385	.095	16	10
	17	7.000	Censored	.	.	16	9
	18	7.000	Censored	.	.	16	8
	19	7.000	Censored	.	.	16	7
	20	7.000	Censored	.	.	16	6
	21	7.000	Censored	.	.	16	5
	22	7.000	Censored	.	.	16	4
	23	7.000	Censored	.	.	16	3
	24	7.000	Censored	.	.	16	2
	25	7.000	Censored	.	.	16	1
	26	7.000	Censored	.	.	16	0

Upper East Region	1	2.000	Dropped	.	.	1	13
	2	2.000	Dropped	.857	.094	2	12
	3	5.000	Dropped	.	.	3	11
	4	5.000	Dropped	.	.	4	10
	5	5.000	Dropped	.	.	5	9
	6	5.000	Dropped	.571	.132	6	8
	7	6.000	Dropped	.	.	7	7
	8	6.000	Dropped	.429	.132	8	6
	9	7.000	Censored	.	.	8	5
	10	7.000	Censored	.	.	8	4
	11	7.000	Censored	.	.	8	3
	12	7.000	Censored	.	.	8	2
	13	7.000	Censored	.	.	8	1
	14	7.000	Censored	.	.	8	0
Upper West Region	1	7.000	Censored	.	.	0	7
	2	7.000	Censored	.	.	0	6
	3	7.000	Censored	.	.	0	5
	4	7.000	Censored	.	.	0	4
	5	7.000	Censored	.	.	0	3
	6	7.000	Censored	.	.	0	2
	7	7.000	Censored	.	.	0	1
	8	7.000	Censored	.	.	0	0

Table A4: Percentage Differences in annual enrollment levels between two adjacent years from 2000/2001 to 2006/2007 Academic years.

Districts	P1-P2		P2-P3		P3-P4		P4-P5		P5-P6		P6-JHS1	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Bole (N/R)	**35.74**	32.62	7.65	10.14	8.53	12.42	55.66	**50.95**	3.73	0.25	-5.71	6.66
East Gonja(N/R)	32.08	21.86	5.70	13.03	6.10	11.66	3.79	-7.03	3.75	19.36	23.33	25.82
East Mamprusi(N/R)	28.15	23.03	-4.25	-1.10	11.43	13.51	**61.06**	**61.55**	2.69	6.86	10.20	26.47
Gushiegu-Karaga(N/R)	**35.50**	33.01	15.99	16.67	3.47	5.04	-0.54	-10.44	15.38	22.55	25.10	23.98
Nanuba(N/R)	**37.78**	22.01	-3.80	11.29	-1.32	2.78	2.93	2.71	-2.10	6.26	11.98	23.93
Saboba-Chereponi(N/R)	**35.68**	22.78	7.32	16.58	1.27	5.44	6.07	0.00	6.46	9.51	2.20	4.28
Savulugu-Nanton(N/R)	14.73	10.61	7.73	5.94	-1.81	-14.36	5.78	14.60	2.99	5.32	20.84	23.83
Tamale(N/R)	19.07	12.16	2.83	4.40	-2.96	-1.92	-6.31	-8.14	5.96	9.00	0.11	4.57
Tolon-Kumbungu(N/R)	**35.46**	22.09	17.20	19.62	3.08	11.82	14.70	4.32	-2.73	12.99	3.15	7.27
West Gonja(N/R)	27.43	23.54	11.85	10.81	2.09	-1.91	**45.11**	**53.75**	14.59	14.79	-4.86	-6.81
West Mamprusi(N/R)	23.66	22.48	1.14	0.36	-1.08	1.09	1.60	2.93	10.92	11.23	-0.17	-3.19
Yendi(N/R)	**42.75**	**38.35**	2.56	0.36	11.49	18.96	-4.07	-19.54	17.36	39.93	-3.08	-5.14
Zabzugu-Tatale(N/R)	**44.28**	**36.60**	-2.10	7.76	4.63	2.62	0.68	6.37	-6.83	13.16	29.07	21.42
Bawku East(U/E)	17.21	11.24	4.76	11.12	4.72	1.34	**43.48**	**42.67**	0.15	0.38	22.47	25.29

Bawku West(U/E)	8.90	10.87	10.57	17.96	3.60	7.16	7.59	5.53	4.55	5.08	8.61	5.19
Bolgatanga(U/E)	27.12	20.77	0.58	3.12	3.58	7.95	**43.60**	**37.58**	13.10	9.14	0.94	-6.98
Bongo(U/E)	26.66	19.33	5.94	13.73	15.67	6.65	10.30	10.51	10.53	11.66	-2.43	-7.43
Builsa(U/E)	30.58	30.00	14.47	19.26	6.75	-1.82	6.54	6.56	17.32	22.24	12.04	4.69
Kassena-Nankana(U/E)	19.88	21.78	12.26	11.13	4.96	5.31	8.86	11.59	2.75	9.99	14.15	7.05
Jirapa-Lambussie(U/E)	20.12	11.16	11.88	15.13	4.89	4.66	-81.05	-85.52	**50.70**	**50.18**	-7.79	-9.33
Lawra(U/W)	23.10	19.22	4.89	11.06	1.52	-11.23	4.64	4.04	0.51	12.27	-8.76	-16.98
Nadawli(U/W)	23.49	16.89	16.21	16.70	3.30	-0.25	12.07	8.69	7.34	5.30	-10.07	1.14
Sissala(U/W)	19.32	11.49	9.20	16.44	-0.82	-2.26	-1.29	-0.28	6.69	11.11	4.33	-1.96
Wa(U/W)	23.72	15.85	1.69	5.51	5.49	6.31	9.54	0.00	-9.06	4.43	-11.30	-7.43

Why the chosen threshold of 35%

The table above gives the percentage (%) differences in enrollment levels between two adjacent years for the various districts in the three Northern Regions of Ghana. A percentage difference which exceeds the threshold of 35% is considered as a drop in enrollment and below the threshold is not considered as a drop in enrollment. The assumption is made because if you scan through the values in the table, a threshold at 10%, 20% cause almost all the districts to experience a drop in enrollment at the end of year one and thus making the result trivial. Especially, almost all the districts in the Upper West Region would experience a drop in Enrollment. Also, a threshold of 30% causes most of the districts to get a drop in enrollment which does not give optimal result to the situation at hand. After series of considerations the most likely feasible result was the one with the threshold of 35%. Since the districts are needed for the comparison of the three Regions.

Table A5: Data on Basic Schools Annual Enrollment levels for the various districts in the three Northern Regions of Ghana (2000/01 to 2006/07 academic years).

Districts	Prim_1		Prim_2		Prim_3		Prim_4		Prim_5		Prim_6		JHS_1	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Bole	2115	1596	1359	1075	1255	966	1148	846	509	415	490	405	518	378
East Gonja	2765	2347	1878	1834	1771	1595	1663	1409	1600	1508	1540	1216	1181	902
East Mamprusi	3666	2735	2634	2105	2746	2147	2432	1857	947	714	922	665	828	489
Gushiegu-Karaga	1755	1030	1132	690	951	575	918	546	923	603	781	467	585	355
Nanuba	2922	1976	1818	1541	1887	1367	1912	1329	1856	1293	1895	1212	1668	922
Saboba-Chereponi	1847	1484	1188	1146	1101	956	1087	904	1021	904	955	818	976	783
Savulugu-Nanton	1684	961	1436	859	1325	808	1349	924	1271	789	1233	747	976	569
Tamale	5458	4393	4417	3859	4292	3689	4419	3760	4698	4066	4418	3700	4413	3531
Tolon-Kumbungu	2981	1675	1924	1305	1593	1049	1544	925	1317	885	1353	770	1311	714
West Gonja	2173	1767	1577	1351	1390	1205	1361	1228	747	568	638	484	669	517
West Mamprusi	1720	1428	1313	1107	1298	1103	1312	1091	1291	1059	1150	940	1152	970
Yendi	2933	1820	1679	1122	1636	1118	1448	906	1507	1083	1490	895	1536	941
Zabzugu-Tatale	1617	1240	901	786	928	725	885	705	879	661	939	574	666	451
Bawku East	4724	3496	3911	3103	3725	2758	3549	2721	2006	1560	2003	1554	1553	1161
Bawku West	1090	1012	993	902	888	740	856	687	791	649	755	616	690	584
Bolgatanga	4048	3573	2950	2831	2933	2743	2828	2525	1595	1576	1386	1432	1373	1532

Bongo	2202	2095	1615	1690	1519	1458	1281	1361	1149	1218	1028	1076	1053	1156
Builsa	1573	2040	1092	1428	934	1153	871	1174	814	1097	673	853	592	813
Kassena-Nankana	3329	3251	2667	2543	2340	2260	2224	2140	2027	1892	1951	1703	1675	1583
Jirapa-Lambussie	1655	1451	1322	1289	1165	1094	1108	1043	2006	1935	989	964	1066	1054
Lawra	1437	1488	1105	1202	1051	1069	1035	1189	987	1141	982	1001	1068	1171
Nadawli	1750	1758	1319	1461	1122	1217	1085	1220	954	1114	884	1055	973	1043
Sissala	1159	1436	935	1271	849	1062	856	1086	867	1089	809	968	774	987
Wa	3254	2612	2482	2198	2440	2077	2306	1946	2086	1946	2275	1858	2532	1996
Total Enrollment	**59857**	**48664**	**43647**	**38698**	**41139**	**34934**	**39477**	**33522**	**33848**	**29765**	**31539**	**25973**	**29828**	**24602**

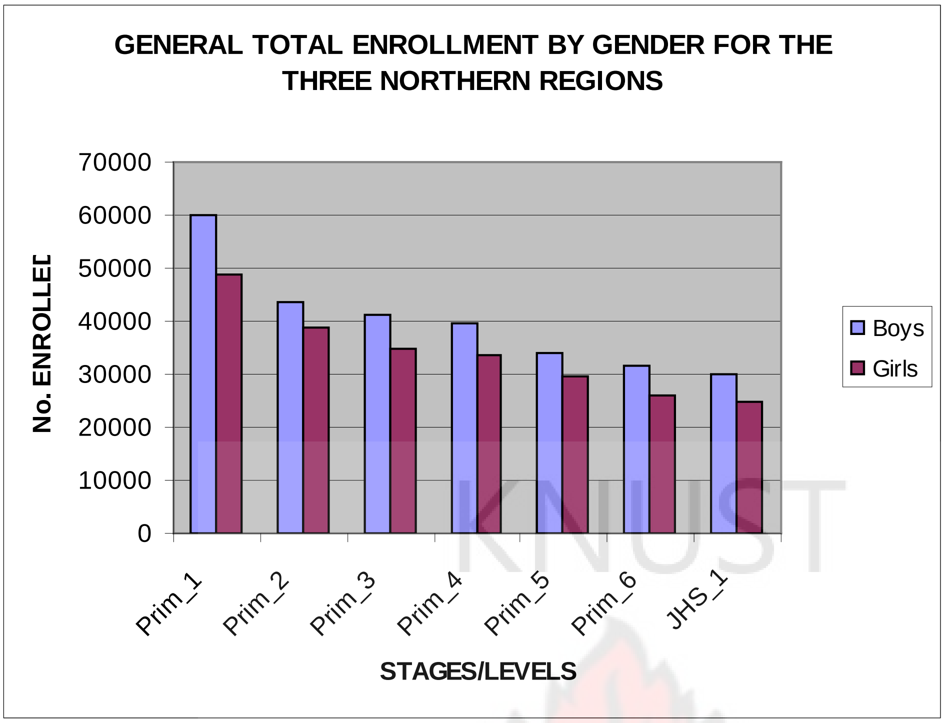

Figure A3: Graph of enrollment by gender in the three (3) northern regions of Ghana

From the graph it is indicated that the general enrollment of boys is greater than that of the girls but the girls maintain their number better than that of the boys. In other words the drop in population of the boys is little bit sharper than that of the girls.