KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI

MODELING HEALTH INSURANCE CLAIM SEVERITY IN NHIS USING PARAMETRIC PROBABILITY DISTRIBUTION

(A Case Study OF Amansie East Municipal)

BY

Essuman Felicia

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MSc. ACTUARIAL SCIENCE

April, 2016

DECLARATION

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

Essuman Felicia

Student(PG8733112)	....	....
Student	Signature	Date

Certified by:

Dr. A.Y. Omari Sasu	....	....
Supervisor	Signature	Date

Certified by:

Prof. S. K Amponsah	....	....
Head of Department	Signature	Date

DEDICATION

I dedicate this work to God Almighty and my family for their prayer and support

ABSTRACT

General insurance companies face two major problems when they want to use past or present claim amount in forecasting future claim severity. First they have to find the appropriate probability distribution for the large volumes of claim amount. Then test how best these distributions fits the claim data The purpose of this study is to fit a particular distribution suitable for the National Health Insurance Scheme in the Amansie

East Municipality. Secondary data was collected from Amansie Municipal Insurance Scheme. The exploratory data analysis technique was used to assist in the identification of the family of distribution which the data might follow. Akike Information Criterion and KolmogorovSmirnov used to test the goodness of fit. The diagnostic test probability plot was to used to graphically demonstrate the goodness of fit the distribution. It was found that the Lognormal distribution was appropriate distribution for Fee For Service for all categories of service in the district. However the Burr distribution were considered to be the best distribution for G-DRG in the District level, CHPS compound, Community clinics, Health centers and CHAG hospital. Also lognormal distribution was the best for G-DRG modeling for Private, Chemical and Pharmacy shops and finally the fisk distribution for Public clinics and Maternity. Management at all municipal and district health insurance schemes should be able to apply the appropriate statistical distribution used in this research for management policy prescription to improve their performance.

ACKNOWLEDGMENTS

No vision or great work is accomplished without help from other people. Many individuals have supported and contributed to this work and I am grateful to each of them Firstly, I am grateful to the Almighty God who has kept us till now, equipped us with his wisdom and understanding. Secondly, I am forever grateful to my supervisor Dr. A.Y. Omari Sasu for his piece of advice, and constructive criticism and the time he had to supervise this work in spite his busy schedule. Also am so much grateful to Albert Ayi Ashiagbor, Anyanumeh Herbert, Anyomi Vincent, and the NHIS management staff at Amansie. Finally to all my

siblings

CONTENTS

LIST OF ABBREVIATIONS

NHIS .... National Health Insurance Scheme FFS

.... Fee For Service

AIC .... Akakais Information Criterion

G-DRG .... Ghana-Diagonostic Related Groupings

KS ....Kolmogorov-Smirnov

CHAG ....Christian Health Association of GHANA

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Background of the study

Globally, the complexity of the insurance industry is not in doubt (Ashiagbor,2014). The contribution of health insurance to national development, its regulations, level of competition due to globalization have attracted policy makers and academic researches into the evaluation of the appropriate models for available data set in claim management in the health insurance sector (Migon,& Moura,2005; Gschlobl & Czado, 2007; Goovaerts, Dhaene, & Rachedi, 2010). The enactment of the Ghana National Health Insurance (NHIS) Act 2003, Act 650 and subsequent amendment in 2012 has led to a significant change in health financing. The Act stipulates among other things that any resident of Ghana must belong to the District Mutual Scheme, the Private Mutual Scheme or the Commercial Insurance Scheme. The expenditure on NHIS claims has increased steadily since 2011, and with a total of GHC 968.5 million spent on claims in the year 2014. That is, the amount has increased by more than 75% in nominal terms over the four years (see Table 1).

Table 1.1: NHIS expenditure by line items 2010-2014.

	2010	2011	2012	2013	2014	% of ttl in 2014
Claim	395.06	548.71	616.21	783.36	968.48	75.6%
Admin & logistic support	12.04	17.2	6.93	4.31	4.48	0.3%
Support to MOH	75.52	147.33	74.67	31.68	29.16	2.3%
Operating	31.26	36.75	71.35	140.02	128.46	10%
NHIS ID Card	23.69	9.62	20.05	27.69	76.57	6.0%
IDA(WB)	0.25	3.29	9.07	0	0	0
Loan Payment	0	0	0	0	73.61	5.7%
Total	537.82	762.9	798.28	987.06	1,280.76	

Source: NHIA

Predictability and adequacy of funds for NHIS is a major challenge for timely reimbursement. The average annual claims expenditure per member can be calculated to be GHC 94.4 in the year 2014. The average premium paid by active members in the informal sector is calculated to be about GHC 11 (National Health Insurance Authority, 2014). The high proportion of claims poses a financial risk to the scheme and also limits the options for further expanding. Guided by the changing needs of the Ghanaian economy and health care financing, there is pressing need to undertake a comprehensive study on claims and premium policies in the country. The question however is, what is the appropriate actuarial model for claims management in the Municipal health insurance schemes in Ghana? This study therefore sought to end some answers to these questions and to evaluate the level of appropriate actuarial model best t for the given claim data set of the municipal health insurance schemes. This study therefore applied statistical distributions to assess the appropriate actuarial model for claims management in the mutual health insurance scheme. The history of statistical distributions started in the 18th century (Patil, Kotz & Ord, 1974; Famoye & Lee, 2014).

1.2 Problem Statement

There is still on going research in investigating appropriate distributions, which describe fee for services and Ghana diagnoses related groupings for the large volume of claims, since a good understanding and interpretation of probability distribution is the back bone all decisions made in the insurance industry.

1.3 Purpose of Study

The purpose of this study is to identify comprehensively the appropriate actuarial model for the claim amounts and severity for the categories of service providers of the NHIS.

1.4 Objectives of Study

The research methodology and scope would be designed to achieve the following objectives:

1. To determine the appropriate probability distribution for Fee for Services in modeling claims for drugs across the various services providers.

2. To determine the appropriate distribution that is useful for modeling the G-DRG claims for services for the Amansie East Municipality NHIS.

1.5 Research Questions

In order to achieve the stated objectives, this study will seeks to answer the following questions.

1. Does log-normal distribution provides a better probability model for the claim amounts of fee for services in NHIS?

2. Does log-normal distribution provides the correct probability model for the claim amounts of G-DRG?

3. What is/are the appropriate recommendation(s) on payment of claims to be made base on the known pattern of distribution?

1.6 Significance of the study

It is believed that the results of this study when completed can:

1. Help the management of all municipal and district health insurance schemes managers in crafting the corporate strategy of their jurisdiction as far as claim payment in health insurance is concerned. Management of the various municipals and district health insurance schemes will be in a position to apply the managerial policy prescriptions to improve on their performances

2. The National Health Insurance Authority in Ghana as the regulator of the health insurance sector may also be able to use this study to inform its policies and regulations, as it embarks on its objective of making healthcare accessible to the general public
3. Serve as a source of evaluation for further study into municipals and district health insurance schemes in Ghana and elsewhere. That is, it may serve as a knowledge base for stakeholders and researchers working on related topics. This study is therefore significant in the sense that no previous study, to the best our knowledge, has carried out a similar methodological analysis on the Ghanaian health insurance schemes.

1.7 Overview of Research Methodology

This Research will use secondary source of data from Amansie East Municipal Health Insurance in Ghana. The Researcher will use exploratory data analysis technique to assist in the identification of family of distribution which the data might follow. The Akaike Information Criterion and Kolmogorov-Smirnov test will be used to test the goodness of fit.

The various parametric probability distributions that will be used are the weibull, normal, Pareto, Burr and lognormal. The variables applied in the study were the loss amount incurred by the service providers of NHIS in the district. All variables are identical and independently distributed. Basic statistical package will be used

1.8 Organization of the Study

This study is organised into five chapters. Chapter one is the introduction of the study. It includes the background of the study, problem statement, objectives of the study, research questions, significant and organization of the study

Chapter reviews relevant literature both theoretical and empirical that underlines this present study. The chapter is divide into various sections that includes theories and empirical studies on parametric probability distributions as relevant in actuarial applications.

Chapter three discuses the methodology of the study, research design on data collection. Chapter four is data analysis, presentation and discussion of findings. Chapter five deliberate on summary of findings, conclusions and recommendations .

CHAPTER 2

Literature Review

2.1 Introduction

The purpose of this study is to analyze comprehensively the appropriate actuarial model for the claim amounts and severity for the categories of service providers of the NHIS. This chapter focus on literature review related to the current study, this afford us opportunity to investigate which statistical model best t the actuarial model for the claims, based on basic concepts actuarial claim models and also be able to examine the various categories of services rendered by providers. This chapter therefore reviews relevant literature, both theoretical and empirical that underlines this present study. The chapter is divided into various sections that include theories and empirical studies on statistical distributions as relevant in actuarial applications. Specifically, first section of this chapter provides an overview of the provision of health care under the national health insurance schemes in Ghana. Second section explores theoretical framework of statistical distribution models and how they are relevant to current study. Third section of this chapter focus on empirical perspective and review of relevant literature related to this current study.

2.2 The provision of health care under the national health insurance schemes in Ghana.

Ghana's National Health Insurance Scheme (NHIS) is a fusion of the traditional Social Health Insurance and Mutual Health Insurance (Gobah & Zhang, 2011). The scheme is designed to promote social health protection through risk equalization, cross subsidization, solidarity, equity and quality care.

The National Health Insurance Authority (NHIA) was established under the National Health Insurance Act 2003, Act 650, as a body corporate, with perpetual succession, an Official Seal, that may sue and be sued in its own name. As a body corporate, the Authority in the performance of its functions may acquire and hold movable and immovable property and may enter into a contract or any other transaction.

The National Health Insurance Authority (NHIA) is located in Accra. They have regional and district offices considered as an extension of the operational division of NHIA and are to monitor and evaluate the performance of the administrative centres of NHIS in each region and district.In 2004, Ghana started implementing a National Health Insurance Scheme (NHIS) to remove cost as a barrier to quality healthcare (Sodzi-Tettey, Aikins, Awoonor-Williams & Agyepong, 2012). Providers were initially paid by fee - for - service.

In May 2008, this changed to paying providers by a combination of Ghana - Diagnostic Related Groupings (GDRGs) for services and fee - for - service for medicines through the claims process.

A new law, Act 852 has replaced ACT 650 in October 2012 to consolidate the NHIS, remove administrative bottlenecks, introduce transparency, reduce opportunities for corruption and gaming of the system, and make for more effective governance of the schemes. Section 39 of Act 852 established the National Health Insurance Fund (NHIF) and places responsibility of its management on the shoulders of the Board. The object of the Fund is to provide finance to subsidize the cost of provision of health care services to members of National Health Insurance Scheme.

The objective of the Authority is to attain universal health insurance coverage in relation to

- persons resident in the country
- persons not resident in the country but who are on a visit to this country and to

- provide access to health care services to the persons covered by the Scheme

The vision of the NHIS in Ghana is to be a model of a sustainable, progressive and equitable social health insurance scheme in Africa and beyond (NHIA, 2010). And its mission is to provide financial risk protection against the cost of quality basic health care for all residents in Ghana, and to delight subscribers and stakeholders with an enthusiastic, motivated, and empathetic professional staff who share the values of accountability in partnership with all stakeholders (NHIA, 2010).

Their core values are Integrity, Accountability, Empathy, Responsiveness and Innovation (NHIA, 2010). The corporate goals of the National Health Insurance Scheme are:

- To attain a financially sustainable health insurance scheme;
- To achieve universal financial access to basic health care.
- To secure stakeholder satisfaction (NHIA, 2010) Provider payment methods in use in Ghana currently are:

1. Itemized Fee for service (FFS) to pay for medicines for insured clients
2. Diagnosis Related Groupings (DRG) for insured clients (Services only)
3. Capitation

2.2.1 Itemized Fee for service

In an itemized fee for service payment method, the provider typically lists the different services that they have provided for the client and the cost of each service and requests payment. To use an illustration from day to day life, it is rather like picking up the items you want from a supermarket shelf and then going to the payment counter for the individual cost of each item to be entered into the cash register and added up so that you pay your final bill. The difference between purchasing health care services by fee for service and purchasing items in the supermarket is that because of the specialized knowledge of the health service provider, which the client or patient often does not share, the service provider chooses the ’items’ for the client. The advantage of the itemized fee for service payment method is that the provider has no incentive to leave anything off the ’shopping list’. Whatever they think the client needs will go into the list of items supplied unless the client does not have the ability to pay. The disadvantage of itemized fee for service is that because, the provider is also often the ’owner’ of the shop and also the one choosing the items to be purchased for the client; it is possible for the provider to provide unnecessary services, medicines and diagnostics to maximize profit. Experience all over the world shows that fee for service payment methods can lead to very rapid inflation of costs and threaten the sustainability of health insurance. Countries such as Germany that use Fee for Service successfully in their health insurance scheme, often devise very complicated methods to counteract this tendency and control cost inflation .

At the start of the National Health Insurance Scheme implementation, Ghana was using itemized fee for service to pay for everything including medicines. In 2007/08 the system was reformed to use a diagnostic related groupings payment for service, but medicines continued to be paid for by itemized fees for each medicine supplied (NHIA, 2010).

2.2.2 Diagnosis Related Groupings (DRG)

In the DRG payment method, related diagnoses are grouped together and the average cost of treatment in that group determined. Providers are paid this average cost according to the diagnosis they give their client. Many developed countries e.g. USA and U.K, use DRG as part of their payment systems.

Currently Ghana uses the DRG system to pay for services to insured clients while continuing to pay for medicines by an itemized fee for each medicine supplied. Under the Ghana DRG system providers have to fill claims forms for reimbursement after providing the services. The claims made by the providers are checked (vetted) for accuracy and genuineness by the schemes and the NHIA before payment. The process is administratively complicated and makes a heavy demand on the time of both provider and scheme staff and the NHIA.

The DRG payment method also still holds some incentives for cost escalation though what is known as 'creeping'. This is observed all over the world where DRG is used and is not unique to any one country. In 'creeping' the provider may deliberately give a diagnosis that attracts a higher fee e.g. instead of diagnosing simple malaria they diagnose complicated malaria. Generally however cost inflation under a DRG payment systems are less than under an itemized fee for service system. Under the Ghana DRG system medicines at all levels continue to be paid for by an itemized fee for medicine payment method and the potential of major cost escalation remains strong.

2.2.3 Capitation

Capitation is a provider payment method in which providers are paid, typically in advance, a pre-determined fixed rate to provide a defined set of services for each individual enrolled with the provider for a fixed period of time. The amount paid to the provider is irrespective of whether that person would seek care or not during the designated period. The fixed amount is typically expressed on a Per Member per Month (PMPM) basis. The member refers to active NHIS subscribers assigned to the accredited providers. Under this payment system, the member or subscriber selects a preferred primary provider (PPP) to provide all the services under the capitation basket in exchange for the capitation rate. The capitation basket refers to the services and medicines that are to be paid for by the per capita rate. The total capitation amount is transferred to the provider at the beginning of the service period. The amount is calculated based on the total number of active members who have selected a given provider (www.myjoyonline.com, 2013).

After service provider provides services to a member:

- Service Provider completes a copy of Health Fee Advisory Board (HFAB) (ticks the referral portion of the form where applicable)

- Service Provider compiles a Schedule of the total cost of service, supplies and medicines; attaches a copy of the Referral to HFAB (where applicable). Include any other attachments.
- Service Provider presents the Schedule and the original copies of the HFAB (with attachments), if any) to the Scheme Office where the service is provided (NHIA, 2010). The scheme has a positive effect on health seeking behaviour and utilization of health care services by removing significant financial barriers to access.

Although the insurance has a positive aspect and impact on the state of the citizens of this nation and other foreign nationals, if care is not taken it can easily collapse due to the lack and delay of funds to pay the claims of the various service providers on time (NHIA, 2011). The scheme in recent times has being experiencing some challenges on its claim system thereby raising comments and questions about the future sustainability of the scheme.

Claims are usually made by service providers after a patient undergoes treatment and then submitted to the district claim processing centers monthly for vetting and payment using the Ghana-Diagnosis Related Group (G-DRG) rates for services and Fee-For-Service (FFS) for medicines. Claims turnaround time has an average of 60-90 days since is basically vetted using paper base process (manual) (NHIA, 2011). Due to delay in vetting and lack of proper planning in funds arrangement, payments are not made on time, hence drawing the attention of international bodies like the World Bank about the sustainability and future prospects of the scheme.

Electronic Claims Processing was a strategy adopted by management in 2013 to address logistical challenges associated with paper claims management, boost efficiency in claims processing, offer transparency to providers and provide credible claims data for analysis. In April 2013, a pilot of e-claims processing was instituted in 47 health care facilities with support from the Health Insurance Project (HIP). E-claims submission is expected to be scaled up in the coming year.

Policy holders must first file an insurance claim before any money can be disbursed to the hospital or other contracted service. The insurance company may or may not approve the claim, based on their assessment of the circumstances after vetting (www.wisegeek.com, 2013). Many researchers have come out with different types of probability distribution models linking them with different types of dataset most especially in finance and insurance (life and non-life). A search of literature has revealed a large number of studies on the application of modeling that have been applied in different aspect of insurance or in actuarial literature (Thomas & Samson, 1987; Renshaw,1994; Haberman & Renshaw, 1996; Karen,Yip, Kelvin & Yau, 2005).

The concept of statistical distribution in modeling claims and cost has been widely discussed from the view of many researchers and institutions relating it to different probability distribution (Adeleke and Ibiwoye, 2011). The lognormal distribution is often skewed to the right or heavily skewed and is often useful in modeling claim size (Hogg et al, 2005). The Weibull distribution has also been found to be particularly useful in non-life insurance for modeling the size of reinsurance claims (Beirlant and Jozef, 1992; Boland, 2007).

The exponential distribution has a memory-less property, a property that is shared by no other continuous distribution. This unique property characterizes the family of exponential random variables. Since much of the theory of the generalized linear models is derived from this distribution, it is a very important distribution in modeling insurance claim counts (Boland, 2007). The Gamma distribution has been found to be extremely useful for risk analysis modeling, most especially for claim size modeling (Hogg et al, 2005). The claims data appears to follow an exponential distribution. However in economics prime examples are the distributions of incomes follows the "Pareto's law" (Reed, 2000).This pattern also known as Pareto distribution might not be valid for all claims. It is not clear cut which distributions are suitable for modeling claim amounts as claims can take on large number of values in different types of insurance business (David, 2006).

Prieto et al (2013) Modeled major failures in power grids in the whole range. Their aim was to find a probability distribution that they could use to model electricity transmission networks reliability data. The size of major failures in terms of energy not supplied (ENS), total loss of power (TLP) or restoration time (RT) using data from European power grid operated by UCTE was fitted to six alternative models: Pareto II, Fisk, Lognormal, Pareto (Power Law), Weibull and Gamma distributions. The distributions were fitted to the data by maximum likelihood; they then compared these models by the Bayesian information criterion; tested the goodness-of-fit of those models by a Kolmogorov-Smirnov test method based on bootstrap resampling; and validated them graphically by rank-size plots.

A two parametric probability distribution was hypothesized to be the fit best in modeling the major events reliability data of electricity transmission networks in the entire domain. Pareto II model fits very well the pattern of Energy not Supplied (ENS) and Total Loss of Power (TLP) data and is the best of the six models considered for Restoration Time (RT) data. Additionally, they found other two models with two parameters: the Fisk (also known as Log-logistic distribution) and the lognormal distributions, adequate specially for Total Loss Power data (Prieto et al, 2013). The dataset was very few and hence bootstrapping technique was applied in simulating their data.

Gilbert et al, (2005) also fitted a linear mixed model and a Bayesian hierarchical model to data provided by an insurance company located in the Midwest of Wisconsin. The model was then used to t the 2004 data to predict health insurance claims costs for 2005.

Adeleke & Ibiwoye (2011) used claims data from the most prominent lines of non-life insurance business in Nigeria to determine appropriate models for claim amounts by fitting theoretical distributions to the various data. The risk premiums for each class of business were also estimated. They collected data on the following insurance categories: Fire Motor, Property, Theft, and Armed Robbery, because these are the classes that fit more into the personal lines. They found the Exploratory Data Analysis (EDA) techniques very useful in investigating the suitability of certain families of distributions for particular data in attempting to fit the various claims data. These fitted distributions include exponential, Lognormal, Weibull and Gamma distributions. Since these techniques for analyzing fit are exploratory, they needed to use one or more of the traditional classic methods to test the goodness of their fit.

In Ghana, there is limited number of study in the health insurance sector. For instance, Sodzi-Tettey, Aikins, Awoonor-Williams & Agyepong (2012), assessed the challenges in provider payment under the Ghana national health insurance scheme in two districts in

Ghana under the scheme. Adei, Osei & Diko (2012) made an assessment of the Kwabre District Mutual Health Insurance Scheme in Ghana. They indicate that there is a low internal fund generation as a factor of excessive disenrollment resulting from membership non-renewal. Based on this premise, the scheme may not be sustainable in the long run as a Mutual Health Insurance Scheme since the schemes are dependent on the collective pool of resources. Their recommendation was that Government should boldly implement the one-time premium on a progressive and reasonable premium affordable to all.

Gobah & Zhang (2011) also conducted a study to assess the effect of the Scheme on access to and utilization of healthcare services in the Akatsi District of the Volta region of Ghana. They indicate that the scheme has a positive effect on health seeking behaviour and utilization of health care services by removing significant financial barriers to access.

Dadey, Ablebu, & Agboda (2011) compared the Poisson distribution and the negative binomial distribution models to determine which distribution best fits motor insurance claim in Ghana. The study revealed that the Negative Binomial distribution appear to be more effective than poisson distribution for fitting insurance claims and therefore, provides somewhat reliable estimates for planning, decision making as well as estimation in insurance administration.

Various reasons for fitting a distribution to a set of data have been summarized by Hahn

& Shapiro (1967) to include: the desire for objectivity, the need for automating the data analysis, and interest in the values of the distribution parameters.

Although various some studies using statistical distributions already exist, e.g., the Pearson system and the Johnson system and the Burr distribution, we are empirically analysing the various statistical distribution in Ghanaian context.

CHAPTER 3

Research Methodology

3.1 Introduction

As stated above, the main purpose of this study is to comprehensively analyse the appropriate actuarial model for the claim amounts and severity for the categories of service providers of the NHIS. This chapter gives the reader a clear view of how this research was carried out and more essentially by linking each research questions with methods. The research method is a strategy of enquiry, which moves from the underlying assumptions to research design, and data collection (Myers, 2009).

This chapter includes research design, data source, target population, sampling technique, and exploratory data analysis technique. Finally, the actuarial modeling process, data analysis method and the statistical techniques used.

3.2 Research Design

This study adopts quantitative research techniques in its approach in order to achieve the main objective of this study. We employ a quantitative technique in this present study because it is more reliable and objective, that is, the research results are relatively independent of the researcher (e.g., statistical significance) and also have higher credibility with many people in power (e.g., administrators, managers).

The basic building blocks of quantitative research are variables. Variables (something that takes on different values or categories), which are the opposite of constants (something that cannot vary, such as a single value or category of a variable). The study use descriptive study to gather information and to discover whether there is any relation between the variables used in the study.

constants (something that cannot vary, such as a single value or category of a variable). The study use descriptive study to gather information and to discover whether there is any relation between the variables used in the study.

3.3 Target Population and Sampling Technique

The population of interest for this study is defined as all NHIS service providers in the Amansie East Municipal Health Insurance Scheme during the period under consideration. A purposive sampling technique used to select target population in order to achieve the stated objectives of the study.

3.4 Data Sources and Data Analysis of the Study

The department of Mathematics of the Kwame Nkrumah University of Science and Technology give an introductory letter which was used to facilitate the collection of secondary data (Claims data). The study involves the use of computer statistical packages include SPSS and R to perform the tests and to plot the necessary graphs.

3.5 Exploratory Data Analysis Technique

This section includes the descriptive statistics and visualizations which aid us in the prior selection of the possible candidate model or family of distribution for the research. Some of which are as follows; Measures of location (mean, median, mode), Measures of variation (range, variance, standard deviation),Measures of symmetry (skewness), Distributional plots (histograms)

3.6 Actuarial Modeling Process

Here, the modeling process was described. That is, steps that was followed in fitting a statistical distribution to the claim. That is, the steps that were taken in the actuarial modeling process are as follows; Selection of Certain Family of distributions, Estimation of the model parameters, Criteria selection to choose the model,Testing the Goodness-of-Fit, and Check the best model that fit the data

3.6.1 Selection of Certain Family of Distributions

The exploratory data analysis assisted in the selection of the following distribution for the modeling process. In this step of modeling process considerations were made of a number of parametric probability distributions as potential candidates for the data generating mechanism of the claim amounts and claim frequency. Fiete (2014) states that most data in general insurance data is skewed to the right and therefore most distributions that exhibit this characteristic can be used to model the claim severity. Due to the rightly skewed nature of the dataset the Log-normal distribution, the Weibull distribution, the Gamma distribution, the Exponential distribution, the Burr distribution and the Pareto distribution were selected as the candidate models after testing a wide range of distributions with the an R software

3.7 Theoretical Framework of Statistical Distribution Models

The claim amounts to the insurance company can be described by discrete and by continuous random variables. In the case of continuous distributed claims, the basic is to find an adequate model for the claim amount. The probability distributions are separated in two families - light tailed and heavy tailed distributions.

3.7.1 The Probability Density Function (PDF)

The density functions are useful when considering continuous random variables. In probability theory, a probability density function describes the relative likelihood for this random variable to take on a given value. A continuous probability distribution is usually defined by its distribution function or by its density function. The percentile function is simply the inverse of the distribution function. This concept is statistical distribution is particularly useful in modeling studies because of the result (Ramberg, Tadikamalla, Dudewicz & Mykytka, 1979).

The density functions are useful when considering continuous random variables. Probability density functions is either discrete or continuous, variable are denoted p(r) and f(x), respectively. They are assumed to be properly normalized such that

$$\sum_r p(r) = 1 \quad \text{,and} \quad \int_{-\infty}^{\infty} f(x)dx = 1 \tag{3.1}$$

where the sum or the integral are taken over all relevant values for which the probability density function is defined. Statisticians often use the distribution function or more often call the cumulative function which is defined as

$$P(r) = \sum_{i=-\infty}^{\infty} p(i), F(s) = \int_{-\infty}^{\infty} f(x)dx = 1 \tag{3.2}$$

algebraic moments of order r are defined as the expectation value

$$\mu_t' = E(x') = \sum_k k' p(k) \int_{-\infty}^{\infty} x' f x dx \tag{3.3}$$

Obviously $\mu_0' = 1$ from the normalization condition and μ_1' is equal to the mean, sometimes called the expectation value, of the distribution. Central moments of order r are defined as

$$\mu_T = E(k - E(k))', \mathbf{or}, E(x - E(x))' \tag{3.4}$$

of which the most commonly used is μ_2 which is the variance of the distribution. Instead of using the third and fourth central moments, one often defines the coefficients of skewness λ_1 and kurtosis γ_1

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}, \gamma_2 = \frac{\mu_4}{\mu_2^2} = 3 \tag{3.5}$$

where the shift by 3 units in γ_2 assures that both measures are zero for a normal distribution. Distributions with positive kurtosis are called leptokurtic, those with kurtosis around zero mesokurtic and those with negative kurtosis platykurtic. Leptokurtic distributions are normally more peaked than the normal distribution while platykurtic distributions are more at topped.

For a distribution in a continuous variable x the Fourier transform of the probability density function

$$\phi(t) = E(e^{\alpha t}) = \int_{-\infty}^{\infty} e^{(\alpha t)} f(x)dx \tag{3.6}$$

is called the characteristic function. The characteristic function is related to the moments of the distribution by

$$\phi_x(t) = E(e^{\alpha t}) = \sum_{n=0}^{\infty} \frac{(it)^n E(x^n)}{n!} = \sum_{n=0}^{\infty} \frac{(it)^n (\mu_x)}{n!} \tag{3.7}$$

3.7.2 Exponential Distribution Model

Another useful continuous distribution is the exponential distribution, which has the following probability density function:

$$f(x, a) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}} \tag{3.8}$$

where the variable x as well as the parameter α is positive real quantities. The exponential distribution is typically used to model time intervals between "random events". The cumulative distribution function of the exponential distribution function is

$$F(x) = \int_0^x f(x)dx = 1 - e^{-\frac{x}{\alpha}} \tag{3.9}$$

And it is thus straightforward to calculate probability content in any given situation. The moments, that is, the expectation value, variance, and lowest order central moments are given by

$$E(x) = \alpha, Var(x) = \alpha^2, \mu_3 = 2\alpha^3, \mu_4 = 9\alpha^4, \mu_5 = 44\alpha^5, \mu_6 = 256\alpha^6$$

More generally algebraic moments are given by $\mu_n = \alpha^n n!$ The characteristic function of the exponential distribution is given by

$$\phi(t) = E\left(e^{\alpha t} = \frac{1}{\alpha}\right) \int_0^\infty e^{(it-\frac{1}{\alpha})x dx} = \frac{1}{1 - it\alpha} \quad (3.10)$$

3.7.3 The Gamma Distribution Model

The Gamma distribution is a two-parameter family of continuous probability distributions and has a right skewed distribution. It is very useful for risk analysis modeling, particularly, for claims size modeling (Hogg et al, 2005). The Gamma distribution is given by

$$f(x, a, b) = \frac{a(ax)^b - 1e^{-ax}}{\Gamma(b)} \quad (3.11)$$

where the parameters a and b are positive real quantities as is the variable x. Note that the parameter a is simply a scale factor. For b≤1 the distribution is J-shaped and for b≥1 it is unimodal with its maximum at x = $\frac{b-1}{a}$ In the special case where b is a positive integer this distribution is often referred to as the Erlangian distribution.

For b = 1 we obtain the exponential distribution and with a = $\frac{1}{2}$and b =$\frac{n}{2}$with n an integer we obtain the chi-squared distribution with n degrees of freedom.

The distribution has the following moments. That is, expectation value, variance, third and fourth central moments are given by

$$E(x) = \frac{b}{a}, Var(x) = \frac{b}{a^2}, \mu_3 = \frac{2b}{a^3}, \mu_4 = \frac{3b(b+2)}{a^2}$$

The coefficients of skewness and kurtosis is given by

$$\gamma_1 = \frac{2}{\sqrt{b}}$$

$$\gamma = \frac{6}{b}$$

More generally algebraic moments are given by

$$\mu_n' = \int_0^\infty x^n f(x)dx = \frac{a^k}{\Gamma(b)} \int_0^\infty x^{n+b-1} e^{ax} dx$$

$$\frac{a^k}{\Gamma(b)} \int_0^\infty \left(\frac{y}{a}\right)^{n+b-1} e^{-y} \frac{dy}{a} = \frac{\Gamma(n+b)}{a^n \Gamma(b)} = \frac{b(b+1...(b+n-1))}{a^n}$$

where we have made the substitution in simplifying the integral. The characteristic function is

$$\phi(t) = E(e^{itx}) = \frac{a^b}{\Gamma(b)} \int_0^\infty x^{b-1} e^- x(a-it)dx$$

$$= \frac{a^b}{\Gamma(b)} \cdot \frac{1}{(a-it)^b} \int_0^\infty y^{b-1} c - ydy = \left(1 - \frac{it}{a}\right)^{-b}$$

where we made the transformation y = x(a-it) in evaluating the integral. In order to calculate the probability content for a Gamma distribution we need the cumulative (or distribution) function

$$F(x) = \int_0^x f(x)dx = \frac{a^b}{\Gamma(b)} \int_0^{ax} u^{b-1} du$$

$$= \frac{a^b}{\Gamma(b)} \int_0^{ax} \left(\frac{v}{a}\right)^{b-1} e - v\frac{dv}{a} = \frac{1}{\Gamma(b)} \int_0^{ax} v^{b-1} dv = \frac{\lambda(b,ax)}{\Gamma(b)}$$

where (b,ax) denotes the incomplete gamma function.

3.7.4 Normal Distribution

The normal distribution is important in insurance and nance since it appears like limiting distribution in many cases. The normal distribution is applicable to a very wide range of phenomena and is the most widely used distribution in statistics. It was originally developed as an approximation to the binomial distribution when the number of trials is large and the Bernoulli probability p is not close to 0 or 1. It is also the asymptotic form of the sum of random variables under a wide range of conditions.

The normal distribution was first described by the French mathematician de Moivre in 1733. The development of the distribution is often ascribed to Gauss, who applied the theory to the movements of heavenly bodies.

The normal distribution is the most important distribution in statistics, since it arises naturally in numerous applications. The key reason is that large sums of (small) random variables often turn out to be normally distributed. The normal density function is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}(\frac{x-\mu}{\sigma})^2}, -\infty \leq x \geq \infty \quad (3.12)$$

The notation X~ $N(\mu,\sigma)$ has implies that X has a normal distribution with parameters μ and σ . The mean and the variance are E(X) = μ and V ar(X) = σ^2. The most important property is the random variable Z = $\frac{X-U}{\sigma}$ is normally distributed with parameters 0 and 1 Z~N(0; 1) The distribution function is denoted φ and

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-x^2}{2}} dx \quad (3.13)$$

3.7.5 The Log Normal Distribution Model

The lognormal distribution is applicable to random variables that are constrained by zero but have a few very large values. The resulting distribution is asymmetrical and positively skewed.

The application of a logarithmic transformation to the data can allow the data to be approximated by the symmetrical normal distribution, although the absence of negative values may limit the validity of this procedure. In probability theory, a log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable is log-normally distributed, then it has a normal distribution.

A random variable X is said to have the lognormal distribution with parameters μR and $\sigma \leq$ 0 if ln(X) has the normal distribution with mean μ and standard deviation σ. Equivalently, X = e^y where Y is normally distributed with mean μ and standard deviation σ. The lognormal distribution is used to model continuous random quantities when the distribution is believed to be skewed, such as certain income and lifetime variables. The lognormal distribution is a continuous probability distribution that has intriguing theoretical and practical properties. The lognormal distribution is often skewed to the right or heavily skewed and is often useful in modeling claim size (Hogg et al, 2005). As the Probability Density Function (PDF) suggests, the lognormal distribution is the distribution of a random variable x in log space. If the data size is too large then in most cases it is assumed to be approaching Normal A random variable x, is said to have a lognormal distribution with parameter μ and σ, if Y=$lnx = N \sim (\mu,\sigma^2)$. The log-normal distribution is therefore given by

$$f(x:\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{ln(x-\mu)}{\sigma}\right)^2} \tag{3.14}$$

where the variable x≤0 and the parameters μ and σ ≤0 all are real numbers. It is sometimes denoted $\Lambda(\mu,\sigma^2)$ in the same spirit as we often denote a normally distributed variable by N(μ,σ^2) . If u is distributed as N(μ,σ^2) and $\mu = lnx$ then x is distributed according to the log-normal distribution. Also, if x has the distribution $\Lambda(\mu,\sigma^2)$ then $e^a x^b$ is distributed as $\Lambda(a + b\mu b^2\sigma^2)$ The figure below show the log-normal distribution for the basic form, with

μ=0 and σ=1

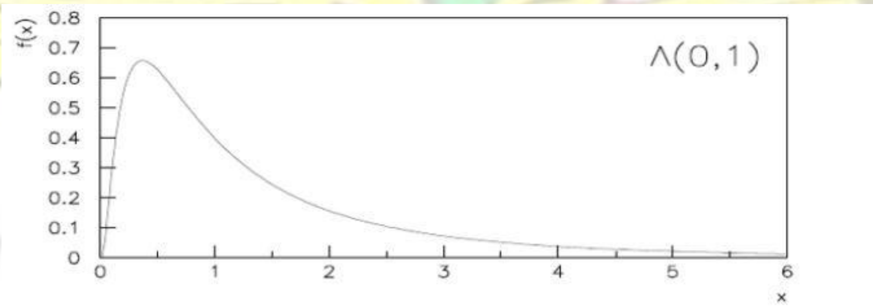

Figure 3.1: Log-normal distribution with μ = 0 and σ = 1

The cumulative distribution, or distribution function, for the log-normal distribution is given by

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_0^x\frac{1}{t}e^{\frac{-1}{2}\left(ln\frac{(t-\mu)}{\sigma}\right)^2}dt, =\frac{1}{\sigma\sqrt{2\pi}}\int_0^{lnx}\frac{1}{t}e^{\frac{-1}{2}\left(\frac{ln(y-\mu)}{\sigma}\right)^2}dy \tag{3.15}$$

Where we have put $z = (lnx-\mu)/\sigma$ and positive sign is valid for z ≤0 and the negative sign for z ≥ 0

3.7.6 Weibull Distribution Model

The Weibull is a very general and popular failure distribution that has been shown to apply to a large number of diverse situations. This distribution is also used to represent nonnegative task times that are skewed to the left. This distribution is named after Waloddi Weibull, a Swedish physicist, who used it in 1939 to represent the distribution of the breaking strength of materials. The distribution has also been used in reliability and quality control. The density function of the three parameter Weibull distribution is given by:

$$f(x) = \frac{\alpha}{\beta - \epsilon}\left(\frac{x-\epsilon}{\beta - \epsilon}\right)^{(\alpha-1)} exp\left[-\left(\frac{x-\epsilon}{\beta-\epsilon}\right)^{\alpha}\right], \alpha \geq 0, \beta \geq 0, 0 \leq \epsilon \leq \infty \quad (3.16)$$

The shape of the distribution depends primarily on the shape parameter, α. The scale parameter is β and the delay/displacement parameter is . If 1=α, the Weibull distribution reduces to the exponential distribution. As increases, the Weibull distribution tends to the normal distribution. For α=1, the distribution becomes the Rayleigh distribution. The Weibull distribution is also known as the bounded exponential distribution. The cumulative distribution function is given by:

$$F(x) = 1 - exp\left[-\left(\frac{x-\epsilon}{\beta-\epsilon}\right)^{\alpha}\right] \quad (3.17)$$

By using the transformation y = $\left(\frac{x-\epsilon}{\beta-\epsilon}\right)^{\alpha}$, the tables of e^{-y} can be used to determine F(x). The mean and variance of the distribution are:

$$E(x) = \epsilon + (\beta - \epsilon)\Gamma(1 + \frac{1}{\alpha}) \quad (3.18)$$

$$Var(x) = (\beta - \epsilon)^2\left[\Gamma\left(1 + \frac{2}{\alpha} - \Gamma^2\left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)\right)\right] \quad (3.19)$$

The coefficient of skew is given by;

$$\gamma = \frac{\Gamma(1+\frac{3}{\alpha}) - 3\Gamma(1+\frac{2}{\alpha})\Gamma^3(1+\frac{1}{\alpha}) + 2\Gamma^3(1+\frac{1}{\alpha})}{[\Gamma(1+\frac{2}{\alpha}) - \Gamma^2(1+\frac{1}{\alpha})]^{\frac{3}{2}}} \quad (3.20)$$

3.7.7 Beta Distribution Model

A distribution that has both an upper and a lower bound is the beta distribution. Generally the beta distribution is defined over the interval 0 to 1. It can, however, be transformed to any interval a to b. If the limits of the distribution are unknown, they become parameters of the distribution making it a four parameter rather than a two parameter distribution. The beta distribution is often used in Bayesian statistics because it is a good model for one's prior belief about the population proportion p, $0 \leq p \geq 1$ The beta density function is given by

$$f(x) = x^{\alpha-1}/\beta(\alpha,\beta), 0 \leq x \leq 1 \tag{3.21}$$

$\alpha > 0$ and $\beta > 0$ are the two parameter of the distribution. This is the standard deviation form of the beta distribution. The function

$$B(\alpha\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx \tag{3.22}$$

is called the beta function. It can be shown that

$$B(\alpha\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{3.23}$$

The moment of Beta distribution model random variable:

$$EX^n = \frac{B(n+\alpha\beta)}{B(\alpha\beta)} = \frac{(n+a)\Gamma(\alpha+\beta)}{\Gamma(n+\alpha)\Gamma(\alpha)} \tag{3.24}$$

The mean and the variance of the beta distribution are

$$E(X) = \frac{\alpha}{\alpha+\beta}, Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{3.25}$$

respectively.

The mean and variance can be used to get the moment estimators for α and β. The beta distribution can assume variety of different shapes depending on the values of its parameters. If $\alpha=\beta= 1$, the beta distribution reduces to uniform distribution in the range[0,1]. But, if one parameter equals one and the other equals two, then it turns into triangular distribution. If both the parameters greater than one, the mode of the distribution is

$$\frac{\alpha - 1}{\alpha + \beta - 2} \tag{3.26}$$

If one parameter equals unity and the other is greater than unity, then there is only one point of inflexion. The distribution is symmetrical if $\alpha=\beta$, skewed to the right if $\alpha>\beta$, and skewed to the left if $\alpha<\beta$

3.7.8 The Pareto Distribution Model

One of the most frequently used analytic claim size distributions is the Pareto distribution model. Experience has shown that the Pareto formula is often an appropriate model for the claim size distribution, particularly where exceptionally large claims may occur (Daykin, Pentikainen,& Pesonen, 1994). The Pareto distribution is given by

$$f(x, \alpha, k) = \frac{ak^{\alpha}}{x^{\alpha+1}} \tag{3.27}$$

Where the interval variable $x \geq k$ and the parameter $\alpha > 0$ are real numbers. As is seen k is only a scale factor. The distribution has its name after its inventor the Italian Vilfredo Pareto(1848-1923) who worked in the fields of national economy and sociology (professor in
Lausanne, Switzerland). It was introduced in order to explain the distribution of wages in society. The cumulative di distribution is given by

$$F(x) = \int_k^x f(u)du = 1 - \left(\frac{k}{x}\right)^{\alpha} \tag{3.28}$$

The algebraic moments of the Pareto model are given by

$$E(x^m) = \int_k^{\infty} x^n f(x) = \int_k^{\infty} x^n \frac{\alpha k^{\alpha}}{x^{\alpha+1}} = \left[\frac{\alpha k^{\alpha}}{x(\alpha - n + 1)}\right]_k^{\infty} = \frac{x^{\alpha+1}}{\alpha - n} \tag{3.29}$$

3.7.9 Burr Distribution Model

The 3-parametric Burr Distribution is a right skewed continuous probability distribution. This distribution is often used to model incomes and other quantities in econometrics (Wright, 2005). A random variable x with parameters k, α and β is said to be a three parametric burr distribution if its probability density function is given by

$$f(x) = \frac{\alpha k}{\beta} \frac{(\frac{x}{\beta})^{\alpha-1}}{(1+(\frac{x}{\beta})^{\alpha})^{k+1}} \quad (3.30)$$

where $x > 0$ and the CDF is given by

$$F(x) = 1 - \left(1 + \frac{x^{\alpha}}{\beta}\right)^{-k} \quad (3.31)$$

When k=1 the the three parametric Burr distribution reduces to the Fisk distribution. When k-continuous shape parameter ($k > 0$), α-continuous shape parameter ($\alpha > 0$) and β continuous scale parameter $\beta > 0$

3.8 Empirical Literature Related to this Current Study

In view of the economic importance of health insurance in developing countries,many attempts have been made in the actuarial literature to nd a probabilistic model for the distribution of the claim amounts reported by insured patients (Denuit, et al 2007). An insurance claim is the actual application for benefits provided by an insurance company. el that fit the data.

3.8.1 Statistical Distributions

- Exponential Distribution Model

$$f(x : \alpha) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}} \quad (3.32)$$

- Gamma Distribution Model

$$f(x,a,b) = \frac{a(ax)^b - 1e^{ax}}{\Gamma(b)}$$
$$E(x) = \frac{b}{a}, Var(x) = \frac{b}{a^2}, \mu_3 = \frac{2b}{a^3} \quad (3.33)$$

- Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}, -\infty < x < \infty \quad (3.34)$$

- The log Normal Distribution Model

$$f(x:\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(ln\frac{x-\mu}{\sigma})^2} \quad (3.35)$$

- Weibull Distribution Model

$$f(x) = \frac{\alpha}{\beta}\left(\frac{x-\epsilon}{\beta-\epsilon}\right)(\alpha-1)exp\left[-\left(\frac{x-\epsilon}{\beta-\epsilon}\right)^{\alpha}\right], \alpha \geq 0, \beta \geq 0, 0 \leq \epsilon \leq \infty$$

$$F(x) = 1 - exp\left[-\left(\frac{x-\epsilon}{\beta-\epsilon}\right)^{\alpha}\right]$$

$$E(x) = \epsilon + (\beta-\epsilon)\Gamma(1+\frac{1}{\alpha}) \quad (3.36)$$
$$Var(x) = (\beta-\epsilon)^2\left[\Gamma\left(1+\frac{2}{\alpha} - \Gamma^2\left(\Gamma\left(1+\frac{1}{\alpha}\right)\right)\right)\right]$$

- Beta Distribution Model

$$f(x) = x^{\alpha-1}/\beta(\alpha,\beta), 0 \leq x \leq 1 \quad (3.37)$$
$$E(X) = \frac{\alpha}{\alpha+\beta}$$

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- Pareto Distribution Model

$$f(x,\alpha,k) = \frac{ak^{\alpha}}{x^{\alpha+1}} \quad (3.38)$$

- Burr Distribution Model

$$f(x) = \frac{\alpha k (\frac{x}{\beta})^{\alpha-1}}{\beta \, (1+(\frac{x}{\beta})^{\alpha})^{k+1}}$$

$$F(x) = 1 - \left(1 + \frac{x^{\alpha}}{\beta}\right)^{-k} \qquad (3.39)$$

3.8.2 Estimation of the Model Parameters

The maximum likelihood estimation method was used to estimate the individual parameters because of its numerous over the other methods of parameter estimation. With the aid of R 3.0.2 statistical software the parameters were computed. Given any model, there exists a great deal of theories for making estimates of the model parameters based on the empirical data (Achieng, 2010), in this case, the claims were used to compute the maximum likelihood estimates of the seven sampled distributions.

Although the log the inverse and square root of the data were taken to minimize the deviation so as to get a very good model there were some non-positive values showing in the log, double log and triple log of the data making it difficult to model with some of the selected distributions. The square root and the inverse of the data also added not much difference to the deviation hence the data was used without any transformation. The first step used in fitting a model to a claims data was done by finding the parameter estimates of their statistical distribution. When the parameters of any distribution have been obtained using the claims data, then literally, the statistical distribution has been fitted to the claims data (Achieng, 2010).

3.8.3 Criteria Selection to Choose the Model

Testing the Goodness-of-Fit is describing how best the statistical distribution fit a set of observation. Akaike Information Criteria (AIC) . AIC = -2(maximized log-likelihood) + 2(number of parameters estimated). The distribution with the smaller value is the preferred one.

3.9 Chapter Conclusion

This chapter has clearly explained the basic concept of methodology that we applied in this current study. An awareness of a clear methodological concept enables better understanding of the concepts and results presented hereafter. The chapter has therefore highlighted the need to undertake this study systematically.

CHAPTER 4

Analysis and Discussion

4.1 Introduction

This chapter four focuses on data presentation, analysis, results and discussion with respect to how it answers the research questions and how it link with literature reviewed. The chapter therefore discusses the distributions of claim data from Amansie East Municipal for July, 2011. The chapter used exploratory data analysis (histogram, mean, skewness, maximum value, minimum value, standard deviation, and 1^{st} and 3^{rd} quartiles) to assist in the identification of the family of distribution which the data might follow. The diagnostics test Probability plot was used to graphically demonstrate goodness of fit to the Log-normal distribution, the Fisk distribution, the Weibull distribution, the Gamma distribution, the Exponential distribution, the Burr distribution and the Pareto distribution. The Goodnessof-Fit tests (Kolmogorov-Smirnov and Anderson-Darling tests) were used to statistically test fitness of the distributions.

4.2 The Exploratory Data Analysis for G-DRG and FFS Claims

Claims are made by service providers after a patient undergoes treatment and then submitted to the district schemes monthly for vetting and payment using the Ghana-Diagnosis Related Group (G-DRG) rates for services and Fee-For-Service (FFS) for medicines. The exploratory data analysis is in two parts the Descriptive statistics and the visualization plots. The sections below explain them.

4.3 Descriptive Statistics

A total of 30509 claim (14981 claims on service using the G-DRG and 15528 claims on drugs using the FFS) for July 2011 submitted to the Amansie East Municipal Health Insurance Scheme from all the various types of service providers in the municipality were used in the analysis.

The scheme currently serves 26 health service providers in the municipality which includes the following; 1 CHAG CHPS Compound, 2 CHAG district hospitals, 5 Chemical shops, 4 public CHPS compounds, 1 public Clinic, 2 District Level Hospitals, 2 Health Centres, 1 Private Maternity Home, 1 Pharmacy, 4 Private Clinics, 1 Private District Hospital, 2 Ultrasound Scan centres. Seven service providers were selected based on their equal categories of tariffs. The service provider with the highest in terms of number of claims and highest amount of claims submitted monthly were selected for the model. They were Bekwai Municipal Hospital, Ahmadiya Muslim Hospital, Dunkrah Health Centre, SDA hospital Dominase, Abenkyiman clinic, Titan Chemist, and Health Frontier Limited. Bekwai Municipal Hospital being the biggest facility

Facility Selected	Facility Types	Claims	Sample Size	Min	Max	Mean	SD	Skew.	Sum
BKW	DLH	G-DRG	5869	0.20	362.50	10.70	23.11	5.75	62,792.12
		FFS	5396	0.03	161.40	8.73	9.79	4.00	47,128.73
AMH	PDH	G-DRG	1530	0.06	333.40	10.92	32.55	6.23	28,255.02
		FFS	1540	5.55	396.64	18.35	22.73	8.51	16,741.23
DUN	HC	G-DRG	1812	1.50	12.10	2.34	1.17	4.11	4,244.80
		FFS	1870	0.30	28.50	6.94	3.56	0.98	12,970.32
SDA	CHAG	G-DRG	3079	1.10	247.90	10.69	16.01	6.89	32,899.82
		FFS	3002	0.01	222.28	11.00	15.34	5.15	3.96 33,027.76
ABN	CL	G-DRG	2432	6.26	213.20	8.67	11.72	12.49	21,080.12
		FFS	2369	5.36	219.60	10.05	8.86	7.41	23,809.32
TIT	PHR	G-DRG	1351	0.15	210.00	11.87	12.63	4.58	16,035.53
HFL	USC	FFS	259	3.26	81.65	29.88	22.18	0.58	7,738.29

in the municipality was chosen for the modeling to represent the District Level hospitals. The Ahmadiya Muslim hospital was selected to represent the private hospitals in the model. Dunkrah Health Centre was selected for CHPS compound, community clinics and Health Centres category which also have the same tariffs. Dominase SDA hospital was selected to represent CHAG hospitals. The Public clinics and maternity homes also have the same tariffs and Abenkyiman clinic was selected for the modeling. The chemical and pharmacy shops were of one category and had the same tariffs; Titan Chemist was selected for the modeling. Finally, the Ultra Scan centres in the municipality having the same tariffs had Health Frontier Limited chosen to represent them in the model. The results in Table 4.1 suggest that the average claim in Ahmadiya Muslim Hospital which is under the private hospital was the highest thus GHs 18.35 and GHs 10.92 for both the FFS for drug and G-DRG for service respectively due to the high Tariffs allocated to them. The standard deviation was as high as GHs 32.55 for G-DRG and GHs 22.73 for FFS indicating high dispersion from one sample to another. However, the sum of claims for both the G-DRG and FFS was GHs 28,255.02 and GHs 16,741.23 respectively. Bekwai Municipal Hospital which is the biggest facility in the municipality submitted the highest claims in terms of the total number of claims and amount of claims (monetary value). The mean however was GHs 10.70 and GHs 8.73for G-DRG for service and FFS for drug claims respectively. The standard deviation was GHs 23.11 and GHs 9.79 for G-DRG for service and FFS for drug claims respectively also indicating high dispersion from one sample to another especially the G-DRG. The sum of claims for both the G-DRG and FFS was GHs 62,792.12 and GHs 47,128.73 respectively. Bekwai municipal hospital submitted about 35.16% of all claims in the municipality. All the claims both the G-DRG and FFS were positively skewed from Table 4.1 indicating that the data is asymmetric. From Figure 4.1 below it can be clearly seen that almost all the service providers in the municipality had higher claims on Drugs than Service except Bekwai Municipal Hospital and Ahmadiya Muslims Hospital due to the numerous nature of services they render and the referral rate to those Hospitals because of the health expects (number and type of doctors) and facilities available

Table 4.1: Descriptive Statistics of the Data

BKW:Bekwai Municipal Hospital AHM:Ahmadiya Muslims,Hospital DUN:Dunkrah Health Centre SDA:Dominase SDA, Hospital ABN:Abenkyiman Clinic TIT:Titan Chemist HFL:Health Frontier LtdDLH:District Level Hospital, PDH:Private District Hospital HC:Health Centre CHAG: Christian Health Association of Ghana CL: Clinic PHR:, Pharmacy USC:Ultra Scan Centre G-DRG: Ghana Diagnosis, Related Grouping FFS: Fee for Service

4.3.1 Visualization Plots

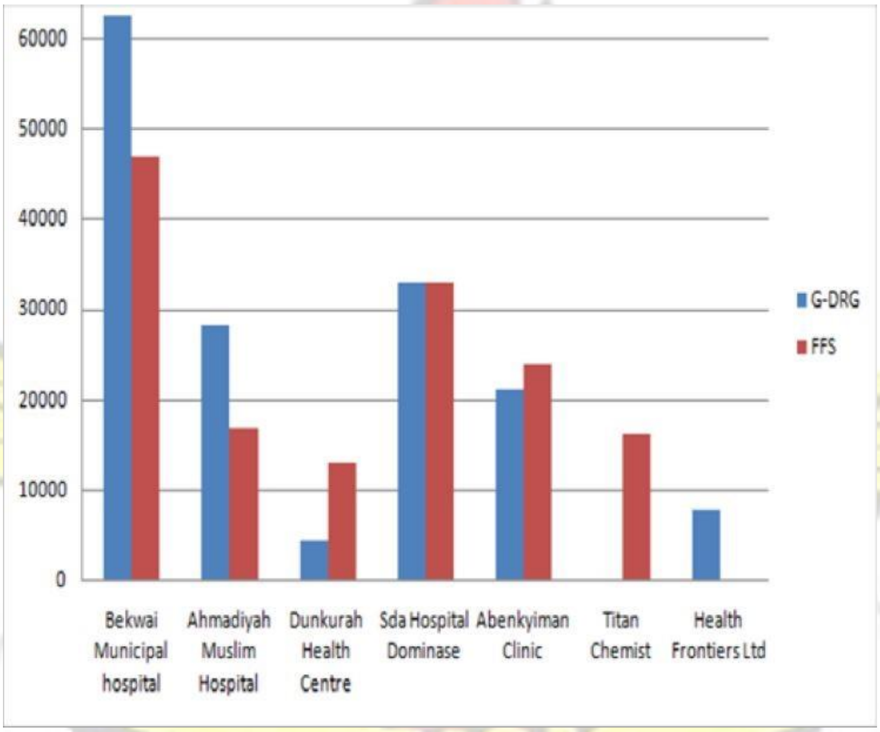

Figure 4.1: A Chart showing the amount of claims submitted for July 2011 The

The histograms for the claims data for both the G-DRG for service claims and FFS for drug claims across the various service providers in the municipality were used to give a clear picture of how the data is skewed. This also assisted in the selection of the family or class of. The histogram in Figures 4.2, 4.3, and 4.4 for all the claims show clearly and confirm the early descriptive analysis that the data is asymmetric

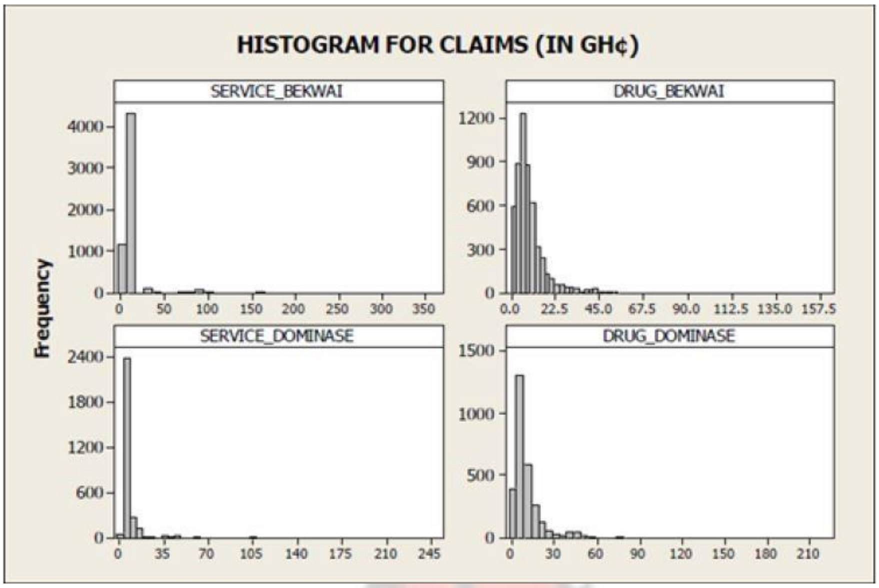

Figure 4.2: Histogram of claims data for Bekwai and Dominase hospital for both FFS and G-DRG

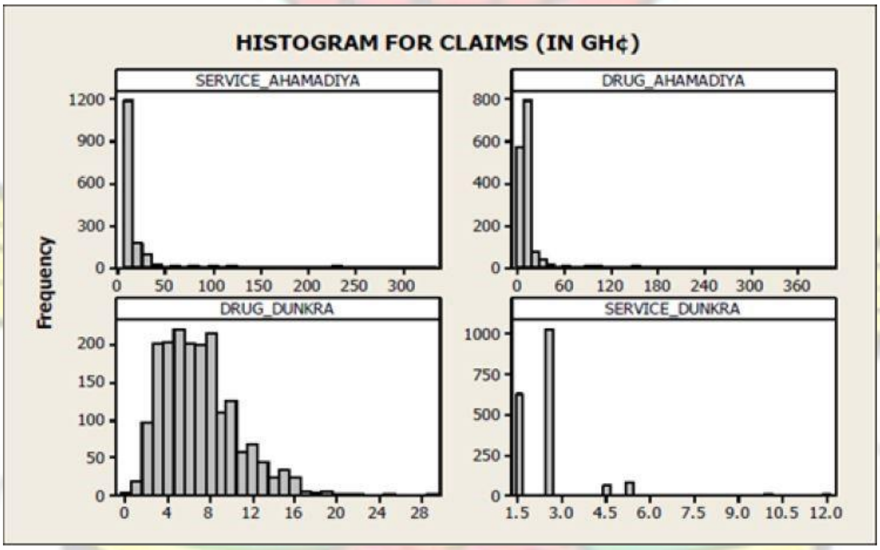

Figure 4.3: Histogram of claims data for Ahmadiya and Dunkrah hospital for both FFS and G-DRG

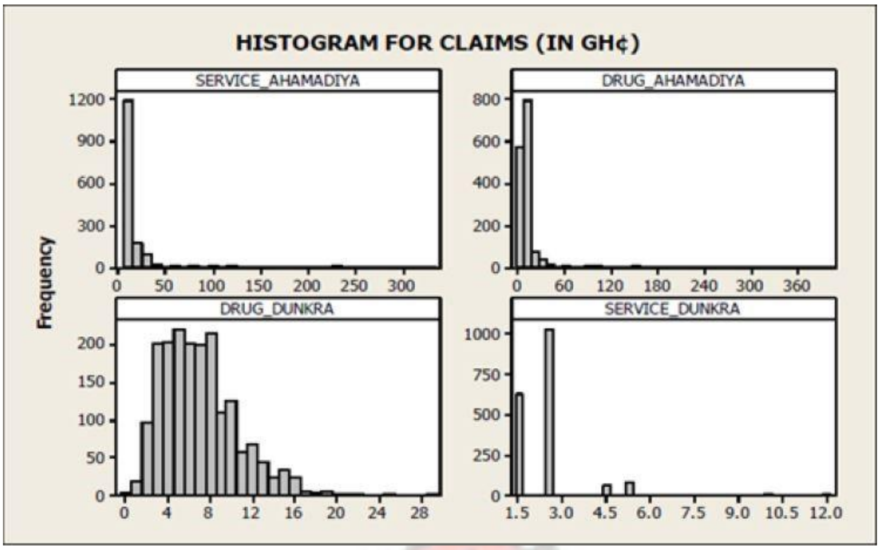

Figure 4.4: Histogram plots of claims data for Abenkyiman Hospital, Health frontier service and Titan Chemist both FFS and G-DRG

In Figure 4.2 the claims for Bekwai Drugs (FFS) and Dominase SDA drugs had very few outliers as compare to their respective G-DRG claims dataset. Dunkrah and Ahmadiya FFS claims also had fewer outliers than their respective G-DRG claims in Figure 4.3. Abenkyiman and Health frontier service had no outlier both in the FFS and G-DRG claims however, Titan chemist saw a very few outliers in the dataset as shown in Figure 4.4. The outliers that were observed in the G-DRG claims for Bekwai, Dominase, Dunkrah and Ahmadiya could be as a result of the variations in the tariff charges across the providers and the various diagnosis or services rendered to patients however the tariffs for drugs remains almost the same for the various health service providers

4.3.2 Estimation of Model Parameters

The maximum likelihood estimation method was used to estimate the individual parameters because of its numerous advantages over the other methods of parameter estimation. With the aid of R 3.0.2 statistical software the parameters were computed in Table 4.2 and 4.3 below.

Table 4.2: Parameter estimates from the Pareto, Fisk, Gamma and Exponential models to the G-DRG and FFS across the various health service providers claims datasets by maximum likelihood

Dataset	Claims	Parameter Estimate			
		Pareto	Fisk	Gamma	Exponential

		α	β	α	α	α	β	α
BKW	G-DRG	0.288	0.200	1.680	6.445	0.213	52.390	0.090
	FFS	0.506	0.920	2.169	6.632	0.910	10.358	0.106
AHM	G-DRG	1.203	5.550	2.270	12.716	0.318	57.760	0.055
	FFS	0.218	0.060	1.571	5.863	0.231	47.263	0.092
DUN	G-DRG	2.695	1.500	4.409	2.172	4.000	0.586	0.427
	FFS	0.333	0.300	3.210	6.031	3.790	1.830	0.144
SDA	G-DRG	0.500	1.100	2.517	8.106	0.445	23.991	0.094
	FFS	0.155	0.010	1.588	6.367	0.515	21.382	0.091
ABN	G-DRG	5.052	6.260	3.894	7.620	0.547	15.848	0.115
	FFS	0.206	0.060	2.266	7.706	1.287	7.809	0.100
TIT	FFS	0.255	0.150	1.686	7.514	0.883	13.437	0.084
HFL	G-DRG	0.534	3.260	1.881	21.097	1.814	16.468	0.033

Source: Author(2015)

Table 4.3: Parameter estimates from the Pareto, Fisk, Gamma and Exponential models to the G-DRG and FFS across the various health service providers claims datasets by maximum likelihood

Dataset	Claims	Parameter Estimate						
		Burr			Weibull		Lognormal	
		α	β	K	α	β	μ	P
BKW	G-DRG	10.840	0.236	0.023	0.888	12.341	1.864	0.719
FFS	2.114	6.931	1.059	1.490	9.766		1.893	0.829
AHM	G-DRG	46.135	8.846	0.059	1.285	19.894	2.545	0.626
	FFS	1.664	7.994	1.351	1.109	9.850	1.771	1.115
DUN	G-DRG	6.787	1.836	0.559	2.803	2.667	0.776	0.353
	FFS	2.474	11.582	3.326	2.282	7.761	1.798	0.554
SDA	G-DRG	15.624	6.189	0.213	1.459	12.030	2.094	0.554
	FFS	1.566	8.822	1.375	1.124	10.632	1.852	1.117
ABN	G-DRG	3288.111	6.251	0.003	2.168	9.939	2.032	0.345

	FFS	2.231	9.926	1.430	1.611	11.017	2.043	0.780
IT	FFS	1.397	19.545	2.604	1.205	12.111	2.019	1.048
HFL	G-DRG	1.342	9.462	20.140	1.344	32.215	3.054	0.895

Given any model, there exists a great deal of theories for making estimates of the model parameters based on the empirical data (Achieng, 2010), in this case, the claims were used as it is to compute the maximum likelihood estimates of the seven sampled distributions. Although the log the inverse and square root of the data were taken to minimize the deviation so as to get a very good model there were some non-positive values showing in the log, double log and triple log of the data making it difficult to model with some of the selected distributions. The square root and the inverse of the data also added not much difference to the deviation hence the data was used without any transformation. The first step used in fitting a model to a claims data was done by finding the parameter estimates of their statistical distribution. When the parameters of any distribution have been obtained using the claims data, then literally, the statistical distribution has been fitted to the claims data (Achieng, 2010). Tables 4.2 and 4.3 show the parameter estimates from the seven models considered: the Pareto distribution (α and β parameters); the Fisk distribution (α and β parameters); the Lognormal distribution (μ and σ parameters); the Exponential distribution (β parameter); the Weibull distribution (α and β parameters); the Burr distribution (α,β and k) and the Gamma distribution (α and β parameters); were fitted to the Ghana Diagnosis Related Groupings (G-DRG) claims for service, and Fee For Service (FFS) claims for medicine in the whole range of health service providers in the municipality. The parameters obtained were then used in the computation of the log-likelihoods statistics of the seven distributions as shown in Table 4.4.

4.3.3 Computation of Log-likelihood Statistics

The log-likelihood for the claims dataset was then calculated by the maximum likelihood method after the parameters were obtained as shown in Table 4.4 below.

Table 4.4: Log-Likelihood Statistics for Pareto, Fisk, Gamma, Exponential, Burr, Weibull and the Lognormal Distribution Fitted for the Claims Dataset

The G-DRG claims for BKW recorded least and highest log-likelihood values at Pareto and Burr

Dataset	Claims	Pareto	Fisk	Gamma	Exponential	Burr	Weibull	Lognormal
BKW	G-DRG	-13903.30	-2092.40	-1209.55	-5627.33	954622.10	-4770.63	7775.06
	FFS	-13649.70	598.49	-4553.14	-5004.02	408.79	-6180.207	4128.62
AHM	G-DRG	-8632.24	2298.3	1 -498.50	-1542.81	1490.40	-1764.27	3148.79
	FFS	-3307.8	-1393.01	-363.95	-1534.38	-865.87	-1886.98	1946.74
DUN	G-DRG	-5197.18	1453.75	-7222.85	-1812.87	3088.90	-2449.08	501.09
	FFS	-4483.04	3689.33	-7063.85	-1872.00	2519.77	-2027.75	2426.28
SDA	G-DRG	-9673.00	4536.53	-1379.21	-3081.46	76687.58	-3697.37	4906.72
	FFS	-6424.33	-2568.35	-1551.82	-3004.29	-2039.84	-3363.97	4058.64
ABN	G-DRG	-29895.70	10264.89	-1336.65	-2434.17	-Inf	-7350.41	3726.32
	FFS	-5839.67	1544.10	-3048.45	-2371.34	1929.17	-2668.07	3655.63
TIT	FFS	-3426.29	-797.72	-1196.89	-1353.47	-789.10	-1454.53	2051.4923
HFL	G-DRG	-1213.52	177.40	-467.62	-262.40	7.20	-267.82	660.77

distributions respectively. The lognormal distribution had the highest log-likelihood value across the entire FFS claims indicating the best model among the seven candidate model. The Fisk and burr distributions also recorded the highest log-likelihood statistics for the G-DRG claims for ABN and (DUN and SDA) respectively whilst, the lognormal recorded the highest log-likelihood values for the G-DRG claims for AHM and HFL.

4.4 Criteria for Choosing the Best Fit Distribution

The selection criterion used was AIC (Akaike, 1973) and SBC (Schwarz,1978) to choose the best out of the candidate model. However, the AIC and SBC in this cases provided equivalent results to the model as shown in Table 4.5 and 4.6.

4.4.1 The Schwarz Criterion

The SBC as was computed after the log likelihoods statistics values for the various distributions were obtained. The results are shown in Table 4.5 below

Table 4.5: SBC Statistics for Seven Candidate Models, Fitted for the Claim Datasets in the Entire Domain

Dataset	Claims	Pareto	Fisk	Gamma	Exponential	Burr	Weibull	Lognormal
BKW	G-DRG	-13912.00	-2101.08	-1218.23	-5631.67	954609.00	-4779.31	7766.38
	FFS	-13658.30	589.90	-4561.73	-5008.32	395.90	-6188.79	4120.02
AHM	G-DRG	-8639.57	2290.98	-505.83	-1546.48	149051.40	-1771.61	3141.46
	FFS	-3315.22	-1400.35	-371.29	-1538.05	-876.88	-1894.32	1939.40
DUN	G-DRG	-5204.68	1446.25	-7230.36	1816.62	3077.65	-2456.59	493.59
	FFS	-4490.57	3681.80	-7071.39	-1875.76	2508.47	-2035.29	2418.75
SDA	G-DRG	-9681.03	4528.50	-1387.24	-3085.48	76675.50	-3705.40	4898.68
	FFS	-6432.34	-2576.36	-1559.83	-3008.30	-2051.85	-3371.98	4050.63
ABN	G-DRG	-29903.50	10257.10	-1344.45	-2438.07	-Inf	-7358.20	3718.53
	FFS	-5847.44	1536.33	-3056.22	-2375.22	1917.52	-2675.847	3647.86
TIT	FFS	-3433.50	-804.93	1204.10	-1357.07	-799.91	-1461.74	2044.28
HFL	G-DRG	-1219.08	171.84	-473.18	-265.18	-1.14	-273.37	655.22

The Lognormal was best for the entire range of FFS claims across the various service providers since it produced the highest SBC value. BKW had the highest SBC value for the FFS claims to be 4120.02, AHM 1939.4, DUN 2418.75, SDA 4050.63, ABN 3647.86 and TIT 2044.283 all occurring under the Lognormal distribution. However, the G-DRG claims for service were best model by different distributions across the various service providers. The G-DRG claims for BKW, DUN and SDA recorded the highest SBC value of 954609, 3077.65 and 76675.5 respectively at the Burr distribution. However the G-DRG claims for ABN recorded the highest SBC value of 10257.1 at the Fisk distribution. The lognormal had the highest value of SBC for both

G-DRG claims for AHM and HFL with values 3141.46 and 655.216 respectively. The FFS claims tariffs across the various service providers' remains almost the same so it could have been a reason why they all fit a particular distribution and the large data size could have been a reason why it is following the lognormal distribution. However, the tariffs for the GDRG claims vary across the various service providers giving rise to the different types of distribution being best

Dataset	Claims	Pareto	Fisk	Gamma	Exponential	Burr	Weibull	Lognormal
BKW	G-DRG	27810.54	4188.81	2423.09	11256.66	-1909238.00	9545.26	-15546.10
	FFS	27303.29	-1192.98	9110.28	10010.03	-811.58	12364.40	-8253.20
AHM	G-DRG	17268.48	-4592.63	1001.00	3087.62	-298118.80	3532.55	-6293.60
	FFS	6619.76	2790.01	731.91	3070.76	1737.70	2790.01	-3889.50
DUN	G-DRG	10398.36	-2903.50	14449.71	3627.74	-6171.80	4902.17	-998.18
	FFS	8970.08	-4848.56	14131.70	3746.00	-5033.54	4059.50	-7374.70
SDA	G-DRG	19350.00	-9069.06	2762.42	6164.93	-153369.00	7398.74	-9809.43
	FFS	12852.66	5140.70	3107.65	6010.58	4085.68	6731.95	-8113.30
ABN	G-DRG	59795.49	-20525.78	2677.30	4870.35	Inf	14704.81	-7448.64
	FFS	11683.34	-3084.19	6100.90	4744.67	-3852.34	5340.14	-7307.30
TIT	FFS	6856.58	1599.44	2397.78	2708.94	1584.20	2913.06	-4099.00
HFL	G-DRG	2431.04	-350.80	939.25	526.80	-8.39	539.63	-1317.50

for different type service providers.

4.4.2 The Akaike Information Criterion

The AIC as was computed after the log likelihoods values for the various distributions were calculated as shown in Table 4.6.

Table 4.6: AIC Statistics for Seven Candidate Models, Fitted for the Claim Datasets in the Entire Domaint

The Lognormal was best again for the entire range of FFS claims across the various service providers since it produced the least AIC values. BKW had the least AIC value for the FFS claims to be -8253.2, AHM with least AIC of -3889.5, DUN with -7374.7, SDA with AIC least value of -8113.3, ABN with least AIC value of -7307.3 and TIT with the least AIC value of -4099 all occurring under the Lognormal distribution. The G-DRG claims for service were best model by different distributions across the various service providers.

The G-DRG claims for BKW, DUN and SDA recorded the least AIC value of -1909238, -6171.8 and -153369 respectively at the Burr distribution. However the G-DRG claims for ABN recorded the least AIC value of -20525.78 at the Fisk distribution. The lognormal had the least value of AIC for both G-DRG claims for AHM and HFL with values -6293.6 and -1317.5 respectively. Comparing the results obtained in Table 4.7 and 4.8 it confirms that the FFS claims tariffs across the various service providers' remains almost the same so it is the reason why they all fit a particular distribution and the large data size is a reason why the data for the FFS is following the lognormal distribution. However, the tariffs for the G-DRG claims vary across the various service providers rise to the different types of distribution been best for different types of service providers. The NHIS does not have any different forms of portfolios or packages since it is a pro-poor policy not aimed at generating profit. Here regardless of the premium paid or not the benefit package is the same across the client or the insured. The only thing that differs is the tariffs across the various health service providers.

4.5 Testing of the Goodness of Fit

The Anderson-Darling Test and the Komogorov-Smirnov (K-S) test were used in testing the goodness of fit of the distribution to know whether indeed the model selected in the previous sections really fit the supposed chosen distributions. This is shown below in Table 4.7 and 4.8.

Table 4.7: Empirical KS statistics for the seven candidate models in the entire domain of the G-DRG and FFS claims with their respective p-values (in parenthesis) at 0.05 level of

significant

Table 4.8 illustrates the Kolmogorov-Smirnov goodness of fit test with their respective Pvalues in parenthesis. The P-value is calculated as a fraction of synthetic data set with a K-S Statistic greater than the empirical K-S Statistic. The Null hypotheses is then rejected at 5% level of significance if p-value < 0.05. The K-S test shows that the lognormal actually fits all the FFS claims perfectly at 5% significant level which confirms the analysis done earlier with the AIC and SBC selection criterion. The p-values for FFS claims for BKW, AHM, DUN, SDA, ABN and TIT were shown be 0.7467, 0.9956, 0.8028, 0.1779, 0.8399, and 0.6125 respectively at 0.05 level of significant. This clearly shows that in fitting the FFS claims to the lognormal distribution is not significant at 5% significance level and the hypothesis is not rejected. Table 4.8 further reveals that the burr distribution is also the best and most accurate in fitting the G-DRG claims from BKW, DUN, and SDA with respective their p-values of 0.9628, 0.9746 and 0.9379. The G-DRG claims for AHM and HFL with the best p-values of 0.9995 and 0.9154 respectively were also confirmed to have been best fitted by the lognormal distribution. The Fisk distribution fitted the G-DRG claims for ABN

Data	Claims	Fisk	Gamma	Exponential.	Burr	Lognormal
BKW	G-DRG	0.4254(0.0303)	0.6490(0.011)	0.4638(<0.0010)	0.0071(0.9628)	0.4540(0.0100)
	FFS	0.0241(0.0049)	0.1503(0.0201)	0.1303(0.1302)	0.0239(0.0041)	0.0096(0.7467)
AHM	G-DRG	0.3837(0.0011)	0.594(0.033)	0.3875(0.011)	0.5277(0.0021)	0.0091(0.9995)
	FFS	0.1378(0.0033)	0.4254(<0.0000)	0.1763(<0.0000)	0.0925(0.0012)	0.0104(0.9956
DUN	G-DRG	0.3132(0.0012)	0.3339(0.0089)	0.4729(0.0100)	0.0112(0.9746)	0.3116(0.0491)
	FFS	0.0655(0.0012)	0.0316(0.0090)	0.2481(<0.0001)	0.0353(0.0498)	0.0148(0.8028)
SDA	G-DRG	0.3970(0.0067)	0.5665(0.0097)	0.4422(<0.0002)	0.0096(0.9379)	0.4134(0.0358)
	FFS	0.0971(0.0079)	0.2412(<0.0001)	0.0958(0.0000)	0.0698(0.0497)	0.1779 (0.1779)
ABN	G-DRG	0.0139(0.7327)	0.5929(0.0000)	0.5143(0.0000)	0.4902(0.0432)	0.4291(0.0427)
	FFS	0.1036(0.0113)	0.1719(0.0013)	0.2186(<0.0001)	0.0746(0.0355)	0.0126(0.8399)
TIT	FFS	0.1260(0.0041)	0.1237(<0.0000)	0.0960(<0.0001)	0.0813(0.0069)	0.0205(0.6125)
HFL	G-DRG	0.1303(0.0499)	0.1692(0.0123)	0.1243(0.0001)	0.1555(0.0111)	0.0340(0.9154)

best than all the other distributions with a p-value of 0.7327. This shows clearly that the lognormal is very powerful and stands best among the entire seven candidate model in fitting distribution to the FFS claims across all the various types of providers. The burr distribution, the lognormal and the Fisk distribution were also good in fitting distributions to the G-DRG claims but however varies across service providers due to the difference in tariffs and services rendered. The other distributions (Gamma, Weibull, Pareto, and Exponential) were all rejected by the test.

Table 4.8: Comparative analysis of the K-S and A-D goodness of fit test for the fitted distribution to the claims dataset

Claims data	α value	Fitted Dist.	K-S	P-Value	Critical Value	A-D	P-Value	Critical Value
BKW-G-DRG	0.05	Burr	0.00713	0.96281	0.01938	0.83	0.034	0.6508
BKW-FFS	0.05	Lognormal	0.00957	0.74667	0.01923	0.314	0.545	2.5018
AHM-G-DRG	0.05	Lognormal	0.00913	0.99946	0.03881	0.161	0.947	2.5018
AHM-FFS	0.05	Lognormal	0.01042	0.99563	0.03461	0.184	0.909	2.5018
DUN-G-DRG	0.05	Burr	0.01122	0.97455	0.0319	0.985	0.014	0.3142
DUN-FFS	0.05	Lognormal	0.01478	0.8028	0.0314	0.584	0.128	2.5018
SDA-G-DRG	0.05	Burr	0.00957	0.93799	0.02447	0.201	>0.250	2.5018
SDA-FFS	0.05	Lognormal	0.17789	0.17789	0.02479	1.866	<0.005	1.2333
ABN-G-DRG	0.05	Fisk	0.01392	0.7327	0.02764	0.368	>0.250	2.5018
ABN-FFS	0.05	Lognormal	0.01262	0.83992	0.0279	0.247	0.755	2.5018
TIT-FFS	0.05	Lognormal	0.02052	0.61252	0.03695	0.742	0.053	2.5018
HFL-G-DRG	0.05	Lognormal	0.03401	0.91536	0.08438	0.247	2.5018	2.5018

Table 4.9 shows the comparative analysis of the Kolmogorov-Smirnov and Anderson- Darling goodness of fit test for the fitted distribution to the claims dataset across the entire service providers in the municipality. This was done because of the advantages and the shortfalls of each of these test cited in chapter three to know actually whether the distributions selected by the information criterion really fit our dataset. A-D Test is more sensitive than the K-S Test near the tail of the distribution than the centre. The hypothesis regarding the distributional form (the data follows the specified distribution) is rejected at 0.05 significance level if the test statistic AD or D (for both A-D Test and K-S Test respectively) is greater than the critical value obtained from a Table or at 5% level of significant if the p-value is less than 0.05. The K-S goodness of fit test rejected none of the distributions earlier selected as best fit by the AIC and BIC this reaffirm the claim that the best distribution for the FFS claims is the lognormal distribution whilst the Burr,

Fisk and the Lognormal were the best for G-DRG claims across the various service providers. The burr distribution was the best for

G-DRG claims from BKW, DUN and SDA. The Fisk distribution was also best for the G-

DRG claims from ABN and the lognormal again was best for G-DRG claims from AHM and HFL. The A-D goodness of fit test however rejected the claim that the Burr distribution was the best for G-DRG claim for BKW and DUN, and also the lognormal distribution was the best distribution for FFS claims from SDA. However the distributions were finally tested graphical by an exploratory data analysis technique (Probability plots to conclude on our modeling process). Table 4.9 displays estimates of the parameters of the fitted distribution and 95% confidence interval of the estimates for the various dataset across the various service providers.

Table 4.9: Estimated Parameters of the Fitted Distribution by MLE

Dataset	Claims	Fitted Distribution	Fitted Parameter	
BKW	G-DRG	Burr	$\hat{\alpha}$	107.350±0.209
			$\hat{\beta}$	4.822±0.001
			$\hat{k}$	0.032±0.005
BKW	FFS	Lognormal	$\hat{\mu}$	1.893±0.008
			$\hat{\sigma}$	0.820±0.0022
AHM	G-DRG	Lognormal	$\hat{\mu}$	2.545±0.009
			$\hat{\sigma}$	0.626±0.016
AHM	FFS	Lognormal	$\hat{\mu}$	1.771±0.014
			$\hat{\sigma}$	1.115±0.086
DUN	G-DRG	Burr	$\hat{\alpha}$	6.787±0.860
			$\hat{\beta}$	1.836±0.021
			$\hat{k}$	0.150±0.008
DUN	FFS	Lognormal	$\hat{\mu}$	1.798±0.003
			$\hat{\sigma}$	0.554±0.005
SDA	G-DRG	Burr	$\hat{\alpha}$	15.624±0.507
			$\hat{\beta}$	6.180±0.003

			$\hat{k}$	0.051±0.006
SDA	FFS	Lognormal	$\hat{\mu}$ $\hat{\sigma}$	1.852±0.0015 1.117±0.098
ABN	G-DRG	Fisk	$\hat{\alpha}$ $\hat{\beta}$	3.894±0.002 7.620±0.001
ABN	FFS	Lognormal	$\hat{\mu}$ $\hat{\sigma}$	2.043±0.008 0.780±0.020
TTT	FFS	Lognormal	$\hat{\mu}$ $\hat{\sigma}$	2.019±0.015 1.048±0.074
HFL	G-DRG	Lognormal	$\hat{\mu}$ $\hat{\sigma}$	3.05±0.030 0.895±0.101

The lognormal distribution was used in fitting 8 out of 12 datasets, it was used in modelling all the FFS claims and the G-DRG claims for AHM and HFL. The μ had higher values across the entire providers than the $\hat{\sigma}$. The Burr distribution however was used to model the G-DRG claims from BKW, DUN and SDA. The $\hat{\alpha}$ which is the first shape parameter is higher than all the remaining parameters with estimated values of 107.350, 6.787, and 15.624 for BKW, DUN and SDA respectively. The scale parameter $\hat{\mu}$ had 4.822, 1.836, and 6.189 for estimates of BKW, DUN, and SDA respectively. Moreover, the second shape parameter k had the least parameter estimates for the burr distribution expect for G-DRG claim for BKW. The k-cap values for BKW, DUN and SDA were 0.032, 0.159, and 0.051 respectively. Finally, the Fisk distribution was used in modeling the G-DRG claim for ABN with its $\hat{\alpha}$ and $\hat{\mu}$ values as 3.894 and 7.620 respectively.

4.6 Validating the Best Fit Models Graphically

This section is interested in the post Model selection first to affirm the selected models. Achieng (2010) argued that the central problem in analysis is the kind of model one needs to use for making inferences from the claim dataset. This is known as the model selection problem (Achieng, 2010). As already shown and proven in both the selection criterion and the goodness of fit test in the earlier sections 4.3 and 4.4 respectively, the log-normal distribution emerged as having the highest log-likelihood, least AIC value and the highest BIC values amongst the seven distributions for all FFS claims across the entire providers in the municipality, also it was not rejected by the K-S non-parametric goodness of fit test for all FFS claims. However, the A-D test rejected the claim that the lognormal was the best for FFS claims of SDA. The Burr, Fisk and the Lognormal were the best for G-DRG claims across the various service providers. The Burr distribution was the best for G-DRG claims from BKW, DUN and SDA. The Fisk distribution was also best for the G-DRG claims from ABN and the lognormal again was best for G-DRG claims from AHM and HFL. Again the A-D goodness of fit test however rejected the claim that the Burr distribution was the best for G-DRG claim for BKW and DUN, and also the lognormal distribution was the best distribution for FFS claims from SDA. This could not have meant they were the best statistical distributions to model the FFS claims and G-DRG claim data across the entire service providers.

As already argued by Nolan (2003) that the P-P plots allow a comparison over the range of the data whiles Q-Q plots not as satisfactory for comparing heavy tailed data to the proposed fit and for technical reasons he recommend the "variance stabilized" P-P plot (as cited in Kallah-Dagadu (2013)), it was necessary to validate goodness of fit test in order to select a statistical distribution that best fits the data. In this study, this was established by the Probability plots to affirm the selected models graphically. The Probability Plots for each of the fitted distribution were constructed using Minitab 16, and the selection criterion was based on the specific hypothesis for each specified model. That is;

H_0: The statistical distribution provides the correct statistical model for the claims data H_1: The statistical distribution does not provide the correct statistical model for the claims data

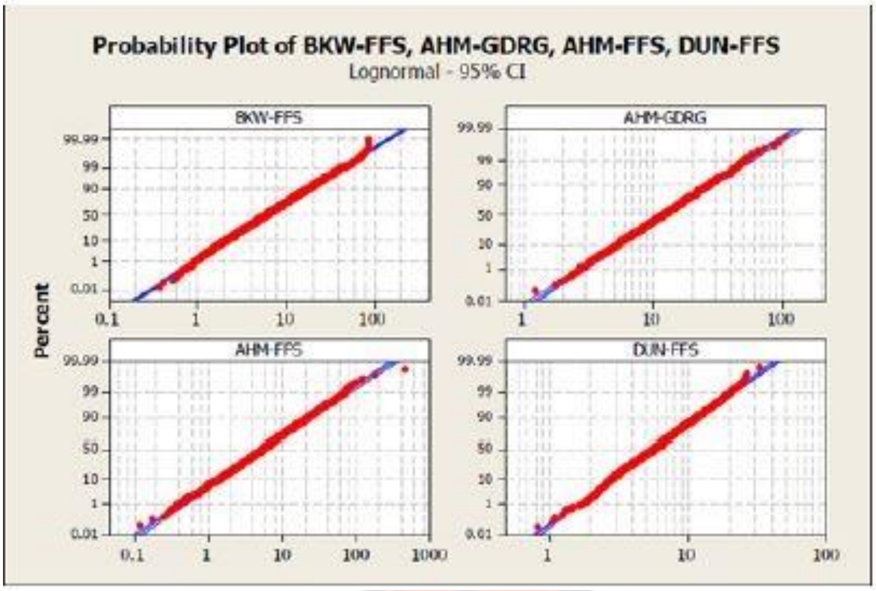

Figure 4.5: Probability plot of BKW-FFS, AHM-G-DRG, AHM-FFS and DUN-FFS

The Probability plots in Figure 4.6 depicts a very good fit for the BKW-FFS, AHM-G-DRG, AHM-FFS and DUN-FFS claims with almost all the data points falling onto or around the reference line. The null hypothesis was therefore not rejected in both plots and a conclusion was made that at 95% confidence level, that the lognormal distribution does provide the correct statistical model for the BKW-FFS, AHM-G-DRG, AHM-FFS and DUN-FFS claims data.

The Probability plots in Figure 4.7 depicts a very good fit for the SDA-FFS, ABN-FFS, TIT-FFS and HFL-G-DRG claims with almost all the data points falling onto or around the reference line. The null hypothesis was therefore not rejected

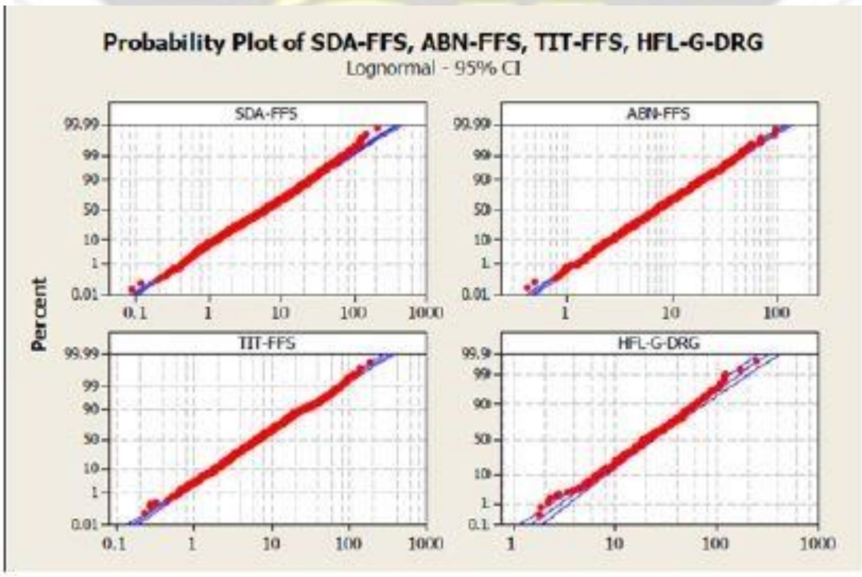

Figure 4.6: Probability plot of SDA-FFS, ABN-FFS, TIT-FFS and HFL-G-DRG

in both plots and a conclusion was made that at 95% confidence level, that the lognormal distribution does provide the correct statistical model for the SDA-FFS, ABN-FFS, TIT-FFS and HFL-G-DRG claims data.

The Probability plots in Figure 4.8 depicts a very good fit for the BKW-GDRG, DUNG-DRG, and SDA-G-DRG claims with almost all the data points falling onto or around the reference line with very few outliers at the lower tail end. The null hypothesis was therefore not rejected in both plots and a conclusion was made that at 95% confidence level that the Burr distribution does provide the correct statistical model for the BKW-G-DRG,

DUN-G-DRG, and SDA-G-DRG claims data.

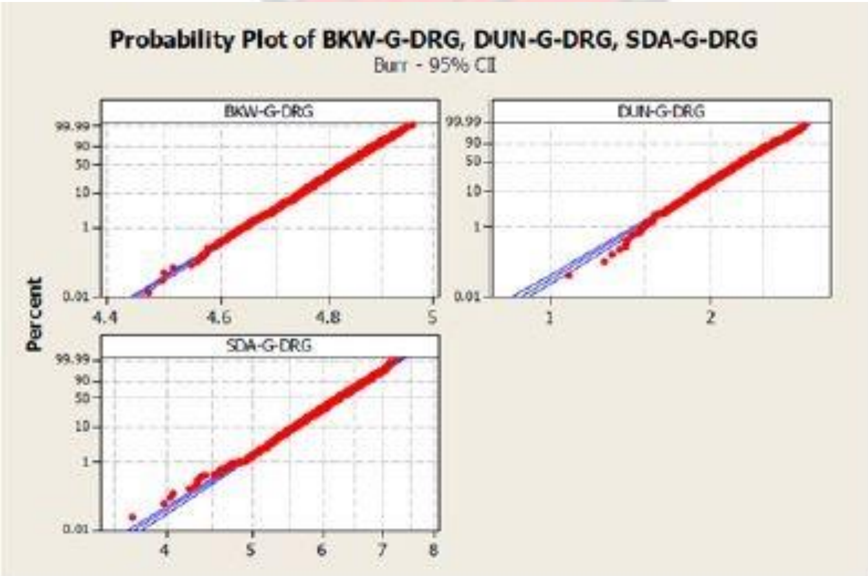

Figure 4.7: Probability plot of BKW-G-DRG, DUN-G-DRG, and SDA-G-DRG claims

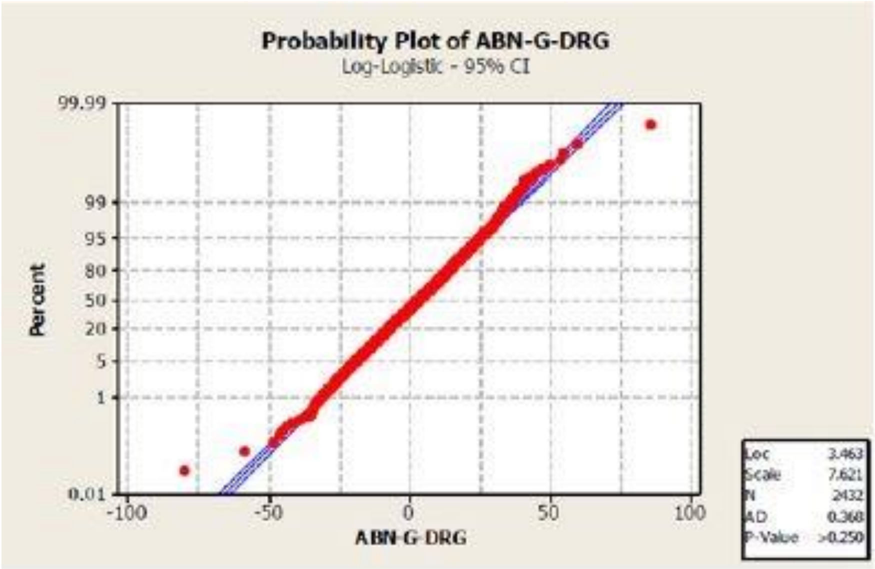

Figure 4.8: Probability plot of ABN-G-DRG claims

The Probability plots in Figure 4.10 depicts a very good fit for the ABN-GDRG claims with almost all the data points falling onto or around the reference line. The null hypothesis was therefore not rejected in both plots and a conclusion was made that at 95% confidence level, that the Fisk distribution does provide the correct statistical model for the ABN-G-DRG claims data. According to NHIA (2011) claims that are been submitted should have at least a turnaround time of an average of 60-90 days. Thus claims are supposed to be vetted and paid to allow service providers operate effectively and efficiently within the stipulated turnaround time. However, this is not the case due to the lack of human resource and the paper base process (manual) of claims vetting in almost all the Schemes. When there is a delay in claims vetting by some NHIS and funds are available for payment the scheme management pay in most cases 100 percent of total claims submitted and deductions are then from the subsequent payments if there were some rejected or error claims. This policy however affects the service providers adversely in their subsequent budget and administrative duties creating a large gap should there be large rejections in the previously paid un-vetted claims.

This main aim of this study was to use claims data from the National health insurance scheme in Amansie East Municipality to determine appropriate statistical models for claim amounts by fitting theoretical distributions to the G-DRG and FFS claims data. The study considered seven different service providers namely; Bekwai Municipal Hospital representing district level hospitals, Ahmadiya Muslim Hospital to representing private and Ahmadiya

Hospitals, Dunkrah Health Centre to represent CHPS compound, community clinics and Health Centers, SDA hospital Dominase to represent CHAG hospitals, Abenkyiman clinic to represent Public clinics and maternity homes, Titan Chemist to represent the chemical and pharmacy shops, and Health Frontier Limited to represent Ultra Scan and diagnostic centers for July, 2011. The result of the study demonstrates that some service providers are indeed better modeled with different distributions. The study revealed that the claim data are not normally distributed but rather asymmetric as seen in the exploratory data analysis and the histogram plots. The Gamma, the Weibull, the Pareto, and the Exponential distributions were rejected by the entire selection criterions (AIC and SBC). The Kolmogorov-Smirnov and Anderson-Darling goodness of fit Test also rejected the Gamma, the Weibull, the Pareto, and the Exponential distributions. It was revealed that the claims data considered, come from a leptokurtic, heavy tails and asymmetry distribution, and the FFS claims for all service providers are log-normally distributed. The GDRG claims however was best modeled with three different distributions; namely the Lognormal, the Burr and the Fisk distribution. The

Kolmogorov-Smirnov goodness of fit tests failed to reject the Lognormal and the other two (thus the Burr and the Fisk) distributions as being the best fit model to the claims data considered. The Anderson-Darling goodness of fit Test however rejected the Burr distribution as being the best for G-DRG claim for BKW and DUN, and also the lognormal distribution as being the best distribution for FFS claims from SDA. However the distributions were finally tested graphical by an exploratory data analysis technique (Probability plots) to conclude on the modeling process. The FFS claims data considered were modeled with the lognormal distribution with the following estimates; BKW μ= 1.893±0.008 and σ^2= 0.829±0.022,

AHM μ=1.771±0.014 and σ=1.115±0.086, DUN μ=1.798±0.003 and σ=0.554±0.005, SDA μ=1.852±0.015 and σ=1.117±0.0098, ABN μ=2.043±0.008 and σ=0.780±0.020 and TIT μ=2.019±0.015 and σ=1.048±0.074 The G-DRG claims data considered were also then modeled with lognormal distribution with the following estimates; AHM μ=2.545.009 and σ=0.626±0.016 and HFL μ=3.054±0.030 and σ=0.895±0.0101; it was again modeled with the Fisk distribution for the G-DRG Claims for ABN with estimates α= 3.894±0.002 and β= 7.620±0.001, Finally, the Burr distribution were used to model the remaining G-DRG claims and the following estimates were observed for BKW α= 107.350±0.209 and β= 4.822±0.001 and k =0.032±0.005, DUN α= 6.787±0.860 and β=1.836±0.0021 and k =0.159±0.008 and

SDA α= 15.624±0.507 and β= 6.189±0.003 and k =0.051±0.006

CHAPTER 5

Summary, Conclusion and Recommendation

5.1 Introduction

This chapter provides a summary of the entire study. The main findings as well as the conclusion drawn from the study are highlighted. That is, this chapter presents summary of the findings from the study, and recommends rational measures for stakeholders, actuaries and insurance managers. The chapter further provides policy recommendation based on the findings of the study and suggestions for further studies. The chapter provides the concluding statements of the research based on the findings.

This main aim of this study was to use claims data from the National health insurance scheme in Amansie East Municipality to determine appropriate statistical models for claim amounts by fitting theoretical distributions to the G-DRG and FFS claims data. The study considered seven different service providers namely; Bekwai Municipal Hospital representing district level hospitals, Ahmadiya Muslim Hospital to representing private and Ahmadiya

Hospitals, Dunkrah Health Centre to represent CHPS compound, community clinics and Health Centers, SDA hospital Dominase to represent CHAG hospitals, Abenkyiman clinic to represent Public clinics and maternity homes, Titan Chemist to represent the chemical and pharmacy shops, and Health Frontier Limited to represent Ultra Scan and diagnostic centers for July, 2011. The result of the study demonstrates that some service providers are indeed better modeled with different distributions.

The study revealed that the claim data are not normally distributed but rather asymmetric as seen in the exploratory data analysis and the histogram plots. The Gamma, the Weibull, the Pareto, and the Exponential distributions were rejected by the entire selection criterions (AIC and SBC). The Kolmogorov-Smirnov and Anderson-Darling goodness of fit Test also rejected the Gamma, the Weibull, the Pareto, and the Exponential distributions. It was revealed that the claims data considered, come from a leptokurtic, heavy tails and asymmetry distribution, and the FFS claims for all service providers are log-normally distributed. The GDRG claims however was best modeled with three different distributions; namely the Lognormal, the Burr and the Fisk distribution. The Kolmogorov-Smirnov goodness of fit tests failed to reject the Lognormal and the other two (thus the Burr and the Fisk) distributions as being the best fit model to the claims data considered. The Anderson-Darling goodness of fit Test however rejected the Burr distribution as being the best for G-DRG claim for BKW and DUN, and also the lognormal distribution as being the best distribution for FFS claims from SDA. However the distributions were finally tested graphical by an exploratory data analysis technique (Probability plots) to conclude on the modeling process.

The FFS claims data considered were modeled with the lognormal distribution with the following estimates; BKW μ= 1.893±0.008 and σ^2= 0.829±0.022, AHM μ=1.771±0.014 and σ=1.115±0.086, DUN μ=1.798±0.003 and σ=0.554±0.005, SDA μ=1.852±0.015 and σ=1.117±0.0098, ABN μ=2.043±0.008 and σ=0.780±0.020 and TIT μ=2.019±0.015 and σ=1.048±0.074 The G-DRG claims data considered were also then modeled with lognormal distribution with the following estimates; AHM μ=2.545.009 and σ=0.626±0.016 and HFL μ=3.054±0.030 and σ=0.895±0.0101; it was again modeled with the Fisk distribution for the G-DRG Claims for ABN with estimates α= 3.894±0.002 and β= 7.620±0.001, Finally, the Burr distribution were used to model the remaining G-DRG claims and the following estimates were observed for BKW α= 107.350±0.209 and β= 4.822±0.001 and k =0.032±0.005, DUN α= 6.787±0.860 and β=1.836±0.0021 and k =0.159±0.008 and SDA α= 15.624±0.507 and β= 6.189±0.003 and k =0.051±0.006

5.2 Conclusion

This study has examined the categories of health service providers' claims data from the National health insurance scheme in Amansie East Municipality and has fitted appropriate theoretical distribution to each G-DRG and FFS claims data. After carrying out each step in the actuarial modeling processes with diligence and accuracy the study has established that two parameter probability distribution(lognormal) was good in modeling the FFS claims for drugs across the various service providers. The three parameter Burr distribution was useful in modeling the G-DRG claims for service for the District level hospital (BKW), the CHPS compound, community clinics and Health Centers (DUN) and CHAG hospitals (SDA). It was found that the two parameter Fisk distribution was also used in modeling the G-DRG claims for Public clinics and maternity (ABN), the two parameter lognormal was again used in modeling the G-DRG claims for private and Ahmadiya Hospitals (AHM) and chemical and pharmacy shops (TIT).

5.3 Recommendation

Management at all municipal and district health insurance schemes should be able to apply the appropriate statistical distribution used in this research for management policy prescription to improve their performance. Further studies should be conducted to use the probabilistic distribution to model appropriate severity and claim to make payment of claims easy and fast.

Reference

Adei, D., Osei, V. K., & Diko. S.K. (2012). An assessment of the Kwabre district mutual health insurance scheme in Ghana. Current Research Journal of SocialSciences 4(5): 372-382.

Annual Report (2014) on the National Health Insurance Scheme. National Health Insurance Authority (NHIA), Accra.

Ashiagbor, A.A. (2014). Bootstrapping malmquist productivity change indices of Ghanaian insurers.Unpublished thesis.School of graduate studies, University of Ghana.

Consul, P.C., Famoye, F. (2006). Lagrangian probability distributions. Birkhauser, Boston, USA

Dadey.E, Ablebu. G, & Agboda. K, (2011). Probability modeling and simulation of insurance claim: Koforidua Polytechnic. Unpublished manuscript

Daykin, C.D. Pentikainen, T. Pesonen, M. (1994). Practical risk theory for actuaries, Chapman & Hall, London. Dimakos, X. & Frigessi, A. (2002). Bayesian premium rating with latent structure. Scandinavian Actuarial Journal, 2002(3):162184.

Gobah, F. F., & Zhang, L. (2011). The national health insurance scheme in Ghana: Prospects and challenges: a cross-sectional evidence. Global Journal of Health Science

Government of Ghana (GoG), 2003. National Health Insurance Act 650 of 2003.

Government of Ghana (GoG), 2004. National Health Insurance Regulations,

Goovaerts, M., Dhaene, J., & Rachedi, O. (2010). Actuarial premium principles.In Encyclopedia of Quantitative Finance. John Wiley and Sons.

Gschlobl, S. & Czado, C. (2007). Spatial modeling of claim frequency and claim size in non-life insurance. Scandinavian Actuarial Journal, 2007(3):202225.

Hanhn, G. J. & Shapiro, S. S. (1967). Statistical models in engineering. New York: John Wiley & Sons, Inc.

Migon, H. & Moura, F. (2005). Hierarchical Bayesian collective risk model: an application to health insurance. Insurance: Mathematics and Economics, 36(2):119135.

Migon, H. & Gamerman, D. (1993). Generalized exponential growth modelsa Bayesian approach. Journal of Forecasting, 12(7):573584

Migon, H. & Moura, F. (2005). Hierarchical Bayesian collective risk model: an application to health insurance. Insurance: Mathematics and Economics, 36(2):119135.

Migon, H. & Penna, E. (2006). Bayesian analysis of a health insurance model.Journal of Actuarial Practice, 13:6180.

Ramberg,J.S.,Tadikamalla,P.R.,Dudewicz, E. J.,Mykytka, E. F.(1979). A probability distribution and its uses in fitting data.Technometrics, Vol. 21, No. 2 pp 201-214Published. http://www.jstor.org/stable/1268517

Sodzi-Tettey,S., Aikins,M. Awoonor-Williams, J. K. & Agyepong, I. A.(2012). Challenges in provider payment under the Ghana national health insurance scheme:A case study of claims management in two districts. Ghana Medical Journal Volume 46, Number 4

Souza, D., Moura, F., & Migon, H. (2009). Small area population prediction via hierarchical models. In Survey Methodology, volume 39 of Catalogue No.12-001-X, pages 203214. Statistics Canada.