MODELLING MEAN LENGTH OF STAY AT THE HOSPITAL. A CASE STUDY OF KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY HOSPITAL, MALE MEDICAL WARD.

By

Adobah Emmanuel

(B.Ed)

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi

in partial fulfilment of the requirement for the degree of

Master of Science

Industrial Mathematics

INSTITUTE OF DISTANCE LEARNING

October, 2011

DECLARATION

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

Emmanuel Adobah(PG 3006309) 02/05/12

Student's Name and ID Signature Date

Certified by:

Nana Kena Frempong 02/05/12

Supervisor's Name Signature Date

Mr. F.K.Darkwah KFDarkwah 23/5/2012

Head of Department's Name Signature Date

Prof. I.K. Dontwi

Dean of Institute of Distance Learning's Name I.K. Dontwi 23/5/2012

Signature Date

ABSTRACT

The purpose of this study was to model mean length of stay (LOS) given several factors at KNUST Hospital (Male ward), Kumasi. The data used was secondary data; it consisted of all patients who had been on admission from January 2005 to December 2009 (male medical ward). Total number of observation for that period was 2206. S PLUS, SPSS and Excel were the statistical packages used for the analysis. It was observed that the mean length of stay over the said period was 2.71 (approximately 3 days). Again, July 2008 recorded the highest admission cases with 79 patients and April 2007 also recorded the least admission cases with 6 patients admitted at the ward (male). Four generalized linear models were formulated and it was the model with the natural log of length of stay which was chosen because it had the lowest Akaike Information Criteria (AIC) value of 4823.5. Four predictor variables were significant at 5% level of significance (age, hernia, pneumonia and sickle cell disease). Patient's age has an influence on his LOS at the ward; aged patient spend more days in the ward while young patient spend few days in the ward. Again patients admitted with pneumonia, hernia and sickle cell disease would also spend more days in the ward.

ACKNOWLEDGEMENT

With my great appreciation and sincere thankfulness, I would like to acknowledge the following people for their help, advice, corporation, encouragement and support which help in finally complete of my study.

First, I would like to express my gratitude to my supervisor; N.K. Frempong for his kindness, guidance, comment, supervision and encouragement throughout the study.

Next, my special thanks go to Dr. S.K. Amponsah, Former Head of Department (Mathematics) for his assistant, advice and encouragement throughout the study. I would like to thank the staff of the University Hospital especially, Medical Director and the nurses of male and female ward for their support and assistant in collection of data to make this study possible. In addition, I would also like to express my great thank to Mr. Nsoyameye, accountant of Nyinahin Senior High School (SHS) and Mr. Owusu Afriyie, postgraduate student of University of Cape Coast (UCC) for their special assistant and support to me throughout my educational carer.

Then, profound thanks to my classmates in the Master's Degree program at KNUST for their love, encouragement and assistant throughout the study.

Finally, my deepest thanks go to my family particularly my beloved mother, brothers, not forgetting my dear uncle Mr. Kwaku Ofori – Duro, Principal of Young Women Christian Association (YWCA) and all dear friends for their understanding, inspiration, consultant and encouragement.

DEDICATION

I dedicate this work to my family especially my late sister Mercy Owusuaah.

TABLE OF CONTENTS

	PAGE
Declaration	i
Abstract	ii
Acknowledgment	iii
Dedication	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
List of Acronyms/Abbreviations	xi
CHAPTER ONE	**1**
Introduction	1
1.1 Background of the Study	1
1.1.1 Brief History about the Study Area	5
1.2 Purpose of the Study	6
1.3 Objective of the Study	7
1.4 Methodology	7
1.5 Significant o the Study	7
1.6 Scope of the Study	8
1.7 Limitation	9
1.8 Organization of the Study	10
CHAPTER TWO	**11**
Literature Review	11
Introduction	11
2.1 Length of Hospital Stay	11

2.2 Management and Motivation of Hospital Staff 16

2.3 Hospital Polices 21

2.4 Regression Analysis and Length of Stay 24

CHAPTER THREE **28**

Methodology 28

Introduction 28

3.1 Data Collection Technique 28

3.3 Sample Size 29

3.2.1 Assumptions 29

3.3 Component of a Regressions Model 29

3.4 Type of Regression Model 30

3.4.1 Simple Regression Model 30

3.4.2 Multiple Regression Models 30

3.4.3 Non – Linear Regression Models 31

3.5 Definition and Feature of Simple Linear Regression Model 31

3.5.1Assumptions of the Model 32

3.5.2 Alternative Model for Simple Linear Regression 32

3.6 Correlation Coefficient 33

3.6.1 Pearson Product-Moment Measure Correlation Coefficient 33

3.7 Multiple Linear Regression Model 33

3.7.1 Assumptions 34

3.7.2Assumed Regression Model 35

3.8 Estimation of Parameters 35

3.8.1 Least Square Estimation 35

3.8.2 Hypothesis Test 38

3.9 Coefficient of Determination 39

3.10 Multiple Coefficient of Determina1ion 40

3.11 Confidence Interval in Multiple Regressions 41

3.11.1Confidence Intervals on the Regression Coefficients 41

3.12 Transformation (Box-Cox Transformation) 42

3.13 Generalized Linear Model 44

3.13.1 Component of Generalized Linear Model 44

3.13.2 The Model 45

3.13.3 The Exponential Family 46

3.13.4 The Link Function 47

3.13.5 Gamma Model with Log Link 47

3.13.6 Normal Log with Identity Link 47

3.14 Gamma with Inverse Link 48

3.16 Akiake Information Criterion 48

CHAPTER FOUR **50**

Data Analysis and Discussion of Results 50

Introduction 50

4.1 Descriptive Analysis of the Data 50

4.2 Interpretation of Output 53

4.2.1 Normal Model with LOS 53

4.2.2 Normal Model with Natural Log on LOS 55

4.2.3 Gamma Model with Log Link on LOS 57

4.2.4 Gamma Model with Inverse Link on LOS 58

CHAPTER FIVE **60**

Discussion, Conclusion and Recommendation 60

5.1 Discussion 60

5.2 Conclusion 62

5.3 Recommendation 63

Reference **64**

Appendices **70**

LIST OF TABLES

	PAGE
Table 4.2: Parameter estimate for model 1	53
Table 4.3: Parameter estimate for model 2	55
Table 4.4 Parameter estimate for model 3	57
Table 4.5 Parameter estimate for model 4	58

LIST OF FIGURES

PAGE

Figure 4.1: Histogram for LOS **51**

Figure 4.2: LOS against types of illness in male ward **52**

LIST OF ACRONYMS/ABBREVIATIONS

AIC	Akaike Information Criterion
ARIMA	Auto Regressive Integrated Moving Average
BPN	Back-Propagation Neural Network
CHI	Community Health Insurance
ED	Emergency Department
FFS	Free For Service
FNN	Feed Forward Neural Network
GLM	Generalized Linear Model
HAI	Hospital Association Infection
ICU	Intensive Care Units
IPD	Inappropriate Patient Days
KATH	Komfo Anokye Teaching Hospital
KNUST	Kwame Nkrumah University of Science and Technology
LOS	Length of Stay
MHO	Mutual Health Organization
MI	Minor Infection
PHI	Private Health Insurance
PHIS	Private Health Insurance Scheme
PRN	Professional Register Nurses
RNs	Register Nurses
SHS	Senior High School
SOA	Systemic Opioid Analgesis
SOM	Self-Organization Map
SPSS (PASW)	Predictive Analytic Software
UCC	University of Cape Coast
VARMA	Vector Autoregressive Moving Average
WHO	World Health Organization
YWCA	Young Women Association

CHAPTER ONE

Introduction

In this chapter we discuss the length of stay of medical patients in the KNUST hospital as the study area and rationale for the study.

1.1 Background of the study

One of the most serious public health problems characterizing the developing world is the length of hospital stay, which is persistent and worsening in some regions despite increasing institutional, national and international efforts aimed at eliminating or reducing the days spent at the hospital.

During the past few years there had been great interest in quality assurance and the setting of standards for health care delivery to ensure that patient's expectations are met. Recently the increasing demand for health care services (due to ageing population and increased prevalence of chronic diseases) together with the increasing cost of providing these services (technological advances, over specialization and in appropriate medical use) supports the need for a reconsideration of the existing practices. Furthermore greater attention has been given by health care providers to the effectiveness and quality of medical care and methods for quality improvements.

Advances in medical diagnosis and treatment have placed a heavy burden on our medical care system. The public, while demanding the benefits of these improved services, is simultaneously unwilling to pay for the concurrent cost increases. In the resulting efforts to find better ways to allocate that facility, equipment and manpower resources necessary for hospital operation, techniques have been developed for scheduling elective admissions, predicting bed needs and measuring bed utilization. One component in these techniques is an accurate prediction of how long a patient will stay in the hospital with respect to certain common diseases and an understanding of the factors that influence his stay.

A hospital is an institution for health care providing patients with treatment by specialized staff and equipment and often but not always providing for longer term patient stays. Patient with certain kinds of diseases or illness tend to stay longer than necessary. It is very common to see hospital management rejecting sick patients who are to be admitted at the ward due to limited facilities at the ward.

In all health care systems, the use of hospital beds is a concern for policy-makers, managers and practitioners. For several years hospital managers have been under pressure to reduce hospital expenditures since it is the largest single source of health service expenditures in most countries. On the other hand, numerous studies have shown that hospital admissions and inpatient days may be unnecessary in certain situations. Thus, improving the efficiency of hospital services through a preventive strategy for inappropriate hospital admission. This stay can lead to increase the productivity of the hospital and reduce the waiting list and optimal use of existing health care facilities without compromising the quality of care. Furthermore, hospital environments increase the risk of Hospital Associated Infection (HAI) for many reasons. The longer a patient is exposed to hospital environment, may increase higher rate of HAI. Eliminating inappropriate hospital stay reduces the cost of healthcare delivery and the risks of conical infection and leaves available resources for patients with more critical conditions.

One of the most interesting approaches to improve the quality and efficiency of hospital use was to evaluate whether their services were used appropriately. The assessment of hospital use is justified by awareness that some of this use may be inappropriate, patients receive either services giving no significant benefit or which would be more suitably given at a different level of care. To date, few studies have been conducted in different countries at different times to determine the length of hospital stay. The current health care climate promotes a decreasing length of stay, with early discharge and care at home actively encouraged.

Despite performing these kinds of studies at different countries, hospital use for presentation of health care services had not been analyzed documentarily in different countries. In view of the importance and necessity of healthcare services quality improvement and with a view to review, the existing structures of hospital management, staff and other policies were surveyed for appropriateness or inappropriateness of hospitalization for this study.

Overcrowding in the emergency department (ED) of a hospital is a common occurrence. Most of us have attended an ED either as a patient or carer and there is widespread appreciation of the problems ~~and~~ issues. Patients and carers are distressed—it is after all an emergency. Doctors and nurses seem to be run off their feet, perhaps at the end of a long shift. Decisions are often made quickly.

Given the increasing demands being placed upon the health services and the likelihood of significant staff shortages, there are serious consequences in both economic resource allocation and patient (and population) health outcomes if decisions about future health service structures are incorrect. Given the recent advances in computing power and the need to improve decision making, there has never been a more opportune time to apply modelling to facilitate improved decision making in the health care sector. One aspect where modelling will become increasingly important in the health sector is in relation to modelling decisions around hospital beds.

Studies on over-hospitalization have relied on measurement of unnecessary hospital admissions and Inappropriate Patient Days (IPD). Appropriate admission is defined as "those patient for whom there is no alternative to admission to the hospital with high-technology facilities .This would be the case even if lower-technology alternatives to hospital admission existed". In other word, there may potentially be a lower-technology alternative to admission to hospital for patients whose admissions are determined as inappropriate (Gertman and Restuccia ,1981).

We decided to avoid defining "short hospitalization" as a fixed number of days, since there is no commonly accepted threshold. Rather, "short hospitalization" was defined as an LOS shorter than the mode for each of the hospital ownership groups. Appropriate hospital stay can be defined as inpatient stay requiring continuous and active medical, nursing or paramedical treatment which could not be provided through external care, day care or outpatient care.

Although most studies worldwide have looked at appropriateness and quality of hospital care, up to now, no study has been focused in this issue in other countries too. In this study an attempt has been made to look at the extent of inappropriate hospital admission and hospital stay in Kwame Nkrumah University of Science and Technology Hospitals in Kumasi. Determining and identification of the leading causes of inappropriate patient stay were another objective followed throughout this study.

1.1.1 Brief History of KNUST Hospital

Kwame University of Science and Technology Hospital was established at 1952 as dressing station which later developed to full status of hospital. It is a department of the University (KNUST) and it is manage by the University. The core responsibility of the hospital is to provide service to senior and junior members as well as communities around.

The hospital has number of sections and units, among the sections includes dental, X-ray, laboratory, pharmacy, nursing, administration and account section. The units also includes records, national health insurance scheme office, public health, theatre and out patient department. It just started a 24 hours service.

KNUST hospital had four wards which were made up male, female, children and maternity ward but in this study, only male medical ward was considered. Children within 12 years and above were admitted at male or female ward. The hospital current had about 96 beds with 27 beds for female ward, 23 beds for male ward, 33 beds for children ward, 11 beds for maternity and rest for emergency ward. All cases beyond the hospital were referred to Komfo Anokye Teaching Hospital (KATH).

1.2 Problem of the study

Length of hospital stay (LOS) is a common parameter used to indicate health resource utilization, health care cost and severity of disease. It can be concluded that patient demographics and hospital characteristics were the two major factors that determine patient LOS. Among demographic characteristics, studies have reported that LOS varied according to age and disease group, where as among hospital characteristics, LOS has been reported to vary by region, hospital size, and health care service which was the main challenge to patient hospital length of stay (Clarke, 2002).

Increasing numbers of emergency admissions put pressure on acute hospital service often precipitating bed crises. This is a major challenge for hospital administrators that need to develop more efficient working practices. Although 'admission avoidance' receives much publicity, on-call doctors are often reluctant not to admit patients, not least because it is quicker and safer to admit rather than discharge a complex patient in whom relevant information to determine safe discharge may not be immediately available.

LOS can be terminated by cure, transfer or death discharge. Many studies have been concerned only with LOS for patients with cured discharge. However, hospital stay terminated by death is not also an important outcome event. Patient care in hospital is the most expensive way of providing palliative care. Longer stays are more likely to indicate physicians' decisions or administrative inefficiencies than patients' need. Reducing LOS is health policy in many countries. Setting up a proper palliative care or providing opportunity for patients to decide where they want to stay during the last stages of their life can reduce unnecessary hospital LOS.

Many people visit hospitals with various diseases which sometimes make them stay longer. Length of stay in hospitals is a major challenge to the patient(s) and the hospital management in terms of occupancy, high cost of medical bills and problem of co-infection.

1.3 Objective of the study

1 To fit generalized linear models to model the effect of age, survival status and types of illness on mean length of stay in the hospital (LOS).

2 To select the best model that explains the variations in length of stay (LOS)

1.4 Methodology

Data for this study was secondary sourced from Kwame Nkrumah University of Science and Technology Hospitals, Kumasi between 2005 and 2009. It was collected from patient admission and discharge books. The generalized linear model(s) was formulated for the analysis. The software package use for statistical analysis was SPLUS, Excel and SPSS.

1.5 Significance of the study

Resource are scarce, as a result, every organization must ensure that effective and efficient utilization of its resources is achieve to yield maximum benefit of its purpose. The core responsibility of the hospital was to ensure that adequate quality health care delivered was given to senior members, junior members and as well as other members of the University community and its environment. The hospital has number of units and departments but the purpose of this study was focused on medical ward only. This is where common diseases like malaria, abortion, typhoid, fever etc are treated. It is our hope that we determine some of these common diseases that prolong the length of stay in the medical ward.

1.6 Scope of the study

Kwame Nkrumah University of Science and Technology hospital (University Hospital) is situated in the Kumasi Metropolitan as capital town in Ashanti region –Ghana sharing boarders with central region to south, Eastern Region to the east, Brong Ahafo Region to the north and Western Region to the west.

Kumasi is the second biggest city of Ghana and it is inhabited by a heterogeneous people from different ethnic groups. It is approximately 250km away from Ghana's National capital, Accra.

Kumasi metropolitan has catchment's estimated population of 1468608 (2000 population and housing censuses) with female forming about 54% of the population and 46% forming male. The main form of occupation in the metropolitan is trading since Kumasi is commercial zone. Social amenities is not a challenge to the metropolitan since it can boast of teaching hospital and one of leading university in the country (KNUST) and number of second cycle institutions as well as Technical and vocational schools.

The study is to examine the length of stay of medical patient in hospital over the period of 2005 to 2009. It considered variables like identity of patient, length of stay, gender, age, ward and year of admission. The research was been carry at male medical ward only.

1.7 Limitation

No study is perfect in design or methodology, and this research is no exception. The results of this study have several limitations.

Due to time (as duration of the programme is two years) and logistics constrains, sample size of five districts hospitals and a teaching hospital in Kumasi only Kwame Nkrumah University of Science and Technology Hospital was considered. Generally, the data collected was restricted to patient admission and discharge books at the medical ward. Again there was no personal interview or questionnaires to ascertain views of patient opinions on the length of stay in the hospital.

The researcher had intension of collecting ten years data but he laid hands on five years data which may increased the error margining of the data. Constrains mentioned above were the main limitations why the research was conducted at KNUST Hospital as well as medical ward.

1.8 Organization of the study

Every research ~~must~~ appropriat~~ely be~~ organised to allow readers to follow the sequence of the study. Accordingly, Chapter one of the study encompasses the background of the study; Statement of the problem; the objective of the study; significance of the study; the scope of the study; limitation; the methodology and organization of the study.

Chapter two reviews the related literature regarding the length of stay in the hospital. Following the review of the literature, chapter three deals with the research methodology, mathematical tools used and computing software used.

Chapter four focuses on data analysis and modelling finally, chapter five present the discussion conclusion and recommendation.

CHAPER TWO

LITERATURE REVIEW

Introduction

The management of every organization has a fundamental responsibility to develop and maintain the image of the organization. The proper stewardship of the organization's resource is an essential responsibility of managers and staff. Hence, management must ensure that the organization programmes are implemented and resources are used efficiently and effectively to achieved desired objectives. The programmes must be implemented and resources used consistently with the organization mission, in compliance with law and regulations and minimal potential for waste, fraud and management (Hiamah, 2007).

Accordingly, in this chapter, literature relating to length of stay in the hospital, the hospital policies, component of hospital, management and motivation and regression analysis related to length of stay were review to the study.

2.1 Length of hospital stay

There is broad agreement that the health service delivery system is not functioning well, but there is considerable disagreement about how to fix it. Some argue for turning over the entire system to a free market. Others recall the benefits of a centrally planned health system.

"The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition" (World Health Organisation (1946)). Thus, in its Constitution, the WHO places substantial responsibilities on the States that are signatories to ensure the effective delivery of health care services in their countries.

According to (Norman, 2003) reviewed that rising demand and a reduction in the number of available beds have greatly increased the turnover in hospital. The pressure to increase this even further has led to the concept of inappropriately delayed discharge.

Hospital occupancy levels have also increased in recent years. The ability to provide services to the growing wave of baby boomers that are expected will not exist, ceterus paribus, unless capacity is increased. Additional capacity, in terms of capital infrastructure, can be created relatively quickly. There is, however, little point in doing so if there are in sufficient staff available to provide services to patients.

Beside average LOS, total numbers of bed days provide useful information on the financial cost of patients who die in hospital. Thus the use of health care services and resources can be identified by considering this aspect, allowing health planning and budget allocations to be fairly distributed to the most suitable groups. In the United States, hospital inpatient stays accounted for nearly one-third of total health care expenses (Machlin and Carper, 2004).

According to (Greta, 2007) reviewed that availability of register nurses (RNs) and their professional nursing skills may also influence patient length of stay (LOS) and hospital costs. Several researchers analyzed the relationship between nurse staffing and LOS. By increasing the number of RNs and providing more RN hours per patient day, patient LOS decreased.

Before proceeding to examine each of these clinical conditions independently, it is important to consider conceptually the ways in which LOS and health outcomes might be related. One must first recognize that the duration of a hospital stay is not a directly manipulable factor in patient management. If LOS is shortened, then treatment schedules must be altered in very specific ways. Some treatments must be foregone, others changed, and others shortened in duration. For example, if a patient with pneumonia is sent home early, one might have to decrease the number of days during which intravenous antibiotics are given. The minor infection (MI) patient may be required to get out of bed and walk sooner. The surgical patient might have to begin a normal diet sooner and perhaps leave the hospital with his or her sutures still in place.

The number of calendar days (LOS calendar) was most frequently used to determine length of stay. While these data are easy to obtain from existing documents, this method can lead to overestimation (Marik and Hedman, 2000). Marik and Hedman suggest estimating LOS from the number of midnight bed occupancy days or total hours.

Nurse staffing variables measured in the included studies were the ratio of RNs to patients, hours per patient day, staff hours, the ratio of RNs to other nursing staff, and RN qualification. These variables are commonly found in the research literature examining the impact of nurse staffing on patient mortality, patient falls, medical errors, or other patient outcomes (Aiken et al., 2002). Each of these variables provides different information about how nurses are assigned to care for their patients and work with each another. Using RN-to-patient ratios, hours per patient day and staff hours provides information about the appropriate utilization of nurse staffing in terms of the number of nursing staff per patient day. Skill mix (expressed in the ratio of RNs to other nursing staff) provides information about the appropriate utilization of nurse staffing in terms of workload shared between RNs and non-licensed nurses. RN qualification (expressed in the ratio of total number of RN years of experience to the total number of staff) indicates the efficiency of using RNs with more years of nursing experience. Theoretically, RNs with more years of nursing experience are expected to provide higher quality care and thus positively affect patient outcomes.

The effect of nurse staffing on LOS and cost of care was reported in each study. Results suggest that higher proportions of RNs optimize quality of care, and decrease LOS and cost of care. However, most studies examining the relationship between nurse staffing, cost, and LOS were limited to patients admitted to general wards. Few studies examined intensive care units (ICUs) and operating rooms, highly specialized areas of the hospital and widely different then general wards. Therefore, further research is needed to validate results and study a variety of clinical areas.

Intensive care units are highly specialized areas of the hospital where critically ill patients are monitored and often treated using complex equipment. Critical care nurses have the ability to promote the quality of clinical practice procedures and decrease ICU length of stay and costs of intensive care. However, studies examining the relationship between nurse staffing, costs of ICU care, and ICU length of stay are limited to patients admitted to surgical ICUs. Thus, further research is needed to validate results and to study patients admitted to a variety of specialty ICUs.

According to (Carmen 2009) ageing of the population will imply a greater demand of health services, as a consequence of the high rates associated morbidity. Among the diseases of the elderly population are chronic-degenerative conditions that make them fragile and turns them into a population that utilize many health services. However, it is well known that part of the hospital resources are used inadequately, either because the patients receive assistance that does not turn into health benefits or because care services could be provided at different institutional levels representing lower costs.

Evaluation of resource use in hospital health systems allows establishing the necessary actions which will correct the identified organizational problems. Unjustified hospital admissions and stays of elderly patients, do not only increase costs but are also related to poor health service and higher mortality resulting from several complications that come together, for instance, hospital infections, pressure ulcers or venous thrombosis, among others. Thus, the permanent evaluation of hospital service utilization is an essential topic that must be considered to improve resource assignment and to increase the quality of medical assistance in institutions that render service.

According to (Mark, 1983) it was reviewed that hospitals can be hazardous to one's health. Complications of hospital treatment are many, including no so comical infections, adverse drug reactions (which may also occur with outpatient treatment, but those that occur in association with inpatient intravenous drug use are more frequently very serious), complications related to procedures, and others. Clearly, one's probability of experiencing one of these adverse effects of hospital care increases directly with one's exposure; the greater the LOS, the greater the chance of infected with other diseases.

A number of studies provide evidence that mutual health organization (MHO) membership is associated with higher utilization of modern health care, in the form of outpatient visits or hospitalization. At least one study found the surprising result that MHO members were less likely than non-members to seek care when ill (Gumber, 2001).There is growing but still limited evidence on the effect of MHOs as a vehicle for reducing out-of-pocket health care expenditures, particularly the catastrophic expenditures associated with hospitalization or surgery. The available evidence indicates that MHO members tend to have lower out-of-pocket payments compared with non members and MHOs that cover inpatient care can reduce the percentage of Hospitalizations resulting in impoverishment. The literature reviews by (Jakab and Krishnan, 2001) and (Ekman, 2004) also conclude that there is consistent evidence that MHO membership is associated with lower out-of-pocket payments for health care.

During the post World War II era, healthcare delivery has experienced tremendous change. First, new technologies have changed the way healthcare is delivered by introducing over time more sophisticated techniques for diagnosing and treating illnesses (Weisbrod, 1991).

2.2 Management and motivation of hospital staff

An area where health care sector competition may prove more straightforward, it had input on the markets. For example, in many countries there is a competitive in labour market for hospital managers, who attract similar compensation packages from both for-profit and non-profit hospitals. Similarly, competition in the markets for physicians, nurses, and medical equipment and materials, and support services such as maintenance, catering, cleaning, and laundry can help to allocate resources in a way that rewards, and thus stimulates, improved performance. Input markets are generally less prone to 'market failure', since they often feature organized purchasers and suppliers with similar information and market power (Adam et al., 2006).

The prices paid by fee-paying patients are set by government, with the dominant concern being to make sure that basic services are affordable to the whole population. Prices have tended to be set below cost for simple and non-invasive care, and above cost for more complex care. The intention is that patients who need 'basic' care receive it, while patients who want the less basic and more expensive care pay enough for it to enable the provider to cross-subsidize the basic care from their profits on the more expensive care. Government subsidies have been based not on performance or throughput, but rather on staff numbers and the stock of beds. This encourages hospitals to expand their workforce and their bed stock, but not to improve the quality of their care or their efficiency. Health insurers—the final payer—vary in the way they pay providers in China. However, only some have moved beyond fee-for-service (FFS) (Adam et al., 2006).

Previous studies demonstrated that significant cost reductions were possible through effective staff management. Increasing the proportion of hours provided by RNs would yield cost savings resulting from avoiding death and reduced LOS (Mattke, 2006). Many studies encouraged reducing costs by increasing the caring activities of RNs (Amaravadi et al., 2000). Although increasing the number of RNs was positively related to increasing hospital expenditures, it did not affect hospital profit significantly. In contrast, greater non-registered nurse hours led not only to higher hospital expenditures but also lower profits (McCue et al., 2003).

Nurses play an important role in pain management. They assess pain and decide whether to administer medication (Manias, 2003). Pain assessment includes the fact that pain is identified, recognized as legitimate, quantified, documented and used to evaluate interventions. Documentation of assessments is the key to adequate management of pain (McGuire, 1992). Systemic opioid analgesics (SOA) are mostly prescribed as a variable dose and given by nurses on a PRN basis. Thus, the nurses make the decisions concerning medication for pain relief. Nurses are expected to, within the prescribed dosage range; use their professional judgement concerning the amount of analgesic administered to patients to avoid under-medication (Hunter, 2000). The rationale is to describe nurses´ approaches to pain management.

Nurse administrators are responsible for allocating nursing staff to meet quality of care standards and hospital budget requirements. On the surface it may appear that a staff mix with a higher level of RNs would be more costly. However, studies reported that the sufficient numbers of skilled RNs may actually reduce LOS and costs. A nurse staffing model with a lower number of RNs may ineffectively prevent adverse patient events resulting in patients having to stay longer than necessary. Furthermore, this may increase nursing turnover costs, and additional costs incurred from poor retention and use of overtime. The evidence suggests that replacing professional nurses with unlicensed assistive personnel is inappropriate to achieve cost-containment objectives. Rather than decreasing the number of RNs, hospitals should consider increasing the proportion of RNs as higher levels of knowledge and skills can reduce patient mortality (Aiken et al., 2002) as well as lower patient resource consumption.

Patient satisfaction survey is becoming the primary tool of accessing this aspect of health care. As stated in the Health Evidence Network report August 2003 by WHO, measurement is central to the concept of hospital quality improvement, it provides means to define what hospitals actually do, and to compare that with the original targets in order to identify opportunities for improvement. Likewise the WHO 2000 report further stated that, the organization, configuration and delivery of services impact on the performance of the overall health system (WHO, 2000).

Patients are the best and sometime the only source of information when it comes to evaluating health care services. The truth about patient satisfaction survey is that they can help identify ways of improving the practices to better service the patient in the future. Ultimately, that translates into better care and happier patients (White, 1999).

Costumer satisfaction is an important measure of quality service in health care organization (Gadallah et al., 2003). From a management perspective, patient satisfaction with health care is important for several reasons where management can identify sources of patient dissatisfaction and can organized to address system's weakness. On the other hand this benefit the patients as well as they are getting better quality service. Previous studies have shown that satisfied patient s are more likely to follow specific medical regimens and treatment plan for better outcome (Gadallah et al., 2003).

Motivation has always been an essential factor in managing personnel. It is a crucial variable in creating a high performance organization. Both private and public sector scholars are convinced of this relationship. Therefore, motivation has become one of the big questions in public administration, as it already has been for a long time in private sector management. However; the attention devoted to this big question of motivation brought forth little public administration research. Contrary to research of private sector management, public administration research has largely ignored motivation as a topic. It should therefore not come as a surprise that the bulk of our knowledge concerning motivation in the public sector is in fact theories that originate from private sector management research (Perry, 1996).

(Halepota, 2005) defines motivation as "a person's active participation and commitment to achieve the prescribed results''. Halepota further presents that the concept of motivation is abstract because different strategies produce different results at different times and there is no single strategy that can produce guaranteed favourable results all the times."

According to (Antomioni, 1999) "the amount of effort people are willing to put in their work depends on the degree to which they feel their motivational needs will be satisfied. On the other hand, individuals become de-motivated if they feel something in the organisation prevents them from attaining good outcomes. It can be observed from the above definitions that, motivation in general, is more or less basically concern with factors or events that moves, leads, and drives certain human action or inaction over a given period of time given the prevailing conditions. Further more the definitions suggest that there need to be an ''invisible force'' to push people to do something in return. It could also be deduced from the definition that having a motivated work force or Creating an environment in which high levels of motivation are maintained remains a challenge for today's management. . This challenge may emanate from the simple fact that motivation is not a fixed trait –as it could change with changes in personal, psychological, financial or social factors.

(Maslow, 1943) suggests that human needs can be classified into five categories and that these categories can be arranged in a hierarchy of importance. These include physiological, security, belongings, esteem and self-actualisation needs. According to him a person is motivated first and foremost to satisfy physiological needs. As long as the employees remain unsatisfied, they turn to be motivated only to fulfil them. When physiological needs are satisfied they cease to act as primary motivational factors and the individual moves "up" the hierarchy and seek to satisfy security needs. This process continues until finally self actualisation needs are satisfied. According to Maslow the rationale is quite simple because employees who are too hungry or too ill to work will hardly be able to make much a contribution to productivity hence difficulties in meeting organisational gaols.

2.3 Hospital policies

Policy in any organization is tool that distinguished one organization from others and health service is not exceptional. Every hospital has its own rules and regulations that govern its activities.

In sequential policy changes examined in the cost to government through subsidies for private health insurance premiums were introduced before the policy with no cost to government in and of itself (lifetime community rating). Yet those policies with a cost to government appear to have had either no impact on private health insurance coverage (PHIS) or a modest impact (the 30% rebate), while the other policy appears to have induced a major response at virtually no cost to government. Ironically, a government-funded reduction in premiums appears to have had a much more muted effect on private health insurance uptake than an unfunded announcement of an increase in premiums (Karnon et al., 2009).

A consequence of the sequencing of policy changes discussed in the government expenditures on health insurance subsidies have increased substantially. This is not because the subsidies actually induced a major uptake of private health insurance but because lifetime community rating induced a major uptake and those insurance policies then qualified for a subsidy. In other words, the large increases in expenditures on subsidies were more likely an effect rather than a cause of increased demand for private health insurance in other countries.

The decision to discharge a patient from the hospital by the physician is a selective process that distinguishes, on a daily basis, between patients who are able to continue the recovery process outside the hospital and those who are not yet ready to leave. This selective process, however, is potentially influenced by a variety of factors, not exclusively clinical considerations. For example, changes in there imbursement systems for hospitalizations (giving incentives for reducing the length of stay [LOS]) may influence the discharge policy. With cost considerations gaining increasing importance, the weight of administrative factors in the decision making process grows. We were especially interested in the interface between providers such as hospital clinicians versus administrators (Clarke and Rosen, 2001).

For the role of discharge in relation to patient outcomes to be fully appreciated, a long-term follow-up, beyond hospital stay, of patients is required. In fact, discharge can be viewed as an informative intermediate event in the long-term process of care and recovery of hospitalized patients. On the one hand, a variety of clinical and institutional factors influence the decision to discharge a particular patient on any given day. On the other hand, discharge implies changing environments reflecting changes in the level of medical supervision and, thus, may modify the risks of adverse outcomes such as mortality, complications, and rehospitalisation's.

To describe the distribution of discharge days, the modal day of discharge, rather than the mean or median of the distribution, was chosen. By definition, the mode is the day with the highest frequency of discharges. Thus, the mode captures the most common patterns of discharge and hence is likely to reflect more strongly a hospital policy, while the mean is strongly influenced by extreme values relevant to few patients with long duration of stay.

According to (Jonny et al., 2008), community health insurance (CHI) schemes are growing in importance in low income settings, where health systems based on user fees have resulted in significant barriers to care for the poorest members of communities. They increase revenue, access and financial protection, but concerns have been expressed about the equity of such schemes and their ability to reach the poorest. Few programmes routinely evaluate equity impacts, even though this is usually a key objective. This lack of evidence is related to the difficulties in collecting reliable data on utilization and socio-economic status.

Despite numerous efforts to establish functioning health care systems, most people in developing countries must still rely on direct payments to finance their health care needs. In some regions, these out-of-pocket payments can account for up to 80 percent of total health expenditure.

Private prepaid programs, such as community-based health insurance schemes, are often the only possible way for poor people to participate in risk-pooling programs. Evidence so far suggests that private schemes can improve access to health care and offer financial protection even to marginalized groups (Jütting, 2005). Despite the growing importance of private health insurance (PHI), however, surprisingly little is known about its role in national health systems in the developing world (Sekhri and Savedoff, 2005).

Sustainable instruments for health financing are urgently needed to reduce the high amount of out-of-pocket payments and the incidence of catastrophic health shocks in the developing world (World Health Organization [WHO], 2006). Several factors have recently stimulated the development of private insurance mechanisms as a means to finance healthcare in low- and middle-income countries. These factors include difficulties with traditional ways of health care financing, diversified consumer demand in the course of economic development, and intensified trade in the health-services sector, which has introduced foreign insurance providers to developing countries.

2.4 Regression Analysis and Length of Stay

A mathematical model can help to predict the length of stay in hospital while KNUT Hospital exceptional.

(Tu, 1993) compared a neural network and logistic regression models for predicting length of stay in the intensive care unit (ICU) following cardiac surgery. This is important from an actuarial perspective because improving the efficiency of hospitals has a direct impact on insurance costs. Structurally, the training set and test set consisted of approximately 700 patients each, and five unique risk strata were created using each model. The advantages and disadvantages of each modelling technique was explored, with the conclusion was that either model could potentially be used as a risk stratification tool for counselling patients and scheduling operations.

(Liu, 2000) combines the Self-Organizing Map (SOM) and Back-Propagation Neural Network (BPN) and used the Neural Network to inquire risk-adjustment problem of policy in the capitation of NHI. (Shih-Fang, 2002) use BPN and statistical models to create an individual risk adjustment model, to explore two different analysis tools of the appropriateness and feasibility, on hopes of keeping medical expenditure controlled within the acceptable range of the insurer.

(Chi-Hung, 2002) use opposite direction transmission Neural Networks and multiple regression analysis to the construct the prediction model of medical expenditures, BPN was more accurate compared to multiple regression analysis.

(Jian-Shen et al., 2006) compares the logistic and Neural Network model to fine out when the cost basis for assessment is misjudged; the Neural Network model has better classification ability. (Shuo-Fen et al., 2006) combine artificial intelligence methods of Self Organization - Map SOM and BPN to predict personal medical expenditures allocation based on health insurance budget. Its findings result that in conjunction with SOM and BPN is more predictive capability improved than comparing to only using BPN or Linear regression models.

(Khashei and Bijari, 2010) research Autoregressive Integrated Moving Average (ARIMA) and Back-Propagation Neural Network (BPN) to predict the Wolf's sunspot, the Canadian Lynx data and the British Pound/United States dollar exchange rate data and they get the result of ARIMA is worse than from BPN model. Therefore, we propose to adopt the quick calculation and fault tolerance of BPN, to find the optimum forecasting scenario for medical expenditures, then comparing analysis with ARIMA, to construct a compliable a forecasting model for health insurance medical expenditures.

(Lin et al., 2006) used ARIMA, Vector Autoregressive Moving Average (VARMA), and Holt-Winters exponential-smoothing models to forecast the monthly discharges and differences in occupancy, across several hospitals. He found that in most cases, ARIMA performed as well as or better than either VARMA or the Holt-Winters method.

(Jones et al., 2002) described several ARIMA models of daily hospital occupancy resulting from emergency admissions. They found that whereas external covariates such as weather were significantly correlated with the number of admissions, they added little to the model's forecasting ability.

(Channouf et al., 2007) applied ARIMA models to forecasting calls to emergency medical services. They report that for forecast horizons of up to about two weeks, the best performing model was a mixed-effects regression model with an autoregressive model of the residuals. Beyond that time, no model had any forecasting benefit.

In the United States, mathematical models are familiar, everyday tools in engineering, business and military applications in most sciences. They represent hypotheses about underlying mechanism that generate observed phenomena or options for action and potential consequence (McKenzie, 2004).

(Contreras et al., 2003) developed a model for predicting the next-day electricity prices in mainland Spain and California markets using an ARIMA model. Their developed model was able to forecast the 24 market clearing prices of tomorrow. The ARIMA model is an effective tool for forecasting time series.

A good model is to be developed for forecasting the length of stay in the hospital. A model fitting quality is defined as the sum of the residuals' squares divided by the sample size. Its objective is to measure the model's capacity to produce the sample data (i.e. to verify how similar the modelled series and the actual series really are) (Guerrero, 2003).

(Zhang, 2003) pointed out that no single method is best in every situation and that combining different models is an effective and efficient way to improve forecasting accuracy, giving examples of previous work in hybrid methods that use neural networks. His paper proposed a hybrid feed forward neural network (FNN)-ARIMA methodology where ARIMA is used to model the linear component and FNNs modelled the forecasting errors. The hybrid method outperforms the two component methodologies when they are used separately.

A recent initiative of several large companies in the food industry, which aimed to improve forecasting practice, identified that 48% of food companies are poor at forecasting (Adebanjo & Mann, 2000).

The methodologies that have been used in sales forecasting are typically time series algorithms that can be classified as linear or nonlinear, depending on the nature of the model they are based on. Linear models, like autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) (Box, Jenkins, and Reinsel, 1994) are the most popular methodologies, but their forecasting ability is limited by their assumption of a linear behavior and thus, it is not always satisfactory (Zhang, 2003).

Combined statistical methodologies and artificial intelligence technologies, like fuzzy logic have also been applied for selecting the appropriate input variables in forecasting problems (Mastorocostas, Theocharis, and Petridis, 2001).

CHAPTER THREE

METHODOLOGY

Introduction

This section focuses on the procedures or steps that would be carried out in order to gather and analyse data relevant for the study. The type of data used for this research was secondary data. The secondary data was collected from Admissions and Discharge Books of KNUST Hospital. The collected data consist of all patients who had been on admission from 2005 to 2009 at male medical ward. Analysis of the data was done by the statistical software S PLUS vision 6, SPSS vision 16 and Microsoft Excel. It is case study of length of stay of patient on male medical ward in the hospital. Statistical software used were S PLUS vision 6, SPSS vision 16 and Microsoft Excel.

3.1 Data Collection Technique

The administrative section of the hospital had admission and discharge books in various wards which had records on patients date of admission (month), type of illness, the sex of the patient, age, and date of discharge minus date of admission (length of stay) and outcome of the treatment were also recorded (survive or not survive). All these variables were coded into Excel spread sheets, variables like age, types of illness, month and etc were given numeric codes. Months were coded (1-12) for January to December, survive and non survive were coded (0,1) respectively and various illness were also given different codes. SPSS was used for descriptive analysis and S PLUS was also used for analysis of generalized linear models. Thirteen diseases were considered for the study and among them were abdominal pain (abd), alcoholic (alco), asthma (asth), cholera (chol), dematus manatus (dm), hernia (hern), malaria (mal), pelvic inflammatory disease (pid), pneumonia (pneu), respiratory infection (rat), sickle cell disease (scd), thyroid (thy).

3.2 Sample size

There was 2206 observation collected from 2005 – 2009 at male medical ward at KNUST – Hospital for the analysis.

3.2.1Assumptions

1. We assumed that each patient reports to the hospital with only one particular illness and not more.
2. Treatment is the same for a particular disease. Patients suffering from the same disease are given the same treatment.
3. We assumed that patient admitted with particular disease would be discharge without infected with other diseases.

3.3 Component of a Regressions Model

The variable of interest, (y) called response or dependent variable. It represents the effect or response resulting from combination of factors. The factors on which response variable depends are called predictors or independents variables. They are represented by the other variables, x, w, z etc. An error term denoted by ε, which cater for the error due to chance and neglected factors which we may assume not important.

The regression model, thus take the form:

$$y = f(x, w, z) + \varepsilon \tag{3.1}$$

Where $f(x, w, z) = E(y / x, w, z)$, the mean response due to specified predictor values x, w, z.

ε the error term which represents a random deviation of a typical response value from the mean response.

3.4 Types of Regression Models

Regression models are classified according to the number of predictor variables and also the form of the regression function.

3.4.1 Simple Regression Model

This is model which has a single regressor x and the relationship between the response y and x a straight line. The model is given by $y = B_0 + B_1 x + \varepsilon$ where the intercept B_0 and the slope B_1 are unknown constant and ε is random error component.

3.4.2 Multiple Regression Models

It involves more than one predictor variable. It is given by $y = B_0 + B_1 x_1 + B_2 x_2 + \varepsilon$

The parameter β_1 indicates the expected change in response (y) per unit change in x_1 when x_2 is held constant, β_2 measures the expected change in y per unit change in x_2 and x_1 is held constant. The response y may be related to k regressor variables. It is given by

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \tag{3.2}$$

The parameter β_j, $j=0,1,...k$ are called the regression coefficients, β_j represent the expected change in the response y per unit change in x_j when all the remaining regressor variables $x_i (i \neq j)$ are held constant.

3.4.3 Non-Linear Regression Models

These models contain one or more predict on variables and or with a second or more order degree. Example of non-linear regression model:

$y = a + bx + cx^2 + \varepsilon$ **(Quadratic Polynomial)**

3.5 Definition and Features of Simple Linear Regression Model

The simple linear regression model is given by

$$y_i = B_0 + B_1 x_i + \varepsilon \qquad (3.3)$$

Where

y_i is the value of the response variable in the *ith* observation (random and continuous)

x_i is the known value of the predictor variable in the *ith* observation

ε is the random error term which cater for the errors due to chance and neglected factors which are assumed not important.

β_0, β_1 are the parameters of the model. β_0 represents the parameter for the intercept on the $y-$axis. β_1 represents the parameter that measures the slop of the linear model.

The model can be broken into two components, namely the constant term, $E(y/x) = \beta_0 + \beta_1 x_i$ and the error term, ε_i.

3.5.1 Assumptions of the model

- The value of y_i are random and independent of each other.
- The mean of the error term $E(\varepsilon_i) = 0$
- The variance of the error term is constant, δ^2 that is $\text{var}(\varepsilon_i) = \delta^2$ for all x_i.
- The probability distribution of $\varepsilon_i \sim N(0, \delta^2)$
- The random errors ε_i and ε_j are independent, that is $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$
- From the above assumptions, the following are established;

$$E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i$$

$$\text{var}(y_i) = \text{var}(B_0 + B_1 x_i + \varepsilon_i)$$

$$= \text{var}(\varepsilon_i) = \delta^2, \forall x_i$$

$$y_i \sim N(B_0 + B_1 x_i, \delta^2)$$

$$\text{cov}(y_i, y_j) = 0$$

3.5.2 Alternative model for Simple Linear Regression

The simple linear regression model given by $\boldsymbol{y = \beta_0 + \beta_1 x + \varepsilon}$ can be express in the form;

$$\cancel{y_i = \beta_0} + \beta_1(x - \overline{x}) \tag{3.4}$$

Where $\boldsymbol{\beta_0} = \beta_0 + \beta_1$x **and** $\boldsymbol{x = \frac{1}{2}\sum_{i=1}^{n} x}$

3.6 Correlation coefficient

This is a quantitative measure of strength of linear relationship between two variables say x and y. There are two types of measures. Pearson product-moment measure and Spearman coefficient.

3.6.1 Pearson product-moment measure Correlation Coefficient.

This is used for quantitative data measured on interval or ratio scale. It is defined by

$$r = \frac{\operatorname{cov}(x, y)}{\delta_X \delta_Y} = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} \tag{3.5}$$

i. y_i is the response variable we wish to estimate or predict $x_{i1}, x_{i2}, \ldots x_{ik}$

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})(y_1 - \bar{y}) \tag{3.6}$$

$$= \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)$$

$$SS_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_1\right)^2$$

$$SS_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i}^{n} y_i^2 - \tfrac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

3.7 Multiple Linear Regression Model

The general linear model for a multiple regression analysis takes the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i \tag{3.7}$$

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_i x_j + \varepsilon_{i,} i = 1, 2, \ldots n$$

Where:

i. y_i is the response variable we wish to estimate or predict $x_{i1}, x_{i2}, \ldots x_{ik}$

ii. $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the predictor variables that are measured without error.

iii. $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are parameter to be determined.

iv. ε_i is random error for any given set values of $x_{i1}, x_{i2}, \ldots, x_{ik}$

3.7.1 Assumptions

i. $E(\varepsilon) = 0$ for $i = 1, 2, \ldots, n$ that is a given set values $x_{i1}, x_{i2}, \ldots, x_{ik}$

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$

ii. $Var(\varepsilon_i) = \delta^2$, which is constant for all value of x. That is $\mathrm{var}(y_i) = \delta^2$, a constant for each

recorded value of y_i

iii. The random error ε_i is independent, that is their covariance is equal to zero.

$$\mathrm{cov}(\varepsilon_i, \varepsilon_k) = 0 = E(\varepsilon_i, \varepsilon_k) = 0, i \neq k \tag{3.8}$$

iv. The random error $\varepsilon_i \sim N(0, \delta^2)$, that is

$$y \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k, \delta^2).$$

We adopted the following techniques in analysing the various types of data collected.

3.7.2 Assumed Regression Model

$$\boldsymbol{Los = \beta_0 + \beta_1 age + \beta_2 survive + \beta_3 time + \beta_4 abd + \beta_5 asth + \beta_6 cho + \beta_7 dm + \beta_8 hern +}$$

$$\boldsymbol{\beta_9 hpt + \beta_{10} mal + \beta_{11} pid + \beta_{12} pneu + \beta_{13} rat + \beta_{14} scd + \beta_{15} thy + \varepsilon} \tag{3.9}$$

3.8 Estimation of Parameters

In estimating parameter of the multiple regression models, we take into consideration the least square method.

3.8.1 Least Squares Estimation

The regression estimate, are obtained by the least square method as follows:

The least square function or SSE,

$$s(\beta_0,\beta_1,\ldots,\beta_k) = \sum_{i=1}^{n}(y_1 - \hat{y}_i)^2 = \sum_{i=1}^{n}\varepsilon_1^{\,2} = \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij})^2 \tag{3.10}$$

To find the regression estimate, we minimize the sum of the squares of the deviations of the points from any proposed plane with respect to $\beta_0,\beta_1,\ldots,\beta_k$. The regression estimates must satisfy the equation:

$$\tfrac{ds}{d\beta_0} / \hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k - 2\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij}) = 0$$

$$\tfrac{ds}{d\beta_j} / \hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k = -2\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{k}\beta_j x_{ij})x_{ij} = 0$$

Simplifying, we obtained the least squares (k+1) normal equations.

$$n\hat{\beta}_0 + \hat{\beta}_1\sum_{i=1}^{n}x_{i1} + \hat{\beta}_2\sum_{i=1}^{n}x_{i2} +, ..., \beta_k\sum_{i=1}^{n}x_{ik} = \sum_{i=1}^{n}y_i$$

$$\hat{\beta}_0\sum_{i=1}^{n}x_{i1} + \hat{\beta}_1\sum_{i=1}^{n}x_{i1}^2 + \hat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{i2} + ... + \beta_k\sum_{i=1}^{n}x_{ik} = \sum_{i=1}^{n}x_{i1}y_i$$

$$\vdots$$

$$\hat{\beta}_0\sum_{i=1}^{n}x_{ik} + \hat{\beta}_1\sum_{i=1}^{n}x_{ik}^2 + \hat{\beta}_2\sum_{i=1}^{n}x_{i1}x_{ik} + ... + \beta_k\sum_{i=1}^{n}x_{ik}x_{ik} = \sum_{i=1}^{n}x_{ik}y_i$$

With the solution being the regression estimates of;

$$\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$$

In matrix form, the least squares function,

$$s(\beta) = \sum_{i=1}^{n}\varepsilon_i = \varepsilon^T\varepsilon = (y - x\beta)^T(y - x\beta) = y^T y - 2\beta^T x^T y + \beta^T x^T x^T \beta \tag{3.11}$$

Since $\boldsymbol{\beta^T X^T y}$ is scalar and its traspose $(\boldsymbol{\beta^T X^T y}) = \boldsymbol{y^T X \beta}$ is also scalar. The least squares estimates must satisfy

$$\tfrac{\partial s}{\partial \beta} / {}^{n}_{\beta} = -2x^T y + 2x^T x\hat{\beta} = 0$$

This simplifies to:

$$x^T x\hat{\beta} = x^T y$$

As equivalent normal equation in the matrix form of the equation is given by;

$$
\begin{pmatrix}
n & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2}\cdots & \sum_{i=1}^{n} x_{ik} \\
\sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2}\cdots & \sum_{i=1}^{n} x_{i1}x_{ik} \\
. & . & .... & . \\
. & . & .... & . \\
\sum_{i=1}^{n} x_{ik} & \sum_{i=1}^{n} x_{ik}x_{i1} & \sum_{i=1}^{n} x_{ik}x_{i2}\cdots & \sum_{i=1}^{n} x_{ik}^2
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_0 \\ \hat{\beta}_1 \\ . \\ . \\ \hat{\beta}_k
\end{pmatrix}
=
\begin{pmatrix}
\sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{i1}y_i \\ . \\ . \\ \sum_{i=1}^{n} x_{i1}y_i
\end{pmatrix}
$$

$$x^T x \qquad \hat{\beta} = \qquad x^T y$$

From this we obtained least squares estimators,

$$\hat{\beta} = (x^T x)^{-1} x^T y$$

The fitted multiple regression model corresponding to the variables

$$x^T = (1, x_1, \dots, x_k) \text{ is}$$

$$\hat{y} = x^T \hat{\beta}$$

$$= \hat{\beta}_0 + \sum_{j=1}^{k} \beta_j x_j$$

The vector of fitted values y_i corresponding to the observed values y_i is

$$\hat{y} = x\hat{\beta}$$

$$= x(x^T x)^{-1} x^T y$$

$= \boldsymbol{Hy}$ where $\boldsymbol{H} = \boldsymbol{x(x^T x)^{-1} x^T}$ is a $(\boldsymbol{n \times n})$ matrix.

The vector of residuals

$$\hat{\varepsilon} = y - \hat{y}$$

$$= y - x\hat{\beta}$$

$$= y - x(x^T x)^{-1} x^T y$$

$$= \left[I - x(x^T x)^{-1} x^T y \right] y$$

$$= [I - H]\, y_i$$

Which satisfy the conditions:

$$x^T \hat{\varepsilon} = 0 \text{ and}$$

$$\hat{y}^T \hat{\varepsilon} = 0$$

The residual or error sum of squares

$$SSE = s(\beta) = \hat{\varepsilon}^T \hat{\varepsilon}$$

$$= (y - x\hat{\beta})(y - x\hat{\beta})$$

$$= y^T y - \beta^T x^T y - y^T x\hat{\beta} + \hat{\beta}^T + \hat{\beta}^T x^T x\hat{\beta} \tag{3.12}$$

$$= y^T y - 2\beta^T x^T y + \hat{\beta}^T x^T x\hat{\beta}$$

$$= y^T y - \hat{\beta}^T x^T y$$

3.8.2 Hypothesis test for $\beta_1, \beta_2, \ldots, \beta_k$

a) Hypothesis

$$H_0 : \beta_i = 0$$

$$H_0 : \beta_i \neq 0$$

b) Test statistics

$$T(\hat{\beta}_i) = \frac{\hat{\beta}_i}{se(\beta_i)} \text{ where } se(\hat{\beta}_i) \text{ is standard error of } \hat{\beta}_i$$

c) Decision rule

Reject H_0 if $\left|T(\hat{\beta}_1)\right| \geq t_{\alpha/2}(n-k)$ or $T(\hat{\beta}_1) < -t_{\alpha/2}(n-k)$ or $T(\hat{\beta}_1) < t_{\alpha/2}(n-k)$

Where k is the number of predictor variables.

3.9 Coefficient of Determination

From the total variation:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$1=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

Let ε_i represent $(y_i-\hat{y}_i)^2$

$$1=\frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+\frac{\hat{\beta}_i^2\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$1=\frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+\left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}\right]$$

$$1=\frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}+r^2$$

$$r^2=1-\frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$r^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2-\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$r^2 = \frac{SS_{yy} - SSE}{SS_{yy}} = \frac{SSR}{SS_{yy}} \qquad (3.13)$$

r^2, here is called the coefficient of the determination which is the explained variation expressed as a fraction of the total variation. It is also defined as the square of the correlation coefficient. Thus r^2 determine the percentage of the total variation in x_i.

3.10 Multiple Coefficient of Determination

The quality, R^2 is called the multiple regression model, y_i is evaluated by R^2. It determines the proportion of the total variation in y_i that is explained or attributable to the predictor variables, $x_{i1}, x_{i2}, \ldots, x_{ik}$.

If $R^2 = 1$, the fitted model passes through all the data point and also

$\hat{\varepsilon}_i = 0$ **for all** $i = 1,2,3,\ldots,n$. If $R^2 = 0$, $\hat{\beta}_0 = \bar{y}$ **and** $\beta_1 = \beta_2 = \ldots \hat{\beta}_k = 0$

In this case the predictor variables $x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ik}$ have no imfluence on the response **variable** $\hat{y}_i$. If more predictor variables are introduced in the model R tend increase it value. This is because adding more predictor variables to the model causes the prediction error to be small and thus reducing SSE, that is from SST = SSV + SSE, when SSE.Get smaller, SSR, must get larger, causing $\mathbf{R}^2 = 1 - \frac{SSE}{SST}$ to increase.

Since R^2 is often by including larger number of predictor variables, it is sometime suggested to modified $\mathbf{R}^2$ to account for the number of independent variables in the model, The adjusted multiple coefficient of determination, denoted, R^2 is defined by

$$R_a^2 = 1 - \left(\frac{n-1}{n-\rho} \right) \frac{SSE}{SST} \tag{3.14}$$

$$= 1 \frac{(1-R^2)(n-1)}{n-k-1}$$

3.11 Confidence interval in multiple regressions

In confidence interval in multiple regression model we normally consider; Confidence intervals on the regression coefficient, confidence interval estimation of the mean response and prediction for new observation.

3.11.1 Confidence Intervals on the Regression Coefficients

To construct confidence interval estimates for the regression coefficient s β_j we must assume that the error ε_i are normally and independently distributed with the mean zero and variance σ^2 1. Therefore the observations y_i are normally and independently distributed with the mean $\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$ and variance σ^2. Since least squares estimator $\hat{\beta}$ is a linear combination of the observations, it fellows that $\hat{\beta}$ is normally distributed with mean vector β and covariance matrix $\sigma^2 (x'x)^{-1}$. This implies that the marginal distribution of any regression coefficient $\hat{\beta}_j$ is normal with the mean β_j and variance $\sigma^2 c_{jj}$ where c_{jj} is the jth diagonal element of the $(x'x)^{-1}$ matrix. It statistics is given by;

$\dfrac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 c_{jj}}}, j = 0,1,\ldots,k$ is distributed as t with $n-p$ degree of freedom, where σ^2 is the estimate of the error .

The $100(1-\alpha)$ percent confidence interval for regression coefficient $\beta_j, j = 0,2,\ldots,k$ is given by;

$$\hat{\beta}_j - t_{\alpha/2,n-p}\sqrt{\sigma^2 c_{jj}} \le \beta \le \hat{\beta}_j + t_{\alpha/2,n-p}\sqrt{\sigma^2 c_{jj}}$$

We usually call the quantity $se(\hat{\beta}_j) = \sqrt{\sigma^2 c_{jj}}$, the standard error of the regression coefficient is $\hat{\beta}_j$.

3.12Transformation (Box-Cox transformation)

The Box – Cox transformation is one particular way of parameterising a power transformation that has advantageous properties. Power transformation is define as continuously varying function with respect to the power λ, in piece-wise function from that, make it continuous at the point of singularity $(\lambda = 0)$.

For data vector $(y_i \ldots y_n)$ in which each $y_i > 0$, the power transformation is

$$y_i^\lambda = \begin{cases} \dfrac{y_i^\lambda - 1}{\lambda (GM(y))^{\lambda-1}}, & if\ \lambda = 0 \\ GM(y)\log y_i & if\ \lambda = 0 \end{cases} \qquad (3.15)$$

Where $GM(y) = (y_1, ..., y_n)^{\frac{1}{n}}$ is the geometric mean of the observations $y_1, .. y_n$

$(\lambda - 1)^{th}$ power of the geometric mean in the denominator, implies that the unit of measurement do not change as λ changes.

Box-Cox family of transformations correcting for;

Skewness

Non-normality

Unequal variance

Nonlinearity

The procedure for choosing a transform from the power family;

$$Y' = Y^{\lambda}$$

It means

$$Y' = \begin{cases} Y^{\lambda} & if\ \lambda \neq 0 \\ In\ Y & if\ \lambda = 0 (by\ def \end{cases} \tag{3.16}$$

$\hat{\beta}_0, \hat{\beta}_1, \beta, \hat{\sigma}^2$ and $\hat{\lambda}$ by maximizing the

Likelihood function

$$\prod_{i=1}^{n} f_{Y_i^{\lambda}}(Y_i^{\lambda}, x_i, \beta_0 . \beta_1, \sigma^2)$$

$$= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i^{\lambda} - \beta_0 - \beta_1 x_i)^2 \right]$$

One can show that the ML estimate $\hat{\lambda}$ equals

$$\hat{\lambda} ML = \arg\min \quad SSE(W / X)$$

Where

$$W = \begin{cases} K_1(Y^{\lambda} - 1) & if\ \lambda \neq 0 \\ K_2 InY & if\ \lambda = 0 (by \quad def) \end{cases} \tag{3.17}$$

For specific constants K_1 and K_2.

3.13 Generalized Linear Models

Generalized Linear Models (GLM) are regressions for data that are not reasonably assumed to be normally distributed. Such models are also sometimes discussed as "probability models" or model of specific types like Logistic regression, Poisson regression, or other models for count data. Though GLM can be flexible enough to be applied to a wide range of real-life situations, they retain most of the power of normal linear regression models.

To construct a generalized linear model we decide on response and explanatory variables for the data and choose an appropriate link function and response probability distribution.

An important assumption is that the *yi* are normal and independent with standard deviation σ.

Linearity assumption is not enough, as the dependent variable might not be normally distributed (typically bounded or discrete).

Examples of GLM include traditional linear model, Logistic Regression, Poisson regression in log linear model and Gamma model with log link.

3.13.1 Components of Generalized Linear Model

The basic components of GLM are stochastic component and systematic component;

Stochastic component:

Y is the random or stochastic components which remain distributed according to a specific exponential family distribution with mean μ. This component is sometime called the 'error structure', or 'response distribution.

Systematic component:

$\theta = X\beta$ is the systematic component producing the linear predictor. So explanatory variables, X, affect the observed outcome variable, Y, only through the functional form of the $g(\cdot)$ function.

3.13.2 The Model

Let $y_i,\dots,y_n$ denote n independent observations on a response. y_i is treat as a realization of a random variable Y_i. In the general linear model we assume that Y_i has a normal distribution with mean μ_i and variance σ^2

$$Y_i \sim N(\mu_i, \sigma^2),$$

and we further assume that the expected value μ_i is a linear function of p predictors that take values $x_i' = (x_{i1},\dots,x_{ip})$ for the $i-th$ case, so that

$$\mu_i = x_i'\beta, \tag{3.18}$$

where β is a vector of unknown parameters.

3.13.3 The Exponential Family

It is assume that the observations come from a distribution in the exponential family with probability density function

$$f(y_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\alpha_i(\phi)} + c(y_i,\phi)\right\} \tag{3.19}$$

Here θ_i and ϕ are parameters and $\alpha_i(\phi), b(\theta_i)$ and $c(y_i, \phi)$ are known functions. In all models considered in this, the function $\alpha_i(\phi)$ has the form

$$\alpha_2(\phi) = \phi / pi,$$

where p_i is a known prior weight.

The parameters θ_i and ϕ are essentially location and scale parameters.

It can be shown that if Y_i has a distribution in the exponential family then it has mean and variance

$$E(Y_i) = \mu_i = b'(\theta_i)$$
$$\operatorname{var}(Y_i) = \sigma_i^2 = b''(\theta_i)\alpha_i(\phi),$$

where $b'(\theta_i)$ and $b''(\theta_i)$ are the first and second derivatives of $b(\theta_i)$.

When $\alpha_i(\phi_i) = \phi / pi$ the variance has the simpler form.

$$\operatorname{var}(Y_i) = \sigma_i^2 = \phi b''(\theta_i) / p_i \tag{3.20}$$

The exponential family includes special cases of normal, binomial, Poisson, exponential, gamma and inverse Gaussian distributions.

3.13.4 The Link Function

The second element of the generalization is that instead of modelling the mean, as before, we will introduce a one-to-one continuous differentiable transformation $g(\mu_i)$ and focus on

$$\eta_i = g(\mu_i) \tag{3.21}$$

The function $g(\mu_i)$ will be called the link function. It is assume that the transformed mean follows a linear model, so that

$$\eta_i = x_i'\beta.$$

The quantity η_i is called the linear predictor. Since the link function is one-to-one we can invert it to obtain

$$\mu_i = g^{-1}(x_i'\beta)$$

3.13.5. Gamma model with log link

The response variable y_i is a positive value and continuous variable with distribution of gamma. The Link function is given by:

$$g(\mu) = \log(\mu)$$

3.13.6 Normal model with identity link

A regression model for gamma distributed data is a generalized linear model and the model is fitted by maximum likelihood. The model uses the same systematic part as the linear normal model but the random part declares the measurements as gamma-distributed.

Systematic: $E(H_i) = \mu_i = \alpha + \beta D_i + \gamma D_i^2$

Random: $H \sim Gamma(s_i, a)$ with $s_i = \mu_i / a$

The observations are gamma distributed with different means and variances but a common shape a. The residual plot shows that the model is successful because their absolute value is not longer related to the predicted values. This indicates that the gamma model is able to model the variance-mean relationship of the data. The square-roots of the absolute residuals are plotted because they are more symmetrically distributed than the absolute residuals and it is easier to detect a trend.

3.13.8 Gamma with inverse link

The identity link and the log-link to model the data with a gamma distribution. These two links yielded both a linear model in the density. Another important link is the inverse link $1/E(y)$. This link is called the canonical link for gamma generalized linear models.

The inverse of the link function is sometimes called the mean function:

$$g^{-1}(\eta_i) = \mu_i$$

An inverse link function does not have to map $X\beta$ back to the whole range of the mean for a distribution.

3.14. Akaike Information Criterion (AIC)

Akaike's information criterion (Akaike, 1973) was introduced as an approximately unbiased estimator of the expected Kullback-Leibler information of the fitted model. Let $D = \{(y_i, x_{ij})\}$ be the data at hand, where y_i is the response vector and x_{ij} is a set of covariates. Also let M and M^* be a candidate and the true model, respectively. Let $L(\beta;D)$ and $L(\beta^*;D)$ be the log-likelihood functions corresponding to the models M and M^*, respectively, where β and β^* are the corresponding regression parameters. The Kullback-Leibler information, also known as cross entropy, between the models M and M^* is

$$\Delta(\beta,\beta^*) = E_{M^*}(-2L(\beta;D)), \tag{3.22}$$

where the expectation E_{M^*} is taken under the true model M^*. For a given set of competing models, we choose that model as the best model for which the Kullback-Leibler information $\Delta(\beta,\beta^*)$ is the smallest. In practice, both β and β^* are unknown, as an asymptotically unbiased estimator of $E_{M^*}(\Delta(\hat{\beta},\beta^*))$ which is actually the AIC can be used as a model selection criterion, where $\hat{\beta}$ is the maximum likelihood estimator of β under any competing model. Notationally, the AIC can be written as

$$AIC = -2L(\hat{\beta};D) + 2p, \tag{3.23}$$

where p is the order of the vector β. A model which minimizes the AIC is considered to be the "best" model. This definition implies that when there are several models whose values of maximum likelihood are about the same level, we should choose the one with the smallest number of free parameters.

CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF RESULTS

Introduction

This chapter deals with the analysis and the discussion or interpretation of the data from male medical ward of KNUST Hospital.

4.1DESCRIPTIVE ANALYSIS OF THE DATA

In this section, summary statistics including graphs, figures and tables are presented to enable us to identify significant features of the data.

The monthly admission cases from January 2005 to December 2009 were fed into a SPSS, version 16 spread sheets and analysed to obtain the descriptive statistics for medical ward (male).

The data set has an average length of stay (LOS) of 2.71 days (approximately 3 days) and standard deviation of 2.68 days. The least number of patients who were on admission during the period were 6 patients in April, 2007 and the highest numbers of patients were recorded were 79 in July, 2008 (details are shown in appendix B).

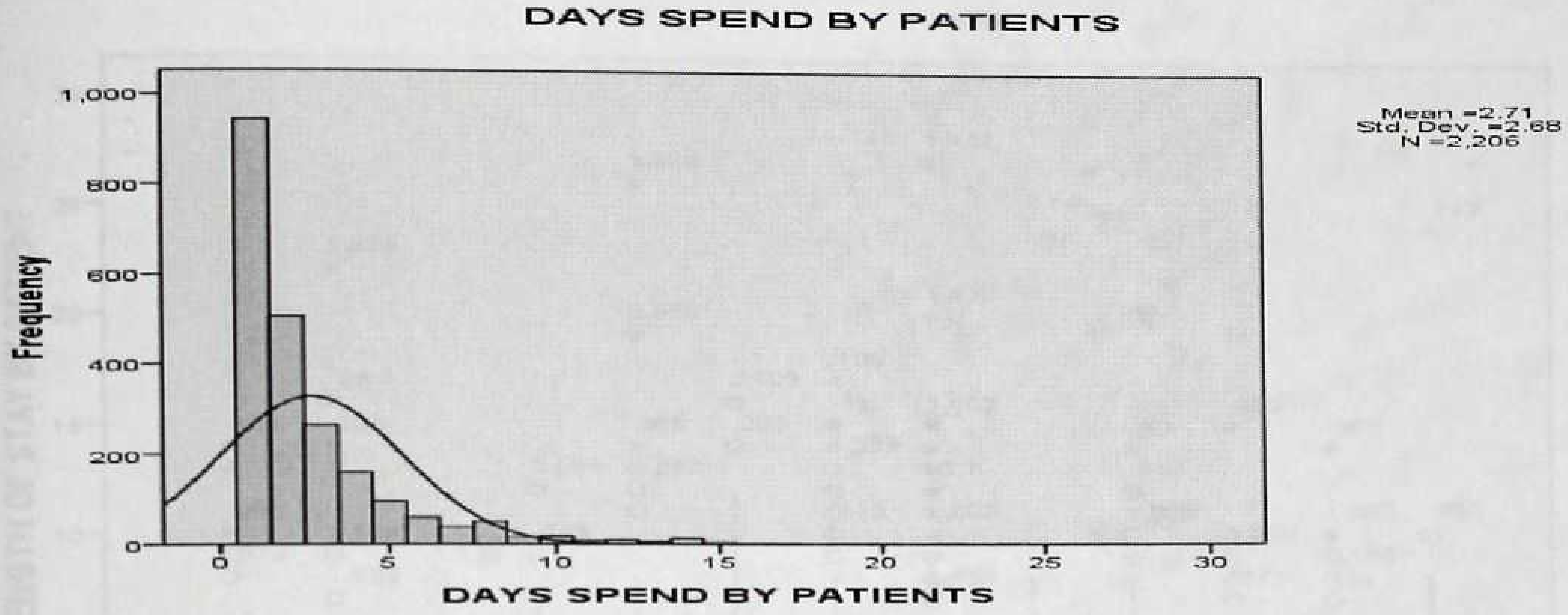

Figure 4.1: Histogram density for Length of Stay (LOS)

From figure 4.1, we observed a positively skewed distribution for LOS. Many patients stayed between 0 – 5 days and few patients stayed beyond 6 days to 17 days. In the male medical ward, the total number of patients who were admitted for the five year period was 2,206 was used for the analysis because it is complete case study. This means that a patient whose illness does not fall within the stated common diseases was not included in the analysis. Average age admitted at the male ward was 39 ± 0.424 years. This may mean that average men in their young adult age normally fall sick and were admitted at the medical ward. From the table 4.1in appendix indicates that within the said period, a disease with the highest frequency was malaria. Out of 2206 patients admitted in the ward, 1,067 reported malaria cases representing 48.4%. The least ailment during the period was Cholera with 28 cases which represent 1.3%. From the said table 4.1, other diseases like pelvic influmentary disease (PID) with 33 cases which present 1.5%, hernia with 38 cases representing 1.7% and typhoid with cases 30 which represent 1.4% does not have high frequency as compared to malaria which shown that almost half of the observed patients suffered said illness. Out of 2,206 male patients who were admitted at the hospital from 2005 to 2009, only 109 which represent about 5% did not survive while remaining 2097 which represent about 95% did survive.

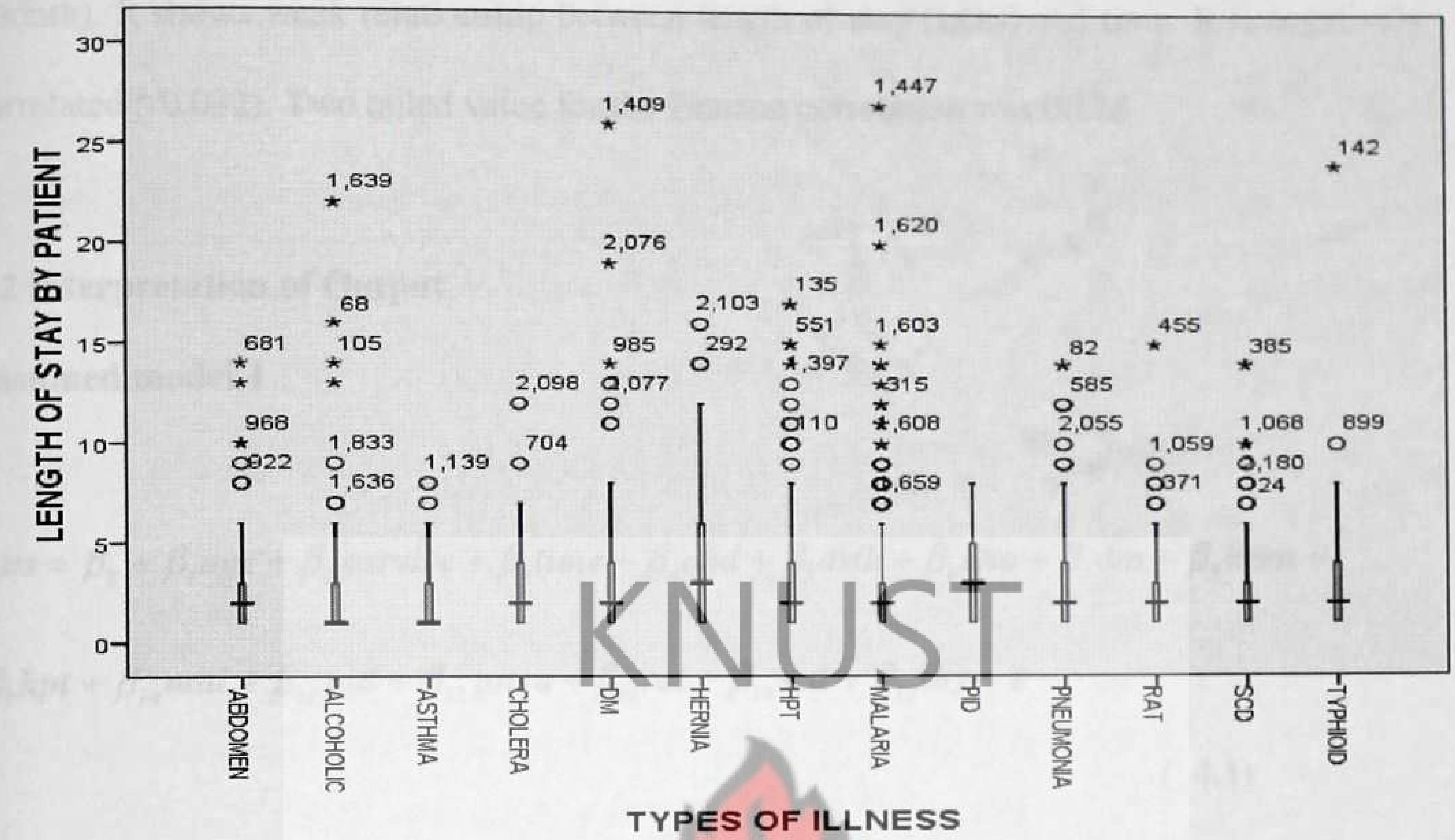

Figure 4.2: LOS against types of Illness in medical male ward

The above box plot in figure 4.2 shows median LOS for each illness and possible outliers' in the data. The figure clearly shows that most of the patients spent less than 10 days in the hospital as they suffered from various illnesses. The box plot also shows the various outliers with data where symbol "0" represents mild outliers and "*" represents extreme outliers. The figure 4.2 also showed that only one patient who suffered from malaria spent about 27 days. It observed from figure 4.2 that the median LOS for Hernia and PID are higher relative to the median LOS for the other eleven illnesses. Asthma and Alcoholic illness have the smallest median of LOS.

From table 4.1 shows in the appendix C, correlation between the lengths of stay and time (month). It shows weak relationship between length of stay (LOS) and time. It is negatively correlated (-0.032). Two tailed value for the Pearson correlation was 0.128

4.2 Interpretation of Output

Assumed model 1

$$Los = \beta_0 + \beta_1 age + \beta_2 survive + \beta_3 time + \beta_4 abd + \beta_5 asth + \beta_6 cho + \beta_7 dm + \beta_8 hern +$$

$$\beta_9 hpt + \beta_{10} mal + \beta_{11} pid + \beta_{12} pneu + \beta_{13} rat + \beta_{14} scd + \beta_{15} thy + \varepsilon \qquad (4.1)$$

4.2.1 Normal Model with identity link for LOS

Table 4.2: Parameter estimates for model 1

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	2.2646	0.3203	7.0710	0.0000
age	0.0120	0.0030	3.9920	0.0001
survive	0.0222	0.2649	0.0840	0.9332
time	-0.0018	0.0036	-0.4990	0.6176
abdom	-0.0054	0.3883	-0.0140	0.9889
asthma	-0.6000	0.4243	-1.4140	0.1575
chol	0.3239	0.5681	0.5700	0.5686
dm	0.6852	0.3883	1.7650	0.0778
hern	1.5460	0.5064	3.0530	0.0023
hpt	0.2890	0.3169	0.9120	0.3618
mal	-0.1586	0.2797	-0.5670	0.5709
pid	0.4131	0.5336	0.7740	0.4389
pneu	0.4683	0.3424	1.3670	0.1716
rat	-0.3042	0.3416	-0.8900	0.3734
scd	0.4649	0.3797	1.2240	0.2210
thy	0.8150	0.5531	1.4730	0.1408

From table 4.2 above, the most significant parameter estimates or predictors variables are the age, and hernia at 0.05 level of significance. They contribute significantly to the model. Survive, time, malaria, hernia, pneumonia, mhpt, asthma, pid, chol, rat, scd, thy, and abd contribute no significance to the model. However, at 10% level of significance, the predictor dm was significant. The Akaike Information Criterion (AIC) of this model was 10571.

The fitted generalized linear model for the predictor variables and the response variable is given as:

$$\hat{Los} = 2.264587 + (0.011955 \times age) + (0.022224 \times survive) - (0.001795 \times time)$$
$$-(0.005427 \times abd) - (0.600017 \times asth) + (0.323923 \times chol) +$$
$$(0.685167 \times dm) + (1.546035 \times hern) + (0.289007 \times hpt) - (0.158581 \times mal) +$$
$$(0.413148 \times pid) - (0.468257 \times penu) - (0.304170 \times rat) + (0.464860 \times scd) -$$
$$(0.814955 \times thy) \quad (4.2)$$

The mean LOS of model 1 was 2.2646 when there are no predictor variables. It was observed that patient age will determined his LOS in the ward. For instance patient who is 52 years old will spend more days in the ward than a patient who is 25 years old. The mean LOS of 52 year old man who admitted at ward will be increased by 0.6217 (approximate a day).

Assumed model 2

$$Los = \beta_0 + \beta_1 age + \beta_2 survive + \beta_3 time + \beta_4 abd + \beta_5 asth + \beta_6 cho + \beta_7 dm + \beta_8 hern + \quad \mathbf{4.3}$$

$$\beta_9 hpt + \beta_{10} mal + \beta_{11} pid + \beta_{12} pne\bar{u} + \beta_{13} rat + \beta_{14} scd + \beta_{15} thy + \varepsilon$$

4.2.2 Normal model with Natural Log of LOS

Table 4.3: Parameter estimates for model 2

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.4931	0.0870	5.6670	0.0000
age	0.0037	0.0008	4.5300	0.0000
survive	-0.0381	0.0720	-0.5300	0.5965
time	-0.0008	0.0010	-0.8540	0.3934
abdom	0.0406	0.1055	0.3850	0.7006
asthma	-0.0926	0.1153	-0.8030	0.4220
chol	0.1729	0.1543	1.1200	0.2628
dm	0.1846	0.1055	1.7500	0.0803
hern	0.3986	0.1376	2.8980	0.0038
hpt	0.1518	0.0861	1.7630	0.0780
mal	0.0416	0.0760	0.5470	0.5845
pid	0.2578	0.1450	1.7790	0.0755
pneu	0.2247	0.0930	2.4160	0.0158
rat	0.0171	0.0928	0.1850	0.8535
scd	0.2165	0.1032	2.0990	0.0360
thy	0.2148	0.1503	1.4290	0.1531

For model 2, the LOS was transformed on the natural log scale. This was as a results obtained from Box-Cox transformation. The fitted generalized linear model for the predictor variables and the response variable is given as:

$$\begin{aligned}
L\hat{N}(Los) &= 0.4930670 + (0.0036849 \times age) - (0.0381116 \times survive) - \\
&(0.008335 \times time) + (0.0405577 \times abd) - (0.0925688 \times asth) + (0.1728551 \times chol) + \\
&(0.1845783 \times dm) + (0.3986456 \times hern) + (0.1517558 \times hpt) + \\
&(0.0415649 \times mal) + (0.2578194 \times pid) + (0.2247075 \times pneu) + (0.0171440 \times rat) + \\
&(0.2164686 \times scd) + (0.2147658 \times thy)
\end{aligned} \tag{4.4}$$

From table 4.3 above four (4) predictors were significant at 0.05 level of significance (ie age, hernia, pneumonia and sickle cell disease). All other variables were not significant but at 10% level of significant dm, hpt and pid were also significant. The mean value of the model is $e^{0.4930670}$ which is 1.6373 (approximate two days) when there are no predictor variables. Point estimate of age is 0.0037 which means that age has an influence on patient's mean length of stay. That is old man who is 69 years old admitted at the ward will spend more days than patient who is 22 years old. Again patients suffering from hernia mean length of stay (LOS) will be increased by 0.3986456 when there are no predictor variables. The model has AIC value of 4823.5.

Assumed model 3

$$Los = \beta_0 + \beta_1 age + \beta_2 survive + \beta_3 time + \beta_4 abd + \beta_5 asth + \beta_6 cho + \beta_7 dm + \beta_8 hern +$$

$$\beta_9 hpt + \beta_{10} mal + \beta_{11} pid + \beta_{12} pneu + \beta_{13} rat + \beta_{14} scd + \beta_{15} thy + \varepsilon \tag{4.5}$$

4.2.3 Gamma Model with Log Link on LOS

Table 4.4: Parameter estimate for model 3

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.8257	0.1147	7.1970	0.0000
age	0.0045	0.0011	4.2080	0.0000
survive	0.0055	0.0949	-0.0580	0.9540
time	-0.0009	0.0013	-0.6970	0.4857
abdom	0.0045	0.1391	0.0330	0.9739
asthma	-0.2326	0.1520	-1.5300	0.1261
chol	0.1324	0.2035	0.6500	0.5156
dm	0.2313	0.1391	1.6630	0.0965
hern	0.4789	0.1814	2.6400	0.0084
hpt	0.1025	0.1135	1.9030	0.3668
mal	-0.0571	0.1002	-0.5700	0.5686
pid	0.1732	0.1912	1.9060	0.3649
pneu	0.1894	0.1227	1.5440	0.1227
rat	-0.1256	0.1224	-1.0270	0.3047
scd	0.1602	0.1360	1.1770	0.2392
thy	0.2972	0.1982	1.5000	0.1339

For model 3, the LOS was modelled using Gamma with log link. The model for predictor variables and the response variable is given as:

$$\hat{L}(Los) = 0.8257061 + (0.0045142 \times age) - (0.0054797 \times survive) - (0.0008977 \times time)$$
$$+ (0.0045461 \times abd) - (0.2326184 \times asth) + (0.1323528 \times chol) + (0.2313099 \times dm)$$
$$+ (0.4789052 \times hern) + (0.1024705 \times hpt) - (0.0571483 \times mal) + (0.1732405 \times pid) +$$
$$(0.1894208 \times pneu) - (0.1256440 \times rat) + (0.1601593 \times scd) + (0.2971664 \times thy) \quad (4.6)$$

From table 4.4 only two predictors were significant at 0.05 level of significant (ie age and hernia). All other variables were not significant. The mean LOS of the model 3 is 0.8257 when there are no predictors. The point estimated for predictors variables were (age = 0.0045 and hernia = 0.4790). Survive, time, abdomen, asthma, cholera, dm, hpt, malaria, pid, pneumonia, rat, scd, and thy contribute no significant to model at 5% level of significant. However, at 10% level of significance, the predictor dm was significance. The model has AIC value of 8368.7.

Assumed model 4

$$Los = \beta_0 + \beta_1 age + \beta_2 survive + \beta_3 time + \beta_4 abd + \beta_5 asth + \beta_6 cho + \beta_7 dm + \beta_8 hern + \beta_9 hpt + \beta_{10} mal + \beta_{11} pid + \beta_{12} pneu + \beta_{13} rat + \beta_{14} scd + \beta_{15} thy + \varepsilon \quad (4.7)$$

4 2.4 Gamma model with Inverse link on LOS

Table 4.5: Parameter estimate for model 4

Variable	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	0.4280	0.0428	10.0070	<2e-16
age	-0.0015	0.0004	-4.0500	0.0001
survive	-0.0011	0.0313	-0.0370	0.9709
time	0.0002	0.0005	0.4450	0.6565
abdom	0.0027	0.0530	0.0500	0.9598
asthma	0.1036	0.0663	1.5630	0.1182
chol	-0.0378	0.0692	-0.5460	0.5850
dm	-0.0684	0.0465	-1.4720	0.1412
hern	-0.1335	0.0512	-2.6050	0.0093
hpt	-0.0336	0.0413	-0.8120	0.4167
mal	0.0261	0.0382	0.6830	0.4946
hid	-0.0490	0.0646	-0.7580	0.4484
pneu	-0.0537	0.0436	-1.2320	0.2181
rat	0.0500	0.0487	1.0260	0.3051
scd	-0.0561	0.0480	-1.1680	0.2429
thy	-0.0929	0.0630	-1.4750	0.1404

Table 4.5 gives information on another model; this is where Gamma Log inverse was computed for LOS. Generalized linear model fitted for the predictor variables and response variable is given as:

$$\hat{Los}^{-1} = 0.4279942 - (0.0015062 \times age) - (0.0011439 \times survive) + (0.0002049 \times time) +$$

$$(0.0026691 \times abd) + (0.1036329 \times asth) - (0.0378013 \times chol) - (0.0684079 \times dm) -$$

$$(0.1334870 \times hern) - 0.0335720 \times hpt) + (0.0260652 \times mal) - (0.0489527 \times pid) -$$

$$(0.0537362 \times pneu) + (0.0499518 \times rat) - (0.0560587 \times scd) - (0.0929367 \times thy) \tag{4.8}$$

From the above model 4 only two (2) predictors were significant at 0.05 level of significance. All the rest of the predictors were not significant at 0.05 level of significant. Point estimate of significance predictors were (age = -0.0015 and hernia = -0.1335). The mean LOS of the model 3 is 0.4280 when there are no predictors. Point estimate of age is -0.0015 which means that the mean length of stay of age will decrease by -0.0015 when all predictor variables remain unchanged.

CHAPTER FIVE

DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

INTRODUCTION

This chapter deals with discussion of results, conclusion and recommendations.

5.1 Discussion

In the study 2206 observations were collected and used for the analysis. In the descriptive statistics, it was observed that the mean length of stay (LOS) was 2.71 (approximate 3 days) with the standard deviation of 2.68. Month which recorded highest admission cases was July 2008 as 79 and month recorded least admission cases was April 2007 as 6 patients admitted at the medical ward. Again majority of the patients spend 0 to 5 days at the medical ward while few spend beyond 6 to 17 days. Average age admitted at the ward was 39. It was shown that almost half of the patients reported with malaria (1068) represent 48.4% and it was few patients reported with cholera (28) representing 1.3%. Box plot shown median LOS for each illness and possible outliers', it was observed that most of the patients spent less than 10 days in the hospital as they suffered from various illnesses. It was only one patient who suffered from malaria spent 27 days in the ward. It was observed that the median LOS for Hernia and PID were higher relative to the median LOS for the other eleven illnesses while Asthma and Alcoholic had the smallest median LOS. There was also indication of no correlation between length of stay (LOS) and time ($r = -0.032$), that shown weak correlation between the LOS and time.

In the study it was observed that model 1 where normal data with LOS was analysed , only two predictors were significance at 0.05 level of significance (ie age = 0.0120 and hernia = 1.5460). The relation between LOS and the explanatory variables were explored using generalized linear models (GLM), since the untransformed LOS conditioned on the explanatory variables did not seem to follow a normal distribution. The Akaike Information Criterion (AIC) value obtained was 10571. Nature of the variable LOS does not allow it to take negative values, implying that it may possibly be more appropriately to use distribution with log to length of stay (LOS).

Model 2 was formulated with natural log computed for LOS to transform the model. It was observed that four predictor variables were significance at 0.05 level of significant (age = 0.0037, hernia = 0.3986, pneu = 0.2247 and scd = 0.2165). However at 0.1 level of significant dm, hpt and pid were also significance. The mean LOS of the model was 1.6373 when there are no predictor variables and its AIC value was 4823.5. Patient age influenced in his LOS; that will still make the patient spend more or few days in the ward depending on his age.

Model 3 was formulated where gamma log with link to LOS was computed. The results shown that only two predictor variables were significance at 0.05 level of significant (age = 0.0045 and hernia = 0.4789). Its mean LOS was 0.8257 when there are no predictor variables and AIC value was 8368.7. It was also observed that among the two significance variables, hernia had the highest estimated value (0.4789) which indicates that patient suffering from hernia will spend more days in the ward than other patient who reported with different illness.

Model 4 was formulated, where gamma inverse was computed for LOS. Only two predictor variables were significance at 0.05 level of significant. It was age and hernia which were significant with estimated parameters of (age = -0.0015 and hernia = -0.1335). Among all the models, it was model 4 with variable age had the smallest standard error value of (0.0004). AIC value was 8374.1.

Among the four Generalized Linear models fitted, model 2 which LOS was transformed on natural log scale was regarded as the optimised one accounting for its lowest AIC value. It had the lowest AIC value of 4823.5; most of its parameters were significance at 0.05 level of significance compare to the other three models.

5.2 Conclusion

Four generalized linear models were formulated and it was the model with the natural log of length of stay which was chosen because it had the lowest Akaike Information Criteria (AIC) value of 4823.5; lower than the gamma log and gamma inverse. It was observed that patient's age has an influence on his length of stay, aged patients spend more days in the ward while young patient spend few days in the ward. Again patients admitted with hernia, pneumonia and sickle cell disease also spend more days in the ward.

5.3 Recommendation

It is recommended that the model with the lowest AIC value should be adopted by hospital in order to minimise length of stay in the hospital.

Researchers who will like to do further work to find out the impact of the LOS in Hospital should extent their work to the whole region or beyond to compare the impact of one Hospital to others. They should also consider interviewing the patients in the wards to find out their opinions about the effectiveness of the work. In this study, only secondary data were used.

Reference

1. Adam W., Li L., Meng Q. and Magnuus L. (2006). Health service delivery in Chain: A literature Review, World Bank Policy Research working paper 3978, Email; awagstaff@worldbank.org
2. Adebanjo D. and Mann R., (2000). Identifying problem in forecasting consumer demand in the fast moving consumer goods sector. *Benchmarking*: *An International Journal vol.* 7, page 223 – 230.
3. Aiken L. H., Clarke S.P., Sloane D.M., Sochalski J. And Silber J. H (2002). Hospital nurse staffing and patient mortality, nurse burn out and job dissatisfaction, *journal of the American medical Association,* vol. 288, page 1987 – 1993.
4. Akaik H., (1973). Information theory and an extension of the maximum likelihood principle. Budapest: Akademiai Kiada.
5. Amaravadi R. K., Dimick J.B., Pronovost P.J. and Lipsett P.A. (2000). Intensive care unit (ICU) nurse- to- patient ratio is associated with complication and resource use after esophagectomy, *Intensive Care Unit,* vol. 26, page 1857 – 1862.
6. Antomioni D (1999). "What motivate middle manager"? *Industrial management*, vol. 41 No 6 page 27 -30.
7. Carmen B. (2009), Direct costs associated with the appropriateness of hospital stay in elderly population, *BMC health service research,* vol. 2 page 472.
8. Channout N (2007). "The application of forecasting techniques to modelling emergency medical system calls in Calgary, Alberta" *Health care managing science,* vol. 10, page 25 – 45.
9. Chi – Hung S. (2002). Apply Fuzzy neural network in hospital staff management under Global Budget system – with a Regional Hospital located at Yunlin country as case study, Industrial Engineering and management in National Yunlin University.

10. Clarke A., (2002). Length of in hospital stay and its relationship to quality of care. *Quality safety health care*, vol. 11, page 209 - 210.

11. Clarke A. and Rosen R., (2001). Length of stay. How short should hospital care be? *European journal of Public health* vol. 11 page 166 – 170.

12. Contreras J (2003). "I E E E Transaction on power system", vol. 18 No. 3, page 1014 – 1020.

13. Earnest A., Chen M.J. Ng D. and Sin L.Y. (2005). "Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in tertiary hospital in Singapore". *BMC Health Service Research*, vol. 5

14. Ekman B. (2004). Community based health insurance in low – income countries; a systematic review of the evidence. *Health policy and planning*, vol.19, page 249-270.

15. Gadallah M., Zaki B., Rady M., Anwer W. And Sallam I. (2003). Patient satisfaction with primary health care service in two districts in lower and Upper Egypt. *La Revne de sant's de la Mediterrane'e orientale* vol. 9, 422 – 430.

16. Gertman P. M. and Restuccia J. D.(1981). The appropriateness evaluation protocol*: a technique for assessing unnecessary days of hospital care. Med care*, vol. 19: page 855-857.

17. Greta G. C. (2007). The impact of nurse staffing on hospital cost and patient length of stay*: A systematic review, nursing economics*. Vol. 25, no. 5 page 261.

18. Gumber A. (2001). Hedging the health of the poor: the case for community financing in India. Health, Nutrition and Population discussion paper. Washington DC: World Bank.

19. Helapota H. A (2005). "Motivational theories and their application in construction" *Cost Engineering,* vol. 47 No3 page 14 – 35.

20. Hiamah J. K. (2007). Evaluation of the internal control system of the central regional health directorate, page 13.

21. Hunter S. (2000). Determination of moral negligence in the context of the under medication of pain by nurses. *Nursing ethics*, vol. 7, page 379 – 391.

22. Jakas M. and Krishnan (2001). Community involvement in health care financing: a survey of the literature on the impact, strengths and weaknesses. World Bank Health, nutrition and population discussion paper, Wishington, DC, World Bank.

23. Jian – Shen C., Chia-Hui L., Mei C., An-Ping W. (2006). The study on auditing the expenditure of National Health Insurance. *Chao – Yang Business and Management Review*, vol. 5, page 111 – 130.

24. Jlitting J. P. (2004). Do community based health insurance scheme improve poor people's assess to health care*? Evidence from Rural Senegal, world Development*, vol. 32, page 273 – 288.

25. Johes S. A., Joy M.P. and Pearson J. (2002). "Forecasting demand of emergency care"*, Medline; Health care managing science*, page 297 – 303.

26. Jonny P. (2008). Health policy and planning: *Equity in community health insurance scheme; evidence and lesson from Armenia* vol. 16, page 209 – 213.

27. Karnon J., Mark M. and Mills T.M. (2009). "Mathematical modelling in health care" *18th World IMACS/ MODSIM congress,* vol.13 page 13 – 17.

28. Liu K. (2000). Establish a Risk – Adjusted model with capitation payment by neural network. Department of Risk management in national Kaosgiung, First University of Science and Technology.

29. Lin C., Elif A. and Murray J. C (2006). "A network flow approach to optimizing hospital bed capacity decision", *Health care managing science,* vol. 9, page 391 – 404.

30. Maslow A.H.(1943). A theory of human motivation, Pscyhological Review, vol. 50, page 370.

31. M c Cue M. Mark B.A., and Harless D. (2003). Nurse staffing, quality and financial performance. *Journal of Health care finance,* vol. 29, page 54 – 76.

32. M c Guire D. (1992). Comprehensive and multidimensional assessment and measurement of pain. *Pain symptom manages* vol. 17, page 312 – 319.

33. M c Kenzie D. (2003). "Measuring inequality with Asset indicators". *Bureau for Research in Economic Analysis of Development working paper*, number 042 Cambridge. M A center for international development, Harvard University.

34. Machla S.R. and Corper K. (2004). Expense for inpatient hospital stay. Statistical brief, number 164.

35. Manias E. (2003). Medication trend and documentation of pain management following surgery, *Nurse Health Science*, vol. 48, page 234 – 242.

36. Marik E. P. and Hedman L., (2000). What is a day? Determining intensive care unit length of stay, *Critical Care Medicine*, vol. 28, 2090 – 2093.

37. Mark R., (1983). Variations in hospital length of stay: The relationship to health outcome, page 37 Washington D.C

38. Mastorocosta P. A., Theocharis J.B. and Petridis V.S (2001). A constrained orthogonal least squares method for generating TSK fuzzy method: Application to short term load forecasting. *Fuzzy set and system*, vol. 118. Page 225 – 233.

39. Mattke S., Needleman J., Peter I.B., Maureen S. and Zelevinsky K. (2006). Nurse – Staff in hospital: Is there a business case for quality. *Journal of Health Affairs*, vol. 25, page 204 – 211.

40. Khashei M. and Bijari M. (2010). An artificial Neural network (p,d,q) model for time series forecasting expert system application, vol. 37, page 479 – 487.

41. Norman V., (2003). Inappropriately delayed discharge from hospital: *What do we know? Education and debate*, vol. 326, page 927.

42. Perry J. K., (1996). Measuring public service motivation; an assessment of construct reliability and validity. *Journal of public administration research and theory*, vol.6, no.1, page 5-23.

43. Sekhri N. and Savedoff, (2005). Private health insurance implications for developing countries. *Bulletin of the World Health organization*, vol. 83, page 127 – 138.

44. Shih – Fang C., (2002). Risk classification and Risk – Adjusted model in capitation payment. Department of Risk Management in National Kaoshiung, First University of Science and technology, Master's thesis.

45. Shuo – Fen H., Chao-Hsin L. and Ya-Ling Y (2006). Risk classification and prediction of individual health expresses. *Management Review*, vol. 25 page 27 – 47.

46. Tu J. V., (1993). "A comparison of Neural Network and Logistic Regression Model for predicting Length of Stay in the intensive care unit Following cardiac surgery". Ph.D Dissertation, University Toronto.

47. Weisbrod B., (1991). "The health care quodrilemmo; An essay on technological change, insurance, quality of care and cost containment". *Journal of Economic Literature*, vol. 15, page 523-552.

48. White B. (1999). Measuring patient satisfaction. How to do it and why bother retrieved, April 28, 2008 from htt// www.acfg.org

49. World Health Organization (WHO) (2004). Regional overview of social health insurance in south – east Asia New Delhi

50. World Health Organization [WHO] (1946). Constitution of the world health organization.

51. World Health Organization [WHO] (2000). Health system: Improving performance. Geneva: World Health Organization.

52. World Health Organization [WHO] (2003). How can hospital performance be measured and monitored.

53. Zhang G.P.,(2003). Time series forecasting using a hybrid ARIMA and Neural Network model, Journal of Neural computing, vol. 50, page 159-175.

APPENDICES

APPENDIX B

Table B.1 DESCRIPTIVE ANALYSIS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ABDOMEN	89	4.0	4.0	4.0
	ALCOHOLIC	98	4.4	4.4	8.5
	ASTHMA	65	2.9	2.9	11.4
	CHOLERA	28	1.3	1.3	12.7
	DM	91	4.1	4.1	16.8
	HERNIA	38	1.7	1.7	18.5

HPT	259	11.7	11.7	30.3
MALARIA	1067	48.4	48.4	78.6
PID	33	1.5	1.5	80.1
PNEUMONIA	155	7.0	7.0	87.2
RAT	156	7.1	7.1	94.2
SCD	97	4.4	4.4	98.6
THYROID	30	1.4	1.4	100.0
Total	2206	100.0	100.0	

Table B.2 DAYS SPEND BY PATIENTS IN THE WARD

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	943	42.7	42.7	42.7
	2	505	22.9	22.9	65.6
	3	266	12.1	12.1	77.7
	4	160	7.3	7.3	85.0
	5	96	4.4	4.4	89.3
	6	60	2.7	2.7	92.0
	7	39	1.8	1.8	93.8
	8	51	2.3	2.3	96.1
	9	16	.7	.7	96.8
	10	19	.9	.9	97.7

11	8	.4	.4	98.1
12	10	.5	.5	98.5
13	6	.3	.3	98.8
14	13	.6	.6	99.4
15	4	.2	.2	99.5
16	2	.1	.1	99.6
17	2	.1	.1	99.7
19	1	.0	.0	99.8
20	1	.0	.0	99.8
22	1	.0	.0	99.9
24	1	.0	.0	99.9
26	1	.0	.0	100.0
27	1	.0	.0	100.0
Total	2206	100.0	100.0	

Table B.3 SURVIVE OR NOT SURVIVE PATIENTS

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	SURVIVE	2097	95.1	95.1	95.1
	NOT SURVIVE	109	4.9	4.9	100.0
	Total	2206	100.0	100.0	

APPENDIX C

Table C.1 Correlations Between LOS Time

		LLOS	TIME
LLOS	Pearson Correlation	1	-.032
	Sig. (2-tailed)		.128
	N	2206	2206
TIME	Pearson Correlation	-.032	1
	Sig. (2-tailed)	.128	
	N	2206	2206

APPENDIX D

Model output

MODEL 1

```
glm(formula = LOS ~ AGE + Survive + Time + ABDOM + ASTHMA + CHOLERA +
    DM + HERNIA + HPT + MALARIA + PID + PNEUM + RAT + SCD + THPHOID,
    family = gaussian(link = "identity"), data = mydata)
```

Deviance Residuals:

```
    Min      1Q  Median      3Q     Max
```

-3.6429 -1.4628 -0.8474 0.6004 24.4505

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.264587	0.320274	7.071	2.06e-12	***
AGE	0.011956	0.002995	3.992	6.75e-05	***
Survive	0.022224	0.264933	0.084	0.93315	
Time	-0.001795	0.003594	-0.499	0.61759	
ABDOM	-0.005427	0.388297	-0.014	0.98885	
ASTHMA	-0.600017	0.424307	-1.414	0.15747	
CHOLERA	0.323923	0.568110	0.570	0.56862	
DM	0.685167	0.388287	1.765	0.07777	.
HERNIA	1.546035	0.506407	3.053	0.00229	**
HPT	0.289007	0.316861	0.912	0.36182	
MALARIA	-0.158581	0.279743	-0.567	0.57085	
PID	0.413148	0.533640	0.774	0.43889	
PNEUM	0.468257	0.342446	1.367	0.17164	
RAT	-0.304170	0.341618	-0.890	0.37336	
SCD	0.464860	0.379726	1.224	0.22101	
THPHOID	0.814955	0.553141	1.473	0.14081	

AIC: **10571**

MODEL 2

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.4930670	0.0870020	5.667	1.64e-08	***
AGE	0.0036849	0.0008135	4.530	6.23e-06	***

Survive	-0.0381116	0.0719688	-0.530	0.59647	
Time	-0.0008335	0.0009763	-0.854	0.39340	
ABDOM	0.0405577	0.1054804	0.385	0.70064	
ASTHMA	-0.0925688	0.1152624	-0.803	0.42200	
CHOLERA	0.1728551	0.1543263	1.120	0.26281	
DM	0.1845783	0.1054776	1.750	0.08027	.
HERNIA	0.3986456	0.1375648	2.898	0.00379	**
HPT	0.1517558	0.0860748	1.763	0.07803	.
MALARIA	0.0415649	0.0759918	0.547	0.58446	
PID	0.2578194	0.1449626	1.779	0.07546	.
PNEUM	0.2247075	0.0930250	2.416	0.01579	*
RAT	0.0171440	0.0928001	0.185	0.85345	
SCD	0.2164686	0.1031522	2.099	0.03597	*
THPHOID	0.2147658	0.1502600	1.429	0.15306	

AIC: 4823.5

MODEL 3

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	0.8257061	0.1147355	7.197	8.44e-13	***
AGE	0.0045142	0.0010728	4.208	2.68e-05	***
Survive	0.0054797	0.0949101	0.058	0.95396	
Time	-0.0008977	0.0012876	-0.697	0.48573	
ABDOM	0.0045461	0.1391042	0.033	0.97393	
ASTHMA	-0.2326184	0.1520044	-1.530	0.12608	

CHOLERA 0.1323528 0.2035206 0.650 0.51556

DM 0.2313099 0.1391006 1.663 0.09648 .

HERNIA 0.4789052 0.1814161 2.640 0.00835 **

HPT 0.1024705 0.1135127 0.903 0.36677

MALARIA -0.0571483 0.1002156 -0.570 0.56856

PID 0.1732405 0.1911721 0.906 0.36493

PNEUM 0.1894208 0.1226785 1.544 0.12272

RAT -0.1256440 0.1223819 -1.027 0.30470

SCD 0.1601593 0.1360338 1.177 0.23918

THPHOID 0.2971664 0.1981581 1.500 0.13385

AIC: 8368.7

MODEL 4

glm(formula = LOS ~ AGE + Survive + Time + ABDOM + ASTHMA + CHOLERA +

DM + HERNIA + HPT + MALARIA + PID + PNEUM + RAT + SCD + THPHOID,

family = Gamma(link = inverse), data = mydata)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.3011	-0.7752	-0.3138	0.2186	3.8346

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.4279942 0.0427682 10.007 < 2e-16 ***

AGE -0.0015062 0.0003719 -4.050 5.29e-05 ***

Survive -0.0011439 0.0312999 -0.037 0.97085

Time 0.0002049 0.0004607 0.445 0.65647

ABDOM	0.0026691	0.0529662	0.050	0.95981	
ASTHMA	0.1036329	0.0663028	1.563	0.11819	
CHOLERA	-0.0378013	0.0692119	-0.546	0.58501	
DM	-0.0684079	0.0464716	-1.472	0.14115	
HERNIA	-0.1334870	0.0512493	-2.605	0.00926	**
HPT	-0.0335720	0.0413276	-0.812	0.41669	
MALARIA	0.0260652	0.0381575	0.683	0.49462	
PID	-0.0489527	0.0645678	-0.758	0.44844	
PNEUM	-0.0537362	0.0436221	-1.232	0.21813	
RAT	0.0499518	0.0486932	1.026	0.30508	
SCD	-0.0560587	0.0479874	-1.168	0.24285	
THPHOID	-0.0929367	0.0630185	-1.475	0.14042	

AIC: 8374.1