Scholars Bulletin *(Engineering and Technology)*

An Official Publication of "Scholars Middle East Publishers",
Dubai, United Arab Emirates
Website: http://scholarsbulletin.com/

ISSN 2412-9771 (Print)
ISSN 2412-897X (Online)

Normality Fit to the Population Data of Health Insurance: A Catalyst for Managerial Decisions

Zakari Abubakari[1], Iddrisu Abubakari[2], Dorcas Kouame[3], Marian Maclean[4]
[1,4]Kumasi Technical University, Kumasi, Ghana
[2]Dormaa Senior High School, Dormaa Ahenkro, Ghana
[3]University of Electronic Science and Technology of China, Chengdu, China

***Corresponding author**
Zakari Abubakari

Article History
Received: 31.08.2017
Accepted: 07.09.2017
Published: 30.09.2017

DOI:
10.21276/sb.2017.3.9.3

Abstract: It is always important to determine distribution of insurance claims in order to estimate future expected values. This study seeks to determine the normality fit to the population data of health insurance. Secondary data collected from Dormaa Municipal Health Insurance Scheme and Dormaa Presbyterian Hospital was analyzed using the Statistical Package for Social Science (SPSS), Excel spreadsheet, and Easy fit. It was found that the population data submitted by 28 health facilities to Dormaa Municipal Health Insurance Scheme follows normal distribution.

Keywords: Normal distribution; health insurance; national health insurance scheme; normal distribution; claim modeling; population data

INTRODUCTION

The socio-economic development of a nation, to a very large extent, depends on a healthy population. A healthy people with the requisite skills and knowledge become the back bone for economic development and transformation. It is in this light that every country in the world places much premium on improved health care for her citizens. It has therefore become the state's responsibility to provide health care to the people of Ghana and this responsibility comes with financial challenges in view of the difficult economic issues that confront the country. Financing an efficient and effective health care system is of a major concern to countries all over the world, especially developing economies. In Ghana, healthcare financing has gone through several phases. After independence in 1957, the provision of health care in Ghana was financed by the state through tax revenue. However, it became obvious that this method of financing health care was not sustainable following the economic difficulties the country experienced from the beginning of the 1960s [1]. As part of efforts to revamp the Ghanaian economy, the state began to reduce expenditure on the provision of social services and the health sector witnessed considerable reduction in state funding. The then government in 1985, introduced user fees for all medical conditions, except certain specified communicable diseases. This system whereby people who access public health facilities pay user fees became known as 'cash and carry' and resulted in several operational challenges as well as people in the country [2]. In order to ameliorate the problems associated with the "cash and carry" system, government introduced the National Health Insurance Law, Act 650 in August, 2003. It sought to provide basic health care services to persons resident in Ghana through mutual and private health insurance schemes, and to establish a National Health Insurance Fund that will provide subsidy to licensed District Mutual Health Insurance Schemes. Health Insurance is an alternative health care financing system which involves resource pooling and risk sharing among members [3]. The Health Insurance Act mandates the creation of district-level Mutual Health Organizations (MHOs) in accordance with national guidelines and the establishment of a National Health Insurance Council (NHIC). The law represents a bold and innovative move by government to provide health insurance coverage to all of its citizens. This is meant to provide financial protection for the entire population and move away from the "Cash and Carry" system which was creating considerable equity concerns, largely due to the non-functional exemption mechanisms.

In view of the economic importance of National Health Insurance Scheme in developing countries, there is a need to use actuarial analysis to model the distribution of the claim amounts presented to the claim center, which can then be used to estimate the most appropriate credibility factor for the best expected claim for the near future.

LITERATURE REVIEW

Literature on actuarial modeling in insurance data

T. Wright [4] and L. Hua [6] used actuarial modeling to fit models to many claim amount drawn from consecutive years. He fitted analytic loss distribution using maximum likelihood estimation for each of the years. He used a lot of models such as Pareto, Burr, Inverse Burr, and Lognormal to fit his data and test to select the best fit one.

R. V. Horg [7] also used Weibull distribution to fit 35 observations of hurricane loss and they found that Weibull distribution performs as well as the lognormal distribution.

J. P. Boucher [5] and A. E. Renshaw [8] used actuarial modeling to fit a statistical distribution to some claim amounts, the statistical distribution he tested were Lognormal, Gamma, and Weibull distributions. He used maximum likelihood estimation to fit the distribution, an idea that was extended to his research, only that he based his methodology on the Bayesian solution. He then used the likelihood functions to calculate the posterior distribution.

P. Cizek *et al* [9] explain that a typical models for insurance risk has two main component: one characterizing the frequency of claims (or incidence) of events, these are analyzed by discrete models such as poison distribution, binomial distributions, and another describing the severity (size or amount) of loss resulting from the occurrence of the event, these are also analyzed by continuous models such as Normal, lognormal distribution, Gamma distribution, Weibull distribution etc.

The normal distribution

Normal distribution is the type of continuous distribution that is most extensively used. The normal distribution is continuous and has 2 parameters, μ and σ; they determine the location and scale, respectively [10]. It is, therefore, described by its mean and variance. The mean of the distribution is referred to as the location parameter, and the (standard deviation) measures how the distribution is spread out (known as the scale parameter). Normal distribution is important in actuarial science and has a single peak at the center of the distribution. Generally, finance related random variables follow normal distribution, so knowledge about normal distribution is vital in understanding portfolio theory. The normal curve falls off smoothly in either direction from the central value which is the mean of the distribution. So the probability density function is bell-shaped and symmetrical about the mean. Its formula is given by:

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

Where $-\infty < x < \infty$ and $-\infty < \mu < \infty$ $\sigma > 0$

Its Mean and Variance are given as $E(X) = \mu$ and $VarX = \sigma^2$

METHODOLOGY

The normal distribution

The normal distribution formula for the research is given by:

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

Where $-\infty < x < \infty$ and $-\infty < \mu < \infty$ $\sigma > 0$

It Mean and Variance are given as $E(X) = \mu$ and $VarX = \sigma^2$

Descriptive statistics

The summary statistics of the aggregate monthly claim amount was used in pointing out the salient features of the data in order to determine the distribution for the population data. The descriptive statistics obtained are the mean, mode, median, skewness, etc. These are obtained by using SPSS and Excel spreadsheets on the 84 data points from the scheme office. The logic behind the best model selection criteria will be based on the results of summary statistics and the test statistics.

Research design and population

Secondary data was used from Dormaa Municipality Health Insurance Scheme, regarding their claims presented from the health center (January 2008-December 2014). Two assumptions were made on the data before using them:

1. All the claims came from the same distribution (they are independent and identically distributed). This means that there is no correlation between reported claims.

2. All future claims were to be generated from the same distribution, that is, the future claims are to follow the same distributions as the past claims. The targeted population for the study was made up of all the 28 accredited health care providers in the Dormaa Municipality that submit claims to the Municipal Health Insurance Scheme.

RESULTS

Table 1: Summary Statistics of population ($\times$ 0000)

N	Mean	Median	Mode	Std deviation	Variance	Skewness	Range	Sum
84	39.312	38.469	36.456	7.9557	63.297	0.028	32.414	3302.2

Source: Dormaa Municipal Health Insurance, Computed by Authors

The 95% confidence level for the mean is $37.586 \leq \mu \leq 41.039$.

After obtaining the descriptive statistics, the histogram of the aggregate claims, probability plot (P-P plot) and goodness of fit test, that is, Kolmogorov-Smirnov test, Anderson Darling test and chi-square, were also used to identify which distribution really fit the aggregate monthly claims. This was done by comparing the p-values at 5% and significant level in the goodness of fit test. Figure 1 and 2, and table 2 shows the histogram, P-P plot and the goodness of fit test respectively.

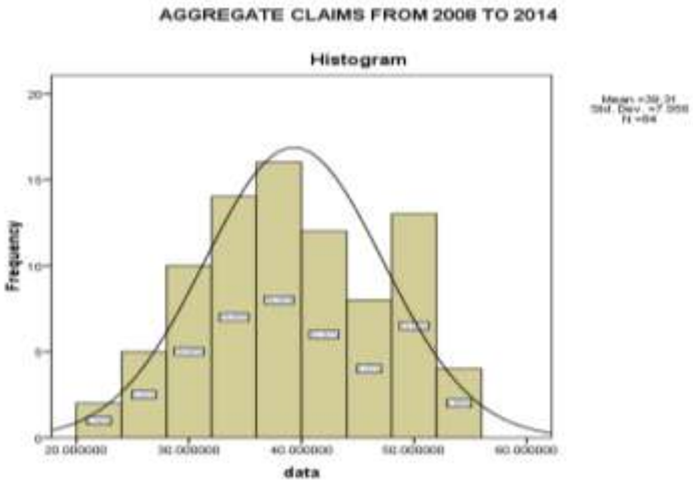

Fig-1: Histogram showing claims data from 2008 to 2014

The histogram above with normal curve superimpose on it indicates normality. The shape of the histogram are also well arranged to depict normal curve.

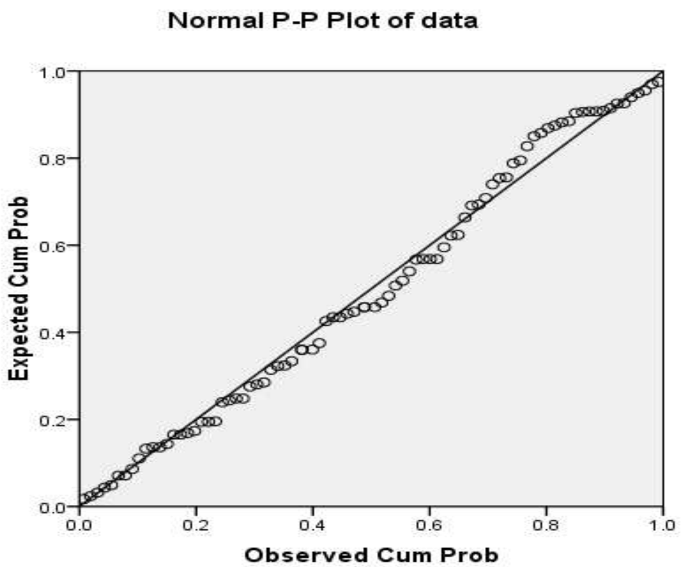

Fig-2: P-Plot showing claims data from 2008 to 2014

Table 2: Goodness of fit test

Normal					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value	84 0.07645 0.68118				
Significant level	0.2	0.1	0.05	0.02	0.01
Critical Value	0.11508	0.13148	0.14605	0.16331	0.17523
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic	84 0.50067				
Significant level	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value	6 5.2062 0.51765				
Significant level	0.2	0.1	0.05	0.02	0.01
Critical Value	8.5581	10.645	12.592	15.033	16.812
Reject?	No	No	No	No	No

DISCUSSION

The histogram in Figure 1 had a normal curve shape and it also shows that the data is not heavily skew in any direction as depicted in the summary statistics, which shows the skewness value of 0.028. From the histogram it can be seen that the original aggregate claims data are centered on the mean. The summary statistics also shows that, the mean, median and the mode are closed to each other. This was interpreted to mean that the claims data had few claims amount at their tail ends, with very low and high values, while most of the claim amounts were centered at the middle. Again, the normal p-plot in figure 2 shows that all the data points are much close to the straight line. In other words, the aggregate claim amounts from the population are much close to the straight line. This shows the evidence of normality.

The results in table 2 show that the p-values for both tests are all greater than the significant value, set at 5%. We can, therefore, conclude that the aggregate claim amount from Dormaa municipal health insurance scheme follows normal distribution.

CONCLUSION

Generally, finance related random variables follow normal distribution, so knowledge about normal distribution is vital in understanding portfolio theory and modeling claims distribution of insurance schemes. The essence of determining the best fit distribution to past data is to enable researchers, financial institutions, and insurance schemes etc. to be able to identify the best prior distribution in estimating the posterior distribution of insurance claims. Choosing the most suitable claim distribution will help Actuaries to use Bayesian's credibility theory to estimate the most appropriate credibility factor for the best expected claim for the near future.

Claim modeling is very important, since a good understanding and interpretation of claim distribution is the back-bone of all the decisions made in the insurance industry regarding expected profits and reserves necessary to ensure profitability and the impact of re-insurance [11]. This study, therefore, seeks to provide further empirical details for the understanding of the distribution of claims amount submitted to National Health Insurance Scheme offices.

REFERENCES

1. Osei-Akoto. (2004). The economics of rural health: the effects of formal and informal risk-sharing schemes in Ghana. in *Development Economics and Policy, 40*.
2. Asenso-Okyere, W. K., Osei-Akoto, I., Anum, A., & Appiah, E. N. (1997). Willingness to pay for health insurance in a developing economy: a pilot study of the informal sector of Ghana using contingent valuation. *Soc. Sci. Med.*, *42*(3), 223–237.
3. Ministry of Health (2003). National Health Insurance Policy Framework for Ghana. Accra.
4. Wright, T. (2005). FIA, Deloitte, UK. COTOR Challenge Round 3, 2005. [Online]. Available: atwww.casact.org/cotor/.

5. Boucher, J. P. (2014). *Regression with count dependent variables. In Predictive Modeling Applications in Actuarial Science; E.W. Meyers, G, Derrig, R.A. Eds.;* Cambridge University Press.
6. Hua, L. (2015). Tail negative dependence and its applications for aggregate loss modeling. *Insur. Math. Econ.*, *61*, 135–145.
7. Horg, R. V., & Klugman, S. A. (1983). On the estimation of long-tailed skewed distributions with actuarial applications. *J. Econom.*, *23*, 91–102.
8. Renshaw, A. E. (2004). Modeling the claim process in the presence of covariates.
9. Cizek, P., Hardle, W., & Weron, R. (2005). *Statistical Tools for Finance and Insurance*. New York, NY: Springer Berlin Heidelberg.
10. Casella, G., & Berger, R. L. (2002). *Statistical Inference*, 2nd Editio. Duxbury Advanced Series.
11. Boland, P. J. (2006). *Statistical Methods in General Insurance*. Chapman and Hall/CRC Inc.