International Journal of Modern Computation, Information and Communication Technology
2020;3(5-6):48-55.
ISSN: 2581-5954
http://ijmcict.gjpublications.com

Research Article

Survival Analysis in Estimation of Attrition Rates among Policyholders in Jubilee and UAP Insurance Companies

Jeffar J. Oburu, Richard Simwa, N. B. Okelo*

School of Mathematics, Applied Statistics and Actuarial Science,
Maseno University, P. O. Box 40105-333, Kenya.

*Corresponding author's e-mail: bnyaare@yahoo.com

Abstract

The present work has used survival models to estimate attrition rates, probabilities, and predictions as well as identify which covariates influence attrition. The study adopts a retrospective explanatory research design. A sample of 149 in-force policies within 1st January 2018 to 30th June 2019 have been used in the analysis. Kaplan Meir curves have been used to show the distribution of survival/retention times.

Keywords: Survival analysis; Attrition; Policy; Insurance.

Introduction

Like many other companies, insurance companies face attrition. Attrition happens when policy holders or the insured switch from one insurance to another for various reasons or simply fail to renew their contract. Many studies have established that customer retention is common concern for insurance companies as it is crucial for increased sales and saving in market. Policyholders' attrition is thus a challenge within the insurance industry. Attrition not only reduces the company's profits but also its growth. Whenever a policy holder withdraws, much effort that was put in place to win the customer goes to waste. There is a significant link between customer satisfaction, customer trust, price perceptions, switching barrier and customer attention. Poor customer satisfaction in the insurance makes the policyholders to switch to competing insurers. A number of studies on service quality and customer retention have found a positive link between service quality and customer retention [1]. However, high quality products at high prices were not acceptable among price sensitive customers. The perception that they were being charged more for a cover, also tend to switch cheaper to insurers for a similar products.

Oketch [2] researched on factors influencing customer's retention at Madison insurance. The study found that there were various measures put in place to retain the customers at the company including a dedicated customer service department, streamlining claims management, renewal notices management and regular follow ups to the clients. However, this study was only concentrated at Madison insurance. Many insurance companies also have so many agents and that brings issues of mistrust as some new clients have been duped to pay premiums and when the risk occurs they realize that they have been conned as they can't be compensated [3].

Other policyholders experience hardships trying to make claims as their compensation is delayed too and hence they tend to switch to other insurances. With the emergence of too many insurance companies, it is very easy for a policy holder to shift from one company to another. Traditionally, classification models such a logistic regression were used to predict customer attrition in insurance companies. However, this convention method of analysis did not differentiate between the mid-term policy cancellation and end-term policy non-renewal. Hence, recently, studies have sought to use survival analysis models to predict attrition. Survival analysis is a panel data method of analysis which tends to estimate time until an event of interest occurs [4]. The event of interest

Received: 10.03.2020; Received after Revision: 31.05.2020; Accepted: 01.06.2020; Published: 13.06.2020

can be death, getting a disease, change of marital status, employment, etc. The time to occurrence of the event or survival time can be measured in days, weeks, months, years, etc. For example, if the event of interest is withdrawal from a policy subscription, time to this event could be measured in terms of days, weeks, months or years. In survival analysis, subjects are usually observed for a specified time and the focus is on the occurrence of the event of interest [5].

Survival analysis techniques have been adopted in establishing various links between covariates and the survival times. Originally, survival analysis was only used in modelling mortality where in the seventeenth century Graunt published the first Weekly Bill of Mortality in London and Healey published the first life table. Life tables were thereafter adopted and used by actuaries, statisticians, government agencies and researchers. Traditionally, survival analysis has been used in the area of medicine, health sciences and engineering to estimate survival times. In medicine, survival analysis is being used to estimate time until a medical event occurs. These events include death of participants of a cohort group, recovery from a disease, addiction cessation among others. In the field of engineering, survival analysis have been used to predict failure in mechanical systems while in public health these techniques have been used in life time analysis on epidemiology. With the emergence of big data and machine learning, survival analysis has been expanded to other usages and not only in estimations of time to death. The techniques are now being tested in banking industry to estimate loan defaults, in marking to estimate churn, in criminology to estimate probing time on prisoners, in marriage lifetimes, insurance claims and also modeling lifetimes of electronic devices among others. Survival analysis has the ability to address whether the policyholder left or not and when. It can also analyze mid-term cancellation and end-term nonrenewal sequentially and therefore provide a better insight on the attrition [5].

The survival models can also take into account time varying covariates and hence can help in understanding how the selected macro-economic factors influence the rate, probabilities and time of attrition. Various studies have applied survival analysis techniques. A study by [6] compared estimation on insurance attrition using survival methods and the use of traditional methods of logistic regression. The study used a panel data from various insurances and established that survival analysis was superior to logistic regression in that it took into account midterm policy cancellation as well as end term nonrenewal of policies whereas the logistic regression generalized this attrition. The study also stated that survival analysis used in attrition estimation was superior to logistic regression methods in that it addressed not only whether the policy will leave but also when it will leave. It also analyzes mid-term cancellation and end-term nonrenewal sequentially and therefore provides a dynamic insight of retention, which improves the static view derived from snapshot data. Another advantage was that survival analysis can take into account time-varying macroeconomic variables, and can help researchers to understand how insurance retention is impacted by the broader economic environment. The superiority of survival analysis in estimating attrition was hence the motivation behind its adoption in this study.

Research Methodology

This empirical study employs survival analysis to study times until policy holders attrite as well as the probabilities of attrition and the covariates that significantly influence attrition. It views occurrences of attrition as an instantaneous change of state, conditioning on being stable prior to time t. In order to describe the likelihood of policyholders' attrition, survival analysis uses two functions namely: survival function $S(t)$ and the hazard function (rate) h(t). In this case, the survival function is the probability that there was no attrition prior to time t.

$$S(t) = Prob(T \geq t) \qquad (1)$$

$$F(t) = 1 - S(t) = \frac{\Delta F(t)}{\Delta t} \qquad (2)$$

Where $F(t)$ is the distribution function. T denotes time to occurrence of attrition. The hazard function (rate) describes the probability of attrition at any given time. Since the hazard is instantaneous, it is modelled as follows.

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t | T > t)}{\Delta t} = \lim_{\Delta t \to 0} \frac{\Pr(t + \Delta t > T > t)}{\Delta t} = \frac{\frac{\delta}{\delta t} F(t)}{S(t)} = \frac{f(t)}{S(t)} \qquad (3)$$

In this case, failure is detected once a policy holder attrites. The most popular survival distributions are the exponential and Weibull. The survival and density functions associated with the exponential distribution are S(t) = e –λt and f(t) = λe −λt , respectively.The hazard function for the exponential distribution is constant, h(t) = λ. The survival and density functions associated with the Weibull distribution are S(t) = e –βtα and f(t) = αβt α−1 e −βtα , respectively. The hazard function of the Weibull distribution is h(t) = αβt α−1 .

Kaplan Meir model

As stated earlier, Kaplan Meir (Product Limit Estimator) is a non-parametric method that has been used to estimate the probability that an insurance policy holder will attrite as well as the duration it takes the policy holder to attrite. The estimator of the survival function S(t) (the probability that attrition time is longer than t is given by;

$$\widehat{S(t)} = \prod_{i:t_i \geq t} (1-\hat{\lambda}) \qquad (4)$$

This originates from the fact that we estimate the hazard within the interval containing event time t_i as:

$$\hat{\lambda} = \frac{d_i}{n_i} \qquad (5)$$

It is possible to show that this estimate arises as a maximum likelihood estimate. The likelihood of the data can be written:

$$\prod_{i=1}^{k} = \lambda_i^{di} (1-\lambda_i)^{n_i - d_i} \qquad (6)$$

effectively this formula is being used for all the other intervals as well, but as $d_i = 0$ in all these intervals, the hazard will be zero .It is possible to show that this estimate arises as a maximum likelihood estimate. The log likelihood is:

$$\ln L = \sum_{i}^{k} \{d_i \ln \lambda_i + (n_i - d_i)\ln(1-\lambda_i)\} \qquad (7)$$

Differentiating with respect to λ_1:

$$\frac{\Delta \ln L}{\Delta \lambda_i} = \frac{d_i}{\lambda_i} - \frac{(n_i - d_i)}{(1-\lambda_i)} \qquad (8)$$

Setting this equal to 0:

$$\frac{d_i}{\lambda_i} = \frac{(n_i - d_i)}{(1-\lambda_i)} \Rightarrow d_i - d_i\lambda_i = n_i\lambda_i - d_i\lambda_i$$

$$\Rightarrow d_i = n_i\lambda_i$$

$$\Rightarrow \lambda_i = \frac{d_i}{n_i} \qquad (9)$$

which gives $\hat{\lambda}_i = \frac{d_i}{n_i}$ and it similarly follows that $\hat{\lambda}_i = \frac{d_i}{n_i}$ for i = 2,3,..k .

Where t_i is the time when at least one policy holder has exited, d_i the number of policyholders who exited by time t_i and n_i the policyholders who have not yet exited up to time t_i. Kaplan Meir curve is a series of stepwise declining graph which illustrates the survival function for population under interval. The observed survival function between successive distinct is assumed to be constant. One major advantage of this curve is that it takes into account right censoring. Several series of curves can be used to visualize the distribution of survival functions between groups of populations e.g. male and females. However, to verify if the groups significantly differ in their survival functions, log rank and Wilcoxon tests are applied.

Logrank test of differences in survival distributions

Two or more survival curves can be compared statistically by testing the null hypothesis i.e. there is no difference regarding survival among two interventions. This null hypothesis is statistically tested by test known as log-rank test and Wilcoxon test. In log-rank test we calculate the expected number of events in each group i.e. E1 and E2 while O1 and O2 are the total number of observed events in each group, respectively. The test statistic is:

$$\log - rankteststatistics = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \qquad (10)$$

The log-rank test is used to test whether the difference between survival times between two groups is statistically different or not, but do not allow to test the effect of the other independent variables. The effect can also be tested by use of Cox Proportions Hazard Model.

Research design

The modelling of the attrition was based on events that have already occurred, this study therefore is a retrospective study. Research studies normally falls under the following categories of designs: Descriptive, Explanatory and Exploratory. Exploratory research design aims at shedding some light on the nature of a given situation. Descriptive research design on the other hand reduces components of a complex matter into smaller comprehensive parts while the explanatory research design tells on the association or causality of a given predictors on a given dependent variables. This study focused on data mining. Data mining is described as the process of getting patterns from given data and making insights from patterns arising from the data. The purpose of data mining was to come up with a valid, useful and ultimately comprehensive patterns in a given data, hence making it a useful tool in this study. Data mining uses the approach of exploration and hence this study adopted the exploratory research design. This research is also quantitative since it aimed at identifying problems through testing theories measured with numbers and analyzing them through statistical techniques. The study is not only deductive but also inductive. It is deductive since it tests hypotheses based on existing theories and inductive as it develops theories on insurance attrition based on observations.

Sampling procedure and sample size

This study made use of the sample provided as secondary data. Hence, the study has ensured that the sample size chosen was manageable while still ensuring high level of precision. With the set sample size, the study was able to get data at an affordable cost in terms of time, finances and human resource. The study adopted Cochran (2003) formula in drawing samples of issued and observed policies for the period 1st January 2018 to 30 th June 2019. The formula applied is;

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \qquad (11)$$

Where,

$$n_0 = \frac{Z^2 pq}{e^2}$$

Z- Z-score, 1.96 for 95% Confidence Interval

P- proportion of population with study attributes

q- 1-p

e - The level of precision or the sampling error.

N- Target population

The total number of policies registered for the study period were 624 (Jubilee, 2019; UAP reports, 2019). Since variability in the population was not known, p was taken to be 0.5 and to ensure the highest accuracy and reliability of the data, 95% confidence interval and 5% level of precision was chosen. The resulting sample size was hence 300 of the total policies.

Results and discussions

The present study used primary data from both UAP and Jubilee insurances. It used a total of 149 policy holders' data; 56 from UAP and 93 from Jubilee insurances who had purchased a policy and could easily be tracked in the period 1st January 2018 to 30th June 2019. This study also aimed at establishing how certain demographic factors e.g. gender, employment, education, age and religion influenced attrition rates.

Gender

From table 1, this study found that there was imbalance in gender of the policyholders. 91 (61.10%) were female while 58 (38.9%) were males. The findings are presented in Table:

Age

This study found that majority, 65 (43.6%) of the policyholders were in the age group 36– 45 years. 58 (38.9%) were in the age bracket of 26 - 35 years and 19 in 46-55 years. Only 4 respondents were in the age group 56 years and above. These results are presented in Table 1.

Table 1. Distribution of policyholders by Age

	Frequency	Percentage
18-25	3	2.0
26-35	58	38.9
36-45	65	43.6
46-55	19	12.7
56+	4	2.7
Total	149	100.0

Education level

From table 2 this study also found that majority (81.21%) of the policy holders had under graduate level of education. There were also 14.09% who held post graduate degrees while only 4.03% were secondary school level policyholders. The few number of policyholders with secondary and primary education can be attributed to the fact that they do not fully comprehend insurance and its importance. These findings were presented in table 2

Table 2. Distribution of policy holders by education

	Frequency	Percentage
Postgraduate	21	14.09
Undergraduate	121	81.21
Secondary	6	4.03
Primary	1	0.67
Total	149	100.0

Marital status

Marital status was suspected to influence attrition rates as the singles are perceived to have few burdens while the married are perceived to be more serious than the singles. From table 3, it has also been observed that those who separate encounter financial difficulties and hence this study investigated the marital status of the policyholders in order to assess whether it impacted attrition rates. This study found that majority (58.39%) of the policies holders were married while 35.57% were unmarried. There were 4.03% who were separated and 2.01% widowed.

Table 3. Marital status

	Frequency	Percent
Single	53	35.57
Married	87	58.39
Separated	6	4.03
Widow/Widower	3	2.01
Total	149	100.0

Time with the insurance

Various studies have established that attrition rates are highly associated with customer lifetime value. Newer customers have the highest risk of attrition while the customers who have stayed for long with the company are likely to remain loyal. From table 4 this study established durations which the policyholders had stayed with their respective companies. The study found that majority (30.2%) had been with the insurance company for 2 – 5 years. Other 22.82% had been with the companies for 6-10 years while only 11.41% had been in the contract for more than 10 years. These findings were presented in table 4.

Table 4. Distribution of Policy holders by Time

	Frequency	Percentage
<6 months	21	14.09
1 year	32	21.48
2-5 years	45	30.20
6-10 years	34	22.82
10+	17	11.41
Total	149	100.00

Type of policy

The type of policy held may determine if a policyholder will attrite or not. For this reason, this study grouped the policyholders into life and non-life policy holders with respect to the policy each had. The study established that majority (54.36%) of the policyholders had a life policy in force. Only 45.64%) had a non-life policy. This can be attributed to the fact that more life products have been designed which are much affordable and match many of the people's needs. These range from education, to savings to medical products.

Premium amount paid

The amount of premium paid may influence whether a policyholder will attrite or not. This study grouped the policyholders with respect to the premiums paid and established that majority (34.23%) paid a premium greater than Ksh.8000. There were 24.83% more who paid premiums ranging from Ksh. 5500 to Ksh 8000, and 21.48% of respondents paid betweem Ksh. 2500 – Ksh. 5000 while only 19.46% paid a premium less than Ksh. 2500.

Repayment method by month

The two insurances offered their policyholders various ways to pay their premiums. This study intended to assess if the payment method used to fund the policy influence attrition rates. From the secondary data obtained, the study established the majority 45.64% of the policyholders used MPESA Pay Bill while 30.87% used the Cheque payments. Other 19.46% of the policyholders used direct debit methods while only 4.03%

policyholders used check-off system. These findings were presented in table 5.

Table 5. Distribution of policyholders by premium amount

	Frequency	Percentage
Direct Debit	29	19.46
Check-off	6	4.03
Cheque payments	46	30.87
MPESA paybill	68	45.64

Employment

The study also established that majority (41.61%) of the policy holders were from the formal sector while 27.52% were from the informal sector. There were 14.09% who were self-employed and other categories accounted for 16.78% of the policyholders.

Inferential analysis

Kaplan Meir curves were used to show the survival distribution of the policyholders. Survival in this context was failure to attrite. Generally, this study observed that attrition rates increased with time. There were fewer policyholders leaving the contract at initial stages. Originally, there were 149 policyholders but after a 100 days elapsed, only 122 policy holders had remained. By the end on 200 days, there were 86 policyholders and by June 2019 only 9 policies remained in force.

From Figure 1, the study also established that there was a higher attrition rate in UAP insurance as compared to Jubilee insurance. These could be attributed to the fact that Jubilee insurance has diversified their products and hence a higher customer retention. Initially, there were 56 policies in force at UAP and 93 in Jubilee insurance. After 100 days, only 40 policies remained at UAP while there were 82 in Jubilee. The number of policies drastically dropped after the 400 days and only 11 policies remained in force at UAP and 14 at Jubilee.

From Figure 2, females were observed to have a higher attrition rates as compared to males. Initially, there were 91 male policyholders and 58 female policyholders. After 100 days, only 77 male policyholders remained while the female policyholders remained at 45. 57 male policyholders remained after 200 days while only 29 females remained. At the end of study duration, only 6, males had remained while the female were 3.

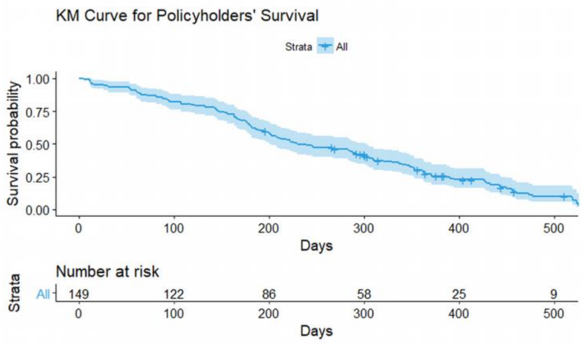

Figure 1. KM curve for policyholders' survival

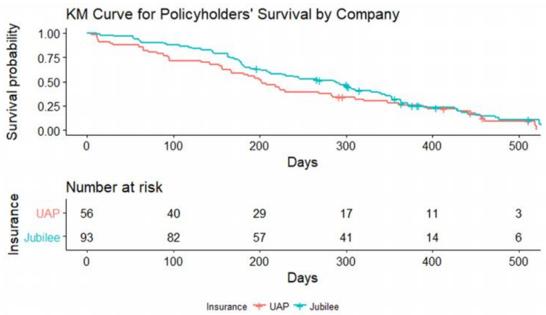

Figure 2. KM Curves for policyholders' survival by company

From Figure 3, this study also deemed it important to assess how policyholder's age impacted attrition. The study observed that the youths (18-25 years) and the elderly (56+ years) were likely to attrite as compared to their counterparts on the other age brackets. This can be attributed to the fact that both the age groups are financially stable.

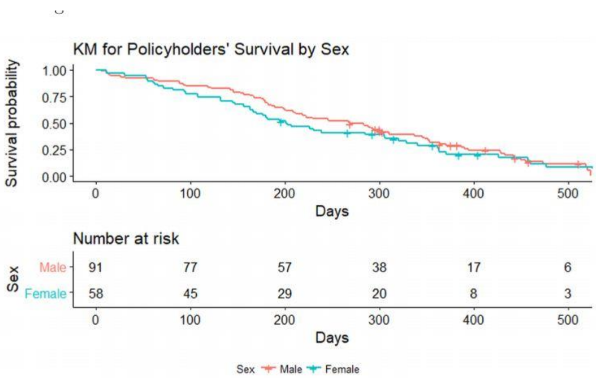

Figure 3. KM Curves fir policyholders' survival by sex

The study also found that policyholder's level of education influenced attrition. It was observed that there were high attrition rates among the primary and secondary school certificate holders. However, this could be attributed to the smaller number of people insured from these education categories. The study also observed a high attrition among the undergraduates.

From Figure 4, the study also observed the duration that the policyholders had stayed with their respective insurances. The study found that those who had lesser years in the contract were likely to attrite as compared to those who had stayed for long. The policyholders who had only one year were the most who dropped followed by those who had less than 6 months in contract.

From Figure 5, this study was able to collect aggregated secondary data on policy type which had been summarized into either life policy or general policy. Life policies included n-year endowment, term insurance, health insurance, whole life insurance among others. General insurance included comprehensive, third party and property insurances. The study presented attrition in both the policies using Kaplan Meir curves. The curves illustrates survival probability and hence the declining trend indicates a declining likelihood of a policyholder remaining in an insurance contract as time changes. The decline implied attrition rates from the insurance rates either by contract non-renewal or failure to pay premiums. This study observed a steep attrition in life policies after the 400 th hundred day to the 500 days. Initially, there were 81 life policies which reduced to 65 after the first 100 days. At the end of the study period, only 5 policies that had been tracked were in force.

Steep attrition in non-life policies after the 365^{th} day was also observed. Initially, there were 47 life policies which reduced to 36 after the first 100 days. At the end of the study period, only 5 policies that had been tracked were in force. This is as a result of most non-life policies which matures after a year, and if a policy holder decides to renew the contract, it will be treated as a different entry. Given that, in this case, the event of interest is the failure to retain the contract after maturity or before due to mid-term cancellation. These findings were presented in the Kaplan Meir curve below. The study also sought to see if the amount of premium charged contributed to attrition. The study found that premiums of ksh 2500-5000 were associated with the highest rates of attrition. Initially, 32 policies in force charged a premium of Ksh 2500-5000. After 100 days, they number declined to 24 and 17 after 200 days. There was also high attrition rates associated with the premium amounts greater than ksh 8000.

From Figure 6, the study also established that direct debit method of premium payments contributed the highest rate of attrition as compared to other methods. There was also high attrition with the check payments and this could be attributed to its low convenience. In as much as there were few policyholders with a check off system, the method had the least attrition rates. These findings are depicted in Figure 8. The study also established that at the initial stages of the observation period, there was a high rate of attrition among the self-employed people. Those in formal employment had the least level of attrition.

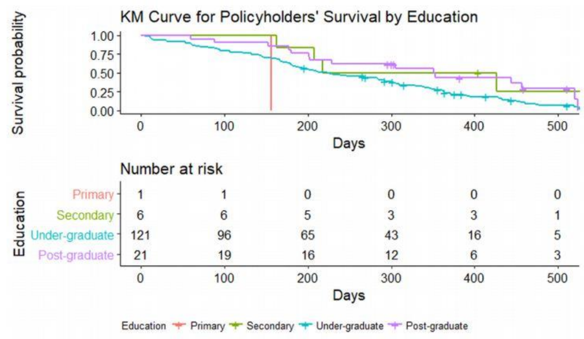

Figure 4. KM curves for policyholders' survival by education

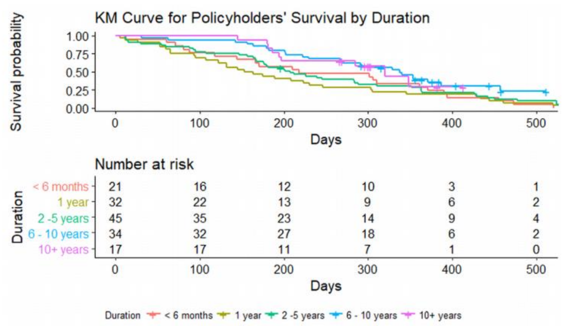

Figure 5. KM curves for policyholders' survival by duration

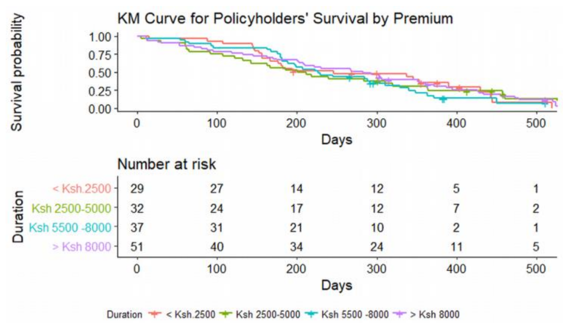

Figure 6. KM curves for policyholders' survival by premium

Conclusions

Log rank test was used to assess whether the observed differences in attrition rates among groups were significant. The study found that there was no significant difference in attrition rates between males and females. However, there was a significant attrition rates between UAP and Jubilee insurances. The study also found insignificant differences in attrition rates between the various age groups. There was insignificant difference in attrition rates between the marital status levels as well as the payment methods. There was also no significant attrition rate between life and non-life products. However, there was a significant attrition rate between the various premium amounts paid and the levels of education.

Conflicts of interest

Authors declare no conflict of interest.

References

[1] Ranaweera C, Neely K. Service Quality and Customer Retention. Int J Anal 2020;2(4):122-33.

[2] Oketch S. Challenges of Customer Retention by Madison Insurance Company Limited in Kenya. Master of Business Administration: University of Nairobi; 2017.

[3] Hasanthika H, Jayasekara F. Analyzing the Customer Attrition using Survival Techniques. International Journal pg Statistics and Probability 2020;6(6):85-90.

[4] Mohammed Z, Mohammed S. Analysis and Estimation of Customer Survival Time in Subscription-based. Business Statistics 2018;2(4) 23-34.

[5] Fu L, Wang H. Estimating Insurance Attrition Using Survival Analysis. Variance 2014;8(1):55-72.

[6] Marin P, Maria A. Survival methods for the analysis of customer lifetime duration in insurance. J Retailing 2019;3(4):155-64.

[7] Oakes D. Survival Analysis. Biometrika Centenary 2001;88:99-142.

[8] Peterson RA, Albaum G, Ridgway NM. Consumers Who Buy from Direct Sales Companies. J Retailing 1989;65:273-86.

[9] Tanser J. Conversion and Retention Modeling. Ratemaking and Product Management Seminar. Casualty Actuarial Society, Chicago; 2010.

[10] Thomas LC, Edelman DB, Crook JN. Credit Scoring and its Applications. SIAM Monographs on Mathematical Modeling and Computation, Philadelphia, SIAM; 2002.

[11] Van den Poel D, Lariviere B. Customer Attrition Analysis for Financial Services using Proportional Hazard Models. European Journal of Operational Research 2004;157(1):196-217.

****

