

Modeling of Average Survival Time for a Loss to Be Handled in Insurance Company

James Akuma Bogonko, George Orwa, Anthony Wanjoya

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

Jamesakuma440@gmail.com (J. A. Bogonko), gorwa@jkuat.ac.ke (G. Orwa), awanjoya@gmail.com (A. Wanjoya)

To cite this article:

James Akuma Bogonko, George Orwa, Anthony Wanjoya. Modeling of Average Survival Time for a Loss to Be Handled in Insurance Company. *American Journal of Mathematical and Computer Modelling*. Vol. 5, No. 1, 2020, pp. 18-21. doi: 10.11648/j.ajmcm.20200501.13

Received: October 28, 2019; **Accepted:** November 21, 2019; **Published:** February 18, 2020

Abstract: Most insurance companies find it hard and hectic to pay claims that had not being anticipated. In order for the companies to be able to make enough reserves to cater for the claims, the average survival time for a claim to occur and then settled in an automobile insurance companies need to be determined. Therefore, the project used survival analysis techniques to analyze this problem. The techniques that were employed include both the product limit estimator and the cox proportional hazard model. The variables that were analyzed in this study were primarily; type of vehicle ownership, type of policy issued, nature of the claim, size of the vehicle and place of residence for the respective customers. The objectives of the study was to compare statistically and graphically the Kaplan Meier survival graphs of different covariate groups and the time a certain vehicle takes for a loss to occur mostly after occurrence of the insured risk and also used a cox-regression to test for their significance. The study used on secondary data that was acquired from one of the insurance Company in Kenya. The review was motor vehicle claims data for 2018 where the information was coded and analyzed using descriptive statistics. The study showed that ownership and residence were significant risk factors that contribute to the occurrence of a loss but they are insignificant in claim settlement using Cox regression model and log rank test. The size of the vehicle and the type of policy given out were significant covariates that influence claim settlement time.

Keywords: AKI - Association of Kenya Insurers, Reserves, Product Limit Estimator, Cox Regression

1. Introduction

Protection is a system by which a company can trade its vulnerability for more noteworthy conviction. Also, this suspiciousness may be, when will a misfortune happen, when will it occur, how dangerous will it be and how many people might be affected in a year, this is what is known as insurance by reference [1]. When the policyholder demand for payment of an occurred loss is referred as a claim. These claims are arranged by accident date that was reported [2]. Motor vehicle insurance in Kenya is well developed helped by its compulsory Department of State according to Business Monitor International. Thus the number of vehicles tends to increase day by day and the need of protection has to be become more articulate because this is the area where majority citizen expect to see health progress [3]. The primary functions in an insurance company are handled by Marketing, Underwriting and Claims divisions. Marketing division involves with the process of identifying clients, distributing products to clients and conducting sales

activities. The underwriting department involves with assessing and classifying the risk represented by the proposed individual and coming up with decision to decline or accept the risk, reference [4]. Claims division ascertains that the sum to be paid as compensation is within the scripted agreement. This claim segment is anticipated to at-least surpass or satisfy the standards of service laid down and to operate within the budget of the insurance [5]. Thus, the claims division is all-important in the success of an insurance company. Reference [6] considered factors that influence how automobile insurance premiums are laid out in the auto insurance companies so as to avoid unexpected claims as by assessing area of operation of the vehicle. He concluded that rates are supposed to be high in geographical areas where number of claims filed is high. Auto insurance accommodates an increased interest because it manages both accidents and the number of insurable vehicles with a broad diverseness of risks. Therefore, with the increased road accidents in Kenya, the auto insurance companies handle a lot of motor vehicle insurance claims from their respective

policy holders. Section 203 Cap 487 (of Insurance Act) requires an insurance company to settle claims within 90 days after financial obligation is established [7]. Therefore, insurance companies have been facing a problem of determining the average survival times to claim when a financial loss has occurred to the insured. Because of this outcome, the researcher modeled the average time for a claim to occur and then settled in the automobile insurance industry.

2. Methodology

The research used administrative data, collected from one of the insurance Company in Kenya in association with AKI. The data contained the number of insured that purchased insurance policy for their automobiles commencing from January, 2018 and expiring on December, 2018. The covariates used in the study of this project was the size of vehicle possessed by the insured, type of insurance policy bought, nature of the claim, ownership and residence. Other covariant used in this study is time; this covariant was characterized as the length of having an insurance policy until a loss occurs and when it was paid.

2.1. Survival Analysis

Survival analysis is an assemblage of statistical procedures for data analysis. Here the outcome variable of concern is time until the occurrence of event [8]. It is the study of time between entry into observation and a succeeding event. The term 'survival analysis' came into being from first studies, where the event of concern was death [9]. The response is often referred to as survival time. Reference [10] described the term 'survival time' as the duration of time taken for failure to take place.

2.2. Describing the Distribution of Time to Event

Let T be the time variable with a continual distribution and taken as a real positive valued variable. The covariant T describes time between the outcomes. The distribution of T (survival time) can be described in the same way as; the cumulative distribution function [11]:

$$F(t) = \text{prob (a claim is made before } t) \quad (1)$$

$$F(t) = p[T \leq t], \quad (2)$$

$$T \geq 0 \text{ which is continuous} \quad (3)$$

$F(t)$ is the probability that a randomly selected from the population will make a claim before time t .

2.3. Survival Function

If T is the failure time, then survival function; $s(t) = p(T > t)$ where $t = [0; \infty)$ and as a probability $s(t) = [0; 1]$:

$$S(t) = p[T > t] = 1 - F(t) \quad (4)$$

$S(0) = 1$ and $s(\infty) = 0$ which means that everyone will finally experience the event if there was no censoring [12].

2.4. Censoring

Censoring occurs where observation is not complete. Observation can be incomplete where time of entry is unknown and this is referred to as Left censoring. The data can also be incomplete where we do not have the information after the end of the observations. This is referred to as Right censoring. Here, the subject is followed until sometime, at which the event has yet to take place, but then takes no further contribution in the study [13]. Interval censoring occurs when the precise time the event occurs is not known exactly, but an interval bounding this time is known. In the data, right censoring was present for some of the motor vehicles because information after end period was unknown for the remaining vehicles to make a claim.

2.5. Kaplan-Meier Estimate

It's non parametric estimator that finds estimates of survival data from none or right censored data [14].

These estimates include the survival estimate ($S(t)$) and cumulative estimate ($F(t)$).

$$S(t) = \prod (1 - \lambda_k) \quad (5)$$

$$\text{And } \hat{F}(t) = 1 - S(t) \quad (6)$$

Where λ_k is the hazard function.

$$\lambda_k = d_k/n_k \quad (7)$$

Where d_k is the number of vehicles that made a claim at time t_k .

And n_k is the number of motor vehicles that are at risk of making a claim before time t_k .

Other denotations used in Kaplan Meier include;

Censored observations before time t_k denoted as C_k .

$$\text{Where } N - M = \sum C_k \quad (8)$$

Where N is the total number of motor vehicles under investigation and

M is the total number of motor vehicles that made a claim during our period of study.

Assumptions of Kaplan Meier model

The assumption here is that the type of censoring present is right censoring.

2.6. Cox-proportional Model

Cox-proportional hazard model describes the effect of each categorical or quantitative variable on survival. In the data, the distribution is unknown. So the cox model used was semi-parametric proportional model under which the hazard function modeled stated [15] is given by;

$$h(t, z_k) = \lambda_0(t) \exp[\beta Z_k] \quad (9)$$

Where $h(t, z_k)$ is the hazard function:

1. β is the regression parameter.

2. Z_k is the covariates and $k = 1, 2, 3, 4$ and 5 .

The force of mortality is proportional to the baseline

hazard $\lambda_0(t)$ at time t .

In the investigation, covariates were as follows;

Z1 = ownership, where

1. If public.

2. If private.

Z2 = size, where

1. If big (a vehicle with more than 4 wheels).

2. If small (4-wheel vehicles).

Z3 = residence, where

1. If urban.

2. If rural.

Z4 = Policy type, where

1. If Comprehensive.

2. If Third party.

Z5 = Nature of the claim

The regression parameters were estimated by maximizing the partial likelihood.

$$L(\beta) = \prod \exp[\beta] / Z_k \sum \exp[\beta Z_k] \quad (10)$$

Solving the equation, got the maximum partial likelihood estimates of β is.

3. Results and Discussion

This chapter contains analysis that was obtained by using different statistical models and procedures described above. The raw data was summarized using descriptive analysis. The main results were obtained by the product limit approach and cox regression model based on 95% confidence interval.

3.1. Results

Table 1. Time from the beginning of an insurance policy until the time of claim occurrence.

| Claim | 75% | | 50% | | 25% | | Mean | | 95% C. I. | |
|---------|--------|------|--------|------|--------|-------|--------|------|-----------|--------|
| | Est. | S. E | Est. | S. E | Est. | S. E | Est. | S. E | L. B | U. B |
| Yes | 269.00 | 5.58 | 205.00 | 7.28 | 42.00 | 6.51 | 178.24 | 4.43 | 169.73 | 187.74 |
| Overall | 365.00 | 0.00 | 269.00 | 7.10 | 124.00 | 15.53 | 245.49 | 4.02 | 238.46 | 253.40 |

Table 2. Time from the beginning of claim report date to time of payment.

| Claim | 75% | | 50% | | 25% | | Mean | | 95% C. I. | |
|---------|--------|-------|--------|------|--------|-------|--------|------|-----------|--------|
| | Est. | S. E | Est. | S. E | Est. | S. E | Est. | S. E | L. B | U. B |
| Yes | 344.00 | 10.92 | 269.00 | 9.13 | 124.00 | 20.98 | 254.87 | 6.33 | 242.80 | 268.12 |
| Overall | 344.00 | 10.92 | 269.00 | 9.13 | 124.00 | 20.98 | 254.87 | 6.33 | 242.80 | 268.12 |

Table 3. Time from the beginning of claim report date to time of payment for the policy type given out.

| Policy | 75% | | 50% | | 25% | | Mean | | 95% C. I. | |
|---------|--------|-------|--------|-------|--------|-------|--------|------|-----------|--------|
| | Est. | S. E | Est. | S. E | Est. | S. E | Est. | S. E | L. B | U. B |
| Comp | 399.00 | 15.68 | 304.00 | 10.78 | 203.00 | 19.89 | 279.26 | 8.11 | 264.75 | 297.37 |
| T. p. | 329.00 | 8.28 | 225.00 | 14.25 | 44.00 | 14.25 | 220.54 | 9.45 | 199.73 | 238.71 |
| Overall | 344.00 | 10.92 | 269.00 | 9.13 | 124.00 | 20.98 | 254.87 | 6.33 | 242.80 | 268.12 |

Est. = Estimate.

S. E. = Standard Error.

C. I. = Confidence Interval.

Comp = Comprehensive.

T. P. = Third Party.

Table 4. Analyzed hazard ratios for cox-regression.

| Risk factor | Coeff (β) | Pr ($> z $) | Hazard Ratios | 95% c. i. for h. r |
|-----------------------|-------------------|---------------|---------------|--------------------|
| Residence | -0.0413 | 0.4715 | 0.9548 | 0.8562-1.0690 |
| Ownership | 0.0175 | 0.9121 | 1.01416 | 0.8125-1.2861 |
| Type of policy | 0.4850 | 4.5 e-06 *** | 1.5690 | 1.3864-2.1140 |
| Nature of claim | -0.0474 | 0.0765 | 0.9356 | 0.8841-1.0990 |
| Size of motor vehicle | 0.0794 | 0.0101 * | 1.0916 | 1.0314-1.1742 |

C. I = Confidence Interval.

Coeff = Coefficient.

H. R. = Hazard Ratio.

There is 56.9% increase in the expected hazard relative to 1 year increase in the type of policy holding size of vehicle fixed. Equivalently, the expected hazard ratio is 1.0916 times higher for the size of the vehicle, holding type of policy fixed. The 95% confidence intervals for significant risk factors do not include null value while for non-significant

risk factors (nature of claim, residence and ownership) include the null value.

3.2. Discussion

The preliminary analysis suggested that; According to the policy, comprehensive policy where mostly bought compared

to third party policy. Also, according to the residence, motor vehicles operating in the citified areas were at higher risk compared to rural residence thus, survival of vehicle operating in rural areas was higher than those operating in the urban areas. This majorly because in urban areas there was more traffic compared to the rural areas hence vehicles were more prone to accidents in urban areas than in rural areas. Regarding to the size of the vehicle, most small vehicles involved in accidents compared to big sized motor vehicle resulting to higher survival of big vehicles than that of small vehicles hence more premiums should be charged on small vehicles than big vehicles. In ownership, private motor vehicles were largely involved in accident compared to Public motor vehicles. Generally, public vehicles had a higher survival estimates compared to private vehicles and thus they should be charged lower premiums than the private vehicles.

4. Conclusion and Recommendations

4.1. Conclusion

The study found out that the average time for a loss to occur is within 209 operative days and the estimate for buying a motor vehicle insurance without a loss occurring is 365 days. The average time to settlement of motor vehicle insurance claims is within 269 operative days. The average time to settlement of motor vehicle third party policy is within 225 operative days whereas the average time for comprehensive policy claims to be settled is within 304 operative days. The variables (ownership, nature of claim and residence) were insignificant risk factors that affect settlement of a claim. Type of policy given out and Size of the vehicle involved in accident were highly significant variables that influenced the time of claims settlement.

4.2. Recommendations

From the conclusions above, it is recommended that the IRA should ensure that more study on the average time to claim governance in the motor vehicle insurance industry is done especially, in other insurance companies in Kenya so that suitable control measures could be implemented to control its failure. Also actuaries should come up with premium software in which claims processed from different insurance companies can be examined by the IRA to prevent delayed claims settlement.

References

- [1] Atkins, D. and Bates (2004). Risk, regulation and capital adequacy. The Chartered Insurance Institute Course book.
- [2] McClenahan, C. L. (2001). Ratemaking. Foundations of Causality Actuarial Science.
- [3] International, B. M. (2014). Kenya insurance report. Business Monitor International.
- [4] Brown, J. (1997). Insurance administration, life management institute Loma. Atlanta, Georgia
- [5] Wedge, P. and Handley, D. (2003). Claims management study. The Chartered Insurance Institute, Great Britain.
- [6] Kiebach, A. (2014). Five factors that affect car insurance rate. Lancaster Red Rose Credit Union.
- [7] AKI (2015). Motor insurance claims. Insurance guidebook, pages 1–66.
- [8] Czado, C. and Rudolph, F. (2002). Application of survival analysis methods to long term care insurance. Center of Mathematical Sciences, Munich University of Technology, Germany.
- [9] Ritesh, S. and Keshab, M. (2011). Survival analysis in clinical trials. National center for biotechnology information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA.
- [10] Stevenson, M. (2009). An introduction to survival analysis. PhD thesis, Epicenter, IVABS, Massey University.
- [11] Wilesmith, J., Stevenson, M., King, C., & Morris, R. (2003). Spatio-temporal epidemiology of foot and mouth disease in two counties of Great Britain in 2001. Preventive Veterinary Medicine, 61 (3), 99- 170.
- [12] Tsiatis, A. and Zhang, D. (2005). Analysis of survival data. Statistics, North Carolina State University, Department of Statistics.
- [13] Dohoo, I., Martin, S., & Stryhn, H. (2003). Veterinary Epidemiologic Research. Charlottetown, Prince Edward Island, Canada: AVC Inc.
- [14] Kaplan, E. and Meier, P. (1958). Non-parametric estimation from incomplete observations. American statistical Association, 53, (282): 457-481.
- [15] Cox, D. and Oakes, D. (1984). Analysis of Survival Data: Monographs on Statistics and Applied Probability. London, New York. Chapman and Hall.