Analysis of Factors Influencing Agent's Perspective towards the Lapsation of Life Insurance Policies

Fernando MSD[1]
Dharmawardane PMN[2]
Napagoda NADN[2]

ABSTRACT

Life insurance is a contract between people and a life insurance company, which provides death benefits to the dependents of the policyholder, when a death is occurred during the contract period. If the policyholder survives till the end of the period, he/she will claim maturity benefits. The policyholder will not be eligible to claim benefits unless he/she has been paying the due premiums properly. When it occurs, it is called a policy lapse. Lapses are created when the policyholder ignores to pay due premiums at due time. Customers seem to be the cause for the policy lapses. However, it is not the only reason. There is another aspect that has to be considered. Agents in the branches of life insurance involve in seeking customers who are in need of life insurance policies. Some of them, for some reason may not be able to find the suitable people and this may be the root cause for lapses. Therefore in this paper, it is expected to find, due to what reasons and qualities of agents, lapses are created. Moreover, to build a multiple regression model for predicting number of lapses in a year according to some pre-determined factors is expected from this. Here we choose number of policies being lapsed in a year, since it has been revealed that more than 75% of policy lapses are first year lapses. Sales agents are always reminded to minimize the first year policy lapses as much as possible. Moreover, it is investigated to find reasons when such lapses occur.

KEYWORDS: Life Insurance, Multiple Regression, Policy Lapses, Sales Agents

INTRODUCTION

Life insurance always comes up with the term, life insurance policy. A life insurance policy is a contract with an insurance company.

In exchange for premiums (payments), the insurance company provides a lump-sum payment, known as a death benefit, to beneficiaries in the event of the insured's death.

According to Fernando & Napagoda (2014), XYZ* life insurance is one of the leading companies in the field of insurance in Sri Lanka. During the short period of time, it has achieved so many goals. Currently it has the highest market share of 28% and the life fund is close to 60 billion rupees. While the customers' satisfaction is rewarded enough, their professionalism, innovation, product development and customer service has been recognized and awarded repeatedly over the years both locally and internationally.

RESEARCH OBJECTIVE

The main objective of this research is to identify the factors of agents, which affect to increase in policy lapses and to build a model to predict the lapse percentage in a year according to identified factors. (Lapses are according to the agents' perspective).

[1]Graduate, Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka

[2]Senior Lecturer, Department of Mathematical Sciences, Faculty of Applied Sciences, Wayamba University of Sri Lanka

XYZ* is not the real company's name

LITERATURE REVIEW

In recent years, there have been several studies concerning the life insurance and other related factors. According to Fier & Liebenberg (2013), the life insurance policy lapses are detrimental to issuing insurers when lapses substantially deviate from insurer expectations. Further, Fier & Liebenberg (2013) found that the age is an important moderating factor in the lapse decision. Changes in income appear to more directly affect the decision to lapse for younger households, while they are generally unrelated to the lapse decision for older households.

The decision to lapse a life insurance policy can have far-reaching effects on the issuing insurance company. From the insurer's perspective, excessive policy lapse activity adversely impacts costs, investment returns and mortality experience, each of which negatively affects the financial stability and wellbeing of the insurer (Kenneth & Skipper, 2000).

According to Fang & Kung (1990), the life insurance market is large and important. Policyholders purchase life insurance to protect their dependents against financial hardship when the insured person, *i.e.* the policyholder, dies.

There are two main types of individual life insurance products; Term Life Insurance and Whole Life Insurance. A Term Life Insurance policy covers a person for a specific duration at a fixed or variable premium for each year. If the person dies during the coverage period, the life insurance company pays the face amount of the policy to his/her beneficiaries, provided that the premium payment has never lapsed. The most popular type of term life insurance has a fixed premium during the coverage period and is called the Level Term Life Insurance. A Whole Life Insurance policy, on the other hand, covers a person's entire life, usually at a fixed premium.

Lapsation is an important phenomenon in life insurance markets. Both Life Insurance Marketing and Research Association (LIMRA) and Society of Actuaries consider that a policy lapses if its premium is not paid by the end of a specified time (often called the grace period). According to the LIMRA, International, the life insurance industry calculates the annualized lapsation rate as follows:

$$\text{Annualized Policy Lapse Rate} = \frac{100 \times \text{Number of Policies Lapased during the Year}}{\text{Number of Policies exposed to Lapse during the Year}}$$

The number of policies exposed to lapse is based on the length of time that the policy is exposed to the risk of lapsation during the year. Termination of policies due to death, maturity, and conversion are not included in the number of policies lapsing and contribute to the exposure for only the fraction of the policy year they were in force. The lapsation rates of individual life insurance policies can be calculated using the above formula.

METHODOLOGY

In this research, it was expected to build a multiple regression model which predicts first year lapsed policies according to the insurance agent's status. In this model, the response variable is number of lapsed policies and the predictor variables are education, experience, working hours etc. The regression model is,

$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \alpha$; where x_i, (i =1,2,3,4) are the independent variables and α is the random error, β_is are the unstandardized regression coefficients. The significance of these variables was tested according to the following hypothesis at 5% significance level.

$H_0: \beta j = 0$; where j = 1,2,3,4
$H_1: \beta j \neq 0$;

The Pearson correlation coefficient was used to analyze the correlation between predictor and response using following hypothesis.

H_0: There is no relationship between the two variables.

H_1: There is a relationship between two variables.

The Spearman's Rank correlation was used to find the correlation of ordinal variables under following hypothesis.

H_0: There is no association between the two variables.

H_1: There is an association between two variables.

DATA COLLECTION AND ANALYSIS

Secondary data from 2013 to 2014 were collected for this study from 74 agents at 9 branches of XYZ Life Insurance at Gampaha, Kurunegala and Puttlam districts. The lapsed policy reports for the above branches describe the due reasons.

Table 1: Test for Correlation

Relationship between	P value	Result
Lapsed % vs. Experience of the Agent	0.003	Reject H_0
Lapsed % vs. Education	0.000	Reject H_0
Lapsed% vs. Age	0.000	Reject H_0
Lapsed% vs. Working hours	0.000	Reject H_0

Here the association of agent's lapsed policy percentage and education level is the only one that is measured under Spearman's Rank correlation. Since it is an ordinal variable its hypothesis is tested under Spearman's Rank correlation. The other three variables are measured under Pearson's correlation coefficient.

Since all p values are less than 0.05, null hypothesis is rejected at 5% significance level; all these variables are correlated with the response variable.

Initially, the assumptions of multiple regression are checked before fitting the model.

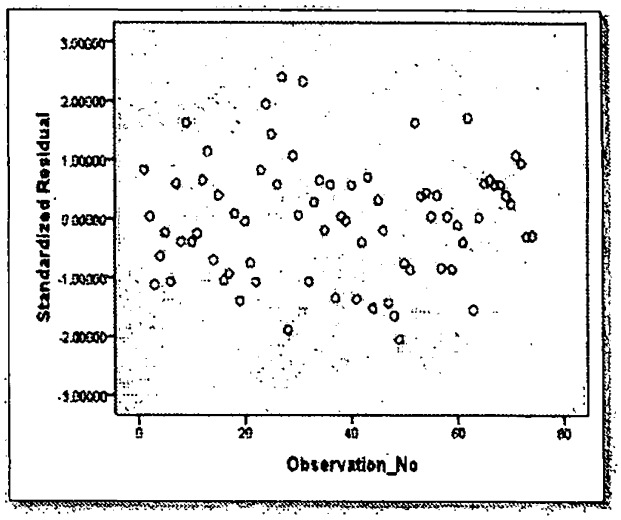

Figure 1: Scatter Plot for Standardized Residuals

According to Figure 1, more than 95% of data points are within 2 and -2 ranges. From statistical perspective, it is accepted more than 95% data within 2 and -2 range is a good data distribution that can be used to make predictions for the population. Furthermore, the data have followed randomness.

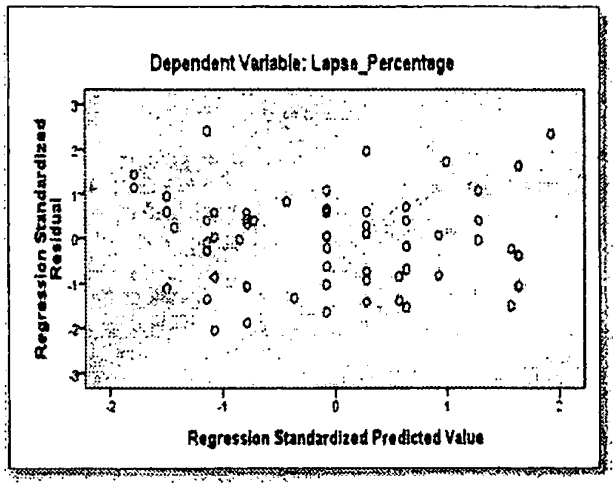

Figure 2: Scatter Plot for Homoscedasticity

Figure 2 depicts the data has been spread in a vast area and no patterns were identified. Therefore, the variance of the data set is homoscedasticity.

The normality of the errors are checked to test the following hypothesis.

H_0: Errors are normally distributed.

H_1: Errors are not normally distributed.

Table 2: Test for Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Standardized Residual	.065	74	.200*	.986	74	.574

At 5% significance level, null hypothesis does not rejected and it can be concluded errors are normally distributed. (Since p values are greater than 0.05)

Table 3: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.837	.700	.696	6.62753	
2	.914	.835	.831	4.94633	1.993

As shown in Table 3, the satisfied combination can be taken from these predictor variables. 83.5% variance of dependent variable is explained by the multiple regression model.

Significance of the model is checked at 5% significance level under the following hypothesis.

H_0: Model is not significant

H_1: Model is significant

Table 4: Anova

Model		SS	Df	MS	F	Sig.
1	Regression	7389.921	1	7389.921	168.243	.000[a]
	Residual	3162.542	72	43.924		
	Total	10552.464	73			
2	Regression	8815.364	2	4407.682	180.154	.000[b]
	Residual	1737.100	71	24.466		
	Total	10552.464	73			

Considering Table 4, model is significant at 5% significance level because p values are less than 0.05.

Significance of the coefficients can be tested under following hypothesis.

H_0: Coefficients are not significant

H_1: Coefficients are significant

Table 5: Unstandardized, Standardized Coefficients Collinearity Statistics for each Model

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	T	Sig.	Tolerance	VIF
1	(Constant)	54.238	1.742		31.137	.000		
	Experience	-4.998	.385	-.837	-12.971	.000	1.000	1.000
2	(Constant)	70.029	2.443		28.661	.000		
	Experience	-3.905	.321	-.654	-12.154	.000	.801	1.248
	Education	-7.093	.929	-.411	-7.633	.000	.801	1.248

According to Table 5, coefficients are significant, because p value is less than 0.05. In addition to that their tolerance values are also greater than 0.2. Therefore there is no multicollinearity between those variables.

After all the assumptions are completed, a regression model is built predicting lapsed percentage.

Lapsed Policy Percentage in the first year of the agent = 70.029 - (3.905 * Education Level of the Agent) – (7.093 * Experience in the industry in years of the Agent)

Equation 1: Regression Model

RESULTS AND DISCUSSION

The study was focused to find out the factors of agents affect to decide the lapsed policy percentage of agents' works at XYZ Life Insurance. The data were collected for one year. They were needed for the first year lapsed policies.

Before analyzing the data, they were checked for the assumptions needed to be satisfied under multiple regression. Initially it was checked for outliers and randomness by plotting standardized residuals against each observation. There were more than 95% of data points between 2 and -2 ranges,

and no data points beyond 3 and -3, so no outliers and there was randomness in the data. Then it was tested for the normality of the errors. Errors were normally distributed according to the Shapiro-Wilk and Kolmogorov-Smirnov tests. Then it was checked for the multicollinearity using collinearity diagnostics. Tolerance values were greater than 0.2. This is implied that multicollinearity doesn't exist. Homoscedasticity was also checked using a constant variance graph with no order/pattern in the data points.

Having checked assumptions for the multiple regression, analysis part was initiated. First it was checked for the coefficient of determination. The second model in Table 5 is the best suited and it can be suggested that 83.5% variance of dependent variable is explained by this model. Then it was checked for the significance of the model using Fisher's (F) test under null hypothesis. Model was significant at 5% significance level. Later it was checked for the significance of the coefficients in the same manner and they too were significant.

Having completed all the analysis part, regression equation was built. The fitted equation is given by Equation 1.

Lapsed policy percentage is negatively correlated with education and experience. Therefore it can be reduced by increasing education level and experience of the agent in the industry. Education level is a categorical variable where level 1 denotes the lowest rank and level 4 denotes the highest rank.

Table 6: Rank Table for Education

Education Level	Rank
Below O/L	1
O/L	2
A/L	3
Diploma or higher	4

The regression model shows that for those with low education level have high percentage of lapse policy percentage and with education level is increased, lapse policy percentage is decreased.

By recruiting much educated agents or giving them (for those with low education level) enough knowledge and education about related fields, lapsed policy percentage can be reduced.

According to the Equation 1, experience has more powerful force than education. The number of lapses occur in a year can be found by multiplying lapsed policy percentage value with total number of policies done in a year.

CONCLUSION

Multiple regression model implies that lapsed policy percentage of "XYZ Life Insurance Company" is decided by the experience of the agent and education level of the agent. They are two more crucial parts that the agent needs to have to show high quality performance from them. The other factors are functions of these ones.

The age of an agent is a variable that everybody thinks that it relates with lapsed policy percentage and then it needs to be fitted in the model. Age is clearly related with experience of the agent (positively related). Because in general agents are recruited at their early 20's or before. Therefore when they reach their 30's, they have completed a service period of 10 years. But there are some situations where some agents are recruited at their 40's or above. A person recruited at that age should have a sound knowledge related to the corresponding field. Therefore most of them are supposed to do better in the field. However the lack of human resource in remote areas, these criteria is partially obeyed. In that case age of the agent may influence the lapse model.

Moreover, for larger data set, covering a whole region, there can be a slight influence from age of the agent towards lapse policy percentage.

REFERENCES

- Fang, H., & Kung, E. (2012) Why Do Life Insurance Policyholders Lapse?: The Roles of Income, Health and Bequest Motive Shocks, Working Paper, Princeton University
- Fernando, M.S.D. & Napagoda, N.A.D.N. (2014). Study on the Major factors Affect the Premium of Insurance Cover of XYZ Life Insurance, STAT 4†24 Project Report submitted for B.Sc. (Joint) Major Degree, Wayamba University of Sri Lanka.
- Fier, S. G. and Liebenberg, A. P. (2013). Life Insurance Lapse Behavior, *North American Actuarial Journal*, 17(2), 153-167.
- Kenneth, B. (Jr.), & Skipper, H.D. (Jr.) (2000). *Life and Health Insurance.* Prentice-Hall, Inc.