Research article

Logistic regression model This is used for analyzing the retirement of life insurance policies Logistic Regression Model for Lapse Analysis of Life Insurance Policy

Arin Boonmeeka * , Winai Phosuwan and Thidaporn Subhaphakorn

Arin Boonmeekham * , Winai Bodhisuwan and Thidaporn Supapakorn

Department of Statistics, Faculty of Science Kasetsart University

Statistics Department, Faculty of Science, Kasetsart University

Received: 28 December 2018

Revised: 21 February 2019

Accepted: 9 May 2019

Abstract

Objective of this research is to create a logistic regression model for forecasting policy retirement.
Life insurance with a dependent variable is policy termination, which is a bilateral qualitative variable. And there is an independent v
There were 18 qualitative or quantitative variables, divided into 2 sets of data to be used to construct the model.
The forecast equation was 1,864 policies and the test data set of 466 policies.
It consists of a total of 6 independent variables: age, amount of life insurance between 50,001 - 100,000 baht, period of time.
More than 3 years of premium payments. Square root of the coverage period. Occupation Class 3 Class and Occupation Class 4
From this equation, it can predict that the insured is 31.76% of policy termination and is effective in
The forecast was accurate to 66.95 percent.

Keywords: policy termination, life insurance, logistic regression, psychoactive response function

*** Corresponding author. E-mail:** arin.knotztu.1412@gmail.com

Research article

Abstract

The purpose of this research is to create the logistic regression model for forecasting life insurance lapse. policy. The dependent variable is lapse class that is dichotomous qualitative variable. There are 18 independent variables which are qualitative or quantitative. The data set is divided into 2 parts; training set of 1,864 policies for building the predicting equation and testing set of 466 policies. The results show that the predicting equation consist of 6 independent variables which are age, face amount (between 50,001 - 100,000), duration of payment (more than 3 years), square root of duration of protection and occupation class (class 3 and class 4). The logistic regression equation predict that the lapse rate of insured is 31.76% and the accuracy performance of forecasting is 66.95%.

Keywords: policy lapse, life insurance, logistic regression, logit response function

Introduction

Insurance business The insurance is divided into two types. Including non-life insurance And life insurance from
Forecasting the growth trend of the life insurance business That will continue to have a continuous growth rate of approximately 4-6
This is a result of the expectation of the domestic economy to expand approximately 3.6-4.6 percent from the world economic factor
Expansion trend And domestic driving force from the export sector Travel Public and private investment
As well as private consumption As a result, the economy has better circulation, resulting in income distribution. People have purchas
More and more awareness The importance of financial planning and risk management of the future (Thai Life)
Assurance Association, 2018)

At present, Thailand has divided the ordinary life insurance (Ordinary insurance) into 4 types:
Whole life insurance, term life insurance, insurance
Annuity life insurance and Endowment life insurance (Office of
Insurance Commission, 2016), where each life insurance product has different coverage periods and benefits.
This makes the profitability of each type of life insurance product different. During the first period of insurance policy issuance, the i
Life will be burdened with higher costs than The amount of premium received such as pension (Commission), which some policies
The pension value is as high as 40% and decreases in the following year. Operating expenses (Administrative expense) and
However, sometimes the insured receives a life insurance policy that does not meet the needs of the policyholder.
Such as the policy conditions do not meet the agreed upon request for insurance or other reasons that the insured.
Need to cancel the policy Which policy termination (lapse) of the insured arises from the insured
Cancel the policy by expropriating the policy before the policy expires, so if the life insurance company
The policy expropriation rate is high during the first year, causing the insurance company to have insufficient income to cover the ex
And causing the life insurance company to lose profit in the business industry today is more competitive and complex.

Therefore, forecasting technique is one of the most important techniques. It is important to the development and affects the progress
The need for forecasting friends And the results of various business decisions. The regression model
Logistic regression model was used in forecasting policy termination (Zian *et al*. , 2016).
They are suitable for logistic regression models, ie gender, policy model. Premium Form of payment
Insurance coverage period GDP, unemployment, financial crisis Total interest rate
To the calendar year, etc.

This is because the retirement of a life insurance policy is the main reason The price of life insurance products is also availabl
Direct impact on the duration of the payment. Premium payment and results The profit of the life insurance company Therefore, the
Model for forecasting termination of life insurance policy using logistic regression model.

Logistic regression analysis

Logistic regression analysis is a qualitative statistical analysis technique.

The objective is to study the relationship between dependent variables. Which is a qualitative variable that has at least two values an
Which could be every quantitative variable Or every qualitative variable Or there are both quantitative and qualitative variables.
Chances of occurrence of an event of interest Logistic regression models can be divided into two categories (Sinsomboonthong,
Binary logistic regression (Binary logistic regression) with a *Y* variable was a qualitative variable.
Only 2 values (Dichotomous variable) 2) Multinomial logistic regression variable according to *Y*
Is a qualitative variable with 2 or more values while the independent variable *X* is either a quantitative or qualitative variable or both.
Quantitative and qualitative

Bilateral logistical regression

Binary logistic regression is the study of the relationship of dependent variables. Which is a qualitative variable with only 2 v
Assigned that the variables *Y* and *X* were correlated under the binary logistic regression model, where *Y* was the Bernoulli distributio
(Bernoulli Distribution) and has a number of *p* independent variables , the probability of occurrence of an event of interest is π as foll

$$Y = \begin{cases} 1; \text{ When an event of interest occurs With probability } \pi \\ 0; \text{ When the event of interest does not occur With probability } 1-\pi \end{cases} \quad (1)$$

Therefore, P (an event of interest occurred)

$$= P(Y = 1)$$
$$= E(Y)$$
$$= \pi$$
$$= \frac{e^{(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}}$$
$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}} \quad (2)$$

Equation (2) found that the relationship between the chance of an event of interest and the independent variable was not linear
Therefore adjusted to be in a linear form (Kleinbaum, 2002) by requiring

$$\text{Odds Ratio} = OR = \frac{P(Y = 1)}{P(Y = 0)} = \frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p} \quad (3)$$

If *OR* is greater than 1, it indicates a greater likelihood of an event of interest than that of a non-event of interest.
Equation (3) for the Logit response function (*Logit*), it was found that *Logit* was in a linear form.

(4)

$$\begin{aligned} Logit &= \log_e (OR) \\ &= \ln (OR) \\ &= \ln (e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}) \\ &= \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \end{aligned} \quad (4)$$

Basic assumptions for binary logistic regression

The binary logistic regression analysis had a total of two basic assumptions (Hilbe, 2019; Tabachnick, 2013):

1) Continuous independent variables have a linear relationship with the psychoactive response function.

2) Independent variables do not have multiple linear relationships. (Multicollinearity)

Method of conducting research

Operations in this research Can write in the form of a flow chart as in Figure 1.

Prepare data

Create a logistic regression model

Select the appropriate independent variables.

Examine the initial assumptions.

Check for suitability and
Accuracy of the subject

Not suitable

appropriate

Forecast of new units

Figure 1 Chart of research work

Data preparation

Prepare the data for analysis Using data from life insurance companies for the year 2017, totaling 2,330 policies with variable Total 18 divider as

1) Two quantitative variables are age (X_1) and period of coverage (X_4)

2) 16 qualitative variables, ie the sum of life insurance ($X_{2A} - X_{2B}$) Payment period Insurance (X_3) Type of payment for insurance premium ($X_5 - X_{5B}$) Occupation level ($X_{6A} - X_{6D}$) Gender (X_7) pawn

Insurance (X_{10A}) Distribution channels and (X_{10B}) policy format Details are as shown in Appendix 1 Table
In which the qualitative variables, the researcher has Set the value as a dummy variable with each possible value 0 and 1.
That is, if that qualitative variable is divided into k groups, the robot variable must be assigned a number of $k-1$.
Example of amount
Insurance is divided into 3 groups: group 1 is less than or equal to 50,000 baht, group 2 50,001-100,000 baht and group 3.

More than 100,000 baht, so the researcher created 2 robot variables, Both X_{2A} and X_{2B} X_{2A} and X_{2B} is only possible value
0 and 1 only, assigned to group 1 Less than or equal to 50,000 baht as a reference group

if	X_{2A} and X_{2B}	0 0	Replace the amount of life insurance less than or equal to 50,000 baht.
	X_{2A} and X_{2B}	1 0	Replace the amount of life insurance 50,001 - 100,000 baht.
and	X_{2A} and X_{2B}	0 1	Replace the sum of more than 100,000 baht for life insurance.

By using all of the data to create a regression model and using that regression model to predict the original data set, it might n
The proportions of the forecasts were accurate, high and gave the students the confidence to apply them. The model was used to fore
When it comes to forecasting a new unit, it may not be correct. Therefore, it should be divided into 2 parts:

1) Training data is the data Let's create a forecast model by giving m_1 represents the amount of information.
practice

2) Testing data is the data It is used to examine or test by using the model from the data.
The apprentice comes to forecast the new units in this test data by giving m_2 represents the number of test data and n represents the total number of data.
Herein $n = m_1 + m_2$ Where in this research the ratio between the practice data set and the test data is equal to 80:20, that is
$m_1 = 1,864$ And $m_2 = 466$

Building a logistic regression model

Once the data is prepared, put those variables into the logistic regression model. Will get a response function
The logic of the full model is shown in Equation (5).

$$Logit = \beta_0 + \beta_{1A}X_{1A} + \beta_{2A}X_{2A} + \beta_{2B}X_{2B} + \beta_{3A}X_{3A} + \beta_4X_4 + \beta_{5A}X_{5A} + \beta_{5B}X_{5B} + \beta_{5C}X_{5C} + \beta_{5D}X_{5D} + \beta_{6A}X_{6A} + \beta_{6B}X_{6B} + \beta_{6C}X_{6C} + \beta_{6D}X_{6D} + \beta_7X_7 + \beta_8X_8 + \beta_9X_9 + \beta_{10A}X_{10A} + \beta_{10B}X_{10B} + \varepsilon \quad (5)$$

Selection of suitable independent variables

A good psychoactive function equation should contain appropriate independent variables, not under-fit or over-fit.
To make the forecast value of the probability that the event of interest is as close as possible to the actual value. Which determines th
What influences the occurrence of an event of interest, there are several ways. In this research, backward stepwise method was select
This research examines the selection of independent variables from the Akaike Information Criteria (AIC).
With the independent variable that gives the lowest AIC value That is, the model tends to give a predictive value close to the true val
Selection of the independent variable starts with creating the full character. After that will 1 independent variable from the regressior
Table 1 found that 10 independent variables were the form of annual and semi-annual premium payments. Premiums are greater than
10,000 baht Policy form, period and accumulation of assets, female Amount

Page 7

More than 100,000 baht of life insurance is taken and the occupation class, occupation class 1 and occupation class 2 are removed fr
In response to the logic, the regression coefficient of the remaining independent variables was estimated. Details are shown in Table
Responding to the logic after selecting the appropriate independent variable as Equation (6)

$$Logit = -16.9881 - 0.0182 X_1 - 0.3465 X_{2A} - 0.9448 X_3 - 0.0643 X_4 + 18.1759 X_{5A} + 17.9908 X_{5B} + 0.4202 X_{6C} + 0.7785 X_{6D} \quad (6)$$

Table 1 Procedures for Selection of Independent Variables

Step	Operation	Degrees of freedom	AIC
1	Full body		1,937.57
2	Field X_{5D}	Out of the psychoactive response function	1,935.57
3	Field X_{5C}	Out of the psychoactive response function	1,933.57
4	Field X_8	Out of the psychoactive response function	1,931.57
5	Field X_9	Out of the psychoactive response function	1,929.57
6	Na $_{10}X$	Out of the psychoactive response function	1,927.66
7	Field X_7	Out of the psychoactive response function	1,925.75
8	Field X_{2B}	Out of the psychoactive response function	1,923.84
9	Field X_{6B}	Out of the psychoactive response function	1,921.99
10	Field X_{6A}	Out of the psychoactive response function	1,920.21
11	Na $_{10}X$	Out of the psychoactive response function	1,918.92

Table 2. The coefficient of logistic regression when there are 8 independent variables (1).

Independent variable	β	*SE*	*Z*	Phi value
Constant	-16.9881	346.6896	-0.05	0.96
X_1	-0.0182	0.0056	-3.22	0.00
X_{2A}	-0.3465	0.1541	-2.25	0.02
X_3	-0.9448	0.1634	-5.78	0.00
X_4	-0.0643	0.2248	-2.86	0.00
X_{5A}	18.1759	346.6895	0.05	0.96
X_{5B}	17.9908	346.6897	0.05	0.96
X_{6C}	0.4202	0.2217	1.90	0.06
X_{6D}	0.7785	0.5050	1.54	0.12

Page 8

Examination of the initial assumptions of logistic regression analysis.

1) Testing of the linear relationship of continuous independent variables with the psychoactive response function. Take the tes
Box-Tidwell (Hilbe, 2009) is achieved by enhancing the interaction term (Interaction regression model too
The independent variable and its interaction terms of that independent variable, or $Logit = \beta_0 + \beta_1 X_i + \beta_2 [X_i \ln(X_i)]$ If diction
The interaction was statistically significant and showed that the independent variable had no linear relationship with the psychoactive
The logistic regression model had two continuous independent variables: age and coverage period.

Table 3 Test of the Linear Relationship of the Age Variable and Duration of Coverage

The test	β	*SE*	*Z*	Phi value
age				
- constants	2.4741	1.1936	2.07	0.04
- X_1	-0.2088	0.1487	-1.40	0.16
- $ln(X_1) \times X_1$	0.0352	0.0320	1.10	0.27
Duration of coverage				
- constants	1.0898	0.2001	5.45	0.00
- X_4	-0.5479	0.0834	-6.57	0.00
- $ln(X_4) \times X_4$	0.1221	0.0250	4.89	0.00

From Table 3, it was found that the age interaction was more significant than the p-value (). Therefore, there is no statistical significance, that is, there is a linear relationship between the independent variable and the psychoac The coverage period of P is less than the significance level 0.05, therefore there is no linear relationship between the independent anc Psychoactive response function Therefore, data must be converted using Box-Tidwell transformation (Hilbe, 2009) as follows:

$$X_4' = X_4 \quad (7)$$

When testing Box-Tidwell, it was found that the interaction of the coverage period had a p value of 0.80, which is greater. The significance level of 0.05 is therefore not statistically significant, ie there is a linear relationship between the independent variabl Responses to the logic, details are shown in Table 4.

Page 9

Research article

Table 4 Test of the 2nd Coverage Period Linear Correlation Variable

The test	β	*SE*	*Z*	Phi value
Duration of coverage				
- constants	3.7333	1.2919	2.89	0.00
- X'	-2.7406	1.0183	-2.69	0.01
- $ln(X_4') \times X_4'$	0.9204	0.5188	1.77	0.08

2) multiple linear relationship test (Multicollinearity) can be achieved by checking the error value. Standard () *SE* of regression coefficients () B (Josephat, 2018) from Table 5 that independent forms of payment. Monthly and quarterly premiums There is a standard error greater than 2.0, so independent variables must be taken. Monthly and quarterly premium payment models from the logistic regression model (Tabachnick, 2013). It was found that the standard error of each independent variable was not greater than 2.0, therefore each independent variable had n Plural linear Details are shown in Table 6. And can create a psycho response function as follows

$$Logit = 1.6004 - 0.0316 X_1 - 0.2396 X_{2A} - 0.8553 X_3 - 0.2785 X_4' + 0.4924 X_{6C} + 1.0946 X_{6D} \quad (8)$$

Table 5 The coefficient of logistic regression when there are 8 independent variables (2).

Independent variable	β	*SE*	*Z*	Phi value
Constant	-16.5211	346.4035	-0.05	0.96
X_1	-0.0182	0.0056	-3.22	0.00
X_{2A}	-0.3456	0.1541	-2.24	0.02
X_3	-0.8713	0.1793	-4.86	0.00
X_4'	-0.3892	0.1328	-2.93	0.00
X_{5A}	18.1904	346.4033	0.05	0.96
X_{5B}	18.0051	346.4035	0.05	0.96
X_{6C}	0.4188	0.2217	1.89	0.06
X_{6D}	0.7769	0.5048	1.54	0.12

Table 6 The coefficient of logistic regression when there are 6 independent variables.

Independent variable	β	*SE*	*Z*	Phi value	*OR*	*OR* confidence interval	
						Bottom edge	Top edge
Constant	1.6004	0.2855	5.61	0.00	4.9551	2.8523	8.7393
X_1	-0.0316	0.0051	-6.21	0.00	0.9689	0.9592	0.9786
X_{2A}	-0.2396	0.1486	-1.61	0.11	0.7870	0.5863	1.0504
X_3	-0.8553	0.1720	-4.97	0.00	0.4252	0.3029	0.5947
X_4'	-0.2785	0.1232	-2.26	0.02	0.7569	0.5910	0.9579
X_{6C}	0.4924	0.2053	2.40	0.02	1.6363	1.0956	2.4546
X_{6D}	1.0946	0.5077	2.16	0.03	2.9879	1.1572	8.7520

Research results

The research results will consist of 3 parts: the investigation of the suitability of the logistic regression model, the Validate the logistic regression model And the new unit forecast as follows

Investigation of the suitability of logistic regression models

Verify the suitability of the logistic regression model from the square statistic (χ^2) Subject evaluation Overall model evaluation by Likelihood ratio test was 0.00 p, which was less than 0.05 level of significance. (Ae, 2013) shows that a model with a free variable is more suitable than a model with just a fixed value. And testing for good health (Goodness of fit test) by means of Hosmer and Limes show. Had a p value of 0.39, which is greater than the significance level 0.05 (Allison, 2014) showed that the logistic regression model was appropriate. Details are shown in Table 7.

Table 7 Examination of the suitability of logistic regression models

Method	Degrees of freedom	Second	Phi value
Probability ratio test	1	256.29	0.00
Hosmer and Limes show	8	8.46	0.39

Validation of the logistic regression model

Validate the validity based on statistical value, Cox and Snell's R and Nagel Kerki (Nagelkerki) expressed the percentage variation of the variable as described by the independent variable. Found that independent var Varies in the range between 12.85% and 17.60%, details are shown in Table 8.

Research article

Table 8 Tests for the validity of the logistic regression model

Method	R^2
Cox and Snail	0.1285
Nagel Kirki	0.1760

New unit forecast

New unit forecast This can be done by using the information in Part 2. Or test data to predict that the insured will ask Cancel the policy or not require

$P(Y = <) 1 0.5$ Will predict that the policy will not be terminated

$P(Y = \geq) 1 0.5$ Will predict that the policy is terminated

From Table 9, it was found that this forecast model predicts the number of insureds who have been terminated for 148 insurar Representing 31.76% and 318 insured persons do not terminate their policy, or 68.24%, which the forecast is correct.

$[(81 + 231/466\ 100\ 66.95\% =$

Table 9 Forecast of policy retirement

Observation	Forecast value		
	Policy expiration	Does not terminate the policy	Total
Policy expiration	81	87	168
Does not terminate the policy	67	231	298
Total	148	318	466

Criticize the findings

From the construction of the policy retirement forecast model using the logistic regression model. With appropriate independe Total 6 characters, including age, amount of life insurance between 50,001 - 100,000 baht , premium payment period More than 3 years square root of the coverage period And the occupation class, occupation class 3 and class 4 occupation, which is c And Xu's research (2015) for similar independent variables. The forecasting model was able to predict the policy retirement. This may be more accurate by using other independent variables related to policy expiration. In creating a predictive model, such as information on the health examination of the insured Interest rate Including the factors of Economics, etc.

Research findings

From the basic information of the insured, it was found that 36.14% of the policy was terminated. Most of them were female. 452 people or 19.40 percent of the amount of life insurance in the range of less than or equal to 50,000 baht, 465 people, which is 19.96% Premium payment period ranged from 1-3 years for 629 persons or 27.00%.

Pay monthly premiums for 818 persons, equivalent to 35.11%. Policy format with savings of 838 persons.

Accounted for 35.97% through the telephone distribution channel of 842 persons or 36.14%.

Equal to 10,000 baht for 842 people, representing 36.14 percent and occupation level level 328, accounting for 14.08 percent.

From selection of independent variables And used independent variables to create a model for forecasting policy termination

$$Logit = 1.6004 - 0.0316 X_1 - 0.2396 X_{2A} - 0.8553 X_3 - 0.2785 X_4' + 0.4924 X_{6C} + 1.0946 X_{6D}$$

The model predicted that 148 insured persons have been terminated, accounting for 31.76 percent.

Ability to accurately forecast test datasets 66.95%

Acknowledgment

Education was accomplished with kindness and support from life insurance companies. Human resource development project Science (Good Scholarships for Science of Thailand) including Department of Statistics, Faculty of Science Kasetsart University That generates a place to use various resources to analyze the data until it is successful.

Reference document

Ae, HP (2013). An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy of Nursing* , *43* (2), 154-164.

Allison, DP (2014). Measures of Fit for Logistic Regression. *SAS Global Forum,* 1-12.

Hilbe, MJ (2009). *Logistic Regression Models*. (1). United States of America: CRC Press.

Josephat, K.P. & Ame, A. (2018). Effect on Testing Logistic Regression Assumption on the Improvement of the Propensity Scores. *International Journal of Statistics and Applications, 8* (1), 9-17.

Kaiyawan, Y. (2012). Principle and Using Logistic Regression Analysis for Research. *Journal of Rajamongala University of Technology Srivijaya, 4* (1), 1-12. (in Thai)

Kleinbaum, GD & Klein, M. (2002) *Logistic Regression: A Self-Learning Text* . (2). United States of America: Springer.

Office of Insurance Commission. (2016). *Life insurance*. Retrieved August 1, 2018, from http://www.oic.or.th/en/consumer / insurance / about / life.

Peng, J., Lee LK & Ingersoll MG (2002). An Introduction to Logistic Regression Analysis and Reporting. *Journal of Educational Research, 96* (1), 3-14.

Sinsomboonthong, S. (2016). *Multivariate Analysis*. (1). Bangkok: Chulalongkorn University Press. (in Thai)

Tabachnick, GB & Fidell, SL (2013). *Using Multivariate Statistics* . (6). United States of America: Courier

Companies.
Thai Life Assurance Association, (2018). Overview of Thai Life Assurance business in 2017 and Trend of Thai. Life Assurance business in 2018.Journal *of Life Assurance* , *38* (1), 1-32. (In Thai)
Xu, R., Lai, D., Cao, M., Rushing, S. & Rozar, T. (2015). Lapse Modeling for the Post-Level Period: A Practical Application of Predictive Modeling. *Society of Actuaries,* 1-22.
Zian, J., Miller, A. & Ducuroir, F. (2016). Lapse Rate Models in Life Insurance and a Practical Method to Foresee Interest Rate Dependencies. *A Reacfin White Paper on Life Insurance* , 1-20.

Annex

Table Appendix 1 General information of the insured

information	variable	Policy expiration Amount (percent)	Does not terminate the policy Amount (percent)	Total (percentage)
age	X_1	842 (36.14)	1,488 (63.86)	2,330 (100.00)
Amount of Life Insurance (Baht)				
1) less than or equal to 50,000	Reference group	465 (19.96)	744 (31.93)	1,209 (51.89)
2) 50,001 - 100,000	X_{2A}	109 (4.68)	354 (15.19)	463 (19.87)
3) more than 100,000	X_{2B}	268 (11.50)	390 (16.74)	658 (28.24)
Premium payment period (years)				
1) 1 - 3	Reference group	629 (27.00)	629 (27.00)	1,258 (53.99)
2) more than 3	X_3	213 (9.14)	859 (36.87)	1,072 (46.01)
Period of coverage (years)	X_4	842 (36.14)	1,488 (63.86)	2,330 (100.00)
Premium payment types				
1) One time payment	Reference group	0 (0.00)	67 (2.88)	67 (2.88)
2) Postpaid	X_{5A}	818 (35.11)	1,075 (46.14)	1,893 (81.24)
3) Quarterly	X_{5B}	24 (1.03)	28 (1.20)	52 (2.23)
4) Semi-annual	X_{5C}	0 (0.00)	163 (7.00)	163 (7.00)
5) Yearly	X_{5D}	0 (0.00)	155 (6.65)	155 (6.65)
Occupation level				
1) Not specified	Reference group	274 (11.76)	754 (32.36)	1,028 (44.12)

information	variable	Policy expiration Amount (percent)	Does not terminate the policy Amount (percent)	Total (percentage)
2) Class 1 occupation	X_{6A}	328 (14.08)	474 (20.34)	802 (34.42)
3) Class 2 occupation	X_{6B}	146 (6.27)	190 (8.15)	336 (14.42)
4) Class 3 occupation	X_{6C}	76 (3.26)	63 (2.70)	139 (5.97)
5) Class 4 occupation	X_{6D}	18 (0.77)	7 (0.30)	25 (1.07)
sex				
1) male	Reference group	390 (16.74)	666 (28.58)	1,056 (45.32)
2) female	X_7	452 (19.40)	822 (35.28)	1,274 (54.68)
Premium (baht)				
1) less than or equal to 10,000	Reference group	842 (36.14)	1,329 (57.04)	2,171 (93.18)
2) More than 10,000	X_8	0 (0.00)	159 (6.82)	159 (6.82)
Distribution channels				
1) by phone	Reference group	842 (36.14)	1,412 (60.60)	2,254 (96.74)

2) The company	X_9	0	(0.00)	76	(3.26)	76	(3.26)
Policy format							
1) lifetime model	Reference group	1	(0.04)	4	(0.17)	5	(0.21)
2) Periodic form	X_{10A}	3	(0.13)	3	(0.13)	6	(0.26)
3) Savings	X_{10B}	838	(35.97)	1,481	(63.56)	2,319	(99.53)
Total		842	(36.14)	1,488	(63.86)	2,330	(100.00)

Note: Occupations in insurance are divided into 4 classes, which are

Occupation Level 1 Most types of work work in the office and do jobs that do not use machines such as executives. Company employees, doctors, pharmacists, nurses, government officials, etc.

Class 2 Occupation Type of work: Most of them are outside the office or have to work outdoors all the time. Is a group of cı Proficiency and skills may sometimes be employed, such as agents / brokers, engineers, carpenters, small business owners. etc.

Occupation level 3, job description that Most of them use heavy machinery. Or is a laborer Or work outside the office Such as workers in the production, transportation, sales staff, actors, guides, journalists, drivers etc.

Occupation 4th level, job description that is most likely to have an accident There is a higher risk than other special classes Stunt performer Construction worker Security guard Courier staff, etc.