Available online at https://journal.rescollacomm.com/index.php/ijqrm/index

International Journal of Quantitative Research and Modeling

e-ISSN 2721-477X
p-ISSN 2722-5046

Vol. 5, No. 2, pp. 148-153, 2024

Analysis of Pet Owners' Willingness to Pay for Pet Insurance Premiums in DKI Jakarta Using Logistic Regression Model

Andhita Zahira Adib[1], Riaman[2], Betty Subartini[3]

[1]*Mathematics Undergraduate Study Program, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia*
[2,3]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia*

**Corresponding author email: andhita20001@mail.unpad.ac.id*

Abstract

Pets provide many benefits to their owners, both physically and mentally. Pet lovers are increasingly aware of the importance of proper health and care for their beloved animals. This has led pet enthusiasts to consider pet insurance. In participating in insurance, there are factors that influence the willingness of pet owners to pay premiums. The objective of this research is to determine the premium for pet insurance and analyze the factors influencing the Willingness To Pay (WTP) of pet owners. This study utilizes choice modeling format by conducting surveys to identify the factors influencing the purchase of pet insurance. Subsequently, binary logistic regression model analysis using the Maximum Likelihood Estimation (MLE) method and the Newton-Raphson Iteration approach is employed to analyze the factors influencing the magnitude of WTP. The research results show that the average willingness to pay for pet insurance premiums is IDR128,574.76 per year. Factors influencing the decision of pet owners include the number of family dependents and awareness of the importance of participating in pet insurance. The likelihood of cat owners being willing to pay pet insurance premiums is 0.8691 or 86.91%.
Keywords: Pet Insurance, Willingness to Pay (WTP), Choice Modeling, Logistic Regression, Maximum Likelihood Estimation (MLE) method

1. Introduction

In the Indonesian dictionary or also known as KBBI, pets or domesticated animals are defined as animals that are raised for pleasure or cultivation (KBBI, 2016). In addition to providing entertainment, keeping pets can bring many benefits to their owners. Pets can make someone feel valuable and needed, thereby eliciting happiness (Erliza and Atmasari, 2022). Similar to humans, the health of pets can also be at risk and may require expensive medical care.

In an effort to provide the best protection and care for beloved animals, the pet insurance industry is rapidly growing. With pet insurance, pet owners don't need to worry about the costs of health and emergency care because insurance companies can compensate for potential risks such as death and others. Risks that can occur to pets vary, including illness, accidents, and even death. However, if the risk occurs within less than 30 days after the insurance policy begins, then generally the insurance will not cover it. Therefore, policyholders should be aware of what is offered by the insurance company.

Prasetyanto and Elkhasnet (2015), in their research, compared the Choice Modeling (CM) and Contingent Valuation (CV) models and concluded that CM generally yields better results than CV, but its questionnaire design is more complex. Becker et al. (2022) examined the willingness to pay of pet owners by developing a model that helps classify pet owners based on their willingness to pay related to their financial capacity. Based on this model, predictions about the limitations and capabilities of pet health insurance can be obtained. In the study conducted by Inna et al. (2023), ordinal logistic regression analysis was used to determine the factors influencing the waiting time for job placement among alumni. Logistic regression analysis was employed to analyze the relationship between one or more independent variables and the ordinal dependent variable.

2. Literature Review

2.1 Pet Insurance

In Presidential Regulation of the Republic of Indonesia No. 48 of 2013 concerning the Husbandry of Pets, pets are defined as animals whose lives depend on humans (Peraturan Pemerintah 2013). As owners, humans are responsible for the well-being of their pets, including physical and mental aspects in accordance with their natural behavior.

In caring for pets, unavoidable risks such as diseases, disabilities, or death may occur. Purchasing insurance policies before these risks arise can provide protection and reduce expensive care costs. Particularly for cats, owners need to be aware of potential diseases and consider insurance to cover expensive treatment expenses.

2.2 Willingness to Pay

Willingness to pay (WTP) is the highest price an individual is willing to pay to obtain the benefits of a good or service (Riana, et al., 2019). Research conducted by Riana et al. (2019) analyzed consumers' willingness to pay using the Contingent Valuation Method to calculate the average maximum WTP and logistic analysis to understand the factors influencing consumers' WTP for organic rice. In calculating the average value that respondents willing to pay are willing to spend, you can use the formula in equation (1).

$$E(WTP) = \frac{\sum_{i=1}^{n} WTP_i}{n}, \tag{1}$$

where $E(WTP)$ is the expected WTP, WTP is the value of willingness to pay of the *i*-th respondent, n is the number of respondents, and i is the i-th respondent willing to pay ($i = 1,2,3, \dots, n$).

One method for data collection is the Stated Preference (SP) method. In this study, the SP method will be used to find WTP, specifically by using choice modeling. Choice modeling directly assesses the characteristics of an attribute and the marginal change in those characteristics as opposed to just assessing whether a product is good or bad overall.

2.3 Logistic Regression Model

Multiple linear regression was used to analyze the factors that influence the WTP value of pet owners. Odds or odds of an event Y symbolized as $O(Y)$ is the ratio between the odds of two outcomes of a binary variable. In other words, Odds is the ratio between the probability of event Y occurring and the probability of event Y not occurring (Harlan, 2018). The general form of the logistic regression equation in equation (2) follows.

$$\text{logit}(Y) = \ln O(Y) = \ln\left(\frac{P(Y=1|X=x_i)}{1-P(Y=1|X=x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \tag{2}$$

with Y is the dependent variable, x_{ip} = pth independent variable where $i = 1,2,\dots,n$, β_0 is the model intercept parameter (constant), and $\beta_1,\dots,\beta_p$ is the regression coefficients in the logistic model.

Then, convert equation (2) into exponent form in both segments to return to chance form and suppose $P(Y = 1|X = x_i) = \pi(x_i)$ where $\pi(x_i)$ is the probability of willingness to pay for the i-th data into the following equation (3) (Riana, et al., 2019).

$$O(Y) = \left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}} \tag{3}$$

The odds equation for logistic regression is in equation (4) (Agresti, 2007).

$$\pi(x_i) = \frac{e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}}{1 + e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}}} \tag{4}$$

2.4 Parameter Estimation of Logistic Regression Model

This study uses indicator variables that are binary or dichotomous in logistic regression and the probability function of discrete random variables from the Bernoulli distribution. According to Hosmer et.al. (2013), the Bernoulli distribution probability density function that can express the contribution to the likelihood function at (x_i, y_i) is:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \tag{5}$$

Substitute equation (4) into the likelihood function of the Bernoulli distribution in equation (5), resulting in the likelihood function of the continuously distributed probability Y in equation (6).

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left[\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}}\right]^{y_i}\cdot\left[1-\frac{e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}}{1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}}\right]^{1-y_i} \quad (6)$$

then, the natural logarithm of the likelihood function in equation (6) becomes:

$$\ln L(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[y_i\left(\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right) - \ln[1+e^{\beta_0+\sum_{j=1}^{p}\beta_j x_{ij}}]\right] \quad (7)$$

where x_{ij} is the representation of the i-th row in the j-th column of the **X** matrix. The likelihood function in equation (7) is used for parameter β_j estimation by maximizing the likelihood function.

2.5 Newton-Raphson Iteration

The Newton-Raphson algorithm approximates the log-likelihood function in an initial guess environment by a polynomial function that has a concave (wavy) parabolic shape making it easier to determine the maximum location of the approximating polynomial (Agresti, 2007). The steps in the Newton-Raphson Iteration method are:

1. Determining the initial estimate, which is:

$$\hat{\boldsymbol{\beta}}^{(0)} = \begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}$$

2. Determining the gradient vector that has first-order derivatives, using the formula:

$$\mathbf{u} = \mathbf{X}^{\mathbf{T}}[\boldsymbol{y_i} - \boldsymbol{\pi}(\boldsymbol{x})] \quad (8)$$

3. Finding the inverse of the Hessian Matrix, which contains second-order partial derivatives from equation (2.20), using the formula:

$$\mathbf{H}^{-1(\mathrm{r})} = -[\mathbf{X}^{\mathbf{T}}\mathbf{V}\mathbf{X}]^{-1} \quad (9)$$

4. Calculating the estimated value $\hat{\boldsymbol{\beta}}$
The value $\hat{\boldsymbol{\beta}}^{(1)}$is obtained to find the value $\hat{\pi}(x_i)^{(1)}$, which then obtains the values $\mathbf{u}^{(1)}$) and $\mathbf{H}^{(1)}$). Subsequently, obtaining the value $\hat{\boldsymbol{\beta}}^{(2)}$ and so on for each *j*. The Newton-Raphson equation is in equation (10) as follows (Agresti, 2007):

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} + \mathbf{H}^{-\mathbf{1}(\mathrm{r})}\mathbf{u}^{(\mathrm{r})} \quad (10)$$

5. The iteration process stops when the estimated value obtained reaches $\left|\hat{\boldsymbol{\beta}}^{(r+1)} - \hat{\boldsymbol{\beta}}_{\boldsymbol{j}}\right| < \varepsilon_0$with ε_0 being an accuracy level of 0.1%. The $\hat{\beta}$ value is obtained for the binary logistic regression model.

2.6 Statistical Hypothesis Test

2.6.1 Likelihood Ratio (LR) Test

The Likelihood Ratio (LR) test function is to test whether all explanatory variables together affect the dependent variable. Kusmiyati and Hakim (2020) explained in their research that the Likelihood Ratio (LR) test is carried out by comparing the calculated Chi-Square value with the Chi-Square value in the table.

The test statistics based on Hosmer and Lemeshow (2013) are:

$$G = -2\ln\left[\frac{(likelihood\ without\ independent\ variables)}{(likelihood\ with\ independent\ variables)}\right] \quad 11)$$

The critical region (H_0 rejection) is H_0 rejected if $G > \chi^2_{(\alpha,v)}$

2.6.2 Wald Test

Wald test in logistic regression is used to determine the effect of each independent variable on the dependent variable partially (Widarjono, 2018). The test statistics based on Hosmer and Lemeshow (2013) are:

$$W = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \quad 12)$$

with $\hat{\beta}_j$ estimator of $\hat{\beta}_j$ and $SE(\hat{\beta}_j)$ is $\sqrt{([\mathbf{X}^{\mathrm{T}}\mathbf{V}\mathbf{X}]^{-1})_{jj}}$, the standard error of the estimator $\hat{\beta}_j$. The critical region (H_0 rejection) is H_0 rejected if $W > \chi^2_{(\alpha;1)}$.

2.6.3 Hosmer and Lemeshow Test

This test shows whether the model can explain the data in the study or not by testing the hypothesis (Nurmalitasari and Purwanto, 2022).

$$\hat{C} = \sum_{k-1}^{g} \frac{(o_k - n_k'\bar{\pi}_k)^2}{n_k'\bar{\pi}_k(1-\bar{\pi}_k)} \tag{13}$$

where g is the number of groups, $\bar{\pi}$ is the expected value in column $\bar{\pi}_k$, m_j is the number of observed values in row $\bar{\pi}_k$, dan $n_k{}'$ is the number of subjects in the k-th group.

3. Materials

The data used in this study is sourced from primary data by conducting a survey in the form of filling out questionnaires to cat owners in DKI Jakarta. The questionnaire distribution process takes one to two months. In the end, 215 respondents filled out the questionnaire.

4. Results and Discussion

4.1 WTP Average Value

Respondents were given three scenario options, namely in option A the premium rate offered was IDR60,000.00/year, option B offered a premium rate of IDR110,000.00/year and option C offered a premium rate of IDR160,000.00/year. What distinguishes the three options is that the benefits or attributes offered are different, of course, adjusting the rates given. The majority of respondents, namely 45%, chose package A with the lowest premium offered. The choice of this attribute also provides an overview to respondents who previously did not know about pet insurance, where insurance companies offer many beneficial attributes for the policy owner's pet.

In the questionnaire results obtained, it was found that the WTP value of each respondent varied. Next, calculate the average using equation (1), as follows:

$$\begin{aligned} E(WTP) &= \frac{\sum_{i=1}^{n} WTP_i}{n} \\ &= \frac{\sum_{i=1}^{121} WTP_i}{215} \\ &= \text{IDR128,574.76} \end{aligned}$$

The result of the calculation of the average WTP value is IDR128,574.76/year, which means that on average the respondents are only willing to pay a pet insurance premium of IDR128,574.76/year.

4.2 Selection of Significant Variables

The next step is to calculate with the help of the IBM SPSS Statistics 26 application to see whether or not there is a relationship between each independent variable and the dependent variable using the multicollinearity test. After the multicollinearity test, it can be concluded that the independent variables x_5, x_{10} and x_{13} do not experience symptoms of multicollinearity. In other words, x_5, x_{10} and x_{13} can be entered into the equation to find the parameter estimation of the logistic regression model because they fulfill the assumption of no multicollinearity symptoms in the data.

After the multicollinearity test and confirming that the data met the multicollinearity assumption, the selected independent variables were parameterized to obtain a logistic regression model of willingness to pay pet insurance premiums. Substitute the variables x_5, x_{10} and x_{13} to the equation (7) then get the following equation.

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i(\beta_0 + \beta_5 x_{i5} + \beta_{10} x_{i10} + \beta_{13} x_{i13}) - \ln\left[1 + e^{\beta_0 + \beta_5 x_{i5} + \beta_{10} x_{i10} + \beta_{13} x_{i13}}\right]\right]$$

4.3 Parameter Estimation Results

Parameter estimation is obtained by maximizing equation (7) using newton raphson iteration in equation (10). The results of parameter estimation assisted with SAS OnDemand software are shown in Table 1.

Table 1: Initial model parameter estimation results

Parameter	Parameter Value	Likelihood	Standard Error	Result
$\hat{\beta}_0$	0.1797	194.277 (without independent variable)	−	−
$\hat{\beta}_5$	−1.2064		0.5109	Significant
$\hat{\beta}_{10}$	0.6483	152.445 (with independent variable)	0.4198	Not significant
$\hat{\beta}_{13}$	2.5296		0.4604	Significant

In the Likelihood Ratio (LR) test, using equation (11.0 the value of $G = 41.832$ is obtained, and the $\chi^2_{tabel} = \chi^2_{(0,05,3)} = 7.815$ so that the value of $G > \chi^2_{tabel}$ and that independent variable that significantly affects the dependent variable. In the Wald test, using equation (12) turns out variable x_{10} does not significantly affect variable y because the Wald value obtained is $< 3.8415 = \chi^2_{(0,05,1)}$ and variable x_{10}will be eliminated. Therefore, with the Wald value of x_5 and $x_{13} > 3.8415$, H_0 is rejected, which means that the variable number of family dependents and the variable awareness of the importance of pet insurance have a significant effect on the willingness to pay pet insurance premiums.

4.4 Final Logistic Regression Model

After the parameter significance test is carried out and the data meets the test criteria, the final logistic regression model is obtained by eliminating independent variables that do not have a significant relationship and assistance from the SAS OnDemand application following equation (2) as follows:

$$\text{logit}(\hat{y}) = 0.5316 - 1.2094x_{i5} + 2.5695x_{i13}$$

for writing the comparison between the odds of two outcomes of a binary variable as in the equation (3), as follows.

$$O(Y) = \left(\frac{P}{1-P}\right) = e^{0.5316-1.2094x_{i5}+2.5695x_{i13}}$$

when written with the odds equation for logistic regression as in equation (4), as follows:

$$\pi(x_i) = \frac{e^{0.5316-1.2094x_{i5}+2.5695x_{i13}}}{1+e^{0.5316-1.2094x_{i5}+2.5695x_{i13}}}$$

Then, the Hosmer-Lemeshow test was carried out using equation (13) and obtained a value of $\hat{C} = 3.851$. The value of the Chi square table with an independent degree of 2 is $\chi^2_{tabel} = \chi^2_{(0,05,2)} = 5.9915$. The value of $\hat{C} < \chi^2_{(0,05,2)}$, then H_0 is accepted. Thus, the model can explain the research data and is declared feasible.
With the odds equation that has been obtained, the odds values of the model parameters are shown in Table 2.

Table 2: Odds value of the logistic regression model parameter

Parameter	Odds Value
$\hat{\beta}_0$	1.7033
$\hat{\beta}_5$	0.2983
$\hat{\beta}_{13}$	13.0592

Based on Table 2., it can be inteIDRreted that x_5 affects the comparison of WTP opportunities for cat owners by 0.2983 when it is 1 (having less than two family dependents), while x_{13} has an effect of 13.0592 when it is 1 (aware of the importance of pet insurance). If the value of x_5 and x_{13} is 0, the comparison of the probability of WTP for cat owners is 1.7033 times.

The equation $\pi(x_i)$ results in a large chance of WTP for cat owners shown in Table 3 as follows.

Table 3: WTP probability of cat owners

$\pi(x_i)$			
$x_5, x_{13} = 0$	$x_5 = 0,$ $x_{13} = 1$	$x_5 = 1,$ $x_{13} = 0$	$x_5, x_{13} = 1$
0.6301	0.9570	0.3370	0.8691

Table 4 shows the results of the calculation of the probability of willingness to pay pet insurance premiums and concludes that:

1. If a cat owner with more than 2 family dependents and does not realize the importance of pet insurance, the probability of willingness to pay pet insurance premiums is 0.6301 or 63.01%.
2. If the owner of a cat with more than 2 family dependents and realizes the importance of pet insurance, the probability of willingness to pay pet insurance premiums is 0.9570 or 95.70%.
3. If the owner of a cat with 2 or less family dependents and does not realize the importance of pet insurance, then the probability of willingness to pay pet insurance premiums is 0.3370 or 33.70%.
4. If the owner of a cat with 2 or less family dependents and realizes the importance of pet insurance, the probability of willingness to pay pet insurance premiums is 0.8691 or 86.91%.

5. Conclussion

From this study, the average WTP value of cat owners for pet insurance premiums is IDR128.574,76/year. There are two variables that influence cat owners' willingness to pay for pet insurance premiums, namely the number of family dependents ($\boldsymbol{x_5}$) and awareness of the importance of participating in pet insurance ($\boldsymbol{x_{13}}$). The probability value of cat owners in DKI Jakarta paying pet insurance premiums is 0.8691 or 86.91%.

References

Agresti, A. (2007) *An Introduction to Categorical Data Analysis Second Edition*. New Jersey: A John Wiley & Sons, Inc.

Becker, M., Volk, H. and Kunzmann, P. (2022) ‘Is Pet Health Insurance Able to Improve Veterinary Care? Why Pet Health Insurance for Dogs and Cats Has Limits: An Ethical Consideration on Pet Health Insurance’, *Animals*, 12(13). Available at: https://doi.org/10.3390/ani12131728.

Erliza, Y. and Atmasari, A. (2022) ‘Pengaruh Pet Attachment terhadap Happiness pada Pemilik Hewan Peliharaan di Kecamatan Sumbawa’, *Psimawa*, 5(1), pp. 54–62. Available at: http://jurnal.uts.ac.id/index.php/PSIMAWA (Accessed: 7 August 2023).

Harlan, J. (2018) *Analisis Regresi Logistik*. I. Depok: Gunadarma.

Hosmer, D.W., Lemeshow, Stanley. and Sturdivant, R.X. (2013) *Applied logistic regression.* 3rd edn. Canada: Wiley.

Inna, S., Mariana, D. and Manaqib, M. (2023) ‘Analisis Regresi Logistik Ordinal Faktor-Faktor Yang Memengaruhi Waktu Tunggu Kerja Alumni’, *SIGMA*, VIII(2), pp. 75–83.

Kusmiyati, S.D. and Hakim, M.Z. (2020) ‘Pengaruh Ukuran Perusahaan, Profitabilitas, Cash Holding, Debt To Equity Ratio Dan Net Profit Margin Terhadap Perataan Laba (Studi Empiris Pada Perusahaan Manufaktur Sektor Industri Barang Konsumsi yang Terdaftar di Bursa Efek Indonesia Tahun 2014-2018)’, *Profita: Komunikasi Ilmiah Akuntansi dan PeIDRajakan*, 13(1), pp. 58–72. Available at: https://doi.org/10.22441/profita.2020.v13i1.005.

Nurmalitasari and Purwanto, E. (2022) ‘Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik’, *Jurnal Derivat*, 9(2), pp. 145–152.

Prasetyanto, D. and Elkhasnet (2015) ‘Perbandingan Nilai Kesediaan Membayar Menggunakan Pendekatan Stated Preference Contingent Valuation dan Stated Preference Stated Choice’, *Jurnal Transportasi*, 15(2), pp. 99–106.

Riana, E.T., Mukson and Roessali, W. (2019) ‘Analisis Kesediaan Membayar (Willingness to Pay) Konsumen terhadap Berbagai Jenis Beras Organik di Kota Semarang (Kasus di Pasar Modern Gelael Signature)’, *Jurnal Ekonomi Pertanian dan Agribisnis*, 3(4), pp. 689–700. Available at: https://doi.org/10.21776/ub.jepa.2019.003.04.4.

Widarjono, A. (2018) *Analisis Regresi dengan SPSS*. 1st edn. Yogyakarta: UPP STIM YKPN.