ESTIMATION OF LOGISTIC REGRESSION MODEL ON THE RISK OF ACUTE RESPIRATORY INFECTION TODDLERS AT CIKAMPAK TORGAMBA PUBLIC HEALTH CENTER USING THE MAXIMUM LIKELIHOOD METHOD

Dian Maya Sari[1], Fibri Rakhmawati[2], Hendra Cipta[3]
[1,2,3]Universitas Islam Negeri Sumatera Utara Medan
Jln. Lapangan Golf, Desa Durian Jangak, Pancur Batu, Deli Serdang, Sumatera Utara
[1]dianmaya1103@gmail.com

Abstract

According to the World Health Organization (WHO), the incidence of Acute Respiratory Infections (ARI) in the world by number of toddler who died achieve ± 13 million each year. Researchers will estimate the model equations logistic regression using maximum likelihood. The purpose of this study was to determine the logistic regression model against respiratory infection in infants and the test results of the method of maximum likelihood against respiratory infection in infants in the health center Cikampak Torgamba. Logistic regression Model against the toddler obtained results that the model to be used and the test results of the method of maximum likelihood that the model is significant.

Keywords: *Acute Respiratory Infections*; regression logistics; maximum *likelihood*.

Introduction

Housing sanitation and the environment is one of the main activities in the program of sanitation. Sanitation and the environment itself preferably in areas that have a high risk of possible transmission of diseases such as, diarrhea, Pulmonary tuberculosis, helminthiasis, dengue hemorrhagic fever (DHF) and Acute Respiratory Infections (ARI) with prevent homes that do not meet the health requirements (Sri, 2017)

Acute Respiratory Infections are a group of complex and heterogeneous caused by a variety of factors supporting the risk that attack every part of respiratory tract starting from the upper tract (nose) to the channel bottom in the human respiratory system. ARI is also a disease that often occurs in children, especially toddlers, because the body's defense system which is still low so prone to diseases. Clinically ARI is a signs and symptoms of acute as a result of infection that occurs in any part of the respiratory tract. As for that belongs to the category of ARI infection is influenza, measles, pharyngitis, tracheitis, brounchitis acute, bronchiolitis, and pneumonia (Bambang, 2016).

Some of the risk factors that may be associated with the incidence of ARI are divided over the factors of intrinsic and extrinsic factors. Factors intrinsic included age, gender, nutritional status, low birth weight, the status of immunization, and breastfeeding. Extrinsic factors such as air pollution, house ventilation, smoking, use of fuel, lack of access to clean water and so forth. ARI begins with a hot and accompanied by one or more of the symptoms that a sore throat or pain on swallowing, runny nose, dry cough or phlegm. Prevalence of ARI is calculated in the period of one month (Kemenkes RI, 2022). Based on the results of preliminary observations of researchers in the health center Cikampak Torgamba Kabupaten Labuhanbatu Selatan many of the parents of the toddler habit of smoking, especially the father. The influence of smoking habits of parents and family members living in one house with the incidence of ARI in children in Phc Cikampak. In addition, one of the environmental factors that is the smoke of the burning of biomass fuels also be the cause of ARI infection in infants.

Housing is not equipped with a good air ventilation will cause the air circulation is not smooth and the room temperature is not appropriate. If the room temperature is not appropriate, it will cause the radiation of heat directly towards the body, or the loss of body heat due to the cold air. The factors of the home environment such as ventilation also causes the transmission of ARI infection, where ventilation can maintain the condition of air that is healthy for humans (Sri, 2017).

Logistic regression is one of the multivariate analysis are useful to predict the dependent variable based on independent variable. Logistic regression is not as linear regression ordinary, logistic regression is a regression non-linear models where specified will follow the pattern of data in the form of categorical data. Logistic regression has several advantages that do not have the assumption of normality over the dependent variable used in the model, variable-the dependent variable in the logistic regression can be a mixture of continuous variables, discrete and dichotomous (Sembiring, 2018). Logistic regression is a regression analysis method with the response variable is a binary variable or categorical, for the response variable is binary or dichotomous, which consists of two categories, namely 0 and 1. The symbol used is the two numbers 0 or 1 to replace the tags on the response variable. The analysis is used to solve the problem is a binary logistic regression model. Its regression analysis using to obtain the relationship between the X as a variable descriptors with Y as the variable response. Regardless of the value of X when disubtitusikan into the logistic function the result will be in the range between 0 and 1 (Sutrianah, 2018).

Estimation is the discussion of statistics that deals with estimating the values of parameters based on the empirical data derived from a random sample. The parameter is a constant that characterizes the characteristics of the population. Activity estimation seeks to mengaproksimasi these parameters using a measure derived from the sample (Panji, 2018). Parameter estimation in this model using the method of maximum likelihood with hypothesis testing using test likelihood ratio. The maximum likelihood method is considered to more powerful on the results of the estimator with the nature of statistics. In addition, this method is also more efficient for the measurement uncertainty through the boundaries of belief. The parameters obtained from the function mestimasi maximum likelihood is the value of the actual (Diandra, 2018).

The maximum likelihood method is one of the methods most good to acquire an estimate of the single. This method cannot be used if the distribution is not known. The maximum likelihood method is useful to determine the parameters of maximum likelihood of the sample data. Although its methodology for maximum likelihood including simple but the implementation of the math very strong. Budget parameters obtained from the function estimation maximum likelihood is the value of the approach to the true value. It is clear that the sample size determines the accuracy of the estimator (Muhammad, 2016).

According to Research conducted by Alfiah Safitri (2017), titled "Binary Logistic Regression Model on the Level of Open Unemployment in the Province of West Sulawesi in 2017, the results of its research indicate that the variable is non significant on the Level of Open Unemployment in the Province of West Sulawesi is the percentage of the gender with the male category and the percentage of businesses with the category of the farm where the variable gender has a p-value 0,017<5% and the variable field of business have a p-value of 0.006<5% can be interpreted that the two variables the effect on the rate of open unemployment.

Research conducted by the Sutrianah Burhan (2018) with the title "Parameter Estimation of Linear Regression Logistic with the Method of Maximum Likelihood Local On the Risk of Breast Cancer in Makassar" the results show that processing by the method of Newton Raphson method with the values of the parameter estimation converges after the seventh iteration, obtained $Y_0 = 0.588$, $Y_1 = 0.0012$, $Y_2 = 0.3206$, $Y_3 = 1.157$. While using the initial value of the ordinary least square method, the values of the parameters will converges to the same value with the iteration of the fifth. The same thing is also obtained by replacing the measure of central tendency than the mean to the mode, even the median of the data. Through the test of the feasibility of the model that has been conducted on estimating the parameters obtained for the

risk of cancer for women in the City of Makassar, the obtained form of the regression logistics is

$$\bar{y} = \frac{\exp(0.5880 + 0.0012(x_1 - \bar{x}) + 0.3206(x_2 - \bar{x}) + 1.157(x_3 - \bar{x}))}{1 + \exp(0.5880 + 0.0012(x_1 - \bar{x}) + 0.3206(x_2 - \bar{x}) + 1.157(x_3 - \bar{x}))}$$

Research Methods

This research was held in the health center Cikampak Riau, Torgamba district, Labuhan Batu Selatan Regency, North Sumatera. The data used in this research is secondary data obtained from the medical records of ARI in the health center Cikampak Torgamba in 2019. The type of data used in is quantitative data. The research variables used in this study is the response variable (Y) and variablel predictor (X) The response variable used is the status of the risk of ARI infection in infants, while the predictor variables used are shown in table as follows:

Table 1. Research Variables

Variables	Description	Category	Scale
Y	The Status of the risk of ARI	Positive or negative	Nominal
X_1	Age	0 : < 5 years 1 : 5 years	Ordinal
X_2	Gender	0 : girl 1 : boy	Nominal
X_3	Nutritional Status	0 : enough 1 : normal	Ordinal
X_4	A Family History Of Smoking	0 : yes 1 : no	Nominal

The procedure of research:

1. Creat the form data in data category
2. Doing estimating of thr parameter logistic regression
3. Determine the logistic regression model

$$p(x) = E(y \mid x) = \frac{\exp(Y_0 + Y_1X_1 + Y_2X_2 + Y_3X_3)}{1 + \exp(Y_0 + Y_1X_1 + Y_2X_2 + Y_3X_3)}$$

4. Doing significant parameters with concurrent test to determine the significance parameters of the predictor variable on the response variable, if it meets proceed to the partial test. If it does not meet then redoing the data collection.
5. Testing of the ratio likelihood. The test is performed by modeling all variables significant predictor to determine whether the predictor variables affect real or not.
6. Applying the test suitability of the model to evaluate whether or not suitable for the model with the data.
7. Make the conclusions from the results obtained.

Results and Discussion

Research conducted by the (Sutrianah Burhan ,2018) with the title "Parameter Estimation of Linear Regression Logistic with the Method of Maximum Likelihood Local On the Risk of Breast Cancer in Makassar" the results show that processing by the method of Newton Raphson method with the values of the parameter estimation converges after the seventh iteration, obtained $Y_0 = 0.588$, $Y_1 = 0.0012$, $Y_2 = 0.3206$, $Y_3 = 1.157$. While using the initial value of the ordinary least square method, the values of the parameters will converges to the same value with the iteration of the fifth. The same thing is also obtained by replacing the measure of central tendency than the mean to the mode, even the median of the data. Through the test of the feasibility of the model that has been conducted on estimating the parameters obtained for the risk of cancer for women in the City of Makassar, the obtained form of the regression logistics is

$$\bar{y} = \frac{\exp(0.5880 + 0.0012(x_1 - \bar{x}_1) + 0.3206(x_2 - \bar{x}_2) + 1.157(x_3 - \bar{x}_3))}{1 + \exp(0.5880 + 0.0012(x_1 - \bar{x}_1) + 0.3206(x_2 - \bar{x}_1) + 1.157(x_3 - \bar{x}_3))}$$

To estimating binary logistic regression model with the dependent variable ARI has two value categories Yes and No as well as the independent variables are age, gender, nutritional status, and family history of smoking, Testing the suitability of the model was conducted to determine the significance of the influence of independent variables simultaneously on the dependent variable using the ratio test likelihood as well as testing the significance of the model parameters individually by using wald.

1. Significance Test The Best Model

Ratio test likelihood is performed using a statistical test of G, the value of G used is 215.3982. If the value of the $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ then the model is said to be not significant and if the value of the $\beta_i \neq 0$ then the model is said to be significant with the region of rejection if $G > X^2_{\alpha;df=p=5.}$ The results of significance test the best model with a significance level of $\alpha = 5\%$ obtained that the value of the $X^2 = 11.0705 (< G = 215.3982)$. Based on these results, the obtained model are significant.

2. Test The Significance Of The Parameters (Partial Test)

The test statistic used is the value of Z is shown in table 2 below.

Tabel 2. Statistics of the Wald Test and P-Value

Parameters	Estimase	Std. Error	Z Value	Pr $(\geq \lvert z \rvert)$
(Intercept)	24.1933	2974.3709	0.0081	0.9935
Age (5 years)	-6.3536	0.8314	-7.6417	0.0000
Gender (girl)	0.2601	0.7076	0.3676	0.7131
Nutritional status (enough)	-4.7772	3669.2057	-0.0013	0.9990
Nutritional status (normal)	-20.1883	2974.3707	-0.0068	0.9946
A family history of smoking	-0.8897	1.3053	-0.6816	0.4955

Based on table 2, the results obtained if the value of the $\beta_j = 0$ the parameter j is said to be not significant, if the value of the $\beta_j \neq 0$ said parameters *j* significantly with the rejection region if the value of the $p_{value} > \alpha$. The results of the significance test parameters at the level of significance $\alpha = 5\%$ values obtained $p_{value} > \alpha = 0.000(< \alpha)$. Based on these results, obtained significant parameters are age, while the parameters gender, nutritional status, and family history of smoking was not significant.

3. Test The Feasibility Of The Model

Test the feasibility of the model is used to determine whether the selected model to be used. Testing is done to test the Hosmer-Lemeshow. The statistical test used is the value of X^2 $H_{HL}^2 = 1.3918$ and the value of the $p_{value} = 0.9944$. If the value of the $p_{value} < \alpha$ then the model is not fit for use, if the value of the $p_{value} > \alpha$ then the model is fit for use. The results of the feasibility test of a model with a significance level of $\alpha = 5\%$ values obtained $p_{value} = 0.994(> \alpha)$ so the model is feasible to use.

4. Model, Ods Ratio and the Accuracy of the Prediction

The model obtained based on the output in table 2 is expressed in the equation:

$$P(Y_i = Y\alpha \mid X_i = \frac{\exp(24.1933 - 6.3536X_i + 0.2601X_2 - 4.7772X_3 - 20.1883X_3 - 0.8897X_4}{1 + \exp(24.1933 - 6.3536X_1 + 0.2601X_2 - 4.7772X_3 - 20.1883X_3 - 0.8897X_4}$$

where $P(Y_i = Y\alpha \mid X_i)$ means the opportunities for positive patients with ARI infection with the value of the independent variable has. With X_1, X_2, X_3, and X_4 are age, gender, nutritional status, and family history of smoking. To interpret how the influence of significant variables that can be used the value of the ods ratio of each variable. The value of the ods ratio is a value exponential of the coefficient parameters of the regression are shown in table 3.

Table 3. The Value Of The Odds Ratio

Parameters	Estimase	*Ods* Rasio
Age 5years	-6.3536	0.0017

Based on table 3, it is known the value of the ods ratio of 0,0017 showed a trend of patients who have 5 years of age to positive suffering from acute respiratory infections. Furthermore, the prepared matrix confusion to find out the results of the predictions of the model and determine the level of accuracy of the predictions.

Table 4. Matrix Confusion Prediction Results of the Status of ARI

Prediction	Actual		Total
	Negative	Positive	
Negative ARI	71	7	78
Positive ARI	2	143	145
Total	73	150	223

The explanation of the table is a matrix of confusion above:

1) 73 is the data actually negative ARI
2) 150 is the data actually positive ARI
3) 78 is the data predicted negative ARI by model
4) 145 is a data predicted positive ARI by model
5) 71 is the data actually negative and predicted negatively by the model (right)
6) 2 is the actual data and negative predicted positive by the model (any)
7) 7 is the actual data and positive predicted negative by the model (any)
8) 143 is the data actually positive and predictable positive by the model (right)

Based on table 4, the model can predict the ARI with the correct total of 143 patients from 150 patients positive ARI and negative ARI correctly and by 71 patients from 73 patients who were not suffering from ARI. The value of the accuracy is calculated as follows:

$$Akurasi = \frac{truepositif + truenegatif}{N} = \frac{71+143}{223} = \frac{214}{223} = 0.9596.$$

The value of 0,9596 means that the model can predict the data accurately as 95.96%.

Conclusion and Suggestions

Based on the analysis, the conclusion that logistic regression model acute respiratory infections in the health center of Cikampak Torgamba can be applied. From the result test using the maximum likelihood method obtained significant results.

There are some suggestions from the researchers to the development and further research, the results test using the maximum likelihood method can be continued with using software as its analysis.

Reference

Azizah, S. Masrofatul. (2017). Model Regresi Logistik Pada Faktor-faktor Yang Mempengaruhi Imunisasi Lengkap Balita. *Jurnal Ilmiah Teknosains*. 3(2). pp. 73-76.

Burhan, S. Andi, K. J. (2018). Penaksiran Parameter Regresi Linier Logistik dengan Metode Maksimum *Likelihood Local* Pada Resiko Kanker Payudara di Makassar. *Jurnal Matematika*. 14(2). pp: 159-165.

https://pusdatin.kemkes.go.id. diakses pada 19 februari 2022 Pukul 20.00 WIB.

Kurniawan, Panji. S. (2018). Maximum Likelihood Estimator Untuk Mengestimasi Model Regresi Isotonik Dengan Pendekatan Polinomial Bernstain Pada Kasus Satu Variabel Independen. Skripsi. Yogyakarta: Universitas Islam IndonesiaYogyakarta.

Maakh, Yorida, F. (2017). Profil Pengobatan Infeksi Saluran Pernapasan Akut (ISPA) Pada Balita di Puskesmas Rambangaru Tahun 2015. *Jurnal Infos Kesehatan*. 15(2): 435-450.

Larasati, R., Diandra. (2018). *Penentuan Harga Opsi Eropa Menggunakan Metode Trinomial dengan Volatilitas Menggunakan Metode Maximum Likelihood Estimation (MLE)*. Skripsi. Makassar: Universitas Islam Negeri Alauddin.

Rahmat, Muhammad, F. (2016). *Penggunaan Metode Maximum Likelihood Dalam Estimasi Parameter Model ARCH(r)-Mean*. Thesis. Makassar: Universitas Negeri Makassar.

Safiah, F. Muhammad, N. (2019). Model Regresi Logistik Biner Pada Tingkat Pengangguran Terbuka di Provinsi Sulawesi Barat Tahun 2017. *Journal of Statistics and Its Application on Teaching and Research*. 1(2). pp: 1-6.

Saputra, Bambang, A. (2016). *Faktor-faktor Yang Mempengaruhi Jumlah Penderita Infeksi Saluran Pernafasan Akut (ISPA) di Kota Medan Menggunakan Analisis Jalur*. Skripsi. Medan: Universitas Sumatera Utara.

Sembiring, K. (2018). *Faktor-faktor Yang Mempengaruhi Kesejahteraan Rumah Tangga Di Kabupaten Deli Serdang Dengan Pendekatan Regresi Logistik*. Skripsi. Medan: Universitas Sumatera Utara.

Wahyuningsih, S. Sitti, R. Syahrul, B. 2017. Infeksi Saluran pernapasan Akut (ISPA) Pada Balita di Wilayah Pesisisr Desa Kore Kecamatan Sanggar Kabupaten Bima. *Jurnal Kesehatan Masyarakat*. 3(2): 98-105.