

Generalized Linear Models in Determining Factors Affecting the Number of Community Visits to Health Service with Bayesian Inference Approach

Selfinia¹, Dodi Devianto^{2*}, Ferra Yanuar³

^{1,2,3}Department of Mathematics and Data Science, Andalas University, Indonesia

selfinia95@gmail.com¹, ddevianto@sci.unand.ac.id², ferrayanuar@sci.unand.ac.id³

ABSTRACT

Article History:

Received : 14-05-2023

Revised : 14-07-2023

Accepted : 17-07-2023

Online : 18-07-2023

Keywords:

Generalized Linear

Models;

Health Service;

Bayesian Inference.



Public health plays an important role in achieving the Sustainable Development Goals (SDGs) set by the United Nations. The SDGs are a series of global targets and commitments aimed at addressing various challenges facing the world today, such as poverty, hunger, gender inequality, climate change, and others. Public health, as one of the important aspects of the SDGs, is closely linked to several sustainable development goals. Efforts made to achieve the SDGs in the health sector are to improve health services. The objective of this study was to identify factors that influence the number of community visits to health services. The data used is a small sample size as one hundred community respondents in Padang City, West Sumatra Province. In this study, the number of respondents' visits to health service was the measured variable, while the predictor variables consisted of five variables, namely the status of the implementation of clean and healthy living behavior, health history, distance to health services, type of insurance owned, and consumption patterns. The generalized linear models is used to identify predictor variables that have significance using the Bayesian inference approach. It was found that there are two predictor variables that are significant in influencing the number of community visits to health services, namely the consumption patterns of respondents and the health history of respondents. These two variables have a very dominant effect on the number of visits to health service facilities in Padang City. This result indicates the community has to pay attention to their consumption patterns and living behavior to prevent periodic disease outbreaks and take care of their health history factors.



<https://doi.org/10.31764/jtam.v7i3.15186>



This is an open access article under the **CC-BY-SA** license

A. INTRODUCTION

Sustainable Development Goals (SDGs) are sustainable development efforts that aim to maintain the improvement of peoples economic welfare. This development aims to maintain the sustainability of peoples social life and environmental quality, as well as ensuring justice and the implementation of good governance, so that there is an improvement in the quality of life from one generation to the next (Xu et al., 2023). The SDGs are global and national commitments to improve people's well-being by including several goals, such as eradicating poverty, overcoming hunger, achieving a healthy and prosperous life, improving the quality of education, realizing gender equality, providing access to clean water and proper sanitation, obtaining clean and affordable energy, creating decent work and economic growth, developing industry, innovation and infrastructure, reducing inequality, building sustainable cities and

settlements, promoting responsible consumption and production, dealing with climate change, protecting ocean ecosystems and terrestrial economies, promoting peace, justice and strong institutions, and establishing partnerships to achieve these goals (Kagesten et al., 2023; Ravn & Gonz, 2023).

One of the programs implemented to the SDGs in the health sector is to improve health services. To achieve the objectives of the program, it is necessary to know what factors influence the number of visits by the public to healthcare facilities. Health is very important for human survival. Health can be defined as the overall state of one's mind and body that indicates the absence of disease, injury, or health problems. To regain health, it is necessary to make visits to health services. In terms of quality aspect, it is found that the utilization rate of health service facilities by the community is still not running optimally. Health behavior includes all activities or actions taken by a person, whether directly observable or not, that are related to maintaining and improving health (Xu et al., 2023). Therefore, health behavior can be divided into healthy behavior, which is the actions people take to maintain or improve their health, and treatment-seeking behavior, which is the actions of people who experience illness or health problems to get healing or overcome their health problems.

Factors that affect individual or community health consist of two things, namely behavioral factors and non-behavioral factors. Behavior is influenced by factors such as age, type of work, education level, knowledge, health history, physical environmental factors and distance to health facilities, as well as reinforcement factors through the support of family and community leaders. On the other hand, non-behavioral factors have a very specific scope with many different indicators (Kagesten et al., 2023; Ravn & Gonz, 2023). In this study, we will discuss what factors influence the behavior of community visits to healthcare facilities. The response variable is the number of respondents' visits to health service, and the predictor variables are the type of health insurance, the distance from home to the health service, the respondents' consumption pattern, the respondents' medical history and the status of the house with clean and healthy living.

The statistical structure how to analyze the multivariate factor affecting the number of healthcare visits with linear effect can be performed by using generalized linear models (GLM). The GLM is used to assess and measure the relationship between dependent variables and independent variables (Avanzi et al., 2020). This model is an extension of the linear regression model by assuming that the predictors have linear effects, but it does not specify a particular distribution of the response variable. This model is used when the response variable belongs to the exponential family (Azreena et al., 2023; Bacha & Tadesse, 2019; Chen et al., 2023; Eguchi, 2018; Gohain & Jansson, 2022; Han, 2021; Jayaraman & Ramu, 2023; Khodadadi et al., 2022; Li et al., 2021; Rodriguez et al., 2022; Shimizu, 2023; Yang et al., 2023; Yanuar et al., 2019). In GLM, the dependent variable does not have to have strictly a normal distribution, but it must belong to an exponential family such as binomial, negative binomial, Poisson, gamma, or inverse Gaussian distribution (Diwidian et al., 2020). The GLM has several areas of application ranging from medicine to economics, quality control and sample surveys (Khuri et al., 2006). The GLM is a statistical method that has been widely used in research, such as to compare parameter estimation results with frequentist estimation methods, or models using Bayesian methods. In addition, the GLM is also applied to model Covid-19 cases in Indonesia, where the variables

used in this study are the number of confirmed Covid-19 cases per day and the number of deaths caused by Covid-19 per day. The purpose of this study is to observe how much impact the confirmation of Covid-19 cases has on the number of deaths that occur every day (Saidi et al., 2021).

There are two methods used to estimate parameters, namely the classical method, which is a classical statistical approach where parameter estimation is based on information from sample data and ignores initial information (*Prior*) and the Bayesian method (Hosack et al., 2017). In the Bayesian method, population parameters are viewed as variables that have an initial distribution (*prior*). Distribution *prior* is a subjective distribution based on one's beliefs about the parameters. In general, the main advantage of using the Bayes method is that it simplifies the classical method that requires the use of complex integrals in calculating the marginal model (Shimizu, 2023). The Bayesian method combines the information contained in the sample with the information contained in the prior distribution, and then produces a posterior distribution. If the posterior distribution cannot be calculated analytically, then a Markov Chain Monte Carlo (MCMC) approach is used. This algorithm uses an acceptance and rejection mechanism to generate a line of random samples. In the MCMC simulation, iterations are carried out until the random samples reach convergence. Samples can be seen from regular patterns on the Dynamic Trace Chart (Shi et al., 2023). In this study, the response variable as the number of visits to healthcare facilities has a Poisson distribution with a small sample size. The proposed model is Bayesian GLM to perform parameter estimation with an algorithm *Markov Chain Monte Carlo* (MCMC) by using Winbugs Software.

B. METHODS

In many statistical models, the generally dependent variable is assumed to be normally distributed, but in reality, it is often found that the dependent variable is not normally distributed (Haupt & Bittner, 2018). One solution to overcome this problem is by developing a linear model such as the GLM (Saputri & Devianto, 2020). The GLM is a generalized form of linear model that is often used Saidi et al. (2021), this model is a more flexible form of ordinary linear regression that describes the linear relationship between several independent variables and one dependent variable through a linking function (Rodriguez et al., 2022). The GLM is a statistical model extended from the linear regression model with the assumption that the predictor effects are linear, but does not require the assumption of a particular distribution of the response variable. This model is usually used when the response variable belongs to the exponential family (Bacha & Tadesse, 2019). The flow chart of the GLM process with Bayesian inference is described as in Figure 1.

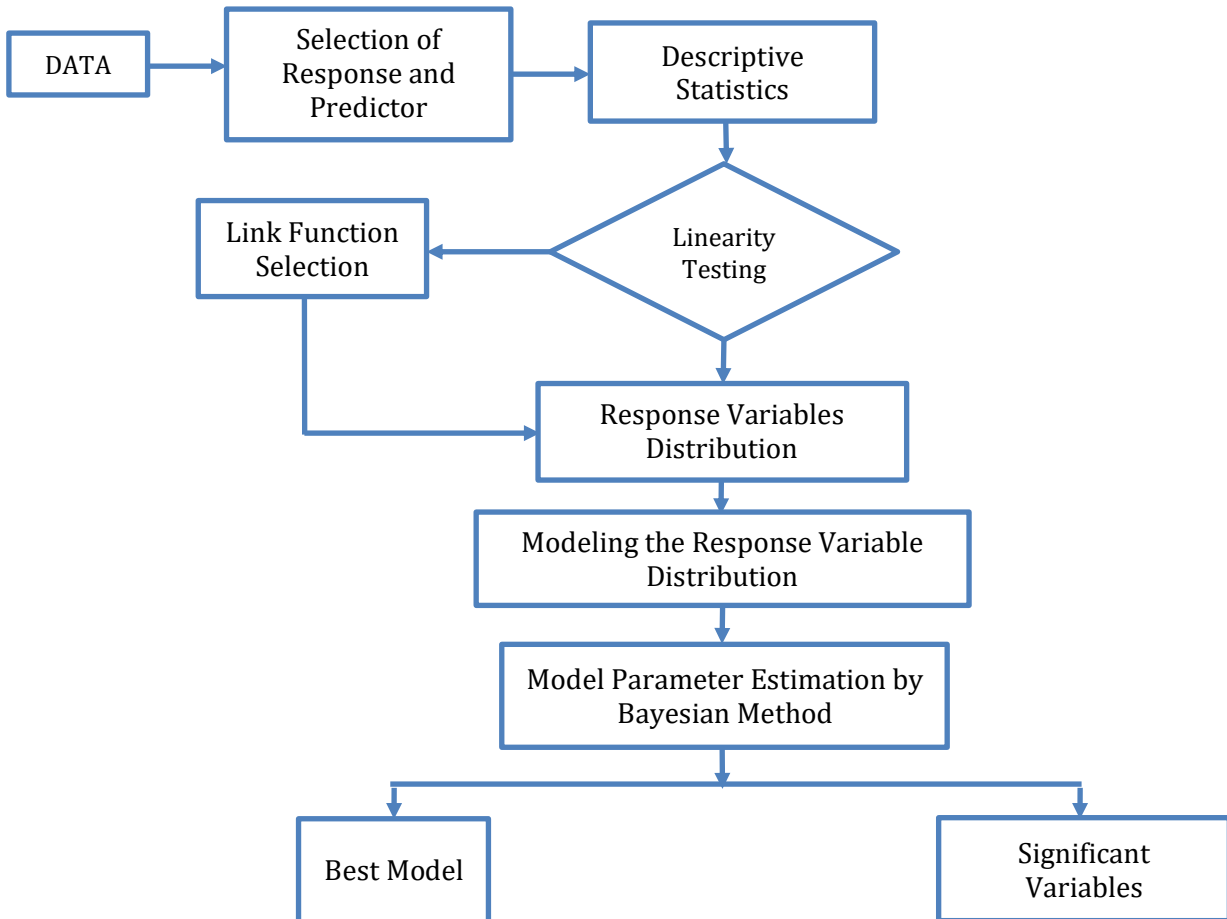


Figure 1. The flow chart of the GLM with Bayesian Estimation.

The GLM was illustrated with examples relating to four distributions; Normal, Binomial (probit analysis), Poisson (contingency table) and Gamma (variance component) (Saputri & Devianto, 2020). Given the formula form of the GLM, i.e. :

$$g(\mu_i) = \eta_i = \alpha + \sum_{j=1}^k \beta_j x_{ij}, \quad (1)$$

with link function $g(\mu_i)$ where $i = 1, 2, \dots, n$ are subjects for individual observations, j is the number of independent variables in the observation, x_{ij} is the i -th observation of the j -th independent variable, and the parameter β_j is the unknown coefficient of the observation. The GLM combines various methods in statistical sciences such as regression, ANOVA, and models for categorical data are special cases of one good model. There are 3 components in the GLM, namely random components, systematic components, and connecting functions (Diwidian et al., 2020). The random component of the GLM identifies the component consisting of the dependent variable Y expressed on mutually independent observations (y_1, y_2, \dots, y_n) and selects the distribution of the study data. In some cases each observation is a count data having a distribution from the exponential family which has the following probability density function

$$f(y; \theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\theta)} + c(y, \phi)\right), \tag{2}$$

with θ_i is canonical parameter, ϕ_i is dispersion parameter, where the value of parameter θ_i with $i = 1, 2, \dots, n$, depends on each value of the independent variable. The systematic component of the GLM specifies the dependent variable linearly as the predictor on the right-hand side of the model equation. In this case connecting vectors $(\eta_1, \eta_2, \dots, \eta_n)$ for the independent variables, suppose x_{ij} is an independent variable value, with $j = 1, 2, \dots, k$ for $i = 1, 2, \dots, n$ then it is obtained

$$\eta_i = \alpha + \sum_{j=1}^k \beta_j x_{ij} \tag{3}$$

This linear combination of independent variables is associated with linear independence or mutual independence which are referred to as linear predictors. The third component is the linking function that connects the random and systematic components. The linking function defines a function $g(\cdot)$ that relates μ_i to a linear predictor η_i denoted $g(\mu_i) = \eta_i$ so that it has the form as follows

$$g(\mu_i) = \eta_i = \alpha + \sum_{j=1}^k \beta_j x_{ij} \tag{4}$$

The link function $(\mu_i) = \mu_i$, is said to be the identity link, if $\eta_i = \mu_i$. This implies a linear model for the mean response, which is expressed as follows

$$\mu_i = \alpha + \sum_{j=1}^k \beta_j x_{ij} \tag{5}$$

If the link function $g(\mu) = \ln(\mu)$ models the log probability, it is called the logit function. The link functions for the GLM model are stated as in Table 1.

Table 1. Link function for the GLM model.

Distribution	Link	Link Function	Link Function Invers	Variance
Normal	Identity	Identity	Identity	1
Binomial	Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\exp(\eta) / [1 + \exp(\eta)]$	$\mu(1-\mu)/\mu$
Poisson	Log	$\ln(\mu)$	$\exp(\eta)$	μ
Gamma	Invers	$1/\mu$	$1/\eta$	μ^2

In this study, to estimate the parameters of the GLM, we used the Bayesian due to relatively a small sample size for community visits to healthcare facilities at West Sumatra. The Bayesian

approach to parameter estimation has the added value that unlike traditional statistical methods, it integrates different types of information into the estimation process, such as expert judgment and statistical data (Jayaraman & Ramu, 2023). This method combines the likelihood function with the prior distribution, and the Bayesian inference method provides posterior probability estimates (Ota et al., 2023). The simulation Method with the MCMC algorithm is performed with the Winbugs application as a powerful tool to analyze the framework of the Bayesian theory for complex probabilistic Markov models (Youn et al., 2022). The data used in this research originating from one hundred respondents in Padang City, West Sumatra Province, who filled out questionnaires in 2022. In this article, the response variable is the number of visits by respondents to healthcare facilities, The variables used as predictors in this article can be found in Table 2.

Table 2. Definition of Predictor Variables.

Variable	Description
X_1	Types of health insurance consist of government insurance, other insurance, not having health insurance.
X_2	The distance from home to health services consists of less than 2 km, 2 km-5 km, and more than 5 km.
X_3	Consumption patterns consist of always eating healthy food, occasionally eating fast food, and there are no rules for always eating healthy.
X_4	The medical history consists of never being seriously ill, ever seriously ill, and often seriously ill.
X_5	The status of the house with clean and healthy living.

For some predictor variables in Table 2 that are categorical variables, it is necessary to convert the variables into dummy variables. This conversion can be seen in Table 3.

Table 3. Definition of Predictor Dummy Variables.

Variable	Description
$X_{1,1}$	Dummy variable first types of health insurance
$X_{1,2}$	Dummy variable second types of health insurance
$X_{2,1}$	Dummy variable first distance from home to health service
$X_{2,2}$	Dummy variable second distance from home to health service
$X_{3,1}$	Dummy variable first consumption pattern
$X_{3,2}$	Dummy variable second consumption pattern
$X_{4,1}$	Dummy variable first medical history
$X_{4,2}$	Dummy variable second medical history
X_5	Dummy variable status of the house with clean and healthy living

The following are some of the steps taken in data processing to determine the predictor variables that are significant to the response variable of the number of community visits to health services in West Sumatra.

1. Perform descriptive statistical analysis of the response variable regarding the number of community visits to health facilities.
2. Determine the factors that influence the number of visits made by the community to healthcare facilities using the GLM with parameter estimation using the Bayesian inference approach, and the stages are as follows.

- a. Identify the prior distribution and likelihood function of the data used;
 - b. Determine the posterior of the GLM;
 - c. Estimating parameters with the MCMC and the Gibbs sampler method as many as T iterations;
 - d. Test the convergence of each model parameter. If the model parameters have not converged, then the MCMC process is repeated by adding iterations until all the parameters of the GLM have converged;
 - e. Test the significance of the model parameters;
 - f. Significance test of predictor variables which aims to check whether the predictor variable has a significant influence on the response variable.
 - g. Test the convergence of each model parameter;
 - h. After knowing which parameters have a significant effect on the response variable, then conduct convergence testing again as in section d.
 - i. The selection of the best model.
 - j. After the convergence and significance conditions are met, parameter estimates will be obtained to form the best model.
3. Make conclusions on the data analysis that has been done based on Bayesian inference approach of the GLM.

C. RESULT AND DISCUSSION

In this study, to analyze the factors that influence the number of respondents visits to health service facilities, a regression analysis method of the GLM is used. This method is used to find out what factors affect the number of respondents visits to health service. The logit function is used as a link function to connect p_i to the linear predictor ($X_i\beta$) in the process of analyzing the data in this case as follows:

$$\begin{aligned}
 g(\mu_i) &= \ln(\mu_i) \\
 \alpha + \sum_{j=1}^k \beta_j x_{ij} &= \ln(p_i) \\
 \exp\left(\alpha + \sum_{j=1}^k \beta_j x_{ij}\right) &= p_i
 \end{aligned} \tag{6}$$

One of the methods used to estimate the GLM parameters is the Bayesian method. The Bayesian method is a method used to estimate parameters by combining the information contained in the sample with other information that has been previously available. In the Bayesian method, when a population follows a certain distribution with a parameter in it (suppose in this case θ), then the probability of the parameter θ follows a certain probability distribution called the prior distribution. Bayesian methods are based on the Bayesian theorem. The Bayesian method combines the likelihood function and the prior distribution of the parameters to obtain the posterior distribution which then becomes the basis for estimating the parameters. Markov Chain Monte Carlo (MCMC) is a simulation technique commonly used

in the Bayes method. MCMC is a simulation method to obtain sample data of a random variable with sampling technique based on Markov chain. One of the famous MCMC method techniques is Gibbs Sampler. Gibbs Sampler uses conditional distribution to generate sample data of random variables. The MCMC method is effective enough to determine the estimated value of the parameters of the posterior distribution which is very complex and quite difficult if solved by other methods. Monte Carlo simulation is an approach to estimate the distribution function of a random variable. The simulation method is Gibbs Sampler, which is a method that uses a fully conditional distribution that is associated with a stationary distribution. The Gibbs Sampler process takes steps such as selecting the initial value then simulating the sampling of random variables based on the full conditional distribution by iterating n times (Bacha & Tadesse, 2019). Test whether the response variable, namely the number of visits made by respondents to health service, follows a Poisson distribution. This test is conducted using the Kolmogorov-Smirnov test, with the following hypothesis.

H_0 : response variable data with Poisson Distribution

H_1 : response variable data not with Poisson Distribution

In this testing process, SPSS software was used to obtain the statistic test of the Kolmogorov-Smirnov as follows:

$$D_{count} = \text{Maximum } |F_o(x) - F_r(x)| \quad (7)$$

where $F_o(X)$ is observed cumulative frequency distribution of a random sample of n observations and $F_r(X)$ is the theoretical frequency distribution, in this case as the Poisson distribution. The results of statistic test as in Table 4.

Table 4. Poisson Distribution Test Results

D_{count}	Asymp Sig (2-tailed)
0.066	0.771

Based on Table 4, the values of $D_{count} = 0.066$ and values of Asymp Sig. (2-tailed) = 0.771. The value of $D_{table} = 0.134$ with a test level of $\alpha = 0.05$ and $n = 100$. So that we have the value of $D_{count} < D_{table}$ and the value of Asymp Sig. (2-tailed) $> \alpha$, then it can be concluded that H_0 does not reject. Thus, it can be concluded that the data regarding the number of visits made by respondents to health facilities has a distribution that follows the Poisson Distribution.

1. Multicollinearity Test

The way to determine the presence of multicollinearity in the predictor variables is by looking at the VIF and Tolerance values of each predictor variable, such as in the output in Table 5.

Table 5. The Tolerance Value and VIF for Each Predictor Variable.

Variabel	Tolerance	VIF
$X_{1,1}$	0.893	1.120
$X_{1,2}$	0.930	1.076
$X_{2,1}$	0.956	1.046
$X_{2,2}$	0.907	1.103
$X_{3,1}$	0.812	1.232
$X_{3,2}$	0.716	1.397
$X_{4,1}$	0.726	1.378
$X_{4,2}$	0.531	1.884
X_5	0.471	2.124

From Table 5 above, it can be seen that all VIF values for these variables are $X_{1,1} = 1.120$, $X_{1,2} = 1.076$, $X_{2,1} = 1.046$, $X_{2,2} = 1.103$, $X_{3,1} = 1.232$, $X_{3,2} = 1.397$, $X_{4,1} = 1.378$, $X_{4,2} = 1.884$, $X_5 = 2.124$, it can be concluded that there is no multicollinearity in all predictor variables, because all VIF values are less than the specified value, as seen in Table 5 above. Other than that, all grades Tolerance value of each predictor variable is smaller than 10 and all the VIF values of each predictor variable are greater than 0.10, so it can be concluded strengthens the assumption that it can be seen that all predictor variables are mutually independent, multicollinearity does not occur, so that it can be continued with the stages of model analysis to obtain the proper modeling.

2. Data Analysis

In this study, regression analysis with the GLM method was used to determine the factors that influence the number of respondents visits to health service. Based on Equation 6, a logit function is obtained which is used as a link function to connect p_i to linear predictors in the process of analyzing the data of this case as follows.

$$p_i = e^{\alpha + X_{1,1}\beta_{1,1} + X_{1,2}\beta_{1,2} + X_{2,1}\beta_{2,1} + X_{2,2}\beta_{2,2} + X_{3,1}\beta_{3,1} + X_{3,2}\beta_{3,2} + X_{4,1}\beta_{4,1} + X_{4,2}\beta_{4,2} + X_5\beta_5} \tag{8}$$

From the obtained equation form, the next step is to estimate the model parameters by using the Bayesian method. The Bayesian method on Generalized Linear Models is done using Winbugs Software. The simulation process begins with building a model and then iterating with the Gibbs Sampler through the Winbugs Application. The parameter estimation of the number of respondent visits to healthcare facilities is displayed on the density in Figure 2.

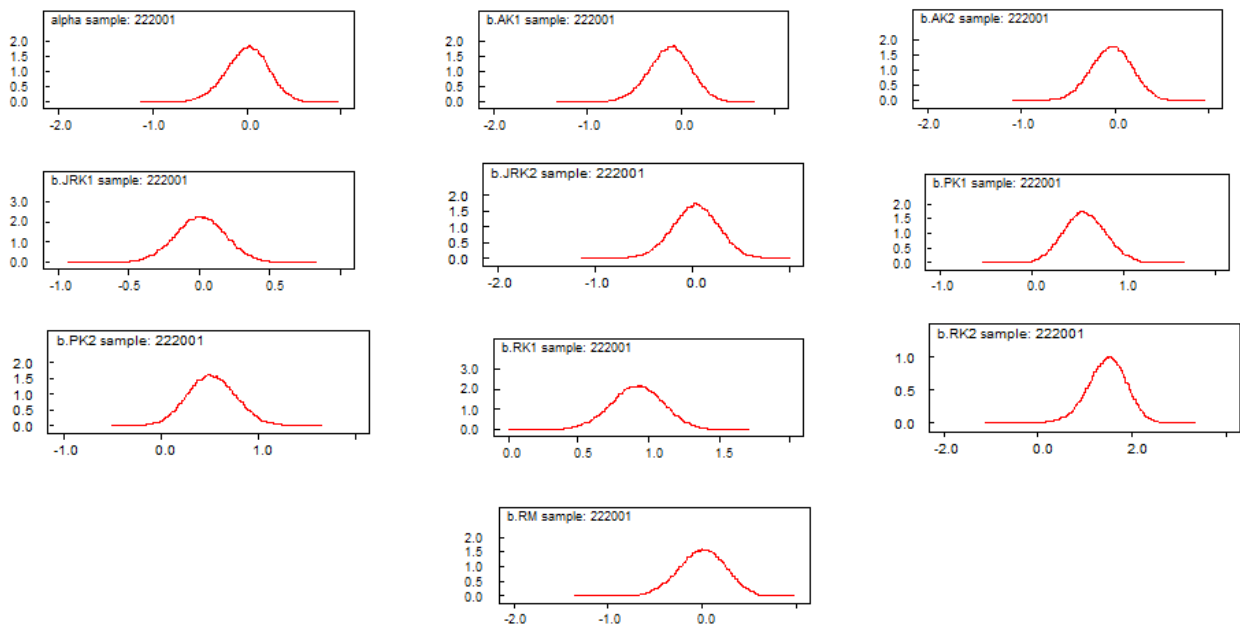


Figure 2. The Posterior Density Plot for Estimating Factors That Affect The Number of Visits to Healthcare Services.

In Figure 2 above it can be seen that *Plot Density* for constants (α), parameters two dummy variables for the type of health insurance ($b.AK1$, $b.AK2$), parameters two dummy variables for the distance of the respondent's house to the health service ($b.JRK1$, $b.JRK2$), parameters the two dummy variables for consumer pattern ($b.PK1$, $b.PK2$), parameters the two dummy variables for health history ($b.RK1$, $b.RK2$), and parameters respondent's home status has quite good ($b.RM$) results because it has a pattern that tends to be smooth and has the shape of a bell curve the posterior distribution formed for parameters in the GLM model is normally distributed. Furthermore, the convergence test was carried out using Trace Plot. Trace Plot is one way to test the convergence of the resulting parameter estimates for the number of iterations performed, as shown in Figure 3.

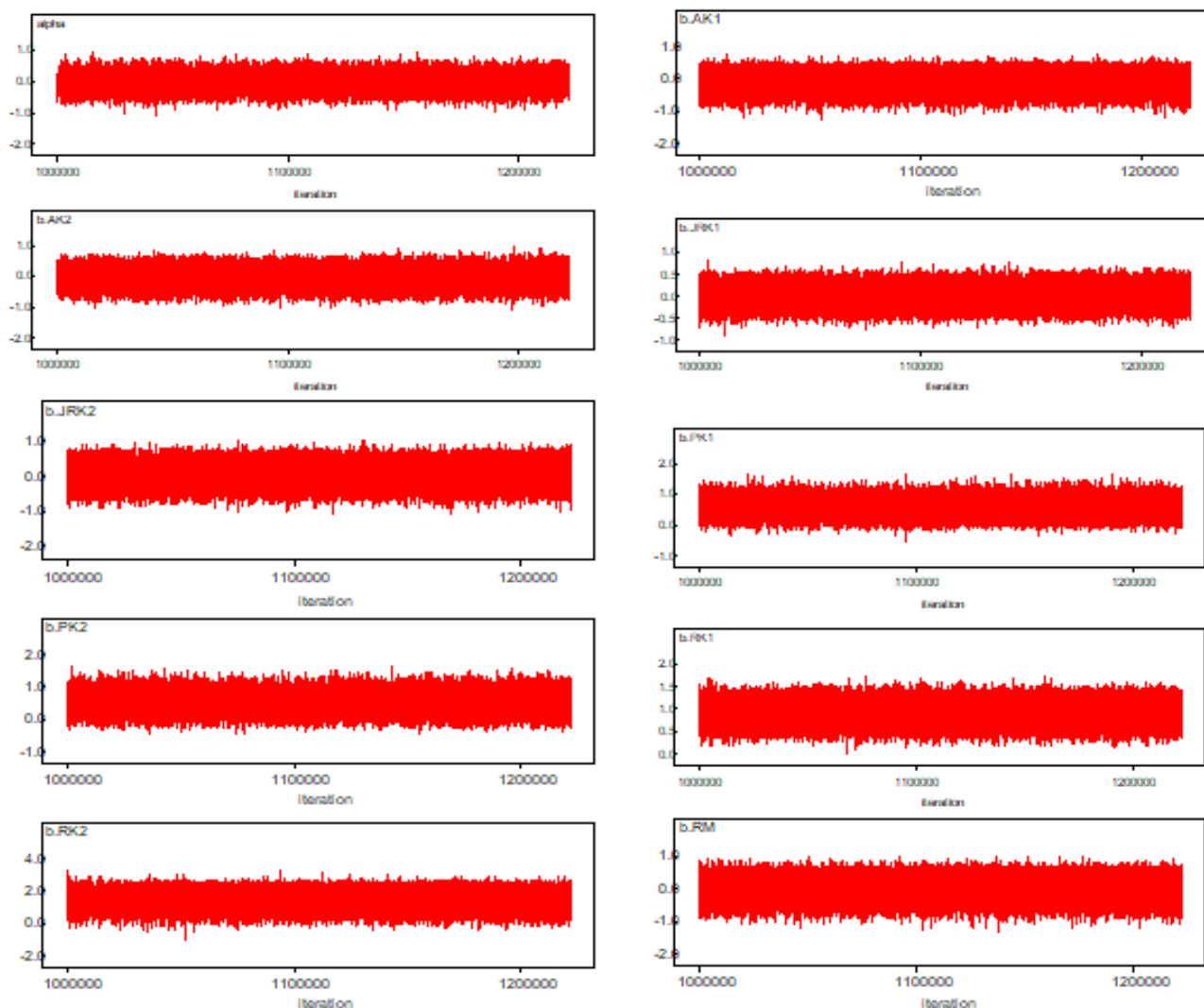


Figure 3. The History Trace Plot for Estimating Factors That Affect the Number of Community Visits to Healthcare Services.

Based on Figure 3, it can be seen that the distribution of the data can be observed for constants (α), parameters two dummy variables for the type of health insurance ($b.AK1$, $b.AK2$), parameters two dummy variables for the distance of the respondents house to the health service ($b.JRK1$, $b.JRK2$), parameters the two dummy variables for consumer pattern ($b.PK1$, $b.PK2$), parameters the two dummy variables for health history ($b.RK1$, $b.RK2$), and parameters for respondent’s home status has quite good ($b.RM$) has been constant and stable which lies between the estimation intervals, so that it can be concluded that the convergence test is fulfilled. The next step is to test the significance of the predictor variables, with the aim of checking whether the predictor variables have a significant influence on the response variable. The significance of parameters in Bayesian can be evaluated by looking at confidence intervals. Based on the analysis, the results of the parameter significance test are shown in Table 6.

Table 6. Parameter Significance Test Result for the First Model.

Parameters	Confidence Interval		Description
	5%	95%	
$alpha (\alpha)a (\alpha)$	-0.367	0.358	Not Significant
$b.AK1 (\beta_{1,1})$	-0.489	0.237	Not Significant
$b.AK2(\beta_{1,2})$	-0.407	0.324	Not Significant
$b.JRK1 (\beta_{2,1})$	-0.285	0.295	Not Significant
$b.JRK2 (\beta_{2,2})$	-0.362	0.420	Not Significant
$b.PK1(\beta_{3,1})$	0.204	0.971	Significant
$b.PK2(\beta_{3,2})$	0.126	0.947	Significant
$b.RK1(\beta_{4,1})$	0.608	1.213	Significant
$b.RK2(\beta_{4,2})$	0.774	2.112	Significant
$b.RM(\beta_5)$	-0.428	0.404	Not Significant

Based on Table 6 above, stating that the parameters $\beta_{3,1}$, $\beta_{3,2}$, $\beta_{4,1}$, $\beta_{4,2}$ do not contain zero in the confidence interval, meaning that the variables $X_{3,1}$, $X_{3,2}$, $X_{4,1}$, $X_{4,2}$ has a significant influence on the response variable Y . After knowing the parameters that have a significant influence on the response variable, the next step is to find a new regression model by eliminating the variables that are not significant to the response variable. This process involves 111,000 iterations. Furthermore, the significance of the parameters that were significant in the previous stage was tested again, with the results obtained in Table 7.

Table 7. Confidence Interval of Parameters for Final Model.

Parameters	Confidence Interval	
	5,0%	95,0%
$b.PK1(\beta_{3,1})$	0.398	0.776
$b.PK2(\beta_{3,2})$	0.331	0.729
$b.RK1(\beta_{4,1})$	0.620	1.179
$b.RK2(\beta_{4,2})$	0.820	2.079

In Table 7 above, it can be seen that influence testing can be seen from the interval value 5,0% up to 95,0%. In this range, it can be concluded that the predictor variable has a significant influence on the response variable because there is no zero value in these parameters. Furthermore, the convergence test of the parameters that have been significant is carried out by paying attention to the history trace plot so that the following output is obtained, as shown in Figure 4.

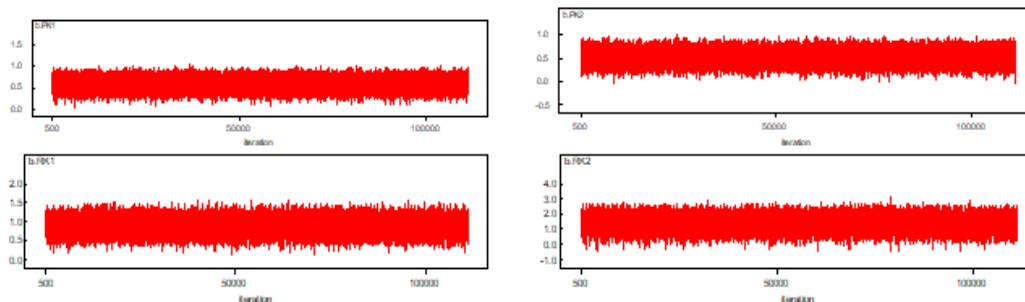


Figure 4. The History Trace Plot Parameters that have been Significant

Based on the history trace plot above, it is known that the variables have converged, then the variable parameter estimation is carried out with the help of Winbugs software, the parameter estimation results are obtained in Table 8.

Table 8. Parameter Estimation MC error for Final Model.

Parameters	Mean	Sd	MC error
$b.PK1(\beta_{3,1})$	0.590	0.115	0.385
$b.PK2(\beta_{3,2})$	0.534	0.121	0.401
$b.RK1(\beta_{4,1})$	0.905	0.170	0.629
$b.RK2(\beta_{4,2})$	1.489	0.385	0.001

Based on the parameter estimation results above, the Generalized Linear Models equation for the opportunity for community visits to healthcare facilities is obtained as follows:

$$p_i = \exp(0,5898 X_{3,1} + 0,5338 X_{3,2} + 0,9051 X_{4,1} + 1,488 X_{4,2}) \tag{9}$$

with $b.PK1(\beta_{3,1})$ and $b.PK2(\beta_{3,2})$ is a parameter for the dummy variables of the consumption pattern as well and $b.RK1(\beta_{4,1})$ and $b.RK2(\beta_{4,2})$ is a parameter for the dummy variable of the respondent’s health history. It can be interpreted that the predictor variables that have a significant influence on the number of community visits to health services are respondents consumption patterns and medical history.

In previous studies, Generalized Linear Models have been widely used to analyze various problems including the analysis of the number of tuberculosis patients and the identification of factors that cause malnutrition in the nutritional status of children in Ethiopia. Therefore, the author conducted a study to analyze the factors that have a significant influence on the number of community visits to health services. Based on the results of the analysis, the respondents consumption patterns and medical history variables have a very dominant influence on community visits to health centers in Padang City. Respondents' consumption patterns have an influence because poor consumption patterns can increase the risk of developing various dangerous diseases. The more people who consume with bad patterns, the more people will suffer from diseases, so the number of visits to health services will be higher than usual. Respondents who have a history of chronic diseases will use health services more to maintain

their health. In addition, for respondents who have a health history in the form of illness, it can have an influence on the number of visits to health services because respondents will more often visit existing health services so that respondents can better control their health.

These two variables have very dominant affect the community visits to health service in Padang City. Respondents consumption patterns have an influence because consumption patterns that are not good can increase the risk of various dangerous diseases. The more people consume by wrong patterns, the more people will suffer from diseases, and then the number of visits to health services will be higher than as usual. Respondents who have a history of chronic disease will use more health service places to maintain a healthy body. In addition, for respondents who have a medical history in the form of illness, it can have an influence on the number of visits to health service providers because respondents will visit existing health service locations more frequently so that respondents can exercise better health control.

D. CONCLUSION AND SUGGESTIONS

Based on the analysis using the GLM with Bayesian parameter estimation, it can be concluded which of only two variables tested has a significant influence on the number of respondents visiting health service facilities in Padang city among five hypothesized variables, namely the consumption patterns of respondents and the health history of respondents. These two variables have very dominant affect the community visits to health services in Padang City, and this indicates the community has to pay attention to their consumption patterns behavior, since the worse behavior of consumption patterns then it will increase the risk of outbreaks of dangerous diseases. Health history also has an impact on the frequency of visits to healthcare centers, as individuals with more medical history tend to have their health conditions checked more often, so they have to take better health control during that time.

ACKNOWLEDGEMENT

The authors would like to express the gratitude to the steering reviewers for their valuable suggestions and inputs that have helped to improve the quality of this article. These inputs on this article is very helpful and it can be organized better and more completely.

REFERENCES

- Avanzi, B., Taylor, G., Anh, P., & Wong, B. (2020). Insurance : Mathematics and Economics A multivariate evolutionary generalised linear model framework with adaptive estimation for claims reserving. *Insurance: Mathematics and Economics*, 93, 50–71.
<https://doi.org/10.1016/j.insmatheco.2020.04.007>.
- Azreena, N., Agostino, C. D., Laureshyn, A., & Várhelyi, A. (2023). *Modelling of motorcyclists ' risky behaviour at an urban T-junction using generalised linear model : An exploratory study*. 47, 94–104.
<https://doi.org/10.1016/j.iatsr.2023.02.001>.
- Bacha, R. H., & Tadesse, M. (2019). Bayesian generalized linear model for identifying predictors of child nutritional status in Ethiopia. *Biometrics & Biostatistics International Journal*, 8(2), 65–74.
<https://doi.org/10.15406/bbij.2019.08.00273>
- Chen, X., Li, H., & Zhang, J. (2023). Complete subset averaging approach for high-dimensional generalized linear models ☆. *Economics Letters*, 226, 111084.
<https://doi.org/10.1016/j.econlet.2023.111084>.
- Diwidian, F., Kurnia, A., & Sadik, K. (2020, January 21). *Analysis of Bayesian Generalized Linear Models on the Number of Tuberculosis Patients in Indonesia with R*.

- <https://doi.org/10.4108/eai.2-8-2019.2290479>.
- Eguchi, S. (2018). Econometrics and Statistics Model comparison for generalized linear models with dependent observations. *Econometrics and Statistics*, 5, 171–188.
<https://doi.org/10.1016/j.ecosta.2017.04.003>.
- Gohain, P. B., & Jansson, M. (2022). Scale-Invariant and consistent Bayesian information criterion for order selection in linear regression models. *Signal Processing*, 196, 108499.
<https://doi.org/10.1016/j.sigpro.2022.108499>.
- Han, H. (2021). Exploring the association between compliance with measures to prevent the spread of COVID-19 and big five traits with Bayesian generalized linear model. *Personality and Individual Differences*, 176. <https://doi.org/10.1016/j.paid.2021.110787>.
- Hosack, G. R., Hayes, K. R., & Barry, S. C. (2017). Prior elicitation for Bayesian generalised linear models with application to risk control option assessment. *Reliability Engineering and System Safety*, 167(June 2016), 351–361. <https://doi.org/10.1016/j.res.2017.06.011>.
- Haupt, J. W., & Bittner, J. L. (2018). Analyzing thresholds and efficiency with hierarchical Bayesian logistic. *Vision Research*, 148(March), 49–58.
<https://doi.org/10.1016/j.visres.2018.04.004>.
- Jayaraman, D., & Ramu, P. (2023). L-moments and Bayesian inference for probabilistic risk assessment with scarce samples that include extremes. *Reliability Engineering and System Safety*, 235(July 2022), 109262. <https://doi.org/10.1016/j.res.2023.109262>.
- Kagesten, A. E., Wallengren, E., Newby, H., Moller, A., Marsh, A. D., Regina, G., Garrison, M. M., & Orlich, F. (2023). Platform Abstracts / Journal of Adolescent Health 72 (2023) S28 e S71. *Journal of Adolescent Health*, 72(3), S48.
<https://doi.org/10.1016/j.jadohealth.2022.11.102>.
- Khodadadi, A., Tsapakis, I., Shirazi, M., Das, S., & Lord, D. (2022). Derivation of the Empirical Bayesian method for the Negative Binomial-Lindley generalized linear model with application in traffic safety. *Accident Analysis and Prevention*, 170(September 2021), 106638.
<https://doi.org/10.1016/j.aap.2022.106638>.
- Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: A review. In *Statistical Science* (Vol. 21, Issue 3, pp. 376–399).
<https://doi.org/10.1214/088342306000000105>.
- Li, N., Peng, X., Kawaguchi, E., Suchard, M. A., & Li, G. (2021). Journal of Statistical Planning and Inference A scalable surrogate L0 sparse regression method for generalized linear models with applications to large scale data. *Journal of Statistical Planning and Inference*, 213, 262–281.
<https://doi.org/10.1016/j.jspi.2020.12.001>.
- Ota, Y., Jiang, Y., & Maki, D. (2023). Parameters identification for an inverse problem arising from a binary option using a Bayesian inference approach. *Results in Applied Mathematics*, 17, 100353.
<https://doi.org/10.1016/j.rinam.2022.100353>.
- Ravn, E., & Gonz, A. (2023). *Motivating a change in environmental assessment practice: Consultant perspectives on SDG integration*. 101(April).
<https://doi.org/10.1016/j.eiar.2023.107105>.
- Rodriguez, E. V., Dinora, A., & Chavez, G. (2022). Application of the generalized linear model to enable refractive index measurement with thermal sensitive interferometric sensors. *Optics Communications*, 524(April), 128765.
<https://doi.org/10.1016/j.optcom.2022.128765>.
- Saidi, S., Herawati, N., & Nisa, K. (2021). Modeling with generalized linear model on covid-19: Cases in Indonesia. *International Journal of Electronics and Communications System*, 1(1), 25–33.
<https://doi.org/10.24042/ijecs.v1i1.9299>.
- Saputri, U., & Devianto, D. (2020). The model of life insurance claims with actuarial smoothing approach by using GLM Poisson regression. *AIP Conference Proceedings*, 2296.
<https://doi.org/10.1063/5.0032179>.
- Shi, P., Yang, T., Yong, B., Xu, C., Li, Z., Wang, X., Qin, Y., & Zhou, X. (2023). Some statistical inferences of parameter in MCMC approach and the application in uncertainty analysis of hydrological simulation. *Journal of Hydrology*, 617(PA), 128767.
<https://doi.org/10.1016/j.jhydrol.2022.128767>.

- Shimizu, K. (2023). Asymptotic properties of Bayesian inference in linear regression with a structural break. *Journal of Econometrics*, 235(1), 202–219.
<https://doi.org/10.1016/j.jeconom.2022.03.006>.
- Xu, Z., Peng, J., Liu, Y., Qiu, S., Zhang, H., & Dong, J. (2023). Exploring the combined impact of ecosystem services and urbanization on SDGs realization. *Applied Geography*, 153(February), 102907.
<https://doi.org/10.1016/j.apgeog.2023.102907>.
- Yang, Y., Yu, C., & Zhong, R. Y. (2023). Advanced Engineering Informatics Generalized linear model-based data analytic approach for construction equipment management. *Advanced Engineering Informatics*, 55(August 2022), 101884.
<https://doi.org/10.1016/j.aei.2023.101884>.
- Yanuar, F., Zetra, A., Muharisa, C., Devianto, D., Putri, A. R., & Asdi, Y. (2019). Bayesian Quantile Regression Method to Construct the Low Birth Weight Model. *Journal of Physics: Conference Series*, 1245(1). <https://doi.org/10.1088/1742-6596/1245/1/012044>.
- Youn, J. H., Petersohn, S., Gittfried, A., Jevdjevic, M., Ainsworth, C., & Piena, M. (2022). C0174 A Comparison of Stan Versus WinBUGS Software for Conducting Bayesian Network Meta-Analyses of Binary Outcomes. *Value in Health*, 25(12), S51.
<https://doi.org/10.1016/j.jval.2022.09.250>.