Bayesian logistic regression and its application for hypothyroid prediction in post-radiation nasopharyngeal cancer patients

P A Lukman¹, S Abdullah¹ and A Rachman²

¹Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
²Department of Medical Education, Faculty of Medicine, Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: sarini@sci.ui.ac.id

Abstract. Logistic regression models are commonly used to model response variables in the form of categorical variables with several predictor variables. The contribution of the predictor variable to the response variable is expressed through a regression coefficient (β). Therefore, it is necessary to estimate β. This study discusses the estimation of β using the Bayesian method. Bayesian approach utilizes a combination of information from sample data and prior information about the characteristics of the parameters of interest, resulting in the updated information, namely the posterior. Bayesian method thus can overcome the problem if the quality of the sample data does not support observation. Bayesian logistic regression method will be used in analyzing post-radiation nasopharyngeal cancer (NPC) patient data, using measurement on Zulewski’s score components. Markov Chain Monte Carlo with Gibbs Sampling were used to obtain the sample from posterior distribution. Convergent estimates were obtained, and the result showed that Zulewski’s component scores only were not enough to explain the hypothyroidism in NPC. Additional information is required in order to explain the incidence of hypothyroidism in NPC.

Keywords: Bayesian logistic regression, Gibbs sampling, logistic regression, Markov Chain Monte Carlo, nasopharyngeal cancer.

1. Introduction

Regression analysis is a process of looking for mathematical models that are most suitable with data that aims to study the form of the relationship between one or more explanatory variables with one variable commonly called response variable [1]. One type of regression analysis model is the logistic regression model. A logistic regression model is a regression model with a response variable that has only two possible values, (i.e. a binary response variable), that is 0 and 1. The logistic regression model can be used if the response variable is a categorical variable. For example, the response variable in the logistic regression model has a value of 1 indicates success and 0 indicates failure.

Logistic regression models are often used to analyze medical data in particular, classification and prediction of medical data. Peretti et al. used logistic regression to predict breast cancer tumors in data.
sets with 569 cases [2] and obtain an overall accuracy of 85 %. Barco et al. used logistic regression in data sets from 1254 breast cancer patients to predict high tumor burden / HTB [3].

In regression model, the regression coefficients play an important role, as the calculation of relative risk and model interpretation are based on these coefficients. Thus, estimation of these parameters should be conducted in a way that the resulting estimated values could produce accurate prediction and insightful interpretation. There are two methods of parameter estimation, namely the frequentist method and the Bayesian method. The frequentist method is a parameter estimation method that only uses information from sample data, so the frequentist method is highly dependent on data quality.

When the data used does not support observation, the accuracy of the frequentist method is questionable. Therefore, it is necessary to use other information that can support the data, namely prior information about the parameters to be estimated. Bayesian approach uses both sources of information: the sample data and prior information. By this approach, if there is a lack of confidence in sample data, then the incorporation of prior distribution using Bayes rule still produce an optimal result than just relying on the sample, as proved in [4–7].

In the Bayesian method, the regression coefficient is treated as a random variable. Information from observational data is summarized in the likelihood function. Then prior information and data information will be combined and produce posterior information. Furthermore, drawing conclusions about the parameters to be estimated will be based on the posterior information. Posterior information can be obtained in the form of closed form and non-closed form [8]. If posterior information is a probability density function of a particular distribution, posterior information is said to be closed-form. The advantage of obtaining closed form posterior information is that the Bayes estimator can be obtained without using complex computational techniques. If non-closed form posterior information is obtained, computational techniques are needed to obtain Bayes estimators such as the Markov Chain Monte Carlo (MCMC) method.

MCMC is a class of sampling techniques from probability distributions and can be used to estimate the distribution of parameters if a set of observations is given. The purpose of this approach is to produce a sample of unknown parameter values from the posterior distribution. There are three MCMC algorithms that are commonly used to estimate the posterior distribution of parameters in the Bayesian MCMC model, namely Metropolis, Gibbs, and Hamiltonian [9]. Gibbs algorithm will be used in this paper. Gibbs Sampling is favored in settings because decomposition into conditionals is easy to implement and fast to run.

The method described will be applied to data of nasopharyngeal cancer (NPC) patients who undergone radiation therapy. Nasopharynx is one part of the upper throat which is located behind the nose and behind the roof of the oral cavity. Nasopharyngeal cancer (NPC) is included in the type of head and neck cancer. NPC occupies the fourth position of the most cancers in Indonesia, after cervical cancer, breast cancer, and skin cancer with a number of new cases around thirteen thousand per year [10].

According to the National Guidelines for Nasopharyngeal Cancer Medicine Services (PNPKNF) Ministry of Health of the Republic of Indonesia, therapies that can be performed for NPC are radiotherapy, chemotherapy, and a combination of both [11]. Radiotherapy alone or supplemented with surgery or chemotherapy, results in a significant increase in the cure rate for head and neck cancer. However, high doses of radiation over large areas, including the oral mucosa, can cause several undesirable reactions that occur during or after the completion of therapy [12]. Radiation given for NPC can cause endocrine complications because it can affect the pituitary gland (pituitary) and thyroid gland which can cause long-term complications such as hypopituitarism and hypothyroidism.

In this paper Zulewski's score will be used to diagnose thyroid dysfunction. Zulewski's scoring measurements consists of 12 components: 7 components measuring symptoms and 5 components measuring signs (score 1 for yes and 0 for no). The seven symptoms are reduced sweating, dry skin, paraesthesia, weight gain, constipation, hoarseness, and deafness. The five signs are slow motion, rough skin, cold skin, periorbital puffiness, and delayed ankle reflexes.

In this paper the Bayesian Logistic Regression method will be used to model post-radiation NPC patient data to assess which Zulewski’s score component has a significant effect in predicting the
presence or absence of hypothyroidism. Because parameter estimation is done using the Bayesian method, specification of prior information is required. To limit the scope of the paper, we decided to use a non-conjugate, non-informative prior. That is, the prior for each of the regression coefficient is a normal distribution with mean $\mu_0$ and variance $\sigma_0^2$, with $\sigma_0^2$ being chosen large enough. Bayes Estimator is obtained using the MCMC method with Gibbs Sampling.

2. Materials and method
This section contains a discussion of the theoretical basis used in this study including logistic regression model and Bayesian method, as well as the data and variables used. Discussions on the Bayesian method include the Bayes theorem, prior distribution, and construction of the posterior distribution.

2.1. Logistic regression model
The logistic regression model is a regression model with a binary response variable. Therefore, it can be assumed that the response variable is a Bernoulli random variable, and if we assume there are $n$ observations, the response variable for the $i$-th observation is $y_i$ ($y_i$ follows a Bernoulli distribution).

The linear regression model as the following

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

with $\varepsilon_i \sim NID(0, \sigma^2)$. Since $E(\varepsilon_i) = 0$, then

$$E(y_i|x_{1i}, x_{i2}, \cdots, x_{ik}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

From the Bernoulli probability distribution, $Pr(y_i = 1) = \pi_i$ dan $Pr(y_i = 0) = 1 - \pi_i$, obtained

$$E(y_i|x_{1i}, x_{i2}, \cdots, x_{ik}) = \pi_i,$$

so that

$$\pi_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

There is a problem in the above equation. That is, the probability of $\pi_i$ on the left side must be between 0 and 1, yet $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$ on the right side can have any real number value. There is no guarantee that the value of which is predicted to be in the correct range unless complex restrictions are imposed on the coefficient. A solution to this problem is to transform the probability to eliminate the range of constraints and model the transformation as a linear function of $x_{1i}, x_{i2}, \cdots, x_{ik}$.

The transformation can be done in the following two steps:

- Calculate the odds of $\pi_i$. Odds is the ratio of a probability to its complement.

$$odds_{\pi_i} = \frac{\pi_i}{1 - \pi_i}$$

- Calculate the logarithm of $odds_{\pi_i}$ commonly called logit transformation or log-odds of probability $\pi_i$, i.e.

$$\eta_i = logit(\pi_i) = ln\left(\frac{\pi_i}{1 - \pi_i}\right)$$

It can be seen that if the probability of $\pi_i$ drops towards 0 then $odds_{\pi_i}$ will go to 0 and $logit(\pi_i)$ will go to $-\infty$. Whereas if the probability of $\pi_i$ goes to 1 then $odds_{\pi_i}$ will go to $\infty$ and $logit(\pi_i)$ will also go to $\infty$. Thus, logit maps the probability of the range (0,1) to the entire real line. After that, the inverse transformation should be performed, as to obtain the following,

$$\pi_i = logit^{-1}(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$
The equation above is a logistic function that has a S-shaped (reverse S-shaped) nonlinear curve value of between 0 and 1.

Suppose the logit of $\pi_i$ is a linear predictor, i.e.

$$
\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}.
$$

Because $E(y_i|x_{i1}, x_{i2}, \ldots, x_{ik}) = \pi_i$ then

$$
E(y_i|x_{i1}, x_{i2}, \ldots, x_{ik}) = \pi_i = \logit^{-1}(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}}
$$

The equation above is called a logistic response function [1].

2.2. Likelihood function

Let $Y_1, Y_2, \ldots, Y_n$ is a random sample with the probability density function $f(y|\beta)\;\beta \in \Omega$ where $\beta$ is a parameter whose value is unknown and $\Omega$ is a parameter space. Suppose that the observation gives the result $Y_i = y_1, Y_2 = y_2, \ldots, Y_n = y_n$. The information from that observation will be summarized into the likelihood function. The likelihood function can be expressed as a joint probability density function of $Y_1, Y_2, \ldots, Y_n$, that is

$$
L(\beta) = f(Y_1, Y_2, \ldots, Y_n|\beta)
$$

Since $Y_1, Y_2, \ldots, Y_n$ are random samples, then the likelihood can be rewritten as the following:

$$
L(\beta) = f(Y_1, Y_2, \ldots, Y_n|\beta) = \prod_{i=1}^{n} f(y_i|\beta)
$$

2.3. Bayesian theorem

Suppose the $C$ sample space is partitioned into $k$ independent events $C_1, C_2, \ldots, C_k$ with $P(C_i) > 0, i = 1, 2, \ldots, k$ and $P(C_1) + P(C_2) + \cdots + P(C_k) = 1$. Then, apply

$$
P(C_i|C) = \frac{P(C_i)P(C|C_i)}{P(C)} = \frac{P(C_i)P(C|C_i)}{\sum_{i=1}^{k} P(C_i)P(C|C_i)}
$$

This theorem states that, starting with the prior information $P(C_i)$ then adding more information from sample data $C$ through the likelihood $P(C|C_i)$, we obtain the updated information in the form of posterior density $P(C_i|C)$, after division by the normalizing constant $P(C)$ to ensure that the distribution property applies.

2.4. Prior distribution

In the Bayesian method, the parameter to be estimated, $\beta$, is treated as a random variable and has a value in the domain of $B$. The prior distribution is denoted by $f_{\beta}(\beta)$, stating information about the parameter before making an observation. Prior distributions can be obtained from prior information about the estimated parameters, for example from previous research or from available theories [13]. Choice of prior distribution greatly influences the results of the assessment using the Bayesian method [14].

The prior distribution is divided into two namely prior conjugate and non-conjugate prior. Priors that do not meet the definition of prior conjugates are non-conjugate priors. The prior distribution is said to be the prior conjugate distribution for a particular model if the resulting posterior distribution comes from the same family as the prior distribution [15]. The selection of prior conjugate distributions is based on the similarity of functional forms to the likelihood model. With prior conjugates, posterior information can be obtained in closed form. Posterior information is said to be closed-form, if posterior information is a function of the probability density of a particular distribution, making it easier to do
analysis / inference. However, it is not always possible to obtain prior forms that are similar to likelihood forms. Therefore, there is another alternative in determining priors, called non-conjugate priors. The selection of non-conjugate prior distributions is based on information from parameter characteristics. In this case, the similarity of functional form with the likelihood model is no longer a major consideration.

2.5. Formation of posterior distribution

The posterior distribution is a conditional probability density function of $\beta$ given $Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n$ is denoted by $f(\beta | y_1, y_2, ..., y_n)$, so based on the Bayesian theorem can be written,

$$f(\beta | y_1, y_2, ..., y_n) = \frac{f(y_1, y_2, ..., y_n | \beta) f_\beta(\beta)}{\int_{\beta} f(y_1, y_2, ..., y_n | \beta) f_\beta(\beta) d\beta}$$

where $f_\beta(\beta)$ is a prior distribution and $f(y_1, y_2, ..., y_n | \beta)$ is a likelihood function. Since the denominator is just a normalizing constant which does not affect the shape of the distribution, we can simplify problem by just considering the components that are proportional to the posterior, that is

$$\text{posterior} \propto \text{likelihood} \times \text{prior}.$$

2.6. Data and variables

Secondary data from post-radiation nasopharyngeal cancer (NPC) patients were used in this study. The data was obtained from one hospital in Jakarta, was collected from November 2015 to March 2016. Data on 97 patients with age of 18–71 years old were recorded. For data on symptoms and signs of hypothyroidism, Zulewski’s score is used by measuring the history of physical examination. The measurement scale for symptoms and signs of hypothyroidism is grouped into 2, presence (notated by 1) and absence (notated by 0). The target, or response variable, is the presence or absence of hypothyroidism. Whereas 12 Zulewski’s score components as predictor variables and are included in the types of indicator variables, as follows:

$$x_{ij} = \begin{cases} 
0, & \text{the absence of the } j_{th} \text{symptom measured by Zulewski instrument} \\
1, & \text{the presence of the } j_{th} \text{symptom measured by Zulewski instrument} 
\end{cases}$$

where the symptoms are reduced sweating, hoarseness, paraesthesia, dry skin, constipation, deafness, weight gain, slow motion, delayed ankle reflexes, rough skin, periorbital puffiness, and cold skin, respectively.

3. Results and discussion

Let $n, k$ be the sample size and number of explanatory variables, respectively. Then, $Y_i$ is Bernoulli distributed with $\pi_i = \Pr(Y_i = 1), i = 1,2, ..., n$. Then the logistic regression model for this data is

$$\text{logit}(\pi_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$\pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}}}.$$ (2)

The likelihood function according to the model is

$$L(\beta) = \prod_{i=1}^{n} \left( \frac{e^{\beta_0 x_{i1} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}} \right)^{y_i} \left( 1 - \frac{e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}} \right)^{1-y_i}.$$ (3)
3.1. Prior distribution specifications
The \( \beta_j \) parameter, \( j = 1, 2, ..., k \), can be in the range \( (-\infty, \infty) \) and there is no information regarding previous studies regarding \( \beta \). Therefore, the prior distribution for \( \beta \) is assumed to be normally distributed with mean \( \mu_0 = 0 \) and variance \( \sigma_0^2 \), with \( \sigma_0^2 \) taken large enough, that is

\[
f(\beta_j) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[ -\frac{(\beta_j - \mu_0)^2}{2\sigma_0^2} \right], \quad -\infty < \beta_j < \infty,
\]

(4)

All \( \beta_j \)'s are assumed to be independent, so the joint prior distribution for all the regression coefficients can be written as:

\[
f(\beta_0, \beta_1, ..., \beta_k) = \prod_{j=0}^{k} \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[ -\frac{(\beta_j - \mu_0)^2}{2\sigma_0^2} \right]
\]

(5)

3.2. Construction of posterior distribution
Denote \( \text{data} = \{d_1, d_2, ..., d_n\} \), where \( d_i = \{x_i, y_i\} \), with \( x_i = \{x_{i1}, x_{i2}, ..., x_{ik}\} \), for \( i = 1, 2, ..., n \), and \( \beta = \{\beta_0, \beta_1, ..., \beta_k\} \). Then the posterior distribution can be denoted by \( f(\beta|\text{data}) \).

From equation 3 and equation 5, then

\[
f(\beta|\text{data}) \propto \prod_{i=1}^{n} \left( \frac{e^{\beta_0 + \beta x_{i1}^t + \cdots + \beta x_{ik}^t}}{1 + e^{\beta_0 + \beta x_{i1}^t + \cdots + \beta x_{ik}^t}} \right)^{y_i} \left( \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[ -\frac{\beta_0^2}{2\sigma_0^2} \right] \right)^{1-y_i} \prod_{j=0}^{k} \frac{1}{\sigma_j \sqrt{2\pi}} \exp \left[ -\frac{\beta_j^2}{2\sigma_j^2} \right]
\]

(6)

The posterior distribution (in equation 6) is a non-closed form since it does not form a particular distribution. Thus, computational techniques are needed to obtain the Bayes estimator (in this case \( \beta \)). MCMC simulation with Gibbs sampling will be used to obtain the Bayes estimator.

Let \( \beta \) be the vector of parameters to be estimated and information about \( \beta \) is measured based on \( f(\beta|\text{data}) \). Gibbs Sampler algorithm to produce \( \beta^{(s)} \) from \( \beta^{(s-1)} \) is as follows:

1. Determine the initial point \( \hat{\beta}^{(0)} = [\beta_0^{(0)}, \beta_1^{(0)}, ..., \beta_k^{(0)}] \)
2. Determine the number of iterations \( (M) \)
3. In iteration \( s \), the value of \( \beta_j, j = 0, 1, ..., k \) is updated by:
   - Sample \( \beta_0^{(s)} \sim f(\beta_0|\beta_1^{(s-1)}, ..., \beta_k^{(s-1)}, \text{data}) \),
   - Sample \( \beta_1^{(s)} \sim f(\beta_1|\beta_0^{(s)}, \beta_2^{(s-1)}, ..., \beta_k^{(s-1)}, \text{data}) \),
   - Sample \( \beta_k^{(s)} \sim f(\beta_k|\beta_0^{(s)}, \beta_1^{(s)}, ..., \beta_{k-1}^{(s-1)}, \text{data}) \).
4. Repeat step 3 until the iteration is finished
5. Repeat steps 1 through 4 for the next chain, if more than 1 chain is needed (optional).

The sampling distribution of \( \beta^{(s)} \) will approach the target distribution when \( M \to \infty \).

The purpose of the MCMC approximation is to obtain a sequence of parameter values \( [\beta^{(1)}, \beta^{(2)}, ..., \beta^{(M)}] \) such that, for a function \( g \), it is expected that the empirical average of \( [g(\beta^{(1)}), ..., g(\beta^{(M)})] \) can estimate the expected value of \( g(\beta) \) based on the distribution target probability \( f(\beta|\text{data}) \).
3.3. Data analysis

Furthermore, the Bayesian logistic regression method will be used to analyze NPC patients post-radiation data. Based on logistic regression,

\[ \pi_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{12} x_{112}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_{12} x_{112}}} \]  

(7)

with \( \pi_i = Pr(Y_i = 1, hypothryoid), i = 1,2,\ldots, 97 \). Based on equation 3, likelihood function of \( Y_1, Y_2, \ldots, Y_{97} \) is

\[ L(\hat{\beta}) = \prod_{i=1}^{97} \left( \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}} \right)^{y_i} \left( 1 - \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}} \right)^{1-y_i} \]  

(8)

Because the value of \( \beta \) in this study can be in \((-\infty, \infty)\) range and there is no information about previous research, the prior distribution for \( \beta \) can be assumed to be \( N(\mu_0, \sigma_0^2) \), large number is chosen for \( \sigma_0^2 \). In this paper, \( \mu_0 = 0 \) and \( \sigma_0^2 = 10^6 \) are chosen, and thus the joint probability density function of \( \beta \) can be written as

\[ f(\beta_0, \beta_1, \ldots, \beta_{12}) = \prod_{j=0}^{12} \frac{1}{10^3 \sqrt{2\pi}} \exp \left[ -\frac{(\beta_j)^2}{2(10^6)} \right] \]  

(9)

From equation 8 and equation 9, the posterior distribution can be written as:

\[ f(\hat{\beta} | \text{data}) \propto \prod_{i=1}^{97} \left( \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}} \right)^{y_i} \left( 1 - \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{12} x_{112}} \right)^{1-y_i} \]  

\[ \times \prod_{j=0}^{12} \frac{1}{10^3 \sqrt{2\pi}} \exp \left[ -\frac{(\beta_j)^2}{2(10^6)} \right] \]  

(10)

Using Gibbs algorithm with \( k = 12 \), and 2 chains with the number of iterations per chain is 100,000 with the first 50,000 iterations for burn-in, MCMC simulation was conducted in R [16]. The convergence of the Markov chain was examined through traceplots, autocorrelation plots, and Gelman et al. plots. From the three plots used, it was found that the simulation converged, and thus the results can be used for further insights [17].

Figure 1. Bayes estimator results
It can be seen from figure 1 in the mean column that the average value of the sampling results using the MCMC method is positive for all beta except $\beta_3$. This can be interpreted that all components of the Zulewski’s score except slow motion are associated with an increase in the probability of developing hypothyroidism. While slow motion is associated with a decrease in the probability of hypothyroidism. Furthermore, from the quantiles for each variable table, information can be obtained that the 95% confidence interval for all betas is between a negative value and a positive value, so there is a possibility that the regression coefficients are zero. If this happens then there is no sufficient evidence that predictor variable could explain the hypothyroidism incidence. Therefore, variable selection procedure was carried out to support the results of the Bayes estimator that has been obtained.

The Bayesian solution for variable selection is if it is believed that there are many regression coefficients whose values are potentially equal to zero, then use a prior distribution that reflects this possibility. This can be achieved by specifying that each regression coefficient has some non-zero probability of zero.

The regression coefficient for variable j can be written as follows,

$$b_j = I_j \times \beta_j, \quad I_j \in \{0,1\}, \quad j = 1,2,...,12$$

(11)

where $I_j$ is an indicator variable whose probability states the probability of a variable to be included in the model. By using this parameterization, the regression equation becomes

$$\pi_i = \frac{e^{\beta_0 + \beta_1 I_{i1} + \beta_2 I_{i2} + ... + \beta_{12} I_{i12}}}{1 + e^{\beta_0 + \beta_1 I_{i1} + \beta_2 I_{i2} + ... + \beta_{12} I_{i12}}}, \quad i = 1,2,...,97.$$  

(12)

$I_j$s are independent for all $j$ and follow a Bernoulli distribution with mean $P(I_j = 1)$. George and McCulloch suggest a value of 0.5 for $P(I_j = 1)$ which makes all models have the same probability of occurring [18]. Estimation of parameters by the variable selection method gives the result that all predictor variables have a small inclusion probability but cannot be ignored.

4. Conclusion

Parameter estimation which is a logistic regression coefficient, using Bayesian method with prior distribution for $\beta$ is $N(\mu_0, \sigma_0^2)$, with $\sigma_0^2$ selected large enough, resulting in posterior distribution in non-closed form. A computational technique is needed to obtain an estimator for $\beta$, the MCMC with Gibbs sampling.

Data analysis on post-radiation NPC patients’ data showed that no components of the Zulewski’s score that was more significant between one another. Thus, we recommend that additional information is needed from measurements other than the Zulewski’s score component to predict hypothyroidism in post-radiation NPC patients.

Acknowledgments

This work was financially supported by Universitas Indonesia under research grant PITTA B with grant contract number .NKB-0665/UN2.R3.1/HKP.05.00/2019.

References

[1] Mendenhall W and Sincich T 2012 *A Second Course in Statistics Regression Analysis* 7th edition (United States of America: Pearson Education, Inc.)

[2] Peretti A and Amenta F 2016 *Breast. Can. Curr. Res.* 1 111

[3] Barco I et al. 2017 *Clin. Transl. Oncol.* 19 1393-9

[4] Cox R T 1946 *Amer. J. Phys.* 14 1-13

[5] Cox R T 1961 The Algebra of Probable Inference (Baltimore: The Johns Hopkins Press)

[6] Savage L J 1954 *The Foundations of Statistics* (New York: John Wiley & Sons Inc.)
[7] Savage L J 1972 *The Foundations of Statistics* (New York: Dover Publications Inc.)
[8] Hoff P D 2009 *A First Course in Bayesian Statistical Methods* (New York: Springer)
[9] Ford P 2018 MCMC Algorithms v0.5 available at https://www.casact.org/library/studynotes/mcmc_algorithms_v0.6.pdf
[10] Adham M, Antonius N K and Jaap M M 2012 *Chin. J. Cancer* **31** 185-96
[11] Adham M et al. 2017 *Pedoman Nasional Pelayanan Kedokteran Kanker Nasofaring* (Jakarta: Kementrian Kesehatan Republik Indonesia)
[12] Tolentino E D S, Centurion B C, Rubira I R F R 2011. *J. Appl. Oral Sci.* **19** 448-54
[13] Box G E P and Tiao G C 1992 *Bayesian Inference in Statistical Analysis* (Canada: John Wiley & Sons, Inc)
[14] Tse Y K 2009 *Nonlife Actuarial Models: Theory, Methods, and Evaluation. Research Collection School of Economics* (Singapore: Singapore Management University)
[15] Klugman S A, Panjer, H H, and Willmot G E 2012 *Loss Models from Data to Decisions* 4th edition (Canada: John Wiley & Sons, Inc)
[16] R Core Team 2019 R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* available at URL https://www.Rproject.org/
[17] Gelman A and Rubin D B 1992 *Statistical Science* **7** 457-511
[18] George E I and McCulloch R E 1993 *J. Am. Stat. Assoc.* **88** 398-409