BAREKENG: Journal of Mathematics and Its Applications
September 2023 Volume 17 Issue 3 Page 1725–1738
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss3pp1725-1738

A STATISTICAL ANALYTICS OF MIGRATION USING BINARY BAYESIAN LOGISTIC REGRESSION

Devi Azarina Manzilir Rohmah[1*], Ani Budi Astuti[2], Achmad Efendi[3]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Science, Brawijaya University Veterans Street, Malang 65145 East Java, Indonesia*

Corresponding author's e-mail: * **deviazarina@gmail.com**

ABSTRACT

Article History:

Received: 23rd May 2023
Revised: 16th August 2023
Accepted: 25th August 2023

Keywords:

Bayesian method;
Binary logistic regression;
Migration.

Binary logistic regression is utilized in research to understand the relationship between multiple independent variables and a binary response variable. In logistic regression modelling, parameter estimation is regarded as a vital stage. The performance of this estimation is often affected by the sample size and data characteristics, and to deal with this problem, the Bayesian method can be employed as an estimation. This research aims to use Regression Logistic with Bayesian estimation to figure out the determinant of recent in-migrants status in Special Region of Yogyakarta 2021, where Yogyakarta's recent in-migrants in 2021 took the first position in Indonesia, whereas this city has the lowest regional minimum wage in Indonesia. The Bayesian method was used in this study to obtain a better estimate than previous studies using maximum likelihood estimation, because Bayesian is unbiased for unbalanced cases which are often found in logistic regression. This research results show that particular variables such as resident age, resident marital status, resident main activities, resident latest education, and resident homeownership have significant effect on resident migrating to Special Region of Yogyakarta, Indonesia

How to cite this article:

D. A. M. Rohmah, A. B. Astuti and A. Efendi., "A STATISTICAL ANALYTICS OF MIGRATION USING BINARY BAYESIAN LOGISTIC REGRESSION," *BAREKENG: J. Math. & App.*, vol. 17, iss. 3, pp. 1725-1738, September, 2023.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Migration is the act of an individual moving from one location to another. Within the field of demographics, migration is referred as a movement that causes a long-term or permanent shift in the person's customary abode. This movement serves as the response to the variation in the condition of a neighbourhood where the population resides. There are various pull and push factors that influence migration [1]. Economic factors like better housing, higher living standards, and employment opportunities are examples of its pull factors. Recent migration, one of the types of migration, refers to recent migrants whose province in the past five years was different from the province during the census. In other words, the recent in-migrants in the Special Region of Yogyakarta refer to the residents living in another province other than Special Region of Yogyakarta before the census took place. The data on recent migrants were obtained from National Socio-Economic Survey held in March 2021.

Results of National Socio-Economic Survey held in March 2021 shows that there are 5.9% recent in-migrants in Special Region of Yogyakarta, which makes Special Region of Yogyakarta has the highest number of recent in-migrants among other provinces in Indonesia. This number indicates that there are 5.9% civilians in Yogyakarta had a place of residence outside this province five (5) years before the enumeration took place in 2021. It is fascinating due to fact that Special Region of Yogyakarta had the lowest Province Minimum Wage (UMP) for years, including in 2021. The prevalent theory that stated economic motivation is one of the determinants of migration is in direct opposition to this. To find out the determinants of the status of migrants entering research to D.I. Yogyakarta, then logistic regression analysis can be used.

Logistic regression is a data analysis technique in statistics that is used to determine the effect of multiple variables where the response variable can be categorical (ordinal and nominal are accepted) with predictor variables that can be categorical or continuous [2]. The logistic regression comprises multinominal, ordinal, and binary logistic regressions. Multi-categorical outcomes can be analyzed with multinomial logistic regressions, whereas ordinal variables should be analyzed with an ordinal logistic regression model [3]. Furthermore, binary logistic regression is utilized in research to understand the relationship between multiple independent variables and a binary response variable [4]. The response variable consists of dichotomous qualitative data, with a value of 1 indicating that a characteristic exists and a value of 0 indicating that it does not. This research aims to figure out the determinant of recent in-migrants in Special Region of Yogyakarta, Indonesia, in 2021. Binary logistic regression is used to analyse this topic due to its dependent variable which consist of dichotomous qualitative data. Previous research using logistic regression with Maximum Likelihood estimation showed that the effect of education and gender is significant on a person's decision to migrate [5]. Other research shows that age, marital status, occupation, and home ownership also have an effect on migration [6][7][8].

In this binary logistic regression modelling, parameter estimation is regarded as a vital stage. The performance of this estimation is often affected by the sample size and data characteristics. An unbalanced dependent variable is often seen in binary logistic regression when one of the classes determined is uncommon [9]. This condition could affect the performance of the estimation method used, and to deal with this problem, the Bayesian method can be employed as an estimation. Bayesian method is unbiased for unbalanced data, because the parameter estimation that refers to Bayesian yields a more relevant result than the conventional method often used in parameter estimation in binary logistic regression, namely the Maximum Likelihood, to model the case with an unbalanced dependent variable [10]. In light of this theory, this research adopted the Bayesian method to estimate binary logistic regression parameters to identify the factors that affect the status of newcomers migrating to the Special Region of Yogyakarta in 2021 with the variables that have been done previously. The Bayesian method is used to obtain better parameter estimates than previous studies. By knowing the factors that influence this migration, it is hoped that it will be able to provide insight to the government regarding the determinants of incoming migrants to research in the Special Region of Yogyakarta so that the government is able to make precise policies to develope its region.

2. RESEARCH METHODS

This research employed secondary data gathered from Migrant Profiles of Socio-Economic Survey Results KOR March 2021, which is a survey conducted once a year, using Bayesian Binary Regression Logistic.

Logistic regression is a basic classification method initially intended for response variable or dependent variable with two classes namely binary logistic regression, which further develops with a dependent variable that consists of Multinomial Logistic Regression. Binary logistic regression is a method of analysis used to find out the connection between a dependent variable that is binary or dichotomous and a predictor variable or independent variable that is polychotomous [11].

If there is an observation (***X*, *Y***) in which the ***X*** represents the independent variable and ***Y*** is the dependent one as stated in **Equation (1)**, then it can be transformed as **Equation (2)**.

$$\pi(\boldsymbol{X}) = \frac{\exp(\beta_0+\beta_1 x_1+\cdots+\beta_m x_m)}{1+\exp(\beta_0+\beta_1 x_1+\cdots+\beta_m x_m)} \quad (1)$$

$$ln\left(\frac{\pi(\boldsymbol{X})}{1-\pi(\boldsymbol{X})}\right) = (\beta_0+\beta_1 x_1+\cdots+\beta_m x_m) = \beta^T\boldsymbol{X} \quad (2)$$

One of the methods commonly used to estimate the parameter in binary logistic regression is the maximum likelihood (MLE) and Bayesian method. The Bayesian method was developed according to the Bayes theorem. In this case study, the application of this method is intended to compound information coming from data with prior probability in terms of model validity level, so that the best model can be selected with the highest posterior probability and an average sum is obtained [12].

In Bayes theorem, B_i, with $i = 1, 2, \dots, n$ as sample space S with $P(B_i) \neq 0$ and representing an independent event; thus, for random event A where $P(A) \neq 0$, probability B_i with condition A is given as follows:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(B_i)P(A|B_i)} \quad (3)$$

Furthermore, if $\sum_{i=1}^{n} P(B_i)P(A|B_i)$ is regarded as constant, **Equation (3)** will turn to **Equation (4)**:

$$P(B_i|A) \propto P(A|B_i)P(B_i) \quad (4)$$

In addition to the model $f(x|\theta)$ or the likelihood, Bayesian requires the distribution for θ, or known as prior [13]. Non-informative prior can be referred to without initial knowledge of the parameter distribution to determine the prior distribution [14], while the prior distribution for the parameter in the binary logistic regression model of Bayesian follows the normal distribution [15]. Estimating parameter θ may refer to the distribution $f(x|\theta)$ and $\pi(\theta)$, with $\pi(\theta)$ as prior distribution for θ [16]. This refers to θ given X (observed data) called posterior distribution given in the following formula:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{g(x)} \quad (5)$$

so:

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \quad (6)$$

The convergence checking in Bayesian method is intended to figure out whether the data is relevant to the prior distribution. This convergence checking method, namely Markov Chain Monte Carlo (MCMC), can refer to Trace Plot, MC error, and Autocorrelation [17].

- Trace Plot; if the model has converged, the trace plot result does not form a particular pattern.
- MC *error*; if the model has converged, the score of MC Error is very low (less than 5% of standard deviation).
- Autocorrelation; if the model has converged, the first lag the autocorrelation score is close to one and the following lag shows an autocorrelation score heading for zero.

Moreover, parameter testing is required to investigate whether the predictor variable significantly affects the response variable. In the Bayesian method, the parameter test refers to *credible intervals* of 2.5% and 97.5% quantile of the distribution. If the *credible intervals* do not indicate a score of 0, the predictor variable significantly affects the response variable.

2.1 Data and Variable Outlines

The criteria of respondents covered 17 to 64-year-old Indonesian citizens residing in Yogyakarta during the census. The variables and indicators used in this research are presented in **Table 1**.

Table 1. Variable Outlines

Variable	Answer	Data Scale
Recent Migrants (Y)	(0) No (1) Yes	Nominal
Age (X_1)	(0) 17-40 years old (1) 41-64 years old	Ordinal
Sex (X_2)	(0) Men (1) Women	Nominal
Marital Status (X_3)	(0) Not yet married (1) Ever/currently married	Nominal
Latest Education (X_4)	(0) Not going to school (1) Primary and Secondary School (2) High School (3) University Qualifications	Ordinal
Main Activities (X_5)	(0) Others (1) Studying (2) Working (3) Taking care of the household	Nominal
Homeownership (X_6)	(0) Not under their ownership (1) Under their ownership	Ordinal

2.2 Method

1. Checking for outliers using Cook's Distance,
2. Checking multicollinearity assumptions,
3. Determine the model of binary logistic regression as stated in **Equation (1)**

$$\pi(\boldsymbol{X}) = P(Y = 1|\boldsymbol{X}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}$$

4. Determine the prior distribution of the parameters,
5. Determine the likelihood of binary logistic regression,
6. Get the posterior distribution according to **Equation (6)**,

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$$

7. Obtain parameter estimation from β using Gibbs Sampling Markov Chain Monte Carlo (MCMC),
8. Checking convergence based on autocorrelation, trace plot, and MC Error plots,
9. Performing parameter testing using credible intervals,
10. Testing the accuracy of classification predictions,
11. Determine the equation model of bayesian binary logistic regression,
12. Calculating the value of the odds, and
13. Interpretating the model obtained.

To make it clearer, the flowchart of this research is shown in **Figure 1**.

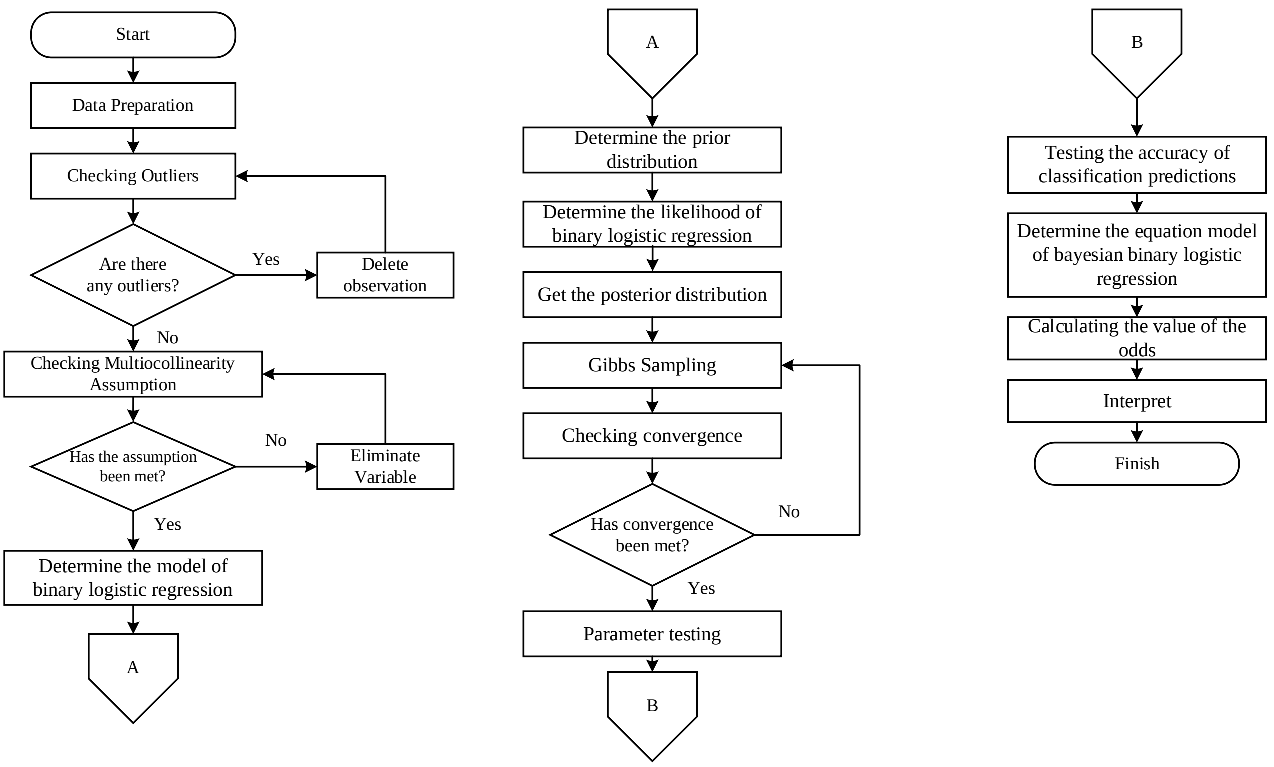


Figure 1. Flowchart

3. RESULTS AND DISCUSSION

3.1 Outlier Detection

In this research, Outlier Detection test was performed using Cook's Distance plot. The Cook's Distance plot can be seen in **Figure 2** below.

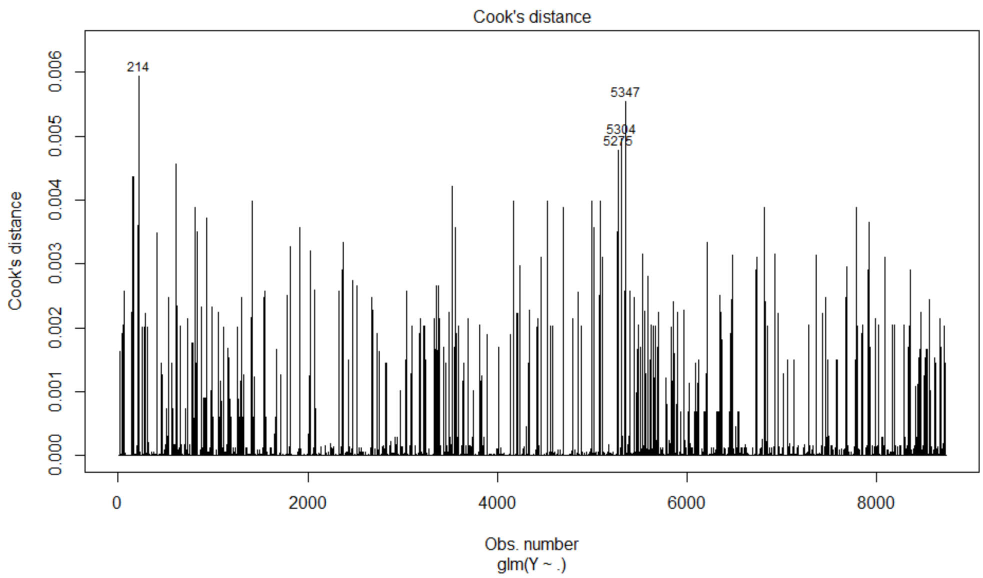


Figure 2. Cook's Distance Plot

In this research, any point with a Cook's Distance over 4/n is considered as outlier and will be removed from the data set and no longer used.

3.2 Multicollinearity Test

The multicollinearity test was performed by referring Pearson-Spearman correlation score. The correlation score of each variable can be seen in **Table 2**, while it's correlation plot can be seen in **Figure 3**.

Table 2. Pearson-Spearman Correlation Score

Corr.	Y	X_1	X_2	X_3	X_4	X_5	X_6
Y	1	-0.149	-0.003	-0.131	0.092	0.081	-0.256
X_1	-0.149	1	0.011	0.451	-0.256	-0.100	0.152
X_2	-0.003	0.011	1	0.159	-0.007	0.255	0.008
X_3	-0.131	0.451	0.159	1	-0.159	-0.099	0.098
X_4	0.092	-0.256	-0.007	-0.159	1	0.037	-0.119
X_5	0.081	-0.100	0.255	-0.099	0.037	1	-0.053
X_6	-0.256	0.152	0.008	0.098	-0.119	-0.053	1

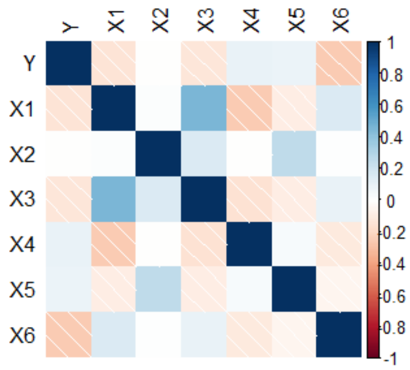


Figure 3. Correlation Plot

Table 2 and **Figure 3** shows that there was no correlation between variables higher than 0.6, meaning that the assumption of non-multicollinearity was fulfilled.

3.3 Parameter Estimation using Binary Logistic Regression

The results of the parameter estimation from binary logistic regression analysis using maximum likelihood are shown in **Table 3**.

Table 3. Parameter Estimation

Parameter	Estimate	Std. Error	z value	Pr(>|z|)	Result
(Intercept)	-1.952	0.187	-10.46	< 2e-16	
X_1	-0.868	0.127	-6.86	6.94E-12	Significant
X_2	-0.080	0.099	-0.80	0.423	Insignificant
X_3	-0.432	0.096	-4.47	7.68E-06	Significant
X_4	0.270	0.064	4.20	2.69E-05	Significant
X_5	0.327	0.060	5.49	4.07E-08	Significant
X_6	-1.717	0.094	-18.27	< 2e-16	Significant

3.4 Bayesian Logistic Regression

Bayesian method aims to obtain posterior distribution from the multiplication of prior distribution and likelihood. Several previous studies using the Bayesian method often used non-informative prior to determining prior distribution, considering that there has not been any prior knowledge regarding the parameter distribution. With the absence of this parameter, previous studies referred to normal distribution. This type of distribution was picked due to the two parameters, namely mean (μ) showing the true parameter score and standard deviation (σ) showing the uncertainty of the score of a parameter. Therefore, in this research, the prior was determined to have a normal distribution with the mean zero and variance 1. In this research, Gibbs sampling was used as the MCMC algorithm with the iteration of 500,000 + 250,000 burn in + 2 thin.

Convergence Test

In the Bayesian method, a convergence test is required to find out if the generated score is in accordance with the posterior distribution. The convergence test in MCMC used a trace plot, MC Error, and autocorrelation plot. The convergence test using MC Error is presented in **Table 4**.

Table 4. Convergence Test using MC Error

Parameter	SD	1% SD	MC Error	Result
β_0	0.186	0.00186	0.0017890	Convergence
β_1	0.118	0.00118	0.0004014	Convergence
β_2	0.101	0.00101	0.0003831	Convergence
β_3	0.065	0.00065	0.0002696	Convergence
β_4	0.064	0.00064	0.0005498	Convergence
β_5	0.059	0.00059	0.0003690	Convergence
β_6	0.094	0.00094	0.0003771	Convergence

The trace plot from the analysis result is presented in **Figure 4**.

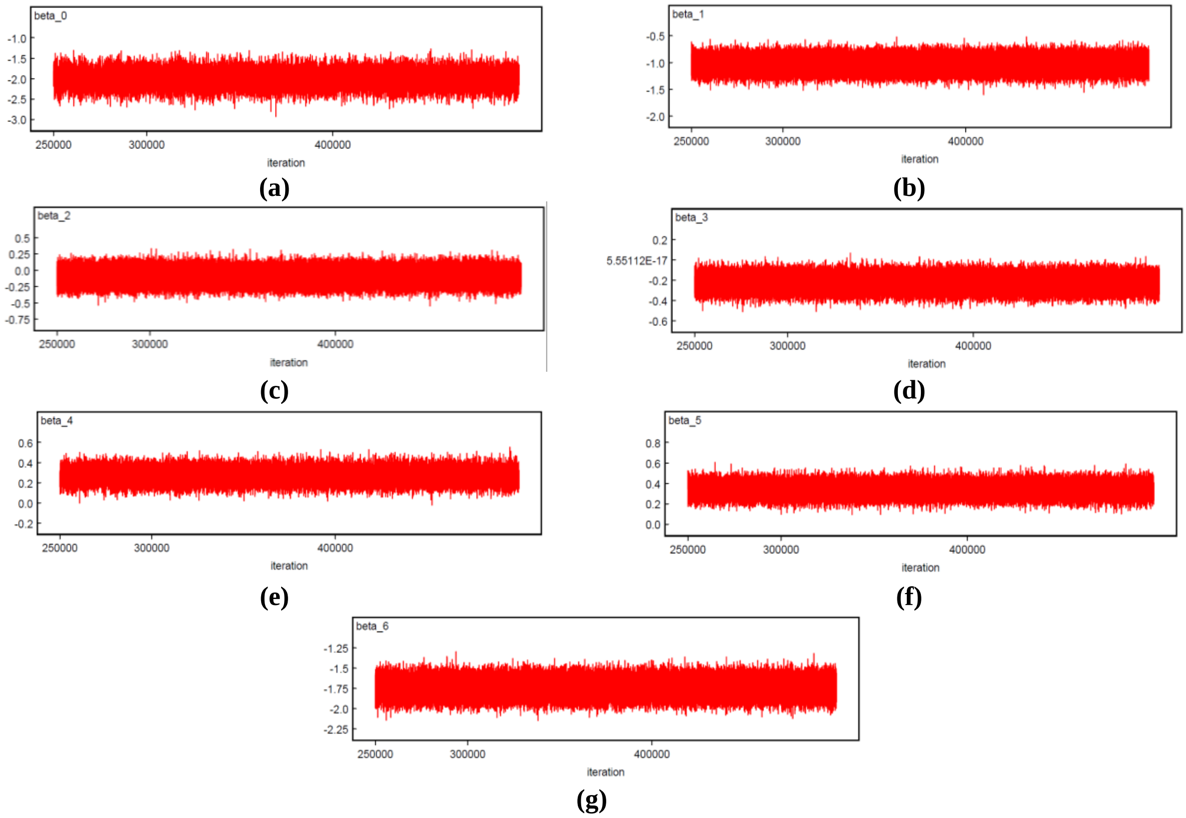
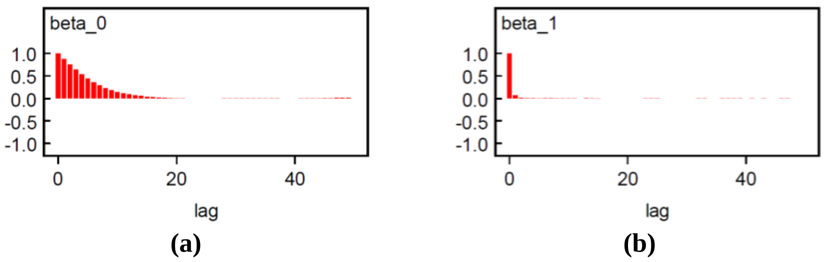


Figure 4. Trace Plot (a) Beta 0 (β_0), (b) Beta 1 (β_1), (c) Beta 2 (β_2), (d) Beta 3 (β_3), (e) Beta 4 (β_4), (f) Beta 5 (β_5), (g) Beta 6 (β_6)

Figure 4 shows that the trace plot did not show any strong pattern or periodicity. The autocorrelation plot is presented in **Figure 5**.



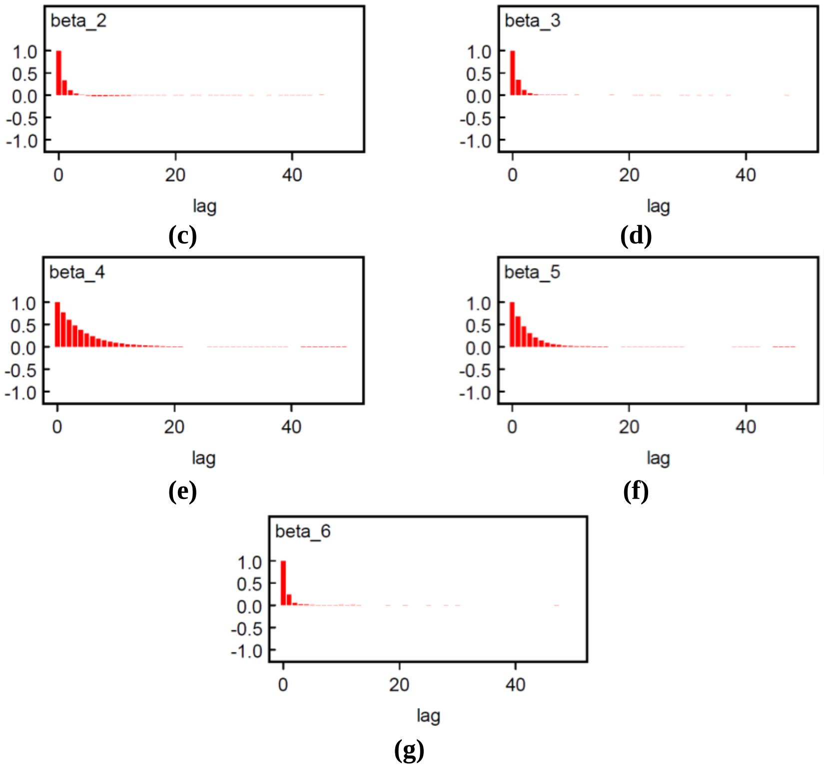


Figure 5. Autocorrelation Plot (a) Beta 0 (β_0), (b) Beta 1 (β_1), (c) Beta 2 (β_2), (d) Beta 3 (β_3), (e) Beta 4 (β_4), (f) Beta 5 (β_5), (g) Beta 6 (β_6)

Figure 5 shows that autocorrelation between parameters was low, resulting in generated independent samples. According to the result of the convergence test of MCMC, each parameter was convergent or the generated samples were from the expected posterior distribution.

Parameter Significance Test

In a Bayesian method, a parameter significance test is performed by examining the credible interval, where significant variables could be determined at the 5% significance level. Values from 2.5% to 97.5% quantiles provide 95% credibility interval for every given variable. A parameter is deemed to be significant if the credible interval does not show zero in the percentile interval of 2.5% to 97.5%. The credible interval for each parameter is presented in **Table 5**.

Table 5. Credible Interval

Parameter	2.50%	97.50%	Result
β_0	-2.366	-1.638	Significant
β_1	-1.260	-0.798	Significant
β_2	-0.290	0.103	Insignificant
β_3	-0.354	-0.097	Significant
β_4	0.146	0.396	Significant
β_5	0.223	0.454	Significant
β_6	-1.922	-1.555	Significant

Table 5 shows that of the six variables used, only one variable was proven insignificant based on the credible interval, namely variable X_2 or sex, while all the five other variables respectively at the 5% significance level are significant based on credible interval.

Classification Accuracy

The accuracy of a model in classifying data is useful to find out the goodness of a Bayesian logistic regression model. The higher the classification of a model is formed, the better the model is obtained. This accuracy in the binary logistic regression is presented in **Table 6**.

Table 6. **Classification Accuracy**

Classification Accuracy		Prediction Class		Precisely Predict
		Not as recent migrants	Recent Migrants	
Actual Class	Not as recent migrants	6575	43	0.949
	Recent migrants	353	10	0.118
	%			0.9433

This table indicates that the model can precisely predict the research subject not as recent migrants, accounting for 6,575 or 94.9% (sensitivity) and this model can precisely predict the recent migrants as the research subject for as much as 10 or 1.18% (specificity). Overall, this model can give an accurate prediction of 94.33%.

In this research, the ROC curve was used to test the relevance of the model used in addition to the analysis using a classification table. The ROC curve of the analysis result is presented in **Figure 6**.

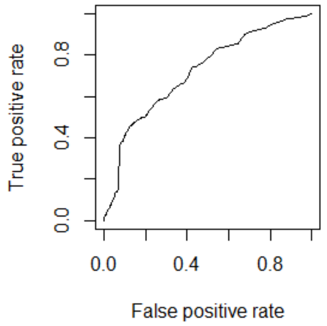


Figure 6. **ROC Curve**

Figure 6 indicates that the model is relevant since the curve generated was close to one. This is in line with the area under the curve or commonly abbreviated as AUC for as much as 0.717. Of this score, this model is deemed to be appropriate to explain the model with a fair discrimination category.

Bayesian Logistic Regression Equation Model

Of the result of Bayesian logistic regression analysis with the iteration of 500,000 + 250,000 burn in + 2 thin, the logistic regression model was obtained as following:

$$g(x) = -2 - 1.026X_1 - 0.224X_3 + 0.271X_4 + 0.339X_5 - 1.738X_6$$

$$ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = -2 - 1.026X_1 - 0.224X_3 + 0.271X_4 + 0.339X_5 - 1.738X_6$$

$$\left(\frac{\pi(x)}{1-\pi(x)}\right) = exp(-2 - 1.026X_1 - 0.224X_3 + 0.271X_4 + 0.339X_5 - 1.738X_6)$$

$$\hat{\pi}(x) = \frac{exp(-2 - 1.026X_1 - 0.224X_3 + 0.271X_4 + 0.339X_5 - 1.738X_6)}{1 + exp(-2 - 1.026X_1 - 0.224X_3 + 0.271X_4 + 0.339X_5 - 1.738X_6)}$$

Interpretation

The analysis result based on the odds Ratio is presented in **Table 7**.

Table 7. **Odds Ratio**

VARIABLE		CATEGORY	ODDS RATIO
X_1	0	17-40 years old	1 (reference)
	1	41-64 years old	0.24 (0.19, 0.29)
X_2	0	Men	1 (reference)
	1	Women	0.98 (0.82, 1.16)
X_3	0	Not yet married	1 (reference)
	1	Ever/currently married	0.34 (0.28, 0.41)

***X*₄**	0	Not going to school	1 (reference)
	1	Primary/Secondary school	2.45 (1.31, 4.58)
	2	High school	6.05 (3.3, 11.1)
	3	University	5.31 (2.884, 9.93)
***X*₅**	0	Others	1 (reference)
	1	Studying	2.06 (0.65, 6.5)
	2	Working	11.61 (3.66, 36.81)
	3	Taking care of household	3.1 (0.97, 9.91)
***X*₆**	0	Not under their ownership	1 (reference)
	1	Under their ownership	0.14 (0.12, 0.17)

3.5 Discussion

Based on the results of the analysis, it is known that age (X_1) and recent in-migrant status are found to be significantly correlated in the parameter significance test, yielding odds ratio of 0.24. This figure shows that people in the age range of 41-64 years are 0.24 times more likely to become migrants in the Yogyakarta Risen than residents in the age range of 17-40 years. In other words, residents in the age range of 17-40 years are 4.17 times more likely to become migrants in Yogyakarta. This is pertinent to previous study that found a negative and significant correlation between age and migration [7][18], indicating that older individuals have a lower likelihood of migrating. In addition, other research found that compared to other age groups, young adults, defined as those aged 18 to 34, have the highest rate of migration and reached their highest point of migration rates between the ages of 20 and 30 [19][20]. Once people reach the limit of productive age, another wave of residents will remigrate because retired residents will decide to return to where they came from [7]. This study's findings are in line with the UNECE Policy Brief on Ageing (2016), which suggests that, despite the main activities being one of the main reasons people migrate, aging is another reason people return to their home town or family, not just because of money. This is also due to the difficulties that elderly migrants encounter when migrating. Elderly migrants are more likely to return to their families and relatives than to migrate due to health conditions or attempts to re-adjust to the environment. The migration wave among retirees is not discussed in this study because its scope is limited to 64-year-old residents only.

Moreover, the result of the odds ratio shows that women were 0.98 more likely to become recent migrants than men. In other words, men were 1.02 more likely to be recent migrants than women. This is in line with the fact that worldwide, the quantity of female transients was two times just that high of guys in all nations besides in Africa and Asia [21]. This was because of discriminative social and social standards and the approaches that had an inconsequential commitment to the issue expecting to safeguard the weakness of ladies. When most people moved, they were affected by gender discrimination, harassment, and violations of women's rights. Gender inequality in migration status in Southeast Asia was highlighted on another research [22]. It shows that young women have the opportunity to fashion themselves in the image of consumerist modernity by migrating to the cities of Southeast Asia. With more prominent independence from the parental look and local area command over their own monetary assets, Southeast Asian ladies can 'makeover' themselves', in this manner growing work market choices, creating certainty and upgrading economic wellbeing. However, gendered identity politics may manifest differently in urban settings; also, female migrant who gain specific advantages in terms of consumption practices, status or self-esteem outside the factories or employers' homes may experience exploitation inside the gates of the factory or house where they are often unable to negotiate work conditions in these constrained settings. As a result of patriarchal norms, filial piety, and familial expectations, young women may be required to marry and become parents, assuming traditional responsibilities for both their natal and marital families. In some instances, the "makeover" may also be temporary [22].

Marital status also contributes significant influences to the status of recent migrants. Residents who ever/currently married had an odds ratio of 0.34, indicating that residents who ever/currently married were 0.34 more likely to migrate than residents who were not yet married. On the other words, residents who were not yet married were 2.94 times more likely to migrate than those who ever/currently married. This is consistent with a previous study that found marital status has a negative and significant impact on Indonesians' decision to migrate internally. This indicates that respondents with married status have a lower likelihood of migrating than respondents with other statuses. The phenomenon of migration occurs in developing nations because the decision to migrate is made on an individual level. As a result, individuals who have not married will likely be freer because they will not have any dependents and will only have responsibilities to themselves; consequently, they will have a greater chance of migrating [23]. Another factor causing marital status is related to migration is due to the cost of moving. Before migrants decide to migrate, they will have to pay for a lot of things like transportation, meals, the cost of buying new assets in the destination, and lodging. The migrants' costs are not too high if they are single; however, if they are married and have children, the costs are even higher [23].

Furthermore, resident latest education is the variable with a significant score. The odds ratio showed that residents with a minimum of a primary and secondary school were 2.45 times more likely to migrate recently than those who did not go to school. In addition, the resident whose latest education was High School was 6.05 times more likely to migrate recently than the resident without a primary education, and the resident whose latest education was University was 5.31 times more likely to migrate recently than the resident without a primary education. Education is frequently regarded as a crucial factor in preventing people from seeing migration as a threat and fostering an openness to diversity [24]. Higher education likewise cultivates people's information handling capacities and, accordingly, better educated people might be better positioned to interpret and evaluate migration phenomena, empowering them to consider the potential positive financial impacts that migration can bring [24]. Furthermore, the literature suggests that young people between around 15 to 25 years old are the group who most consider migrating for education, in accordance with their essential concerns of getting great capabilities prompting a steady employment. Educations is a chance to improve their and their family's financial situation and a way out of poverty [25].

The working category had the highest odds ratio for the variable of main activities with 11.61 odd ratio score, out of the above four categories. This indicates that those who were currently working were 11.61 times more likely to be recent migrant than those whose primary occupation was not employment, nor studying, nor taking care of household. The previous study figured out that, between push factors (factors that make individuals want to leave) and pull factors (factors that attracts individuals to live in an area), pull factors impact greater movement when compared with the push factors. Economic factors (prospects for higher wages, improved living standards, personal development, job opportunities, good welfare standards, and labour demands) were found to have the greatest influence on migration among the pull factors that were taken into consideration [26]. Meanwhile, those who currently studying had a chance of 2.06 more likely to be recent migrant in Yogyakarta. There are universities available in many young people's hometowns. However, young people living in underdeveloped areas frequently believe that these schools provide a subpar education that will not lead to successful careers. The assumption for studying on in a more developed city has turned into a vital piece of certain social orders, and signals a change to adulthood alongside capabilities [25]. A reasonable connection between resident latest education and the tendency to mitigate is for somebody who is taking higher education. Besides that, the education factor can also be linked to the modern theory of migrant internal migration as consumers. The intended consumption is the availability of amnesties, particularly public goods like schools or educational institutions. As a result, the decision to migrate is not just driven by the level of education completed; it may also be motivated by better education [23]. The last category of X_5 accounted for 3.1 more likely to be recent-migrant, indicating that those in charge of households were 2.35 times more likely to migrate than those primary occupation was not employment, nor studying, nor taking care of household. Essentially, there are a few powers setting off individuals to stay in their place of beginning and to leave their place of beginning. In this instance, taking care of household acts as a driving force in their favour [27]. On the other hands, researches had explained that while immigration is decreasing the unemployment of men, it is increasing the unemployment of immigrant women accompanying their spouses [28].

With an odds ratio of 0.14, the variable of house ownership also has significant effects on recent migrants. This figure shows that the inhabitant was 0.14 bound to be ongoing transients than those not claiming houses. That is, compared to people whose homes were owned, those who did not own their homes were 7.14 times more likely to migrate recently. It is showed on prior research that migrants mostly choose to rent a living place in the transition period when they first arrive in a new and unfamiliar city [29]. Other research argue that it is generally difficult for people to make the decision to migrate, and that there are a variety of factors that keep some people in one location. In addition, a number of studies suggest that house ownership has a negative impact on a person's decision to move, primarily due to the fact that moving out of a house one owns incurs more expenses than moving out of a house one rents. This is because transactional expenses are frequently assumed to be the responsibility of house owners [30].

4. CONCLUSION

This study finds that, of all six variables that have been determined, five have significant effects on the status of recent in-migrants migrating to the Special Region of Yogyakarta, according to the Bayesian technique in logistic regression with 500,000 iterations, two thinning intervals, and 250.000 burn-in. These five variables consist of resident's age (X_1), resident's marital status (X_3), resident's latest education (X_4), resident's main activities (X_5), and resident's homeownership status (X_6). Of these five variables, resident in the range of age of 17-40 years old, the resident that not yet married, the resident whose their latest education is high school or equal, the resident whose their main activity are currently working, and the resident renting a house are more likely to be recent in-migrants in Special Region of Yogyakarta. According to 0.717 AUC score, this model is deemed to be appropriate to explain the model with a fair discrimination category.

This research is expected to provide benefits to local governments in anticipating matters related to migration to Special Region of Yogyakarta such as improving school systems and facilities, procuring student- and worker-friendly public transportation, and others. Suggestions for future research are to use rare event logistic regression analysis to reduce the level of bias in the prediction of minority data, or latent class regression analysis can also be considered as another option. In addition, the use of methods that are robust to outliers can also be used to obtain better estimates. From the other side, further researchers can examine from an economic perspective the influences of migrants to migrate to Special Region of Yogyakarta, such as living comfort, low cost of living, etc.

REFERENCES

[1] M. B. Toney and A. K. Bailey, "Migration, an Overview," in *Encyclopedia of Quality of Life and Well-Being Research*, Dordrecht: Springer Netherlands, 2014, pp. 4044–4050. doi: 10.1007/978-94-007-0753-5_1804.

[2] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic Regression Model Optimization and Case Analysis," in *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2019*, Oct. 2019, pp. 135–139.

[3] J. Liang, G. Bi, and C. Zhan, "Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R," *Ann Transl Med*, vol. 8, no. 16, pp. 982–982, Aug. 2020, doi: 10.21037/atm-2020-57.

[4] N. Srimaneekarn, A. Hayter, W. Liu, and C. Tantipoj, "Binary Response Analysis Using Logistic Regression in Dentistry," *Int J Dent*, vol. 2022, pp. 1–7, Mar. 2022, doi: 10.1155/2022/5358602.

[5] M. Syairozi and K. Wijaya, "Migrasi Tenaga Kerja Informal: Studi Pada Kecamatan Sukorejo Kabupaten Pasuruan," *Seminar Nasional Sistem Informasi (SENASIF)*, vol. 4, no. 1, pp. 2383–2394, 2020.

[6] T. Andias and W. Syafitri, "Analisis Faktor-Faktor yang Mempengaruhi Keputusan Migran Bekerja di Dalam Negeri dan Luar Negeri (Studi Kasus Kecamatan Gondanglegi Kabupaten Malang)," *Jurnal Ilmiah Mahasiswa FEB Universitas Brawijaya*, vol. 2, no. 2, 2014.

[7] A. Zaiceva, *The Impact of Aging on The Scale of Migration*. IZA World of Labor, Institute of Labor Economics (IZA), 2014.

[8] C. Dustmann and J. Mestres, "Savings, Asset Holdings, and Temporary Migration," *Ann Econ Stat*, no. 97/98, p. 289, 2010, doi: 10.2307/41219119.

[9] A. B. Owen, "Infinitely Imbalanced Logistic Regression," 2007.

[10] G. King and L. Zeng, "Logistic Regression in Rare Events Data," *Political Analysis*, vol. 9, no. 2, pp. 137–163, 2001.

[11] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, Third Edition. New Jersey: John Wiley & Sons, 2013.

[12] S. T. Rachev, J. S. Hsu, B. S. Bagasheva, and F. J. Fabozzi, *Bayesian Methods in Finance*. John Wiley & Sons, 2008.

[13] J. K. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Science dan Business Media, 2007.

[14] H. Liu and D. A. Powers, "Bayesian Inference for zero inflated poisson regression modes. ," *Journal of Statistics : Advance in Theory and Aplications*, vol. 7, no. 2, pp. 155-188., 2012.

[15] A. Genkin, D. D. Lewis, and D. Mandigan, "Large Scale Bayesian Logistic Regression for Text Categorization," *Technometrics*, vol. 49, no. 3, pp. 291–304, 2007.

[16] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability & Statistics for Engineers & Scientist*, 9th ed. Pearson Education, Inc, 2012.

[17] I. Ntzoufras, *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, 2011.

[18] M. Guo, M. Stensland, M. Li, X. Dong, and A. Tiwari, "Is Migration at Older Age Associated With Poorer Psychological Well-Being? Evidence from Chinese Older Immigrants in the United States," *Gerontologist*, Jun. 2018, doi: 10.1093/geront/gny066.

[19] M. J. Benetsky, C. A. Burd, and M. A. Rapino, "Young Adult Migration," 2015.

[20] A. Zeki Al Hazzouri *et al.*, "Quantifying Lifecourse Drivers of International Migration," *Epidemiology*, vol. 32, no. 1, pp. 50–60, Jan. 2021, doi: 10.1097/EDE.0000000000001266.

[21] United Nations General Assembly, "The impact of migration on migrant women and girls: a gender perspective," Apr. 2019.

[22] B. S. A. Yeoh, "Migration and Gender Politics in Southeast Asia," *Migration, Mobility, & Displacement*, vol. 2, no. 1, 2016.

[23] D. W. Guntoro, S. Indartono, and I. Sholekhah, "Understanding gender, marital status & education as internal migration factors in developing countries," *Jurnal Pendidikan Geografi*, vol. 24, no. 2, pp. 98–107, Jun. 2019, doi: 10.17977/um017v24i22019p098.

[24] F. Borgonovi and A. Pokropek, "Education and Attitudes Toward Migration in a Cross Country Perspective," *Front Psychol*, vol. 10, Oct. 2019, doi: 10.3389/fpsyg.2019.02224.

[25] E. Browne, “Evidence on education as a driver for migration,” 2017.
[26] M. Urbański, “Comparing Push and Pull Factors Affecting Migration,” *Economies*, vol. 10, no. 1, p. 21, Jan. 2022, doi: 10.3390/economies10010021.
[27] I. B. Mantra, *Demografi Umum*. Yogyakarta: Pustaka Pelajar, 2012.
[28] A. K. Sahin and S. Aytac, “International Migration of Highly Qualified Housewife Mothers and Their Career Concerns,” *International Journal of Business and Management Research*, vol. 9, no. 1, pp. 45–50, Feb. 2021, doi: 10.37391/IJBMR.090107.
[29] S. Zhang, C. Hou, and J. Chen, “Homeownership, city integration, and the sense of happiness of migrants in urban China,” *Frontiers of Business Research in China*, vol. 13, no. 1, p. 1, Dec. 2019, doi: 10.1186/s11782-019-0050-5.
[30] A. C. Helderman, M. Ham, and C. H. Mulder, “Migration and Home Ownership,” *Tijdschrift voor Economische en Sociale Geografie*, vol. 97, no. 2, pp. 111–125, Apr. 2006, doi: 10.1111/j.1467-9663.2006.00506.x.