Statistika, Vol. 23 No. 1, 48-56
Mei 2023

Premarital Sex Behavior Model with Lasso Generalized Linear Mixed Model and Group Lasso Generalized Linear Mixed Model

KHALILAH NURFADILAH[1], ASFAR[2], KHAIRIL A. NOTODIPUTRO[3], BAGUS SARTONO[4], AZLAM NAS[5]

[1] Faculty of Science and Technology, UIN Alauddin Makassar, Indonesia
[2] Patompo University, Makassar, Indonesia
[3,4] Department of Statistics and Data Science, IPB University, Indonesia
[5] Statictics Of Takalar Regency, Indonesia
e-mail: khalilah@uin-alauddin.ac.id

ABSTRACT

Premarital sexual behavior is sexual behavior that is carried out between men and women without legal marriage. As the number of premarital sex increases, efforts need to take. One that can do is to identify the main factors contributing to reducing or increasing premarital sex behavior by a Regression model. In the context of sexual behavior, environmental influences cannot be ignored. GLMM is used to model data that is grouped into certain Groups, include environment effect that is modeled as mixed effect in GLMM. In terms of parsimony, the LASSO method can do selection variables. This research uses GLMM LASSO and GLMM Group LASSO as a model to approach the data. The best model that describes premarital sex behavior in South Sulawesi is the GLMM Group LASSO model based on the greatest AUC value. The variables that significantly influence the model are Type of Residence (X_1), Education Level (X_2), Literacy (X_3), Internet use (X_4), Knowledge of Contraceptive Methods (X_6), Health Insurance Ownership (X_7), Employment Status (X_8), Knowledge of Sexually Transmitted Diseases (X_9). By knowing the factors that influence premarital sex behavior, the government is expected to take the appropriate action for handling it.

Kata Kunci: GLMM LASSO, GLMM Group LASSO, Premarital Sex

1. INTRODUCTION

Premarital sexual behavior is sexual behavior that is carried out between men and women without legal marriage. he age at which adolescents engage in premarital sex varies across countries, ranging from 12 to 17.5 years, with an average age of onset at 15 (Guiella and Madise, 2007) (Adogu et al., 2014). Utomo and McDonald (2009) found that approximately 25% to 51% of adolescents have engaged in premarital sex. This finding aligns with the results of the Indonesian Youth Reproductive Health Survey (SKRRI) conducted in 2007, which showed that 6.4% of male and 1.3% of female adolescents had premarital sex (Rahyani et al., 2012).

Premarital sexual behavior among adolescents is a pressing issue that demands attention, as approximately 26.67% of individuals in the young age group (10-24 years) have engaged in premarital sex.(Berg et al., 2014) (Adogu et al., 2014) (Erlinda, 2014). Efforts need to be made to reduce the prevalence of premarital sex among adolescents. One of the efforts that can do is to identify the main factors contributing to reducing or increasing premarital sex behavior, that is, by the Regression model.

In the context of sexual behavior, environmental influences cannot be ignored. Individuals who are in the same environment have relatively similar behavior. In Statistics, data like this is termed hierarchical data. Several approaches can be used to model hierarchical data. One way is to use the Generalized Linear Mixed Model (GLMM). GLMM is used to model data that is grouped into certain groups. GLMM is the development of a linear model that includes random effects and fixed effects in models with response variables (outcomes) that do not have to be normally distributed and are usually used when the data is not independent and has a hierarchical structure.

One of the characteristics of the GLMM model is the link function. This function is a combination of fixed factors and random factors, which are defined as follows;

$$g(\mu_{it}) = \eta_{it} = x_{it}^T\beta + z_{it}^T b_i = \eta_{it}^{par} + \eta_{it}^{rand} \quad \text{…(1)}$$

Where I denote cluster (census block), t denotes observation (household), $\boldsymbol{\mu_{it}} = \boldsymbol{E}(\boldsymbol{Y_{it}}|\boldsymbol{b_i})$, $\boldsymbol{x_{it}^T}$ is the row vector of the explanatory variable, $\boldsymbol{\beta}$ is the parameter vector of the fixed effect, $\boldsymbol{b_i}$ is the random effect vector. Where $\boldsymbol{b_i} \sim N(\boldsymbol{0}, \boldsymbol{Q})$, and $\boldsymbol{Q}$ are the variance matrices of size $q\ x\ q$, and $\boldsymbol{z_{it}^T}$ are the row vectors of the covariates. Often, the random effect is univariate, as in this case, resulting in the equation $\boldsymbol{z_{it}^T} = \boldsymbol{1}$ (Agresti, 2007)(Agresti 2013). $\boldsymbol{\eta_{it}^{par}} = \boldsymbol{x_{it}^T\beta}$ β is arranged in a parametric linear equation in the parameter vector $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \ldots, \beta_p)$ and $\boldsymbol{\eta_{it}^{rand}} = \boldsymbol{z_{it}^T b_i}$ is the random effect of census blocks. g is a monotone and link function that depends on the distribution of the response variables used.

In GLMM, it is assumed that the conditional probability of y_{it} over the explanatory variable $\boldsymbol{x_{it}}$ and the random effect $\boldsymbol{b_i}$ is a form of the Exponential family.

$$f(\boldsymbol{y_{it}}|\boldsymbol{x_{it}}, \boldsymbol{b_i}) = \exp\left\{\frac{\left(\boldsymbol{y_{it}\theta_{it}} - k(\boldsymbol{\theta_{it}})\right)}{\boldsymbol{\phi}} + c(\boldsymbol{y_{it}}, \boldsymbol{\phi})\right\} \quad \text{…(2)}$$

$\boldsymbol{\theta_{it}} = \boldsymbol{\theta}(\boldsymbol{\mu_{it}})$ is a natural parameter, $k(\boldsymbol{\theta_{it}})$ is a specific function depending on the type of Exponential family, $c(.)$ is a log normalized constant and $\boldsymbol{\phi}$ is a dispersion parameter.

One method for maximizing GLMM can be done by applying Penalized Quasi Likelihood (PQL) (Breslow and Clayton (1993); Lin and Breslow (1996); Breslow and Lin (1995) and Muslim et al. (2018). The matrix of variance $\boldsymbol{Q}(\boldsymbol{\varrho})$ on random effects $\boldsymbol{b_i}$ depends on the unknown parameter vector $\boldsymbol{\varrho}$. On the basic concept of penalization, the combined probability function is defined by the parameter vector of the variance structure ϱ together with the dispersion parameter $\boldsymbol{\phi}$ in the form $\boldsymbol{\gamma^T} = (\boldsymbol{\phi}, \boldsymbol{\varrho}^T)$. The parameter vector $\boldsymbol{\delta}^T = (\boldsymbol{\beta}^T, \boldsymbol{b}^T)$ is the fixed and random effects parameter vector. Thus, the log function is obtained as follows:

$$l(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \log\left(\int f(\boldsymbol{y}_i|\boldsymbol{\delta}, \boldsymbol{\gamma}) p(\boldsymbol{b_i}, \boldsymbol{\gamma}) d\boldsymbol{b_i}\right) \quad \text{…(3)}$$

Where $p(\boldsymbol{b_i}, \boldsymbol{\gamma})$ is the probability function of random effects. Equation (3) can be approximated using the Laplace method so that the following Equation (4) is obtained:

$$l^{app}(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \log\left(f(\boldsymbol{y}_i|\boldsymbol{\delta}, \boldsymbol{\gamma})\right) - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{Q}(\boldsymbol{\varrho})^{-1}\boldsymbol{b} \quad \text{…(4)}$$

With the penalty form $\boldsymbol{b}^T\boldsymbol{Q}(\boldsymbol{\varrho})^{-1}\boldsymbol{b}$ for using the Laplace method approach (Gad and Kholy, 2012).

Data on premarital sex behavior were obtained from 2016 Indonesian Demographic and Health Survey (IDHS), and there were 12 explanatory variables identified as variables that influence premarital sex behavior among adolescents. One of the principles in statistical modeling is model simplicity or the parsimony principle. The simpler the model, the easier to interpret. Therefore it is necessary to make a variable selection. One of the variable selection methods that can be used is the Least Absolute Shrinkage and Selection Operator (LASSO). The LASSO method can select variables by reducing the regression coefficient to zero (Tibshirani, 1996). Adding the LASSO penalty to the GLMM equation is hoped to reduce the coefficient estimator to exactly zero, making it easier to interpret the model (James et al., 2021). However, when the explanatory variables used in modeling are categorical, the Group LASSO method can be used. The Group LASSO method is a development of the LASSO method, used when the explanatory variables have large dimensions and form Groups. When the variables used are categorical, the group is based on the dummy variable design formed and is called a dummy variable Group (Yuan and Lin, 2006).

LASSO, proposed by Tibshirani (1996), is a regression analysis method that performs variable selection and regularization to improve the prediction accuracy and interpretation of the resulting statistical model. The basic idea of this method came from Nonnegative Garrotte (Breiman, 1995). It started to gain attention after Efron, Hastie, Johnstone, and Tibshirani developed the Least Angle Regression (LAR) algorithm (2004). LASSO uses the L_1 penalty, i.e., $P(\beta) = \lambda\sum_{i=1}^{p}|\beta_i|$.

Groll and Tutz (2012) propose that the GLMM LASSO is formed by inserting the L_1 penalty element into the GLMM equation, so the following Equation will be obtained:

$$l^{Lasso}(\beta, b, \gamma) = l^{Lasso}(\delta, \gamma) = l^{app}(\delta, \gamma) - \lambda \sum_{i=1}^{p} |\beta_i| \quad \ldots(5)$$

Parameter estimation can be obtained by optimizing Equation (6) as follows:

$$\hat{\delta} = \underset{\delta}{\operatorname{argmax}}\, l^{Lasso}(\delta, \hat{\gamma}) = \underset{\delta}{\operatorname{argmax}} \left[l^{app}(\delta, \hat{\gamma}) - \lambda \sum_{i=1}^{p} |\beta_i| \right] \quad \ldots(6)$$

The optimization problem in Equation (9) can be solved using the gradient ascent algorithm Goeman (2010) developed.

The Group LASSO method was developed by adding Group constraints to the LASSO method. The Group LASSO method assumes that the explanatory variables are grouped and will intuitively drive all weights in one group to zero simultaneously, thereby leading to group selection (Yuan and Lin, 2006). Several previous studies, including Bakin (1999), proposed the LASSO Group and computational algorithms and were further developed by Yuan and Lin (2005). Another researcher who studies the theoretical nature of Group LASSO is Bach (2008), who shows that Group LASSO is a consistent group selection method in a random design model for a fixed parameter, p. Based on the description above, this study aims to identify the factors that influence premarital sex behavior in South Sulawesi Province using the GLMM LASSO and GLMM Group LASSO models.

Group LASSO (Yuan and Lin, 2006) is a generalization of LASSO for selecting Group variables. Grouping variables aims to facilitate the selection of variables that have similar characteristics. This method selects Grouped variables in large sizes with unequal numbers and gives better results than the LASSO method (Huang and Zhang (2009) and Lounici et al. (2011)). The LASSO Group uses the penalty $\sum_{j=1}^{J} \left(\beta_j^T K_j \beta_j\right)^{\frac{1}{2}}$, where $\lambda \geq 0$ is the specified parameter and $K_1, K_2, \ldots, K_J$ is the positive definite matrix. The Group LASSO estimator is asymptotically consistent even when the model complexity increases with increasing sample size (Nardi, 2008).

By substituting the penalty of the Group LASSO to GLMM, the Equation below is obtained:

$$l^{GLasso}(\boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\gamma}) = l^{GLasso}(\boldsymbol{\delta}, \boldsymbol{\gamma}) = l^{app}(\boldsymbol{\delta}, \boldsymbol{\gamma}) - \lambda \sum_{j=1}^{J} \left(\beta_j^T K_j \beta_j\right)^{\frac{1}{2}} \quad \ldots(7)$$

Parameter estimation can be obtained by optimizing Equation (8) as follows:

$$\hat{\boldsymbol{\delta}} = \underset{\boldsymbol{\delta}}{\operatorname{argmax}}\, l^{GLasso}(\boldsymbol{\delta}, \hat{\boldsymbol{\gamma}}) = \underset{\boldsymbol{\delta}}{\operatorname{argmax}} \left[l^{app}(\boldsymbol{\delta}, \hat{\boldsymbol{\gamma}}) - \lambda \sum_{j=1}^{J} \left(\beta_j^T K_j \beta_j\right)^{\frac{1}{2}} \right] \quad \ldots(8)$$

2. METHODS

2.1 Data

The data used in this study is from BPS's 2016 IDHS survey in South Sulawesi. The variables studied were male premarital sexual behavior. A total of 251 observations were included in the model. In variable Y, the data used is binary: have and have never had premarital sex. This information is obtained from the difference in age at marriage and the age at which you first had sex. If the age at first having sex is younger than the age of marriage, then the individual is categorized as having had premarital sex, symbolized as 1, and 0 otherwise. Data on this variable is limited to events in the last 3 years. For independent variables, the existing categories are those made in the IDHS questionnaires.

Table 1. Variables

Variables	Description	Scales
Y	Premarital sex behavior	Nominal 0 = never had premarital sex 1 = ever
Fixed Factor		
X_1	Type of Residence	Nominal 1 = urban 0 = rural
X_2	Level of education	Ordinal 0 = Never went to school 1 = Primary 2 = Secondary 3 = Higher
X_3	Literacy (reading ability)	Ordinal 0 = illiterate 1 = Able to read some sentences 2 = Able to read whole sentences 3 = Blind
X_4	Internet Use	Nominal 0 = Never 1 = in the Last 12 months 2 = less than 12 months 3 = rare use
X_5	Wealth Index	Ordinal 1 = Very Poor 2 = Poor 3 = Moderate 4 = Rich enough 5 = Very rich
X_6	Knowledge of contraceptive methods	Nominal 0 = don't know 1 = know
X_7	Having Health insurance	Nominal 0 = None 1 = Have minimum a health insurance
X_8	Profession	Nominal 0 = Jobless 1 = Worker
X_9	Knowledge about the Sexually transmitted disease	Nominal 0 = None 1 = Have enough knowledge
X_{10}	Knowledge about AIDS	Nominal 0 = None 1 = Have enough knowledge
Random Factor		
Z	Block Census	Nominal

2.1 Analysis

The process of data analysis carried out in this study is as follows:

1. The data exploration focused on examining the general characteristics of the variables;
2. Data Preparation, by making the categorical type variable a dummy variable in binary form (0,1) as in the Y variable, if the age at first having sex is younger than the age of marriage, then the individual is categorized as having had premarital sex symbolized 1, 0 otherwise;
3. Modeling data by GLMM LASSO;
4. Modeling data by GLMM GROUP LASSO;
5. Determine the best model by Goodness of Fit Criteria;
6. Interpret the best model.

3. RESULT AND DISCUSSION

3.1 Data Preparation

Data preparation steps are as follows:

1. Remove the missing data
2. Variable reduction:
   a. Variable Literacy, X_3, firstly defined in 4 categories and then reduced to 3 categories according to the respondent's responses in the survey.
   b. Variable Internet Use, X_4, firstly defined in 4 categories and then reduced to 3 categories according to the respondent's responses in the survey.
3. Transform the categorical variables into dummy variables with binary values (0.1), as shown in Table 4 below.

Table 2. Dummy Variables

Explanatory Variables	Initial notation	Dummy Notation	Description
Type of Residence	X_1	D_{11}	0 = Rural 1 = Urban
Education Level	X_2	D_{21}	0 = otherwise 1 = Primary
		D_{22}	0 = otherwise 1 = Secondary
		D_{23}	0 = otherwise 1 = Higher
Literacy	X_3	D_{31}	0 = otherwise 1 = Able to read some sentences
		D_{32}	0 = otherwise 1 = Able to read whole sentences
Internet Use	X_4	D_{41}	0 = otherwise 1 = in the Last 12 months
		D_{42}	0 = otherwise 1 = less than 12 months
Wealth Index	X_5	D_{51}	0 = otherwise 1 = Poor
		D_{52}	0 = otherwise 1 = Moderate
		D_{53}	0 = otherwise 1 = Rich enough
		D_{54}	0 = otherwise 1 = very rich
Knowledge of contraceptive methods	X_6	D_{61}	0 = otherwise 1 = know
Having Health insurance	X_7	D_{71}	0 = None 1 = Have minimum a health insurance
Profession	X_8	D_{81}	0 = Jobless 1 = Worker
Knowledge about the Sexually transmitted disease	X_9	D_{91}	0 = None 1 = Have enough knowledge
Knowledge about AIDS	X_{10}	D_{101}	0 = None 1 = Have enough knowledge

3.2 Modeling by GLMM LASSO

One of the most important steps that should follow before modeling by the GLMM LASSO regression is the process of tuning the parameter λ. The tuning parameter λ functions to control the explanatory variable coefficients. The greater the value of λ, the simpler the regression model that is formed, but the prediction error tends to be large and otherwise. Optimum λ determination was carried out using a 10-fold Cross Validation (CV) with a measure of the mean cross-validated error (cvm). After 100 iterations, the optimum λ values in Table 3 are obtained as follows:

Table 3. Optimum λ optimum based on the *deviance* in GLMM LASSO

λ **type**	λ	***Deviance***
Lambda.1se	1.0302266774	0.0865478
Lambda.min	0.0209136016	0.1247398

Determining the value of λ is important because it affects the number of variables to be selected. A good λ is a λ that can produce the smallest model deviation. Based on Table 3, the value of the tuning parameter λ used in the GLMM LASSO modeling in the case of premarital sex behavior in South Sulawesi is Lambda.1se (λ= 1.0302266774) with a deviation value of 0.0865478. Following are the results of parameter estimation of the GLMM LASSO model using the tunning parameter λ = 1.0302266774. Thus, the GLMM LASSO model is obtained as follows:

$$\begin{aligned} g(\boldsymbol{\mu}_{it}) = -2.6827550 - 0.2695400D_{11} + 0.5201711D_{21} - 0.0035514D_{22} \\ - 0.1538596D_{23} + 0.1640216D_{32} + 1.7693479D_{42} - 0.0036412D_{53} \\ - 0.1088079D_{54} + 0.1816142D_{61} - 0.3324082D_{71} - 2.0829153D_{81} \\ + 0.5115377D_{91} - 0.0056984D_{101} \end{aligned} \quad \ldots(9)$$

3.3 Modeling by GLMM Group LASSO

As with the GLMM LASSO method, the step that must be taken before modeling by the GLMM Group LASSO model is to determine the value of λ by tuning the parameter λ using a 10-fold Cross Validation. The main objective of this study's GLMM Group LASSO method is to identify functional groups that influence premarital sex behavior by selecting groups of variables. The group of variables used in this study is based on categorical variables. After 100 iterations, the optimum λ values in Table 4 are obtained as follows:

Table 4. Optimum λ based on deviance in GLMM Group LASSO

Jenis λ	**Nilai** λ	**Nilai** ***Deviance***
Lambda.1se	1.0302266774	0.08326013
Lambda.min	0.0105083121	0.12446880

Thus, the GLMM Group LASSO method can simplify the model by selecting variables based on the variable group. *The fixed effect* model with the best λ by GLMM Group LASSO is as follows:

$$\begin{aligned} g(\boldsymbol{\mu}_{it}) =- 2.5353735 - 0.2939631D_{11} + 0.4027095D_{21} - 0.0500352D_{22} \\ - 0.1261231D_{23} - 0.0029288D_{31} + 0.0096018D_{32} + 0.0870814D_{41} \\ + 1.7436159D_{42} + 0.2326118D_{61} - 0.4482858D_{71} - 2.2106361D_{81} \\ + 0.6299373D_{91} \end{aligned} \quad \ldots(10)$$

3.4 Goodness of Fit

The size of the goodness of the model can be measured using the Area Under Curve (AUC), which is the area under the Receiver Operating Characteristic (ROC) curve. ROC is a plot between sensitivity and (1- specificity). The higher the AUC value, the better the prediction (Agresti 2013). The AUC value criteria are divided into several categories according to Gorunescu (2011), as in Table 2 below:

Table 5. AUC Value Criterion

Nilai AUC	**Interpretation**
0.9 – 1.0	Excellent classification
0.8 – 0.9	Good Classification
0.7 – 0.8	Fair Classification
0.6 – 0.7	Poor Classification
0.5 – 0.6	Failure

From the GLMM LASSO and GLMM Group LASSO models previously obtained, it can be determined which model is better at selecting variables in modeling cases of premarital sex behavior in South Sulawesi. The goodness of fit model can be measured by calculating the AUC value of each model obtained.

Table 6. Determination of the AUC value of each model

Model	AUC
GLMM LASSO	0.9039634
GLMM Group LASSO	0.9197154

Table 6 shows that the GLMM Group LASSO model provides a higher AUC value than the GLMM LASSO model, with an AUC value of 0.9197154, which indicates that the GLMM Group LASSO model is capable of very good classification accuracy. So the best model to be used in modeling premarital sex cases in South Sulawesi is the GLMM Group LASSO model. Although the AUC values of the two models are close and the number of selected variables is only 1 different, the two models provide very good classifications. This is clarified by looking at the ROC plots of the two models.

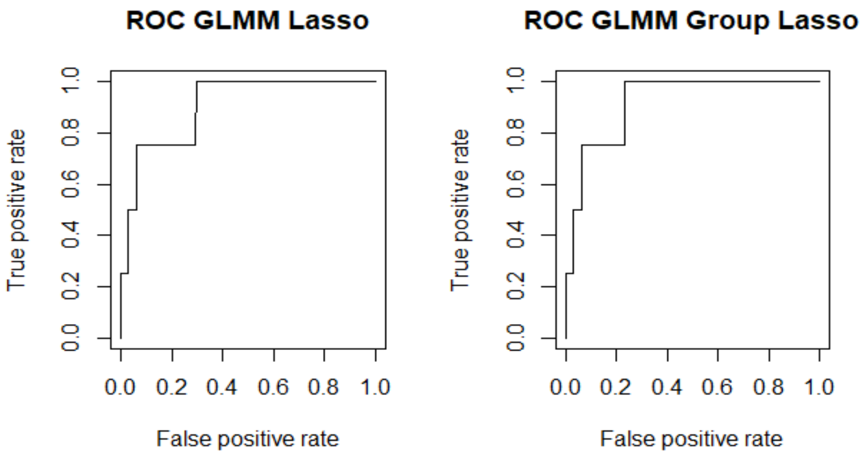

Figure 1. The plot of ROC in GLMM LASSO and GLMM Group LASSO

Based on Figure 2 above, it can be seen that the GLMM LASSO and GLMM Group LASSO models have similar plot shapes. And both plots show that the GLMM LASSO and GLMM Group LASSO models predict more negative values incorrectly.

3.4 Discussion

The results of the previous analyses show that the model used in describing premarital sex behavior in South Sulawesi is the GLMM Group LASSO model. The GLMM Group LASSO model is formed from two effects: the fixed effect and the random effect. To investigate what factors have an influence, we can look at the fixed effect model. The fixed effect model shows that the variable group selected in the model is the Wealth Index variable (X_5) and the Knowledge about AIDS variable (X_{10}). So that the variables that significantly influence the model are Type of Residence (X_1), Education Level (X_2), Literacy (X_3), Internet use (X_4), Knowledge of Contraceptive Methods (X_6), Health Insurance Ownership (X_7), Profession (X_8), Knowledge of Sexually Transmitted Diseases (X_9).

Type of residence (X_1) significantly affects premarital sex behavior. This is due to the influence of the development of information technology, support from the mass media, and the weakening of parental supervision and control from society. Technological advances are not only in urban areas but have penetrated rural areas making it easier to access information. Education level (X_2) has a significant role in premarital sexual behavior. Low education or even dropping out of school will result in a person not getting the correct information, especially regarding issues of sexuality and reproductive health. A good education will improve a person's understanding of information, and they will be able to consider the positive and negative things from that knowledge. This is also related to literacy skills (X_3); the higher one's education, the better one's literacy skills will be.

Internet use (X_4) is one of the causes of premarital sex behavior among adolescents. This is because when someone surfs the internet, it will be very easy to be influenced by advertisements or pornographic shows, affecting premarital sex behavior. In addition, knowledge of contraception methods (X_6) can make a person easy to engage in premarital sex. In addition, the ease of obtaining information on contraceptive methods also plays a role in increasing premarital sex behavior among adolescents.

Knowledge about sexually transmitted diseases (X_9) is one of the factors that plays a very important role in reducing premarital sex behavior in the community. Understanding STDs makes a person more careful in having sex with other than a partner (husband or wife), especially when having sex freely and having multiple partners.

Next is the random effect, namely the census block parameter, which is the standard deviation of the census block. The standard deviation shows the variation between census blocks within a city/district. The higher the value, the more diverse or different the conditions between census blocks within a city/regency. The standard deviation value of the census block in the GLMM Group LASSO model is 0.3318650. The standard deviation value of the GLMM Group LASSO model is relatively small, and this indicates that the conditions between census blocks tend to be more uniform as a result of the fact that premarital sex behavior in South Sulawesi tends to be evenly distributed.

4. SIMPULAN DAN SARAN

The best model that describes premarital sex behavior in South Sulawesi is the GLMM Group LASSO model. The variables that generally have a significant effect on cases of premarital sex behavior are the Wealth Index (X_5) and the Knowledge of AIDS variable (X_{10}). So that the variables that significantly influence the model are Type of Residence (X_1), Education Level (X_2), Literacy (X_3), Internet use (X_4), Knowledge of Contraceptive Methods (X_6), Health Insurance Ownership (X_7), Employment Status (X_8), Knowledge of Sexually Transmitted Diseases (X_9).

The standard deviation of the census block in the GLMM Group LASSO model is relatively small, indicating that premarital sex behavior in South Sulawesi is evenly distributed.

In the next study, in conducting the GLMM Group LASSO analysis, to incorporate spatial effects into the model and compare it with other variable selection methods such as Elastic Net Fused LASSO.

DAFTAR PUSTAKA

Agresti, A. 2007. 'An Introduction to Categorical Data Analysis Second Edition'. Canada (CA): John Wiley and Sons.

Agresti A. 2013. 'Categorical Data Analysis, Third Edition. Canada (CA)': John Wiley and Sons.

Bach, F. R. (2008). 'Consistency of the group Lasso and multiple kernel learning'. J. Mac. Learn. Res., to appear.

Bakin, S. (1999). 'Adaptive regression and model selection in data mining problems'. Ph.D. Thesis. Australian National University, Canberra.

Balitbang Kemenkes RI. 2013. 'Riset Kesehatan Dasar'; RISKESDAS. Jakarta: Balitbang Kemenkes RI.

Breiman, L. 1995. 'Better subset regression using the nonnegative garrote'. Technometrics, Vol. 37, No 4, pp. 373-383.

Breslow, N.E dan Clayton, D.G. 1993. 'Approximate inference in generalized linear mixed models'. J. of the American Statistical Association 88, 9-25.

Breslow, N. E., & Lin, X. (1995). 'Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion'. Biometrika, 82(1), 81–91. https://doi.org/10.2307/2337629

E.Gbar, E., W.Stroup, W., S.McCarter, K., Durham, S., J.Young, L., Christman, M., Kramer, M. (2012). 'Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences'. USA: Book and Multimedia Publishing Committee.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). 'Least Angle Regression. The Annals of Statistics', 32(2), 407–451. http://www.jstor.org/stable/3448465

Erlinda. 2014. 'Upaya Peningkatan Anak dari Bahaya Kekerasan, Pelecehan dan Eksploitasi'. Jakarta: KPAI.

Gad A M, Kholy R S, 2012, 'Generalized Linear Mixed Models for Longitudinal Data', International Journal of Probability and Statistics 2012, 1(3): 67-73 DOI:10.5923/j.ijps. 20120103.03, Cairo.

Goeman JJ. L1. 'penalized estimation in the Cox proportional hazards model'. Biom J. 2010 Feb;52(1):70-84. doi: 10.1002/bimj.200900028. PMID: 19937997.

Gorunescu F. 2011. 'Data Mining: Concepts, Models, and Techniques. Berlin': Springer

Guiella, G., & Madise, N. J. (2007). 'HIV/AIDS and Sexual-Risk Behaviors among Adolescents: Factors Influencing the Use of Condoms in Burkina Faso'. African Journal of Reproductive Health / La Revue Africaine de La Santé Reproductive, 11(3), 182–196. https://doi.org/10.2307/25549739

Groll A, Tutz G. 2012. 'Variable selection for generalized linear mixed models by L1-penalized estimation'. Statistics and Computing. 24: 137-154. DOI:10.1007/s11222-012-9359-z

Huang J, Zhang T. 2009. 'The Benefit of Group Sparsity'. Ann Statist.38:5277-5286.

James G, Witted D, Hastie T, Tibshirani R. 2021. 'An Introduction to Statistical Learning with Application in R, Second Edition'. New York (US): Springer.

Kachman, S. D. 1998. 'An Introduction to generalized linear mixed models'. Dept. of Biometry, Univ. of Nebraska-Lincoln.

Kurnia, A. 2000. 'Pendekatan GEE dan Quasi-Likelihood dalam Generalized Linear Mixed Models'. Bogor: Institut Pertanian Bogor

Lounici, K., Pontil,M., Van, GS., and Tsybakov, AB. 2011. 'Oracle Inequalities and Optimal Inference Under Group Sparsity'. Ann Statist 39:2164-2204.

Muslim A, Hayati M, Sartono B, Notodiputro KA. 2018. 'A Combined Modeling of Generalized Mixed Model and LASSO Technique for Analyzing Monthly Rainfall Data'. IOP Conference Series: Earth and Environmental Science.

Nardi, Y., and Rinaldo, A. 2008. 'One of the Asymptotic Properties of The Group LASSO Estimator for Linear Models. Electronic Journal of Statistics'. ISSN: 1935-7524.

Tibshirani, R. (1996). 'Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society': Series B (Methodological), 58(1): 267–288.

Utomo ID, McDonald P. 'Adolescent reproductive health in Indonesia: contested values and policy inaction'. Stud Fam Plann. 2009 Jun;40(2):133-46. doi: 10.1111/j.1728-4465.2009.00196.x. PMID: 19662805.

Yuan M, Lin Y. 2006. 'Model Selection and Estimation in Regression with Grouped Variables'. Journal of the Royal Statistical Society Series B 68(1): 49-67

Yuni Rahyani, Ni Komang And Utarini, Adi And Agus Wilopo, Siswanto And Hakimi, Mohammad (2012) 'Perilaku Seks Pranikah Remaja. Jurnal Kesehatan Masyarakat Nasional, Kesmas', 7 (4). pp. 180-185.

Wolfinger, R.D., & O'connell, M.A. (1993). 'Generalized linear mixed models a pseudo-likelihood approach'. Journal of Statistical Computation and Simulation, 48, 233-243.