Analysis of Multidimensional Stunting Intervention Factor Using Mixed Model

D N Agustina[1,2]*, **B Sartono**[1], **K A Notodiputro**[1]

[1] Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
[2] Statistics Indonesia

*Corresponding author: agustinadin.nurika@apps.ipb.ac.id

Abstract. The mixed model combines fixed effect for all groups and random effect representing the diversity inter groups in the model (province) to increase the model precision. This study provides information on the significance of multidimensional stunting intervention factors (predictor variables) on stunting prevalence (response variables as indicator 2.2.1 Sustainable Development Goals /SDGs) with district/city as observation units. Using official data from Statistics Indonesia (National Socio Economic Survey) and Ministry of Health (Basic Health Research), this study expects to be one basis of information for the government, stakeholders, and further research to accelerate Indonesia's SDGs targets in 2030. Comparison of classical linear mixed model method and linear mixed model with Least Absolute Shrinkage and Selection Operator (Lasso) variable selection conduct with relatively better results of mixed linear modelling with Lasso. The results showed that the predictor variables, namely complete immunization, ease of access to health facilities, diversity of food intake, improve water, food expenditure per capita, children's participation in early childhood education, maternal education, and ownership of National Health Insurance for toddlers, significantly affected the stunting prevalence decrease. The predictor variables, namely low birth weight, households with social protection cards, and the percentage of poor people, significantly increase the stunting prevalence.
Keywords: *stunting, susenas, riskesdas, Indonesia 2018, mixed model*

1. Introduction

1.1. Background

Stunting is defined by the World Health Organization (WHO [1] as the percentage of children aged 0 (zero) to 59 months (children under five years), with a height of less than -2 (minus two) standard deviations from the child growth standard. It is generally referred as stunting medium stunting and heavy stunting. Malnutrition in early childhood, indexed by stunting, is associated with poor cognitive function [2]. They are also stunting one of the obstacles to human development efforts and efforts to break the poverty chain. Indications of failure to thrive in children under five (stunting) are associated with the cause of suboptimal brain development and a higher risk of suffering from chronic diseases in adulthood, thereby inhibiting maximum contribution to country development.

United Nations (UN) Decade of Action on Nutrition 2016-2025 (“Nutrition Decade”) to accelerate achieve the global nutrition [3]. It is in line with multidimensional efforts to prevent and overcome stunting also accelerate reducing its prevalence launched by The Indonesian Government (specific nutrition interventions and sensitive nutrition interventions). The interventions interpret as nutritional and indirect nutrition that effectively addresses maternal and child malnutrition [4].

The statistical model builds to determine the significance and influence of stunting intervention factors (predictor variables) on stunting prevalence (response variables). The data used in this study is an essential indicator data (indicator 2.2.1) of Sustainable Development Goals (SDGs) with 513 districts/cities unit of observation. Based on exploration, stunting prevalence data by district/city has a normal distribution of data with almost the same characteristics intra-province.

Each unit of observation in the classical linear modeling is treated based on the assumption of independent and identical data and containing only fixed effects in the model. In one of the classical linear model developments, each observation unit with the same character tendency in a group (cluster) gets more treatment. The group is treated as a random effect in the model as an addition to a fixed effect. Modeling methods containing random effects and fixed effects are called mixed-effect linear models or linear mixed models.

The mixed linear model accommodates data characteristics that have similarities in a group by combining the fixed effect parameters and random-effects parameters in the model to minimize undefined residuals in the model [5]. Fixed effects are assumed to be well-defined and repeatable and are valid for the entire population in the study unit. Meanwhile, random effects apply to experimental units or observations in certain groups of units in the study to represent the variation inter-group.

1.2. Objectives
This study intends to build a model to analyze the significance and influence of stunting intervention factors (predictor variables) on stunting prevalence (response variables). The application of classical linear mixed modeling method and linear mixed modeling with Lasso variable selection (Least Absolute Shrinkage and Selection Operator) in this study aims to increase understanding of model variations and as an effort to present alternative considerations of the goodness of the model. Furthermore, the study result also expects to be helpful for the government, policymakers, and further research to identify the significance of multidimensional stunting intervention factors to achieve an accelerated reduction in the prevalence of stunting according to Indonesia's SDGs target in 2030, which is 10 percent.

2. Methods

2.1. Data
This study explores data from the 2018 National Socio-Economic Survey (Susenas) and data from the 2018 Basic Health Research Results (Riskesdas) with 513 regencies/cities as observation units. Susenas 2018 and Riskesdas 2018 were carried out in an integrated manner or other words, Susenas 2018 respondents were also interviewed as respondents to Riskesdas 2018. The survey methodology was a two-stage one-phase stratified sampling [6] with a target sample of 325,000 households in all districts/cities in Indonesia to estimate statistics at district/city, province, and national level.

United Nations Children’s Fund (UNICEF) states that children at a younger age, children living in poor households, and those living in disadvantaged areas are the most vulnerable to having the worst eating patterns [7]. Previous studies have also stated that the economy and education [8], health status, and nutrition [9], also access to improve water and improve sanitation [10], significantly influence the prevalence of stunting. Considering the multi-dimensional stunting intervention factors in the Special Stunting Handling Index [11], the predictor variables were determined, namely complete immunization and ease of access to health facilities, use of long-term contraceptive methods, low birth weight, diversity of food intake, improve water, improve sanitation, percentage of poor population, food expenditure per capita, child participation in early childhood education, mother's education, ownership of social protection card (kartu perlindungan sosial/KPS) or prosperous family card (kartu keluarga sejahtera/KKS) in the household, and the privilege of National Health Insurance Card for children under five.

The details structure of the predictor variables (fixed effect) and the response variables in this study are as shown in table 1.

Table 1. List of variable names, descriptions, and data sources.

Variables Names		Descriptions	Data Source
Response Variables:			
Stunting Prevalence	(Y)	Percentage of children 0-59 months (under five years old) with a height < -2 standard deviations from the child growth standard (according to the World Health Organization/ WHO).	Riskesdas
Predictor Variables:			
Complete immunization	(X1)	Percentage of children 12-23 months who received minimal complete immunization (BCG=1, DPT=3, Polio=3, HB=3, and Measles=1)	Riskesdas
Ease of access to health facilities	(X2)	Percentage of households who think it is easy access to community health centers, mobile health centers, or village midwives	Riskesdas
Long-term contraceptive methods use	(X3)	Percentage of women of childbearing age (15-49 years) currently married and using long-term contraceptive methods (Male Sterile, Female Sterile, IUD, injections, or implants)	Susenas
Low birth weight	(X4)	Percentage of children 0-59 months with low birth weight (< 2.5 kg)	Riskesdas
Diversity of food intake	(X5)	Percentage of children 6-23 months with diverse food intake, which fulfills more than 4 (four) categories (with categories: cereals and tubers, nuts, milk and their products, meat, eggs, vegetables and fruit sources of vitamin A, as well as other vegetables and fruits)	Riskesdas
Improve water	(X6)	Percentage of households that have access to safe water, namely the main water source from pipes, rainwater, or wells (bore wells/pumps or protected wells or protected springs with a distance to the final disposal site of 10 m and more).	Susenas
Improve sanitation	(X7)	Percentage of households having access to proper sanitation, namely having defecation facilities that are used by household members alone or with other household members on a limited basis, type of goose-neck toilet and final disposal of feces in the form of a septic tank or wastewater treatment plant.	Susenas
Percentage of poor population	(X8)	Percentage of population below the poverty line	Susenas
Food expenditure per capita	(X9)	Average per capita expenditure on food in a month (units of tens thousands rupiah)	Susenas
Child participation in early childhood education	(X10)	Percentage of participation of children aged < 10 years in Early Childhood Education	Susenas
Mother's education	(X11)	Percentage of women 15-49 years of age who have ever been married (married, divorced, divorced) with a minimum senor high school/equivalent	Susenas
Ownership of social protection card	(X12)	Percentage of households that have a Social Protection Card/Prosperous Family Card	Susenas
Privilege of JKN for children under five	(X13)	Percentage of children aged 0-59 months who have a National Health Insurance Card	Susenas

2.2. Linear mixed model

The observations of this study are a number of m = 34 provincial groups (i = 1,2,…,m) and a number of n = 513 districts/cities (j = 1,2,…,n). Each research unit of the response variable (district/city stunting prevalence) is expressed in vector form as $y_{ij}^T = (y_{11}, \ldots, y_{mn})$. Then, $x_{ij}^T = (1, x_{ij1}, \ldots, x_{mnp})$ is a fixed effect vector associated with a number of p = 13 predictor variables. Furthermore, $z_{ij}^T = (z_{ij}, \ldots, z_{mn})$ is a vector associated with a random effect.

Referring to the explanation by Groll, A. and Tutz Gerhard [12], research observation of y_{ij} are assumed to be conditionally independent with a mean of $\mu_{ij} = E(y_{ij}|b_i , x_{ij}, z_{ij})$ and variance $var(y_{ij}|b_i) = \varphi\upsilon(\mu_{ij})$, where $\upsilon(.)$ is a known variance function and φ is scalar parameter. The linear mixed model for the variables of this study is stated as follows:

$$y_{ij} = x_{ij}^T \beta + z_{ij}^T b_i = \eta_{ij}^{par} + \eta_{ij}^{rand} \tag{1}$$

where $\eta_{ij}^{par} = x_{ij}^T \beta$, and β is a vector of regression coefficients including the intercept $\beta^T = (\beta_0, \beta_1, \ldots, \beta_p)$. $\eta_{ij}^{rand} = z_{ij}^T b_i$, with b_i s a random effect parameter.

The mixed-effects model combines two vectors of random variables, namely the observed response variable (Y) with n dimensions and the unobserved random effect vector $\mathbb{B}$ with q dimensions. Bates, D [13] states that the mixed model is described in an unconditional distribution of $\mathbb{B}$ and a conditional distribution of Y with a specific $\mathbb{B}$ $(Y|\mathbb{B})$ equal to b.

The unconditional distribution of $\mathbb{B}$ s a Gaussian multivariate or q-dimensional normal with a mean of 0 (zero) and a parameterized variance (covariance) matrix as follows:

$$\mathbb{B} \sim \mathcal{N}(0, \sigma^2\Lambda(\theta)\, \Lambda'(\theta)) \tag{2}$$

where the scalar value σ is called the general scale parameter.. The matriks $\Lambda(\theta)$ of size q x q is the left factor of the diversification matrix (when σ=1) or the relative variance matrix (when $\sigma \neq 1$).

The conditional distribution Y with a specific $\mathbb{B}$ $(Y|\mathbb{B} = b)$ is a particular form of a multivariate Gaussian distribution, where I_n is an identity matrix of size n, and the conditional mean $\mu_{Y|\mathbb{B}}(b)$ s a linear predictor of $X\beta + Zb$, which can be written with the formula:

$$(Y|\mathbb{B} = b) \sim \mathcal{N}(X\beta + Zb, \sigma^2 I_n) \tag{3}$$

Conditional distribution $(Y|\mathbb{B} = b)$ depends on b through a linear predictor. The linear transformation of the "spherical" Gaussian q-dimensional random variable u of $\mathbb{B}$ with a central mean of 0 (zero) is:

$$\mathbb{B} = \Lambda(\theta)\boldsymbol{u},\ \boldsymbol{u} \sim \mathcal{N}(0, \sigma^2 I_q) \tag{4}$$

Linear predictor as a function of u is written as:

$$\gamma\, (\boldsymbol{u}) = Z\, \Lambda(\theta)\boldsymbol{u} + X\beta \tag{5}$$

Furthermore, the emphasis on model parameters, namely θ and β in the formulation γ, will be written as linear predictors $\gamma\, (\boldsymbol{u}, \theta, \beta)$.

The observed value is y, while b or u are not observed so that for statistical inference purposes, a conditional distribution of y is used, namely $(\boldsymbol{u}|y)$ or equivalent to $(\mathbb{B}|y)$. This conditional distribution is a continuous distribution with the conditional probability density function $f_{\boldsymbol{u}|y}(\boldsymbol{u}|y)$. Decide $f_{\boldsymbol{u}|y}(\boldsymbol{u}|y)$ as the product of the unconditional density $f_{\boldsymbol{u}}\, (\boldsymbol{u})$, and the conditional density function or probability mass function $f_{y|\boldsymbol{u}}(y|\boldsymbol{u})$ expressed as the conditional unnormalized density (since the conditional density is proportional concerning $h(\boldsymbol{u}|y, \theta, \beta, \sigma)$ particularly:

$$h(\boldsymbol{u}|y, \theta, \beta, \sigma) = f_{\, y|\boldsymbol{u}}(y|\boldsymbol{u}, \theta, \beta, \sigma)\ f_{\boldsymbol{u}}\, (\boldsymbol{u}|\sigma) \tag{6}$$

Function h needs to be normalized by dividing by the integral value to obtain the normalized conditional density, by the equation:

$$L(\theta, \beta, \sigma|y) = \int_{\mathbb{R}^q} h\, (\boldsymbol{u}|y, \theta, \beta, \sigma)\, du \tag{7}$$

The value of $L(\theta, \beta, \sigma|y)$ is the likelihood equation for the parameters θ, β, and σ ased on specific observed y data. The maximum likelihood estimate of the parameter is the value that maximizes L.

Linear mixed model in this study was built by the lme4 program package on the r data processing software. Bates, D. et al. [14] stated that the lmer formula in the lme4 program package [15] calculates the maximum likelihood of a linear mixed model involving repeated application of the penalized least square (PLS) method. In particular, the PLS problem is to minimize the number of weighted residual squares that are penalized for obtaining an estimate of the regression coefficient parameter (β) of a linear mixed model at the estimated value of the optimum variance parameter $\theta = \hat{\theta}$. Then, for the estimation of variance σ^2 restricted maximum likelihood (REML) is applied.

2.3. Linear mixed model with lasso variable selection

The use of linear mixed models is usually limited to a few predictor variables because many predictor variables result in unstable estimates. The approach presented for fitting a general linear mixed model includes the penalty term L1 [16]. It stated that the approach presented for fitting a general linear mixed model includes the penalty term L1, which simultaneously applies variable selection and depreciation. The algorithm used allows a model that maximizes the penalty log-likelihood to produce a model with reduced complexity. High-dimensional model settings where many predictor variables are available, including potentially influential ones, can use this method.

Representation approach for the equation (1) for all observations in the matrix design is:

$$g(\mu) = X\beta + Zb \tag{8}$$

where $X^T = [X_1^{T,}, \ldots, X_n^T]$, and the block-diagonal matrix (*block-diagonal*) $Z = (Z_1, \ldots, Z_n)$. The random effect vectors $b^T = (b_1^{T,}, \ldots, b_n^T)$ each follow a normal distribution with the block-diagonal covariance matrix $Q_b = \mathrm{diag}(Q, \ldots, Q)$.

Considering the linear mixed model, then the assumption of conditional density y_{ij} of certain predictor variables and random effects b_i has the form of an exponential family:

$$f(y_{it}|x_{it}, b_i) = \exp\left\{\frac{(y_{it}\,\theta_{it} - \varkappa(\theta_{it})}{\phi} + c(y_{it}, \phi)\right\} \tag{9}$$

where $\theta_{it} = \theta(\mu_{it})$ is a natural parameter. $\varkappa(\theta_{it})$ is a specific function related to the type of the exponential family (in this study, it means normal function). ϕ is the dispersion parameter. c(.)s the log normalization constant and is the distribution parameter.

Considering the linear mixed model, covariance matrix $Q(\varrho)$ of the random effect b_i depends on the unknown q vector parameter. The penalization-based concept of the shared likelihood function is specified based on the parameter vector of the covariance structure ϱ and the dispersion parameter ϕ, combined in $\gamma^T = (\phi, \varrho^T)$ and parameter vector $\delta^T = (\beta^T, b^T)$. The corresponding likelihood logs are:

$$l(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \log\left(\int f(y_i|\,\boldsymbol{\delta}, \gamma)\, p(b_i, \gamma)\, \boldsymbol{\partial} b_i\right) \tag{10}$$

where $p(b_i, \gamma)$ is the density of the random effect. Based on the statement of Breslow and Clayton (1993) in Groll, A., and Tutz Gerhard[6], the approximation or approach of the derivative is:

$$l^{\mathrm{app}}(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \log\left(\int f(y_i|\,\boldsymbol{\delta}, \gamma)\right) - \frac{1}{2} b^T Q(\varrho)^{-1} b \tag{11}$$

part penalty-$\frac{1}{2} b^T Q(\varrho)^{-1} b$ s an approximation based on the Laplace method.

The log-likelihood function (10) is extended to include the penalty portion $\lambda \sum_{i=1}^{p} |\beta_i|$. penalized log-likelihood approximation becomes:

$$l^{\mathrm{app}}(\boldsymbol{\delta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \log\left(\int f(y_i|\,\boldsymbol{\delta}, \gamma)\right) - \lambda \sum_{i=1}^{p} |\beta_i| \tag{12}$$

for a certain $\hat{\boldsymbol{\gamma}}$, in order to optimize the function, the equation is reduced to:

$$\hat{\boldsymbol{\delta}} = \mathrm{argmax}_{\delta}\, l^{\mathrm{app}}(\boldsymbol{\delta}, \hat{\boldsymbol{\gamma}}) = \mathrm{argmax}_{\delta}\left[l^{\mathrm{app}}(\boldsymbol{\delta}, \hat{\boldsymbol{\gamma}}) - \lambda \sum_{i=1}^{p} |\beta_i|\right] \tag{13}$$

The penalty used in equations (12) and (13) is a partially penalized approximation when considering all vector parameters $\delta^T = (\beta^T, b^T)$.

The linear mixed model with Lasso variable selection in this study uses R data processing software with the glmmLasso program package. The algorithm on glmmLasso seeks to maximize the log-likelihood penalty $l^{\mathrm{app}}(\boldsymbol{\delta}, \boldsymbol{\gamma})$ n equation (12) by utilizing iterations of the Fisher scoring procedure, where the initial value $\hat{\beta}^0$, $\hat{b}^0$, $\hat{Q}^0$ the first iteration according to a simple global intercept model with random effects, namely $g(\mu_{it}) = \beta_0 + z_{ij}^T b_i$ which is obtained easily using the function in the R data processing software, namely glmmPQL. The estimation of the covariance matrix $Q_b^{(l)}$ obtained after obtaining the results of the calculation of $\hat{\delta}^{(l)}$ from the results of the lth iteration, as an estimate with the restricted maximum likelihood (REML) approximation for Q_b.

3. Results

3.1. Data exploration

Globally, stunting affected about 144,0 million children or 21.3 percent in 2019 [17]. Although prevalences of moderate-and-severe stunting declined in developing countries [18], Riskesdas 2018 reports that the prevalence of stunting in children under five in 2018 in Indonesia was around 30.8 [19]. The need to accelerate stunting reduction in Indonesia is explicitly stated since the prevalence of stunting as SDGs global indicator targets reducing to 19 percent in 2024 and 10 percent in 2030.

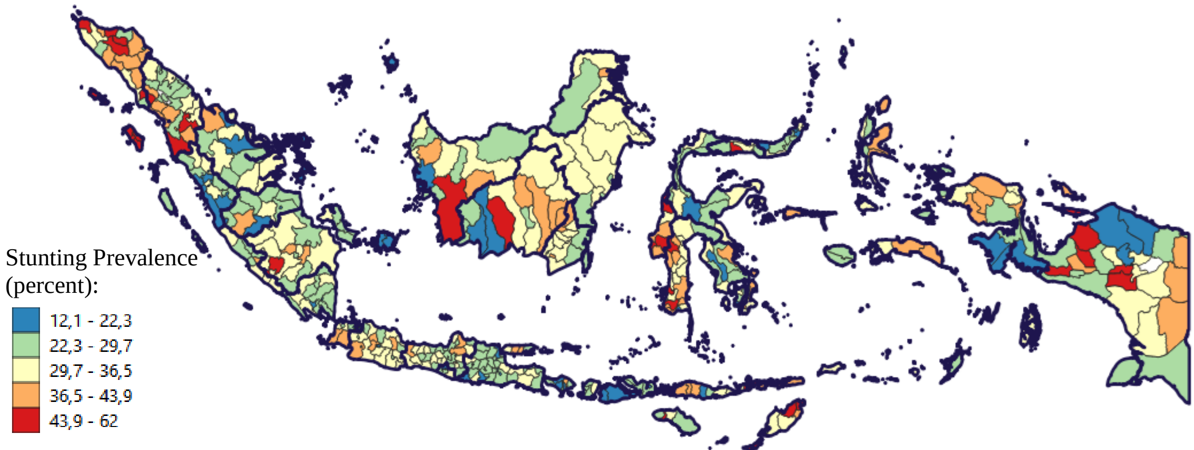

Figure 1. Stunting prevalence by district/cities in Indonesia, 2018

The prevalence of stunting at the estimated district/city level fluctuates with a range of around 49.97 percent (the complete distribution can be seen in figure 1). It shown in Figure.1 shows that the prevalence of stunting in a province is relatively uniform. It indicates that the area can be the basis for grouping and as a random effect model.

Figure 2 represents the estimated stunting level plots for districts/cities with a boxplot diagram sorted by the average of each province. Information on the stunting prevalence range in each province and outliers can be obtained from the boxplot.

The correlation between the research variables (variables explanation in table 1.) can be seen in figure 3. It informs the estimation of the direction and magnitude of the relationship between the stunting variable and the stunting intervention variable in the model built. The large number of predictor variables for multidimensional stunting intervention factors reviewed in this study and the presence of several predictor variables with strong correlations indicates the need for modeling with Lasso's variable selection.

3.2. Modeling analysis

The linear mixed model in this study was following the characteristics of the data, namely using the province as a random effect. It is also strengthened by the significance of the chi-square test's probability value for the province as a random effect. In other words, the prevalence of stunting in districts/cities within a province is relatively homogeneous, and the prevalence of stunting in districts/cities between provinces is quite heterogeneous. The difference in the values of the variance parameter and the standard error parameter of the provincial random variable group with the classical linear mixed-method and the linear mixed method of Lasso variable selection occurs due to the different approaches to measuring the parameters of the random variable.

4. Discussion

Comparison of the results of classical linear mixed modeling and linear mixed modeling with Lasso variable selection (as shown in table 2) shows the difference in the significance of the predictor variables. Using the lme4 program package in the R data processing program, the classical linear mixed

IOP Conf. Series: Earth and Environmental Science **948** (2021) 012067 doi:10.1088/1755-1315/948/1/012067

model presents all regression coefficients for all predictor variables (both significant and insignificant). The linear mixed model with Lasso variable selection with the glmmLasso program package on the R data processing program presents the regression coefficients only for the influential predictor variables. The classical linear mixed model shows fewer predictor variables that significantly affect the response variable than the linear mixed model using the Lasso variable selection method. It is possible because many predictor variables that tend to be multi collinear in the model will be reduced in the classical linear mixed model. The predictor variables, namely complete immunization, diversity of food intake, access to safe water, maternal education, and ownership of the National Health Insurance for children under five, significantly affect the decrease in stunting prevalence. The only predictor variable of low birth weight was significant in increasing the prevalence of stunting.

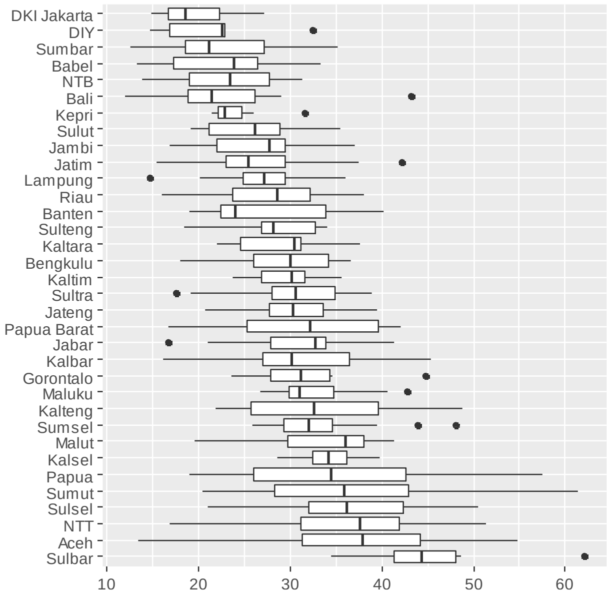

Figure 2. Boxplot of stunting prevalence at regency/city estimation by province.

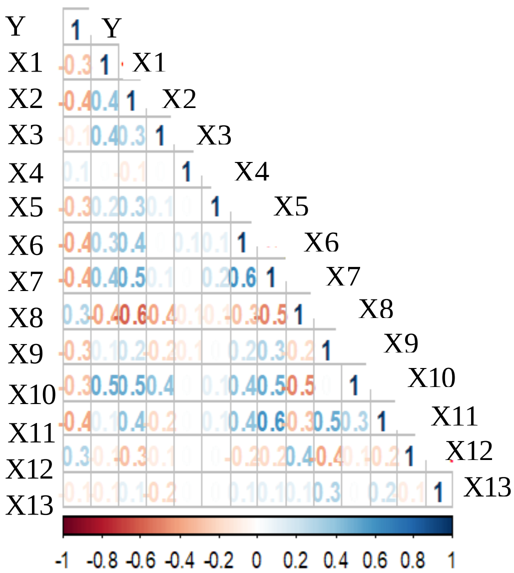

Figure 3. Correlation of research variables.

Simultaneous selection of variables and shrinkage, also known as Lasso, can be used for setting high-dimensional models with a large number of potentially influential predictor variables, making it more flexible to see the effect of many variables in the model. It can be seen in table 2 where the Lasso linear mixed model shows more significant predictor variables. The predictor variables of complete immunization, ease of access to health facilities, diversity of food intake, improve water, food expenditure per capita, children's participation in early childhood education, maternal education, and National Health Insurance for toddlers significantly affect the reduction stunting prevalence. The predictor variables, namely low birth weight, households with social protection cards, and the percentage of poor people, increase the prevalence of stunting substantially.

The linear mixed model with Lasso variable selection seems to have better performance in this study. The mean squared error generated by the linear mixed model with the Lasso variable selection is relatively smaller than the classical linear mixed model. A linear mixed model with Lasso variable selection in this study also provides the advantage of the flexibility of analysis, where more information on predictor variables, because this method allows predictors that have the potential to affect stunting prevalence are also included in the model.

Table 2. Comparison of research variable modeling results with province random effect according to linear mixed model and linear mixed model with lasso variable selection

Variables Name		Model Specification			
		Linear Mixed Model		Linear Mixed Model with Lasso	
		Estimation	*P Value*	Estimation	*P Value*
(Intercept)		52,185	1,4 e-30	51,759	< 2,2 e-16
Complete immunization	(X1)	-0,032	0,06	-0,024	< 2,2 e-16
Ease of access to health facilities	(X2)	-0,044	0,12	-0,039	8,3 e-16
Long-term contraceptive methods use	(X3)	0,021	0,56	0,000	
Low birth weight	(X4)	0,078	0,09	0,070	< 2,2 e-16
Diversity of food intake	(X5)	-0,087	0,00	-0,085	< 2,2 e-16
Improve water	(X6)	-0,095	3,3 e-0	-0,097	< 2,2 e-16
Improve sanitation	(X7)	-0,007	0,72	0,000	
Percentage of poor population	(X8)	0,113	0,10	0,150	< 2,2 e-16
Food expenditure per capita	(X9)	-0,028	0,37	-0,012	0,025
Child participation in early childhood education	(X10)	-0,050	0,19	-0,063	< 2,2 e-16
Mother's education	(X11)	-0,062	0,04	-0,067	< 2,2 e-16
Ownership of social protection card	(X12)	0,042	0,30	0,036	1,2 e-07
Privilege of JKN for children under five	(X13)	-0,032	0,01	-0,041	< 2,2 e-16
Random Effect		Pr(>Chisq) : 3.133e-14			
Residual		variance: 37,76 std error: 6,145		variance: 34,1 std error: 5,866	
Mean Square Error		34.858		34.348	

5. Conclusion

Lasso variable selection has the best performance applied to the data due to comparing the mean square error, which is relatively more minor. Another advantage of using Lasso is more flexible analysis, where we get more information on predictor variables that have or can influence stunting prevalence. Thus, the results of the study stated that the stunting intervention factors, namely complete immunization, easy access to health facilities, diversity of food intake, access to decent water, per capita food expenditure, children's participation in early childhood education, maternal education, and ownership of National Health Insurance for children under five years, were significant effect the reduction in the prevalence of stunting. Furthermore, the variables of low birth weight, households with social protection cards, and the percentage of poor people significantly increase the prevalence of stunting.

5.1. Suggestion

The research concludes that a linear mixed model with Lasso variable selection builds a good model to analyze the significance and influence of stunting intervention factors (predictor variables) on stunting prevalence (response variable). However, this research can still be developed by exploring specific variables related to children under five also find an alternative for more precise statistical methods.
Based on the influence of stunting intervention factors, social protection programs correlate with the percentage of the poor, which means that the coverage of this program is appropriate, namely targeting the poor. As poverty contributes positively to the increase in stunting prevalence, the results of this study recommend integrating social protection programs to pay attention to the mechanism of nutritional sustainability of children under five. The research result also suggests the need to improve the ownership

of national insurance cards for the low-income family in the household, especially with the presence of children under five years.

Acknowledgments
This paper prepared by Din Nurika Agustina under supervision of Bapak Dr. Bagus Sartono and Bapak Professor Khairil Anwar Notodiputro. We thank the Government of Indonesia for funding education through Statistics Indonesia so that the first author participated in a doctoral program at the Department of Statistics, Bogor Agricultural University. We express our appreciation to Statistics Indonesia and the Indonesian Ministry of Health, who have opened access to data to carry out this research.

References
[1] WHO 2015 Stunting in a Nutshell. WHO Departmental news [online].
[2] Berkman D S *et al*. 2002 *The Lancet* **359** 564
[3] World Health Organization (WHO) 2016 *General Assembly proclaims the Decade of Action on Nutrition* (Geneva: WHO Departmental news) [online].
[4] Keats E C *et al*. 2021 *The Lancet Child & Adolescent Health* **5** 367
[5] Agresti A 2015 *Foundations of Linear and Generalized Linear Models* (New Jersey: John Wiley & Sons, Inc).
[6] Badan Pusat Statistik (BPS) 2017 Pedoman Kepala BPS Provinsi, Kepala Bidang Statistik Sosial, dan Kepala BPS Kab/Kota. (Jakarta: BPS).
[7] United Nations Children's Fund (UNICEF) 2021 *Fed to Fail? The Crisis of Children's Diets in Early Life. 2021 Child Nutrition Report* (New York: UNICEF)
[8] Verma P and Prasad J B 2021 *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **15** 1
[9] Kurnia I D, Krisnana I, and Rahmadhani T D 2021 *Enfermería Clínica* **31** S33.
[10] Cameron L *et al*. 2021 *Economics & Human Biology* **40** 100944
[11] Badan Pusat Statistik 2020 *Laporan Indeks Khusus Penanganan Stunting 2018-2019* (Jakarta: BPS).
[12] Groll A and Tutz G 2011 *Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation. Technical Report Number 108* (Munich: Department of Statistics, University of Munich)
[13] Bates D 2003 *Computational Methods for Mixed Models*: *Technical Report No. 1076* (Madison: Department of Statistics, University of Wisconsin)
[14] Bates D, Martin M, Benjamin M B, and Steven W 2015 *Journal of Statistical Software* 67 1
[15] Bates D 2010 *lme4: Mixed-effects Modeling with R*. (Madison: Springer)
[16] Demidenko E 2013 *Mixed Models Theory and Applications with R, Second Edition*. (New Jersey: John Wiley & Sons, Inc)
[17] United Nations Children's Fund (UNICEF), World Health Organization, International Bank for Reconstruction and Development/The World Bank 2020 *Levels and trends in child malnutrition: Key Findings of the 2020 Edition of the Joint Child Malnutrition Estimates* (Geneva: World Health Organization)
[18] Stevens G A *et al*. 2012 *The Lancet* **380** 824
[19] Kementerian Kesehatan 2019 *Laporan Riskesdas 2018 (Provinsi)* (Jakarta: Kementerian Kesehatan).