5th ICRIEMS Proceedings
Published by Faculty Of Mathematics And Natural Sciences
Yogyakarta State University, ISBN 978-602-74529-3-0

Modeling The Number of Infant Mortality in East Java Using Hierarchical Bayesian Approach

Prizka Rismawati Arum[1,a)], Nur Iriawan[2,b)], and Muhammad Mashuri[3,c)]

[1,2,3]*Department of Statistics, Faculty of Mathematics, Computation, and Data Science,*
Institut Teknologi Sepuluh Nopember (ITS),
Surabaya, Indonesia

[a)]Corresponding author: prizka.rismawati@gmail.com
[b)]nur_i@statistika.its.ac.id

Abstract. Improving good health services for children are very important to be realized in Indonesia. This case encourages a millennium declaration known as a Millennium Development Goals (MDGs). The quality of children's healthcan be measured through infant mortality rate. This rate in East Java Province during 2014 was still above the MDGs target. The number of infant deaths at each sub-district in East Java Province has a hierarchical structure. A two-level hierarchical model was developed in this paper to analyze this hierarchical phenomenonby involving two or more levels of relationships between variables influencing the infant mortality in each sub-district and their parameters to the district characteristics. Due to the complexity of this model, a Bayesian hierarchical modeling couple with Markov Chain Monte Carlo (MCMC) approach to this problem is employed to provide a better solution. The results show that the two-level hierarchical model is much better than when all data in each district is modeled as only in one level modeling. The Cross-level interactions between sub-district and district characteristics, on the other hand, can also be found for further exploration in governance management.

INTRODUCTION

Improving good health services for children are very important to be realized in Indonesia. This needs to be done because children are the next generation of the nation that will greatly determine the quality of Human Resources in the future. The importance of good health for children has been recognized by the whole world. This case encourages a millennium declaration known for its Millennium Development Goals (MDGs). The quality of health in children can be measured using indicators that have been determined by the Ministry of Health. An indicator of the final result that can be used, one of them is through infant mortality rate.Infant mortality is a death that occurs between the time after the baby is born until the baby is not exactly one year olds. Many factors are affected by infant death [1].

Based on the results of the 2011-2013 National Social Economic Survey of East Java, Number of Infant Mortality in East Java Province in 2013 amounted to 27.23 per 1,000 live births and in 2014 amounted to 26.66 per 1,000 live births or as many as 5,219 cases of infant mortality. Based on the existing data, Infant Mortality Rate in East Java Province in 2014 has decreased but still above the target set of MDGs.Previous research the number of infant mortality in East Java has been done by Kuntoro and Kurniati using Geographically Weighted Poisson Regression (GWPR) method. Where in the research, the method produces a global model with the spatial effect of the data [2]. The resulting model illustrates the influence of predictor variables at the district/city level, whereas in the case of infant mortality in 2014 each sub-district has their characteristics that are also influential. So, it is necessary to find a model that analyzes the influence of predictor variables at the sub-district.

Datas on the number of infant deaths in East Java Province have a hierarchical structure. Groupings in hierarchical data are built on the principle of equality in members in one group so that members of one group have similar properties [3]. In a hierarchically structured data, the data can be classified into different levels/groups. For example, in this study, the first level is the sub-districtobservations and the second level is observed in the district/city. The issues to be addressed firstly are to look at the characteristics of infant mortality rate at the sub-district level and then to model the variables that affect infant mortality at the district/city level. Therefore, the hierarchical data structure cannot be simply ignored in the analysis process, so the proper solution to this problem is the hierarchical model.Basically, the hierarchical model is built by two sub-models, namely micro model and macro model. In the two-level hierarchy model, the micro model is a regression model that expresses the relationship between the response variable to be observed and the predictor at the first level. While the macro model expresses the relationship between the regression coefficient of the micro model and the predictor variable at the second level[4]. Hierarchical model has several advantages. First,this model can be used to analyze at several different levels simultaneously in one statistical analysis, and second, this model takes into account the variance at each level of the response variance [5]. Comparisons of classical models and hierarchical models on data that are simulated in a hierarchical structureprove that the hierarchical model is better than the classical model expressed by itssmaller mean square error [6].

In the hierarchical model, it will involve many variables causing the model to be complex and will be difficult to be solved by the classical estimation method. The use of modeling through the Bayesian approach to this problem will provide a better solution. The Bayesian method is very flexible and easy to estimate the parameters of complex hierarchical models[7]. The success in modeling with the Bayesian approach is the information distribution pattern of the observational data. In its decision-making, the Bayesian statisticsworks based on the new information from observed data and prior knowledge [8]. Bayesian observational data are said to be derived from a probability distribution having unknown parameters with certainty. Therefore, it is necessary to specify a distribution of each parameter, called the prior distribution.

Based on the background that has been described, this research proposedto model the number of cases of infant mortality in district/city in East Java Province that exactly a complex problem which needs to be solved by using a hierarchical model with a Bayesian approach competing with the uni-level Binomial negative regression approach. The formation of a two-level hierarchy model was performed using the characteristics of sub-districts as a predictor at the first level and the district/city characteristics as a predictor at the second level. While the competing model assumes that all of sub-district are directly modeled as a one level modeling. After applying these two modeling to the data on the number of infant deaths in East Java Province, seeking the best model is based on the smallest Deviance Information Criterion (DIC) value.

DATA AND METHODS

Data

Data were taken from published data of Health Office and Statistics Indonesia on each district or city in East Java, Indonesia. There are 619 sub-districts as the first level observation units and 29 district/cityas the second levelobservation units.The response variable (**Y**) is the number of infant mortality ineach sub-district in East Java Province. While the predictor variables (**X** for the first level predictors and **W** for the second level predictors) used in the modeling will be explained in **TABLE 1**:

TABLE 1. Variables And Data Scales

Variables	Description	Scales
X_1	Percentage of mother who did first check up	Ratio
X_2	Percentage of mother who did complete check up	Ratio
X_3	Percentage of birth assisted by health personnel	Ratio
X_4	Percentage mother who consumption of Fe1 tablets	Ratio
X_5	Percentage mother who consumption of Fe3 tablets	Ratio
X_6	Sum of low birth weight babies	Ratio
X_7	Percentage of baby born with complications	Ratio
X_8	Percentage of neonatus check up	Ratio
X_9	Percentage of baby who has exclusive breast feeding	Ratio
X_{10}	Percentage of baby who has complete immunization	Ratio
W_1	GDP of district/city	Ratio

W_2	Percentage of poverty	Ratio
W_3	Percentage of healthy lifestyle	Ratio
W_4	Percentage of healthy home	Ratio

The following stages are the steps performed in the data analysis of the research.

1. Pre-processing data.
2. Data exploration for all interested variables.
   a) Data exploration of the number of infant mortality by district/city in East Java Province.
   b) Performing Goodness of fit data on the infant mortality by sub-districts using a Chi-Square test to identify the distribution of the response data.
   c) Data explorationof the predictor variables (**X**and **W**) by sub-districtand district/city in East Java Province.
3. Estimating the first model: To model the number of infant mortality in all of sub-district in East Java Province as a one level model without including the second level factors, with steps:
   a) Determine the prior distribution of parameters of model to be estimated
   b) Create a syntax program codes for a one-level model according to a predetermined priordistribution
   c) Estimate the parameters of the model using Monte Carlo Markov Chain (MCMC) and Gibbs Sampling by running the syntax on WinBUGS
   d) Evaluate the model using a credible interval to identify significant predictor variables.
4. Estimating the second model: To model the number of infant mortality in the sub-district in East Java Province using a two-level hierarchy structure with steps:
   a) Determine the prior distribution of parameters in each level of model to be estimated.
   b) Create a syntax program codes for a two-level hierarchy model according to a predetermined priordistribution.
   c) Estimate the parameters of the two-level hierarchical model using MCMC and Gibbs Sampling by running the syntax on WinBUGS.
   d) Evaluate the model using a credible interval to identify significant predictor variables.
5. Selecting the best model and interpretation
   a) Compare the DIC value of each model, selects the best model based on the smallest DIC value.
   e) Make the best model interpretation and get the conclusion.

Methods

Hierarchy Model

The hierarchical model is a model built on hierarchically structured data with one measured response variable at the lowest level, and predictor variable at several levels. The levels of data structures can be infinite, but this study will only use two levels of the data hierarchy. There are $n_m$sub-district that come from 29district/city. For example, $Y_{ij}, Y_{2j}, \dots\ Yn_m j$arethe j-thgroup response variable with $n_m$number of observations. Whereas $X_{1j}, X_{2j}, \dots, X_{10j}$are the number of 10predictor variables at the first level forthe j-thgroup, and $W_1, W_2, \dots, W_4$are predictor variables at the second level. Formation of a two-level hierarchy model is as follows [9]:

1. Model at the first level.
   In normal distribution, the model equations at the first level for each group can be expressed as a multiple regression as in equation (1):
   $Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{lj}X_{lij} + e_{ij}$
   with $i = 1, 2, \dots, n_m$; $j = 1, 2, \dots, 29$; and$l = 1, 2, \dots, 10$, or this equation can be written in the matrix form as follows:
   $$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\beta}_j + \boldsymbol{e}_{ij} \quad (1)$$
2. Model atthe second level.
   The first level model will produce as many as 29 regression models, with regression coefficient valuesvarying between groups. Variations will be explained by regressing each coefficient$\beta_{lj}$with a predictor at the second level. This model is known as a macro model. This model can be expressed as:
   $\beta_{lj} = \gamma_{0j} + \gamma_{1j}W_{1j} + \gamma_{2j}W_{2j} + \cdots + \gamma_{pj}W_{pj} + u_{ij}$
   with$j = 1, 2, \dots, 29$and$p = 1, 2, \dots, 4$, or it can be written in the matrix form as follows:
   $$\boldsymbol{y}_j = \boldsymbol{W}\boldsymbol{\gamma}_j + \boldsymbol{u}_j \quad (2)$$

Equations (1) and (2) are two separate equations of this hierarchical model By combining these two equations can be obtained the single equation as follows [10]:

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{W}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{e}_j, \tag{3}$$

with, $\boldsymbol{X}_j\boldsymbol{W}_j\boldsymbol{\gamma}$ = fixed (deterministic) component in the hierarchical model,
$\boldsymbol{X}_j\boldsymbol{u}_j$ = random component in the hierarchical model.

Equation(1) will be different if data do not follow a normal distribution because they have link function. Based on the generalized linear model, then theirs must be added a link function based on its distribution [11]. Moreover, a chi-square test will be used to check the distribution of the data. The Chi-square test procedure includes determining the initial hypothesis that the observed data follow a particular distribution [12] and specifying statistical test used. In order to determine the test statistic, used equation (4):

$$\chi^2 = \sum_{i=1}^{n_m} \frac{(o_i - e_i)^2}{e_i} \tag{4}$$

Where: χ^2: Chi-Square value; o_i: observation value
n_m: number of observation in each district; e_i: expectation value ofdata observation

The last step is comparing the chi-square statistic with the critical value from the table. The null hypothesis is rejected if $\chi^2 > \chi^2{}_{(n-1)}$. Where $\chi^2{}_{(n-1)}$is the value of the Chi-Square table.

Bayesian Estimation Approach

The Bayesian method looks at each unknown parameter of a model as a random variable having a distribution, called the prior distribution. Based on the combination between the prior distribution andthe observational pattern of data, furthermore, the posterior distribution to obtain Bayesian estimatorcan be determined. Conceptually, the Bayesian method is developed based on the Bayes theorem, by multiplying the prior probability distribution by the likelihood function, and then dividing by the normalizing constant into the posterior probability distribution. Box and Tiao [13] said that if Y is a random variable following a certain distribution pattern with a density function (PDF)$f(\boldsymbol{y}|\boldsymbol{\theta})$, with $\boldsymbol{\theta}$ is the parameter vector with size d or $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)^T$ and$\boldsymbol{y} = (y_1, y_2, \dots y_n)^T$ are sample vectors with size$n$ which are identical and independent distributions, then the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{y}$ can be written in the following equation (5):

$$f(\boldsymbol{y}, \boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) = f(\boldsymbol{\theta}|\boldsymbol{y})f(\boldsymbol{y}) \tag{5}$$

Based on the Bayes theorem, the posterior distribution of θ, $f(\theta \mid \boldsymbol{y})$, can be derived from the equation (5),

$$f(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\boldsymbol{y})} \tag{6}$$

where$f(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function of data that contains information data, and $f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$. While$(\boldsymbol{\theta})$ is a prior distribution function of parameters $\boldsymbol{\theta}$ and $(\mathbf{y})$ is a normalized constant, which can be written as:

$$f(\boldsymbol{y}) = \begin{cases} \int\int \dots \int f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta})d\theta_1 \dots d\theta_d \text{ , if } \boldsymbol{\theta} \text{ is continuous} \\ \sum f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}), \text{ if } \boldsymbol{\theta} \text{ is discrete} \end{cases}$$

Thus equation (6) can be expressed in terms of:

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{7}$$

or also commonly written Posterior $\propto$ LikelihoodxPrior

Prior distribution determination is very important step in the Bayesian method because it affects the posterior distribution. This prior distribution could be determined based on the classification of the types. Those are, conjugate prior or non-conjugate prior, proper prior or improper prior,informative prior or non-informative prior [13],and pseudo prior [14].

In complex models such as hierarchical models, there will be many parameters. For example, using two parameters β and γ,it requires a high degree of marginalization process from the posterior distribution that has to be faced and it is difficult analytically. One solution to solve this problem is using a numerical approach, namely Monte Carlo Markov Chain (MCMC)[14]. The MCMC method requires proper sampling algorithm to get samples from a distribution. The algorithm which is often used as a generator of the random variable in the MCMC is Gibbs Sampling. Gibbs sampling can be defined as a simulation technique to generate random variables from a particular distribution function without having to calculate its density function [15]. Each posterior parameter distribution is estimated from its full-conditional posterior distribution which is submerged in the MCMC-Gibbs sampler algorithm.

The interval estimate of the parameters used in the Bayesian method is known as the credible interval. In addition, a credible interval can be used for confidence interval generation. The credible interval can be used as a significance test for θ with the null hypothesis is $H_0: \theta = 0$ and the alternative hypothesis is $H_1: \theta \neq 0$. The decision would reject H_0 if the credible interval does not contain a zero value.

Computation Estimation Using WinBUGS

WinBUGS is a development of Bayesian Inference Using Gibbs Sampling (BUGS) which is designed based on windows and is an open source software. The WinBUGS is a programming language-based software used to generate a random sample of the Bayesian model's posterior parameter distribution [16]. The WinBUGS gives a simplification for estimating model parameters numerically. It is a developed software that can generate a posterior distribution of model parameters using MCMC. The Deviance Information Criterion (DIC) for each model can be also obtained using the WinBUGS. A models with smaller DIC ~~values~~ indicates a better model for explaining the variation of response variables.

RESULTS AND DISCUSSION

Descriptive Statistics and Response Pattern

There is a difference in the number of infant deaths between district/city. Sidoarjo has the highest infant mortality rate compared to other districts/cities. In addition, the coefficient of variation also shows a high value. This situation indicates a considerable difference between the infant mortality in each sub-district inSidoarjo. Meanwhile, the lowest infant mortality occurred in Lamongan with the lowest coefficient of variation also compared to other districts/cities. The low coefficient of variation can indicate that the number of infant mortality in Lamongan is almost same and evenly distributed in each sub-district.

The goodness of fit test using Chi-Square test to the number of infant deaths between sub-districts shows that the appropriate distributionpattern of the response data is negative binomial distribution pattern.

Estimation One-Level and Model Two-Level Hierarchical Negative-Binomial Regression

It is known that the number of infant deaths in the East Java districts follows the negative binomial distribution. Using two-level hierarchical model, the parameters in the first level of infant mortality are r and $\boldsymbol{\beta}$ which represent the distribution parameters of the negative binomial distribution and also the micro model regression parameters. While parameters in the second level (hyperparameter) are τ and $\boldsymbol{\gamma}$ which represent the precision parameters of the prior distribution and the regression parameters of the macro model. The prior and hyperprior distribution used in this study are a combination of conjugate prior, pseudoprior, and informative prior~~,~~ where:

a) *prior*forβ_{lj}, denoted as$p(\beta_{lj})$, is a normal distribution which isa *pseudo prior*.
b) *prior*for r, denoted as$p(r)$, is a Gamma distribution which isan *informative prior*.
c) *hyperprior*for γ_{pl}, denoted as$p(\gamma_{pl})$, is a normal distribution which is*a pseudo prior* and an *informative prior*.
d) *hyperprior*for $\tau_{[\beta]l}$,denoted as$p(\tau_{[\beta]l})$, is a Gamma distribution which is a *conjugate prior* to parameter $\sigma^2_{[\beta]l}$as variance of the estimated β_{lj}.

One-level Negative-BinomialRegression Model

The one-level negative binomial regression analysis was done using WinBUGS. The iterations were doneby using the scenarios of 20,000 times with thin 20 and burn-in 401 iterations, so the samples used to estimate the parameter characteristics were 19,600 samples. Modeling was done by including the characteristics of the sub-districtand neglecting the information belong to the district/city.The hypothesis testing to the parameters of this one-level negative-binomial were done by employing the credible intervals. If the credible interval of each estimated ~~parameters~~contains a zero value, then it is concluded that the estimated parameter is not significant.Some of the estimatedparameters of negative-binomial regression modelare presented in **TABLE 2**.

TABLE 2. Estimation Parameter of Negative Binomial Regression Model

Node	Mean	Sd	MC error	2.50%	Median	97.50%
$\beta_{0[1]}$	-189.40	10030.00	600.40	-22580.00	-60.71	20130.00
$\beta_{0[2]}$	-1184.00	11230.00	749.30	-31270.00	-29.04	15580.00
...	...	...	...	...	...	...
$\beta_{0[29]}$	637.80	10310.00	630.90	-16550.00	-376.50	28150.00
...	...	...	...	...	...	...
$\beta_{1[1]}$	-219.30	791.30	64.95	-2408.00	-141.60	976.80
$\beta_{1[2]}$	-300.60	840.80	69.83	-1626.00	-431.20	1713.00
...	...	...	...	...	...	...
$\beta_{10[29]}$	1530000.00	2330000.00	167400.00	5191.00	543600.00	8760000.00

Two-Level Hierarchy Negative-BinomialRegression Model

Similar to the one level model, the two-level hierarchical model based on the negative binomial was also done using WinBUGS.The iteration used is 20,000 times with thin 5 iterations and the burn-in is 801 iterations, so the samples used to estimate the parameter characteristics is 19,200 samples. Modeling is done by including all predictors that have been determined, ten predictors of sub-district characteristics and four predictors of district/city characteristics.

It is known that three parameters, i.e. β_7, β_9, and β_{10}, are the micro regression model coefficients that are not significant in almost all district models. Other some parameters also appear to be insignificant in some micro models. By assuming that all considered covariates listed in **TABLE 3**, are important for monitoring the maternal and child health, then all of the related predictors are still considered to be included in the model.Some of the estimatedparameters of the first level regression model are presented in **TABLE 4**.

TABLE 3. EstimationParameter of First Level Regression Model

Node	Mean	Sd	MC error	2.50%	Median	97.50%
$\beta_{0[1]}$	11.1500	2.1970	0.1707	7.402	10.9700	15.8500
$\beta_{0[2]}$	7.5630	2.3310	0.1870	3.449	7.2670	12.1300
…	...	...	...	...	...	...
$\beta_{0[29]}$	4.5270	5.1540	0.4342	0.3525	3.2670	24.8700
…	...	...	...	...	...	...
$\beta_{1[1]}$	-0.2487	0.2555	0.0217	-1.073	-0.1665	0.00094
$\beta_{1[2]}$	0.3365	0.1634	0.0139	0.1165	0.2926	0.7351
…	...	...	...	...	...	...
$\beta_{10[29]}$	-0.0109	0.0255	0.0020	-0.0852	-0.0060	0.0244

Some of the estimatedparameters of the second level regression modelare presented in **TABLE 4**. TABLE 4 shows that not all characteristics of the district/cityare significantly influencing the number of infant mortality through the variability of parameters of thefirst level model. It can be seen that three parameters, i.e. γ_1, γ_2, and γ_4 which are related to GDP, poverty and healthy house variables, are significantly influence the number of infant deaths.

TABLE 4. EstimationParameter of Second Level Regression Model

Node	Mean	Sd	MC error	2.50%	Median	97.50%
$\gamma_{0[1]}$	-23.7500	0.2516	0.0047	-24.2400	-23.7500	-23.2600
$\gamma_{0[2]}$	0.2342	0.3506	0.0071	-0.4608	0.2351	0.9243
…	...	...	...	...	...	...
$\gamma_{0[11]}$	0.0269	0.0939	0.0047	-0.1101	0.0191	0.2010
$\gamma_{1[1]}$	0.0459	0.0173	0.0014	0.0144	0.0428	0.0887
$\gamma_{1[2]}$	-6.00E-04	1.29E-03	4.36E-05	-3.35E-03	-5.15E-04	0.0017
…	...	...	...	...	...	...
$\gamma_{4[11]}$	4.46E-04	1.08E-03	7.16E-05	-7.72E-04	2.78E-04	0.0032

The first model only using predictors at the first level and the second model using all predefined predictors. It is known that the model with DIC value of the first alternative model is 3498,170 and the DIC value of second alternative model is 27883,400. Considering the DIC value, it can be concluded that the first alternative model is better than the second alternative model.

Influence of Sub-district and DistrictCharacteristics on Infant Mortality in East Java

Based on the results presented in TABLE 2 and TABLE 3, the micro model produces 29 regression models that elaborate the effect of sub-district characteristics to the infant mortality in each sub-district. As an example, the micro model for Kediri district can be written as an equation (7).

$$Y \sim \text{Binomial negative } (p, r) \tag{7}$$

where $p = \frac{r}{r+\lambda}$, $r \sim$ gamma (a, b), and

$$Log\ (\lambda) = -0.8612\, X_{1.6} + 0.1453 X_{2.6} + 0.5064 X_{4.6} - 0.2984 X_{5.6} + 2{,}3830\, X_{7.6} - 0.2724 X_{8.6} - 0.0372 X_{9.6}$$

The micro model for each district/city can be written in the same way as an equation (8) based on their significant coefficients. The equation of micro model shows that the chance of infant mortality in Kediri will reduce, when the percentage of mother who does pregnancy check-up is small. The larger percentage check-up lead to case that the mother has problem with her pregnancy. The positive effect is shown by a complete pregnancy check-up; the chances of infant mortality will be smaller when the percentage of mothers who do the pregnancy checkup is larger. A positive influence is also indicated by the consumption of pregnant women to Fe1 tablets, the more percentage of mothers who consume~~d~~ Fe1, the smaller the chances of the baby die. Conversely, the smaller the percentage of check up after birth and the less breastfed infants, the greater chances of the baby die.

Unlike the micro model, the macro model has a broader interpretation to illustrate the factors that influence the number of infant mortality. Parameters in the macro model can explain the influence of predictor of death characteristics or the characteristics of district and even the interaction of both.

TABLE 5. Coefficient Regression of Two Level HierarchyModel on Data Number of Infant Mortality

No	Paramete	Variables	***PosteriorMean***
1.	γ_{00}	Constanta	-23.7500
		District Characteristics	
2.	γ_{10}	GDP	0.0459
3.	γ_{20}	Poverty	0.7564
4.	γ_{40}	Healthy House	0.2383
		Interaction Between Sub-district Characteristics and District	
5.	γ_{17}	GDP*complications	-0.0045
6.	γ_{21}	Poverty * first check up	-0.0448
7.	γ_{24}	Poverty *Fe1	0.0359
8.	γ_{43}	Healthy House * assisted by health personnel	-0.0050
9.	γ_{47}	Healthy House * complications	-0.0193

Based on**TABLE 5**, it can be seen that all predictors of the characteristics of district/city have a positive effect to the number of infant mortality. The number of infant deaths in areas with a greater percentage of poverty tends to have a greater number of infant deaths as well. Poverty, in this case, represents the level of economic ability in the district/city. Thus it can be concluded that district/municipalities with poor economic conditions would lead to increase the number of infant mortality cases in that area. The district GDP and the percentage of healthy homes also have a positive effect on the number of infant deaths. Parameters in the macro model are the main components of forming a single equation model in the hierarchical model. Taking into account to all of the significant regression coefficients of macro models, the complete model for the number of infant deaths in East Java Province can be written as the following equation (8).

$$Log\ (\lambda) = -23.75 + 0{,}0459 W_1 + 0{,}7564 W_2 + 0{,}2383 W_4 - 0{,}0045 W_1 X_7 - 0{,}0448 W_2 X_1 + 0{,}0359 W_2 X_4 - 0{,}005 W_4 X_3 - 0.0193 W_4 X_7 \tag{8}$$

The equation(8) gives an idea of writing a single equation model in the hierarchical model. From the two-level hierarchical model, there are no significant variables at the sub-district level. While for district/city level of GDP variable, poverty, and healthy home give a significant positive influence on infant mortality. It shows that the higher the value of GDP, the smaller the chances of infant mortality. The interaction between GDP variable with complication, poverty variable, and prenatal check-up, and the interaction of healthy house variable and complication have a negative effect to infant mortality. If the district/city is getting poorer (the percentage of poverty is becoming greater), the consumption of Fe1 Tablets in the district will be smaller and the chances of infant mortality become greater. Unlike the case for the interaction percentage of healthy homes and birth percentage helped by health personnel, the smaller the percentage of healthy homes and the smaller the percentages the help of health personnel, the greater the chance of infant mortality.

CONCLUSION

The goodness of fit test to the distribution fitting by using the Chi-Square test shows that the appropriate distribution of the infant mortality data at eachdistrict in East Java is following the negative-binomial distribution. The results of the implementation of a two-level hierarchy model to the number of infant deaths data in East Java Province showsmaller DIC and is the best model compared with the one-level negative-binomial model Two advantages using the two-level hierarchical model in the Bayesian approach are detecting more significant variables in the regression model and describing a cross-level interaction between the characteristics of the sub-district and the characteristics of the district.

ACKNOWLEDGMENTS

The authors thank to Faculty of Mathematics, Computation, and Data Science, InstitutTeknologiSepuluhNopember, also to the parties who have helped the authors in completing this paper.

REFERENCES

1. C. Wiguna, “Causes and Methods for Dealing Baby Death in Bali”,http://ilmukesmas.com/penyebab-dan-cara-penanggulangan-kematian-bayi-di-bali/(2014)(accessed 5 September 2017).
2. Kuntoro dan S. C.Kurniati, “Modeling the Number of infant deaths In East Java WithGeographically Weighted Poisson Regression (GWPR)”, Universitas Airlangga, Surabaya, 2015.
3. H. Goldstein, “Multilevel Statistical Models”, London: Edward Arnold, 1995.
4. N.Iriawan, “Data-Driven Modeling and Analysis Volume 1”, Surabaya : ITSpress, 2012.
5. J. J. Hox, “Multilevel Analysis: Techniques and Applications, Quantitative Methodology Series”, New York: Routledge, 2010.
6. P.Ismartini, N. Iriawan, and B S S. Ulama,“Model Criticism for Log-Normal Hierarchical Bayesian Models on Household Expenditure in Indonesia”, in *IEEE* Computer Society. Published, DOI: 10.1109/ICSSBE.2012.6396521(2012).
7. J. K. Kruschke and W. Vanpaemel, “Bayesian Estimation in Hierarchical Models”, in *The Oxford Handbook of Computational and Mathematical Psychology*, 279(2015).
8. M. C. M. Wong, K. F. Lam, and E. C. M. Lo, “Bayesian analysis of clustered interval-cencored data”, in *Journal of Dental Research*, 84(9), (2005) pp. 817-821.
9. S. W. Raudenbush, and A. S. Bryk, “Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition (Vol. 1)”, London:Sage (2002).
10. J. De Leeuw, E. Meijer, and H. Goldstein, “Handbook of multilevel analysis”, New York: Springer, 2008.
11. W. W. Daniel, “Applied Nonparametric Statistics”, Jakarta : PT. Gramedia, 1989.
12. P. McCullagh and J. A. Nelder, “Generalized Linear Models”, Chapman &Hall/CRC. London, 1989.
13. G.E.P. Box and Tiao. “Bayesian Inference in Statistical Analysis”, Reading, MA, Addison-Wesley, 1973.
14. B. P.CarlinandS. Chib, “Bayesian Model Choice via Markov Chain Monte Carlo Methods”, in *Journal of the Royal Statistical Society. Series B (Methodological),* (1995) pp.473-484.
15. G. Casella, and E. I. George, "Explaining the Gibbs Sampler", in *The American Statistician*, Vol. 46, No. 3, (1992) pp. 167-174.
16. I. Ntzoufras, “Bayesian Modelling Using WinBUGS”,New York: John Willey and Sons, Inc, 2009.