Hierarchical Generalized Linear Model Approach For Estimating Of Working Population In Kepulauan Riau Province

A Muslim1,2, A Kurnia1, K Sadik1

1 Department Of Statistics, Faculty Of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor,16111, Indonesia
2 Statistics Of Kepulauan Riau, Tanjungpinang,29192 Indonesia

Email: anangk@apps.ipb.ac.id

Abstract. Hierarchical Generalized Linear Model (HGLM) extended Generalized Linier Mixed Model (GLMM) in which the distribution of random components are extended to conjugate of arbitrary distributions from the GLM family. HGLM modeling used to model fitting with exponential family random effects. The Extended Likelihood used in HGLM should be describe hierarchical data character. Nasional Labor Force Surveys, a regular survey by Statistics Indonesia, indicates a hierarchical character data. Application of HGLM with Gamma Normal Beta model on the Nasional Labor Force Surveys data, indicate that the proportion of working population in Kepulauan Riau Province in August 2015 are affected by the number of population in age group 25-49, age group 50 years and over, and the number of population with graduate or post graduate degree. The interaction between the number of population by age group and the highest educational attainment completed, together does not affect the percentage of working population. Model without interaction between age group and highest educational attainment provides a better model compared to model with interaction.

1. Introduction

Nowadays, random effects modeling is becoming more and more essential for understanding complex data in empirical sciences. The complexity in the “big data” contains grand challenges in terms of size, structure and interpretation, for which random effects models, a.k.a. linear mixed models, have great potentials to dissect the problems[6].

In statistical modeling and references required information about the data of the population or sample including the type of distribution of the data. In modeling, there are times when the Linear model has only one error normal distribution of these conditions there is some development that can be done, the first development is if the linear model with 2 or more error conditions. The second development of the linear model which is to Generalized Linear Model (GLM). Distribution of this model was developed from a normal distribution where one of parameter is exponential family distribution and the addition of a systemic effect on a monotonic transformation of the average called circuit function (link function). Combined both of these developments is called Generalized Linear Mixed Model (GLMM), where in the linear predictor of GLM allowed to have additional fixed effects, with one or more components are assumed to be normally distributed random. Although the normal
distribution is used to express the correlation between the random effects but other distribution can be used on random effects could to develop the model. Lee and Nelder developed GLMM be Hierarchical GLM (HGLM) with a random component is conjugate of certain distributions from the exponential family. Lee and Nelder use the combined likelihood as an approach to estimate parameters in the model, the combined likelihood also called extended likelihood.

HGLM is a development of GLMM where the distribution of the random component is the development of a family of distribution conjugate GLM / exponential family. HGLM include GLMM which have normal random components. The h-likelihood for inference results in an extension of likelihood inference, providing an efficient fitting algorithm for the various likelihood-ratio tests and systematic model-checking methods. The method leads to statistically reliable and efficient estimators similar to those obtained from marginal likelihood, while having the considerable advantage of not requiring the integrating out of random-effects. Their h-likelihood method is an extension of classical likelihood inference, and so needs neither the use of prior probabilities nor computationally intensive methods such as Monte-Carlo Markov’s chain (MCMC)[6]

This Paper presents a comparative study between linear model as a Frecuentist Model, Bayesian Regression and HGLM when used simulation data and a comparative study HGLM Gamma Linier Betha model when there are interaction and no interactions between random component when using National Labour Force Surveys in Kepulauan Riau Province.

2. Data
For simulated Model, will be simulate five clusters with 20 observations for each cluster. For an average of some models, and the variance of the random effects is $\mu = 0 \sigma^2 = 0.2$ Then we used the National Labor Force Survey data in Kepulauan Riau Province August 2015. The response variable is the percentage of the working population, while the explanatory variables are the age group that are divided into three categories: 15-24, 25-49 years old and 50 years and over. For education are divided into 6 groups: elementary school down, junior high school, General senior high school, vocational school, Academy and University (D4/S1/S2/S3) while the regency in the Riau Islands province is considered as replication where in Riau Islands province, there are 7 municpalities, namely Karimun, Bintan, Lingga, Anambas, Natuna, Batam and Tanjung Pinang.

![Figure 1. Hierarchical character of National Labor force Survey Data](image)

3. Methods
3.1. Hierarchical Generalized Linear Models (HGLM)
Based on the previous information, in statistical modeling and statistics reference need information about the data of the population or sample including the type of distribution of the data. Likelihood is the main concept between modeling and inference. An $y$ containing data observation obtained from a research and when will be used in modeling, the data must be obtained /generated by taking into
account the elements probability. Probability distribution of the response variable and random effects should be known in the application HGLM.

Lee and Nelder (1996) defined HGLM: Let \( y \) be the response and \( u \) be the (unobserved) random component. We consider the following hierarchical model or HGLM.

- The conditional (log-)likelihood for \( y \) given \( u \) has the GLM form

\[
l(\theta', \phi; y \mid u) = \{y\theta' - b(\theta')\}/a(\phi) + c(y, \phi)
\]

where \( \theta' \) denotes the canonical parameter and \( \phi \) is the dispersion parameter, \( \mu' \) for the conditional mean of \( y \) given \( u \), where \( \eta' = g(\mu') \) is the link function for the GLM describing the conditional distribution of \( y \) given \( u \). The linear predictor \( \eta' \) takes the form \( \eta' = \eta + \nu \) and \( \eta = X\beta \) as for a GLM and \( \nu = (u) \) for some strictly monotonic function of \( u \).

- The distribution of \( u \) is assumed appropriately.

The modelling of \( \eta' \), involves not only fixed effects modeling for \( \eta \) but also dispersion modeling for \( \nu \), which describes the overdispersion. In this framework, the mixed linear model is the normal-normla HGLM (with identity link), where the first element refers to distribution of \( u \) and the second to that for \( u(=v) \).

3.2. Hierarchical Likelihood

The \( h \)-likelihood, denoted by \( h \), is defined by

\[
h = l(\theta', \phi; y \mid u) + l(\alpha; \nu)
\]

where \( l(\alpha; \nu) \) is the logarithm of the density function for \( \nu \) with parameter \( \alpha \), and \( l(\theta', \phi; y \mid u) \) is that for \( y \mid u \). The random component \( \nu \) is the scale on which the random effect \( u \) occurs linearly in the linear predictor. We may derive the \( h \)-likelihood from density functions of \( u \) and \( y \mid u \) as well. \( l(\alpha; \nu) \) can be derived from the density function of \( u \) with differential element \( d\nu(u) \) and \( l(\theta', \phi; y \mid u) = l(\theta', \phi; y \mid u) \), the logarithm of the density function for \( y \mid u \), since \( \nu \) is the strictly monotonic function of \( u \).

The \( h \)-likelihood is the logarithm of the joint density function for \( \nu \) and \( y \). When both distributions are normal the \( h \)-likelihood is Henderson’s joint likelihood. When one or both distributions are non-normal, the \( h \)-likelihood is an obvious generalization of the joint likelihood. Clearly the \( h \)-likelihood is not an orthodox likelihood because the \( \nu \) are not observed. We call estimates derived from maximizing the \( h \)-likelihood maximum \( h \)-likelihood estimates (MHLEs); these are obtained by solving \( \frac{\partial h}{\partial \beta} = 0, \frac{\partial h}{\partial \nu} = 0 \).

From the definition of the \( h \)-likelihood, can easily to see that MHLEs for \( \beta \) given \( u \) are obtained by the GLM equations with \( \nu(u) \) as an offset. As with maximum likelihood (ML) estimates, the MHLEs for random effects are invariant with respect to the transformation of random components; for example, estimating equations \( \frac{\partial h}{\partial \nu} = 0 \) and \( \frac{\partial h}{\partial u} = 0 \) result in the same random effect estimate[3].

3.3. \( H \)-likelihood estimation and inference framework

Standard maximum likelihood estimation for models with random effects is based on the marginal likelihood as objective function. The parameters are estimated using a marginal likelihood procedure (MML) and standard errors are computed from the inverse of the negative hessian matrix of the marginal likelihood. In the marginal likelihood approach random effect \( \nu \) are integrated out and only fixed effects in the mean structure \( \beta \) and dispersion parameter \( \lambda \) are retained in the maximized function. For a mixed effects model, conditional likelihood of the \( j \)-th (\( j = 1, \ldots, n_t \)) repeated observation on the \( i \)-th subject (\( i = 1, \ldots, N \)), i.e., \( y_{ij} \), is given by \( f_{\beta, \lambda}(y_{ij} \mid \nu_i) \). The likelihood of the \( i \)-th random effect is denoted as \( f_\lambda(\nu_i) \). Note that \( \lambda \) contains dispersion parameters of the random
component \( \boldsymbol{v}_i \) as well as the parameters describing the residual dispersion (overdispersion) of the response \( y_{ij} \).

The marginal likelihood maximized in the MML procedure is given by

\[
L_M(\beta, \lambda|y) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f_{\beta, \lambda}(y_{ij}|v_i)f_d(v_i)dv_i
\]  

Maximizing \( L_M \) or equivalently the log-likelihood \( \ell_M = \log(L_M) \) yields consistent estimates of the fixed effects parameters. MLE is determined with a Newton-Raphson procedure so integrals need to be computed also for the first and second derivatives.

In Lee and Nelder (1996) another approach to estimating the parameters, use the joint likelihood \( L_E \) for the maximization, which is directly available from the definition of the model. The joint likelihood, called also extended likelihood or \( h \)-likelihood, is then maximized jointly with respect to \( \boldsymbol{v} \) and \( \boldsymbol{\beta} \) given dispersion parameters \( \lambda \). At the maximum, standard errors are obtained in the classical way, the extended likelihood is given by

\[
L_E(\beta, \lambda, \boldsymbol{v}|y, \boldsymbol{v}) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f_{\beta, \lambda}(y_{ij}|v_i)f_d(v_i)
\]  

The logarithm of (4) is called the extended log-likelihood by Lee et al. (2006) and it denoted its logarithm as \( h = \log[L_E(\beta, \lambda, \boldsymbol{v}|y, \boldsymbol{v})] \). This extended likelihood reflects the hierarchical character of the data [5].

3.4. Computing marginal MLEs using the \( h \)-likelihood approach

In some special cases, i.e., when the random effects are on the canonical scale, joint maximization of the extended log-likelihood \( h \) with respect to all parameters \( (\beta, \lambda, v_1, ..., v_N) \) is equivalent to maximizing the marginal likelihood with respect to \( \beta, \lambda \) and taking the empirical Bayes (EB) estimates for \( v_1, ..., v_N \). But, most often the two maximization procedures are not equivalent.

Noh and Lee (2007) using a Laplace approximation to the marginal likelihood (3) is called integral of the function \( k(x, y)\exp[-ng(x, y)] \) with respect to \( x \) can be approximated as follows:

\[
\int k(x, y)\exp[-ng(x, y)] \, dx = \left. \frac{\left( \frac{\partial^2 g(x, y)}{\partial x^2} \right)}{\frac{\partial^2 g(x, y)}{\partial x^2} \frac{\partial^2 g(x, y)}{\partial y^2}} \right|_{x=\hat{x}} \exp[-ng(x, y)]k(x, y) \bigg|_{x=\hat{x}}
\]  

\( \hat{x} \) the value of \( x \) that maximizes \(-g(x, y)\). That is called Laplace approximation of the above integral (at \( \hat{x} \)). Taking in expression (5) \( k(x, y) = 1; \exp[ng(x, y)] = \exp[-h(\theta, v)] \) where by \( \theta = (\beta, \lambda); \boldsymbol{v} \) representing the stacked vector of \( N \) random effects and finally \( n \frac{\partial^2 g(x, y)}{\partial x^2} \frac{\partial^2 g(x, y)}{\partial y^2} \) leads to

\[
L_M(\beta, \lambda|y) = \int \exp[h(\beta, \lambda, v)] \, dv \approx \left. \frac{\partial^2 h(\beta, \lambda, v)}{\partial v^2} \right|_{v=\hat{v}}^{-\frac{1}{2}} \exp[h(\beta, \lambda, v)] \bigg|_{v=\hat{v}}
\]  

With \( \hat{v} \) maximizing the extended likelihood for a given (starting) value of \( \beta \) and \( \lambda \), i.e., \( \hat{v}(\beta, \lambda) \).

Use the logarithm of the previous expression leads to the adjusted profile (log)-likelihood.

\[
p_v(h) = h(\beta, \lambda, v) \bigg|_{v=\hat{v}} - 0.5 \log \left( \frac{|p(h,v)|}{2\pi} \right)_{v=\hat{v}}
\]  

\( p_v(h) = h(\beta, \lambda, v) \bigg|_{v=\hat{v}} - 0.5 \log \left( \frac{|p(h,v)|}{2\pi} \right)_{v=\hat{v}} \)
Where \( D(h, v) = -\frac{\delta^2 h(\beta, \lambda, v)}{\delta v \delta \beta} \), The next step in the iterative procedure is to maximize the adjusted profile (log-) likelihood (7) with respect to \( \beta \). After obtaining \( \hat{\beta}_\lambda \) from maximization of (7) for a given dispersion component \( \lambda \), the estimation algorithm proceeds with estimation of \( \lambda \). The marginal distribution of the data \( y \) be \( f_{\beta, \lambda}(y) \) (marginalized over \( v \)), i.e., the LHS of expression (6) or its approximation, now seen as a probability density function (pdf) of the data. Conditional on the sufficient statistics for \( \beta \), i.e., \( \hat{\beta}_\lambda \), the (marginalized) distribution of the data can be derived from \( f_\lambda(y|\beta_\lambda) = \frac{f_{\beta, \lambda}(y)}{f_{\beta, \lambda}(\beta_\lambda)} \), Where \( f_{\beta, \lambda}(\beta_\lambda) \) is the distribution of \( \beta_\lambda \). To obtain in general the distribution of the ML estimator. Namely,

\[
f_{\beta, \lambda}(\beta_\lambda) = \left[-\frac{1}{2\pi} \frac{\delta^2 \log f_{\beta, \lambda}(y)}{\delta \beta \delta \beta} \right]^{\frac{1}{2}} \frac{f_{\beta, \lambda}(y)}{f_{\beta, \lambda}(\beta_\lambda)} \tag{8}
\]

After substitution of expression (8) into (7) one obtains:

\[
\log[f_\lambda(y|\beta_\lambda)] = \ell_M(\beta_\lambda, \lambda|y) \bigg|_{\beta = \hat{\beta}_\lambda} - 0.5 \log \left[ \frac{1}{2\pi} \right]^{\frac{1}{2}} \frac{f_{\beta, \lambda}(y)}{f_{\beta, \lambda}(\beta_\lambda)} \tag{9}
\]

With \( D(\ell_M, \beta) = -\frac{\delta^2 \ell_M}{\delta \beta \delta \beta} \). In the next step, one replaces everywhere the marginal log likelihood \( \ell_M \) by the adjusted profile log-likelihood \( p_h(h) \) evaluated in \( \beta = \hat{\beta}_\lambda \). This results in:

\[
\log[f_\lambda(y|\beta_\lambda)] = h(\beta, \lambda, v) \bigg|_{\beta = \hat{\beta}_\lambda, v = \theta} - 0.5 \log \left[ \frac{1}{2\pi} \right]^{\frac{1}{2}} \frac{1}{\beta = \hat{\beta}_\lambda} - 0.5 \log \left[ \frac{1}{2\pi} \right]^{\frac{1}{2}} \frac{p_h(h, \beta)}{p_h(h, \hat{\beta}_\lambda)} \bigg|_{\beta = \hat{\beta}_\lambda} \tag{10}
\]

Finally, it is shown that the sum of the last two terms in the above expression is equal to

\[
-0.5 \log \left[ \frac{1}{2\pi} \right]^{\frac{1}{2}} \frac{1}{\beta = \hat{\beta}_\lambda, v = \theta} \quad \text{with} \quad D[h, (\beta, v)] \quad \text{equal to} \quad \begin{pmatrix}
\frac{\delta^2 h}{\delta \beta \delta \beta} & \frac{\delta^2 h}{\delta \beta \delta v} \\
\frac{\delta^2 h}{\delta v \delta \beta} & \frac{\delta^2 h}{\delta v \delta v}
\end{pmatrix}
\]

with dimensions equal to the sum of the dimensions of two adjustment terms in (10). As a result one obtains the following adjusted profile likelihood:

\[
p_{\beta, v}(h) = h(\beta, \lambda, v) \bigg|_{\beta = \hat{\beta}_\lambda, v = \theta} - 0.5 \log \left[ \frac{1}{2\pi} \right]^{\frac{1}{2}} \frac{1}{\beta = \hat{\beta}_\lambda, v = \theta} \tag{11}
\]

The latter adjusted profile likelihood is maximized with respect to \( \lambda \) to obtain \( \hat{\lambda} \). Note that this objective function is “focused” solely on the dispersion parameters. This offers an extension of restricted maximum likelihood (REML) estimation and provides inference for the class of generalized linear mixed models. We show below that in the case of linear mixed models this function is exactly the restricted maximum likelihood[5].

3.5. The Linear Mixed Model Case

The classical linear mixed effects model assumes:

\[
f_{\beta, \lambda}(y_i | v_i) = \mathcal{N}(X_i \beta + Z_i v_i; \Sigma_i), f_{v_i}(v_i) = \mathcal{N}(0; \Lambda_i) \tag{12}
\]

The design matrix \( X_i \) contains fixed effects for the \( i \)-th subject, while \( Z_i \) is a design matrix for the random effects for the \( i \)-th subject. Matrices \( \Sigma_i \) and \( \Lambda_i \) determine the residual variance of \( y_i \) and the variance of random effects \( v_i \) respectively. Note that we have split up \( \lambda \) into \( \lambda_\sigma \) and \( \lambda_\nu \), the dispersion parameters pertaining to the residual variability \( \Sigma_i \) and random effects variability \( \Lambda_i \), thereby showing the role of each of the dispersion components. Denote by \( V \) the total (over all
subjects) marginal variance-covariance matrix \( V = \mathbf{Z}\Lambda\mathbf{Z}^\top + \Sigma \) where \( X \) is the design matrix for fixed effects obtained by stacking \( X_1 \) to \( X_N \), \( Z = \text{diag}(Z_1, ..., Z_N) \) is the design matrix of random effects, \( \Lambda = \text{diag}(\Lambda_1, ..., \Lambda_N) \) is the variance covariance matrix of the random effects and \( \Sigma = \text{diag}(\Sigma_1, ..., \Sigma_N) \) is the residual variance covariance matrix. In this case the adjusted profile likelihood \( p_v(h) \) becomes:

\[
p_v(h) = \ell_M = -\frac{1}{2} \log|2\pi V| - \frac{1}{2} (Y - X\beta)^\top V^{-1}(Y - X\beta)
\]

Which is the expression for the marginal likelihood of a linear mixed model. Further, the adjusted profile likelihood \( p_{\beta, v}(h) \) is equal to:

\[
p_{\beta, v}(h) = \log[f_{\beta}(y|\hat{\beta})] = \ell_M \big|_{\beta = \hat{\beta}} - 0.5\log \left| \frac{X^\top V^{-1}X}{2\pi} \right|
\]

Which is exactly equal to the classical REML likelihood for the linear mixed model [5].

3.6. Application in HGLM

Previous theory refers to the GLM with random effects. For HGLM modeling, the model assumes the distribution of responses \( y_{ij} \) (conditional random effect) is the exponential family.

\[
f_{\beta, \lambda_e}(y_{ij}) = \exp\left[ \frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\lambda_e} + c(y_{ij}, \lambda_e) \right]
\]

The distribution is combined with the distribution of the random component, in which the distribution relates to the family of conjugate Bayesian distribution for exponential families.

\[
f_{\lambda_v}(v_i) = \exp[a_1(\lambda_v)v_i - a_2(\lambda_v)b(v_i) + c_2(\lambda_v)]
\]

3.7. Model Checking

Checking the model after matching the random effects model is the part that can not be ignored. For HGLM, plot checking the model refers to McCullagh P and Nelder (1998) [4]. HGLM estimation method to be used to fit 2 GLM interconnected and GLM model checking method can be used to check various assumptions and kindness HGLM models (Lee Y and Nelder JA, 2001). For checking the model, can use two plots are plots residual deviances already in standardization to the fitted value plot absolute value or residual value deviances against fitted values. (Lee and Nelder JA, 1996). Refers to Lee and Nelder (1996) introduced deviances to test the suitability model, where:

\[
D = D(y, \hat{\mu}) = -2\{l(\hat{\mu}; y|v) - l(y; y|v)\}
\]

The methods section should provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference: only relevant modifications should be described. The method also includes a clear description of the design of the study, including population and sampling, and the type of analysis used, to enable replication. For studies involving human participants, a statement detailing ethical approval and consent should be included in the methods section [4].

4. Results and Discussion

In HGLM application will use some of the conditions among others that use data simulation and real data using National Labor Force Survey (Sakernas) Data in Kepulauan Riau Province August 2015 results, with one and with 2 random effects.
4.1. Simulation HGLM the Poisson Model with Random Effects Gamma Distribution.

For dependent count data that is common to model a Poisson distributed response and Gamma distributed random effects (Lee et al, 2006)[2]. Assuming no overdispersion conditional on u and has a fixed dispersion with the following models:

\[ E(y_i|\beta, u) = \exp(X_i\beta + Z_i v) \] (18)

With a level j in the random effects v, described by \( v_j = \log(u_j) \) and \( u_j \) is iid Gamma distribution with the mean and variance: \( E(u_j) = 1, \operatorname{var}(u_j) = \lambda \). Will be simulated Poisson models with random effects from gamma distribution. Will be simulated five clusters with 20 observations for each cluster. For an average of some models, and the variance of the random effects is \( \mu = 0, \sigma^2_u = 0.2 \). HGLM Model will be computed, and it will be compare with parameter that computed by frequentist model with linear model and Bayesian regression model. For Bayesian model, we use MCMCpack and Bayesian fitting of the same model with non informative prior on the coefficient vector and the default settings for the parameters governing the MCMC algorithm can be accomplished with MCMCpack using the following syntax MCMCregress and because the simulation using Poisson models, with response have a Poisson Distribution, we use MCMCpoisson. Using conjugate prior distributions, we can write a Bayesian Poisson without poisson distribution. \( y_t \sim \text{poisson}(\lambda_t) \)[1]. Comparative between frequentist model with linear model, Bayesian regression model and HGLM, we can see in Table 1.

**Table 1. Coefficient (β), Mean Square Error (MSE) for Linear Regression Model, Bayesian Model and HGLM**

| No | Model                                      | Coefficient (β) | MSE  |
|----|--------------------------------------------|-----------------|------|
| 1  | Linear Regression Model (Frequentist)      | -1.1209         | 3.2654 |
|    | Bayesian regression Model with prior Poisson (MCMCpoisson), quantiles for each variables |                 |      |
| 2  | 2.5%                                       | -0.1416         | 6.4306 |
| 3  | 25%                                        | -0.04599        | 5.9852 |
| 4  | 50%                                        | 0.00002         | 5.7726 |
| 5  | 75%                                        | 0.05087         | 5.5731 |
| 6  | 97.5%                                      | 0.1467          | 5.2326 |
|    | Bayesian regression Model (MCMCregression), quantiles for each variables |                 |      |
| 7  | 2.5%                                       | -0.4081         | 3.6360 |
| 8  | 25%                                        | -0.1421         | 3.3813 |
| 9  | 50%                                        | -0.00036        | 9.3220 |
| 10 | 75%                                        | 0.1363          | 3.4307 |
| 11 | 97.5%                                      | 0.4083          | 3.7857 |
| 12 | HGLM                                       | -0.00813        | 1.7919 |

If we compare, The Mean square error (MSE) from linear regression model (Frequentist Model), Bayesian Regression and HGLM, HGLM give best model than any other because HGLM can give smallest MSE.

4.2. HGLM Gamma Normal Beta Model (National Labor Force Surveys Data in Kepulauan Riau Province, August 2015)

From National Labor Force Survey (Sakernas) Data in Kepulauan Riau Province August 2015, The response variable is the percentage of the working population, while the explanatory variables are the
age group that have three categories: 15-24, 25-49 years old and 50 years and over. For education are six groups: elementary school down, junior high school, General senior high school, vocational school, Academy and University (D4 / S1 / S2 / S3) while the municipality in Riau Islands province is considered as replication. There are 7 municipalities, namely Karimun, Bintan, Lingga, Anambas, Natuna, Batam and Tanjung Pinang, so there are 126 observations about percentage of the working population. Further, there are two random effects, showing repetition, in this case the Municipality \((v_i)\), and population by age group in each municipality \((v_{ij})\). Linear predictor models are defined as follows:

\[
\eta_{ijk} = \text{Intercept} + \text{Age Group}_j + \text{Education}_k + \text{Age Group}_j \times \text{Education}_k + v_i + v_{ij} \quad (19)
\]

Where \(i=1,\ldots,7\) refers to Municipality, \(j = 1, \ldots, 3\) refers to age groups and \(k = 1, \ldots, 6\) educational attainment. In this model the percentage of the working population as a response variable with Gamma Distribution. From the results of hypothesis testing where \(H_0: \) Gamma distribution data, from the output shows that the P-Value is 0.421, with \(\alpha = 5\%\), we can conclude that the response variable has Gamma distribution.

Random effects in the model are \(v_i\) and \(v_{ij}\). The municipality has been determined, it is assumed normal distribution while the combination of municipality by the age group has Beta distribution.

**Table 2.** Test of Gamma Distribution for response variable

| Subject          | Value | Approx. Sig. |
|------------------|-------|--------------|
| Gamma            | .056  | .421         |
| N of Valid Cases | 126   |              |

**Figure 2.** Test of Beta Distribution for Random effects \(v_{ij}\)

The results obtained in Table 3. Explanatory variables that affect the percentage of the working populations in Kepulauan Riau province are the number of population by age group 25-49 years, the age group of 50 years and over and the number of population with graduate or post graduate degree. In addition, The interaction between the number of population by age group and the highest educational attainment completed, together does not affect the percentage of working population.

Based on Table 3, it can be concluded that the interaction of age groups and education attained no effect on the percentage of the working population, therefore it needs to be seen how the model HGLM without interaction and how the results are compared with HGLM interaction.
Table 3. Output HGLM Gamma Normal Beta Model with Interactions

| Information | Estimation | P-value |
|-------------|------------|---------|
| (Intercept) | 4.427409   | <2.00E-16 |
| Age Group   |            |         |
| 25-49 year old | 0.144366  | 0.001319 |
| 50 year old and over | 0.151708 | 0.000738 |
| Education Attainment |     |         |
| Junior High School | -0.07325 | 0.079596 |
| General Senior High School | -0.05828 | 0.16313 |
| Vocational School | -0.02039 | 0.625573 |
| Academy | 0.066814 | 0.10984 |
| University/graduate or post graduate degree | 0.104045 | 0.012778 |
| Interaction |            |         |
| 25-49 years old and Junior High School | 0.044736 | 0.449038 |
| 50 years and over old and Junior High School | 0.067205 | 0.255444 |
| 25-49 years and General Senior High School | 0.069973 | 0.23639 |
| 50 years and over and General Senior High School | 0.052325 | 0.375923 |
| 25-49 years and Vocational School | -0.00839 | 0.887157 |
| 50 years and over and Vocational School | 0.026183 | 0.657722 |
| 25-49 years and Academy | -0.06720 | 0.255468 |
| 50 years and over and Academy | -0.05080 | 0.390039 |
| 25-49 years and graduate or post graduate degree | -0.08797 | 0.136592 |
| 50 years and over and graduate or post graduate degree | -0.09382 | 0.112368 |

HGLM models without interaction is defined as follows:

\[ \eta_{ijk} = \text{Intercept} + \text{Age Group}_j + \text{Education}_k + v_i + v_{ij} \]  

(20)

Table 4. Output HGLM Gamma Normal Beta Model without Interactions

| Information | Estimation | P-Value |
|-------------|------------|---------|
| (Intercept) | 4.43045    | 2.00E-16 |
| Age Group   |            |         |
| 25-49 year old | 0.13561  | 6.45E-09 |
| 50 year old and over | 0.15128 | 9.45E-11 |
| Education Attainment |     |         |
| Junior High School | -0.03532 | 0.1548 |
| General Senior High School | -0.01689 | 0.4963 |
| Vocational School | -0.0143 | 0.5647 |
| Academy | 0.02791 | 0.261 |
| University/graduate or post graduate degree | 0.04426 | 0.0747 |

Further we fitted a Gamma Normal Beta model without interaction and compared it with the original model. If there are without interaction, explanatory variables that affect the percentage of the population worked in Kepulauan Riau Province is the amount of population by age group 25-49 years
and the age group of 50 years upwards while education groups no effect at all on the percentage of the working population in Kepulauan Riau Province.

Figure 3. Diagnostics of Gamma Normal Beta Models With and Without Interaction

Figure 3 shows model obtained either with or without interaction, the next test will be conducted to know the goodness of the model. In figure 3, it appears that in order to plot deviance residual against fitted values, for models with and without interaction does not have a significant difference, but when seen in the plot between absolute deviance residuals against Fitted Value, it appears that the line representing the distribution of absolute deviance residuals model without interaction, have a absolute deviance residuals value smaller than with interaction and smoother lines than the models with interaction, so the plot can be concluded that the model without interaction is better than the models with interaction.

To compare the Gamma Normal Beta models with and without interaction can also be done using likelihood ratio test, the results are as follows:

\[ H_0: \beta_{jk} = 0 \quad \text{(Simplified model)} \]
\[ H_1: \text{at least one } \beta_{jk} \neq 0 \quad \text{(More complex models)} \]

With marginal likelihood comparison: LR test p-value: 0.121055; LR test statistics: 15.31316; LR difference df: 10 Chi-square test statistic with degrees of freedom 10, indicating not confident enough to incorporate interaction into the model (p-value = 0.12), simpler model or a model without interaction is a better model than the model with interaction.

5. Conclusion

In statistical modeling and statistical references required information about the data of the population or sample including the type of distribution of the data itself. Hierarchical Generalized Linear Model (HGLM) extended Generalized Linier Mixed Model (GLMM) in which the distribution of random
components are extended to conjugate of arbitrary distributions from the GLM family. HGLM modeling used to model fitting with exponential family random effects. The Extended Likelihood used in HGLM should be describe hierarchical data character. HGLM give a best model than linear regression model (frequentist model) and bayesian regression because HGLM can give smallest MSE than any other model.

National Labor Force Surveys, a regular survey by Statistics Indonesia, indicates a hierarchical character data. Application of HGLM with Gamma Normal Beta model on National Labor Force Surveys data, indicate that the proportion of working population in Kepulauan Riau Province in August 2015 are affected by the number of population in age group 25-49, age group 50 years and over, and the number of population with graduate or post graduate degree. The interaction between the number of population by age group and the highest educational attainment completed, together does not affect the percentage of working population. Model without interaction between age group and highest educational attainment provides a better model compared to model with interaction.

6. References

[1] Andrew D, Martin et al 2011 J. Statist. Soft. June 2011 42 Issue 9.
[2] Lee Y et al.2006 Generalized Linear Models with Randoms Effect, Unified Analysis Via H-Likelihood (New York: Chapman & Hall) chapter 6 pp 173-202
[3] Lee Y and Nelder J.A 1996 J. R. Statist. Soc. B (1996) 58 No. 4 pp.619-78.
[4] Lee Y and Nelder J.A 1998 The Canadian Journal of Statist 26 No 1 1998 Pages 95-105
[5] Molas Marek dan Lesaffre Emmanuel 2012 J. Statist. Soft. March 2011 39 Issue 13.
[6] Sarpong Et all 2015 Int.J. Statist and App 2015 5(5): 196-207
[7] XiaShen, et al.2014 JSM-Section on Statist Comp. pp 1254-63