Southeast Asian Journal of Sciences: Vol 2, No. 1 (2013) pp. 27-40

APPLICATION OF MIXED EFFECT MODEL FOR ASSESSING DEMOGRAPHIC INDICATORS

Kuntoro, Nunik Puspitasari,
Arief Wibowo, Lutfi A. Salim and **Nurul Fitriyah**

Dept. of Biostatistics and Population Study
Airlangga University, Surabaya, Indonesia
e-mail: kuntoro2@yahoo.com; nuniksay@gmail.com;
arief-w@fkm.unair.ac.id; lut-unair@yahoo.com;fitri1975@yahoo.com

Abstract

Mixed effect models have been implemented widely by researchers who are doing experiments. Most population studies are done by observational studies such as survey research, or by secondary data analysis from census data. Independent variable of province determined by a researcher may be considered as fixed effect component, while random samples of districts or cities drawn from each province may be considered as random effect component.

Three provinces in Java Island, East Java, Central Java, and West Java, are considered as fixed effect component. Fourteen random samples that represent district or city are drawn from each province. A number of demographic indicators from each district or city are collected. The data are obtained from population census performed in 2010 in three provinces. A mixed effect model is constructed and parameters of random effect component are estimated.

1 Introduction

In an experimental research, the subjects of interest are given a definite treatment, then its effect is evaluated[3]. Suppose someone is interested in comparing the effect of two treatments such as the standard and new drugs with

Key words: fixed effect, random effect, census, mixed effect model.

additional treatment such as placebo for increasing internal validity. In this case, he is interested in the fixed effect of three treatments on the subjects. Suppose each treatment group of subjects is considered as a population. Due to limited budget, time, and personnel he does not want to involve all subjects in each population, but he selects randomly the subjects from each population. Hence, he is interested in random effect of three treatments on the subjects. When he analyzes both fixed and random effects, then he is interested in mixed effects of three treatments on the subjects.

There is a nonexperimental research in which someone is interested in analyzing the fixed, random, or mixed effect of certain characteristics on the subjects under study. An example, most population studies are done by a nonexperimental research called an observational research[3]. This includes survey research, nonobtrusive research using secondary data from census. The demographic data such as level of education and divorce rate may be considered as characteristics of interest. When someone is interested in comparing demographic characteristics across provinces, then he prefers the fixed effect model to other model. When he selects randomly a number of districts in each province and he is interested in random effect on demographic characteristics, then he prefers the random effect model to other model. When he is interested in both fixed and random effect on demographic characteristics, then he prefers the mixed effect model to other model.

The data were obtained from Central Board of Statistics[1] based on Indonesian population census in 2010. Three provinces were selected, these were East Java, Central Java and West Java. These provinces are located in Java Island of Indonesia. From each province was selected randomly ten districts and four cities. The data were analyzed by mean of Mixed Effect Model. Maximum Likelihood Estimation was used to estimate both parameters of fixed and random effects.

2 Basic Concepts

2.1 Model

Mixed-effects model with fixed row factors and random column factor may be presented as follows[2]:

$$Y_{ijk} = \mu + \alpha_i + B_j + C_{ij} + E_{ijk} \tag{1}$$

In which each α_i is considered as a constant such that

$$\sum_{i=1}^{r} \alpha_i = 0 \tag{2}$$

Moreover, B_j, C_{ij}, and E_{ijk} are mutually independent random variables that satisfy:

$$
\begin{aligned}
B_j &\sim N(0, \sigma_C^2) \\
C_{ij} &\sim N(0, \sigma_{RC}^2) \\
E_{ijk} &\sim N(0, \sigma^2)
\end{aligned} \tag{3}
$$

in which $i = 1, 2, \cdots, r$ $j = 1, 2, \cdots, c$ and $k = 1, 2, \cdots, n$.

Mixed-effects model with fixed column factors and random rows factors may presented as follows.

$$
Y_{ijk} = \mu + A_i + \beta_j + C_{ij} + E_{ijk} \tag{4}
$$

In which each β_j is considered as a constant such that

$$
\sum_{j=1}^{c} \beta_j = 0 \tag{5}
$$

Moreover, A_i, C_k, and E_{ijk} are mutually independent random variables that satisfy:

$$
\begin{aligned}
A_i &\sim N(0, \sigma_R^2) \\
C_{ij} &\sim N(0, \sigma_{RC}^2) \\
E_{ijk} &\sim N(0, \sigma^2)
\end{aligned} \tag{6}
$$

in which $i = 1, 2, \cdots, r$ $j = 1, 2, \cdots, c$ and $k = 1, 2, \cdots, n$.

2.2 Testing Hypothesis

2.2.1 Row Effect

- Fixed effects. $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$.
- Random effects. $H_0 : \sigma_R^2 = 0$.
- Mixed effects with rows fixed and columns random. $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$.
- Mixed effects with rows random and columns fixed. $H_0 : \sigma_R^2 = 0$.

2.2.2 Column Effect

- Fixed effects. $H_0 : \beta_1 = \beta_2 = \cdots = \beta_c = 0$.
- Random effects. $H_0 : \sigma_C^2 = 0$.
- Mixed effects with rows fixed and columns random. $H_0 : \sigma_C^2 = 0$.
- Mixed effects with rows random and columns fixed. $H_0 : \beta_1 = \beta_2 = \cdots = \beta_c = 0$.

2.2.3 Interaction Effect

- Fixed effects. $H_0 : \gamma_{ij} = 0$, for all i, j.
- Random effects. $H_0 : \sigma_{RC}^2 = 0$.
- Mixed effects with rows fixed and columns random. $H_0 : \sigma_{RC}^2 = 0$.
- Mixed effects with rows random and columns fixed. $H_0 : \sigma_{RC}^2 = 0$.

2.3 Expected Mean Squares (EMS)

2.3.1 Row Effect

- Fixed effects. $\sigma^2 + cn \sum_{i=1}^{r} \frac{\alpha_i^2}{r-1}$.
- Random effects. $\sigma^2 + n\sigma_{RC}^2 + cn\sigma_R^2$.
- Mixed effects with rows fixed and columns random. $\sigma^2 + n\sigma_{RC}^2 + cn\sigma_R^2$.
- Mixed effects with rows random and columns fixed. $\sigma^2 + n\sigma_{RC}^2 + cn\sigma_R^2$.

2.3.2 Column Effect

- Fixed effects. $\sigma^2 + rn \sum_{j=1}^{c} \frac{\beta_j^2}{c-1}$.
- Random effects. $\sigma^2 + n\sigma_{RC}^2 + rn\sigma_C^2$.
- Mixed effects with rows fixed and columns random. $\sigma^2 + n\sigma_{RC}^2 + rn\sigma_C^2$.
- Mixed effects with rows random and columns fixed. $\sigma^2 + n\sigma_{RC}^2 + rn \sum_{j=1}^{c} \frac{\beta_j^2}{c-1}$.

2.3.3 Interactions

- Fixed effects. $\sigma^2 + n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\gamma_{ij}^2}{(r-1)(c-1)}$.
- Random effects. $\sigma^2 + n\sigma_{RC}^2$.
- Mixed effects with rows fixed and columns random. $\sigma^2 + n\sigma_{RC}^2$.
- Mixed effects with rows random and columns fixed. $\sigma^2 + n\sigma_{RC}^2$.

2.3.4 Error

- Fixed effects. σ^2.
- Random effects. σ^2.
- Mixed effects with rows fixed and columns random. σ^2.
- Mixed effects with rows random and columns fixed. σ^2.

Selecting the numerator as well as the denominator is important to select various F statistics required for testing the hypothesis of interest. Moreover, the numerator mean square always corresponds to the factor to be considered[2]. When someone considers "rows" as factor, then the numerator mean square is MSR no matter the type of model. When someone considers "columns " or "interaction" as factor, then the numerator mean square is respectively MSC or MSRC. Furthermore, the denominator selected corresponds to the EMS to which the numerator EMS reduces under the null hypothesis of interest. Suppose someone is interested to test the hypothesis of row effects in a random-effect model. Hence, the numerator EMS

$$(\sigma^2 + n\sigma_{RC}^2 + cn\sigma_R^2) \tag{7}$$

It reduces to

$$\sigma^2 + n\sigma_{RC}^2 \tag{8}$$

under $H_0 : \sigma_R^2 = 0$. Then the denominator mean square is MSRC, since the EMS of MSRC under the Random Effects Model is exactly.

$$\sigma^2 + n\sigma_{RC}^2 \tag{9}$$

Thus, the Ratio of EMS is

$$\frac{EMS(R)}{EMS(RC)} = \frac{\sigma^2 + n\sigma^2_{RC} + cn\sigma^2_R}{\sigma^2 + n\sigma^2_{RC}} \tag{10}$$

It reduces to

$$\frac{\sigma^2 + n\sigma_{RC}}{\sigma^2 + n\sigma_{RC}} = 1 \tag{11}$$

under $H_0 : \sigma^2_R = 0$. Hence, the F statistic $\frac{MSR}{MSRC}$ is the ratio of two estimators of the same variance under H_0.

Suppose someone is interested in testing the hypothesis of Row Effect based on the mixed effect model with the Row Factor is Fixed and The Column Factor is random. The test statistic is

$$F = \frac{MSR}{MSRC}$$

It involves the following ratio of EMS:

$$\frac{EMS(R)}{EMS(RC)} = \frac{\sigma^2 + n\sigma^2_{RC} + cn\Sigma^r_{i=1}\frac{\alpha^2_i}{(r-1)}}{\sigma^2 + n\sigma^2_{RC}} \tag{12}$$

Under $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. This ratio may be simplified to be

$$\frac{(\sigma^2 + n\sigma_R C^2)}{(\sigma^2 + n\sigma_R C^2)} = 1 \tag{13}$$

Hence, the F statistic is the ratio of the two estimators of the variance under H_0.

Suppose someone is interested in testing Row Effect based on the mixed effect model with the Row Factor is Random and the Column Factor is Fixed. Hence, the F statistic is

$$F = \frac{MSR}{MSRC}$$

It involves the following ratio of EMS as follows.

$$\frac{EMS(R)}{EMS(RC)} = \frac{(\sigma^2 + n\sigma^2_{RC} + cn\sigma^2_R)}{(\sigma^2 + n\sigma^2_{RC})} \tag{14}$$

Moreover, under $H_0 : \sigma_R^2 = 0$, this ratio may be simplified as follows.

$$\frac{(\sigma^2 + n\sigma_{RC}^2)}{(\sigma^2 + n\sigma_{RC}^2)} = 1 \quad (15)$$

3 Materials and Methods

An obtrusive or nonreactive research[4] was used in this study by analyzing secondary data from Indonesian Population Census 2010[1]. Three provinces, East Java, Central Java, and West Java, were selected. These provinces are located in Java Island, the most populated island in Indonesia. From each province, fourteen random samples of districts/cities were selected. Demographic characteristics that include percentage of people graduated high school or higher, percentage of widows due to divorce, and percentage of widows due to husband's death are considered as dependent variables. The first indicator was selected since it is one of Human Development Index (HDI) components. The last two indicators were selected since they have strong association with family integrity and family welfare. These family conditions will influence three components of HDI either directly or indirectly. These components include health, education, economic. The fixed effects of provinces, the random effects of random selection of district/cities, and mixed effects on dependent variables were analyzed.

4 Results and Discussion

4.1 People graduated high school or higher

Table 1: Fixed effects of province

Type III Tests of Fixed Effects				
Source	**Numerator df**	**Denominator df**	**F**	**Sig.**
Intercept	1	1.986	429.414	.002
province	2	39.992	6.047	.005

Table 1 shows the results of F statistic for intercept and fixed effect of province on percentage of people graduated high school or higher. There is fixed effect of province on percentage of people graduated high school or higher with p-value equal to .005. That means province, in the context of provincial government with its policy in population program, may influence the percentage of people graduated high school or higher. Since education is one component of human develoment index(HDI), the provincial government allocates the budget for enhancing level of education until high school or higher. Null hypothesis that intercept is equal to zero is to be rejected with p-value equal to .002. That means in mixed effect modeling, intercept should be included in the model.

Table 2: Estimates of fixed effect parameters

Estimates of Fixed Effects							
						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	1.886872	.094008	2.581	20.071	.001	1.558243	2.215501
[province=1,00]	-.177143	.057123	39.992	-3.101	.004	-.292594	-.061691
[province=2,00]	-.010714	.057123	39.992	-.188	.852	-.126166	.104737
[province=3,00]	0	0	.	.	.	.	.

Table 2 shows the results of t statistic for intercept and fixed effect of each province as well as 95 percent of confidence intervals for population paramaters of intercept and fixed effect. East Java province (code = 1.00) may influence the percentage of people graduated high school or higher with its p-value equal to .004. Compared to Central Java and West Java, East Java in the best in educational program for people in the community. In the national level, East Java province is considered as successful indicator of national development. Since the p-value of intercept is .001, intercept should be included in the model.

Table 3: Estimates of random effect parameters

Estimates of Covariance Parameters							
						95% Confidence Interval	
	Parameter	Estimate	Std. Error	Wald Z	Sig.	Lower Bound	Upper Bound
Residual		.022842	.005108	4.472	.000	.014736	.035406
district	Variance	.014188	.015534	.913	.361	.001659	.121307

Table 3 shows the within district variance component as well as district variance component. The estimates of within district and district variance components are respectively .022842 and .014188. Table 4 shows the random effect covariance structure. Random effect covariance represents within district variance component. Again, table 3 shows that there is random effect within district since Wald Z statistic shows its p-value equal to .000.

Table 4: Within district variance component

Random Effect Covariance Structure (G)		
	[district=1,00]	[district=2,00]
[district=1,00]	.014188	0
[district=2,00]	0	.014188

Table 5: Means of fixed effect parameters

Estimates					
				95% Confidence Interval	
province	Mean	Std. Error	df	Lower Bound	Upper Bound
east java	1.710	.094	2.581	1.381	2.038
central java	1.876	.094	2.581	1.548	2.205
west java	1.887	.094	2.581	1.558	2.216

Table 5 shows the estimates of means of percentages of people graduated high school or higher. The estimates of means for East Java, Central Java, and West Java are respectively 1.710, 1.876, and 1.887. Their standard errors are the same. That means variation across provinces are stable.

Table 6: Multiple comparison of means

Pairwise Comparisons							
						95% Confidence Interval for Difference	
(I) province	(J) province	Mean Difference (I-J)	Std. Error Error	df df df	Sig.a	Lower Bound	Upper Bound
central java	east java	,166*	.057	39.992	.006	.051	.282
west java	east java	,177*	.057	39.992	.004	.062	.293

Table 6 shows multiple comparison among means using LSD method. The means between East Java and Central Java, East Java and West Java are significatly different with p-values are respectively .006 and .004.

4.2 Widows due to divorce

Table 7: Fixed effects of province

Type III Tests of Fixed Effects				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	42	2684.247	.000
province	2	42	49.729	.000

Table 7 shows the results of F statistic for intercept and fixed effect of province on percentage of widows due to divorce. There is fixed effect of province on percentage of widows due to divorce with p-value equal to .000. That means province, in the context of provincial government with its policy in woman empowerment and family welfare program, may influence the percentage of widows due to divorce. Woman empowerment program as well as family welfare program are important to reduce divorce among families. Null hypothesis that intercept is equal to zero is to be rejected with p-value equal to .000. That means in mixed effect modeling, intercept should be included in the model.

Table 8: Estimates of fixed effect parameters

Estimates of Fixed Effects							
Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	3.445000	.090908	42	37.895	.000	3.261539	3.628461
[province=1,00]	-.962143	.128564	42	-7.484	.000	-1.221595	-.702690
[province=2,00]	-1.215000	.128564	42	-9.451	.000	-1.474453	-.955547
[province=3,00]	0	0	.	.	.	.	.

Table 8 shows the results of t statistic for intercept and fixed effect of each province as well as 95 percent of confidence intervals for population paramaters of intercept and fixed effect. East Java province (code = 1.00) as well as Central Java province (code = 2.00) may influence the percentage of widows due to divorce with their p-values equal to .000. These provinces have good programs of woman empowerment and family welfare. In the national level, these provinces are considered as successful indicators of these programs. Since the p-value of intercept is .000, intercept should be included in the model.

Table 9: Estimates of random effect parameters

Estimates of Covariance Parameters							
Parameter		Estimate	Std. Error	Wald Z	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Residual		.115701	.025248	4.583	.000	.075438	.177453
district	Variance	.000000	.000000	.	.	.	.

Table 9 shows within district variance component as well as district variance component. The estimates of within district and district variance components are respectively .115701 and .000000. Table 10 shows the random effect covariance structure. Random effect variance represents within district variance component. Again, table 9 shows that there is random effect within district since Wald Z statistic shows its p-value equal to .000.

Table 11 shows the estimates of means of percentages of widows due to divorce. The estimates of means for East Java, Central Java, and West Java are respectively 2.483, 2.230, and 3.445. Their standard errors are the same. That means variation across provinces are stable. West Java province shows the highest value of mean of percentage of widows due to divorce compared to the two other provinces.

Table 12 shows multiple comparison among means of percentage of widows

Table 10: Within district variance component

Random Effect Covariance Structure (G)		
	[district=1,00]	[district=2,00]
[district=1,00]	.000000	0
[district=2,00]	0	.000000

Table 11: Means of fixed effect parameters

Estimates					
				95% Confidence Interval	
province	Mean	Std. Error	df	Lower Bound	Upper Bound
east java	2.483	.091	42	2.299	2.666
central java	2.230	.091	42	2.047	2.413
west java	3.445	.091	42	3.262	3.628

Table 12: Multiple comparison of means

Pairwise Comparisons							
						95% Confidence	
(I) province	(J) province	Mean	Std. Error	df	Sig.a	Lower Bound	Upper Bound
central java	east java	-.253	.129	42	.056	-.512	.007
west java	east java	.962	.129	42	.000	.703	1.222

due to divorce using LSD method. The means between East Java and West Java are significatly different with p-value equal to .000. The means between East Java and Central are not significantly different with p-value equal to .056.

4.3 Widows due to husband's death

Table 13 shows the results of F statistic for intercept and fixed effect of province on percentage of widows due to husband's death. There is fixed effect of province on percentage of widows due to husband's death with p-value equal to .000. That means province, in the context of provincial government with its policy in woman empowerment and family welfare program, may influence the percentage of widows due to husband'd death. Woman empowerment program as well as family welfare program are important to enhance the welfare of widows due to husband's death. Null hypothesis that intercept is equal to zero is to be rejected with p-value equal to .036. That means in mixed effect modeling, intercept should be included in the model.

Table 14 shows the results of t statistic for intercept and fixed effect of each province as well as 95 percent of confidence intervals for population paramaters of intercept and fixed effect. East Java province (code = 1.00) as well as Central Java province (code = 2.00) may influence the percentage of widows due to

Table 13: Fixed effects of province

Type III Tests of Fixed Effects				
Source	Numerator df	Denominator df	F	Sig.
Intercept	1	.960	382.724	.036
province	2	38	22.506	.000

Table 14: Estimates of fixed effect parameters

Estimates of Fixed Effects							
						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	8.721606	.651977	1.847	13.377	.007	5.682046	11.761167
[province=1,00]	3.997143	.597995	38	6.684	.000	2.786566	5.207720
[province=2,00]	2.297143	.597995	38	3.841	.000	1.086566	3.507720
[province=3,00]	0	0	.	.	.	.	.

husband's death with their p-values equal to .000. These provinces have good programs of woman empowerment and family welfare. In the national level, these provinces are considered as successful indicators of these programs. Since the p-value of intercept is .007, intercept should be included in the model.

Table 15: Estimates of random effect parameters

Estimates of Covariance Parameters								
							95% Confidence Interval	
	Parameter	Estimate	Std. Error	Wald Z	Sig.		Lower Bound	Upper Bound
Residual		2.503183	.574269	4.359	.000		1.596664	3.924384
district	Variance	.472068	.874748	.540	.589		.012494	17.835816

Table 15 shows within district variance component as well as district variance component. The estimates of within district and district variance components are respectively 2.503183 and .472068. Table 16 shows the random effect covariance structure. Random effect variance represents within district variance component. Random effects variances in East Java province and Central Java province are .157562. Again, table 15 shows that there is random effect within district since Wald Z statistic shows its p-value equal to .000.

Table 16: Within district variance component

Random Effect Covariance Structure (G)		
	[district=1,00]	[district=2,00]
[district=1,00]	.157562	0
[district=2,00]	0	.157562

Table 17: Means of fixed effect parameters

Estimates					
				95% Confidence Interval	
province	Mean	Std. Error	df	Lower Bound	Upper Bound
east java	12.774	.506	5.168	11.487	14.062
central java	11.074	.506	5.168	9.787	12.362
west java	8.777	.506	5.168	7.490	10.064

Table 17 shows the estimates of means of percentages of widows due to husband's death. The estimates of means for East Java, Central Java, and West Java are respectively 12.774, 11.074, and 8.777. Their standard errors are the same. That means variation across provinces are stable. East Java province shows the highest value of mean of percentage of widows due to husband's death compared to the two other provinces.

Table 18: Multiple comparison of means

Pairwise Comparisons							
(I) province	(J) province	Mean Difference (I-J)	Std. Error	df	Sig.a	95% Confidence Interval for Difference	
						Lower Bound	Upper Bound
central java	east java	-1,700*	.583	39.725	.006	-2.879	-.521
west java	east java	-3,997*	.583	39.725	.000	-5.177	-2.818

Table 18 shows multiple comparison among means of percentage of widows due to husband's death using LSD method. The means between East Java and Central Java are significatly different with p-value equal to .006. Also the means between East Java and West Java are significantly different with p-value equal to .000.

Maximum likelihood estimation method was used to estimate the parameters in this study. This method is better than restricted maximum likelihood estimation method. This method gives all p-values greater than those obtained by restricted maximum likelihood estimation method (the results are kept by the author, not being presented in this paper). However, the two methods still give the results that reject the null hypotheses. There is hierarchical administration system in the regions. Under province, there is a number of districts and cities. Districts and cities as administration system in the regions show variation of decision making process as well as budgeting. These may influence demographic indicators such as people completed high school or higher, widows due to divorce and widows due to husband's death. The first indicator is part of HDI component. The provincial governments as well as district and city governments concern about HDI as a measure of successful regional development and also as indicator of regional government performance. The second indicator is a measure of family integrity that may result in various family and social consequences. The third indicator may indirectly influence family welfare particularly when widows have low education, no jobs, many children although they still keep family integrity. The last two indicators also the problems that should be solved by regional governments.

5 Conclusion and Recommendation

Percentages of people completed high school education or higher, widows due to divorce, and widows due to husband's death may be influenced by province. There are fixed effects on these demographic indicators.

Random effects variance represented by within district variance may influence percentages of people completed high school education or higher, widows due to divorce, and widows due to husband's death.

Maximum likelihood estimation method gives better probability for rejecting null hypothesis than restricted maximum likelihood estimation method.

It is recommended to apply mixed effects models for explaining the population problems in the regions. For estimating the parameters of mixed effects models, it is recommended to apply maximum likelihood estimation method.

References

[1] BPS, *Indonesian Population Census 2010*, Central Board of Statistics, Jakarta (2011).

[2] D.G. Kleinbaum, L. L. Kupper, K. E, Muller, and A. Nizam, *Applied Regression Analysis and Other Multivariable Methods*, Duxbury Press, Pacific Groove (1998).

[3] K. Kuntoro, *Philosophycal Basis of Research Methodology*, 2nd Edition, C.V. Pustaka Melati, Surabaya(2011).

[4] W.L. Neuman, *Social Research Methods - Qualitative and Quantitative Approaches-Sixth Edit.*, Pearson Public., Boston (2006).