Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 116 (2017) 284–291

Procedia Computer Science

www.elsevier.com/locate/procedia

2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017, 13-14 October 2017, Bali, Indonesia

Linear Mixed Model for Analyzing Longitudinal Data: A Simulation Study of Children Growth Differences

Novi Hidayat Pusponegoro[a], Ro'fah Nur Rachmawati[b],

Khairil Anwar Notodiputro[c]*, Bagus Sartono[c]

[a]*Institute of Statistics, Otista Raya No. 64C, Jakarta 13330, Indonesia*
[b]*Bina Nusantara University, Kebun Jeruk Raya No. 27, Jakarta 11530, Indonesia*
[c]*Bogor Agricultural University, Jl. Raya Darmaga Bogor, Bogor 16680, Indonesia*

Abstract

Growth developmental research is one example of the application of longitudinal data that have correlated value over time. Linear Mixed Model (LMM) is an extension of classic statistical procedures that provides flexibility analysis in correlated longitudinal data and allows researcher to model the covariance structures that represent its random effects. This paper briefly describes growth curves model as a single LMM that represent two levels of observation, which focused on modeling its covariance structure to capture correlated information over time of individual performance. We apply LMM and model different types of its covariance structure in the simulation study of children's growth differences based on the feeding methods. We perform simulation scenario using MIXED procedure in SAS system, based on three fit indices (-2RLL, AIC and SBC) and *p*-value significance level, we obtain Unstructured (UN) covariance is always be the best fit in presenting the characteristic of data but not the best choice considering inefficient numbers of parameters while Heterogeneous First-order Autoregressive (ARH(1)) is a proper alternative covariance structure with ease of data interpretation from fewer numbers of estimated parameters.

Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.

Keywords: covariance structure; reapated measurement; *within-subject*; growth curves; marjinal model.

* Corresponding author. Tel.: +62- 251 - 8624535; fax: +62- 251 - 8624535.
E-mail address: khairilnotodiputro@gmail.com

1877-0509
Peer-review under responsibility of the scientific committee of the 2nd International Conference on Computer Science and Computational Intelligence 2017.
10.1016/j.procs.2017.10.071

1. Introduction

Linear model is an approach for modeling the relationship between n independent observations $\mathbf{y} = \left(y_1,...,y_n\right)^T$ with one or more explanatory variables denoted $\mathbf{X} = \left(x_{ij}\right)$ as n x p matrix model, where x_{ij} is the value of explanatory variable j for observation i. So that, in matrix notation the linear model can be expressed as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{\beta}$ is a p x 1 parameter vector with $p \le n$ and $\boldsymbol{\varepsilon}$ is a Gaussian random component having $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathbf{V} = \mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\sigma}^2\mathbf{I}$. Note that in this classical linear model, the variance structure is presumed in ordinary way that are constant variance and uncorrelated data over time. When we work on longitudinal data analysis this model become less appropriate, because by assuming the constant variance and ignoring the correlation structure the statistical inference from linear model might be erroneous.

Linear Mixed Model (LMM) or Gaussian Mixed Model (MM) analysis is a statistics procedure that provides a flexible approach in statistical analyses with correlated longitudinal data. Longitudinal data is described by response variable with repeatedly measured for each unit analysis and time as the repeated factor. In longitudinal data structure, the number of repeated measure may not be equal for each unit analysis and the time of observation may also not in the same spacing or intervals. In other words, longitudinal data characteristics allow for unequal space and missing value observations. LMM covers characteristics of longitudinal data analysis while in Linear Model are hardly applicable.

Growth developmental researchs evolve in wide variety of scientific fields, such as psychology[10], economics[3] and educations[11]. Growth curves in developmental study are usually represented by means of the two-level hierarchical linear model, in which population parameters, individual effects, and *within-subject* variation are defined at the first level and *between-subject* variation defined at the second level[13]. Growth curves are very interested to develop because we can analyze simultaneously individual level effect/individual differences (or *within-subject*) and group level effect (or *between-subject*)[2] that changes over time. Harville D. A. (1977)[7] and Jaume A. et al. (2010)[8] stated that LMM as the integration of two levels (hierarchical) observation (i.e. *within* and *between-subject*) in a single model, so that LMM fulfill the needs of capturing the individual performance information over time accurately because it allows a variety of variance covariance structure or correlation patterns to be explicitly modeled.

As the specification of Generalized Linear Mixed Model (GLMM), the term Mixed in LMM refers to the fixed and random effect simultaneously used and analyzed in the model. The dissimilarities between fixed and random effect are defined by Lee, Y. *et. al*. (2006)[9] that fixed effects can describe systematic mean patterns such as treatment methods, while random effects may describe correlation patterns between repeated measures within subjects or heterogeneities between subjects or both. The correlation can be represented by random effects of subjects, and heterogeneities by saturated random effects. In practice, it is often necessary to have both types of random components. However, what really distinguish the LMM from the classic linear model is it models variance and covariance (random effects), so that the parameters of covariance are employed to model data with specific characteristics.

As part of growth developmental research, this paper briefly describes growth curves as a single LMM that represent two-level observations of longitudinal data. Level-1 presents growth shape to capture within-subject effect and level-2 presents growth parameters that characterized between-subject differences. As we stated before, the excellence of LMM is the variance covariance modeling while in linear model it is presumed as uncorrelated and constant variance over time. In this paper, we model the variance and covariance using different structure types such as Unstructured (UN), Compound Symmetric (CS), Heterogeneous Compound Symmetric (CSH), First-order Autoregressive (AR(1)) and Heterogeneous First-order Autoregressive (ARH(1)). In orders to explore the characteristic of correlated longitudinal data and excavate the individual trend performance, we perform simulation study and apply variance covariance modeling to generated children's weight growth that we divide into two groups based on its feeding methods. Simulation study developed using MIXED procedures in SAS system that allows researcher to analyze both fixed and random effect in LMM.