

https://dspace.org/

Linear mixed effects models: misspecification of the covariance structure studied through simulation and applications to an environmental data set

Kirton, Alecia

2009-07-08T05:51:02Z
http://hdl.handle.net/10539/7076

Downloaded from a DSpace repository. https://dspace.org/

ABSTRACT

The longitudinal experimental design with continuous responses is common in many fields, particularly in the ecological sciences. Due to the complexities in the data, such as the correlation between measurements from the same subject, appropriate forms of analysis need to be implemented. A review of the available models is provided. The linear mixed effects model is a widely accepted method for analysis of repeated measurements. This study explores how this model accounts for correlation in the data set through the inclusion of random effects and modelling of the covariance structure. In particular, the robustness of the linear mixed effects model to misspecification is explored by means of a simulation study.

Data is simulated, based on the Potthoff and Roy dental study data set (Potthoff & Roy, 1964), from models under a selection of covariance structures and random effects. Model fit of linear mixed effects models to the simulated data under the same selection of covariance structures is explored by means of information criteria, and coverage probabilities of the confidence intervals for the estimated fixed effects parameters were used to test the robustness of the linear mixed effects model to misspecification of the covariance structure. The results of this analysis showed the random intercept and slope models with unstructured covariance matrix for the random effects and either independent or autoregressive covariance structure assumed for the random errors had robust specifications for the model covariance. In addition, the linear model with Toeplitz error structure was also shown to be a robust choice, and obtained good model fit statistics. This study showed that models which obtained good model fit criteria did not necessarily obtain satisfactory coverage probabilities.

To demonstrate the implementation of linear mixed effects models in the field of environmetrics, the linear mixed effect model was fitted to a biological control data set. The process for selecting the best fitting model, with respect to both the mean structure and the covariance structure, was carried out on the data set in order to illustrate the diagnostic procedure. A forecasting exercise was also carried out on the data to determine which covariance model predicted new data best. Both the random intercept and slope model with independent errors and unstructured random effects, and the no random effects model with Toeplitz error obtained good model fit statistics and predicted new data well. It was found that under the correctly specified mean structure, the covariance matrices estimated were simpler compared to a less adequate mean structure. Therefore, correctly specifying the mean has the advantage that simpler covariance structures can be chosen. When the mean model is poorly specified, then the covariance estimates attempt to compensate by inflating certain covariances, thereby increasing the standard error of the fixed effects estimates.

The ordinary regression model was compared to models which modelled the covariance structure, and although the mean structure for this model was found to be unbiased, the inferences based on this model were always inferior.

Different methods of estimating the parameters of the linear mixed effects model are discussed. The procedure used to fit models in this study was SAS PROC MIXED (ver. 9.1). This study suggests that more research and development should go into better algorithms for fitting linear mixed effects parameters, such as the EM algorithm or hierarchical Bayes methods, as failure of the generally used Newton-Raphson algorithm to obtain estimates for the covariance structure was a common problem encountered in this study.