Linear Mixed Models and Small Area Estimation

Tatsuya Kubokawa

Graduate School of Economics, University of Tokyo

Abstract
Sample survey data can be used to derive a reliable estimate of the total mean for a large area. When the same data are used to estimate means of small areas like a town, city, or county belonging to the large area, the usual direct estimators such as the sample mean have unacceptably large standard errors due to the small sizes of the samples in the small areas. This is known as the small area problem. To obtain more accurate estimates for given small areas, one needs to “borrow strength” from the related areas. The linear mixed model (LMM) is recognized as an appropriate model for handling such a problem, and the resulting empirical best linear unbiased predictor (EBLUP) can yield a smaller standard error. This article reviews small area estimation based on LMM. In particular, it explains how the structure of common parameters plus random effects in LMM works to get accurate estimates. The estimators of the mean squared errors of EBLUP and the confidence interval based on EBLUP are derived to evaluate the accuracy of EBLUP. Finally, some generalizations and various variants of LMM are described in order to analyze spatial data, and the generalized linear mixed model (GLMM) and its application to mortality rate estimation are explained.

Key words: confidence interval, empirical Bayes method, finite population, generalized linear mixed model, linear mixed model, mean squared error, random effects, small area estimation, variance components model.

E-mail address: tatsuya@e.u-tokyo.ac.jp (Tatsuya Kubokawa)

Received September 27, 2006; Received in final form November 29, 2006; Accepted December 3, 2006.