A comparison of three estimators in GMANOVA model when the sample size is fewer than the dimension

Hirokazu Yanagihara[1] **and Muni S. Srivastava**[2]

[1] *Hiroshima University, Japan, email:yanagi@math.sci.hiroshima-u.ac.jp*
[2] *University of Toronto, Canada, email:srivasta@utstat.toronto.edu*

Keywords: GMANOVA model, high dimensional data, least square estimator, Moore-Penrose inverse, ridge type estimator.

A generalized multivariate analysis of variance (GMANOVA) model proposed by Potthoff and Roy [1] is one of statistical models suitable for a longitudinal data. The matrix form of GMANOVA model is given by

$$Y \sim N_{n\times p}(A\Xi X', \Sigma \otimes I_n),$$

where Y is an $n \times p$ response variables matrix, A is an $n \times k$ between-individuals explanatory variables matrix with the full rank k $(< n)$, X is a $p \times q$ within-individuals explanatory variables matrix with the full rank q $(\leq p)$, and Ξ is a $k \times q$ unknown regression coefficients matrix. It is a known fact that a maximum likelihood estimator (MLE) of regression coefficients Ξ in the GMANOVA model is defined as

$$\hat{\Xi}_{\mathrm{ML}} = (A'A)^{-1}A'YS^{-1}X(X'S^{-1}X)^{-1},$$

where $S = Y'\{I_n - A(A'A)^{-1}A'\}Y/(n-k)$. However, if the dimension p is larger than the sample size n, the MLE of Ξ cannot be defined because the inverse matrix of S does not exist then. Hence, we avoid such an undesirable situation by the following three methods:

(1) We ignore S^{-1}. This is corresponding to the least square estimator (LS) of Ξ, i.e.,

$$\hat{\Xi}_{\mathrm{LS}} = (A'A)^{-1}A'YX(X'X)^{-1}.$$

(2) We replace S^{-1} with the inverse matrix of the ridge type estimator of S, i.e.,

$$\hat{\Xi}_{\mathrm{R}} = (A'A)^{-1}A'YS_{\lambda}^{-1}X(X'S_{\lambda}^{-1}X)^{-1},$$

where $S_{\lambda} = S + \lambda I_p/(n-k)$ and $\lambda = \mathrm{tr}(S)/\sqrt{p}$. This ridge type estimator of S proposed by Srivastava and Kubokawa [2].

(3) We replace S^{-1} with the Moore-Penrose inverse of S, i.e.,

$$\hat{\Xi}_{\mathrm{MP}} = (A'A)^{-1}A'YS^{+}X(X'S^{+}X)^{-1}.$$

An aim of this paper is to compare with above three estimators theoretically and numerically when the dimension p is larger than the sample size n.

References

[1] Potthoff, R. F., Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–325.

[2] Srivastava, M. S., Kubokawa, T. (2008). Akaike information criterion for selecting components of the mean vector in high dimensional data with fewer observations. *Journal of the Japan Statistical Society* **38**, 259–283.