Selection Effects of Common Variables on Statistical Matching

Yukiko Kurihara[1*]

1. Department of Humanities, Hirosaki University, Japan
*Contact: yukuri@cc.hirosaki-u.ac.jp

Keywords: Bayesian regression imputation, Mahalanobis method, Multiple imputation, Sampling experiment

This study verifies the precision of correlation coefficients based on statistical matching and multiple imputation under different matching methods and combinations of common variables. The matching methods for verification are a non-parametric approach based on Mahalanobis distance (package **StatMatch**) and the Bayesian regression imputation method (NIBAS)—a parametric method [5]. Questionnaire data from the Financial Statements Statistics of Corporations by Industry (Ministry of Finance) were used to clarify the effectiveness of matching data created from different sample datasets.

The three main findings are as follows: First, NIBAS enables the estimation of correlation coefficients with lesser bias than those of the Mahalanobis matching method when one aims to obtain only one set of statistics. Second, the primary condition for high-precision estimation is a combination of common variables with both low conditional dependence and strong correlation with target variables. Finally, the confidence interval computed by multiple imputation with NIBAS suitably covers the true value and measures the uncertainty inherent in statistical matching, except in the case of point estimates with extremely large bias.

References

[1] D'Orazio, M., M. Di Zio, & M.Scanu (2006), *Statistical Matching: Theory and Practice*, Wiley, West Sussex.

[2] Goel, P.K. & T. Ramalingam (1980), *The Matching Methodology: Some Statistical Properties*, Springer, Berlin.

[3] Little, R.J.A. & D.B. Rubin (2002), *Statistical Analysis with Missing Data*, Wiley, New York.

[4] Mase, S (2010). *R Programming Manual* (in Japanese), Surikougaku-sha, Tokyo.

[5] Rässler, S. (2002), *Statistical Matching*, Springer, New York.