

公的統計調査におけるロバストな比率の推定による 企業の経理項目の欠測の補定について

独立行政法人 統計センター 坂下 佳一郎, 床 裕佳子, 和田 かず美

目的

一般に統計調査から得られる観測データは欠測することがあり、そのまま集計すれば結果に偏りが生じる場合には補定 (imputation) 処理を行う必要がある。

本報告では、企業の経理項目を対象に、経理項目間の比率を用いて欠測項目の補定を行う場合を想定し、外れ値の影響を受けにくい比率補定 (ratio imputation) の方法を提案する。

ロバストな比率の推定とそれを利用した補定

政府の統計調査の集計実務において、比率補定はよく使用される方法の一つである。これは、複数要素からなる集団があり、各要素が二つの数量項目を持っている場合、要素ごとに求めた二つの数量比率の平均値として得られる比率推定量 (ratio estimator) を用いて欠測を補定するものである。ところがこの推定量はロバストではない。例えば、比率が他と大きく異なる要素が集団に含まれば推定量が大きく変わってしまう。

通常その影響を少なくするには、統計検定あるいは目視などの方法で外れ値を確認し除去する必要がある。その作業の効率化のために、本報告では繰返し加重最小二乗 (IRLS) 法の加重によるロバスト化の考え方を比率推定量に適用し、欠測項目を補定する方法を提案する。

試算には平成 24 年経済センサス - 活動調査のデータを用いる。人為的に欠測を発生させるモンテカルロシミュレーションを行い、2つの比率 (費用総額/売上金額、給与総額/費用総額) を推定し、給与企業の売上高、費用総額及び給与総額について補定を行うことによりこの比率補定法の評価を行う。

なお、補定を行う際は、企業の産業分類などの属性により補定ドメインを設定する必要があるが、このドメインの設定の方法についても検討を行う予定である。

参考文献

De Waal, T., Pannekoek, J., Scholtus, S. (2011) Handbook on Statistical Data Editing and Imputation, Wiley handbooks in survey methodology. Hoboken, New Jersey: John Wiley & Sons.

Ito, T., Abe, Y., and Noro, T. (2013) The Best Stratification to Impute Missing Values of Turnover in Economic Surveys. 59th Session of the ISI World Statistics Congress Proceedings, 25-30, Aug. 2013, Hong Kong, China.

和田 かず美 (2012) 多変量外れ値の検出～繰返し加重最小二乗 (IRLS) 法による欠測値の補定方法～, 統計研究彙報第 69 号, pp. 23-52, 総務省統計研修所