

連鎖方程式による多重代入法と医学研究における応用

統計数理研究所 野間 久史

一般的な統計学の教科書で扱われる統計解析の方法は、原則として「欠測は1つもなく、完全な状態でのデータが得られている」ことを前提としているが、実際の調査・実験研究では、計画されていた通りに完全なデータの測定ができることはほとんどなく、概ね必ずと言ってよいほど、欠測は生じる。これらの欠測は、解析の段階で適切な処理を行わなくては、バイアスや推定精度・検出力の低下をもたらすこととなり、研究の最終的な結論にも重要な影響を及ぼし得るものとなる。欠測データの統計解析の方法は、数理的にも高度で難解なものが多く、標準的な統計ソフトウェアにも、長らく計算モジュールが十分に実装されてこなかった。しかしながら、この数年ほどで、医学分野において、欠測データの取り扱いに関するガイドラインの整備などが進められており、近年、諸科学分野における調査・実験研究の統計解析のプラクティスは急速に変わりつつある。本講演では、2017年の応用統計学会誌に組まれた特集号「医学研究における欠測データの防止と統計解析」を中心として、近年の不完全データの取り扱いに関する統計的諸問題に関する動向を平易に解説する。

また、これらの動向の中、さまざまな応用領域での不完全データの解析のための方法として普及しつつある方法に、連鎖方程式による多重代入法 (multiple imputation by chained equation; MICE) がある。MICE は、欠測が、複数の変数にまたがって、個人ごとに異なるパターンで起こった場合にも、汎用的に用いることができる方法であり、その実践的な有用性から、近年、多くの統計ソフトウェアに実装され、データ解析の実践においても、広く用いられるようになってきている。本講演では、特に、野間 (2017) による解説論文をもとに、MICE の実践的な方法と応用についての解説を行う。

参考文献

野間久史, 五所正彦 (編集) (2017). 応用統計学会誌 特集号「医学研究における欠測データの防止と統計解析」. 応用統計学 46 巻 2 号.

野間久史 (2017). 連鎖方程式による多重代入法. 応用統計学 46, 67-86.