

## 多重代入

多重代入 (MI: multiple imputation) というのはデータの一部が欠損している状態で推定を行うための統計的手法です。通常の Stata コマンドを用いて  $y$  のパラメータを回帰変数  $x_1, x_2, x_3$  で推定する場合、 $y, x_1, x_2, x_3$  のいずれかに欠損値 (missing values) を含む観測データ (observations) は推定で利用されないこととなります。このアプローチは listwise deletion とか casewise deletion と呼ばれます。MI というのはこれらの通常であれば無視されてしまう observations 上の情報も活用しようとする技術ですが、その際、欠損値のパターンは MAR (missing at random) または MCAR (missing completely at random) であることが前提となります。欠損値の生じる確率は観測されたデータに依存することはあっても観測されなかったデータには依存しない場合に、欠損値のパターンは MAR であると言います。さらにその確率が観測されたデータにも依存しない場合には、欠損値のパターンは MCAR であると言います。

MI (多重代入) はその名が示唆するように、欠損値に対して多重に代入を行います。欠損したデータを単に 1 組の値で補完するのではなく、複数組の値で補完します。代入/補完の目的とするところは可能な限り真のデータに近い置換値を生成することではなく、統計的推論が正しく行われるように欠損したデータを取扱うことにあります。

MI による分析は次の 3 つのステップから構成されます。

1. それぞれの欠損値に対して  $M$  個の置換値を生成する。
2. 結果として生成される  $M$  組のデータセットの各々に対し推定対象のモデルをフィットさせる。
3. それら  $M$  組の推定結果を結合し、最終的な単一の推定結果を導く。

MI の機能は `mi` コマンドによって提供されます。それはほとんどの推定コマンドに対して適用できますが、詳細については [MI] `intro` (*mwp-102*) をご参照ください。

■