

実践 “R” 統計学：Dos & Don'ts と一般化線形混合モデル

田中 啓太・森本 元・山口 典之

おかげさまで統計自由集會も三回目となりました。第一回は様々な問題に対処しやすく、かつ無償で配布されている便利な統計ソフト“R”と、“R”によって比較的容易に使えることができるようになった一般化線形モデルという解析方法を紹介しました。前回は一般化線形モデルの基盤ともいえる、従来のt検定や分散分析、回帰分析など、正規分布を用いた様々な統計手法を一つの枠組みとして捉えた一般線形モデル、その一般線形モデルと一般化線形モデルの橋渡しになっている尤度という概念、そして従来の検定とは別のアプローチ、つまり、P値に悩まされる必要のないモデル選択という方法を紹介しました。詳しくはウェブ上で公開されているのでご覧ください (<http://homepage.mac.com/daichis/osjstat/>)。それらを踏まえ、今回はより実践を意識した内容にしていきたいと思えます。

1. 線形モデルにおける Dos & Don'ts

生態学者は、汚いデータの解析を迫られます。「汚い」とは例えば、(1) 正規分布ではなさそうな上に、不等分散を示もの、(2) 単位あたりの個体数や回数(頻度)を測るときに、各データで「分母」が異なるもの(田圃あたりの鳥の個体数という比データで、観察した田圃面積が様々であるとき等)、(3) 同じ個体やグループから何度か計測したものが混在するもの、が挙げられます。このようなデータを、どのように解析するのが適切なのでしょうか。ここでは、これまで、そして現在も良く行われている対処法の利点と欠点を紹介します。

2. 一般化線形混合モデル (GLMM) 概説

一般化線形モデルの発展版ともいえる、一般化線形混合モデルを紹介します。これは解析の目的である変数を母数効果(固定効果)、それ以外の興味はないが無視できない変数を変量効果(ランダム効果)として扱い、擬似反復を避けつつパラメータを推定する方法です。特に個体内での反復を多用する我々のようなフィールド研究者にとってはより実用的と言える解析方法といえるでしょう。

3. GLMM 実践例

鳥学会に所属する若手に苦労話も交えつつ実践例を紹介してもらう予定です。