GMANOVA モデルにおける新たな推定方法と解釈

中京大学 教養教育研究院 永井 勇

n 個の各個体に対して, 全ての個体で測定時点を揃えて p 回測定して得られる経時測定データの分析を本講演では考える．この分析の目的は, データに隠れている経時変動を上手く捉えることある．このようなデータの分析の際には, Pothoff and Roy (1964) で提案された次の一般化多変量分散分析 (GMANOVA) モデルが使われる．

$$\boldsymbol{Y} = \boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' + \boldsymbol{A} \Xi \boldsymbol{X}' + \boldsymbol{\mathcal{E}}, \tag{1}$$

ここで, $\boldsymbol{1}_n$ は n 次元の全てが 1 からなるベクトル, $\boldsymbol{0}_r$ は r 次元の全てが 0 からなるベクトル, $\boldsymbol{Y}$ は各行が各個体で測定して得られる経時測定データからなる $n \times p$ 行列, $\boldsymbol{A}$ は各個体の特徴を表す測定時点に無関係な k 個の変数からなる $\mathrm{rank}(\boldsymbol{A}) = k$ の $n \times k$ 行列とし, $\boldsymbol{A}'\boldsymbol{1}_n = \boldsymbol{0}_k$ (各説明変数で中心化されている) を満たしているとし, $\boldsymbol{X}$ は後述のように各行が測定時点の関数からなる $p \times q$ 行列であり, これらは既知である．また, $\boldsymbol{\mu}$ は q 次元未知ベクトル, Ξ は $k \times q$ 未知行列であり, $\boldsymbol{\mathcal{E}}$ は $E[\boldsymbol{\mathcal{E}}] = \boldsymbol{0}_n \boldsymbol{0}_p'$, $\mathrm{Cov}[\mathrm{vec}(\boldsymbol{\mathcal{E}})] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$ の $n \times p$ 誤差行列とし, $\boldsymbol{\Sigma}$ は正則な $p \times p$ 未知行列とする．このモデルにおいて, $E[\boldsymbol{Y}]$ の部分が経時変動に対応している．

ここで測定時点を $t_1 < t_2 < \cdots < t_p$ とすると, $\boldsymbol{X}$ の i 列目を $(t_i^0, t_i^1, \ldots, t_i^{q-1})$ とすることは, 経時変動を測定時点の $(q-1)$ 次多項式で推定することに対応する．また, この部分を適当な柔軟な関数にすることは, Nagai (2011) にあるように柔軟な関数の重み付き和のような形で経時変動を推定していることに対応する (Nagai (2011) では柔軟な関数を用いた場合に起きる問題点の回避についての推定手法も提案している).

このモデル (1) において, 未知の $\boldsymbol{\mu}$ と Ξ の推定としてよく使われる推定量 $\hat{\boldsymbol{\mu}}$ と $\hat{\Xi}$ は, 次のリスクを最小にする $\boldsymbol{\mu}$ と Ξ を求めることで得られる;

$$\mathrm{tr}\left\{ \left(\boldsymbol{Y} - \boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' - \boldsymbol{A} \Xi \boldsymbol{X}' \right) \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{Y} - \boldsymbol{1}_n \boldsymbol{\mu}' \boldsymbol{X}' - \boldsymbol{A} \Xi \boldsymbol{X}' \right)' \right\}. \tag{2}$$

実際に求めると, それぞれ $(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})\hat{\boldsymbol{\mu}} = \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}'\boldsymbol{1}_n/n$, $\hat{\Xi}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ の解となる．これらの解を得るために, モデル (1) において $\mathrm{rank}(\boldsymbol{X}) = q$ を仮定することが多い．この仮定は $\boldsymbol{X}$ に用いる関数を制限していることとも考えられる．

一方で, $\mathrm{rank}(\boldsymbol{X}) < q$ の場合, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$ に逆行列が存在しないため, この推定量は得られない． そこで本講演では, 永井 (2021, 2022) のアイデアと同様に, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_r)'$, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_r)$, $\Xi = (\Xi_1, \ldots, \Xi_r)$ として, モデル (1) を次のように書き換える;

$$\boldsymbol{Y} = \sum_{i=1}^{r} \boldsymbol{1}_n \boldsymbol{\mu}_i' \boldsymbol{X}_i' + \sum_{i=1}^{r} \boldsymbol{A} \Xi_i \boldsymbol{X}_i' + \boldsymbol{\mathcal{E}}.$$

このように書き換えたモデルにおいても，リスク (2) と同様のリスクを最小にするような $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_r$ や $\Xi_1, \ldots, \Xi_r$ を求めて並べれば, $\mathrm{rank}(\boldsymbol{X}) < q$ でも $\boldsymbol{\mu}$ および Ξ の推定量が得られる．また, 提案する推定量の解釈などについても講演では触れる予定である.

引用文献:

[1] Pothoff, R. F. & Roy, S. N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313–326.

[2] Nagai, I. (2011) Modified C_p criterion for optimizing ridge and smooth parameters in the MGR Estimator for the nonparametric GMANOVA model. *Open Journal of Statistics*, **1**, 1–14.

[3] 永井 勇 (2021) 高次元小標本における多変量線形回帰モデルでの推定法, 2021 年度統計関連学会連合大会

[4] 永井 勇 (2022) 説明変数がランク落ちしている状況での多変量線形回帰における不偏推定量, 2022 年度統計関連学会連合大会