

階層性のある多重比較とメタボロームデータに対する応用

シミック株式会社 データサイエンス事業本部 統計解析第1部 小野 あかり

慶應義塾大学 理工学部 数理科学科 南 美穂子

慶應義塾大学 医学部 衛生学公衆衛生学教室 竹内 文乃

慶應義塾大学 医学部 衛生学公衆衛生学教室 原田 成

1 はじめに

メタボロームとは、生体内の細胞や組織において代謝に伴い生成されたり使われたりする物質である。メタボローム間には複雑な関係があり、また曝露とメタボロームの双方に影響する共変量の存在が考えられる。そのため、メタボロームデータ解析では、本来メタボローム構造を考慮した解析や慎重な共変量調整が必要である。しかし、実際はそのような解析は未だ行われていないようであり、先行研究には (1) 共変量の調整方法が最適であるか議論されていない、(2) 検定の検出力が劣る、(3) メタボローム構造を考慮していない、という問題がある。そこで、以上の (1), (2), (3) を改善した解析手法を提案する。

2 解析手法

本講演では、慶應義塾大学が山形県鶴岡市で実施しているコホート調査「鶴岡みらい健康調査」のベースラインデータをもとに、現在の喫煙習慣が TCA 回路 (Tricarboxylic acid cycle) 上の Citrate, Isocitrate, Succinate, Fumarate, Malate の順で並ぶ 5 つのメタボロームに及ぼす効果を解析した結果を考察する。上記のデータから既往症やアンケートの回答の矛盾、欠測のある被験者を除いた 5,870 人のデータと、そのうち男性 2,798 人のデータに対し、それぞれ次の解析を行い結果を比較する。

推定 1: 共変量を調整をせずメタボロームデータを現喫煙群と非現喫煙群に分け、両側 Welch 検定統計量と p 値を計算する。

推定 2: 各メタボロームを目的変数、喫煙習慣の変数と共変量を説明変数とした線形回帰モデルを当てはめる。ただしモデルに入れる共変量は、性別等 22 種類の共変量候補から AIC を規準としたステップワイズ法で選択する。

推定 3: 傾向スコアにより共変量調整を行い、喫煙習慣による各メタボロームへの平均処置効果を推定する。推定方法として、(i)IPW 推定、(ii) サブクラスの幅を 0.1 に固定した層別解析、(iii) 全サブクラスのサイズが等しく、かつ現喫煙者と非現喫煙者が共に 1 人以上含まれるような最大層数の層別解析 (full subclassification, [2] Wang, L. (2016)) の 3 つの手法を用いる。

検定: 推定 1, 2, 3 の推定結果をもとに、各メタボロームの値に対する喫煙習慣の効果が有意であるか検定する。多重比較には、Bonferroni 法と階層手順を使用し、 $\text{FWER} \leq 0.05$ で検定する。

参考文献

- [1] Wang, L. (2016) "Robust Estimation of Propensity Score Weights via Subclassification"
<https://arxiv.org/pdf/1602.06366.pdf>