

数量化Ⅲ類によるデータ解析について

— 国民性と価値観国際比較データへの適用を中心に —

東洋英和女学院大学 林 文
帝京大学大学院公衆衛生学研究科 山岡 和枝

1. 数量化Ⅲ類

数量化Ⅲ類と言われる方法は、後に開発された、対応分析(多重対応分析)と同じ分析手法であり、既存のいくつかの分析ツールによって、標準化の仕方等々により解の値そのものは異なることがあるが、本質的な解の構造としては同じである。林知己夫が数量化理論を発表してから60年を経て、カテゴリカルデータに対する分析の研究発展、コンピューターの格段の進歩による分析ツールの開発に伴って、簡単に使えるようになった。数量化Ⅲ類は手法としては、そうした分析の基本的な方法に位置づけられるとも言えるが、数量化理論は、手法としてだけでなく、その根底をなすデータに対する考え方そのものを含むものであり、解法は単純である。発展した解法では、出力された解がそのまま解釈できるような仕組みを取り入れて、誰もが使いやすいことを目指しており、普及に繋がっているのは確かであろう。林の数量化理論は、そうした仕組みを取り入れるよりも、データへの適応とその解釈について、深く考えることを求めていると考えている。

2. 数量化Ⅲ類の適用上のいくつかの課題

林知己夫は、統計数理研究所で1953年から始まった日本人の国民性調査や1970年からの国際比較調査のデータを、数量化理論、とくに数量化Ⅲ類を用いて解析してきた。著者らはその考え方に習い、調査データの解析に数量化Ⅲ類を用いている。その中で、反省を含め、さまざまな課題が提起される。

- ・適用するデータは適切か
- ・該当数の少ないカテゴリーの扱い
- ・ボンドサンプルの解析の意味
- ・取り扱う次元の問題と構造の安定性

適用するデータは、該当数の少ないカテゴリーや欠損値の問題も含むが、取り上げる質問項目群については、回答カテゴリーの単純集計、質問相互間のクロス集計の検討の上に、どのような関連を把握しようとしているかについて検討した上で選択的に取り上げることが適切な解析につながる。なぜなら項目間(カテゴリー間)の関連性は、あくまでもとりあげた次元の上での関連性であり、項目が異なればその構造も異なってくるからである。内容においても、仮説検証型に慣れた研究者は、関連の深い質問項目群を取り上げがちであるが、むしろ、大きな構造として捉えることが数量化理論の真髄ともいえ、関連の有無も含めて大局を捉えることに意味が見出せることも少なくない。

該当数の少ないカテゴリーについては、解が飛び離れた値となるため、除外して分析することが多いが、関連性の構造によっては、必ずしも飛び離れるとは限らない。飛び離れること自体がデータの性質であることは把握する必要があり、それを解釈の一つとして生かす方がよいこともある。

ボンドサンプルの解析は、国際比較において、それぞれの国・地域ごとの構造を個別の社会集団の構造として捉えた上で、それらを統合したボンドサンプルについて、国・地域の位置づけを捉えるために行うことがある。しかし、統合する各国・地域ごとの重みをどうするか、国・地域を統合した集団の社会集団としての構造に意味があるのか、その点についての考え方抜きにしては意味がないともいえよう。これらの問題は、データへの適用以前のデータの持つ背景やデータ収集の方法等にも依存する課題といえる。

3. まとめ

上記の課題は機械的に解決できるものではなく、適用には試行錯誤が必要ということであろう。誰でも簡単に使えるような分析ツールがあっても同様であり、データそのものを眺め、疑問を持ち悩みながら解析を進める探索的な解析の課程は重要である。