

非計量データに対する多変量解析法の新展開

中村 裕子

実験や調査などさまざまな場面で観測されるデータは、名義尺度や順序尺度から成る「非計量データ」と間隔尺度や比尺度から成る「計量データ」に大別される。非計量データは、計量データと同様数値を用いて表現されてはいるものの、その数値は実験者もしくは調査作成者が恣意的に割り当てたものであるため、その数値の間隔が意味を成さない、平均値や分散が何を意味するか明確ではない、非線形の関係のある変数の扱いが困難であるといった問題があり、そのため非計量データを計量データと同様に扱って分析を行うことは不適切であるという見方がある。

上記のような問題を回避するため考案された手法が、恣意的に割り当てられた数値の代わりに新たに最適な数値、すなわち数量化得点を計算により導出するというものであり、「数量化法」と呼ばれる。

本論文では、非計量データに対する数量化得点の算出を伴う多変量解析法を新たに2つ提案した。

1つめの手法は、三相非計量データに対する主成分分析法である。三相データとは、複数被験者に対し複数銘柄の商品を提示し、それぞれについて複数の質問項目に答えてもらうといったような三者関係から成るデータである。この例の場合、被験者、銘柄、質問項目の三者関係からデータが成り立っている。主にマーケティングや心理実験等の分野では、このデータが非計量データであることが珍しくない。本論文ではこのような三相非計量データが得られた場合に、数量化得点の算出と主成分分析を同時に行う手法を提案した。本手法はTUCKER3と呼ばれる三相の計量データに対する主成分分析法を拡張したものと見なすこともできる。また、飲料に対する印象評定データへの適用例からも、提案手法の有用性が示された。

2つめの手法は、非計量データに対する判別分析法である。判別分析法とは、個体の所属群が既知である多変量データに対し、各個体の説明変数の値を用いて所属群を判別するための基準を導出し、所属群が未知である個体が現れたときにその判別基準を用いていずれかの群に判別する方法である。本論文では説明変数が非計量データとして得られた場合の判別分析法を提案した。なお、判別分析法の定式化はいくつか存在するが、ここでは線形回帰分析の枠組による定式化を行った。

さて、現在においてもしばしば用いられている手法として、「順序尺度を間隔尺度とみなして計量データと同様の分析を行う」というものがある。そこで本論文では、薬剤効果の判定を行った全て順序尺度から成る非計量データに対して、提案手法を適用した結果と従来法である「順序尺度を間隔尺度とみなして計量的に分析」を行った結果とを比較した。正しく判別できた個体数を比較した結果、提案手法の方が従来法よりも優れていることが示された。個体を2次元空間上にプロットした図(図1)からも、提案手法の方が従来法よりも群ごとのまとまりが良く、群間がよく分離していることが示唆された。以上より、非計量データに対して数量化得点を算出することの有用性が示された。

さて、本論文で提案した2手法において算出した数量化得点は、あくまで限られたサンプルから得られたものであり、サンプルが異なれば得点もある程度変動することも忘れてはならない。とはいえ数量化得点の算出によってカテゴリ間の非線形の関係を取り出すこともでき、提案手法は意義深いものといえよう。ここで、提案手法により算出された数量化得点のプロット例を示す(図2)。ここで示したプロットは、提案手法の一つである「非計量判別分析法」を薬剤効果の判定データに適用した際に得られた数量化得点であり、「副作用」に関する4カテゴリの関連性が2次元平面上に図示されている。プロットを見ると、横軸方

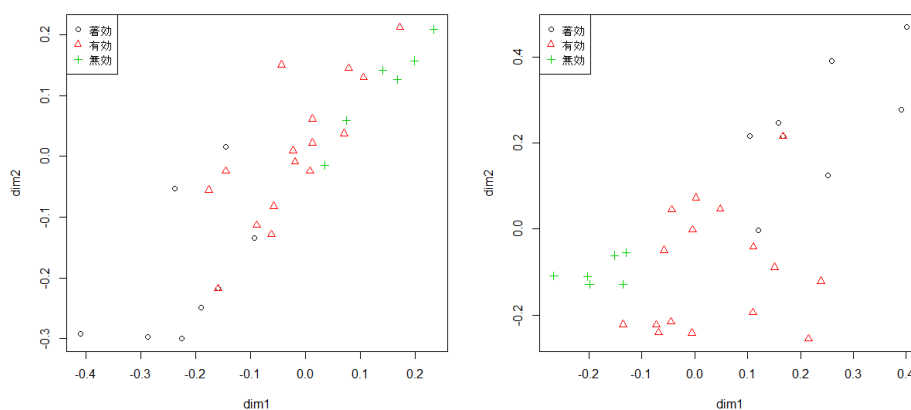


図 1:従来法の個体プロット(左)と提案手法の個体プロット(右)

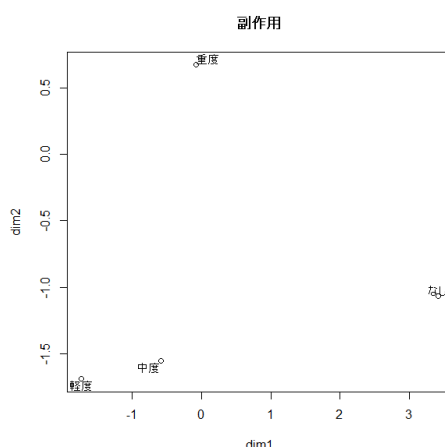


図 2:2次元数量化得点プロット例

向は「副作用の有無」、縦軸方向は「副作用の強さ(正の方向ほど副作用が強い)」を表していることが読み取れる. このように, 数量化得点の算出を行うことによって, 単にカテゴリに対して「1:なし, 2:軽度, 3:中度, 4:重度」という数値を機械的に割り当てたのでは得られない知見を得ることができる.

また, 提案手法は各変数ごとに数量化得点の算出を行うため, さまざまな尺度の変数が混在している場合にも容易に適用可能であるというメリットもある. また, 変数ごとに異なる数量化次元数を設定できるため, 本論文で提案したモデルは柔軟性の高いモデルと言える. 数量化次元数については, 理論的・数学的に最適な次元数を決定するための基準が未だ確立されていないという問題があり, 今後さらなる研究が望まれる. 現状においては, カテゴリ間の関係性あるいは主成分負荷量・判別係数のいずれに興味を持つか, または変数の尺度(名義尺度の場合, 2次元以上の数量化を行うのが望ましい場合が多い), 解釈のしやすさといった事情を鑑みて, 数量化次元数を決定すべきであろう. (行動統計科学)