

多変量都道府県データの比較分析 —スパース主成分分析によるアプローチ—

山下 かおり

本稿における研究目的は、都道府県の特徴抽出と、高次元データへのスパース主成分分析の適用である。近年、増加する訪日客や加速する人口減少など、日本国内における地域のあり方が問われている。地域活性化のためには、各地域の特徴や独自性を把握し、個別的な対応をすることが重要である。一方、統計解析の分野では、近年、遺伝子発現解析や情報工学への適用ニーズが高まり、高次元データの次元縮約が求められている。これらの背景より、本稿では都道府県に関する 441 次元のデータを対象に Unpenalized Sparse Loading Principal Component Analysis (以下、USLPCA) を行い、その結果、抽出された特徴や独自性を解釈し、課題について考察した。USLPCA は、Adachi & Trendafilov (2016) によって提唱された手法である。

USLPCA によって、441 変数は 15 の主成分に分類された。スパース化前に 80.51%であった累積寄与率は、スパース化後には 64.06%となったが、主成分負荷量行列 \mathbf{A} は、6615 ある要素のうち 5602 要素がゼロに変換され、解釈の煩雑さは大いに解消されたと言える。第 1 主成分は全体の 30%以上を説明しており、その特徴から『少子高齢化特性』と命名された。第 2 主成分以降の説明率は全て 10%未満であり、上位の主成分から『生活困窮者割合』、『警察コスト・教育費(一人当たり)』、『多子化特性』、『短命傾向』、『寒冷地域特性』、『総面積』、『被災影響度』、『第二次産業』、『出身県内就職傾向』、『学習機会・安全』、『患者の病床利用率』、『高収入傾向』、『前年度比経済成長度』、『乳児・幼児死亡率』と命名された。主成分が意味する内容は、経済状況や人口形態、健康状態を示唆するものが多く抽出される傾向があった。

エリアごとに抽出される特徴として、北海道・東北地域では、寒冷な気候特性と東日本大震災の影響、そして短命な傾向が浮き彫りとなった。関東地方では、少子高齢化の進行が比較的遅く、東京都に隣接する県では出身県内で就職しない傾向が明らかになった。中部地方は生活困窮者が全国的に最も少なく、近畿地方は、滋賀県を除く 6 府県では少子化の傾向が強い一方、滋賀県のみが多子化傾向を示すことがわかった。中国地方は総じて、第二次産業が盛んであり、四国地方では、特に少子化傾向が示された。九州・沖縄地方では、生活困窮者割合が比較的高い傾向が示された。(行動統計科学)