

テーマ：Rと自己組織化マップによる県民性の分類

名列番号：039 氏名：中村真吾

1. 研究目的

現在、様々な分野で統計解析による分析・予測がなされている。そこで、統計解析がどのようなものかを実際に解析してみることで学ぶ。テーマは以前、県民性の分類についての書籍を読んだことがあるため、そういうことができるものなのかと疑問に思い決定した。

2. 研究手順

8週間ほど R-tips という資料を使って輪講を行い、その後各自のテーマについて研究した。

3. Rとは

「R」とは統計計算とグラフィックスのための言語またはその環境のことで多様な統計手法を可能とする。商用の S 言語・環境をオープンソース化したものであり、無料で導入が可能である。

4. 解析に用いたデータ

下記の URL からダウンロードしたデータを整形して解析した。

<http://www2k.biglobe.ne.jp/~yano/KENMIN.html>

データは都道府県別に6つの特徴量を調べた結果で、各県はそれぞれ「美人度数（得票数）」、「男性の平均結婚年齢」、「女性の平均結婚年齢」、「世帯当たり預貯金（万円）」、「人口当たり所得（万円）」「離婚率」に関する数値を持つ。

都道府県は47個なので、この行列の数値部分は47行6列のデータを持つ。

5. 自己組織化マップについて

特徴量が多数でも、必ず二次元のマップに描画してくれる。マップの大きさを変えると、1マスの中に入る要素数（今回は都道府県の数）が変わる。今回はマップの大きさを6マス×6マスに、マップの種類を四角形（rect）に設定した。

6. 結果

分類した結果を図1に示す。マスの中の折れ線グラフは、そのマスに配置された都道府県の特徴量6個（一番左が美人度、一番右が離婚率）の分布を表している。縦線で範囲が示してある場合は、その特徴量に関してばらつきがあったことを示している。

7. 考察

東京を含む大都市圏は、世帯当たり預貯金にばらつきがあるものの、他の値は全て高かった。マップの上下と離婚率が関係しており、右上の大都市圏と左上の九州および北海道では離婚率が高かった。対照的に、北陸を含む中央下から左下にかけては離婚率が低く、世帯当たり預貯金が高いことがわかった。

8. 今後の課題

時間が足りず、細かい分析まで辿り着けなかったことが残念。

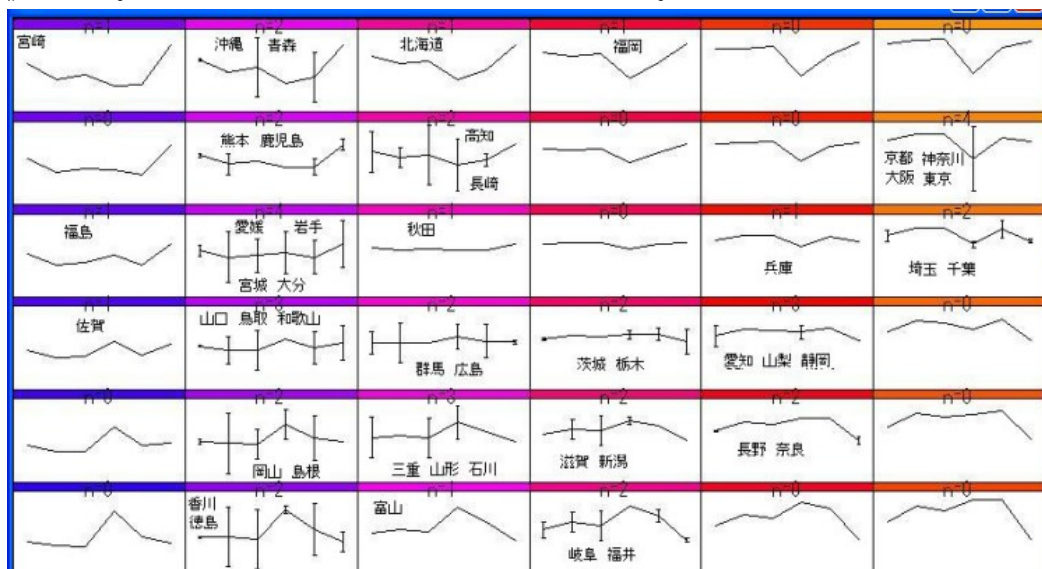


図1. 自己組織化マップを用いた県民性の分類結果