

統計だより

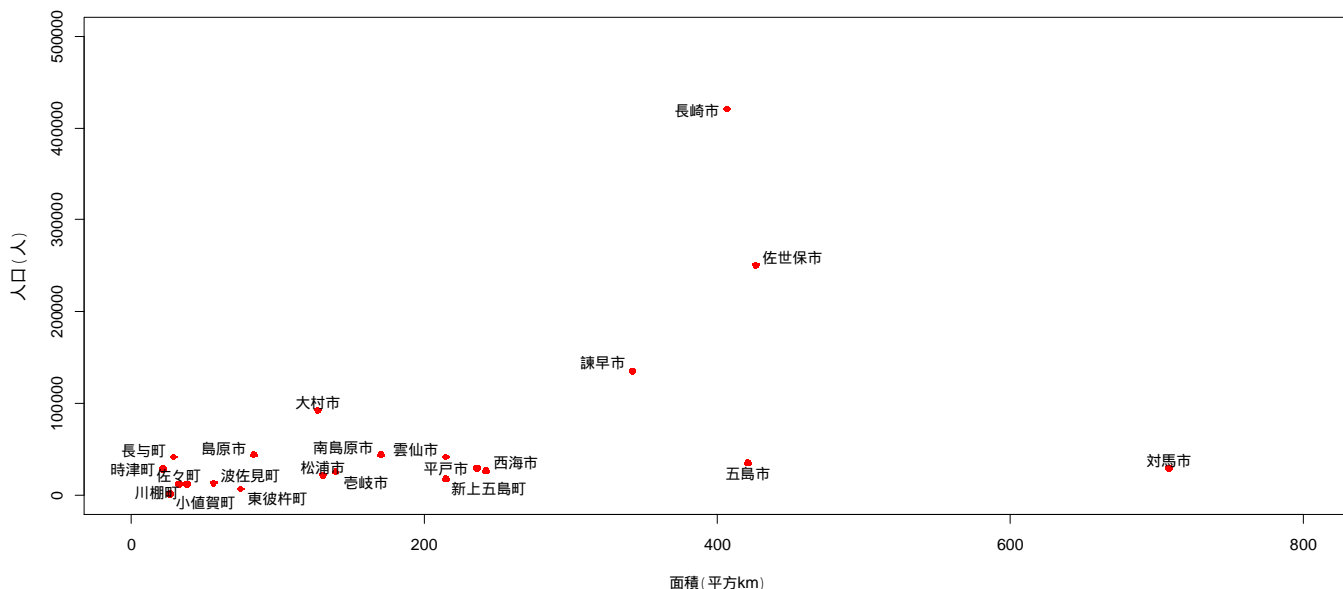
統計データから長崎県内21市町のグループ分け

長崎県県民生活部統計課 商工勤労統計班 山口 雄介

長崎県には13市と8町があります。今回は各市町のデータを利用して、人口構成や産業構造が似ているグループ分けに挑戦しようと思います。

図1は横軸に面積、縦軸に人口をプロットしたもの(注1)です。県内最大の面積を誇る対馬市が右下にあり、真ん中上部に人口が40万人を超える長崎市があり、佐世保市、諫早市と続きます。しかし、これでは多くの市町が左下に固まっていて、グループ分けするのなかなか難しそうです。また、面積と人口だけの2項目しか使わないのも、グループ分けするには、あまりにも単純すぎる気がします。

図1 散布図(長崎21市町)



そこで、項目数を増やしてみます。人口構成や産業構造に関する次の10項目を利用します。

総面積	総人口	生産年齢人口割合	市町民税収入額/人	就業率
第1次産業就業者割合	第2次(同)	第3次(同)	総農家数	持ち家比率

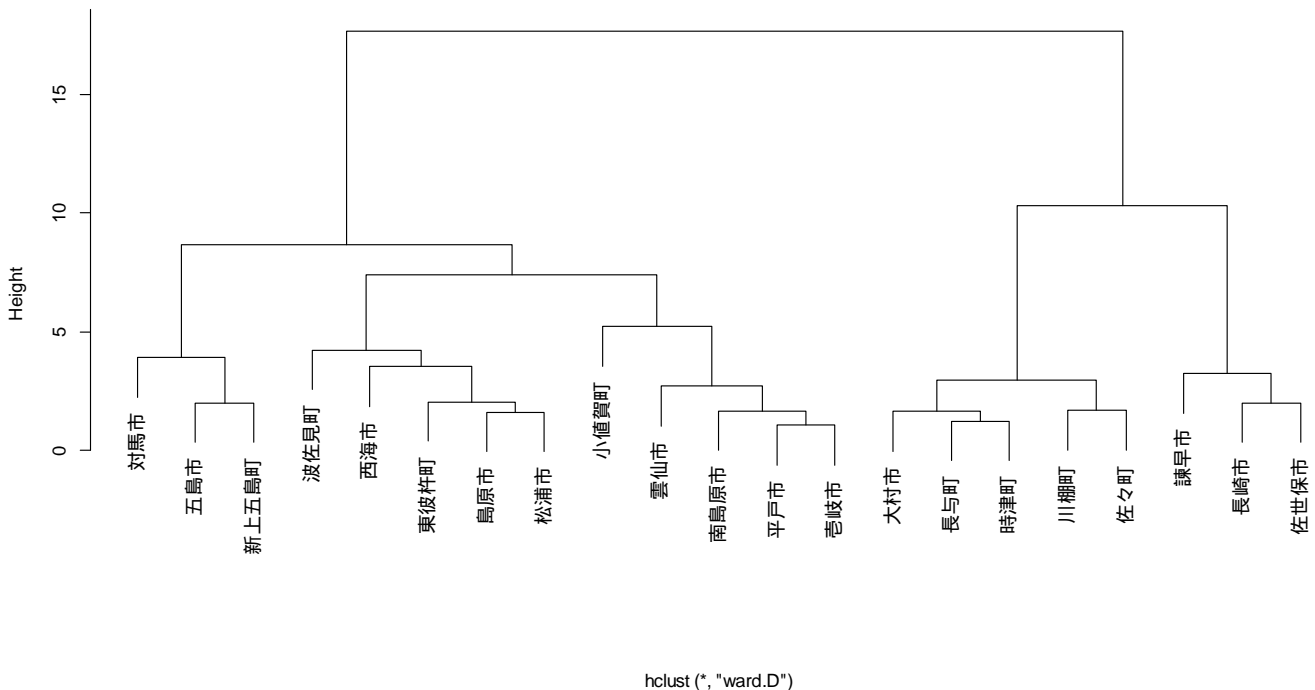
注2 長崎県統計課のホームページ参照 ~ は2017年、 ~ は2016年、 ~ は2015年実績

2項目では横軸(面積)と縦軸(人口)の散布図にできましたが、当然ながら10項目ではできません。そこで登場するのがデータ分析手法です。

ここではデータ分析手法の一つであるクラスター分析を利用します。計算は複雑ですが、2項目で散布図を描いたときに対馬市が右下に離れていて、長崎市が真ん中の上部にあった場合と同じように、10個の項目について21市町同士の似ている位置関係を示すことができます。理屈よりも実際にみてもらった方が早いかもしれません。クラスター分析を行い作成したのが図2のデンドログラムです。日本語では樹

形図と呼ばれます。上下ひっくり返すと枝分かれして、木が茂っているように見えますよね？

図2 デンドログラム(長崎21市町)



デンドログラムの説明ですが、似ている市町同士は低い位置で線がつながっています。[五島市-新上五島町]、[長与町-時津町]、[島原市-松浦市]などは人口構成や産業構造が似ていることがわかります。21市町が似ている順に線につながっていき、最終的には一番上で21市町が全てつながるわけです。

デンドログラムは客観的なデータから得られた証拠であり、強固な説得力を持った資料になります。

例えば、行政職員の立場で考えてみてください。長与町では効果的だった住民サービスが、デンドログラムで真横にある時津町では同様の施策を実施する場合、同様の効果を得られる可能性は高いと考えられます。(注3)

EBPM(証拠に基づく政策立案)という言葉を目にするところがあるかもしれません。以前では「～と考えられる」、「～と推計される」としかできなかった部分が、入手できるデータが膨大になり、多くの統計分析手法が使えるようになったことで、ゆるぎない「データ」という証拠をもって政策立案がされるようになりました。統計調査等によって得られる「データ」が今後より一層政策に反映することが可能になると考えられます。

注1 グラフの作成、データ分析は「オープンソース・フリーソフトウェアR」を利用しています。

注2 参考HP：長崎県統計課ホームページ 長崎100の指標 較べてみれば-

<https://www.pref.nagasaki.jp/bunrui/kenseijoho/toukeijoho/kankoubutsu/100shihyou/>

注3 利用した10項目で行った結果です。実際には施策に応じて項目も選択する必要があります。

参考文献 杉原敏夫・藤田渉著「経済の情報と数理 多変量解析」

舟尾暢男著「The R Tips 第2版: データ解析環境Rの基本技・グラフィックス活用集」

.....
(編集後記)

平成の大合併によって21市町になりましたが、私が小学生のころは「上県郡」や「南高来郡」を読めることを自慢する男子がいました。今ではなくなってしまったこのような地域名を令和時代の小学生は読めたりするのでしょうか。