

都道府県を一文字で識別する方法

黒田 昭

(山形大学農学部農地造成学研究室)

Display Method of Data Analysis on unit of Prefecture

Akira KURODA

Laboratory of Reclamation and Melioration, Faculty of Agriculture,
Yamagata University, Tsuruoka 997, Japan

(Received September 30, 1989)

Summary

This study is concerned to the method how to display for data analysis on unit of prefecture. In Japan, there are forty-seven prefectures, that are constituted by the usually forty-three prefectures so-called "ken" and the four special prefectures which are called by Tokyo-to, Hokkai-do, Osaka and Kyoto-fu.

We desire to display one significant symbol at two dimensional plan to indicate on these prefectures, that are analyzed by the method of correlation analysis or multivariate statistical analysis. Since we have ideographic symbols so-called "kanji" to differ phonetic symbols, we can use kanji to indicate each prefecture. In Fig. 2, the results on this study are shown.

Keywords : prefecture ; display method ; ideographic symbol ; statistical analysis.

I. ま え が き

地域の分析単位として都道府県をとる場合が多い。この分析結果を図示するとき、棒グラフや円グラフのような本質的に1次元的(1変数的)な表示の場合は、都道府県名をそのまま略さずに用いればよいが、2次元的な表示たとえば2種類のデータの相関分析とか、主成分分析の結果(第1主成分と第2主成分の2変数、2次元座標)を表わしたいとき、都道府県を代表する一文字を点代わりとして用いたくなる。

このようにすれば、ただ点だけの分布図とか、都道府県を数字^{a)}で識別した図よりも、認識能を高めることができるからである。

たとえば、アメリカ合衆国などではアルファベットの2文字を用いて、50州(とワシントン D.C.)を識別することが確立しており、郵便^{b)}のあて先などに広く用いられている。

ところで都道府県名には同じ文字を用いている場合が少なくなく、たとえば福岡県と福島県などの場合に困ってしまうわけである。従来までは^{c)}福岡県の方に「福」を用いたら、福島県の方は丸で囲って「福」を用いて、欄外に(注)を付記するなどしてなんとか凌いできたがどうもすっきりしない。

そのうえ、第III章で述べるように識別文字をいろいろな図形で囲って、第3次元的、あるいは第4次元的な意味あいをもたせようと思ってもできないことになる。

そこで、47都道府県名を文字一文字で識別できる方法、そして規範的な規則はないかと考え、都道府県名を文字の相互関係として分析した結果、表示可能な方法を見出したので報告したい。

II. 都道府県名の相互関係分析

1. 分析の前提

47都道府県名の延べ使用文字数は北海道^{d)}と神奈川、和歌山、鹿児島^{e)}の3県が3文字であるから98文字となる。この98文字をもって構成される集団を一つの集合と考え、文字である集合因子の間の相互関係を分析する。98

文字がすべて異なる文字（単因子）から構成されていればなら問題は無いが、表1に示すように15の文字が重因子として、延べ40回にわたって重複使用されている。

表1 47都道府県名で重複して使用される文字

文字	回数	文字	回数
山	6	京	2
島	5	宮	2
川	3	奈	2
岡	3	崎	2
福	3	賀	2
大	2	城	2
知	2	愛	2
長	2	合計	40

したがって、47都道府県名は73文字 (73=98-40+15) の文字種より構成されていることから、都道府県のそれぞれを異なる一文字で識別表示しうる可能性はあることになる。

また、重複使用される文字（これを重字という）は15文字であるから、1回しか使用されない文字（これを単字という）は58文字となる。

ちなみに、使用頻度の多い文字は「山」が6回で「島」が5回である。これからも、日本国は山がちな島国であるということが如実に示されているといえようか。

2. Aランク

都道府県名がすべて単字で構成されるものは表2に示すように15道県である。これらは道県名のどの一文字を使用しても識別が可能となるからこの意味において、Aランクの識別能力があるといえる。

これらは最初の一文字（初字という）を識別文字として使用するのが、最も認識され易いと思われる。しかし、このことは初字以外の文字（次字という）の識別文字としての使用を制限するものではない。

たとえば、千葉県などは初字の「千」とするより、次字の「葉」とした方が、慣用にあっているかもしれないからである。

しかし、本論ではAランクの識別文字としてすべて初字を用いることとした。

このことから、次の規則を得る。

規則1. 単字のみから構成される名前では原則として初字を識別文字とする。

このAランクの15都道府県によって31単字が使用（消

表2

Aランク
(北)海道
(青)森
(岩)手
(秋)田
(栃)木
(群)木
(埼)玉
(千)葉
(新)潟
(岐)阜
(三)重
(兵)庫
(鳥)取
(熊)本
(沖)縄

表3

Bランク
(茨)城
(東)京
(神)奈川
(富)山
(石)川
(静)岡
(滋)賀
(和)歌山
(広)島
(徳)島
(香)川
(高)知
(佐)賀
(鹿)児島

表4

Cランク
大 (阪)
大 (分)
愛 (媛)
長 (野)
島 (根)
福 (井)
山 (形)
山 (梨)
山 (口)
(京)都
(奈)良

表5

Dランク
(愛)知
(長)崎
(宮)崎
宮 (城)
(岡)山
(福)岡
福 (島)

費）されたから、残る27単字と15重字で残された32都府県を識別しなければならないことになる。

3. Bランク

残る32都府県名のうち、初字は単字であるが、次字は重字であるものは表3に示すようにBランクの14都県がある。これらも初字の一文字を用いれば識別可能となる。Bランクの次字の相互関係を中心に図示したものが図1-a, b, c である。図1-c のなかの奈良県の場合はCランクであるから次節で述べる。

規則2. 初字が単字の場合には、初字を識別文字とする。

このBランクの14都県によって16単字が使用（消費）され、そのうえ「川」と「賀」の2重字がCランク以降に関係しないから、残る18府県を11単字と13重字で識別しなければならない。

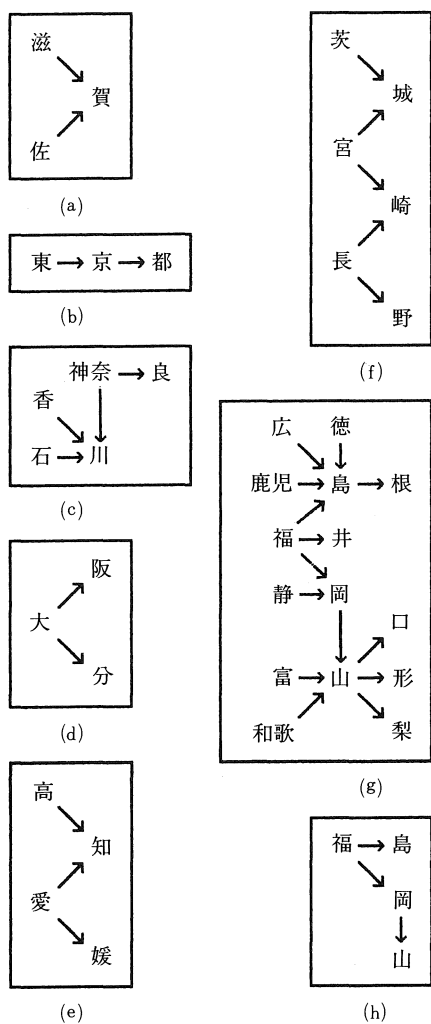
4. Cランク

次に、初字は重字であるが次字が単字の場合である。表4のCランクの11府県がそれである。

この場合には原則として次字を識別文字として使用すればよい。Cランクの府県名に関係する文字の相互関係は図1-c から図1-g に示されている。

ここで、京都府と奈良県についてさらに分析しておきたい。Cランクのなかでこの2府県は、初字が重字とい

図1 都道府県名の文字の相互関係



と混同されることはないであろうし、京都府の識別文字として「都」でなく「京」を用いても東京都と混同されることはないであろう。したがって次の規則を得る。

規則3. 初字は重字であるが次字が単字の場合は、次字を識別文字とする。ただし、初字が他の名前前の次字とのみ重字の場合は初字を識別文字として用いることもできる。

このCランクの11府県によって11単字が使用（消費）され、「京」、「大」、「奈」の重字が次のDランクに関係しないから、残る7県を10重字で識別しなければならないこととなるが、県数より文字数の方が多いから、その可能性は残されているといえる。

5. Dランク

最後まで残ったDランクの7県は表5に示されるとおりであり、県名のすべての文字が重字のみから構成され、延べ14字の10重字である。この7県名の使用されている文字の相互関係は3グループに分かれる。グループ1は図1-eの愛知県のみ、グループ2は図1-fの長崎、宮崎、宮城の3県、グループ3は図1-hの福島、福岡、岡山の3県である。この図1-hは図1-gのうちDランクに関係のあるものだけを抜き出して表示したものである。

グループ1の愛知県の「愛」も「知」も重字であるがまだ識別文字としては使用されていないから、どちらを使ってもよい。このようなときは使用頻度の少ない方を採用する。両字とも使用頻度は2回であるから、このような場合にはやはり初字の方を優先して用いたい。「愛」をみて、これは愛媛県か愛知県を表示していると推定されるし、「媛」が別にあることから愛知県と判断されるであろう。

グループ2のなかの長崎県も同様の理由で初字「長」の方が識別文字として使用できる。

グループ3の岡山県も同様な理由で初字「岡」を使用すれば、福岡県は「福」としなければならず、残る福島県は「島」としなければならない。「島」をみれば島根県と推定するのが普通であるから、この識別文字には欄外に(注)が必要である。

最後の最後まで残った宮崎県と宮城県では識別文字として未使用の「宮」、「崎」、「城」の3文字で2県を識別すればよいから、その方法は3通りある。使用頻度も同じであるからどれを用いてもよい。このような場合には、次字からではなかなか都道府県名が推定されがたいし、また字体の条件からいえば「崎」は埼玉県の「埼」と字面が似ているからなるべく避けたい。

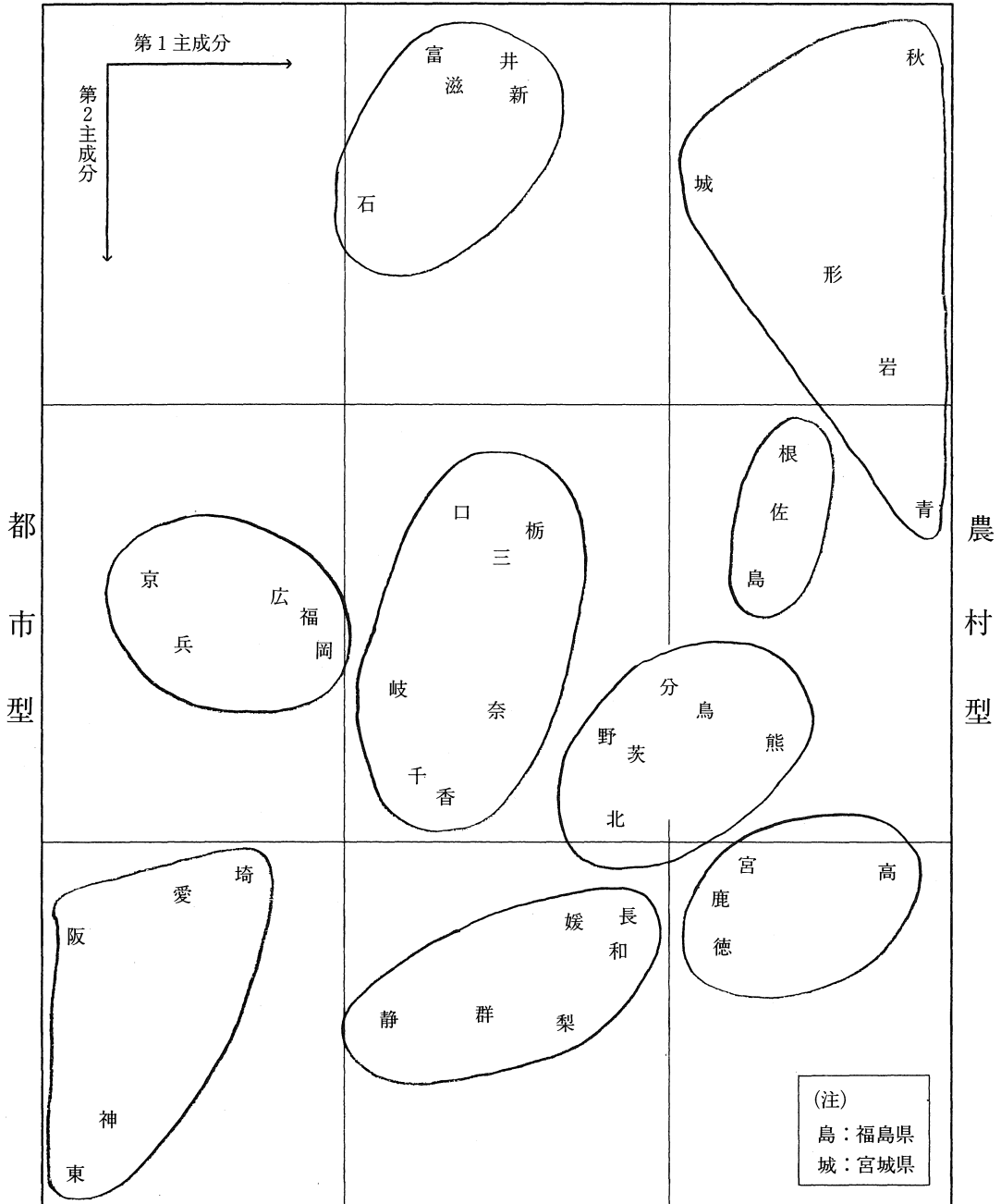
えども他都県名の次字としてのみ重複使用されている。厳密にいいかえれば、Cランクの名前のうち、Bランクの名前とのみと重字となり、Dランクの名前には使用されない場合の識別文字の決定法である。

具体的にいえば、京都府の「京」はBランクの東京都の次字としてのみ重複使用されているし、奈良県の「奈」はBランクの神奈川県の次字としてのみ重複利用されている。

神奈川県の識別文字には「神」を、東京都には「東」を識別文字として使用することは、すでにBランクの節で述べたところであるから、奈良県の識別文字として「良」でなく「奈」を用いても、これがために神奈川県

図2 都道府県単位による主成分々析

稲作型



畑作型

すると、「宮」で宮崎県を識別表示することになるから、「城」で宮城県とならざるを得ない。この「城」の識別文字についても欄外の(注)は必要である。

規則4. 名前のすべてが重字の場合は、文字の相互関係を分析し関係の単純なものから順次識別文字を決定していく。

6. 注意書き

以上の分析で47都道府県は異なる一文字で識別可能となったわけであるが、この識別方法が万人の共通認識となるまでには、やはりいくつかの注意書き、欄外の(注)が必要であろう。

Dランクの7県すべてに(注)を付けるのがもっとも親切であるし、場合によってはCランクを含めた18府県までに凡例が必要とされる場合もあろう。

しかし、なじみの深い都道府県名の場合では、欄外の(注)はつぎの二つ【島=福島県、城=宮城県】で必要十分と思われる。

Ⅲ. 3; 4次元的表示のための図形使用

図2は47都道府県を対象とした主成分々析の結果¹⁾である。第1主成分(都市型か農村型か)を横軸に、第2主成分(稲作型か畑作型か)を縦軸に表わし、都道府県の分類に利用したものである。ここで欲をいえば、この2次元平面に第3主成分そして第4主成分の結果も表示したくなる。そのときの工夫として図3、4のようなことが考えられよう。

図3 3次元的表示のための図形

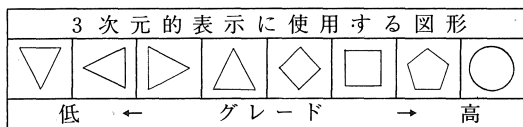
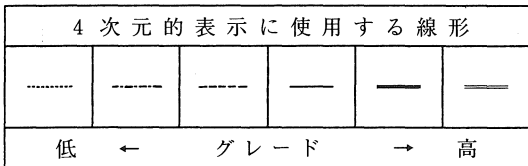


図4 4次元的表示のための線形



各識別文字をいろいろな図形で囲んで第3主成分(第3次元的表示)のグレードを表示する方法である。図形はいろいろ考えられるが、対象が50個程度であれば、4

~6のグレードが識別されれば充分であろう。

そのうえ第4主成分(第4次元的表示)をも示したいときは、この各図形の線を点線、鎖線、一点鎖線、細線、太線、二重線などに変化させて表示することもできる。対象が50個程度であれば第4次元的表示は2~3グレードもあれば充分である。

このような多次元への拡張表示が可能となったのは、47都道府県をすべて一文字で識別表示し得たことの効用である。

Ⅳ. 山形県44市町村への応用

都道府県名を識別表示するために考察した方法と規則を応用して、山形県44市町村の識別表示法について、分析過程を省略し結果のみを表6に示す。

表6 山形県市町村名の識別表示

Aランク	Bランク	Cランク
(鶴) 岡	(三) 川	山 (辺)
(新) 庄	(立) 川	大 (蔵)
(長) 井	(真)室川	大(石)田
(天) 童	(鮭) 川	(河) 北
(東) 根	(村) 山	
(南) 陽	(松) 山	
(温) 海	(中) 山	
(遊) 佐	(金) 山	
(八) 幡	(米) 沢	Dランク
(高) 畠	(尾)花沢	(川) 西
(小) 国	(戸) 沢	(西) 川
(白) 鷹	(酒) 田	(朝)日町
(飯) 豊	(平) 田	朝(日)村
(余) 目	(最) 上	(上) 山
(藤) 島	(寒)河江	山 (形)
(羽) 黒	(舟) 形	大 (江)
(楠) 引		

V. 摘 要

47都道府県を単位として、統計解析の結果を2次元座標に表示するとき、点の代わりとして各都道府県を代表しうる一文字で表示したくなる。そこで都道府県名の文字としての相互関係を分析し、識別能に応じてランク分けし、識別能の高いものから順次決定し、全て相異なる一文字で識別することができた。結果は表2~5に示さ

れており、応用例は図2に示されている。得られた規則を山形県44市町村名に適用したものが表6である。

補 注

- a) 我が国の地方自治体のうち、47都道府県には、北海道の01番から沖縄県の47番まで2桁の番号が振られている。市区町村の方には都道府県番号につづけて3桁の番号が振られている。
- b) 我が国でも郵便番号制度が制定されているが、制度そのものは合理的であるが、郵便番号ははなはだ不合理な決め方がなされている。
たとえば、〒900は沖縄県那覇市であるが、〒910は福井市となり、〒990は山形市となる。
地方自治体番号、郵便番号、それに電話番号の市外

局番の3者はいずれも異なる原理に基づいた識別番号の決定法が用いられており、利用者が使用するにも記憶するにも不便となっている。

- c) たとえば地方自治体「神奈川県」の固有名詞部分が「神奈川」なら、地方自治体「北海道」の固有名詞部分は「北海」とされるべきであるという異説があるが、これは筋は通っている。本論では通説に従ったが、「北海道」がAランクであるために異説に従ったとしても（延べ文字数は97文字となるが）、本論の識別文字の決定には影響することがない。

引用文献

- 1) 黒田昭：主成分分析による県単位分類について（1975），農業土木学会大会講演要旨集，pp 344-345