

# 日本人における姓名の統計的検証

2004MM087 山田夏華

指導教員: 尾崎俊治

## 1 はじめに

近年、年金の記録問題が発覚し、その原因の1つとして名前があげられる。名前は個人を認識するのにとても重要なモノであり、本論文では、漢字二字ずつ計四文字が最も一般的な日本独自の姓名について研究、考察していく。

## 2 順位相関係数

$n$ 個のデータの組のうちのどれか2組  $(x_i, y_i)$   $(x_j, y_j)$  に対して、次のような用語を定義する:

$x_i < x_j$ かつ $y_i < y_j$ , または $x_i > x_j$ かつ $y_i > y_j$  のとき正順位

$x_i < x_j$ かつ $y_i > y_j$ , または $x_i > x_j$ かつ $y_i < y_j$  のとき逆順位

$x_i = x_j$ または $y_i = y_j$ のとき無順位であるという。そして正順位、逆順位、無順位の組に対して、それぞれ+1点、-1点、0点を与えて、すべての組に対して平均点を求める。この値 $\tau$ をケンドールの順位相関係数という。これは次のように求められる。いま正順位、逆順位、無順位のものがそれぞれ $p, q, r$ 組あったとすれば

$$p + q + r = N = nC_2 = \frac{n(n-1)}{2}$$

であって、ケンドールの順位相関係数 $\tau$ は次式で与えられる。

$$\tau = \frac{p - q}{N} = \frac{2p + r}{N} - 1 = 1 - \frac{2q + r}{N}$$

$\tau$ は  $-1 \leq \rho \leq 1$  の範囲にあり、 $x_i, y_i$ の順位がすべて等しいとき  $(u_i = v_i) \rho = 1$  で、順位が完全に逆転するとき  $(v_i = n + 1 - u_i) \rho = -1$  で、それ以外の場合は  $-1 \leq \rho \leq 1$  である[1]。

### 2.1 愛知県の姓の傾向

では実際に、愛知県内の姓の傾向を瀬戸市を基準に順位相関係数を用いて算出していく。このデータは写録宝夢集[2]から引用した。

表 1: 瀬戸市を基準にした順位相関係数 $\tau$

尾張旭市	春日井市	一宮市	稲沢市
0.36383	0.38727	0.27636	0.28141
豊田市	岡崎市	新城市	知多市
0.18686	0.22949	0.15595	0.19292

上記の表1は瀬戸市とその他市町村を比べて順位相関係数を算出したものである。これより、全体的に $\tau$ の値は小さく、低い正の相関があるか、もしくはほとんど相関なしという結果となり、地域ごとでかなり順位が違ってくる。しかし、瀬戸市と隣接する尾張旭市や春日井市は、その中でも高い値を示しているが西側に隣接する豊田市はそれとは違い低い値を示していることに注意し

て頂きたい。また、次の表2は新城市とその近辺にある市町村とで順位相関係数を求めたものである。

表 2: 新城市を基準にした順位相関係数 $\tau$

豊田市	岡崎市	豊橋市	田原市
0.22545	0.21131	0.36929	0.26161

こちらでも新城市と隣接する豊田市や岡崎市では $\tau$ の値が小さいのに比べ、隣接する豊橋市の値は高い結果となった。これらのことから、姓はかなりの小さい範囲や地域で密集しており、その地域の文化と色濃く関係していることがわかる。瀬戸市に隣接する豊田市よりも少し距離のある一宮市や稲沢市の方に高い値が出たのは、瀬戸市と豊田市を境に尾張と東三河という大きな地域区分があるからではないだろうか。愛知県の姓は図1のような名古屋、尾張、知多、西三河、東三河の地域ごとによって独立している傾向が伺える。

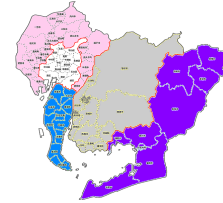


図 1: 愛知県内の地域区分

### 2.2 東京都の傾向

表 3: 新宿区を基準にした順位相関係数 $\tau$

渋谷区	中野区	文京区	世田谷区
0.71717	0.65737	0.66040	0.6830

上記の表3は東京都新宿区を基準に順位相関を求めたものである。全体的に $\tau$ の値は高く、高い正の相関があるという結果となった。これは、それぞれの区の姓順位にあまり変化がなく、似通った姓が多いためである。つまり、愛知県と比べ、姓に独立性がないと言える。東京都は昭和40年から50年にかけて転入者数のピークを迎え、今でも日本の中では最も人の出入りが多い県である。愛知県やその他の地域のように地元で根ざした姓が薄れ、全国平均的な姓が蔓延した結果である。

### 2.3 姓と転入者数の関係

次の表4は全国と各都道府県の上位100位の相関係数を算出したものである。東京都やその周りの地域は全国との相関係数が極めて高い。図2は横軸に表4の相関係数、縦軸に各都道府県の転入者数[3]をプロットした散布図である。相関係数が高い値には転入者数が多く対応し、相関係数が低い値には転入者数が少なく対応する傾向が

表 4: 全国と各都道府県の姓の相関係数ランキング

1	東京	0.96495
2	神奈川	0.94245
3	埼玉	0.90257
4	千葉	0.88692
:	:	:
44	高知	0.31433
45	鹿児島	0.29209
46	佐賀	0.28776
47	沖縄	0.22891

みられ、転入出者数と姓の全国化には大きな関係性があるといえる。

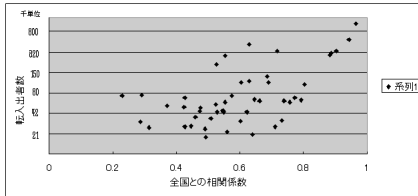


図 2: 各都道府県の相関係数と転入出者数の散布図

### 3 Power's Law

Power's Lawは統計モデルの1つで、日本語ではべき乗則と訳され観測量がパラメータのべき乗に比例する法則のことである。もっとも一般的なPower's Lawは

$$Y = CX^{-\theta}$$

で表わされる。Cとθは定数であり、θはスケーリング指数と呼ばれる。そして、両辺の変数の対数をとると

$$\log Y = -\theta \log X + \log C$$

となる。この式は、この傾きθの線形関係の形をとり、独立変数のスケーリングは関数の上か下かの移動を誘導し関数の形と傾きθの両方が変化しない[4]。

#### 3.1 データにおけるPower's Lawの検証

Power's Lawの興味深い特徴として安定性が挙げられる。これはどの分布を見ても傾向が変わらないということである。また、いずれの分布もべき乗則に従っており縦軸と横軸ともに対数の値をプロットすると直線になることである[5]。では、実際に姓のデータがこれらの特徴に当てはまるかどうかを確認したい。図3は全国とその他都道府県の人口と順位の対数を取って縦軸と横軸にプロットしたものである。

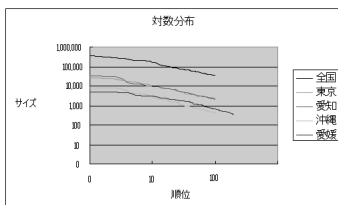


図 3: Power's Law対数分布

人口数や姓順位に違いがあるにも関わらず、全国とその他都道府県の傾向は殆ど変わらず、且つ直線を描くという結果がでた。この事により、姓のデータはPower's Lawに適合することがわかる。しかし、下記の図4を見て頂きたい。これは図3と同じように人口と順位の対数を取って各地域の区町村と全国を比べたものである。

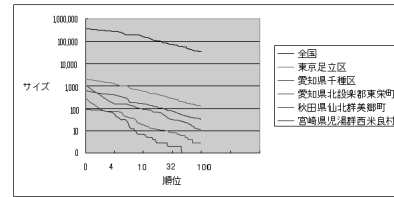


図 4: 全国と区町村の対数分布

東京や名古屋といった大きな都市の区などは都道府県と同じように全国との対数分布の傾向が同じになるが、図4にある東栄町や美郷町、西米良村のように姓のPower's Lawから外れる町や村も少なくない。これは、この土地を開墾した氏族がそのまま定着した結果である。姓のPower's Lawに属さない村や町は上位の姓を名乗る人が極めて多い傾向がある。以上のことから、姓のデータの多くがPower's Lawに適合するが氏族の結びつきが強く、最後まで離散しなかった人々が多い地域だけはこのPower's Lawに適合しない結果となった。そして、全国上位1000位のデータで回帰分析を行った結果、

$$Y = CX^{-\theta}$$

$$\log Y = -\theta \log X + \log C$$

$$\log Y = -(-0.89106) \log X + 0.271151, R^2 = 0.978$$

という式となり、これがこの姓のPower's Lawといえる。

### 4 おわりに

順位相関係数とPower's Lawを活用し姓の傾向について調べた結果、姓の広まりや独立性は人の移動と密接に関係していることが分かった。愛知県の傾向や姓と転入出者数の関係、そして姓のPower's Lawに属さない村や町が未だ数多く存在することから、交通手段が発達し日本国内なら移住も簡単な時代ではあるが、頻繁に人の出入りがあり、姓が全国化をしているのは東京都とその周辺のみであり、その他の地域ではあまり人が動いていないことが見て取れる。この結果から、地元を離れない人も割合多く、離れたとしても意外と県内、もしくは東京都周辺が多いと推測される。

#### 参考文献

- [1] 和田秀三:基本演習 確率統計,サイエンス社(1990) .
- [2] 姓名分布と姓名ランキング『写録宝夢巢』, <http://www2.nipponsoft.co.jp/bldoko/index.asp/>.
- [3] 総務省統計局HP, <http://www.stat.go.jp/index.htm>.
- [4] Wikipedia「冪乗則」 <http://ja.wikipedia.org/>.
- [5] マイケル・J・モーブッシン:投資の科学,日経BP社(2007) .