

## 1. はじめに

日本では 1875 年の平民苗字必称義務令により全国民が苗字を持つことになった。そのため苗字は土地や風土の影響を受けており、苗字の分布には強い地域差がある。近年では人口流動の高まりが影響を与えていると考えられる。本研究では苗字のローレンツ曲線とジニ係数を計算することで、空間的分布がどのようになっているかを分析する。

## 2. 分析の方法

本研究ではローレンツ曲線とジニ係数を用いて苗字の偏在性について検討する。以下ではこの分析方法について説明する。ローレンツ曲線は事象の不平等度を視覚的に表す手法である。より詳しく述べると、

苗字全掲載数を  $T$  とし、

$$T = x_1 + x_2 + \dots + x_n$$

第  $i$  階級までの累積比率を  $r_i$  とし、

$$r_i = \frac{x_1 + x_2 + \dots + x_i}{T}, (i = 1, 2, \dots, n - 1)$$

$xy$  平面上の  $n+1$  個の点

$$(0,0), \left(\frac{1}{n}, r_1\right), \left(\frac{2}{n}, r_2\right), \dots, \left(\frac{n-1}{n}, r_{n-1}\right), (1,1)$$

を結んだ折れ線がローレンツ曲線である。これらの点を結んだ線が対角線に近いほど均等に分布しているとみなすことができる。

ローレンツ曲線と対角線に囲まれた面積を求め、その面積を2倍したものがジニ係数である。0~1の値で求められ、0に近ければ近いほど均等、1に近ければ不均等に分布しているとみなす。

## 3. 分析結果

苗字のランク付けを調査し、全国および都道府県ごとの掲載数をまとめる。このデータからローレンツ曲線とジニ係数を算出し、空間分布の偏在性を求める。苗字掲載数データは写録宝夢巢 Ver.11 から採取する。この統計データは NTT の電話帳をベースに作成されたもので、全国の市町村別に苗字の分布が整理されている。

図1、表1に苗字の偏在度を表すローレンツ曲線とジニ係数を示す。佐藤は全国で最も多い苗字で全体の1.56%を占める。このジニ係数は0.59と全国の分布に対してやや不均等に偏在度が高い。逆に田中や山本、中村はジニ係数が0.5を下回り偏在度は低いという結果になった。地域性が強いと考えられる沖縄1位の比嘉はジニ係数が0.95であり、かなり偏在性が強いことがわかる。

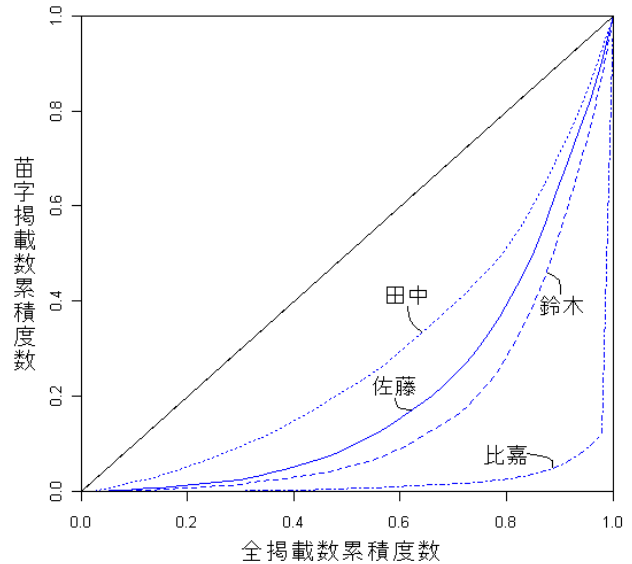


図 1: 苗字の偏在度(ローレンツ曲線)

表 1: 上位苗字のジニ係数

順位	苗字	構成比(%)	ジニ係数
1	佐藤	1.56%	0.59
2	鈴木	1.37%	0.67
3	高橋	1.14%	0.54
4	田中	1.05%	0.42
5	伊藤	0.87%	0.55
6	山本	0.86%	0.44
7	渡辺	0.86%	0.51
8	中村	0.84%	0.41
9	小林	0.82%	0.55
10	加藤	0.69%	0.59
沖縄 1 位	比嘉	0.04%	0.95

全国での上位 10 位までの苗字の構成に近い構成をとる都道府県について考察する。構成比で相関係数を求めると、東京が 0.88 で最も高く、続いて神奈川、北海道で 0.85 を上回り埼玉、千葉と順に高い。首都圏と北海道では人口が激しく流動しているためだと考えられる。

## 4. おわりに

ジニ係数を求めることにより、日本における苗字の空間的分布を把握した。首都圏と北海道に全国分布との類似性があり、ほぼ全国の縮図になっていることが確認できた。これらの地域は全国調査や市場調査のサンプルとしてこれらの地域が適していることを示すと考えられる。

## 文献

- [1] 林利充, 大澤義明, 小林隆史 “苗字の空間的偏在とその変化” オペレーションズ・リサーチ誌, 2009 年 1 月号.