

# 居住者構成のダイナミクスを反映するジオデモグラフィクスの開発

植田雄登\*・中谷友樹\*\*

## Developing Geodemographics reflecting the dynamics of resident composition

Yuto Ueda\*, Tomoki Nakaya\*\*

Geodemographics is a typological indicator describing demographical and social compositions of residents in small geographical areas. Various studies have showed its usefulness in applied geographies highlighting area differences in health, crime and educational attainments. But the neighborhood structure has been temporally changed greatly by urban development/shrinkage and aging population. Thus, geodemographics made from data of a single census year was unable to capture such historical transformations of resident's compositions. In this paper, we developed a new geodemographic typology reflecting the dynamics of resident's compositions, using six census datasets from 1995 to 2020 in Japan.

**Keywords:** ジオデモグラフィクス (Geodemographics), 居住地域分化 (residential segregation), 国勢調査 (Census)

### 1. はじめに

#### 1.1. 背景

ジオデモグラフィクスは、小さな地理的空間に暮らす人々の人口・社会的構成を要約する複合（類型化）指標である (Harris et al., 2005)。その端緒は、チャールズ・ブースが作成した貧困地図やシカゴ学派による都市生態学の社会地図研究とされ (Singleton and Longley, 2009a)、その後 1950 年代の社会地区分析および 1960 年代の因子生態分析を経て居住地域分化の指標を利用した各種の研究へと発展してきた (ノックス・ピンチ, 2013)。これらの研究によって培われた知見は、マーケティングにおける消費者のターゲティングにも有用であると考えられ、ジオデモグラフィクスは商用の地域指標として開発が続けられてきた (Singleton and Spielman, 2014)。近年では健康 (中谷・埴淵, 2013; 村中ほか, 2011; Grubestic et al. 2014; Moon et al., 2019; Grekousis et al., 2021) や犯罪 (上杉ほか, 2018)、教育達成 (上杉・矢野, 2018) など多様な応用地理学分野において、居住者の多面的な属性を捉える分析指標としてジオデモグラフィクスの学術的な有用性も指摘されている。

しかし、人口減少および高齢化、また近年のライ

フフォーの多様化により、時間の経過とともに居住地域構造は大きく変容している (中澤, 2019)。例えば、青井・中澤 (2014) によれば、東京大都市圏では、都心付近での社会経済的な居住者の豊かさがセクター的構造よりも同心円構造を強める変化を指摘している。また、豊田 (2012) では、都市圏を対象として豊かな郊外から豊かな都心へと居住地域構造に変化が生じていることを報告している。このように時間とともに変化を続ける地域に対して、単年次の国勢調査に基づいた類型化では、そのような変化を捉えることは容易ではない。また、日本においては、多くの労働者が 60 歳代で定年を迎えること、また郊外住宅地における居住者の多くが上記の年代に達することから、郊外住宅地の高齢化が急激に進行している。しかし、定年後であっても従前の職歴によって、生活様式が規定されていると想定され、単年次の国勢調査を用いるのみでは、適切に居住者の生活様式に対応した居住地域類型を得ることは難しい。

以上より、本研究では、複数年次の国勢調査データを活用することにより、居住者構成のダイナミクスを反映した地域類型の開発に取り組むことにした。

\* 学生会員 東北大学大学院環境科学研究科 (Tohoku University)

〒980-8576 宮城県仙台市青葉区荒巻字青葉 6-6 E-mail : yuto.ueda.t5@dc.tohoku.ac.jp

\*\* 正会員 東北大学大学院環境科学研究科 (Tohoku University)

## 1.2. 先行研究

ジオデモグラフィクスの分野において、複数年次の国勢調査を用いた地域類型の試みとして、McLachlan and Norman (2021) があげられる。この研究では、1991～2011 年度までの 3 年次の国勢調査を活用し、その調査項目の中でも一貫して利用可能な変数を用いて地域類型を作成している。そして、年次間の比較を行い、異なる国勢調査年次において多く小地域が同じ地域類型に属し続けることを示した。また、ジオデモグラフィクスの分野以外においても、居住地域構造の変容を捉える研究は盛んである。例えば、Delmell (2016) では、1970～2010 年度の国勢調査を活用して地域類型を作成した上で、配列解析を行っている。この方法によって、それぞれの地域が割り当てられる地域類型が時間の経過とともに、どのように変化してきたのかに基づいて、各地域を再分類している。

しかし、以上の研究では、年次間での地域類型の比較や時間的な変化に関心がある一方で、それらの知見を加えた新たな地域類型の作成は行っていない。加えて、地域類型の変化を居住者構成の変容として捉えているが、この場合より詳細な居住者構成の動的な変化が不明瞭になっている可能性が考えられる。

## 2. 方法

### 2.1. 複数年次の変数を導入する手順

複数年次の国勢調査より地域類型を作成する方法は複数考えられる。McLachlan and Norman (2021) では、3 年次の国勢調査データから全ての年次で利用可能な変数群をすべて同時に利用してクラスター分析を実行し、地域類型を作成している。この方法により、年次間での地域類型の基準が統一されるため、年次を越えて地域類型の結果を比較できる。しかし、上記の方法の場合、複数の年次を通して一貫した変数のみを利用しなければならない制約がある。日本において、国勢調査は 5 年ごとに行われている一方で、年次ごとに調査項目が変更されている場合もあり、上記の方法では年次の除外また調査項目の除外という操作を行う必要がある。

また、Delmell (2016) が行ったように、各年次で

割り当てられた地域類型をもとに配列解析を行う操作は、居住地域構造の動的な変化を捉えることができる一方で、居住者がどのような人口統計学的、社会経済的属性を有しているのかに対して不明確である。

そこで、本研究では、2020 年度の国勢調査に基づく居住者特性の変数に加えて、過去複数年次の国勢調査の結果から得られる属性の時間変化の情報を用いることで過去の変化を考慮した居住者の地域類型を作成する。

作成過程で行った国勢調査データの加工や定量的分析においては、R 言語を用い、面積按分や地図化には ArcGIS Pro を利用した。

### 2.2. 作成手順

#### 2.2.1. 利用する変数群

本研究では、政府統計から入手することができる日本全国の国勢調査小地域統計として、1995 年度から 2020 年度までの 6 年次分を利用した。利用した変数群の概要を表 1 に示す。

各年次間での国勢調査データの一貫性を確保するため、国勢調査の年次間で変更が加えられている調査項目（職業大分類や最終学歴など）においては、カテゴリを統合する処理を行った。例えば、職業大分類ではホワイトカラー、ブルーカラー、グレーカラーの 3 種類のカテゴリに単純化している。

また、分析単位が小さくなるほど割合等の指標値に極端な値が発生する可能性が高まり (Vickers and Rees, 2007)、分類結果が不安定になることが想定される。以上から、本研究では、全ての年次において

表 1 利用する変数

調査項目	該当する年度
男女別人口	1995, 2000, 2005, 2010, 2015, 2020
5歳階級別年齢人口	同上
外国人人口	同上
家族類型	同上 (ただし、年度間で変更有)
住宅建て方に関わる変数	同上
住宅所有関係に関わる変数	同上 (ただし、年度間で変更有)
労働力状態	同上
職業大分類	同上 (ただし、年度間で変更有)
最終学歴	2000, 2010, 2020
居住期間	2000, 2010, 2015, 2020

人口および世帯数が 20 以上となる小地域のみを対象とした。また、正確な居住構成を把握できないため、過去に一度でも秘匿地域となった小地域も同様に除外した。

#### 2.2.2. 面積按分および変数の加工

国勢調査における小地域は複数年次を通して一貫した形状をしていることが望ましいものの、市区町村の合併や町丁字の変更等にもとない、形状や区割りに変更が生じている。これに対処するため、本研究では、1995 年度から 2015 年度までの国勢調査の統計データについては、各年次の小地域統計境界を利用し、2020 年度国勢調査による小地域境界を基準とする面積按分処理を行い、統計データの地理的分析単位を 2020 年度の小地域に統一した。

この面積按分処理は、国勢調査メッシュ統計(2020 年度では 250m メッシュ、その他の 5 年次では 500m メッシュ)より人口が 1 以上のメッシュのみを取り出し、上記で作成した小地域統計のポリゴンから重複する箇所を抽出して実施した。これにより、面積按分する際に、本来人口が少ないにも関わらず大きな面積を占める小地域への過剰な按分を回避できる。

按分処理後、全ての観測値を総人口や総世帯数などの分母とした構成比を算出した。さらに、ある年次  $t$  における国勢調査の構成比と、その前回調査の年次  $t - 1$  における国勢調査の構成比の差を求め、これを変数とした。以上で算出された変数群を 2020 年次における国勢調査データに結合し、分析用のデータ行列とした。

#### 2.2.3. 変数の正規化・標準化

国勢調査の小地域統計には、調査項目によってゼロの数値や割合にした時に外れ値となる項目が多く含まれている。こうした外れ値への対処を含めたデータの標準化を行う決定的な方法は現状存在しない (Spielman and Singleton, 2015)。Gale et al. (2016) は、イギリスのオープンジオデモグラフィクスである 2011OAC の作成において、常用対数変換、Box-Cox 変換、逆ハイパボリックサイン変換の 3 つを比較している。これらのうち対数変換および Box-Cox 変換は、変換を行うために数値が 1 より大きい正の値である必要があり、これに対処するために適当な正の

定数を加えることが多いが、その定数値の設定如何により結果が左右される (Gale et al., 2016)。

以上から、本研究ではまず数値が 0 の場合でも算出が可能であり、また定数を指定する必要がない、逆ハイパボリックサイン変換を用いた。ただし、このような変換操作は、大きく歪んだ変数分布における裾の位置に分布する値が外れ値となる影響を小さくすることであるため、歪度の絶対値が 1 以上の変数を対象とした。

以上で正規化した変数群でクラスター分析を行う場合、分散の大きい変数は他の変数と比べて、分析結果により大きな影響を与えることから、変数の標準化を行う必要がある (Vickers and Rees, 2007)。

Gale et al. (2016) では、標準化の方法として、Z 変換、範囲標準化、頑健な範囲標準化を用いて比較している。このうち、範囲標準化は極端な外れ値がある場合うまく機能しない (ONS, 2015)。また、頑健な範囲標準化では、閾値を設定する必要がある。本研究では Z 変換を標準化の方法として採用した。

#### 2.2.4. 分類変数の選択

まず、国勢調査の調査項目の中には人口総数や世帯総数などに対して該当する対象が少ない調査項目カテゴリが含まれている。そのようなカテゴリはノイズとなり、クラスタ分析によって得られる重要なパターンを隠してしまう可能性がある (Grekousis et al., 2021)。そこで、本研究では日本全国で各調査項目を集計した際に、全体を占める割合が 1 % を下回るカテゴリは除外した。

その上で、2020 年度国勢調査データに対して、因子分析を行い (結果の詳細は省略する)、居住地域構造を捉える上で、重要であると考えられる変数群の存在を確認した。

居住地域構造の研究において日本では、社会経済的地位と家族的地位が居住地域分化を示す主要な因子であるという知見が蓄積されている (森川 1975; 高野 1994)。しかし、本研究では変数の統合といった処理を行っているため、社会経済的地位に関する変数が、家族的地位に関する変数と比べてかなり少なくなっている。クラスター分析において互いに相関関係の明瞭な変数群は、それらが共有する次

元を際立たせる効果を持っている (Gale et al., 2016) ことから、取得した変数群に対してそのままクラスター分析を実行すると、出力されたクラスターは社会経済的地位に関する次元の影響力が小さく、家族的地位に関する次元の影響力を大きく受けることになる。

それら変数群の分類におけるバランスを考慮するため、居住者の社会経済的特性に関する変数の数と家族的地位と関連する人口学的特性の変数の数が、ほぼ同数となるようにクラスター分析用の変数群を選択した (因子分析結果における因子負荷量の絶対値の大きいものを参考に選択した)。その上で、選択した変数の過去の調査年次からの差分の変数をすべて加えてクラスター分析を実施した。

#### 2.2.5. クラスター分析手法

##### 2.2.5.1. 分類手法

多変量データを定量的に分類する手法には、階層型クラスター分析や非階層型クラスター分析など多くの選択肢がある (Kogan et al., 2006)。ジオデモグラフィクスに利用するデータセットは大規模であることが多いため、非階層型クラスター分析の *k-means* 法が主に用いられてきた。*k-means* 法では初期条件としてランダムにクラスター代表点を選択し、各データを最近隣の代表点に属させる。そして、所属するデータとクラスター代表点との距離の 2 乗の総和が最小となるように、代表点を改めて算出する。以上の代表点への割り当てと代表点の更新を繰り返すことで、クラスター内の同質性を高める。しかし、この手法はデータに含まれる外れ値に強く影響を受けてしまう (Wu et al., 2008)。

そこで、本研究では、非階層型クラスター分析の一種である *k-medoids* 法を採用した。この分類手法は、先述した *k-means* 法の課題を克服するために提案された (Arora et al 2016)。*k-medoids* 法では、初期条件として各データの中からランダムに代表点を選び、他のデータを最近隣の代表点に属させる処理は *k-means* 法と同様だが、所属するデータ間の距離の総和が最小になるように新たなクラスター代表点 (medoid) を、各データの中から選択する。このとき算出される代表点は、クラスター内の中央値的な

特徴を有する。このように、新たなクラスター代表点を更新する過程で、*k-medoids* 法は 2 乗距離を利用しないことから、外れ値の影響を受けにくいとされる。

ただし、*k-medoids* 法を実施するアルゴリズムは、すべての対象間の類似度を算出する必要があり、分類対象となるデータ数に計算可能な上限が存在する。そのため、本研究ではその代替的手法として開発され、R 言語で実装されている CLARA (Clustering Large Applications) を用いた (Kaufman and Rousseeuw, 1986)。CLARA とは、計算時間とコンピュータのメモリ領域を抑えながら、*k-medoids* 法によるクラスター分析を、膨大な観測値をもつデータセットに適用するために開発された手法である (Kaufman and Rousseeuw, 2005)。まず、分類対象であるデータセットから指定したサイズで観測値をランダムにサンプリングし、サブデータセットを複数個作成する。次に、サブデータセットに対して *k-medoids* 法によるクラスター分析を実施し、残りのデータセットの各観測値を上記の分析で作成されたクラスターの代表点の中で最も類似するものに割り当てる。そして、以上の操作で作成されたクラスターのそれぞれの非類似度を計測し、サブデータセット内での平均値を算出する。最後に、非類似度の平均値をサブデータセット間で比較し、最小の数値であるサブデータセットのみ保持する。以上のサンプリングとクラスター分析、そしてサブデータセットの保持を、指定した回数繰り返し行う。ただし、2 回目以降のサンプリングでは、前回保持されたサブデータセットにおける代表点を必ず含まれるようにする。指定したサンプリングの回数この処理を繰り返した後、最良となったサブデータセットによる結果を出力する。

本研究では、分析実行のために事前に指定するサンプル数とサンプルサイズに関して、前者を 100、後者を 5000 とした。

##### 2.2.5.2. クラスター数の決定方法

非階層型クラスター分析は、事前にクラスター数を決定しておく必要があるが、その決定は基本的に

主観的なプロセスである (Singleton and Longley, 2009b). そこで, 本研究では, 階層型クラスター分析におけるデンドログラムのように, 視覚的にクラスター数を決定できるクラスタグラム (Schonlau, 2002) を用いた. このクラスタグラムは, 近年の地域類型の開発研究で, クラスター数の決定の際に利用されている手法である (Dolega et al., 2019; Cocking et al., 2020; Singleton et al., 2020). クラスタグラムは, 縦軸を主成分分析で抽出される主成分の得点, 横軸をクラスター数として, 各クラスターの代表点をプロットした図である. そして, その図を観察し, 代表点の縦軸方向の分離度や, また代表点の安定性を考慮してクラスター数を決定する.

本研究においてもクラスタグラムを活用してクラスター数を決定する. ただし, クラスター分析において, クラスター数が大きいほど, クラスター間の属性による違いを捉えることが困難になる (Singleton and Spielman, 2014). これを踏まえ, クラスタグラムを作成するにあたって捉えるクラスター数の範囲を 2~20 とした. なお, 先行研究において, 単年次の国勢調査データを利用した 2001OAC や 2011OAC はそれぞれ大分類が 7 および 8 であった (Vickers and Rees, 2007; Gale et al. 2016). 本研究では単年次では捉えきれない居住者構成のダイナミクスを反映させるため, より多くのクラスター数を選択し, 解釈の容易さを検討した上で最終的なクラスター数を決定した.

### 3. 地域類型の結果

選択した変数群に基づいて CLARA を実行し, 作成したクラスタグラムが, 図 1 である. 図 1 より, 代表点の縦軸方向が十分に分離されており, また, 代表点がより多いクラスター数の場合にも安定して抽出されていることから, 本研究ではクラスター数を 18 とした.

この CLARA によって得られた 18 地域類型の地理的分布を図 2 に示す. 得られた類型の 1 例として D 類型をみると, その地理的分布は東京大都市圏では中心から 20km 圏内, および地方都市では市街地

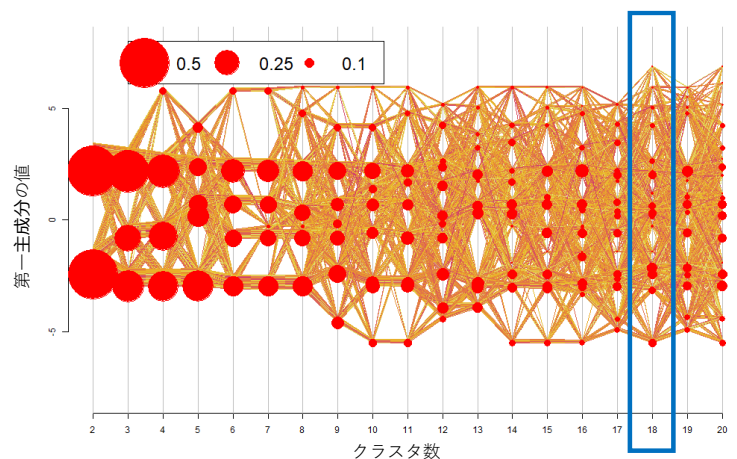


図 1 クラスタグラム

中心部に主に分布している (図 3). この類型に含まれる地区の medoid 特性である各変数の代表値を図 4 に示した. 特徴として, 2020 年現在においてはホワイトカラー割合や大学・大学院の卒業生割合など社会的経済的地位の高い居住者特性である変数の Z 値が高いことが挙げられる. 加えて, 共同住宅 6~10 階建や単独世帯, 民間借家の割合の Z 値も高い. 一方で, 共同住宅 6~10 階建の変化量に関する Z 値が, 2000 年~2005 年および 2005 年~2010 年において高い値となるものの, 2010~2015 年では値がおおよそ 0, 2015 年~2020 年においては負の値となった. 同様の傾向として, 夫婦子供世帯の変化量に関する Z 値が, 2015 年~2020 年において負の値に転換している. 以上の傾向と反対に, 単独世帯の変化量に関する Z 値は, 2010 年~2015 年および以前の年次間においては負の値であったが, 2015 年~2020 年では正の値であった. 以上より, 都心から近郊にあり利便性の高い地区に居住する流動性の高いホワイトカラーの就業者世帯の居住地であるが近年は居住者の単身化が進んでいる地域と考えられる.

別の例として H 類型をみると, この地域類型はその地理的分布が東京大都市圏では, D 類型よりも外側に位置する中心部から 20~40km の領域に主に分布している (図 3). 図 5 より, 変数の特徴として, ホワイトカラー割合や大学・大学院の卒業生割合など社会的経済的地位の高い居住者特性である変数の Z 値が高い点は D 類型と同様である. ただし D 類型

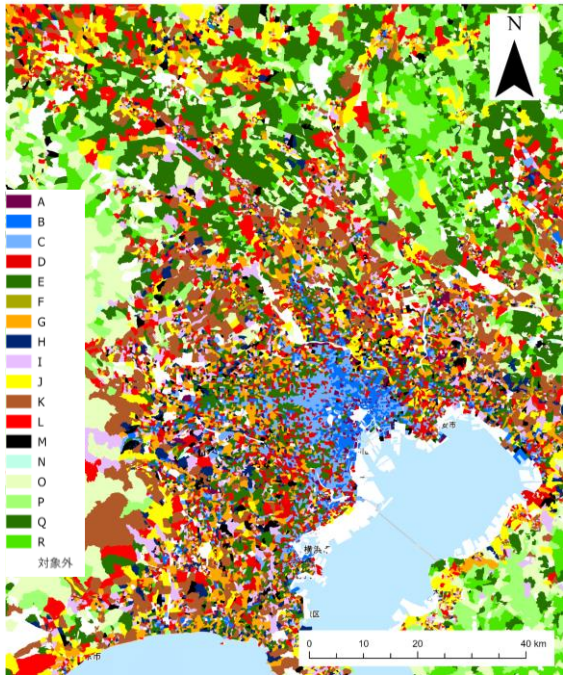


図2 首都圏における地域類型の分布

と異なり、夫婦と子供からなる世帯や持ち家の割合のZ値が高い。居住者特性の変化をみると、2010年から2015年および2015年から2020年にかけてのホワイトカラーの変化量のZ値が負の値となっており、当該の時期にホワイトカラー就業者の割合が減少している郊外住宅地域と解釈できる。

## 4. おわりに

### 4.1. 考察

本研究では、複数年次の国勢調査データを活用し、居住者構成のダイナミクスを捉えた地域類型の開発に取り組んだ。その結果、単年次の国勢調査を利用するのみでは捉えきれない社会的構成の変化を反映した地域類型を得た。

例えば、D 類型は 2020 年現在では経済的にゆとりのある流動性の高い居住者の居住地と理解できる。近年では共同住宅 6~10 階建居住および夫婦と子供からなる居住する世帯は増加の鈍化ないし減少が進む一方で、単独世帯の割合が増加している。こうした地域は、かつて子供のいる世帯の住宅需要に合わせて建設された駅周辺のマンションの多い地区であったが、最近の少子高齢化やライフコースの多様化に伴う需要減による影響を受け、居住者構成の単身

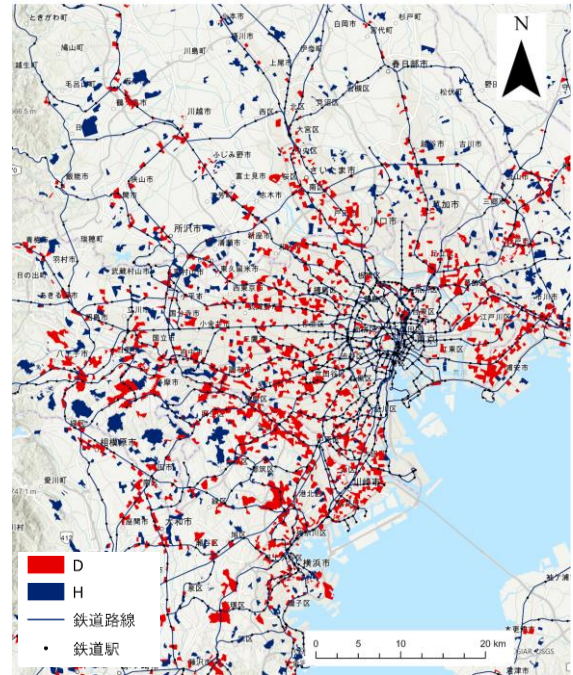


図3 D・H 類型の東京中心部における分布

化が進む地域であると考えられる。

H 類型は、その特徴から経済的に豊かな郊外戸建住宅地であることがわかる。一方で、2020 年度国勢調査においてホワイトカラーの割合が相対的に高いにもかかわらず、他の地域と比べてその割合が前年次より減少する傾向にある。高度成長期以降に開発された住宅地域が多く含まれ、2010 年ごろに定年になる団塊の世代が多く居住しており、この居住者の定年による非労働力化や働き方の変化による居住者構成の変化が顕在化した地域と考えられる。

このように単年次では単独世帯が多く居住する D 類型、あるいは郊外住宅地としての特徴を有する H 類型のような特徴を得るばかりでなく、複数年次の国勢調査データを活用することで、各類型固有の居住者構成のダイナミクス、すなわち 1995 年以降現在も進行する低成長時代の居住者構成の変容を捉えることができたと考えられる。

### 4.2. 今後の課題

本研究では国勢調査の年次間の構成比の差を用いることで居住者構成の変化を捉えようとした。こうした変化には、居住者の継続的な居住によって自然な人口増加または人口減少といった系統的な変化とそうではない変化に区別できる。しかし、現段階で

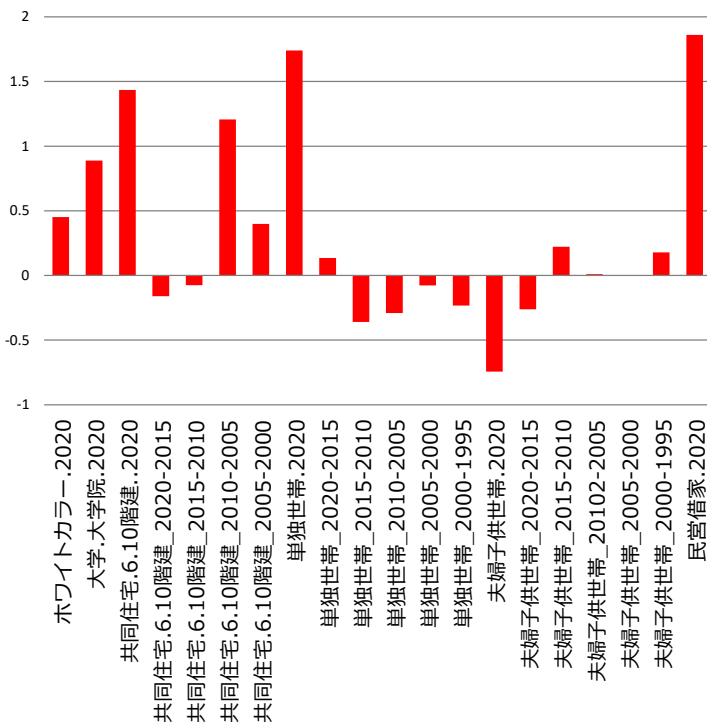


図4 D類型における特徴的な変数のZ値

はその様な区別を考慮した分類はできていない。

また、国勢調査の小地域は、オープンデータとして公開されておりデータ入手の利便性は高いが、必ずしも住所と対応した地理的単位とはなっておらず、社会調査データなどの住所情報を持つデータとのデータリンケージの面で理想的な単位とは言えない。そのため、今後は郵便番号区を基礎的な地理的な単位とする分類の有用性も検討したいと考えている。

### 参考文献

青井新之介・中澤高志 (2014) 東京圏における世帯内単身者とブルーカラー従事者の空間パターンの変容-展開法の応用. 「E-journal GEO」, **9**(2), 12-32.  
 上杉昌也・樋野公宏・矢野桂司 (2018) ジオデモグラフィクスによる社会地区類型を活用した窃盗犯の発生要因に関する小地域分析. 「E-journal GEO」, **13**(1), 11-23.  
 上杉昌也・矢野桂司 (2018) ジオデモグラフィクスを用いた教育水準の学校間格差の評価—大阪市を事例として—. 「人文地理」, **70**(2), 253-271.  
 高野岳彦 (1994) 仙台市における近年の住民属性と

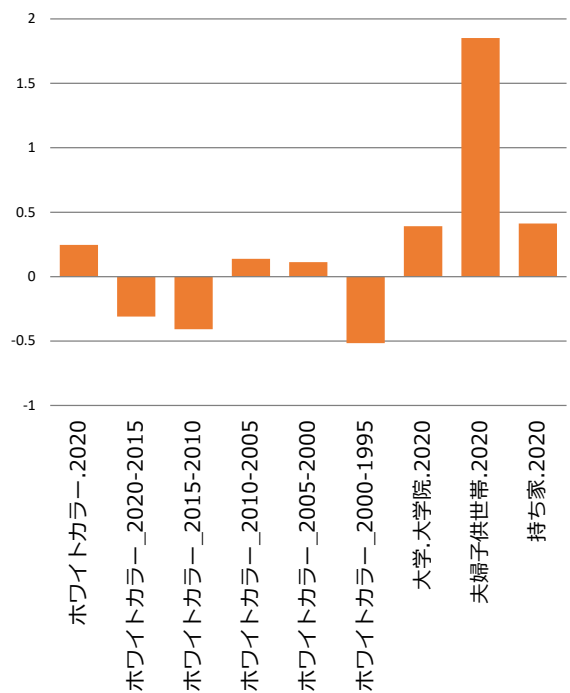


図5 H類型における特徴的な変数のZ値

居住地区分化の変質. 「地理学評論 Ser. A」, **67**(11), 753-774.

豊田哲也 (2012) 所得の地域格差と都市圏の空間構造. 「徳島大学総合科学部 人間社会文化研究」, **20**(51), 62.

中澤高志 (2019) 『住まいと仕事の地理学』. 旬報社.  
 中谷友樹・埴淵知哉 (2013) 居住地域の健康格差と所得格差. 「経済地理学年報」, **59**, 57-72.

村中亮夫・中谷友樹・埴淵知哉 (2011) 社会地区類型に着目した花粉症有病率の地域差. 「GIS—理論と応用」, **19**(2), 71-81.

森川洋 (1975) 都市社会地理研究の進展. 「人文地理」, **27**(6), 638-666.

ノックス, P.・ピンチ, S. 著, 川口太郎・神谷浩夫・中澤高志訳 (2013) 『改訂新版 都市社会地理学』. 古今書院. Knox, P. and Pinch, S. (2010) *Urban Social Geography: An Introduction. Sixth edition.* London: Pearson.

Arora, P., Deepali, and Varshney, S. (2016) Analysis of K-Means and K-Medoids Algorithm for Big Data. *Physics Procedia*, **78**(2016), 507-512.

Cockings, S., Martin, D. and Harfoot, A. (2020)

- Developing a National Geodemographic Classification of Workplace Zones. *Applied Spatial Analysis and Policy*, **13**(4), 959-983.
- Delmelle, E. C. (2016) Mapping the DNA of urban neighborhoods: Clustering longitudinal sequences of neighborhood socioeconomic change. *Annals of the American Association of Geographers*, **106**(1), 36-56.
- Dolega, L., Reynolds, J., Singleton, A. and Pavlis, M. (2019) Beyond retail: new ways of classifying UK shopping and consumption spaces. *Environment and Planning B: Urban Analytics and City Science*, **0**(0), 1-19.
- Gale, C. G., Singleton, A. D., Bates, A. G. and Longley, P. A. (2016) Creating the 2011 area classification for output areas (2011 OAC). *Journal of Spatial Information Science*, (**12**), 1-27.
- Grekousis, G., Wang, R. and Liu, Y. (2021) Mapping the geodemographics of racial, economic, health, and COVID-19 deaths inequalities in the conterminous US. *Applied Geography*, **135**, 102558.
- Grubestic, T. H., Miller, J. A. and Murray, A. T. (2014) Geospatial and geodemographic insights for diabetes in the United States. *Applied Geography*, **55**, 117-126.
- Harris, R., Sleight, P., and Webber, R. (2005): *Geodemographics, GIS and neighbourhood targeting*. Wiley, London.
- Kaufman, L. and Rousseeuw, P. J. (1986) Clustering large data sets (with discussion), In Gelsema, E. S. and Kanal, L. N. eds. *Pattern Recognition in Practice II*, Elsevier/North-Holland, Amsterdam, pp. 45-437. [3.1, 3.4.1]
- Kaufman, L. and Rousseeuw, P. J. (2005) *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.
- Kogan, J., Nicholas, C. and Teboulle, M. (2006) *Grouping multidimensional data*. Springer-Verlag Berlin Heidelberg.
- Moon, G., Twigg, L., Jones, K., Aitken, G. and Taylor, J. (2019) The utility of geodemographic indicators in small area estimates of limiting long-term illness. *Social Science and Medicine*, **227**, 47-55.
- McLachlan, G. and Norman, P. (2021) Analysing socio-economic change using a time comparable geodemographic classification: England and Wales, 1991–2011. *Applied Spatial Analysis and Policy*, **14**(1), 89-111.
- Office for National Statistics (ONS) (2015) *Methodology Note for the 2011 Area Classification for Local Authorities*. July 2015.
- Petersen, J., Gibin, M., Longley, P., Mateos, P., Atkinson, P. and Ashby, D. (2011) Geodemographics as a tool for targeting neighbourhoods in public health campaigns. *Journal of Geographical Systems*, **13**(2), 173-192.
- Schonlau, M. (2002) The Clustergram: A Graph for Visualizing Hierarchical and Nonhierarchical Cluster Analyses. *The Stata Journal: Promoting communications on statistics and Stata*, **2**(4), 391-402.
- Singleton, A.D. and Longley, P.A. (2009a) Geodemographics, visualisation, and social networks in applied geography. *Applied Geography*, **29**(3), 289-298.
- Singleton, A.D. and Longley, P.A. (2009b) Creating open source geodemographics: Refining a national classification of census output areas for applications in higher education. *Papers in Regional Science*, **88**(3), 643-666.
- Singleton, A. D. and Spielman, S. E. (2014) The past, present, and future of geodemographic research in the United States and United Kingdom. *The Professional Geographer*, **66**(4), 558-567.
- Spielman, S. E. and Singleton, A. (2015) Studying neighborhoods using uncertain data from the American community survey: a contextual approach. *Annals of the Association of American Geographers*, **105**(5), 1003-1025.
- Vickers, D. and Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, **170**(2), 379-403.
- Wu, X., Kumar, V., Ross, Q.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J. Ng, A., Liu, B., Yu, P.S.



Zhou, Z.H., Steinbach, M., Hand, D.J. and Steinberg, D.  
(2008) Top 10 algorithms in data mining. *Knowledge  
and Information Systems*, **14**(1), 1-37.