

研究論文

EDA 手法による要介護認定率の地域差の分析

谷口豊*

大和田孝文†

大塚忠義‡

2020 年 11 月 25 日投稿

2021 年 3 月 25 日受理

概要

公的介護保険の持続可能性を確認する上で、市区町村ごとの要介護認定率の特性を分析することは重要である。実際に、公的介護保険制度は全国一律の制度であるが、地域ごとに異なる人口分布の影響を取り除いたとしても、要介護認定率は地域によって大きな差が見られる。

本稿の目的は、市区町村別の要介護認定率の地域差を解明することである。具体的には、市区町村別の公的統計データから活用可能な説明変数をもとに、探索的データ解析に基づく手法により、要介護認定率の高い地域・低い地域を類型化し、介護格差の地域イメージを明らかにすることである。分析の結果、要介護認定率については、大都市圏やその周辺都市で高い傾向であることを確認した。さらに、西日本の海岸沿いの工場地帯、愛知県や滋賀県の企業城下町、農業従事者が多い山村の農村部、完全失業率の低い小さな町・村・島などで特徴的な傾向が確認された。

キーワード：介護格差、標準化要介護認定比、社会的環境因子、探索的データ解析

1 はじめに

公的介護保険は、我が国の社会保障制度を構成する主要な社会保険のひとつであり、国民すべてを対象として全国一律の介護保障を行う世界でも稀有の制度である。2019年3月末時点で、公的介護保険による要支援・要介護認定者は658万人、介護給付・予防給付の費用額は10兆4319億円となっており、増加傾向が継続している¹。

公的介護保険による要介護者の認定基準・認定方式および給付水準・給付方法は全国一律であるが、市区町村を保険者として、市区町村単位で運営されている。そして、要介護者数、要介護認定率および平均介護費用は地域ごとの差異が存在する。これは、地域ごとの人口構成、人口動態の差異に加え、地域の特性や生活習慣の違い等によるものと考えられている。公的介護保険の持続可能性を確認するうえで、市区町村ごとの要介護認定率の特性を分析することは重要である。

* 早稲田大学総合研究機構保険研究所 〒169-8050 新宿区西早稲田 1-6-1 Email: taniguchi.yutaka12@gmail.com

† 損害保険料率算出機構 Email: takafumiowada@gmail.com 本稿の内容はすべて筆者の個人的な見解であり、筆者が所属する団体の見解とは無関係である。

‡ 早稲田大学大学院会計研究科 〒169-8050 新宿区西早稲田 1-6-1 Email: otsukata@aoni.waseda.jp

本稿は早稲田大学特定課題研究助成費(課題番号 2020C-178)による成果である。

¹ 介護保険事業状況報告(厚生労働省, 2019)

要介護指標の分析は個々人の生活習慣の側面からのアプローチと個人を取り巻く社会的環境因子の側面からのアプローチがある。要介護認定率の調査・分析に係る先行研究は特定地域における情報を収集・分析したものが多数を占める。要介護認定率と生活習慣の係に係る分析は、多岐にわたる情報を分析する必要性から特定地域において質問票を収集する利点が大いだが、社会的環境因子にかかる分析について特定地域を対象に行った場合、得られた知見が全国の場合と同一であるか不明である。

本稿の目的は、市区町村別の要介護認定率の地域差を解明することである。具体的には、市区町村別の公的統計データから活用可能な説明変数をもとに、探索的データ解析（exploratory data analysis、以下「EDA」と称する）に基づく手法により、要介護認定率の高い地域・低い地域を類型化し、介護格差の地域イメージを明らかにすることである。目的変数とする要介護認定率は市区町村別に公表されている数値ではなく、市区町村ごとに異なる年齢構成比・男女比の調整を行い比較可能なものとなるよう標準化し指標化した標準化要介護認定比を用いる。

EDA とはデータの特徴を発見するための手段の総称である。データ解析の手法には大きく分けて、検証的データ分析（confirmatory data analysis、以下「CDA」と称する）と EDA とがある。CDA は仮説検証を目的とする分析手法の総称であるのに対し、EDA は大規模、多属的なデータから仮説・知見等を得ることを目的とする分析手法である。EDA では説明変数をあらかじめ選択せず広い対象から知見を求めて探索的に分析する。

これ以降の構成は次のとおりである。第 2 章では先行研究を整理し本稿の新規性を明確にする。第 3 章では分析手法を詳述する。つぎに、第 4 章で要介護認定率の高い地域、低い地域について地理的可視化による分析および記述統計を用いた分析を行う。第 5 章では機械学習により説明変数の選択を、第 6 章でクラスター分析により市区町村の類型化を行う。そして、第 7 章で分析の結果の考察を行い第 8 章で分析によって得られた知見をまとめるとともに残された課題を整理する。

2 要介護認定率の決定要因および地域分類に係る先行研究

先行研究は、分析者が主観的に選択した説明変数を用いて分析する CDA の手法による仮説検証が多くを占めている。それに対して、本稿ではあらかじめ選択せずに幅広い説明変数を用いて分析する EDA の手法をとることで、未知の地域特性の探索を試みる。

また、先行研究の多くは回帰分析などの手法により要介護の要因分析を行っているが、要介護に影響がある社会的環境因子は複合的であり、単一方程式モデルで複合的要素を記述するには限界があると考えられる。そして、要因分析により得られた社会的環境因子は地域特性を表現する代理変数である可能性もある。本稿では地域間の介護格差を類型化により把握することで要介護の複合的要因への接近を試みる。

さらに、地域分類の先行研究においては産業発展に伴う地域間不均衡の分析を目的としたものが主である。医療費格差の分析を目的とした地域分類は少数存在するものの、要介護認定率の格差の分析を目的とした地域分類はほとんどない。

2.1 要介護認定率の決定要因

要介護認定率の決定要因についての分析を行った先行研究は、内閣府（2018）、平井他（2009）、久保寺（2013）などがある。内閣府（2018）では、単相関、固定効果モデルおよび OLS 推定による分析を市区町村単位で実施し、普及啓発等の予防事業の参加人数の増加が、軽度要介護認定率の上昇を抑制させる傾向があるとの結論を得た。平井他（2009）では、地域在住高齢者 9,702 人を 3 年間追跡し、要介護認定のリスク要因について、服薬数が多い、咀

嚼力が低い、歩行時間 30 分未満、友人と会う頻度月 1 回未満などを特定している。安藤（2007）は、要介護認定率に関し、所得段階 1²は正に有意であることを確認し、生活保護世帯などに要介護者が多く存在している可能性や、この階層が要介護認定にアクセスしやすいか認定審査を通過しやすい可能性があることを推測した。久保寺（2013）では、介護サービス種類の選択行動と、サービス利用量の選択行動とを分けて分析を行い、所得水準、世帯構成、産業特性、医師密度、施設定員率、人口密度等を説明変数として市区町村単位の回帰分析を行った。分析の結果、居住介護サービスについては、低所得者が多い地域では居住サービスへのニーズが低いこと、高齢者夫婦世帯割合は負で有意であることを確認した。また、1 人当たりのサービス利用件数については、人口密度、医師密度は正で有意、第一次産業比率は負で有意であり、居住サービスの供給体制が充実している地域ほどサービス利用量が増えると推測した。この他、所得に注目をして、要介護認定割合との関係を分析した先行研究は、近藤他（2012）、齋藤（2018）、佐藤（2016）などがある。近藤他（2012）は、6 自治体に居住する高齢者 14,652 人を 4 年間追跡調査したデータをもとに、要介護認定率と所得との関係性を分析し、最高所得層と最低所得層の間に男性 1.55 倍、女性 1.22 倍の要介護認定率の差が認められることを確認した。齋藤（2018）は、首都圏の A 自治体に居住する介護保険の第一号被保険者約 9 万人のデータをもとに、所得段階を低い順から 11 に分類し、女性では所得段階 1 の要介護発生率は中央値である所得段階 5 の約 2.4 倍、要介護発生率が最も低い所得段階 7 の約 4.6 倍であり、男性では所得段階 1 の要介護発生率は中央値である所得段階 8 の約 2.5 倍、要介護発生率が最も低い所得段階 10 の約 4.8 倍であることを分析した。佐藤（2016）は、同データをもとに、所得段階が 1 段階上がると、要介護危険率（オッズ）が 0.869 倍に低下することを示している。

2.2 市区町村別の要介護認定率

都道府県別の要介護認定率の差について示したデータとしては、厚生労働省（2015）があるが、各都道府県で異なる年齢構成比や男女比の調整は行っておらず、同一基準での要介護認定率の地域差を分析するのは困難である。宮下他（1999）では、年齢構成比および男女比の調整を行った上で都道府県ごとに 1995 年度の 65 歳以上の要介護者数の期待数に対する比の算定を行っているが、東京都や大阪府の比率は高くない結果となっている。石橋、鈴木（2019）では、年齢構成比および男女比の調整を行った「地域包括ケア「見える化」システム」（厚生労働省）をもとに、都道府県ごとに 2017 年度の要介護認定率を示しているが、大阪府は高い数値となっている。

また、都道府県別よりさらに細かい市区町村別の要介護認定率の差について示したデータは「地域包括ケア「見える化」システム」がある。このデータは年齢構成比や男女比の調整を行っている。しかし、このような市区町村間で比較可能となるよう調整されたデータは存在するものの、市区町村間の介護格差を分析した先行研究は管見の限り存在しない。

2.3 地域分類の研究

都道府県単位で統計指標を用いて分類を行った先行研究は数多く存在する。しかし、都道府県単位では、地域ごとの人口構成、人口動態の差異、地域の特性や生活習慣の違い等が十分反映されない。市区町村単位での地域分類の方が地域の特徴がより鮮明に浮き彫りになる。市区町村単位で地域を分類した先行研究としては、小内（1996）、森川（1998）、蓮見（2016）、山本他（2018）がある。小内（1996）は、地域の生産力、生産関係、産業構造、開放性、階層性等の指標を用いてクラスター分析により、経済的基盤の脆弱な産業基盤停滞地域、低生産力商業地域、低生産

² 平成 27 年度介護保険事業状況報告（年報）

力サービス地域、管理・行政的な管理中枢地域に分類した。

森川（1998）は、人口、世帯、産業別・職業別就業者比率、所得額、小売販売額、教育水準といった多様な指標（38 指標）を用いてクラスター分析により地域分類を行っている。38 指標に対し因子分析を行い、固有値 1 以上の 8 因子を抽出した後、説明力の高い 3 つの因子（「社会階層因子」「小売商業 因子」「工業因子」）により 8 つの地域（「地方の工業都市」「地方の中小中心都市」「若干工業機能をもつ農村的都市」「住宅型周辺都市」「工業機能をもつ周辺都市」「大都市および地方の中心都市」「若干商業機能をもつ農村的都市」「農村的都市」）に分類した。

蓮見（2016）は、産業発展に伴う地域の不均衡発展や地域間格差を明らかにする視点で分析を行い、人口増加率、生産性、消費、医療、所得水準、自治体財政等の指標を用いて、指標の条件が上位および下位の市区町村を分類した。

山本他（2018）は、統計指標に対し主成分分析を行い、得られた 8 主成分「人材の集積」「人口の流動性」「特定の仕事の場」「商業集積」「生活(非生産)の場」「拠点立地」「行政需要」「第一次産業の牽引」を用いてクラスター分析より 20 の地域に分類した。

また、機械学習により地域のグループ化を行っている先行研究としては、須田（2018）がある。須田（2018）は、26 の説明変数（都道府県別）の候補からランダムフォレスト、回帰木の手法により説明変数を特定し、選択された説明変数を用いて 1 人当たり医療費の水準について都道府県のグループ分けを行った。

3 分析方法

3.1 使用するデータ

本稿において使用した公的統計は、目的変数については「介護保険事業状況報告（年報）」（厚生労働省）である。説明変数については、別表 1 に列挙した 19 種類の公的統計である。対象年度は、2010 年度と 2015 年度とした。これは説明変数に国勢調査を使用しているため国勢調査の実施年度を対象とした。また、要介護度の認定基準は 2009 年に変更されており、変更前後では一致率が低いことが知られているので³、それ以前の情報を使用しないこととした。

目的変数は「介護保険事業状況報告（年報）」をもとに作成した軽度要介護認定率（65 歳以上の要支援 1、要支援 2、要介護 1）の標準化要介護認定比（軽度）と、重度要介護認定率（65 歳以上の要介護 2⁴、要介護 3、要介護 4、要介護 5）の標準化要介護認定比（重度）である。それぞれの目的変数は、先行研究（内閣府（2018））で要介護認定率が年齢構成と強い正の相関が示されたことを参考にし、地域ごとに異なる年齢構成比、男女比の調整を行い、比較可能なものとなるよう標準化した。標準化とは、当該地域が、全国平均の要介護認定率であった場合の要介護者数に対して、実際の要介護者数を指標化したものであり、地域 A における標準化要介護認定比 R_A は次の式(1)により作成している。

$$R_A = \frac{\sum_{s,t} L_{A,s,t}}{\sum_{s,t} C_{s,t} \cdot P_{A,s,t}} \quad (1)$$

ただし、 $L_{A,s,t}$ は地域 A における性別 s の t から t+4 歳の要介護認定者数、 $C_{s,t}$ は性別 s の t から t+4 歳の全国平均の要介護認定率、 $P_{A,s,t}$ は地域 A における性別 s の t から t+4 歳の人口であるとする。また、年齢階級の t は $t \in$

³ 三宅文枝ほか（2011）

⁴ 先行研究では、要介護 2 を軽度要介護として取り扱うケースも多いが、本稿では介護予防やリハビリ等により健康状態に回復する可能性が比較的高い要介護 1 までを軽度と定義した。

{65,70,75,80,85,90}であり⁵、 s は性別であるため $s \in \{\text{男性}, \text{女性}\}$ である。

各市区町村で男女の割合、高齢者の年齢分布が異なるため、単純に（各市区町村での要介護認定者数）/（第一号被保険者）を算定すると、高齢者の年齢分布が高い地域においては要介護認定率が高く算定される。このような見かけ上の要介護認定率の高低は、上記算式で示した標準化を行うことによって是正され、各市区町村での比較を正確に行えるようになる。

なお、得られた標準化要介護認定比（軽度）・標準化要介護認定比（重度）の基本統計量は次のとおりである。標準化要介護認定比（軽度）・標準化要介護認定比（重度）の平均はともに1未満となっている。この平均は各市区町村の標準化要介護認定比の算術平均でありその平均が1未満ということは、人数規模の大きい市区町村では高く小さい市区町村では低いことを意味している。

標準偏差は軽度の方が重度より大きいので、軽度の方が重度に比べ地域間の格差が大きいことが伺える。ただし、最大値については、重度の方が軽度より大きいので、特定の地域では重度の要介護者が多いことといえる。また、軽度の最小値は0である一方重度の最小値は0でないことから、軽度の要介護者がおらず重度になってようやく認定を受ける地域があることを示唆している。

表1 標準化要介護認定比（軽度）・標準化要介護認定比（重度）の基本統計量

	標準化要介護認定比（軽度）	標準化要介護認定比（重度）
平均	0.901	0.969
最大値	1.905	2.457
最小値	0.000	0.301
標準偏差	0.229	0.151

説明変数は『国勢調査』（総務省統計局）、『地域保健・健康増進事業報告』（厚生労働省）など、社会的環境因子として市区町村別に把握可能な公的な統計値を幅広く候補とする。ただし、説明変数間の相関が高いものはどちらか片方を選択した。結果、候補として採用した説明変数は別表1のとおりである。なお、調査不可能などの理由により福島県の特定期域において欠損値が生じている場合は当該市区町村ごと削除した⁷。また、調査基準が異なるなどの理由により、欠損値が複数の市区町村にまたがっている場合は、当該項目を説明変数の候補から除外した。その結果、分析対象とする説明変数の候補数は79となった。

3.2 地理的可視化による分析および記述統計

目的変数の標準化要介護認定比（軽度）・標準化要介護認定比（重度）について、2015年度データで、それぞれ比率の高い地域、低い地域をヒートマップで示す。可視化により得られた情報をもとに地域の特性を考察する。そして、

⁵ $t=90$ のときは $t+4$ 歳は94歳ではなく最高年齢まで。また、2010年度はデータの制約上、年齢階級は5歳毎ではなく65～74歳、75歳以上の2区分としている。

⁶ 相関が-0.8以下、0.8以上である場合は相関が高いとした。ただし、平均年齢については、高齢者単身世帯数や高齢者夫婦世帯数と相関が高かったが、地域を特徴づける重要な説明変数と考え残すこととした。

⁷ 2015年度データについては、人口が欠損値となっている福島県の富岡町、大熊町、双葉町、浪江町のレコードを削除。世帯数が欠損値となっている福島県飯館村のレコードを削除。平均寿命が欠損値となっている福島県の楢葉町、川内村、葛尾村のレコードを削除した。2010年度についても同様に、福島県の広野町、楢葉町、富岡町、川内村、双葉町、新地町のレコードを削除した。また、2015年度の福島県大熊町の小売店数、飲食店数、大型小売店数、百貨店・総合スーパー数が欠損値となっていたため0とした。

特徴を指摘した地域とそれ以外の地域に分割したうえで、それぞれの地域の目的変数の平均値を算定し、平均値の差異に対して Welch の t 検定を行う。

これにより介護格差の地域特性の視覚情報を獲得したうえで、これらの情報を記述統計で数値化して整理する。視覚情報および記述統計は、機械学習やクラスター分析で得られた地域分類の結果と合わせて分析することで、分析結果の妥当性を確認することに用いる。

3.3 機械学習による説明変数の選択

機械学習の手法を用いて、標準化要介護認定比（軽度）・標準化要介護認定比（重度）を目的変数とする説明変数の候補 79 に対して選択を行う。機械学習の手法として、本稿ではランダムフォレストを使用する⁸。ランダムフォレストはデータをランダムに抽出するアルゴリズムのため、得られる結果も試行の都度異なる。結果の安定性を確保するために、10 回の試行の平均値を採用した⁹。すなわち、各回の試行で得られた %IncMSE、IncNodePurity の結果をそれぞれ標準化し、計 20 の平均値により上位層を採用した。

ランダムフォレストは、学習データの中から復元抽出でランダムにサンプリングしたレコードに対して決定木を複数作り、それらの決定木の予測値の平均を全体の予測値とするモデルである。%IncMSE は、説明変数を除いた場合にモデルの予測精度（平均 2 乗誤差）がどれだけ悪化するかに着目し、その評価をもとに算定された重要度となっている。IncNodePurity は、決定木の分割時にノードの不純度（残差平方和）がどれだけ改善するかに着目し、その評価をもとに算定された重要度となっている。

この他にも、機械学習の手法としては、サポートベクターマシン（SVM）やニューラルネットワークなどがある。しかし、これらの手法は説明変数の選択結果の解釈がブラックボックスであるため、本稿では指標重要度を用いて説明変数の選択が行われることが明瞭な決定木をベースとするモデルを採用した。なお、非線形性があり、誤差の大きいデータに関しては、ランダムフォレストなどの機械学習を行うことで高い予測力を発揮できる場合があることは知られている。

ここで、主成分分析とランダムフォレストの違いは、目的変数に応じた分析ができるか否かである。主成分分析は、複数の目的変数が存在する場合、目的変数ごとに説明変数の使い分けはできず、目的変数とは無関係に説明変数をすべて用いて分析を行ってしまう¹⁰。これに対し、機械学習は目的変数に応じた説明変数の選択が可能である。

3.4 クラスター分析による地域分類

本稿の目的は地域間の介護格差の具体的イメージを持つことなので、要因分析の手法である回帰分析ではなくクラスター分析の手法を用いる。具体的には、ランダムフォレストにより選択した上位層の説明変数に対し、非階層型クラスター分析の手法である k¹¹平均法により市区町村を分類する。

クラスター分析には樹形図状の階層構造を作りながらグルーピングする階層クラスター分析と階層構造を作らない非階層クラスター分析がある。k 平均法は、①データを k 個のクラスターに分け、②各クラスターに割り当てられたデータの重心を計算し、③データと重心との距離を用いて距離が一番近いクラスターに割り当て直す調整を繰り返す。

⁸ 分析は統計ソフトウェア R を用いた。R のバージョンは 4.0.2 を使い、ランダムフォレストは randomForest パッケージの randomForest 関数を用いて算定した。

⁹ 10 回の試行の平均値を採用したが、各試行の結果は安定していることを確認した。

¹⁰ 要介護認定率（軽度）と要介護認定率（重度）のそれぞれについて認定率の高い順に並べたデータを 10 分割し、最上位地域と最下位地域を抽出し、目的変数に応じた主成分分析を別途実施した。その結果は粒度が粗く、軽度と重度とで差異はみられず、ランダムフォレストで得られた詳細な結果は得られなかった。

¹¹ 同様に分析は統計ソフトウェア R を使い、k 平均法は cluster パッケージの kmeans 関数を用いて算定した。

返して、データを分類するアルゴリズムである。k 平均法は階層型構造を持たないデータに対しても有効であり、近年ビッグデータに対しての活用が広がっているクラスタリングの手法である¹²。

4 地理的可視化による分析および記述統計

本章では 3.2 に示した地理的可視化による分析を行う。目的変数の標準化要介護認定比（軽度）・標準化要介護認定比（重度）について、2015 年度データで、それぞれ比率の高い地域、低い地域をヒートマップで示したものが図 1 である。色が濃い地域ほど比率が高く、色が薄いほど比率が低い地域である。図 1 から次の視覚情報が得られる。

視覚情報 1：標準化要介護認定比（軽度）について、比率の高い地域は、大阪や東京などの大都市部で、関西から西の地域が多く、さらに内陸部よりも沿岸地域の方が多い。

視覚情報 2：標準化要介護認定比（軽度）について、比率の低い地域は、内陸部が多く、特に関東地方、中部地方の内陸部で多い。

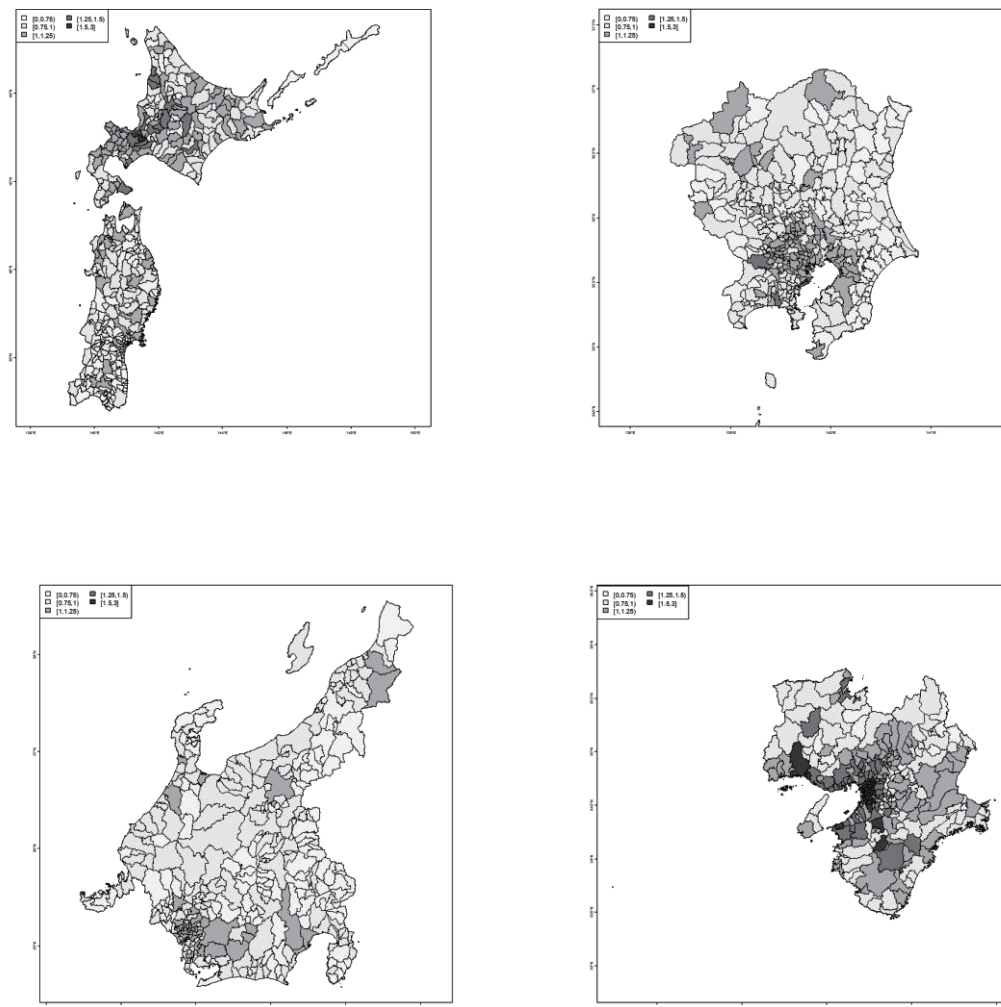
視覚情報 3：標準化要介護認定比（重度）について、大阪や東京などの大都市部で比率が高い。東北地方でも比率が高い地域が一部存在する。

視覚情報 4：標準化要介護認定比（重度）について、北海道地方では比率が低い地域が多い。

¹² k 平均法も試行の都度結果は異なる。そこで試行結果が安定する k-medoids 法により別途分析を行い、k 平均法で得られた結果と大きな差異がないことを確認した。

図1 標準化要介護認定比（軽度）・標準化要介護認定比（重度）に関するヒートマップ¹³

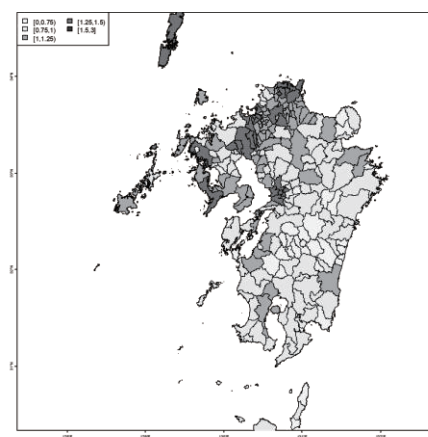
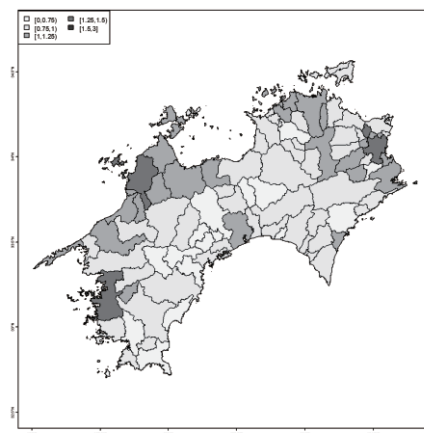
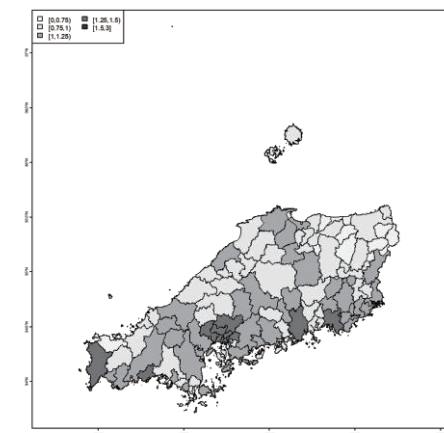
<標準化要介護認定比（軽度）>



地図データ：(c)Esri Japan

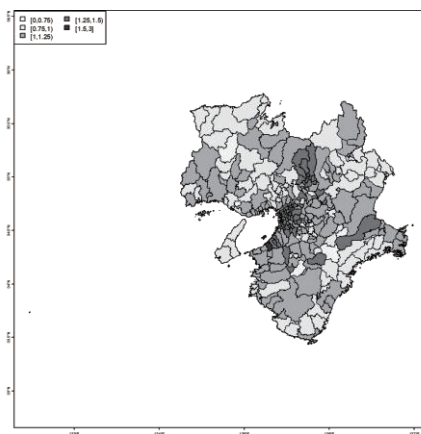
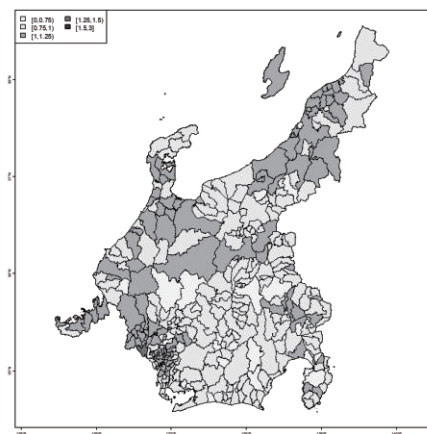
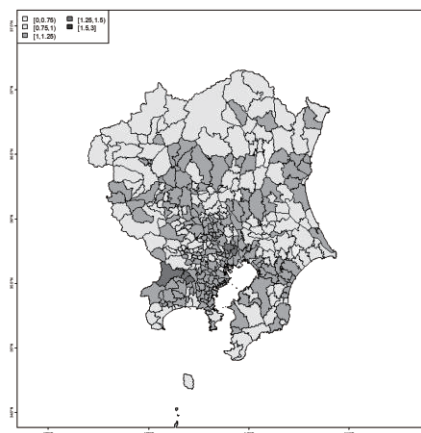
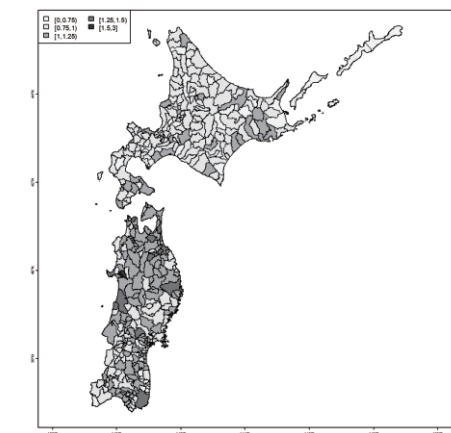
¹³ ヒートマップは、国土地理院発行の数値地図（国土基本情報）と ESRI ジャパンの全国市区町村界データを使用し加工した。

<標準化要介護認定比（軽度）>



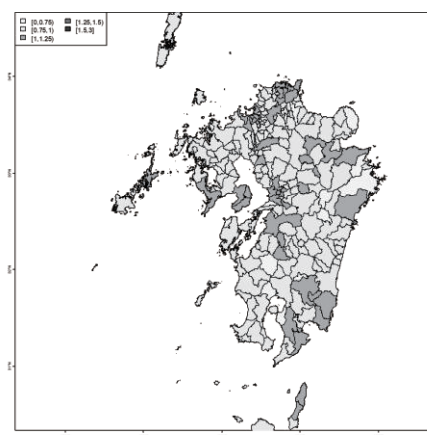
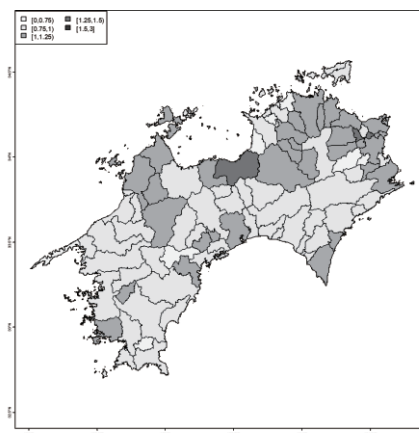
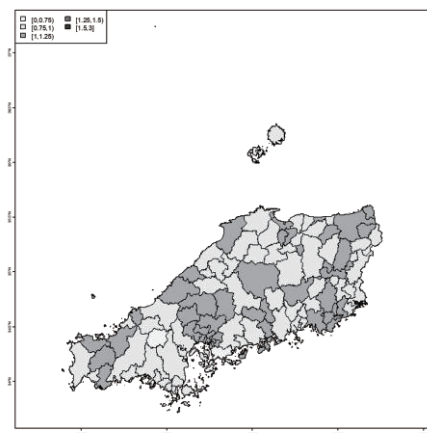
地図データ：(c)Esri Japan

<標準化要介護認定比（重度）>



地図データ：(c)Esri Japan

<標準化要介護認定比（重度）>



地図データ：(c)Esri Japan

視覚情報として列記した5つの特徴について、Welch の t 検定を行った結果を表 2 に示した。標準化要介護認定比（軽度）は大都市圏・都市圏や西日本、沿岸地域で高く、振興山村地域や過疎地域で低い傾向にある。また、標準化要介護認定比（重度）は大都市圏・都市圏で高く、振興山村地域や過疎地域で低い傾向にあることが確認できる。なお、標準化要介護認定比（重度）については、西日本地域および沿岸地域での有意差は確認できない。

ここに、大都市圏・都市圏は、総務省統計局が国勢調査において定義している統計上の地域区分で、大都市圏は中心市及び周辺市町村によって構成され、大都市圏の中心市は、東京都特別区部及び政令指定市である。都市圏は中心市及び周辺市町村によって構成され、都市圏の中心市は、大都市圏に含まれない人口 50 万人以上の市であり、2015 年度は宇都宮市、松山市、鹿児島市である。

また、西日本は、近畿地方より西の地域とした。振興山村地域は山村振興法で指定された地域で、昭和 25 年時点の旧市町村単位で昭和 35 年時点で林野率 75%以上かつ人口密度が一定基準未満の地域である。過疎地域は人口減少が著しく地域社会における活力が低下している地域で、総務省が過疎地域自立促進特別措置法により指定している地域である。

表 2 標準化要介護認定比（軽度）・標準化要介護認定比（重度）の可視化情報による分析

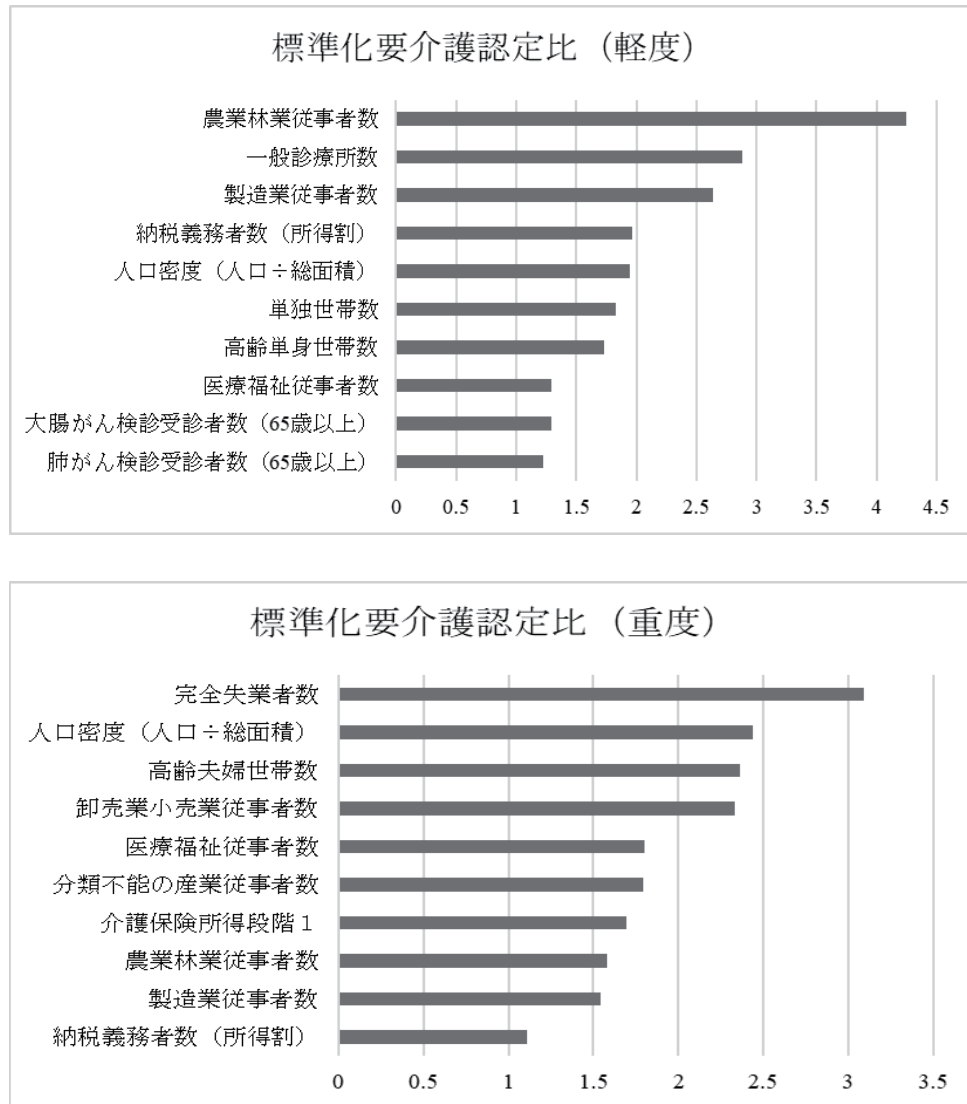
	標準化要介護認定比 (軽度)	Welch 検定			標準化要介護認定比 (重度)	Welch 検定		
①大都市圏・都市圏	1.131				1.079			
②上記以外	0.881	t 値	-10.021		0.960	t 値	-6.593	
①/②	128.4%	p 値	0.0000	***	112.3%	p 値	0.0000	***
③西日本	0.967				0.970			
④上記以外	0.839	t 値	-11.398		0.960	t 値	-1.263	
③/④	115.3%	p 値	0.0000	***	101.0%	p 値	0.2067	
⑤沿岸地域	0.933				0.968			
⑥上記以外	0.863	t 値	-6.088		0.961	t 値	-0.810	
⑤/⑥	108.1%	p 値	0.0000	***	100.7%	p 値	0.4184	
⑦振興山村地域	0.817				0.893			
⑧上記以外	0.898	t 値	4.379		0.973	t 値	6.135	
⑦/⑧	91.0%	p 値	0.0000	***	91.8%	p 値	0.0000	***
⑨過疎地域	0.843				0.931			
⑩上記以外	0.915	t 値	6.282		0.983	t 値	6.252	
⑨/⑩	92.2%	p 値	0.0000	***	94.7%	p 値	0.0000	***

有意水準 *** (0.1%), ** (1%), * (5%)

5 機械学習による説明変数の選択

本章では3.3のランダムフォレストによる説明変数の選択を行う。ランダムフォレストの標準化要介護認定比（軽度）・標準化要介護認定比（重度）の結果は、図2のとおりである。横軸は重要度を示し、上から順番に重要度の高い説明変数から上位10を並べている。先行研究の安藤（2007）や久保寺（2013）などで分析対象とされた高齢者単身世帯、高齢夫婦世帯、人口密度、第一次産業や第二次産業に関する説明変数は、本稿の結果でも同様に選択されている。また、ランダムフォレストの標準化要介護認定比（軽度）・標準化要介護認定比（重度）の結果において、共通して重要度の高い説明変数は、人口密度、医療福祉従業者数、製造業従業者数、納税義務者数となっている。このうち特に人口密度が軽度・重度に共通して重要度が高いことは注目すべき点である。人口密度は高齢者が都市部で生活しているか、地方部で生活しているかを表している。4章で示したように、標準化要介護認定比は軽度・重度ともに大都市圏・都市圏で高く、振興山村地域や過疎地域で低い傾向にある。ランダムフォレストの結果は、4章で得られた情報と整合したものになっている。

図2 ランダムフォレストによる説明変数の一次選択の結果



6 クラスター分析による地域分類

本章では3.4のk平均法による地域分類を行う。まず、k平均法で使用する説明変数を選択する。図2が示す重要度の大きさは、標準化要介護認定比（軽度）では重要度が2以上の説明変数は3つ、1.5以上は7つある。また、グラフから見て説明変数の重要度の長さに不連続な段差が、1, 3, 7番目の説明変数の後にある。そこで、3番目の段差となる重要度が1.5以上の7番目までの説明変数を採用する。なお、上位3または上位7の選択肢があるが、地域分類を行う際に3つの説明変数では十分な分類ができない可能性があるため上位7つとした。

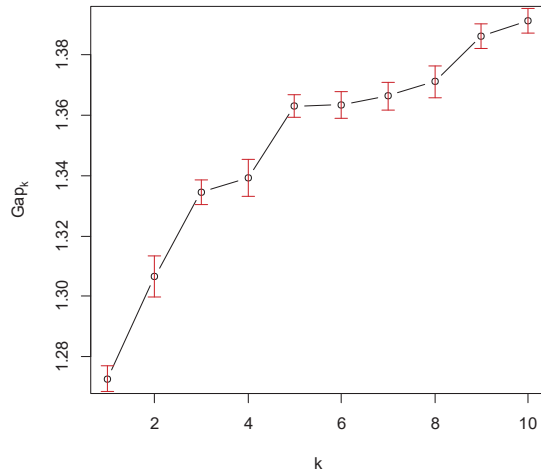
一方、標準化要介護認定比（重度）では重要度が2以上は4つ、1.5以上は9つある。説明変数の重要度の長さの不連続な段差は、1, 4, 9番目の説明変数の後にある。そこで、最初の段差となる重要度が2以上の4番目までの説明変数を採用する。

図2の棒グラフの段差から選択した説明変数を用いてk平均法を実施し地域を分類するが、k平均法は予めクラスター数（分類の数）を指定する必要がある。本稿ではGap統計量を用いて、最適なクラスター数を定める。Gap統計量とは、クラスター数kに対するk平均法で作成したクラスターの評価関数を分母に、クラスター数kに対する一様乱数により作成したクラスターの評価関数を分子とした指標であり、クラスター数kを1つずつ増やしていき、分母の評価関数が小さくなるほどすなわちGap統計量が大きいくほど精度が良いことを意味する。

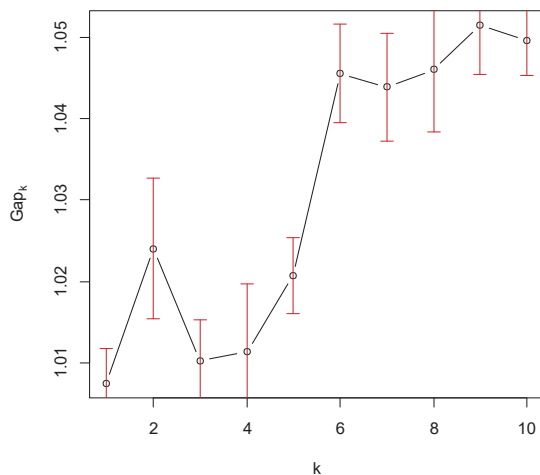
Gap統計量を作成したものが図3である。図3より標準化要介護認定比（軽度）ではクラスター数k=5までは急速にGap統計量が増加し、k=5以降はなだらかに増加していることがわかる。標準化要介護認定比（重度）ではクラスター数k=2のとき山となり、k=3で減少し以降k=6までは急速にGap統計量が増加し、k=6以降はなだらかに増加していることがわかる。よって、クラスター数は、標準化要介護認定比（軽度）は5、標準化要介護認定比（重度）は6を採用する。

図3 Gap 統計量によるクラスター数の決定

<標準化要介護認定比（軽度）>



<標準化要介護認定比（重度）>



上記方法でクラスター数を定めた後に k 平均法を実施し得られた結果が表3である。定めたクラスター数により、標準化要介護認定比（軽度）は5つに、標準化要介護認定比（重度）は6つに分類される。そして、平均標準比介護認定比の最も高い地域1から降順に並べた。表3前半の（各地域における説明変数の平均）では、それぞれの地域に分類される市区町村数と平均標準化要介護認定比、および各説明変数の平均値を記載した。表3後半の（各地域における説明変数のランキング）では、それぞれの説明変数の平均値が高い地域から順に1位、2位とランキングを付番した。最後に、各地域における説明変数のランキングを直感的に理解するためにレーダーチャートを掲載した。

標準比介護認定比（軽度）が最も高い地域1を考察すると、高所得者が多い、人口密度は高い、単独世帯割合が多い、農業従事者数が少ない傾向にある。低い地域5では、製造業従事者数が多い、一般診療所数は少ない、単独世帯・高齢者単身世帯割合が少ない傾向にある。また、やや低い地域4は農業従事者数が多い傾向にある。

標準比介護認定比（重度）が最も高い地域1を考察すると、人口密度が高い、高齢夫婦世帯割合は低い傾向にある。次に高い地域2は完全失業者数が多い傾向にある。低い地域6は、完全失業者数は低い、人口密度は低い、高齢夫婦世帯割合は高い傾向にある。

表3 k 平均法によるクラスター分析

<標準化要介護認定比（軽度）>

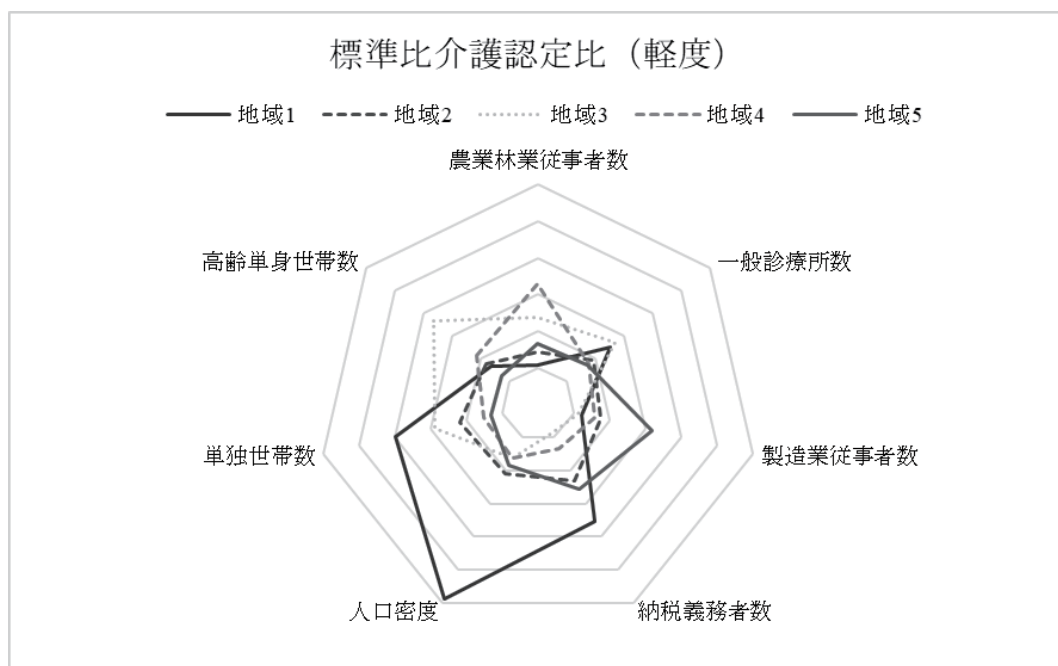
■各地域における説明変数の平均

	地域1	地域2	地域3	地域4	地域5
地域数	137	991	517	608	899
平均標準比介護認定比（軽度）	1.06	0.99	0.91	0.85	0.80
農業林業従事者数	0.0022	0.0197	0.0687	0.1152	0.0317
一般診療所数	0.0011	0.0007	0.0011	0.0006	0.0006
製造業従事者数	0.0494	0.0708	0.0446	0.0642	0.1304
納税義務者数（所得割）	0.4824	0.4215	0.3523	0.3733	0.4348
人口密度（人口/総面積）	11,054	1,305	84	99	640
単独世帯数	0.4285	0.2880	0.3419	0.2341	0.2195
高齢単身世帯数	0.1003	0.1062	0.1870	0.1217	0.0828

■各地域における説明変数のランキング

	地域1	地域2	地域3	地域4	地域5
平均標準比介護認定比（軽度）	1	2	3	4	5
農業林業従事者数	5	4	2	1	3
一般診療所数	2	3	1	4	5
製造業従事者数	4	2	5	3	1
納税義務者数（所得割）	1	3	5	4	2
人口密度（人口/総面積）	1	2	5	4	3
単独世帯数	1	3	2	4	5
高齢単身世帯数	4	3	1	2	5

■レーダーチャート（レーダーチャートの項目は5章の説明変数の選択で得られたもの）



<標準化要介護認定比（重度）>

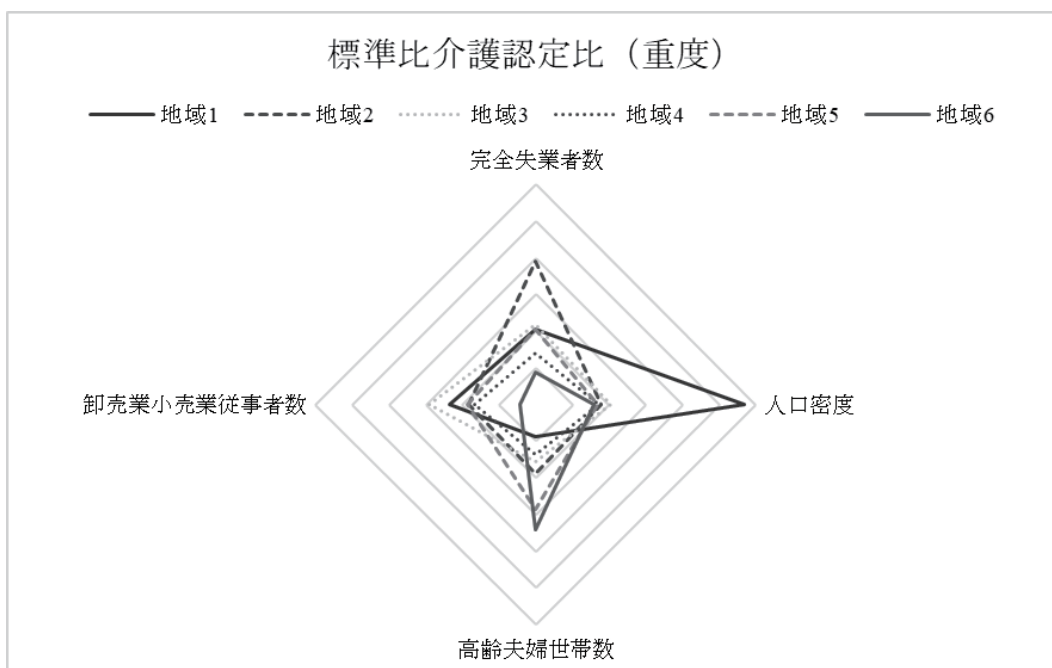
■各地域における説明変数の平均

	地域 1	地域 2	地域 3	地域 4	地域 5	地域 6
地域数	161	305	972	665	684	365
平均標準比介護認定比（重度）	1.05	1.03	0.99	0.95	0.95	0.88
完全失業者数	0.0249	0.0380	0.0255	0.0203	0.0246	0.0164
人口密度（人口/総面積）	10,508	574	1,209	376	267	31
高齢夫婦世帯数	0.0907	0.1283	0.1154	0.1074	0.1638	0.1849
卸売業小売業従事者数	0.0703	0.0634	0.0777	0.0614	0.0638	0.0449

■各地域における説明変数のランキング

	地域 1	地域 2	地域 3	地域 4	地域 5	地域 6
平均標準比介護認定比（重度）	1	2	3	4	5	6
完全失業者数	3	1	2	5	4	6
人口密度（人口/総面積）	1	3	2	4	5	6
高齢夫婦世帯数	6	3	4	5	2	1
卸売業小売業従事者数	2	4	1	5	3	6

■レーダーチャート（レーダーチャートの項目は5章の説明変数の選択で得られたもの）



7 結果の分析

本章では、これまでの分析によって得られた知見（表 4 参照）を考察する。

表 4 「地理的可視化による分析」の視覚情報と「地域分類による分析」の結果の照合

		地理的可視化による分析	地域分類による分析	照合
標準化要介護認定比（軽度）	比率の高い地域	視覚情報 1： 大阪や東京などの大都市部、 関西から西の地域で内陸部 よりも沿岸地域。	大都市圏・都市圏に該当するのは 41%、 大都市圏・都市圏およびその周辺を加え ると 98% 沿岸地域に該当するのは 50%、西日本に 該当するのは 47%	視覚情報と 整合的
	比率の低い地域	視覚情報 2： 内陸部（特に関東地方、中部 地方の内陸部）	愛知・滋賀などの企業城下町が多い 内陸部が多い（特に関東地方、中部地方 の内陸部） 全域が振興山村地域に該当するのは 23%、 一部が振興山村地域に該当するの は 65%、過疎地域に該当するのは 69%、 農林業就業者数が全国平均を上回る地 域は 97%	新しい知見が 得られた 視覚情報と 整合的
標準化要介護認定比（重度）	比率の高い地域	視覚情報 3： 大阪や東京などの大都市部、 東北地方でも一部存在する。	大都市圏・都市圏に該当するのは 32%、 大都市圏・都市圏およびその周辺に該当 するのは 98%	視覚情報と 整合的
	比率の低い地域		青森県、徳島県、高知県、沖縄県に該当 するのは 34%、完全失業率が全国平均を 上回る地域は 100%	視覚情報と 整合的 新しい知見も 得られた
	比率の低い地域	視覚情報 4： 北海道地方が多い。	小さな町や村、島が多く、農業や酪農、 漁業が盛んな地域 全域が振興山村地域に該当するのは 52%、 一部が振興山村地域に該当するの は 74%、過疎地域に該当するのは 97%、 農林業就業者数が全国平均を上回る地 域は 81%、北海道または長野県に該当 する地域は 47%	視覚情報と 整合的 新しい知見も 得られた

7.1 標準化要介護認定比（軽度）

7.1.1 高い地域

標準化要介護認定比（軽度）について、最も高い地域 1 は、納税義務者数（所得割）、人口密度、単独世帯数がランキング 1 位である。納税義務者が多く、人口密度が高く、単身者が多く生活している地域は都市部が多いと考えられる。具体的な市区町村を確認すると、地域 1 には東京都特別区、大阪府大阪市、福岡県福岡市、その周辺の人口の多い地域である神奈川県横浜市、神奈川県川崎市、千葉縣市川市、千葉県船橋市、千葉県浦安市、埼玉県川口市、埼玉県戸田市、埼玉県草加市、大阪府東大阪市、大阪府豊中市、大阪府吹田市などが分類されている。4 章で得られた視覚情報 1 の大都市で比率の高いことと整合的であり、k 平均法で得られた地域分類の結果が妥当性であると言える。なお、地域 1 に分類された市区町村のうち大都市圏・都市圏に該当するのは 41%、大都市圏・都市圏にその周辺を加えると 98%となっている。

比率のやや高い地域 2 には千葉県市原市、神奈川県藤沢市、神奈川県茅ヶ崎市、三重県伊勢市、大阪府貝塚市、大阪府忠岡町、兵庫県姫路市、兵庫県明石市、和歌山県海南市、岡山県倉敷市、広島県福山市、広島県海田町、山口県防府市、山口県光市、香川県坂出市、愛媛県松前町、福島県行橋市などが分類されている。海岸沿いの工場地帯が多く、ベッドタウン¹⁴となっている地域も多い。古くから工業都市である地域もあり、高齢化が進んでいる一方都市への交通アクセスの便利さから若い世帯も共存している地域が多い。地域 1 の都市圏からの人口密度の連続体として地域 2 を捉えれば、軽度の要介護者の比率が高い地域であるということも納得感がある。地域 2 は西日本の沿岸地域が多く分類されていることも確認できる。4 章で得られた視覚情報 1 の関西から西の地域が多く、さらに内陸部よりも沿岸地域の方が多いことと整合的である。なお、地域 2 に分類された市区町村のうち沿岸地域に該当するのは 50%、前述との重複も含め西日本に該当するのは 47%となっている。

7.1.2 低い地域

比率の低い地域 5 には愛知県豊田市、愛知県刈谷市、愛知県安城市、愛知県西尾市、滋賀県甲賀市、滋賀県竜王町、滋賀県日野町が分類されている。地域 2 と同じく工場地帯が多く分類されているが、地域 2 との違いは愛知・滋賀などの企業城下町が多く、若い世帯・外国人が多い地域という特徴がある。愛知県・滋賀県の企業城下町以外で地域 5 に分類されている地域は、製造業以外に農業も盛んな地域が多い。また、日本全国において企業城下町の地域とそれ以外の地域とで軽度介護者認定率を比較すると、企業城下町の方が高い傾向にあることが確認され、必ずしも企業城下町だから低いということではなく、愛知県・滋賀県の企業城下町が低い傾向にあるということだと分析される。また、地域 2 の工場地帯と比べると、地域 2 よりも人口密度が低く田舎的だとも言えるが、納税義務者数の割合は地域 5 の方が多く裕福な地域であるといった特徴がある。これらの内容は 4 章の可視化では得られなかった知見である。

比率のやや低い地域 4 には群馬県嬬恋村、群馬県昭和村、長野県川上村、青森県新郷村、徳島県佐那河内村、熊本県産山村などが分類されている。高齢者が多い地域ではあるが山地で農業が盛んな地域が多い。4 章で得られた視覚情報 2 の内陸部が多く、特に関東地方、中部地方の内陸部で多いことと整合的である。なお、地域 4 に分類された市区町村のうち全域が振興山村地域に該当するのは 23%、一部が振興山村地域に該当するのは 65%、前述との重複も含め過疎地域に該当するのは 69%、農林業就業者数が全国平均を上回る地域は 97%となっている。

7.2 標準化要介護認定比（重度）

7.2.1 高い地域

標準化要介護認定比（重度）について具体的な市区町村を確認すると、比率の高い地域 1 には東京都特別区、大阪府大阪市、その周辺の人口の多い地域である神奈川県横浜市、神奈川県川崎市、千葉県市川市、千葉県船橋市、千葉県浦安市、埼玉県川口市、埼玉県戸田市、埼玉県草加市、大阪府東大阪市、大阪府豊中市、大阪府吹田市などが分類されている。これらの地域は高齢単身者が多い地域というのも特徴であり、千葉県浦安市と埼玉県和光市以外の地域では高齢単身世帯割合の方が高齢夫婦世帯割合よりも多くなっている。比率が高い地域に都市部が多い傾向は標準化要介護認定比（軽度）と同じであり、軽度の要介護状態の認定率が高いだけでなく、重度の要介護状態の認定率も高い結果となっている。4 章で得られた視覚情報 3 の大都市およびその周辺で比率の高いことと整合的であり、k 平均法で得られた地域分類の結果が妥当性であると言える。なお、地域 1 に分類された市区町村のうち大都市圏・都市

¹⁴ 工場地帯、ベッドタウンには特段の定義はないが、一般に使用されている用語であり地域イメージを具体化するために使用する。本章において、同様の目的でいくつかの用語を使用している。

圏に該当するのは32%、大都市圏・都市圏にその周辺を加えると98%となっている。

比率の高い地域2には青森県青森市、青森県八戸市、徳島県阿南市、徳島県鳴門市、高知県四万十市、高知県土佐市、沖縄県沖縄市、沖縄県うるま市、沖縄県浦添市などが分類されている。青森県、徳島県、高知県、沖縄県の地域が多く、これらの地域は完全失業率が高い。4章で得られた視覚情報3の東北地方で比率が高い地域が一部あると整合的である。しかし、徳島県、高知県、沖縄県については可視化では得られていなかった知見である。なお、地域2に分類された市区町村のうち青森県、徳島県、高知県、沖縄県に該当するのは34%、前述との重複も含め完全失業率が全国平均を上回る地域は100%となっている。

7.2.2 低い地域

比率の低い地域6には北海道南富良野町、北海道士幌町、長野県川上村、長野県南相木村、奈良県十津川村、奈良県東吉野村、宮崎県西米良村などが分類されている。小さな町や村、島が多く、農業や酪農、漁業が盛んな地域が多い傾向が見られる。人口が少なく高齢化が進み働き手を外から募集している地域も多いことが特徴である。特に北海道や長野県の地域が多く分類されている。また、軽度と同様に農業が盛んな地域では要介護者認定率が低いことが示されたが、重度の方がより山地でより過疎化が進んだ地域の傾向にある。例えば、軽度が少ない分類には、青森県つがる市、宮城県蔵王町、福島県喜多方市、高知県土佐市など比較的人口が多い地域が含まれているが、重度が少ない分類には含まれず、東京都新島村、広島県江田島市、大分県姫島村、香川県直島町、鹿児島県三島村など軽度には含まれていない離島の地域が含まれている。4章で得られた視覚情報4の北海道地方で比率が低いことと整合的である。しかし、小さな町、村、島でかつ農業や酪農、漁業が盛んな地域であることは可視化では得られていなかった知見である。なお、地域6に分類された市区町村のうち全域が振興山村地域に該当するのは52%、一部が振興山村地域に該当するのは74%、前述との重複も含め過疎地域に該当するのは97%、農林業就業者数が全国平均を上回る地域は81%、北海道または長野県に該当する地域は47%となっている。

8 さいごに

本稿では、地域ごとに異なる年齢構成比、男女比の調整を行い、比較可能なものに指標化を行ったうえで、要介護認定率の高い地域と低い地域の具体的なイメージを示すことができた。特に、軽度および重度の要介護認定率については、大都市圏やその周辺都市で高い傾向であることを確認した。年齢構成比および男女比の調整を行った上で、都道府県単位で比較分析し大阪府が高いことを示した先行研究はあるものの、市区町村単位で分析した結果、大都市圏とその周辺で要介護認定率が高い傾向にあるという具体化されたイメージの知見が得られた。

さらに、軽度の要介護認定率については、西日本の海岸沿いの工場地帯において認定率が高い傾向にあり、愛知県や滋賀県の企業城下町や農業従事者が多い山村の農村部においては低い傾向にあることを確認した。また、重度の要介護認定率については、完全失業率の高い地域である青森県、徳島県、高知県、沖縄県で認定率が高い傾向にあり、完全失業率の低く農業や酪農、漁業が盛んな小さな町・村・島では認定率が低い傾向を確認した。なお、農業が盛んな地域では軽度・重度ともに要介護者認定率が低いことが示されたが、重度の方がより山地でより過疎化が進んだ地域傾向があることを確認した。これらの結果から、要介護認定率について軽度・重度ともに人口密度がひとつの重要な因子であることが伺える。

しかし、標準化要介護認定比（軽度）での一般診療所数や標準化要介護認定比（重度）での卸売業小売業従事者数などは要介護認定率に直接影響を及ぼすような因子とまでは断言できず、むしろ別の要因の代理変数である可能性

が高いと思われる。分類方法が妥当であるかの議論は残っていると考える。介護格差の地域分類の妥当性は引き続き検討課題であろう。また、介護格差に係る地域分類の結果をもとに、どのような施策を行うべきかという検討には至っていない。都市部の要介護認定率の引き下げのために、人口密度を引き下げることが有効な対策でないことは自明である。

このような課題はあるものの、先行研究では得られていない知見を提示することができた。本研究で得られた介護格差の分析結果が、介護格差の是正および公的介護保険制度の持続可能性を高めるための一助となることを祈念する。

別表 1 目的変数・説明変数一覧

項目	目的変数	説明変数	出典	表番号
(要支援 1・要支援 2・要介護 1) 認定者数男女計	○		介護保険事業状況報告 (厚生労働省)	第 4-1-1 表
(要介護 2・要介護 3・要介護 4・要介護 5) 認定者数男女計	○			
所得段階 (2015 年：最低 1 と最高 9, 2010 年：最低 1 と最高 6)		○	介護保険事業状況報告 (厚生労働省)	第 17 表
所得平均		○		
平均余命 (男性)		○	市区町村別生命表 (厚生労働省)	統計表 1
平均余命 (女性)		○		
外国人人口 (対人口)		○	国勢調査 (総務省統計局)	
出生数 (対人口)		○	人口動態調査 (厚生労働省)	
死亡数 (対人口)		○		
転入者数 (対人口)		○	住民基本台帳人口移動報告年報 (総務省統計局)	
昼間人口 (対人口)		○	国勢調査 (総務省統計局)	
核家族世帯数 (対世帯数)		○		
単独世帯数 (対世帯数)		○		
高齢夫婦世帯数 (対世帯数)		○		
高齢単身世帯数 (対世帯数)		○		
婚姻件数 (対人口)		○	人口動態調査 (厚生労働省)	
離婚件数 (対人口)		○		
人口密度 (=人口/総面積(北方地域及び竹島を除く))		○	全国都道府県市区町村別面積調 (国土交通省国土地理院)	
納税義務者数 (所得割) (対人口)		○	市町村税課税状況等の調 (総務省)	
実質公債費比率(市町村財政)		○	市町村別決算状況調 (総務省)	
歳出決算総額(市町村財政) (対人口)		○		
地方税(市町村財政) (対人口)		○		
幼稚園数 (対人口)		○	学校基本調査 (文部科学省)	
中学校数 (対人口)		○		
高等学校数 (対人口)		○		
完全失業者数 (対人口)		○	国勢調査 (総務省統計局)	
役員数 (対人口)		○		
公民館数 (対人口)		○	社会教育調査 (文部科学省)	
図書館数 (対人口)		○		
非水洗化人口 (対人口)		○	一般廃棄物処理事業実態調査 (環境省)	
ごみ計画収集人口 (対人口)		○		
小売店数 (対人口)		○	経済センサス基礎調査 (総務省統計局)	
飲食店数 (対人口)		○		
大型小売店数 (対人口)		○		
百貨店、総合スーパー数 (対人口)		○		
一般病院数 (対人口)		○	医療施設調査 (厚生労働省)	
一般診療所数 (対人口)		○		
歯科診療所数 (対人口)		○		
児童福祉施設等数 (助産施設・児童遊園を除く) (対人口)		○	社会福祉施設等調査 (厚生労働省)	
保育所等数 (対人口)		○		
産業別就業者数_農業、林業 (対人口)		○	国勢調査 (総務省統計局)	
産業別就業者数_漁業 (対人口)		○		
産業別就業者数_鉱業、採石業、砂利採取業 (対人口)		○		
産業別就業者数_建設業 (対人口)		○		
産業別就業者数_製造業 (対人口)		○		
産業別就業者数_電気・ガス・熱供給・水道業 (対人口)		○		
産業別就業者数_運輸業、郵便業 (対人口)		○	国勢調査 (総務省統計局)	
産業別就業者数_卸売業、小売業 (対人口)		○		
産業別就業者数_金融業、保険業 (対人口)		○		
産業別就業者数_不動産業、物品賃貸業 (対人口)		○		
産業別就業者数_学術研究、専門・技術サービス業 (対人口)		○		

項目	目的 変数	説明 変数	出典	表番号
産業別就業者数_宿泊業、飲食サービス業（対人口）		○	国勢調査 （総務省統計局）	
産業別就業者数_生活関連サービス業、娯楽業（対人口）		○		
産業別就業者数_教育、学習支援業（対人口）		○		
産業別就業者数_医療、福祉（対人口）		○		
産業別就業者数_複合サービス事業（対人口）		○		
産業別就業者数_サービス業（対人口）		○		
産業別就業者数_公務（対人口）		○		
産業別就業者数_分類不能の産業（対人口）		○		
健康増進に関する会議の参加機関・団体数（対人口）		○	地域保健・健康増進事業報告 （厚生労働省） 地域保健編	第 3 5 表
集団健康教育の参加延人員（対人口）		○	地域保健・健康増進事業報告 （厚生労働省） 健康増進編	第 0 4 表
健康診査受診者数（65 歳以上）（対 65 歳以上人口）		○		第 0 6 0 1 表
歯周疾患検診受診者数（65 歳以上）（対 65 歳以上人口）		○		第 1 0 表
骨粗鬆症検診受診者数（65 歳以上）（対 65 歳以上人口）		○		第 1 1 表
訪問指導従事者延人員 保健師（対人口）		○		第 1 5 表
訪問指導従事者延人員 管理栄養士及び栄養士（対人口）		○		第 1 5 表
胃がん検診受診者数 胃部エックス線検査（65 歳以上） （対 65 歳以上人口）		○		第 1 6 0 1 表
肺がん検診受診者数 胸部エックス線検査（65 歳以上） （対 65 歳以上人口）		○		第 1 7 0 1 表
大腸がん検診受診者数（65 歳以上） （対 65 歳以上人口）		○		第 1 8 0 1 表
国民年金割合（対年金受給者数）		○	市町村別状況 （厚生労働省）	
地方債（対人口）		○	市町村別決算状況調 （総務省）	
社会福祉費（対人口）		○		
老人福祉費（対人口）		○		
保健衛生費（対人口）		○		
自殺者数（対人口）		○	自殺の統計（内閣府）	
平均課税年収		○	市町村税課税状況等の調 （総務省）	
交通事故件数（対人口）		○	交通事故統計年報 （交通事故総合分析センター）	
介護老人福祉施設数・介護老人保健施設数・介護療養型医療施設数 の合計		○	介護サービス施設・事業所調査 （厚生労働省）	基本票第 1 表
平均年齢		○	国勢調査 （総務省統計局）	
人口増減率		○		

参考文献

- [1] 安藤道人 (2007) 「介護給付水準と介護保険料の地域差の実証分析—保険者データを用いた分析—」『季刊社会保障研究』 Vol.44 No.1 pp.94-109
- [2] 石橋未来, 鈴木隼 (2019) 「「令和」時代の介護, 地域差と要介護女性に視線を注げ」『大和総研調査季報』 春季号 Vol.34, 2019 年 pp.64-81
- [3] 岩沢宏和, 平松雄司(2019)「EDA (探索的データ解析)」『入門 R による予測モデリング—機械学習を用いたリスク管理のために』 東京図書 pp.46-62
- [4] 小内透 (1996)『戦後日本の地域社会変動と地域社会類型—都道府県・市町村を単位とする統計分析を通して』 東信堂
- [5] 久保寺重行 (2013) 「介護サービス需要行動に関する実証分析 —今後の介護保険制度改革に向けて—」 社会福祉学 第 54 巻第 2 号, 2013 年 6 月 pp.70-82
- [6] 近藤克則・芦田登代・平井寛・三澤仁平・鈴木佳代 (2012) 「高齢者における所得・教育年数別の死亡・要介護認定率とその性差—AGES プロジェクト縦断研究—」『医療と社会』 Vol.22 No.1, 2012 pp.19-30
- [7] 齋藤香里 駒村康平編著 (2018) 「介護と貧困」『貧困』 ミネルヴァ書房 pp.130-142
- [8] 佐藤哲彰 (2016) 「要介護状態の発生率は, 所得水準によってどう異なるか—ロジスティック回帰による分析—」『千葉商大紀要』 第 53 巻第 2 号, 2016 年 3 月 pp.93-106
- [9] 須田茂夫 (2018) 「機械学習による都道府県別医療費の分析」『社会保障研究』 Vol.3 No.3, 2018 (pp.403-415)
- [10] 高間康史, 狩野真次 (2011) 「比較分析に着目した時空間的動向情報の探索的分析支援」『人工知能学会論文誌』 26 巻 4 号 pp.494-503
- [11] 内閣府政策統括官 (2018) 「要介護 (要支援) 認定率の地域差要因 に関する分析」 内閣府政策統括官 (経済財政分析担当)
- [12] 蓮見音彦 (2016)『現代日本の地域格差—2010 年・全国の市町村の経済的・社会的ちらばり—』 東信堂
- [13] 飛田英子 (2015) 「介護費用の適正化に向けた課題—保険者データの分析を踏まえて—」 JRI レビュー 2015 Vol.11, No.30, 2015 年 9 月 pp.21-40
- [14] 宮下光令, 橋本修二, 尾島俊之, 中村好一, 林正幸, 加藤昌弘, 福富和夫 (1999) 「高齢者における要介護者割合と平均自立期間 —既存統計にもとづく都道府県別推計—」『厚生指標』 第 46 巻第 5 号, 1995 年 5 月 pp.25-29
- [15] 森川洋 (1998)『日本の都市化と都市システム』 大明堂
- [16] 山本雄三, 高見具広, 高橋陽子 (2018) 「統計指標に基づく市町村分類の試み」(労働政策研究・研修機構) デイスクッションペーパー 18-05, 2018 年 3 月
- [17] 美添泰人 (1999) 「探索的データ解析法の考え方」『Estrela』 No.65, 1999 年 8 月号 pp.2-8
- [18] R. Tibshirani, G. Walther, T. Hastie. (2001) “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411-423

Analysis of regional differences in the nursing care certification rate by EDA method

Yutaka Taniguchi*

Takafumi Owada†

Tadayoshi Otsuka‡

Abstract

In confirming the sustainability of public nursing care insurance, it is important to analyze the characteristics of the nursing care certification rate for each municipality. In fact, the public nursing care insurance system is a uniform system nationwide, but even if the influence of the population distribution that differs from region to region is removed, the certification rate for nursing care varies greatly from region to region.

The purpose of this paper is to clarify the regional differences in the certification rate for nursing care by municipality. Specifically, we will categorize areas with high and low certification rates for nursing care by a method based on exploratory data analysis using explanatory variables that can be used from official statistical data by municipality. This will clarify the regional image of the nursing care disparity. As a result of the analysis, it was confirmed that the certification rate for long-term care tends to be high in metropolitan areas and surrounding cities. Furthermore, it is confirmed that there is a characteristic tendency in industrial areas along the coast of western Japan, company towns in Aichi and Shiga prefectures, rural areas in mountain villages with many agricultural workers, and areas with low unemployment rate and active agriculture, dairy farming, and fishing.

* Waseda University Insurance Research Institute 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050 Email: taniguchi.yutaka12@gmail.com

† General Insurance Rating Organization of Japan Email: takafumiowada@gmail.com

‡ Waseda University Graduate School of Accountancy 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050 Email: otsukata@aoni.waseda.jp