

EDA手法による要介護認定率の地域差の分析

大和田 孝文 損害保険料率算出機構

早稲田大学保険フォーラム 2021/11/13

*本研究は谷口豊(早稲田大学総合研究機構保険研究所)、大塚忠義(早稲田大学大学院会計研究科)との共同研究である。

*本研究の内容はすべて筆者の個人的な見解であり、筆者が所属する会社・団体の見解とは無関係である。

- **動機・背景**
- **先行研究および本稿の新規性**
- **使用データ**
- **分析手法**
- **結果の考察**
- **まとめと課題**

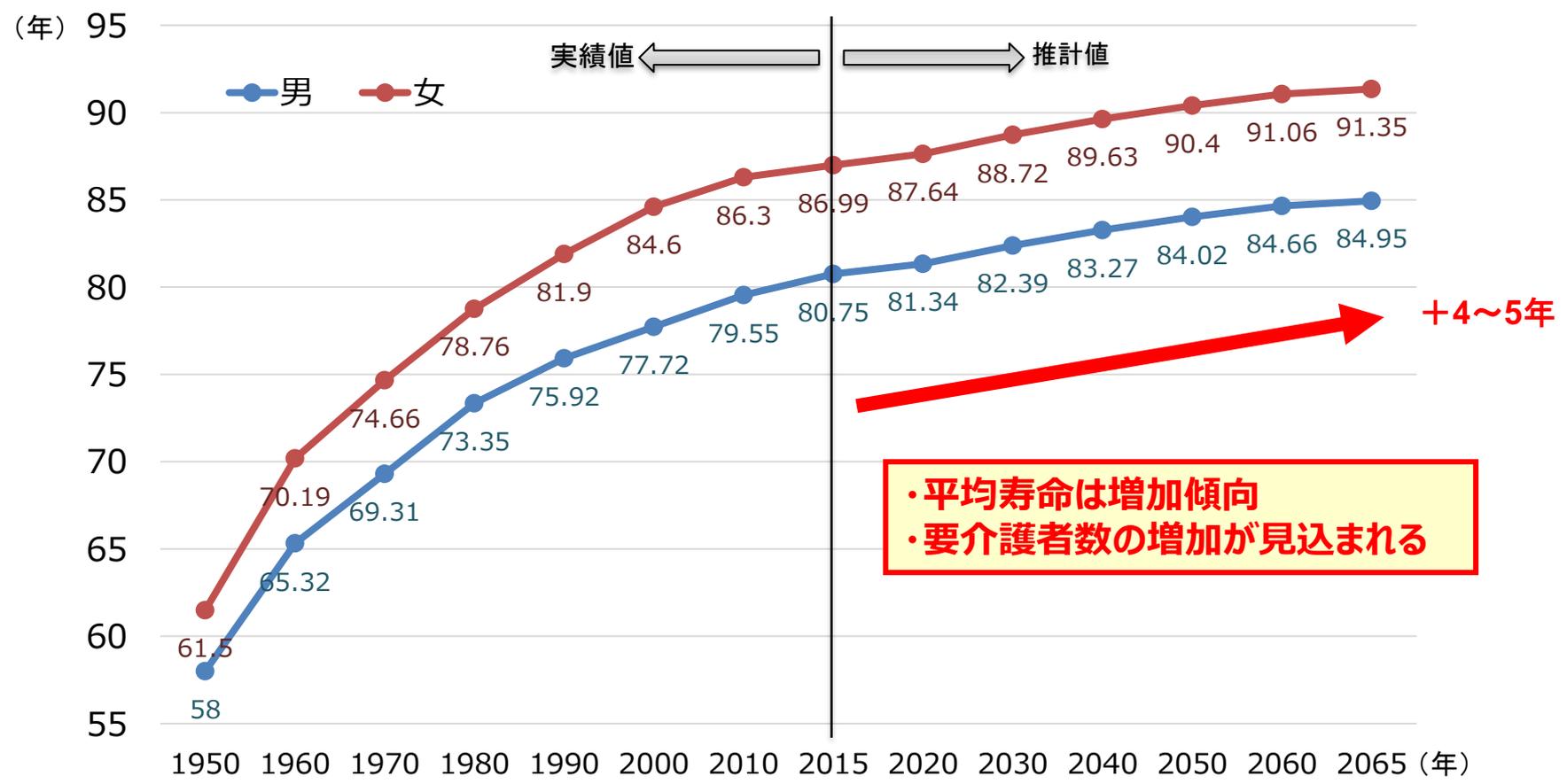
✓ **公的介護保険は市区町村ごとの運営である
運営は**持続可能**か？**

⇒ **市区町村別の要介護の地域差を分析**

⇒ **市区町村別の公的統計データに対し機械学習の手法
により要介護認定率の地域差を類型化**

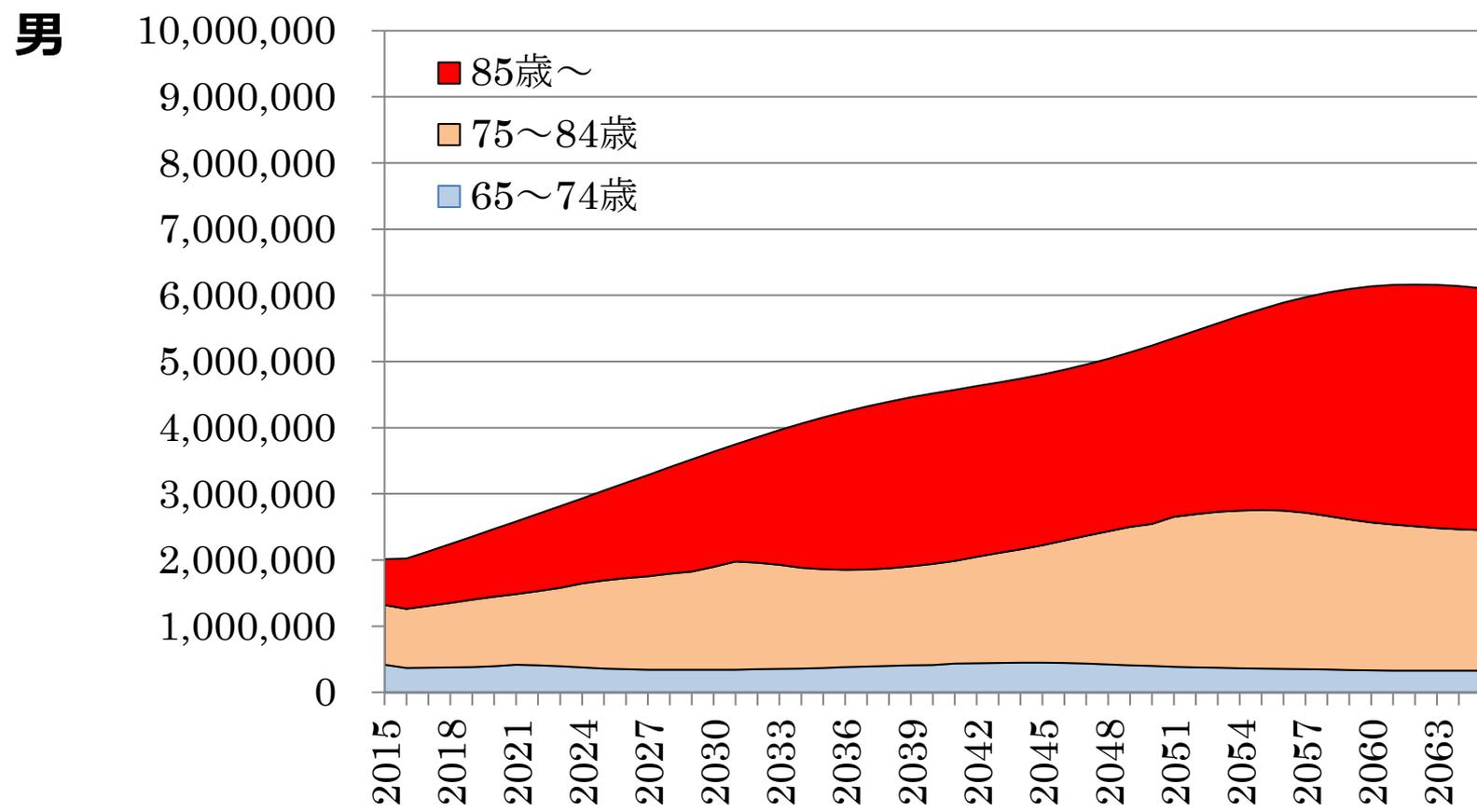
⇒ **介護格差の地域イメージを具体化することにより
格差是正に向けた特定地域への対応が可能**

平均寿命の推移と将来推計



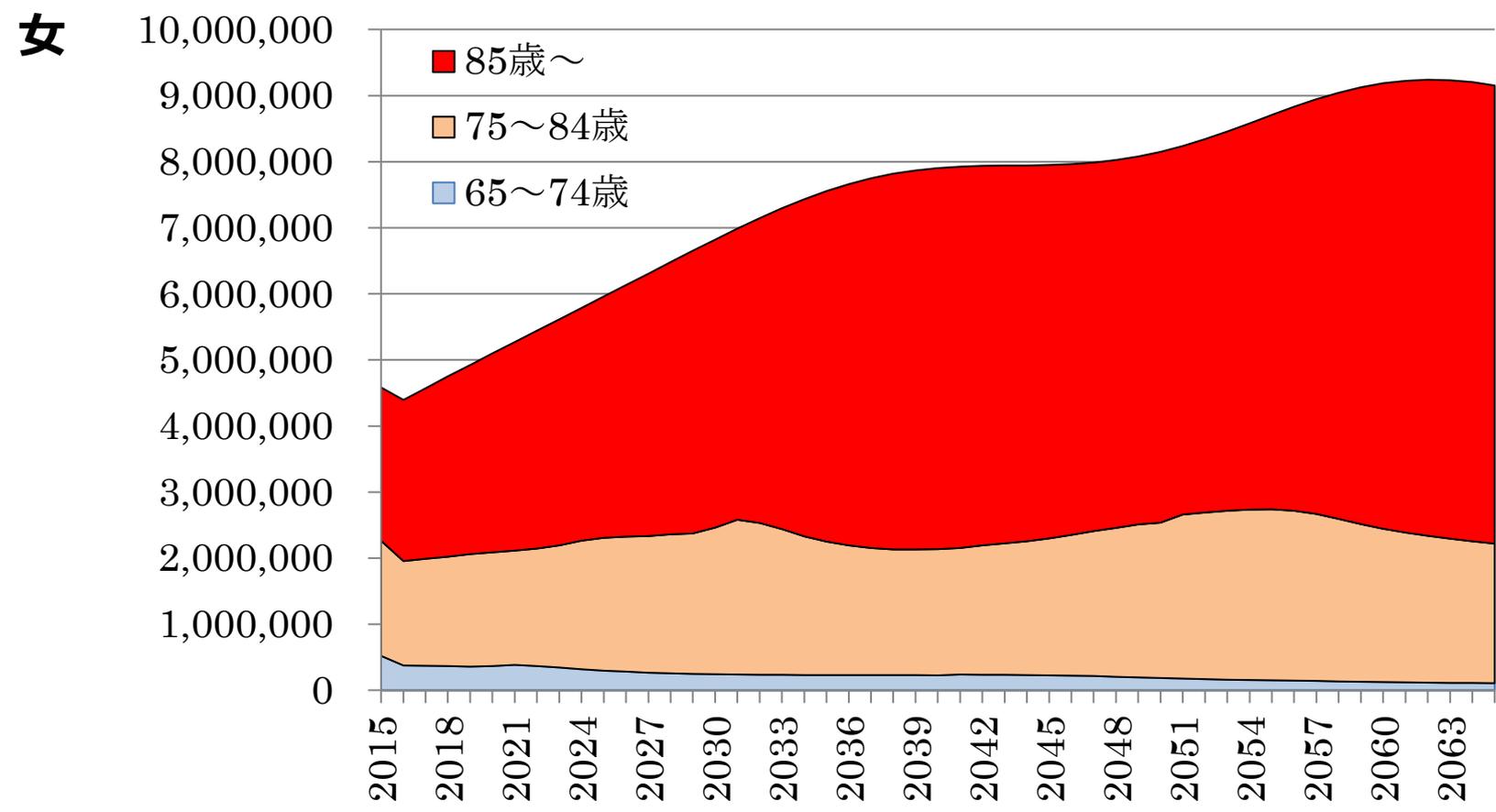
資料：1950年は厚生労働省「簡易生命表」、1960年から2015年までは厚生労働省「完全生命表」、2020年以降は国立社会保障・人口問題研究所「日本の将来推計人口（平成29年推計）」の出生中位・死亡中位仮定による推計結果をもとに筆者作成
(注) 1970年以前は沖縄県を除く値である。

要支援・要介護者数の将来推計



<出典：「健康寿命および要介護者数の将来推計」(大塚・谷口) 2019.3>

要支援・要介護者数の将来推計



<出典：「健康寿命および要介護者数の将来推計」(大塚・谷口) 2019.3>

【要介護認定率の決定要因に関する研究】

- ・単相関、固定効果モデル及びOLS推定による分析を市区町村単位で実施し、予防事業への参加が、軽度要介護認定の上昇抑制効果があることを確認した
内閣府（2018）
- ・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た
平井ほか（2009）
- ・サービス種類・サービス利用量の選択に着目し、市区町村単位の回帰分析を実施
居住サービスは、低所得者が多い地域や、高齢者夫婦世帯が多い地域で低く
サービス利用量は、医師密度が正で有意、第一次産業が負で有意を確認した
久保寺（2013）
- ・要介護認定割合に関し、所得段階1は正に有意であることを確認した
安藤（2007）
近藤他（2012）、齋藤（2018）、佐藤（2016）など

【市区町村別の要介護認定率】

・都道府県別の要介護認定率

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いない

厚生労働省

・都道府県別の要介護認定率

各都道府県で異なる年齢構成比や男女比の調整を行っている

東京都や大阪府が高いことは示されていない

大阪府が高いことを示している

宮下他（1999）

石橋、鈴木（2019）

・市区町村別の要介護認定率 地域包括ケア「見える化」システム

各市区町村で異なる年齢構成比や男女比の調整を行っている

市区町村間の格差を示した先行研究はない

厚生労働省

【地域分類に関する研究】

都道府県単位のもの多数存在。市区町村単位のもの以下

- ・地域の生産力、生産関係、産業構造等の指標を用いてクラスター分析により産業基盤停滞地域、低生産力商業地域、低生産力サービス地域、管理・行政的な管理中枢地域に分類 **小内（1996）**
- ・人口、世帯、産業別・職業別就業者比率、所得額、教育水準といった多様な指標（38指標）に対し因子分析を行い、説明力の高い3つの因子によりクラスター分析を行い8つの地域（「地方の工業都市」「住宅型周辺都市」「大都市および地方の中心都市」「農村的都市」など）に分類 **森川（1998）**
- ・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 **蓮見（2016）**
- ・統計指標に対し主成分分析を行い、得られた8主成分を用いてクラスター分析のより20の地域に分類 **山本他（2018）**

【EDA（機械学習）を活用した研究】

- ・ランダムフォレスト、回帰木により説明変数を特定し、1人当たり医療費の都道府県のグループ分けを行った **須田（2018）**
- ・企業のデフォルト予測モデルについて、機械学習モデルとロジットモデルの精度を比較した **三浦・井實・竹川（2019）**

先行研究

✓ CDA

仮説検証を目的とする分析手法
分析者が主観的に選択した説明変数を用いて分析したものが多数

要因分析が目的であるが、
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

要因分析により得られた社会的環境因子は地域特性の代理変数である可能性

本研究

✓ EDA

*1960年代から1980年代にかけて、
チューキー（J. W. Tukey）を中心として
プリンストン大学統計学部で開発

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
あらかじめ選択せず幅広い説明変数を用いて主観によらない地域特性を発見する可能性

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

先行研究

- ✓ 特定地域について追跡調査したものが多い
- ✓ 説明変数が介護に関係あるものに限定
- ✓ 市区町村単位で要介護認定率を年齢構成比・男女比を調整したものはない
- ✓ 市区町村単位の分類は地域経済の不均衡把握が目的

本研究

- ✓ **全国の市区町村**を調査した
- ✓ 幅広く説明変数を候補にした機械学習による選択
*主成分分析などは複数の目的変数に対応できない
- ✓ 市区町村単位で年齢構成比・男女比を調整して分析
- ✓ 介護格差を目的に地域分類

使用データ

公的統計データの集計値（2015年・2010年の2年分）

介護保険事業状況報告（年報）	厚生労働省
国勢調査	総務省統計局
市町村別決算状況調	総務省
全国都道府県市区町村別面積調	国土交通省国土地理院
市町村税課税状況等の調	総務省
市町村別決算状況調	総務省
住民基本台帳人口移動報告年報	総務省統計局
地域保健・健康増進事業報告	厚生労働省
自殺の統計	内閣府
交通事故統計年報	交通事故総合分析センター

ほか、全19種類

目的変数

『介護保険事業状況報告』をもとに作成した

- ①軽度要介護認定率（65歳以上要支援1、要支援2、要介護1）を標準化
- ②重度要介護率（65歳以上要介護2、要介護3、要介護4、要介護5）を標準化

説明変数

市区町村別に把握可能な統計値を幅広く候補とした
ただし、欠損値を含むデータは、原則レコードごと削除した

→ **説明変数の候補数は79**

目的変数

標準化：

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
この標準化により地域ごとに異なる年齢構成比、男女比を比較可能になる

(地域_Aの標準化介護認定比) =

$$\sum_{\substack{t \in \{65, 70, 75, 80, 85, 90\} \\ x \in \{\text{男性}, \text{女性}\}}} (\text{地域}_A \text{の} t \sim t + 4 \text{歳} X \text{の要介護認定者数})$$

$$\sum_{\substack{t \in \{65, 70, 75, 80, 85, 90\} \\ x \in \{\text{男性}, \text{女性}\}}} (\text{全国平均の} t \sim t + 4 \text{歳} X \text{の要介護認定者数}) \cdot (\text{地域}_A \text{の} t \sim t + 4 \text{歳} X \text{の人口})$$

(注) 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。

ステップ1： 地理的可視化・記述統計による分析

標準化介護認定比（軽度）・標準化介護認定比（重度）について
比率の高い地域、低い地域をヒートマップで示す
介護格差の地域特性を概観し、地域特性の視覚情報を獲得する
視覚情報・記述統計はクラスター分析の結果の妥当性の確認に用いる



ステップ2： ランダムフォレストによる説明変数の選択

ランダムフォレスト（%IncMSE、IncNodePurity）を10回実行し平均値を算定
指標重要度の上位層を採用した



ステップ3： クラスター分析(k平均法)による地域分類

選択した説明変数を用いて、Gap統計量にて最適なクラスター数(分類の数)を決定
各クラスターに分類された地域特性を考察

分析手法

モデル	特徴	
ランダムフォレスト	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	
	%IncMSE	説明変数を除いた場合にモデルの予測誤差（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
	IncNodePurity	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
k平均法	非階層クラスター分析のひとつ 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して、データを分類するアルゴリズム	

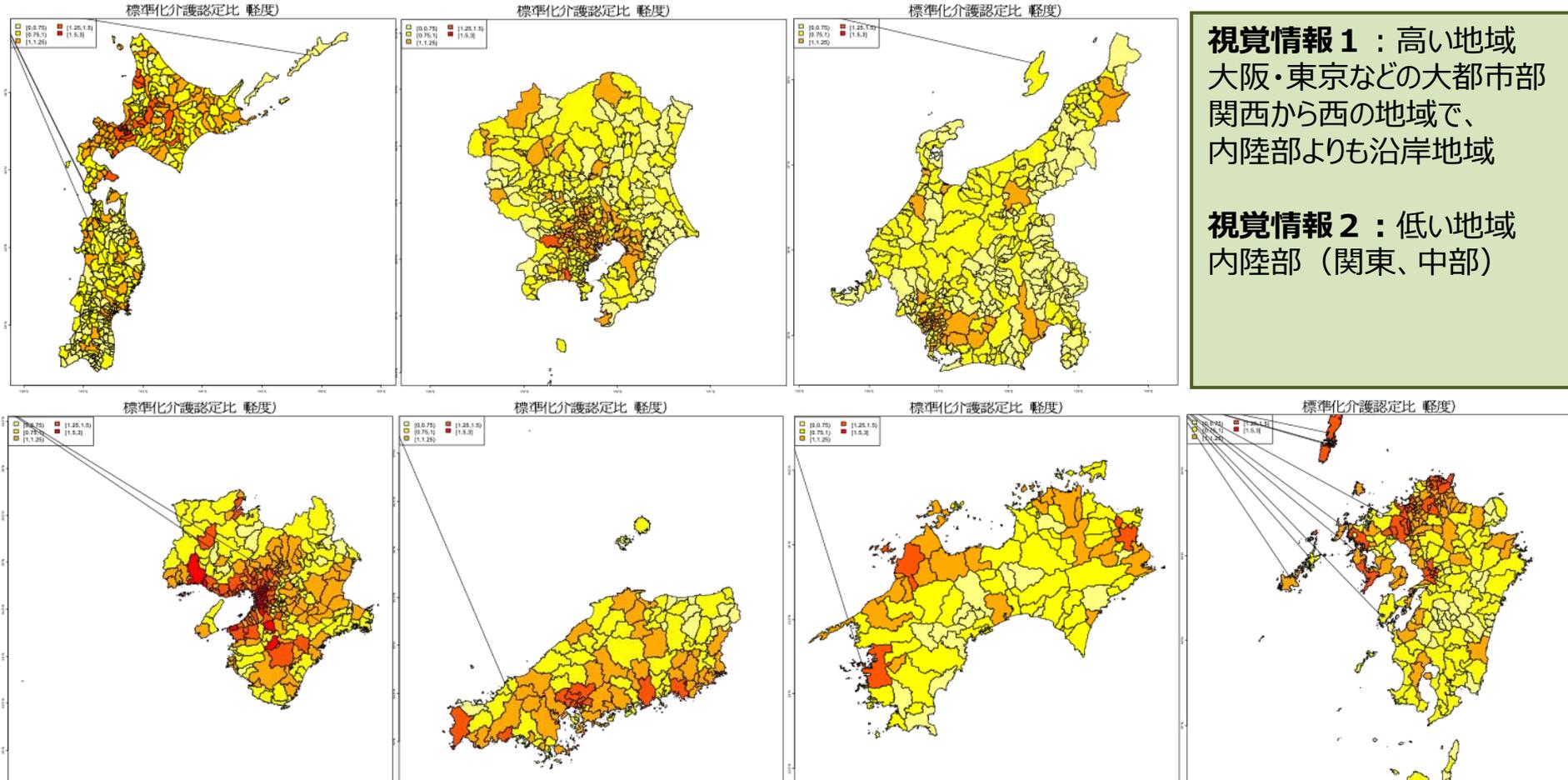
(補足1) 統計処理はR(4.0.2)で行った
ランダムフォレストは、randomForestパッケージのrandomForest関数を用い、
k平均法はstatsパッケージのkmeans関数を用いた
なお、後述するGAP統計量の測定は、clusterパッケージのclusGAP関数を用いた

(補足2) 主成分分析は、目的変数ごとに説明変数の使い分けはない
ランダムフォレストは、目的変数に応じた説明変数の選択が可能

ステップ1： 地理的可視化・記述統計による分析

【標準化介護認定比（軽度）】

2015年データ



視覚情報1：高い地域
大阪・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

視覚情報2：低い地域
内陸部（関東、中部）

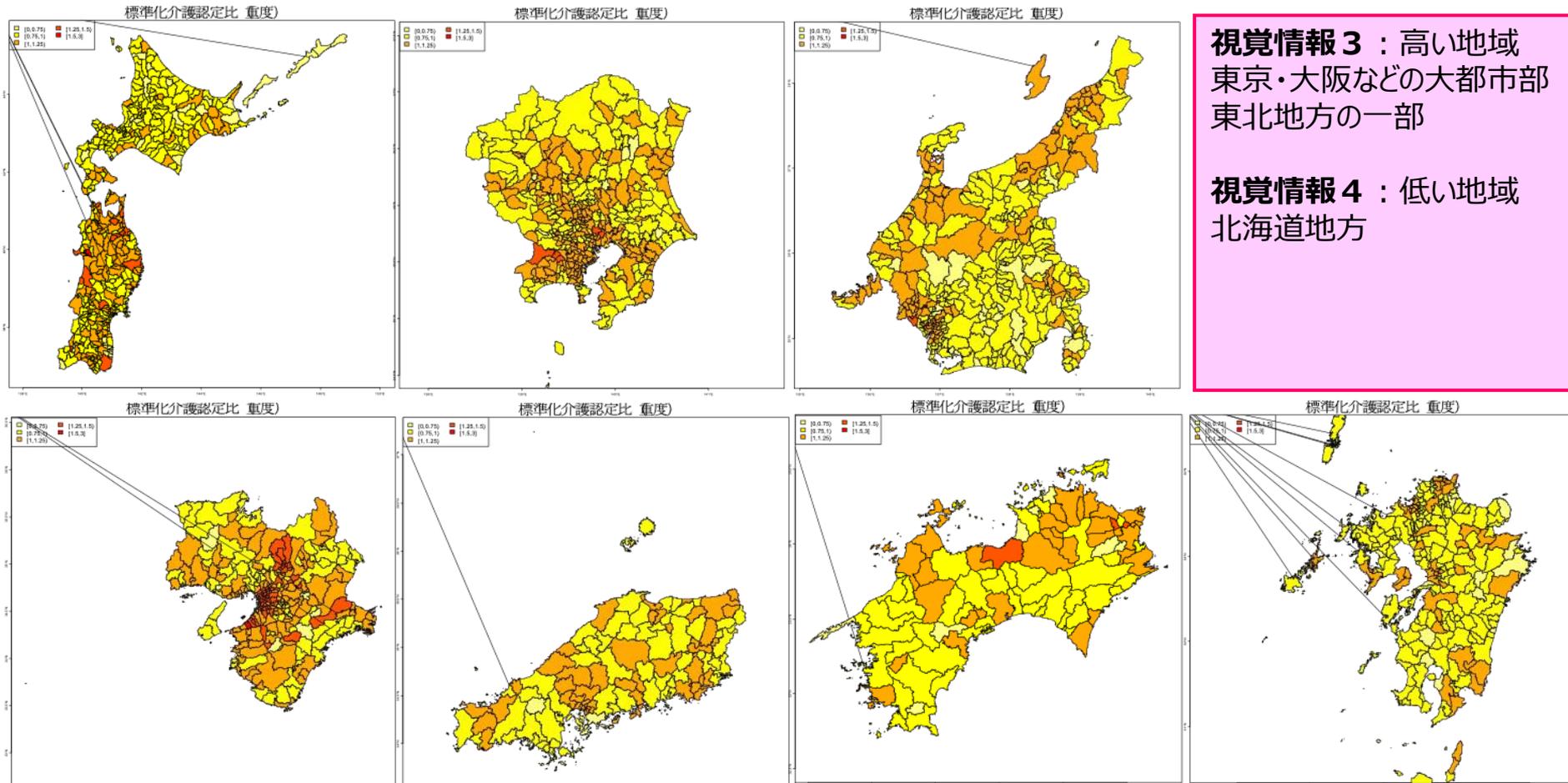
(注)ヒートマップは、国土地理院発行の数値地図（国土基本情報）と ESRI ジャパンの全国市区町村界データを使用し加工した。

地図データ：(c)Esri Japan

ステップ1： 地理的可視化・記述統計による分析

【標準化介護認定比（重度）】

2015年データ



視覚情報3：高い地域
東京・大阪などの大都市部
東北地方の一部

視覚情報4：低い地域
北海道地方

(注)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境界データを使用し加工した。

地図データ：(c)Esri Japan

ステップ1： 地理的可視化・記述統計による分析

2015年データ

	標準化介護認定比 (軽度)	Welch検定		標準化介護認定比 (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 ***

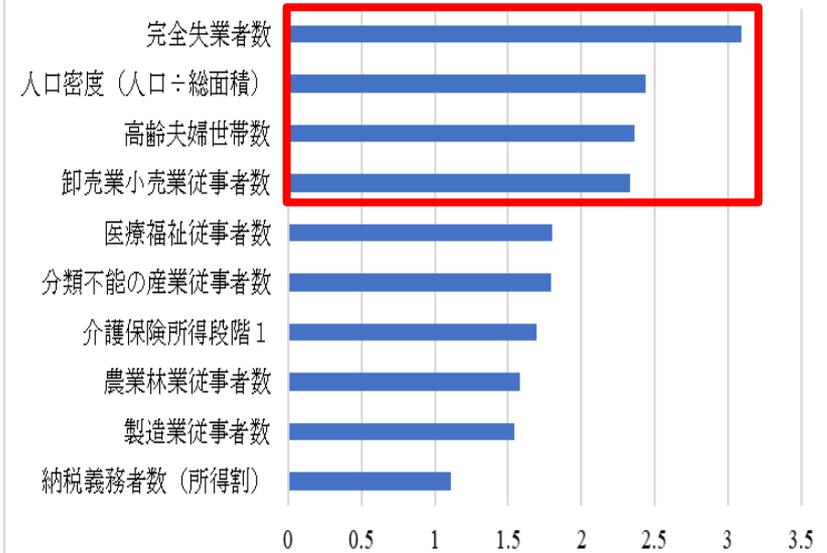
有意水準 *** (0.1%)、** (1%)、* (5%)

ステップ2： ランダムフォレストによる説明変数の選択

標準化介護認定比（軽度）



標準化介護認定比（重度）



説明変数の重要度を棒グラフの長さが示している → 短くなっている段差が確認できる

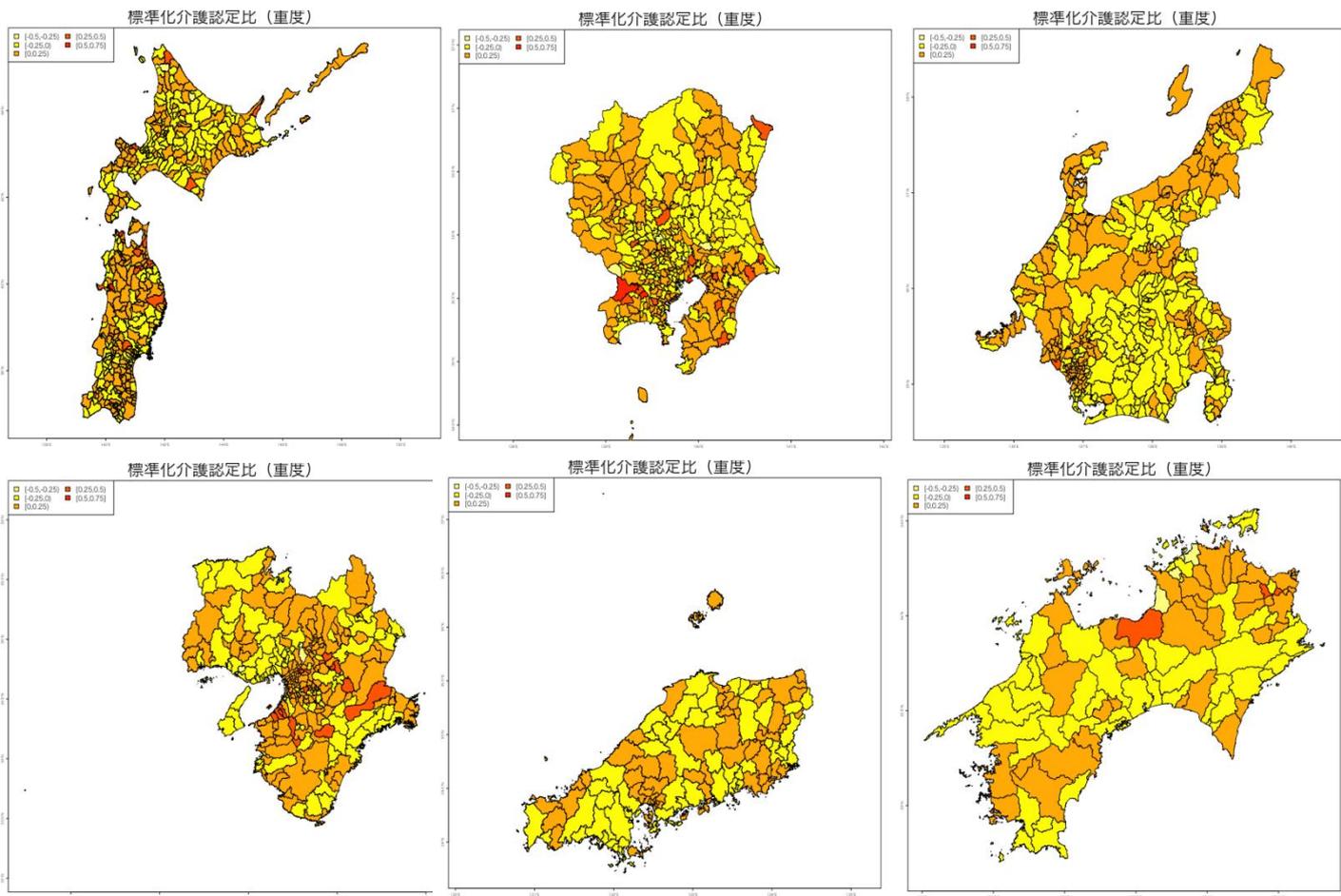
標準化介護認定比（軽度）は上位7まで

標準化介護認定比（重度）は上位4までを k平均法で使用する説明変数として選択する

ステップ2（参考）： ランダムフォレストの予測結果を確認（残差）

【標準化介護認定比（重度）】

2015年データ



視覚情報5：軽度よりも全体的に残差が小さいが、特に中部地方、九州地方の残差が小さい

視覚情報6：軽度よりも強くステップ1のマップと同じ傾向があらわれた

(注)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境界データを使用し加工した。

地図データ：(c)Esri Japan

ステップ2（参考）： ランダムフォレストの予測結果を確認（2乗残差）

【標準化介護認定比（重度）】

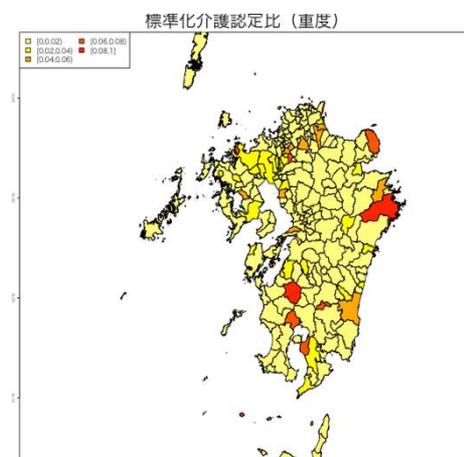
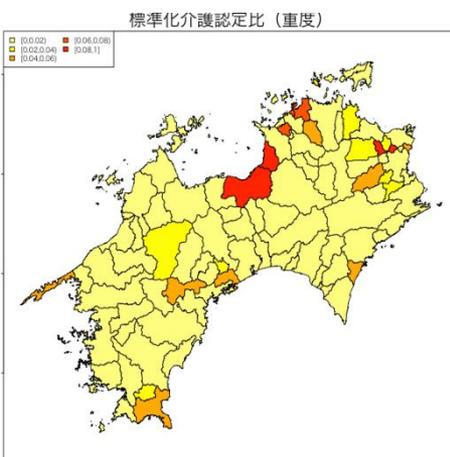
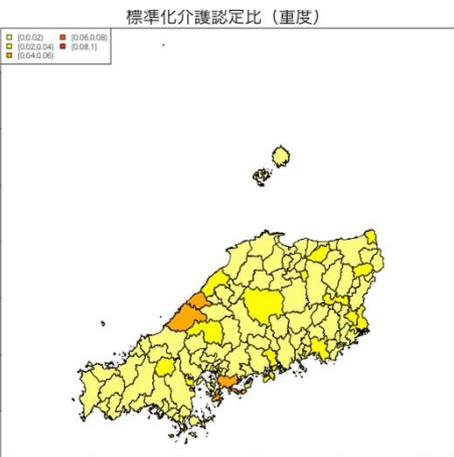
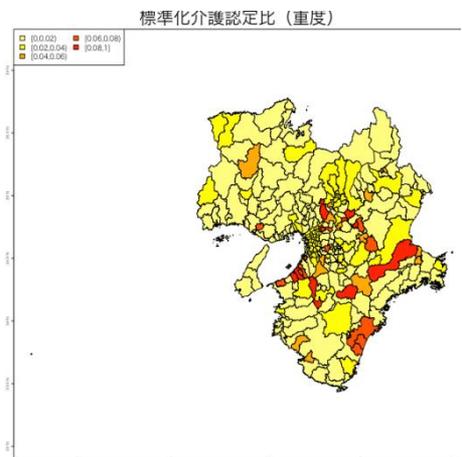
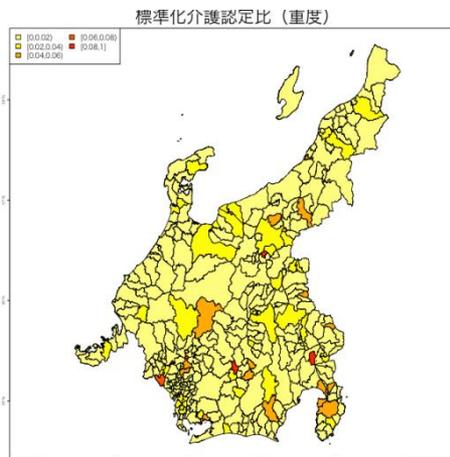
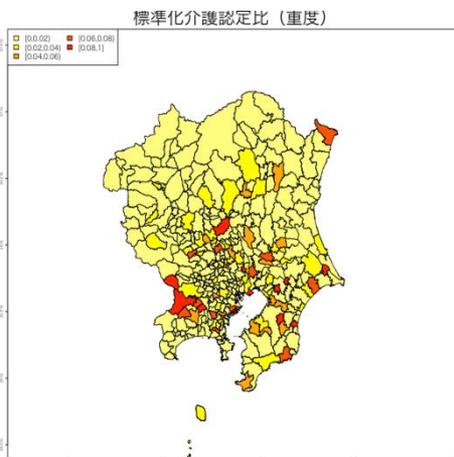
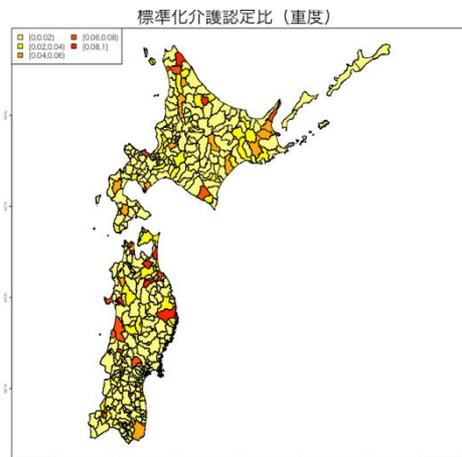
2015年データ

視覚情報7：軽度に比べて2乗残差が全体的に小さい

【平均2乗残差】

（軽度）：0.026141

（重度）：0.016779



(注)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。

地図データ：(c)Esri Japan

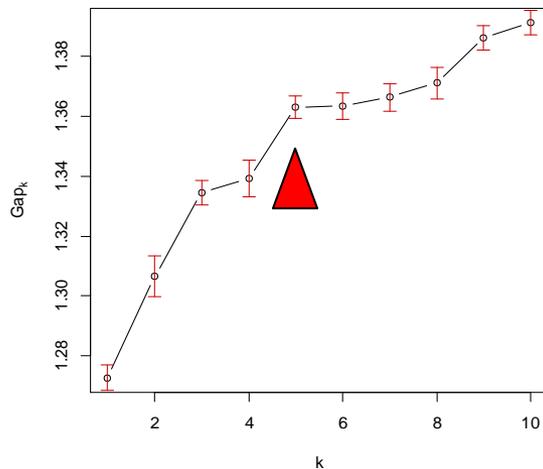
ステップ^o3 : クラスタ分析(k平均法)による地域分類

Gap統計量

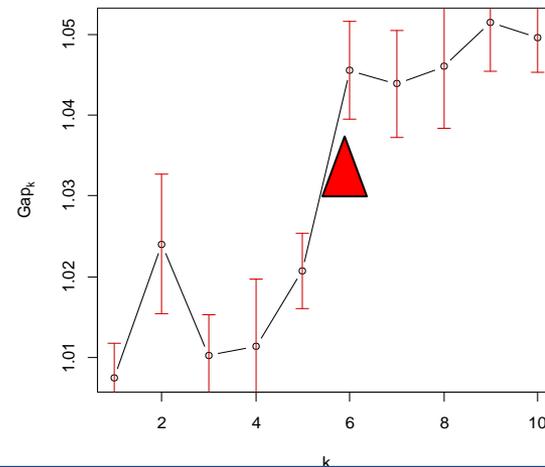
クラスター数 k に対する k 平均法で作成したクラスターの評価関数を分母に、クラスター数 k に対する一様乱数により作成したクラスターの評価関数を分子とした指標

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統計量が大きいほど精度が良いことを意味するもの

標準化介護認定比（軽度）



標準化介護認定比（重度）



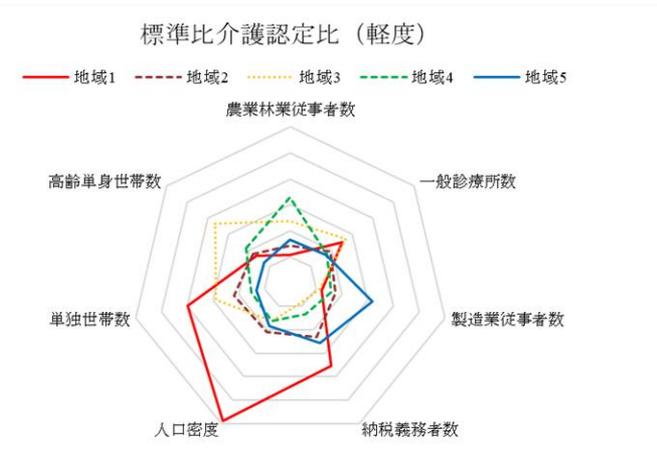
クラスター数は
標準化介護認定比（軽度）は5、標準化介護認定比（重度）は6 を採用した

ステップ3： クラスター分析(k平均法)による地域分類

【標準化介護認定比（軽度）】

	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	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

	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	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



地域1は高い、地域2はやや高い、地域3は標準、地域4はやや低い、地域5は低い

高い地域1 ……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

低い地域5 ……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

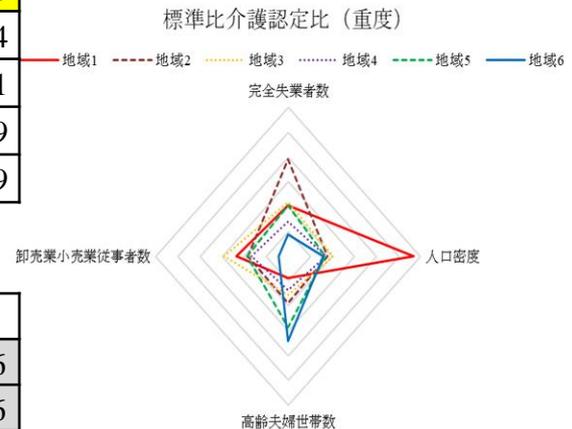
やや低い地域4 ……農業従事者数が多い

ステップ3： クラスター分析(k平均法)による地域分類

【標準化介護認定比（重度）】

	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	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

	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	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



地域1は高い、地域2は高い、地域3は標準、地域4は標準、地域5は標準、地域6は低い

高い地域1 …人口密度が高い、高齢夫婦世帯割合は低い

次に高い地域2 …完全失業者数が高い

低い地域6 …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

結果の考察

【標準化介護認定比（軽度）】

再掲

	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	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	0.4824	0.4215	0.3523	0.3733	0.4348
人口密度	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

	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	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	1	3	5	4	2
人口密度	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

高い地域1 …東京都特別区、大阪府大阪市、福岡県福岡市、神奈川県横浜市、千葉県市川市、大阪府東大阪市などが分類。**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に該当するのは98%

やや高い地域2 …千葉県市原市、神奈川県藤沢市、三重県伊勢市、兵庫県姫路市、岡山県倉敷市、香川県坂出市、愛媛県松前町、福岡県行橋市などが分類。**海岸沿いの工場地帯が多く、ベッドタウンとなっている地域も多い。西日本の沿岸地域**が多く分類。沿岸地域に該当するのは50%、西日本に該当するのは47%

低い地域5 …愛知県豊田市、愛知県刈谷市、愛知県安城市、滋賀県甲賀市、滋賀県竜王町などが分類。地域2と同じく工場地帯が多いが、**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。その他、**製造業以外に農業も盛んな地域**が多い傾向。納税義務者数の割合は地域2より多く裕福な地域。

やや低い地域4 …群馬県嬬恋村、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。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。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%

結果の考察

【標準化介護認定比（重度）】

再掲

	地域1	地域2	地域3	地域4	地域5	地域6		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	1	2	3	4	5	6
平均標準比介護認定比	1.05	1.03	0.99	0.95	0.95	0.88	完全失業者数	3	1	2	5	4	6
完全失業者数	0.0249	0.0380	0.0255	0.0203	0.0246	0.0164	人口密度	1	3	2	4	5	6
人口密度	10,508	574	1,209	376	267	31	高齢夫婦世帯数	6	3	4	5	2	1
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	2	4	1	5	3	6
卸売業小売業従事者数	0.0703	0.0634	0.0777	0.0614	0.0638	0.0449							

高い地域1 ……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。これらの地域は高齢単身者が多い。**軽度と同じ傾向。**

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%

高い地域2 ……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。**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**完全失業率も高い**傾向。

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%

低い地域6 ……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。

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。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。

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%、北海道または長野健に該当する地域は47%

結果の考察

		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 介護認定比 (軽度)	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に該当するのは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
	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 介護認定比 (重度)	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% 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 視覚情報と 整合的 新しい知見も 得られた
	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

まとめ

- ✓ 軽度・重度ともに都市圏やその周辺で高い傾向が確認された。
- ✓ 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が確認できた。
- ✓ 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が確認できた。
- ✓ 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。

課題

- ✓ 標準化介護認定比（軽度）での一般診療所数や、標準化介護認定比（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地域分類に不可欠な因子とまでは断言できず、むしろ過疎地域の代理変数である可能性の方が高いと思われる。
- ✓ ランダムフォレストのヒートマップから、重度に比べて、軽度の分析には更なる改善の余地があると思われる。
- ✓ 分類方法が妥当であるかの議論は残っている。より妥当な介護格差の地域分類はこれからも引き続き検討課題である。
- ✓ 得られた介護格差に係る地域分類の結果を持って何をすべきかの政策的な提言までには至っていない。

機械学習やデータサイエンスの有用性について

機械学習やデータサイエンスの有用性について

最近、新しい保険商品として健康増進型保険、およびテレマティクス保険（走行距離や運転特性等の情報を基に保険料が変わる保険）が発売されている。



保険会社は保険契約後も保険契約者からデータを取得できるようになり、今までよりも詳細な統計データを取り扱うことが可能になった。

統計データを機械学習やデータサイエンスの知見を応用しながら分析・評価することで、今より合理的な保険料率の算出や社会的意義の大きな情報発信等を行うことが可能になる。

機械学習やデータサイエンスの有用性について

学生でも機械学習やデータ分析を行うことはできる？

統計データ（公的統計）

RやPythonなどの無料で
使用できる統計ソフト



回帰分析などの伝統的手法から機械学習までの
様々な分析を行うことが可能！



- Excelでも一通りの分析は可能
- プログラミングに習熟していなくても勉強すればある程度は動かすことができる。

- [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－」『厚生の指標』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) ディスカッションペーパー 18-05 2018年3月
- [17] 三浦翔、井實康幸、竹川正浩 (2019) 「入出金情報を用いた信用リスク評価－機械学習による実証分析－」日本銀行ワーキングペーパーシリーズ 2019年6月
- [18] 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
- [19] 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

ご清聴ありがとうございました
