

2025 年度 統計データ分析コンペティション

審査員奨励賞 [大学生・一般の部]

人口動態の地域差に着目した基礎的分析

山形 陽生

(武蔵精密工業株式会社)

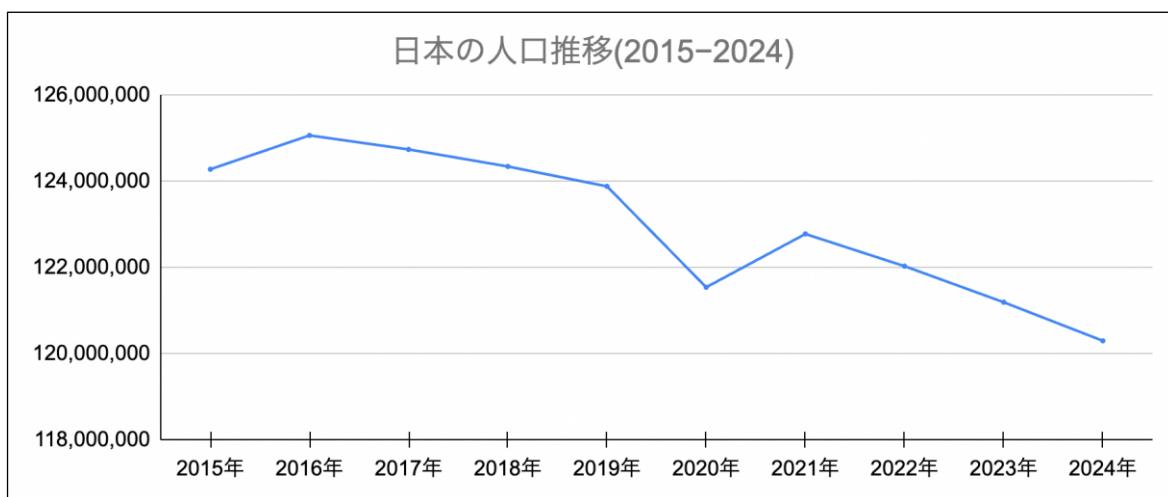
人口動態の地域差に着目した基礎的分析

1. 研究のテーマと目的

日本社会が直面する課題は多岐にわたるが、その中でも人口減少は特に深刻な問題である。出生率の低下や高齢化が主要な要因とされる一方、都市部への人口集中や過疎地域からの人口流出といった地域間の人口移動も、各地域の人口構造やバランスに影響を及ぼしている。こうした人口動態の変化は地域ごとに異なる傾向を示しており、限られた資源の中で有効な政策を立案するには、各地域の特性を正確に把握する必要がある。

本研究では、都道府県ごとの自然増減率と社会増減率に着目し、その関係性を統計的に分析する。自然増減率と社会増減率はいずれも人口変動の主要な構成要素であるが、地域ごとにその影響の現れ方には違いが見られる可能性がある。本研究では、都道府県ごとの自然増減率と社会増減率に着目し、その関係性を統計的に分析する。自然増減率と社会増減率はいずれも人口変動の主要な構成要素であるが、両者の寄与度や相関の強さは地域によって異なる可能性がある。そのため、地域ごとの特性を捉えるためにクラスタリングし、それぞれの傾向を把握した上で、地域特性を踏まえた政策検討の基礎となる知見を提示することを目的とする。

現在の人口対策は全国一律的な傾向が強く、地域ごとの特性まで十分に考慮されていない場合も多い。本研究は、統計的な分析を通じて、地域ごとの傾向を把握することで、今後の施策検討に資する視点を提供したい。



[図1：日本の人口推移]

2. 研究の方法と手順

本分析では、都道府県別の「自然増減率」と「社会増減率」の関係性を明らかにすることで、地域ごとの人口動態の特性を把握し、地域特性を踏まえた政策検討の基礎的な知見を得ることを目的とする。「自然増減率」は出生と死亡の差を表す人口動態の自然的要素を示す指標、「社会増減率」は転入・転出による社会的な人口動態を示す指標として選定した。本分析を行うことで、両者の相関の程度や地域差を統計的に検証し、全国一律ではない政策の必要性を示すことが期待される。主な分析方法は次の通りである。

Step1：分析項目の定義

本研究において分析対象とする指標は、自然増減率と社会増減率である。自然増減率は出生数と死亡数の差による人口変動、社会増減率は転入者数と転出者数の差による人口変動を、それぞれ総人口に対する割合で算出する。これらを基に、地域ごとの人口動態の傾向を把握する。

Step2：データの収集

自然増減率および社会増減率のデータは、e-Stat（政府統計ポータルサイト）に公表されている2023年10月～2024年9月の「都道府県別自然増減率－総人口」「都道府県別社会増減率－総人口」を用いる。さらに、重回帰分析で用いる補助的な変数については、教育用標準データセット（SSDSE）や、e-Statなどから抽出する。これらのデータにより、都道府県ごとの分析用データセットを構築する。

Step3：外れ値の検討

自然増減率および社会増減率の分布をボックスチャートにより可視化し、特異な傾向を示す都道府県が存在するかを確認する。これらの都道府県については、他と区別して独立したクラスタとして扱う。

Step4：クラスタリングの実施

特異な都道府県を除いた残りの都道府県について、自然増減率と社会増減率の関係を基に教師なし学習によるクラスタリングを実施する。クラスタリング後は散布図を用いて結果を視覚的に確認し、クラスタ分けがデータの傾向を適切に反映しているか評価する。

Step5：PLS回帰による重要変数の抽出

特異な都道府県を除いた全体を対象に、PLS（Partial Least Squares）回帰分析を実施した。目的変数は自然増減率と社会増減率で、それぞれ個別にモデル化した。PLSは多重共線性や高次元に対する頑健性があり、影響要因の寄与構造を把握する本分析の目的にも適するため採用した。説明変数にはSSDSEやe-Statなどから抽出した補助的変数を用い、地域ごとの人口動態に対する寄与度をクラスタごとに評価した。極端な特性を示しサンプル数が1に留まる特異な都道府県は、安定性の観点から除外し補足的に扱った。

Step6：PLS分析結果と可視化の評価

PLSの結果については、VIF（Variance Inflation Factor）により多重共線性を確認し、各説明変数のVIP（Variable Importance in Projection）スコアおよびその標準偏差に基づいて寄与の大きさと安定性を評価する。次に、VIPの高い変数を抽出し、それらについてクラスタごとにバイオリンプロットを作成し、分布を可視化することで、クラスタごとの特徴把握と解釈を深める。なお、PLSは各クラスタに対して目的変数を説明するための線形結合を個別に学習しているため、目的変数の分散構造や説明変数間の相関関係が異なり、VIPや係数の重要度の順位がクラスタ間で異なることは自然な結果である。

Step7：考察と解決策の提示

クラスタリングおよびPLSの結果をもとに、地域ごとの人口動態の傾向や特徴を整理する。特に、全国一律の政策と地域ごとの実情との適合性について検討し、地域特性に応じた対応の必要性を分析の観点から明らかにする。得られた結果を踏まえ、各地域の特徴に即した人口政策の方向性を整理し、今後の検討に向けた基礎的な視点を示す。

3. データセットの加工

3.1 目的変数

日本の自然増減率および社会増減率のデータを取得し、人口動態の指標として用いた。各データの出典および対象年度については、表1に示す。

表1 目的変数のデータセット

目的変数	出典	年度
日本の人口推移	統計ダッシュボード	2015～2024
自然増減率	e-Stat 人口推計	2022.10～2023.09
社会増減率	e-Stat 人口推計	2022.10～2023.09

3.2 説明変数

本研究で使用した説明変数とその出典を表2に示す。説明変数は、目的変数である自然増減率および

社会増減率との関連を考慮し、政策領域ごとに分類して選定した。なお、本研究では都道府県単位でクラスタリングを行った上で PLS 回帰分析を実施しており、スマートデータの構造にあることから、モデルの安定性を確保する目的で、各クラスタにおける説明変数の数は最大 10 個に制限している。

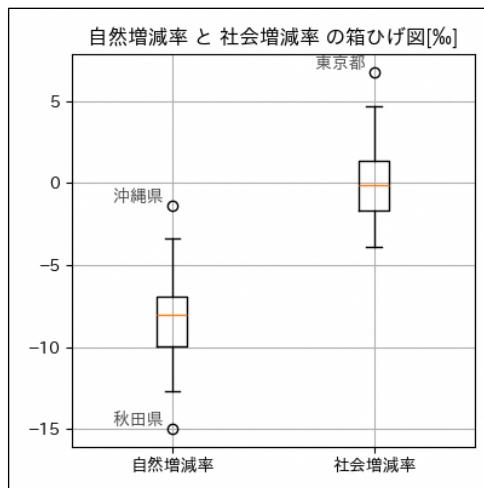
表 2 説明変数のデータセット

目的変数	政策領域	説明変数	説明変数の加工	出典	年度
自然増減率	経済	1人当たり県民所得	-	SSDSE-E-2025	2020
		正常分娩の平均出産費用	-	厚生労働省 保健局	2023
		児童手当支給水準	都道府県別支給額÷都道府県別支給対象児童数	e-Stat 児童手当事業年報	2024
	サービス	保育所定員率	保育所等在所児数 ÷ 保育所等定員数	SSDSE-B-2025	2022
		介護施設整備率	介護施設定員数÷70歳以上人口	e-Stat 介護サービス施設・事業所調査、人口推計	2023
		学校給食	-	SSDSE-C-2025	2024
	健康	人口千人当たりの医師数	医師数÷総人口×1000	SSDSE-D 医師数 e-Stat 人口推計	2022
		高齢者人口当たり病床数	病床数÷70歳以上人口	e-Stat 病院報告令和5年病院報告 二次医療圏編 二次医療圏、人口推計	2023
	社会	人口千人当たり婚姻件数	婚姻件数÷総人口×1000	SSDSE-B-2025	2022
		自殺率	-	厚生労働省自殺対策推進室	2024
社会増減率	経済	1人当たり県民所得	-	SSDSE-E-2025	2020
		可処分所得	-	e-Stat 都道府県データ	2023
	雇用	月間有効求人件数（一般）	-	SSDSE-B-2025	2022
		若年労働力率	就業者数（15～24歳）÷労働力人口	e-Stat 都道府県データ	2020
	交通	可住地駅密度（駅/10km ² ）	-	全国の"駅密度"調査	2024
		通勤・通学	-	SSDSE-D-2025 MG51	2021
	住宅	持ち家比率	持ち家数÷総住宅数	SSDSE-E-2025	2023
	環境	公園面積比率	自然公園面積÷可住地面積	SSDSE-E-2025 SSDSE-E-2024	2022
	文化・社会	人口10万人当たり図書館・博物館施設数	(図書館数+博物館数)÷総人口 *100000	SSDSE-E-2025 SSDSE-E-2023	2021
		人口千人当たり公営の借家数	公営の借家数÷総人口*1000	e-Stat 都道府県データ	2023

4. データ分析の結果

4.1 外れ値の検討とクラスタリングの結果

自然増減率および社会増減率の分布を把握するため、ボックスチャートにより可視化を行った。その結果、沖縄県、秋田県、東京都が他の都道府県と比較して顕著に異なる傾向を示すことが確認された。これらの都道府県については、独立したクラスタとして別扱いとした（図 2）。

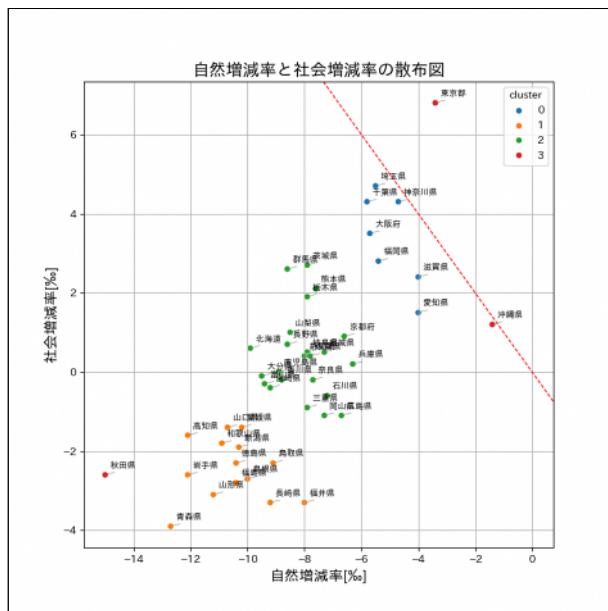


[図 2：自然増減率と社会増減率のボックスチャート]

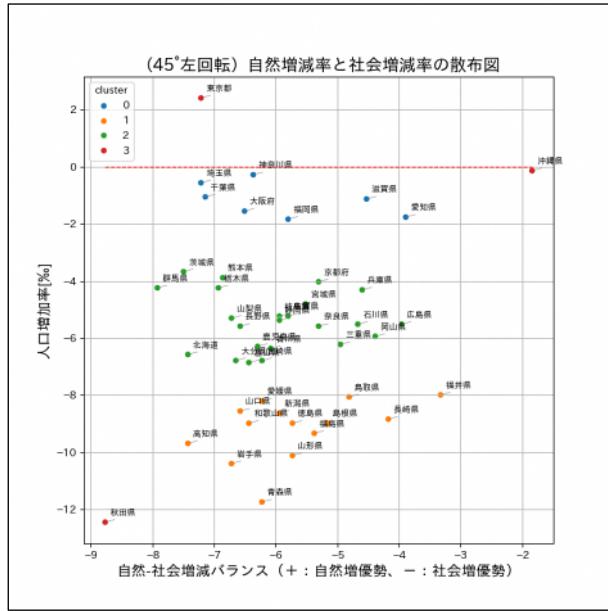
次に、特異な都道府県を除いた残りの都道府県について、自然増減率と社会増減率を基に教師なし学習によるクラスタリングを実施した。手法には Gaussian Mixture Model (GMM) を用いた。これは、データ

の分布が楕円形を呈していること、および各クラスタの密度が一様でないと考えられたことから、この手法が適していると判断したためである。

クラスタリング結果は散布図として可視化し、自然増減率と社会増減率の合計がゼロとなる基準線を赤線で示した。この基準線を上回る点は人口が増加している自治体、下回る点は人口が減少している自治体をそれぞれ示している。さらに、散布図を45度左回転させることで、人口増減の傾向がより直感的に把握できるようにした(図3、4)。クラスタは人口増減率を基準として分類された。クラスタ0は、東京都を除く大都市圏およびその周辺地域が多く含まれており、全体として人口減少が比較的緩やかな傾向を示す。クラスタ1は東北、山陰、四国地方の県が多く、人口減少傾向が顕著な地域が中心となっている。クラスタ2はその中間的な位置づけで、増減率の水準が極端でない県が多く含まれる。



[図3：自然増減率と社会増減率の散布図]



[図4：図3の45° 左回転の散布図]

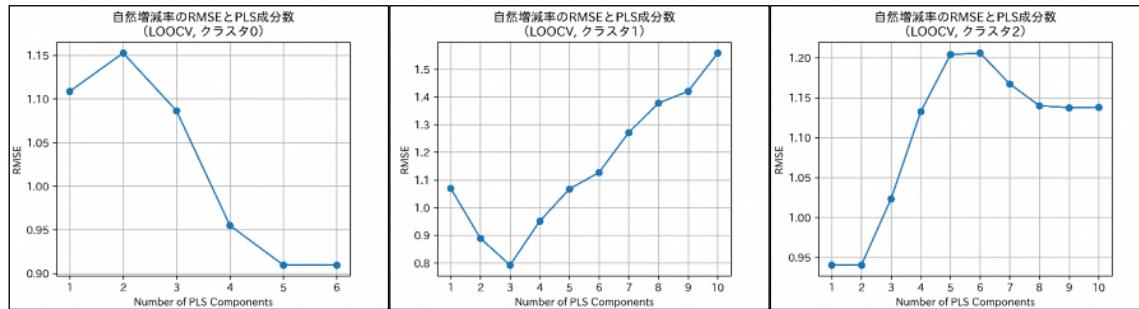
4.2 PLS 回帰分析の結果

クラスタリングによって分類された各クラスタに対して、自然増減率および社会増減率を目的変数とするPLS回帰分析を実施した。目的変数ごとに個別のモデルを構築し、説明変数は標準化したうえで用いた。過学習を避ける目的でLOOCV(Leave-One-Out Cross-Validation)を適用し、PLS成分数はRMSE(Root Mean Squared Error)が最小となる点を選定した。

4.2.1 自然増減率の分析

図5に、各クラスタにおける自然増減率を目的変数としたPLS回帰において、LOOCVを実施し、PLS成分数とRMSEの関係を示す。RMSEの最小値は、クラスタ0で成分数5、クラスタ1で成分数3、クラスタ2で成分数2において確認された。

表3には、自然増減率に関するPLS回帰の分析結果をまとめている。説明変数の影響の大きさはVIPに基づいて評価した。また、VIPの信頼性を確認するため、LOOCVによって得られたVIPのばらつきから標準偏差を算出した。加えて、多重共線性の影響を把握するためにVIFを算出した結果、10を超えるような著しく高い値は見られず、多重共線性の影響は限定的であると判断した。



[図5：クラスタごとの自然増減率のRMSEとPLS成分数]

表3 自然増減率のPLS回帰の分析結果

説明変数	VIF	PLS係数			VIP			VIPの標準偏差		
		クラスタ0	クラスタ1	クラスタ2	クラスタ0	クラスタ1	クラスタ2	クラスタ0	クラスタ1	クラスタ2
1人当たり県民所得	3.14	0.51	0.15	0.04	1.65	0.54	0.79	0.23	0.18	0.09
正常分娩の平均出産費用	3.93	0.19	0.18	0.17	0.78	0.52	1.11	0.18	0.15	0.1
児童手当支給水準	4.90	-0.04	0.18	-0.21	0.78	0.46	1.30	0.14	0.15	0.11
保育所定員率	1.31	0.14	0.32	0.04	0.79	0.78	0.18	0.15	0.18	0.09
介護施設整備率	2.19	-0.01	0.36	0.18	0.80	0.79	0.79	0.27	0.1	0.19
学校給食	1.31	0.13	0.12	-0.17	0.62	0.85	1.05	0.11	0.18	0.14
人口千人当たりの医師数	2.09	0.25	0.19	0.22	0.92	0.96	0.98	0.16	0.17	0.09
高齢者人口当たり病床数	3.28	-0.13	-0.20	-0.06	0.97	0.55	0.81	0.19	0.11	0.09
人口千人当たり婚姻件数	2.82	0.32	0.74	0.32	0.93	2.10	1.54	0.25	0.11	0.12
自殺率	1.25	0.50	-0.32	-0.16	1.32	1.28	0.84	0.21	0.17	0.13

※VIFおよびPLS係数については、各変数の中で最大値を赤字、最小値を青字で示した。

※VIPが1以上となる項目については、セルを着色して強調した。

※VIPの標準偏差が1以上となる項目についても同様にセルを着色した（該当なし）。

VIPが1以上で、クラスタ間で値の高低に明確な差が見られ、その方向性が係数の符号と一致している変数を、特徴的な指標として取り上げた。各クラスタにおいて代表的な変数は以下の通りである。

- クラスタ0：1人当たり県民所得 (VIP=1.65、係数正→クラスタ内および他クラスタとの水準差は小さいが、東京都のみ突出している)
- クラスタ1：自殺率 (VIP=1.28、係数負→他クラスタと比べ水準が高い)
- クラスタ1, 2：人口千人当たり婚姻件数 (VIP=2.10, 1.54、係数正→他クラスタと比べて水準が低い)

以下では、これらの変数について、分布の特徴や他クラスタとの違いを踏まえて考察する。

① 1人当たり県民所得 該当クラスタ：0

クラスタ0では「1人当たり県民所得」がVIP=1.65と高く、重要な説明変数である。分布を見ると、特異な都道府県である東京都が極端に高く、全体のスケールを押し上げている一方で、クラスタ0～2間の水準差は小さい。また、クラスタ0～2内部では大きな格差は見られず、比較的均質な分布となっている。

このことから、所得のわずかな差であっても、クラスタ0内では目的変数に一貫して影響を与える可能性があり、その結果VIPが高くなっていると考えられる。

② 自殺率 該当クラスタ：1

クラスタ 1 では「自殺率」が VIP=1.28 と高く、目的変数に対する重要な説明変数である。係数は負であり、自殺率が高いほど予測される自然増減率が低くなる傾向を示している。

分布を見ると、クラスタ 1 は他クラスタと比較して自殺率が高く、広がりも大きいのが特徴である。高水準かつばらつきの大きい指標であることから、改善の余地が比較的大きい変数といえる。

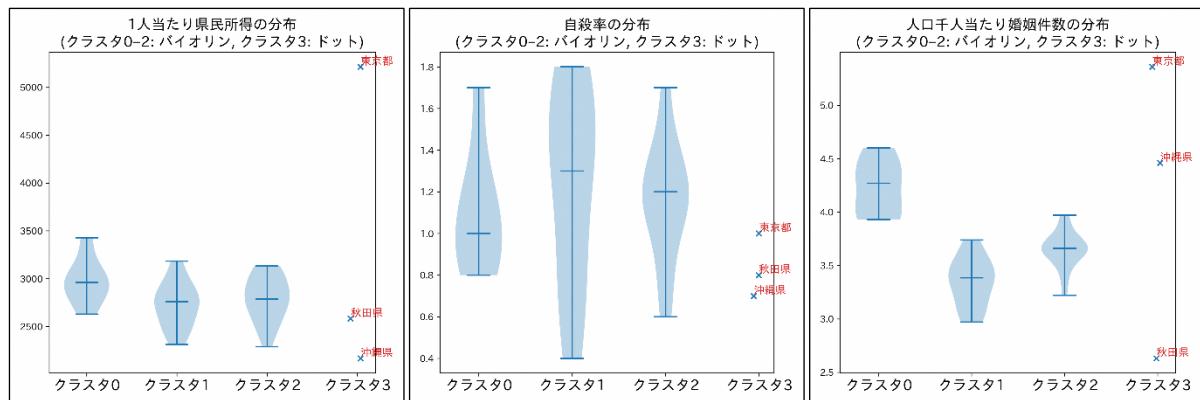
③ 人口千人当たり婚姻件数 該当クラスタ：1, 2

クラスタ 1 および 2 では「人口千人当たり婚姻件数」がそれぞれ VIP=2.10, 1.54 と高く、両クラスタにおいて目的変数に対する重要な説明変数となっている。いずれも係数は正であり、婚姻件数が多いほど自然増減率が高くなる傾向を示している。

クラスタ 1 では、中央値が低く、分布の下方に厚みがあることから、婚姻率の低さがクラスタ全体の課題として色濃く表れている。VIP の値も非常に高く、婚姻件数の引き上げが目的変数改善に直結する可能性が高いクラスタである。

一方クラスタ 2 では、中央値はクラスタ 1 より高いものの、クラスタ 0 と比べれば依然として水準は低めである。

分布も比較的まとまりがあるが、一定の改善余地を残しており、婚姻件数の増加は有効な施策対象と考えられる。

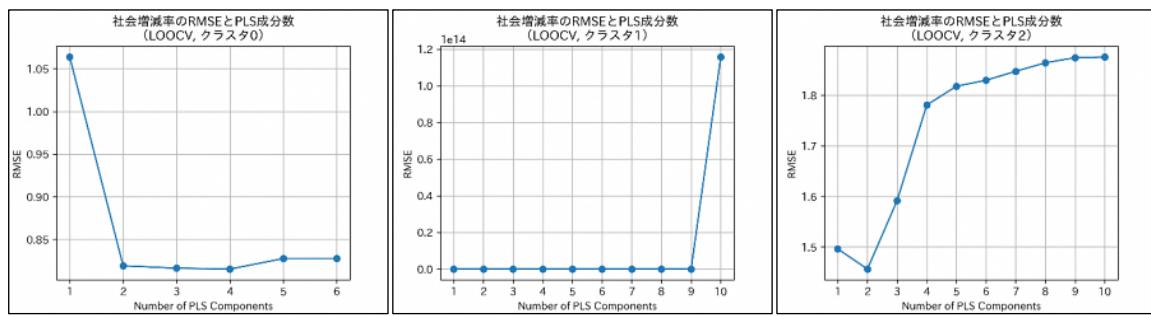


[図 6 : 自然増減率に対する主要説明変数のクラスタ別分布 (外れ値比較を含む)]

4.2.2 社会増減率の分析

図 7 に、各クラスタにおける社会増減率を目的変数とした PLS 回帰において、LOOCV を実施し、PLS 成分数と RMSE の関係を示す。RMSE の最小値は、クラスタ 0 で成分数 4、クラスタ 1 で成分数 1、クラスタ 2 で成分数 2 において確認された。

表 4 には、社会増減率に関する PLS 回帰の分析結果をまとめている。説明変数の影響の大きさは VIP に基づいて評価した。また、VIP の信頼性を確認するため、LOOCV によって得られた VIP のばらつきから標準偏差を算出した。加えて、多重共線性の影響を把握するために VIF を算出した結果、10 を超えるような著しく高い値は見られず、多重共線性の影響は限定的であると判断した。



[図 7 : クラスタごとの社会増減率の RMSE と PLS 成分数]

表4 社会増減率のPLS回帰の分析結果

説明変数	VIF	PLS係数			VIP			VIPの標準偏差		
		クラスタ0	クラスタ1	クラスタ2	クラスタ0	クラスタ1	クラスタ2	クラスタ0	クラスタ1	クラスタ2
1人当たり県民所得	3.51	-0.47	-0.05	0.41	1.20	0.50	1.71	0.39	0.24	0.21
可処分所得	1.71	0.36	-0.13	-0.03	1.76	1.43	0.12	0.17	0.24	0.28
月間有効求人人数（一般）	6.20	0.06	0.00	0.07	0.31	0.03	0.72	0.25	0.15	0.21
若年労働力率	2.23	0.00	0.12	-0.25	0.00	1.32	1.22	0.00	0.38	0.25
可住地駅密度（駅/10km ² ）	3.65	0.27	0.13	-0.34	0.42	1.40	1.73	0.30	0.22	0.22
通勤・通学	3.90	0.52	-0.03	0.20	1.79	0.30	0.85	0.14	0.22	0.13
持ち家比率	2.79	0.19	-0.15	0.13	0.50	1.58	0.53	0.24	0.10	0.18
公園面積比率	1.13	0.07	0.07	-0.16	0.75	0.73	0.60	0.31	0.22	0.20
人口10万人当たり図書館・博物館施設数	2.06	0.09	0.07	-0.18	0.57	0.75	0.77	0.32	0.29	0.19
人口千人当たり公営の借家数	2.83	-0.20	0.05	0.07	0.91	0.55	0.31	0.19	0.26	0.24

※VIFおよびPLS係数については、各変数の中で最大値を赤字、最小値を青字で示した。

※VIPが1以上となる項目については、セルを着色して強調した。

※VIPの標準偏差が1以上となる項目についても同様にセルを着色した（該当なし）。

VIPが1以上で、クラスタ間で値の高低に明確な差が見られ、その方向性が係数の符号と一致している変数を、特徴的な指標として取り上げた。各クラスタにおいて代表的な変数は以下の通りである。

- クラスタ0：可処分所得（VIP=1.76、係数正→分布の中央値は高いが、ばらつきがある）
- クラスタ1：若年労働力率（VIP=1.32、係数正→水準は他クラスタより低い）
- クラスタ2：可住地駅密度（駅/10 km²）（VIP=1.73、係数負→中央値は低めである）

以下では、これらの変数について、分布の特徴や他クラスタとの違いを踏まえて詳細に考察する。

① 可処分所得 該当クラスタ：0

クラスタ0では「可処分所得」がVIP=1.76と高く、目的変数に対して強く寄与する重要な説明変数である。分布を見ると、クラスタ0は他クラスタに比べて中央値がやや高く、全体として水準は良好である一方、上下方向へのばらつきも大きく、一部の都道府県では低水準にとどまっている様子も確認される。

このことから、クラスタ0内には可処分所得の底上げが期待できる自治体が含まれていると推察され、社会増減率の向上を目指す上で、可処分所得の改善は検討に値する方策の一つといえる

② 若年労働力率 該当クラスタ：1

クラスタ1では「若年労働力率」がVIP=1.32と比較的高く、目的変数との関係性が一定程度示唆された変数である。分布を見ると、クラスタ1の若年労働力率は**他クラスタと比べて水準が低く、また分布の集中度が高いことから、地域間での差があまり見られない状況にある。

一方で、クラスタ0（東京都を除く大都市圏とその周辺）が軒並み高い水準を示していることを踏まえると、クラスタ1の地域は、若年層を大都市圏に流出させている可能性があるとも考えられる。

このような構造は、就業機会や生活基盤を求める移動が特定の地域に集中していることを反映している可能性があり、今後の地域人口施策を検討する上で留意すべき視点となる。

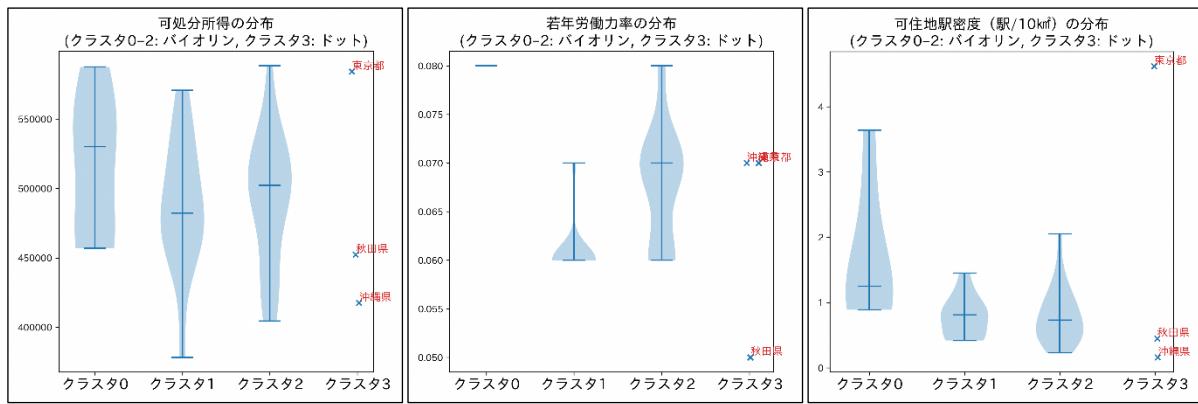
③ 可住地駅密度（駅/10km²） 該当クラスタ：2

クラスタ2では「可住地駅密度（駅/10km²）」がVIP=1.73と高く、目的変数に対する影響が大きい説明変数として抽出された。一方で係数は負となっており、駅密度が高い地域ほど社会増減率が低い傾向が示されるという、直感に反する結果が得られている。一般的に駅が多いことは交通利便性の高さや都市機能の充実を意味し、人口流入に好影響を与えると考えられるが、本分析の対象地域では、かえって他地域への流出を促進する要因となっている可能性も否定できない。

特に、駅ネットワークの発達が人口規模の小さい地域においては、地域外への移動のしやすさに偏り、地元定着の弱さに繋がっているという仮説も成り立つのではと考える。

この結果は、駅の整備そのものを否定するものではなく、交通利便性の活用とあわせて、地域内の生活環境や魅力の向上にも目を向ける必要があることを示唆している。

駅密度の高さが必ずしも人口維持につながらないケースもあることから、地域にとどまる理由づけとなるような要素の充実が、今後の検討課題の一つとなり得る。



[図 7 : 社会増減率に対する主要説明変数のクラスタ別分布 (外れ値比較を含む)]

5. 結果の解釈

本研究では、都道府県レベルのクラスタリングと PLS 回帰を通じて、人口増減に影響を与えている可能性のある変数について探索的な分析を行った。その結果、クラスタごとに異なる変数が比較的高い VIP 値を示し、地域の人口動態の背景には、それぞれ固有の要因が関わっていることがうかがえた。これは、全国一律の人口対策ではなく、地域の状況に応じた施策が必要であるという直感的な理解を、一定程度裏付ける結果となったといえる。

総じて、本分析はあくまで探索的なものであり、因果関係を明示的に示すものではない。しかし、クラスタごとに見られた傾向の違いは、地域特性に着目した施策検討の出発点として、一定の示唆を与えるものであると考えられる。今後は、市町村レベルでのより詳細な分析や、因果推論的アプローチを組み合わせることで、地域ごとの実情に即した施策立案につなげていくことが期待される。

参考文献

- [1] 守田 愛梨:「だし味添加ゼリーの粘弾性特性と官能評価によるおいしさ評価分析法の検討」, 日本食品工学会誌, Vol. 20, No. 2, pp. 71-79, 2019.
URL : https://www.jstage.jst.go.jp/article/jsfe/20/2/20_19545/_pdf
- [2] 藤原 幸一:『スマートデータ解析と機械学習』, オーム社, A5 判, 296 頁, ISBN 978-4-274-22778-3, 2022 年 2 月 19 日.