

EBPM の実現に向けた統計分析による都市の分類

(株)国土開発センター

正会員

○中宮 雅貴

(株)国土開発センター

非会員

新家 哲平

(株)国土開発センター

非会員

稲田 裕介

(株)国土開発センター

非会員

新田川貴之

(株)国土開発センター

非会員

森川 大輔

金沢工業大学

非会員

武市 祥司

1. 研究の目的

近年、ICT の発展により、ビッグデータ及びオープンデータの活用や、AI（人工知能）の実用化に向けた取組が進んでいる。公共分野においても、「電子行政オープンデータ戦略（2012 年 7 月）」の策定を契機として、国や地方公共団体等においてオープンデータ化への取組が活発になっており、公共データの公開とその利活用による経済や地域の活性化等が期待されている。

この背景を受けて、本研究は、一般的な単純集計及びクロス集計等の分析結果が政策判断に用いられている地方公共団体等において、ビッグデータ等のエビデンスを政策立案の根拠とする EBPM（Evidence-Based Policy Making）を推進する際の基礎資料としての活用を目指す。本稿では、「都市の特性を表す様々な統計量が整理されたオープンデータである都市モニタリングシート（国土交通省公表）のデータを使用した多変量解析等の分析による、人口以外の要素にも着目した都市の分類」について述べる。

2. 研究概要

2-1 使用データ

全国 1,719 都市（2017 年度の市区町村数、東京都区部は 1 都市とする）の基礎情報（人口・面積等）・交通・防災・産業・経済・財政・生活の利便性の指標・健康・福祉の指標などの 351 個の統計指標が記載されたオープンデータである都市モニタリングシートを用いる。

2-2 研究の流れ

①データ前処理

統計指標の欠損値の処理、データの標準化及び合成を行う。

②因子の抽出

多くある指標のうち、似た傾向のものを統合して集約するため、指標に影響を与える潜在因子を抽出する因子分析を行う。

③都市の分類

後工程で行う重回帰分析の精度を高めるため、②で抽出した因子間で、同様の傾向を示す都市を分類するクラスター分析（k-means 法）を行う。

④相関分析・モデル検証（重回帰分析）

③で分類された都市のデータを用いて、重回帰分析により指標と因子との関係性を明らかにする。

⑤都市の特徴の洗い出し

④までの分析結果から、特定の都市に着目し、その特徴を見出す。具体的には、重回帰分析により得られた都市のある指標に対する予測式と実測値の差分に着目し、その差分に及ぼす要因に関して分析・考察する。

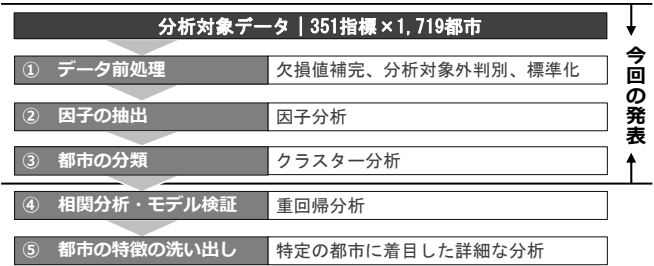


図 1 分析フロー

3. 分析結果

3-1 データ前処理

使用するデータには欠損が存在していたため、欠損が多い指標は分析対象から除外し、欠損が少ない指標は欠損値を平均値又は 0 で補完した。また、性質の似ている指標（例：H22 人口と H27 人口）は合成した。前処理の結果、分析に使用する指標は 86 種となった。

欠損値を補完したデータは、標準化（z-score normalization）を行った。

3-2 因子の抽出（因子分析）

因子分析では、86 種の指標から 17 個の因子が抽出された。ラベル付けを検討した結果、9 個がラベル付け可能であると判断した（表 1）。これにより、分析対象 86 種の指標から抽出された因子は 9 個となった。9 個の因子の寄与率は図 2 の通りである。

キーワード EBPM, ビッグデータ, オープンデータ, クラスター分析, 多変量解析, 重回帰分析

〒924-0838 石川県白山市八東穂 3-7（石川ソフトリサーチパーク内） TEL:076-274-8818, FAX:076-274-8821

表 1 因子抽出結果

因子	因子が影響する主な指標	因子負荷量
因子① 生活利便性	市街化区域人口密度	0.91
	駅バス停等徒歩圏人口密度	0.9
	自動車分担率（通勤通学）	-0.8
因子② 都市の規模	サービス・小売販売数等	0.97
	図書館・大学数	0.96
因子③ 都市の資産・投資	歳出～土木費等	0.88
	住民税・固定資産税等	0.88
因子④ 農業の振興 （活力）	農業産出額	0.81
	農林業経営体数	0.8
	鉄道分担率（通勤通学）	-0.23
因子⑤ 事業費	普通建設事業費	0.99
	単独事業費	0.94
因子⑥ 人口増減と負担	H22～H27人口変化率	0.84
	H22～H27昼間人口変化率	0.68
	1人あたり歳入・歳出・地方税額	-0.7
因子⑦ 土砂災害	土砂災害危険箇所面積	0.75
	H27自治体面積	0.59
	歩道設置率	-0.52
因子⑧ 富裕度合い	一人当たり税収額	0.7
	経常収支比率	-0.57
	実質公債費比率	-0.56
因子⑨ 地価変化	住宅地地価変化率	0.7
	商業地地価変化率	0.67
	津波浸水想定面積	-0.17

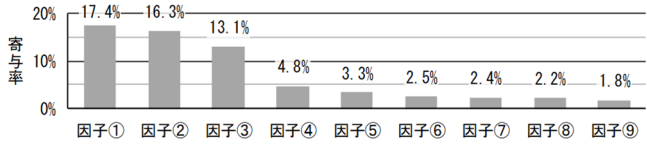


図 2 因子の寄与率

3-3 都市別分類（クラスター分析）

エルボー法による最適クラスター数の判定では、Gap 統計量が減少に転じる 18 クラスターを最適とした。次に、k-means 法によるクラスター分析を行い、各クラスターの特徴を把握した（一部表 2 に掲載）。

その結果、三大都市圏の主要都市で構成されるクラスター（No.1）、主に政令市で構成されるクラスター（No.2）、衛星都市・利便性の高い都市が多いクラスター（No.3,4）等、因子①「生活利便性」、因子②「都市の規模」が主要なファクターとみられる分類があった。

また、原子力発電所やダム等の固定資産税が大きな

施設が立地、あるいは観光産業が活発な小規模都市で構成されるクラスター(No.5,6)も分類された。このクラスターは因子⑧「富裕度合い」が主要ファクターと考えられる。

その他、北海道の炭鉱都市や中山間地等の過疎化・衰退傾向のある都市が多いクラスター(No.7,8)等が分類され、都市規模以外の因子も分類に影響していることが確認された。

さらに、抽出した因子から特徴づけにくい都市で構成されるクラスター(No.9)もあった。

4. 今後の課題と展望

今後は、重回帰分析により得られる因子と指標との関係について、実用性を検証する。

一方、今回の研究では、クラスター分析の結果、人口に強く影響を受けて、中小規模の多くの都市で構成されるクラスターが現れるという課題が明らかになった。今後、同クラスター内で再度クラスター分析を行うことや分析対象データの見直し等、分類の精度を高めるためのデータ処理方法を検討していきたい。

さらに、回帰式から都市の課題となっている因子を導き出し、例えば教育・子育て支援に関する因子と年少人口の関連性の分析等、将来予測に活用するための課題に取組み、本研究が政策立案のエビデンスとして用いるに足る信頼性、有用性を備える事を目指す。

将来的には、都市モニタリングシート以外の統計データや過去の統計データ、画像・音声等の非構造化データを追加し、分析精度の向上を目指していきたい。

表 2 クラスター分析結果（18 クラスターの一部抜粋）

No.	← 各因子が取りうる値 → 1 5 9 低 高	① 生活利便性	② 都市の規模	③ 都市の資産	④ 農業の振興	⑤ 事業費	⑥ 人口増減	⑦ 土砂災害	⑧ 富裕度合い	⑨ 地価変化	クラスター内の都市数				主な都市
											市	区	町	村	
1	三大都市圏の主要都市で構成 (特別区部、横浜、大阪、名古屋)	8	9	9	8	5	6	5	6	8	3	1	0	0	東京都特別区部、横浜市、名古屋市、大阪市
2	主に政令市で構成 (財政規模が大きい都市)	7	8	9	8	5	6	6	5	7	13	0	0	0	札幌市、仙台市、さいたま市、川崎市、浜松市、 豊田市、京都市、神戸市、岡山市、福岡市 等
3	地方の衛星都市・利便性の高い都市 (5〜30万人程度の市町)	6	5	5	4	5	6	5	5	6	135	0	79	5	室蘭市、名取市、春日部市、印西市、 栗東市、三田市、大和高田市、太宰府市 等
4	三大都市圏の衛星都市・利便性の高い都市 (5〜50万人程度)	7	5	5	5	5	6	5	5	6	94	0	2	0	川越市、船橋市、三鷹市、長久手市、西宮市、 府中町、大野城市、那覇市 等
5	原発・観光等での財政が豊かな都市 (1万人程度の町村)	4	5	5	4	5	5	5	8	5	1	0	25	15	占冠村、東通村、刈羽村、美浜町、 おおい町、軽井沢町、白川村、王滝村 等
6	原発・観光等での財政が豊かな都市 (5千人未満の小規模村)	4	5	5	4	5	5	4	9	5	0	0	0	5	泊村、檜枝岐村、上野村、南相木村、飛島村
7	過疎化・衰退傾向のある都市 (旧炭鉱都市など)	4	5	5	5	5	5	5	5	3	39	0	180	10	夕張市、美唄市、三笠市、下仁田町、 大川市 等
8	山間部・離島等に立地する都市 (小規模町村)	4	5	5	5	4	5	5	4	5	15	0	161	84	陸前高田市、宿毛市、木古内町、津南町、池田町、 高野町、周防大島町、阿智村、栄村、姫島村 等
9	抽出した因子から特徴づけにくい都市 (小〜中規模都市)	5	5	5	4	5	5	5	5	5	42	0	203	59	勝浦市、八丈町、見附市、滑川市、養老町、 垂井町、関ヶ原町、福岡町 等

以下略