

平成 28 年経済センサス - 活動調査に関する
地域メッシュ統計を利用した地域分析

II 分析の結果

任意の地域単位の産業別従業者割合の主成分分析による
産業構造の分布と動向

令和 3 年 10 月

公益財団法人 統計情報研究開発センター

目 次

1	分析の概要	1
1.1	分析の背景	1
1.1.1	産業構造の変遷	1
1.1.2	可変単位地区問題	2
1.2	分析の対象と目的	3
2	分析	5
2.1	2変量の主成分分析－知識産業の分布－	5
2.1.1	主成分分析とは	5
2.1.2	2変量の場合の主成分の導出	7
2.1.3	主成分分析の解析の流れ	9
2.1.4	市区町村別集計を利用した主成分分析	9
2.1.5	地域メッシュ統計を利用した主成分分析	17
2.2	2012～16年の動向	28
3	まとめと今後の課題	30
3.1	まとめ	30
3.2	今後の課題	30
	参考文献	31

1 分析の概要

1.1 分析の背景

1.1.1 産業構造の変遷

イギリスの経済学者コーリン・クラークは産業を第一次産業、第二次産業、第三次産業の3部門に分け、労働力の移動に関する分析を行った。第一次産業は採取産業グループで農業、水産業などが含まれ、第二次産業は加工産業グループで、鉱業、建設業、製造業の部門、第三次産業は以上の産業を除いた非物財産業グループで、卸売業、小売業、宿泊業、飲食サービス業などが含まれる。

イギリスの経済学者ウィリアム・ペティはその著作『政治算術』の中で「農業よりは製造業が、また、製造業よりも商業（Merchandize）がずっと多くの利得がある」と指摘し、経済が発展するにつれて利得が大きい産業へ人々が移り、産業構造の比重が第一次産業から第二次産業そして第三次産業へと移り、産業構造が変わっていくことを示唆している。

ペティの考え方をさらに発展させたのがコーリン・クラークで、経済の発展とともに産業間の所得格差が原因となって、第一次産業から第二次産業へ、さらに第三次産業へと労働力が移動していることを、イギリスやアメリカなど10数カ国の統計資料により明らかにした。クラークは著書『経済進歩の諸条件』の中で所得水準の上昇とともに、第一次産業の労働人口の比重は低下し、第二次・第三次産業の労働人口の比重が上昇することを示した。

図1は国勢調査のデータを利用して、日本の1960年～2015年における産業別就業者数の割合の推移をグラフ化したものである。日本においては、第一次産業、第二次産業が減少傾向にあり、第三次産業が増加している様子が確認できる。

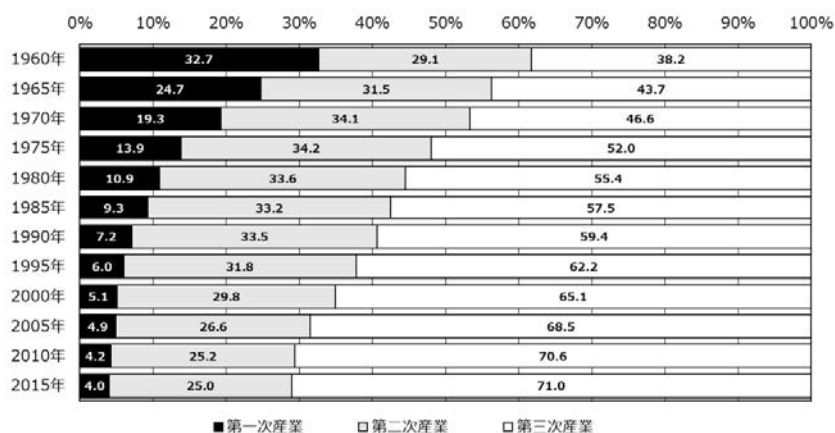


図1 産業別就業者数割合の推移（全国、1960～2015年）

平成 27 年国勢調査のデータを利用して、第三次産業就業者数の全就業者数に占める割合が高い上位 5 都県について確認すると、最も割合が高いのは東京都の 82.1%、次いで沖縄県の 80.0%、神奈川県 of 76.7%、千葉県の 76.5%、福岡県の 75.9%である。5 都県の合計の産業大分類別就業者数の割合に対する各都県の産業大分類別就業者数の割合の特化係数を確認すると、東京都においては「G情報通信業」、「L学術研究，専門・技術サービス業」、「K不動産業，物品賃貸業」が高く、千葉県においては「H運輸業，郵便業」が高い。沖縄県では「M宿泊業，飲食サービス業」、「Rサービス業」、「S公務」が高く、「P医療，福祉」は福岡県、沖縄県で共に高い。第三次産業就業者数の割合が高い都県でも、特化している産業は、地域によって異なる（表 1）。

アメリカの経済学者フリッツ・マハループは「アメリカにおける知識の生産と流通」で産業論の立場で知識を論じている。マハループは企業、研究所、団体などの生産するものが、主として情報ないし知識である場合を「知識産業」と定義し、①教育、②研究と開発、③コミュニケーションのメディア、④情報機械、⑤情報サービスの五つに分類した。そして、知識の総生産額が国民総生産に占める割合を計算し、その割合が年々増加していること、また、国民総生産よりも成長が速いことを示した¹。

本研究では、知識産業として「G情報通信業」、「L学術研究，専門・技術サービス業」に着目しながら、東京圏、名古屋圏、大阪圏における従業者数による産業構造の地域分布と動向の把握を試みる。

表 1 第三次産業の割合が高い都県における産業別就業者数の特化係数

	特化係数				
	千葉県	東京都	神奈川県	福岡県	沖縄県
G情報通信業	0.80	1.34	1.06	0.43	0.39
H運輸業，郵便業	1.28	0.81	1.06	1.08	0.78
I卸売業，小売業	1.06	0.93	1.01	1.12	0.93
J金融業，保険業	1.01	1.20	0.90	0.76	0.60
K不動産業，物品賃貸業	0.85	1.22	0.98	0.73	0.70
L学術研究，専門・技術サービス業	0.80	1.24	1.02	0.68	0.65
M宿泊業，飲食サービス業	0.97	1.00	0.98	0.98	1.37
N生活関連サービス業，娯楽業	1.13	0.95	0.97	1.00	1.07
O教育，学習支援業	0.96	0.99	1.01	1.02	1.17
P医療，福祉	0.96	0.86	1.00	1.33	1.30
Rサービス業（他に分類されないもの）	1.01	1.00	0.99	0.96	1.20
S公務（他に分類されるものを除く）	1.08	0.91	0.91	1.10	1.79

（注）第 3 次産業のうち従業者数が少ない「F電気・ガス・熱供給・水道業」、「Q複合サービス事業」は除外した

1.1.2 可変単位地区問題

経済センサス - 活動調査の集計結果は全国、都道府県、市区町村、町丁大字、地域メッシュな

¹ マハループは限られた資料で算出していることを留意点として挙げている。

どの大きさの異なる様々な地域別に集計（地域メッシュ統計は編成）されている。このような地域別に集計された統計データを利用して分析する際の地域単位をここでは「分析地域単位」とする。また、分析を行うに当たり、分析対象地域を定める必要があるが、この分析対象地域を「分析地域範囲」とする。

地域別の統計データを利用した分析においては、研究の目的に応じて分析地域単位と分析地域範囲を設定するが、地域別の統計データを用いた分布図の作成や、各種の空間データ解析を行う際に、分析地域単位と分析地域範囲の違いが分析結果に違いをもたらすことがある。これらの分析単位によって結果が変わる問題は空間スケール問題といわれ、可変単位地区問題（MAUP：Modifiable Area Unit Problem）のひとつとして知られている（Openshaw(1984)、中谷(2015)）。

可変単位地区問題には、分析地域単位問題や分析地域範囲問題のような空間スケール問題以外にゾーニング問題もある。ゾーニング問題とは、同じ程度の大きさの分析地域単位であっても、その分析地域単位の作り方（境界設定の仕方）によって分析結果が変わってしまうものである。

本研究は、経済センサス - 活動調査の市区町村別集計、地域メッシュ統計を利用して、主成分分析におけるゾーニング問題の影響について、知識産業の地域分布をテーマとして整理する。

1.2 分析の対象と目的

本研究の分析対象項目は平成 24 年、平成 28 年の経済センサス - 活動調査に関する地域メッシュ統計の編成項目のうち、表 2 に示す項目である。なお、「従業者総数」は産業大分類の「A～B 農林漁業」、「S 公務」を除いた「C～R 非農林漁業」の従業者数とする。また、本研究の分析対象地域を東京圏（埼玉県、千葉県、東京都（島嶼部を除く）、神奈川県）の 243 市区町村）、名古屋圏（名古屋市役所から概ね半径 50 km 圏にかかる 106 市区町村（図 2））、大阪圏（大阪市役所から概ね半径 50 km 圏にかかる 157 市区町村（図 3））とする²。

本研究の分析の目的は、半径 4,000m の円内に含まれる 2 分の 1 地域メッシュ別の統計値を距離減衰関数による重みをかけて任意の地域で集計した結果について主成分分析を行い、東京圏、名古屋圏、大阪圏における知識産業の地域分布や動向について把握することである。

² 東京圏は島嶼部を除く 4 都県の全市区町村を含めて分析したが、名古屋圏と大阪圏は各市役所から 50 km 圏にかかる都道府県全域を含めると東京圏と比較して広範囲になるため、市区町村を抽出した。



图2 名古屋圈市区町村



图3 大阪圈市区町村

表 2 分析対象項目

調査項目	統計項目	項目符号	計算処理
従業者 総数	C~R 従業者総数 (総数)	CREM	2012~16 年従業者数増 減率
産業別従 業者数	C05 鉱業、採石業、砂利採取業従業者寄与度 (総数)	0251	2012~16 年産業別従業 者数の寄与度
	D建設業従業者数 (総数)	0254	
	E製造業従業者数 (総数)	0266	
	F電気・ガス・熱供給・水道業従業者数 (総数)	0344	
	G情報通信業従業者数	0359	
	H運輸業、郵便業従業者数	0377	
	I 1 卸売業従業者数	0407	
	I 2 小売業従業者数	0428	
	J金融業、保険業従業者数	0464	
	K不動産業、物品賃貸業従業者数	0479	
	L学術研究、専門・技術サービス業	0515	
	M75 宿泊業従業者数	0533	
	M76 飲食店従業者数	0542	
	M77 持ち帰り・配達飲食サービス業	0545	
	N生活関連サービス業、娯楽業従業者数	0548	
	O教育、学習支援業従業者数	0608	
	P 83 医療業従業者数	0659	
	P 84 保健衛生従業者数寄与度	0674	
P 85 社会保険・社会福祉・介護事業	0677		
Q複合サービス業	0698		
Rサービス業従業者数寄与度	0707		

2 分析

2.1 2変量の主成分分析 - 知識産業の分布 -

2.1.1 主成分分析とは³

表3のようにn個の個体についてp個の特性値(変量)が観測されている。例えば、東京圏の市区町村別、産業別の従業者数の割合などが得られている場合を想定する。

表3 n個体についてのp変量の観測値

	x_1	x_2	...	x_p
1	x_{11}	x_{21}	...	x_{p1}
2	x_{11}	x_{21}		x_{p2}
...		
n	x_{1n}	x_{2n}		x_{pn}

³ 2.1.1 主成分分析の説明については、田中、脇本(1983)を参考にした。

これら $x_1, x_2 \dots x_p$ を代表する総合的指標を求めたい、という場合を考える。総合的指標とは、入学試験の場合ならば各科目の成績を総合した総合的成績、体力テストの場合ならば各種の体力テストの結果を総合した総合的体力、会社を比較する場合ならば種々の財務指標にもとづく企業の評価、を求めたいという場合である。

主成分分析法はこのような目的に利用され、 p 個の変量 $x_1, x_2 \dots x_p$ だけの情報からそれらの変量を代表する合成変量を求める方法である。また、多くの変量が存在する場合に次元を削減して、データの情報をより解釈しやすくするための方法ともいえる。

本項では $p=2$ の場合について考える。地域別の「D建設業」、「E 製造業」従業者数の従業者総数に占める割合のデータ x_1, x_2 があるとき、これらを総合した指標として従業者数の合計の割合が用いられる。

$$(II.1) \quad z = x_1 + x_2$$

合計した値を用いるということは、(II.1) の和が一定の (x_1, x_2) はすべて同じ成績とみなそうという考え方である。

ところで、東京圏の市区町村別「E 製造業」従業者数の割合は 1.02%~53.10%の範囲にばらついているのに、「D建設業」従業者数の割合は 0.66%~14.58%の範囲にかたまっているという場合には、合計の割合の大小は主として x_1 によって定まり、 x_2 の影響の度合いは小さい。このように単純な合計を用いるのは、見かけ上 x_1 と x_2 に等しい重みをかけているように見えるが、 x_1 と x_2 の大きさは一般に等しくなく、そのばらつきの大小に依存しているのである。

主成分分析では、 x_1, x_2 に対する重み（係数） a_1, a_2 を積極的に持ち込んで

$$(II.2) \quad z = a_1 x_1 + a_2 x_2$$

の形の合成変量を考える。ここで、係数 a_1, a_2 はこの z が x_1, x_2 をできるだけ「よく代表する」ように定められる。次にもとの変量 x_1, x_2 をよく代表する z を求めるための基準について述べる。

2.1.2 2変量の場合の主成分の導出⁴

変量 x_1, x_2 の平均、分散及び共分散を

$$(II.3) \quad \left\{ \begin{array}{l} \text{平均} \quad \bar{x}_1 = \frac{1}{n} \sum_{i=1}^n x_{1i}, \quad \bar{x}_2 = \frac{1}{n} \sum_{i=1}^n x_{2i} \\ \text{分散} \quad s_{11} = \frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2, \quad s_{22} = \frac{1}{n} \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 \\ \text{共分散} \quad s_{12} = s_{21} = \frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \end{array} \right.$$

とする。このとき (II.2) の形の合成変量 z の分散は

$$(II.4) \quad \begin{aligned} V_z &= \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \{a_1(x_{1i} - \bar{x}_1) + a_2(x_{2i} - \bar{x}_2)\}^2 \\ &= a_1^2 s_{11} + 2a_1 a_2 s_{12} + a_2^2 s_{22} \end{aligned}$$

と表される。

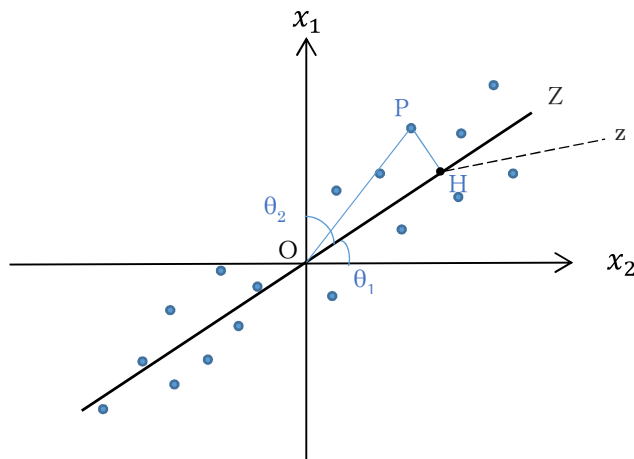


図4 (x_1, x_2) のプロット

図4のように各点 (x_1, x_2) をプロットしたとき、もし観測値の点の集まりが1本の直線に十分近ければ、2次元平面上の点のちらばりはこの直線方向の(1次元)の点のちらばりで代表できると考え、そのような意味でこれらの点のちらばりの最も大きい方向、すなわ

⁴ 2.1.2 主成分の導出については、田中・脇本 (1983)、永田・棟近 (2001) を参考にした。

ち図の OZ の方向をみつけ、各点の直線 OZ 方向の成分を総合的指標として用いると考える。

いま OZ と x_1 軸, x_2 軸のなす角をそれぞれ $\theta_1, \theta_2 (=90^\circ - \theta_1)$ とすれば、 θ_1, θ_2 平面上の任意の点 P (x_1, x_2) の OZ 軸上の座標 z (点 P から OZ におろした垂線の足 H までの原点 O からの長さ) は

$$(II.5) \quad z = x_1 \cos \theta_1 + x_2 \cos \theta_2 = x_1 \cos \theta_1 + x_2 \sin \theta_1$$

と表される。これを (II.2) と比較すると、 a_1, a_2 はそれぞれ $\cos \theta_1, \cos \theta_2$ に当たり、

$$(II.6) \quad a_1^2 + a_2^2 = 1$$

のような条件を満たす。したがって問題は (II.6) の制約条件のもとで (II.4) の V_z を最大にするような a_1, a_2 を求めることに帰着する。この問題は制約付きの最大化問題を求める方法であるラグランジュの未定乗数法を用いて (II.7) に変形でき、求める解 a_1, a_2 は (II.7) を満たさなければならない。

$$(II.7) \quad \begin{aligned} a_1 s_{11} + a_2 s_{12} &= 2 \lambda a_1 \\ a_1 s_{12} + a_2 s_{22} &= 2 \lambda a_2 \end{aligned}$$

ここで、分散共分散行列 V 、係数ベクトル \mathbf{a} を

$$V = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

とおけば、(II.7) は

$$(II.8) \quad \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \text{あるいは } V \mathbf{a} = \lambda \mathbf{a}$$

と表されるが、(II.8) は λ が分散共分散行列 V の固有値であり、求めるべき a_1, a_2 は固有ベクトルであることを示している。ここで、(II.8) に左からベクトル $\mathbf{a} = (a_1, a_2)$ をかけると、

$$(II.9) \quad a_1^2 s_{11} + 2a_1 a_2 s_{12} + a_2^2 s_{22} = \lambda (a_1^2 + a_2^2)$$

を得る。この式の左辺は V_z であり、右辺は (II.6) の制約条件より λ に等しい。すなわち、

$$(II.10) \quad V_z = \lambda$$

である。以上より、(II.4) の V_z を最大化することは、「分散共分散行列 V の固有値問題を解いて、最大固有値に対応する固有ベクトル $\mathbf{a} = (a_1, a_2)$ を求めれば、それが V_z の最大値を与える (a_1, a_2) であり、 V_z の最大値は λ となる」という手続きで実行できる。

V_z の最大値は z の分散の最大値であり、これが第 1 主成分となるので、添え字をつけて z_1 と表す。第 1 主成分 z_1 だけでデータの情報を十分説明できないときは、第 2 主成分 z_2 を考慮する。第 2 主成分 z_2 は、すでに定まっている第 1 主成分 z_1 に含まれない情報を追加するために導入するので、 z_1 と無相関になるように定める。導出は省略するが第 2 主成分も分散共分散行列 V の固有値問題を解くことで得られる。 V_z の最大値は V の 2 番目に大きな固有値 λ_2 となり、第 2 主成分 z_2 の構成には λ_2 に対応する固有ベクトル $\mathbf{b} = (b_1, b_2)$ を用いる。

なお、本稿の主成分分析の計算は、統計ソフトウェアや python プログラムで実行した。

2.1.3 主成分分析の解析の流れ

主成分分析の解析は以下の手順で行った。

- ① 分散共分散行列 V の第 1 固有値 (最大固有値) λ_1 に対応する固有ベクトルから第 1 主成分 z_1 を求める。次に、 V の第 2 固有値 λ_2 に対応する固有ベクトルから第 2 主成分 z_2 を求める。同様にして、多変量の場合は第 k 主成分を求める。
- ② それぞれの主成分の寄与率および累積寄与率を求める。「固有値」が 1 以上ないしは「累積寄与率が 70% を超える」を目安として主成分を選択する。
- ③ 固有ベクトルの値を参考にして、選択した各主成分の意味について考察する。
- ④ 主成分得点を地図化し、地域の特徴付けや分類を行う。

2.1.4 市区町村別集計を利用した主成分分析

東京圏、名古屋圏、大阪圏における平成 28 年経済センサス・活動調査の市区町村別集計を利用して、「G 情報通信業」、「L 学術研究, 専門・技術サービス業」従業者数の割合の総合的指標を「知識産業」として算出する。

① 基本統計量

最初に「G 情報通信業」、「L 学術研究, 専門・技術サービス業」従業者数の割合の基本統計量と統計地図を確認する。表 4 は東京圏、名古屋圏、大阪圏における基本統計量をまとめたものである。

平均についてみると、東京圏における「G 情報通信業」は 1.91% で、「L 学術研究, 専門・技術サービス業」の 2.98% と比較して小さい。名古屋圏、大阪圏は東京圏と比較して水準が低く、名古屋圏は 0.74% と 2.04%、大阪圏は 0.85%、2.08% となっている。

標準偏差についてみると、東京圏における「G 情報通信業」は 3.39 で「L 学術研究, 専門・技術サービス業」の 2.97 よりも大きい。同様に名古屋圏は、1.52、1.32 と「G 情報通

信業」の方が大きいですが、大阪圏は 1.60、1.75 と「L 学術研究，専門・技術サービス業」の方が大きい値を示す。しかしこれらの差はあまり大きくなく、それぞれの平均の水準が異なるのでばらつきの程度を比較しにくい。

これらの比較のために、東京圏、名古屋圏、大阪圏において、標準偏差を平均で割った変動係数について確認する。東京圏の「G 情報通信業」の変動係数は 1.77、「L 学術研究，専門・技術サービス業」の変動係数は 1.00 となっており、「G 情報通信業」のばらつきが大きい。同様に名古屋圏、大阪圏についても変動係数を確認すると、名古屋圏は $2.07 > 0.65$ 、大阪圏は $1.89 > 0.84$ となっており、全ての地域（以下「3 圏」という）で「G 情報通信業」のばらつきが大きい。

最小は 3 圏の両項目においてほぼ 0% となっており、これらの産業の従業者数がほとんどいない市区町村もあることがわかる。また、最大の水準は東京圏が 20% 前後と高いのに対して、名古屋圏、大阪圏は 10% 前後かそれよりも低い値を示しており、名古屋圏、大阪圏の「G 情報通信業」、「L 学術研究，専門・技術サービス業」の従業者数の割合は東京圏の半分程度の水準となっている。

表 4 基本統計量

	東京圏		名古屋圏		大阪圏	
	G 情報通信業	L 学術研究， 専門・技術サービス業	G 情報通信業	L 学術研究， 専門・技術サービス業	G 情報通信業	L 学術研究， 専門・技術サービス業
データ数	243		106		157	
平均	1.91	2.98	0.74	2.04	0.85	2.08
標準偏差	3.39	2.97	1.54	1.32	1.60	1.75
分散	11.52	8.82	2.36	1.75	2.56	3.07
最小	0.00	0.05	0.00	0.19	0.00	0.00
最大	19.23	23.21	10.03	6.49	9.97	11.85
変動係数	1.77	1.00	2.07	0.65	1.89	0.84

②従業者数割合の統計地図

図 5-1、図 5-2 は東京圏における「G 情報通信業」と「L 学術研究，専門・技術サービス業」の従業者数割合の統計地図、図 6-1、図 6-2 は名古屋圏、図 7-1、図 7-2 は大阪圏の統計地図である。全ての統計地図について比較を行いやすいように塗分けの凡例を揃えて表示した。青系→緑系→黄色系→赤系のグラデーションを用いて塗り分け、青系は従業者数の割合が低く、赤系は従業者数の割合が高いことを示している。

図 5-1、図 5-2 の東京圏の統計地図をみると、赤系で塗り分けられた市区町村が分布しており、名古屋圏、大阪圏と比較して従業者数の割合が高い。図 5-1 の「G 情報通信業」は、割合の高い市区町村が都心部に集中しているのに対して、図 5-2 の「L 学術研究，専門・技術サービス業」は、都心部の他に郊外にも分布しており、「G 情報通信業」と比較して青系の色合いも明るく、東京圏に広く分布している。

名古屋圏、大阪圏は、東京圏と比較すると赤系の市区町村の数が少ない。また、「G 情報通信業」の従業者数の割合は、多くの市区町村において1%未満の低い値を示す。

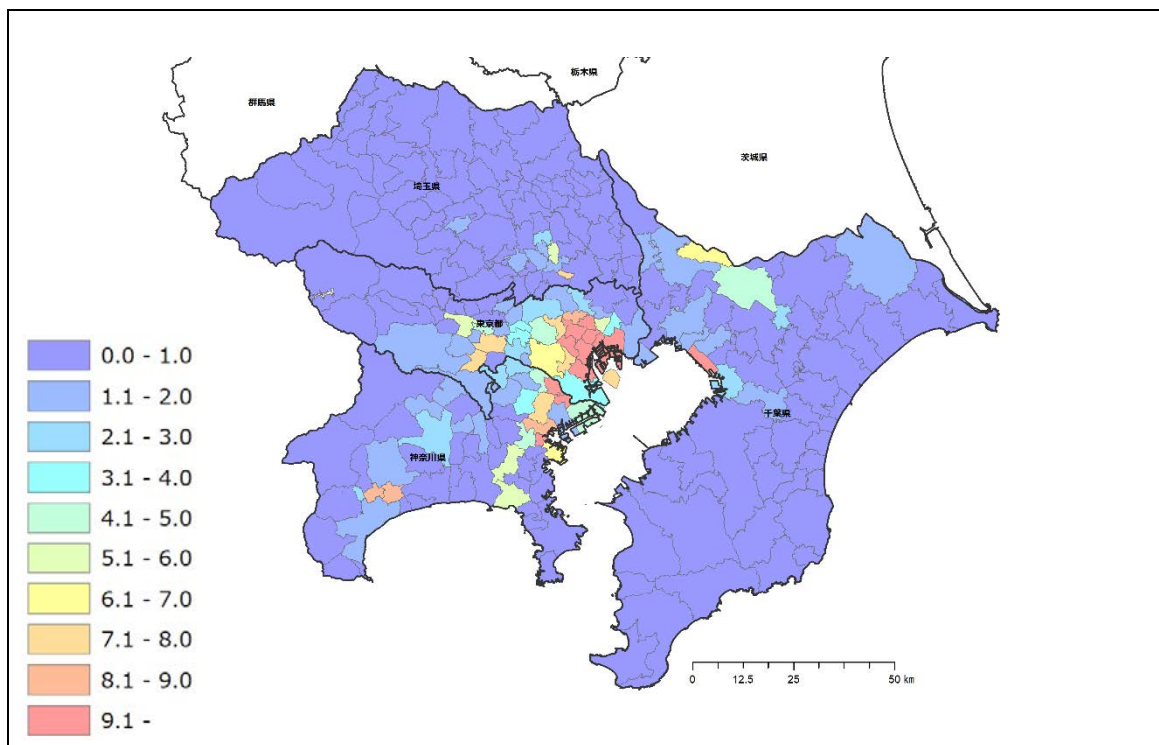


図 5-1 G 情報通信業従業者数割合（東京圏）

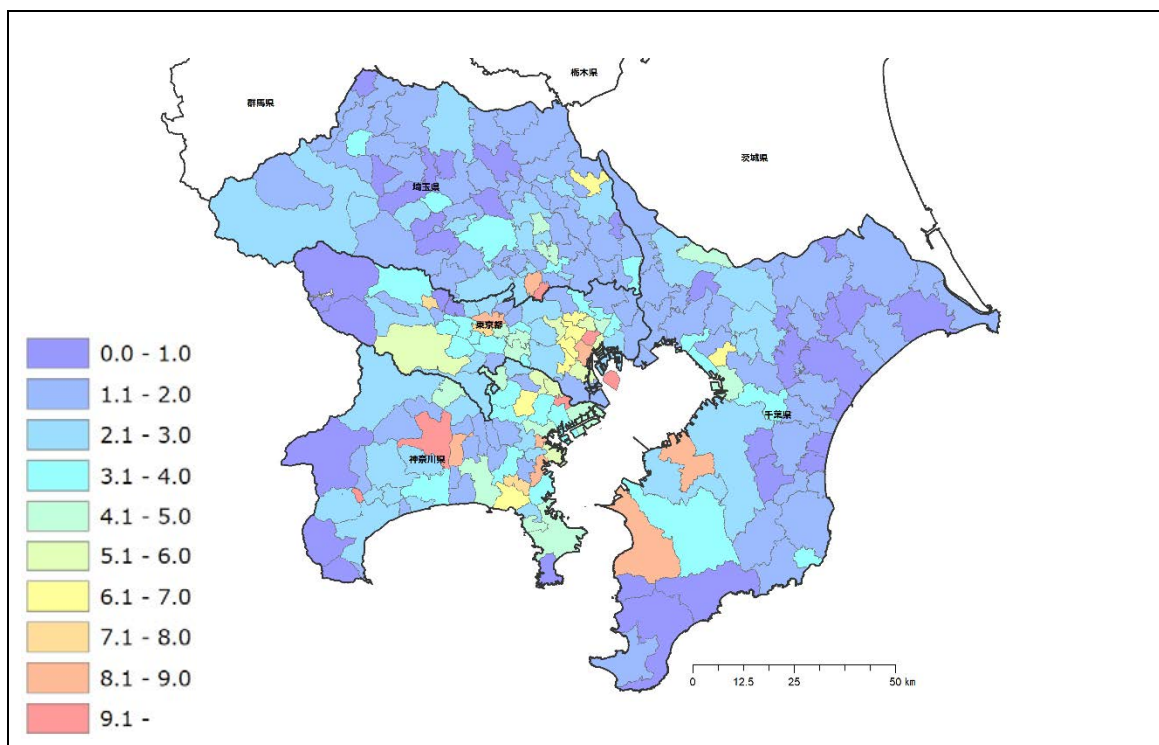


図 5-2 L 学術研究，専門・技術サービス業従業者数割合（東京圏）

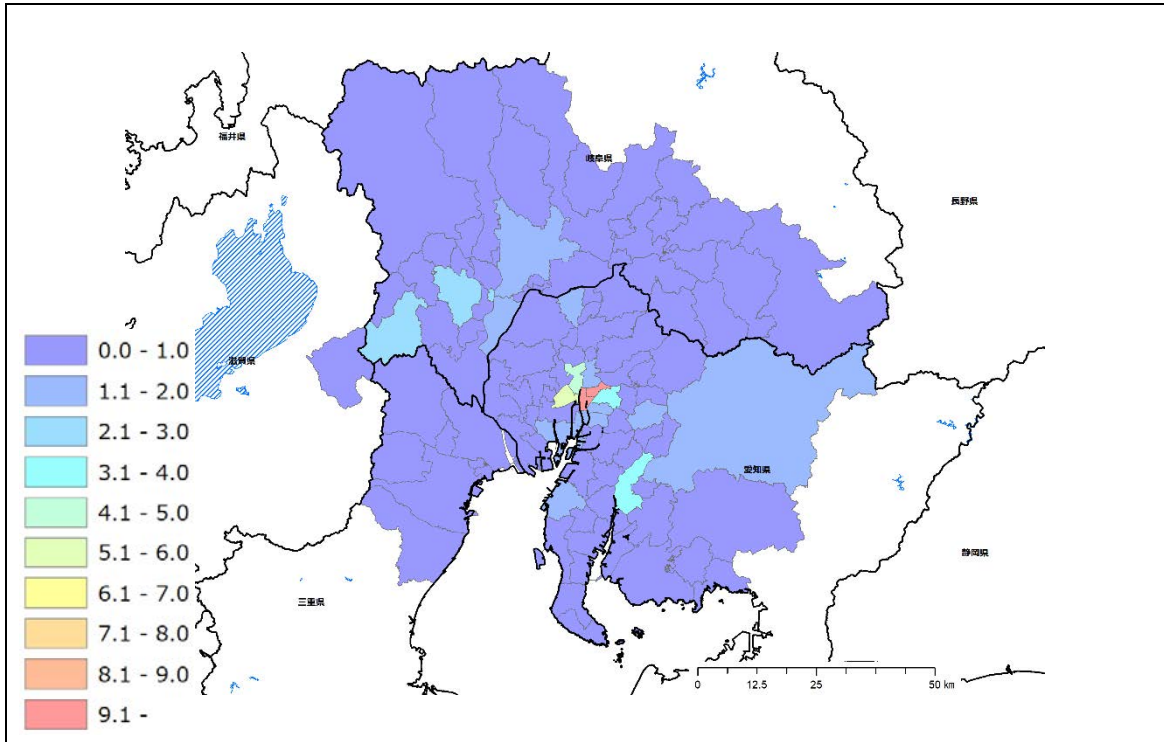


図 6-1 G 情報通信業従業者数割合（名古屋圏）

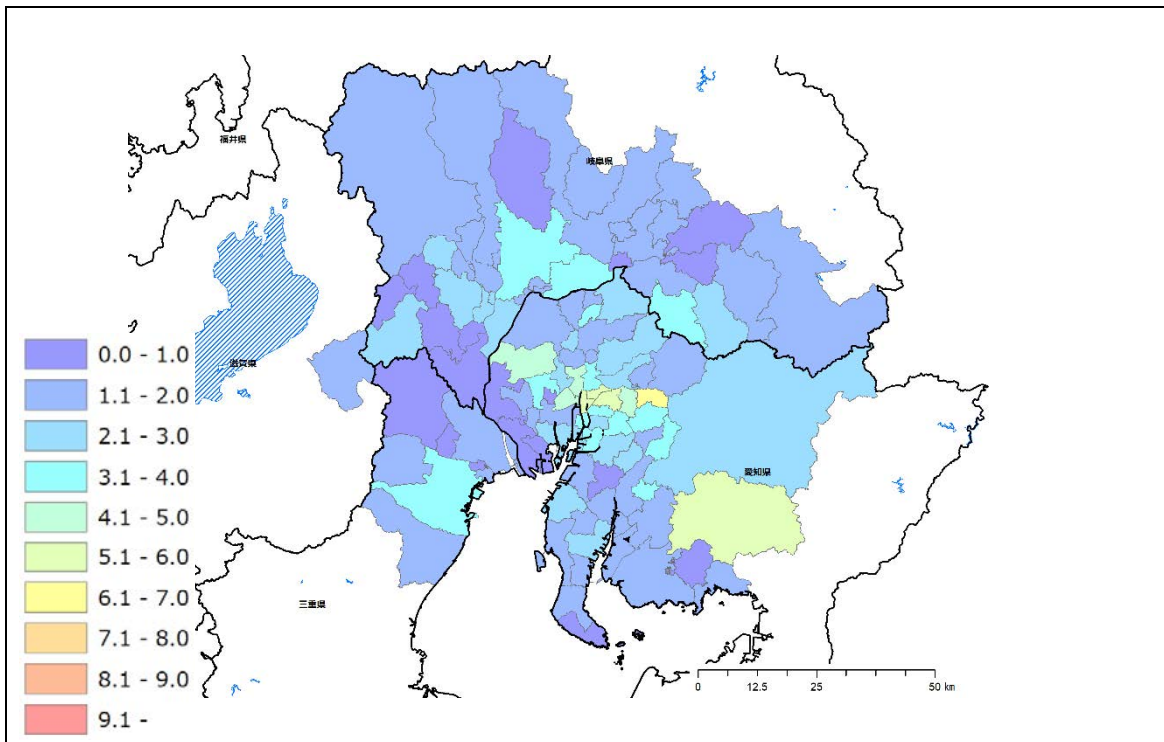


図 6-2 L 学術研究，専門・技術サービス業従業者数割合（名古屋圏）

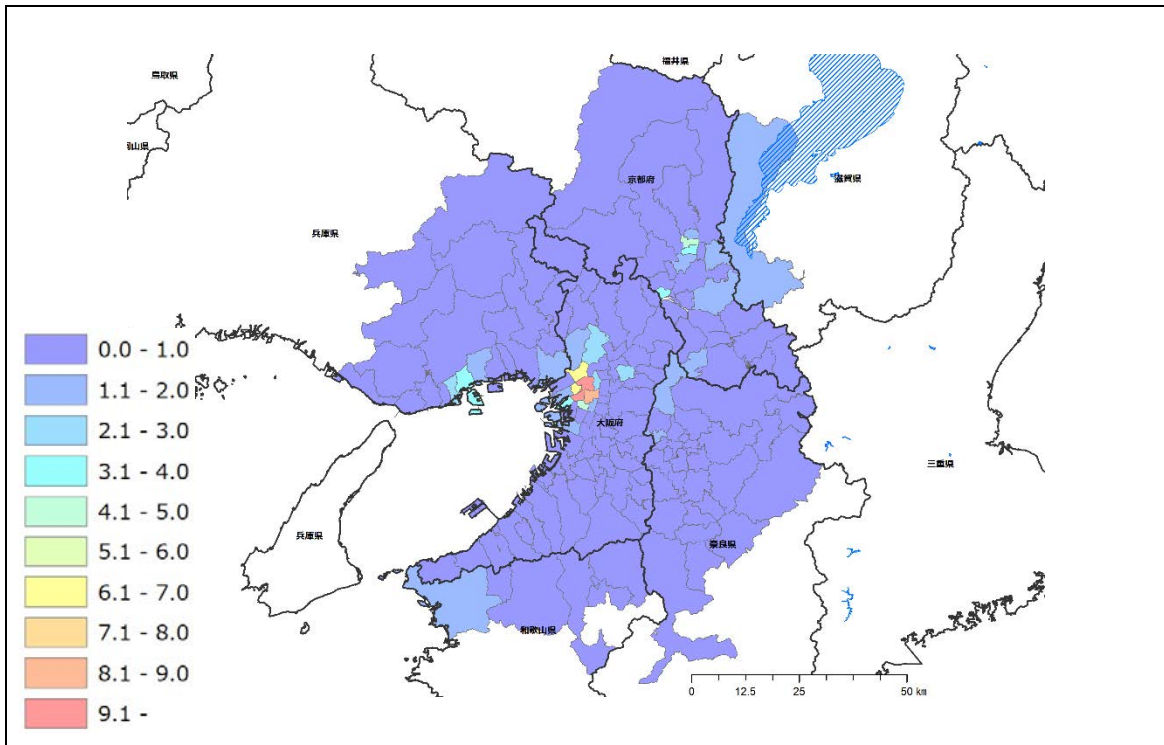


図 7-1 G 情報通信業従業者数割合（大阪圏）

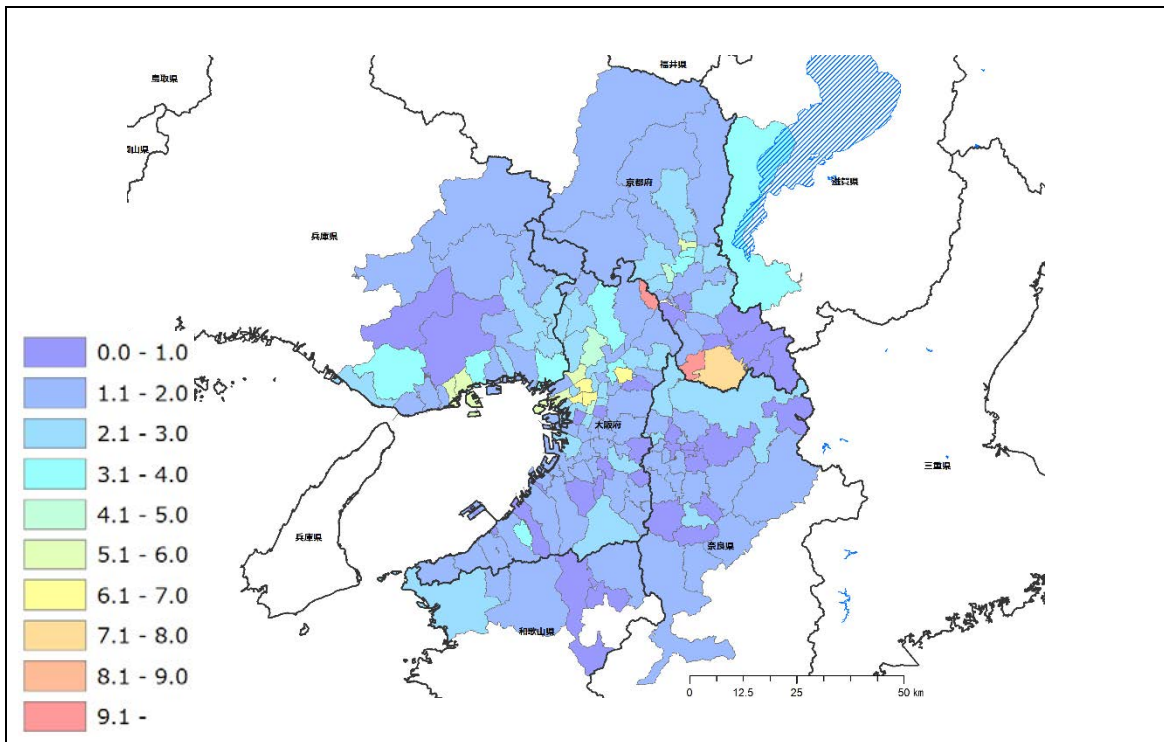


図 7-2 L 学術研究，専門・技術サービス業従業者数割合（大阪圏）

③固有値と寄与率

主成分分析を行う変数の分散が異なる場合、分散の大きな変数が重み係数の決定に大きく影響することになる。このため、データを標準化（平均0、分散1）してから主成分分析を行った。

表5は、分散共分散行列の固有値で、東京圏、名古屋圏、大阪圏の第1主成分の固有値は1.422、1.599、1.554、第2主成分の固有値は0.578、0.402、0.466で、それぞれ第1主成分の固有値が1以上となっていることから、知識産業は第1主成分で代表される。

また、固有値の寄与率をみると、第1主成分の固有値の寄与率は0.711、0.799、0.777となっており、全分散の70%以上を占めている。寄与率の基準からも知識産業の総合的指標として第1主成分が選択される。

表5 分散共分散行列の固有値、寄与率

		固有値	寄与率	累積
東京圏	第1主成分	1.422	0.711	0.711
	第2主成分	0.578	0.289	1.000
名古屋圏	第1主成分	1.599	0.799	0.799
	第2主成分	0.402	0.201	1.000
大阪圏	第1主成分	1.554	0.777	0.777
	第2主成分	0.446	0.223	1.000

④固有ベクトル

固有ベクトルは $z = a_1x_1 + a_2x_2$ の a_1, a_2 に当たる。ある主成分の固有ベクトルについて、ある重み係数が大きな値をとっているならば、その主成分の性格にはもとの変数の影響が強い。本分析のように2変数の主成分分析では、それぞれの変数の重み係数の値は等しく、主成分に与える影響は等しい。3変数以上になると、重み係数の値は多様になる。

表6 分散共分散行列の固有ベクトル

	主成分1	主成分2
G 情報通信業	0.707	0.707
L 学術研究、専門・技術サービス業	0.707	-0.707

⑤主成分得点の地図化

個々のデータに対して各変数の値を標準化して、(II.2) $z = a_1x_1 + a_2x_2$ に代入して得られた値を主成分得点という。もとの変数 x_1, x_2 と主成分 z_1, z_2 との関係は、 x_1, x_2 軸から z_1, z_2 主軸への座標変換と考えることができる。

図8は主成分1の主成分得点の統計地図で、知識産業は都心において集積が高く、千代田区、港区、川崎市幸区では主成分得点が4.0以上となっている。都心から離れるに従って集積の程度が低くなる傾向が見られる。

図8-1、図8-2は「G情報通信業」従業者数の割合、「L学術研究、専門・技術サービス業」従業者の割合を標準化した結果の統計地図である。図8-1「G情報通信業」は都心部に

分布が集中しているのに対して、図 8-2「L 学術研究，専門・技術サービス業」は郊外や山間部、沿岸部にも分布している。図 8-1「G 情報通信業」、図 8-2「L 学術研究，専門・技術サービス業」の地域分布を合成した結果が図 8 に表現されているといえる。なお、図 8、図 8-1、図 8-2 の統計地図は、階級区分を揃えて作成した。

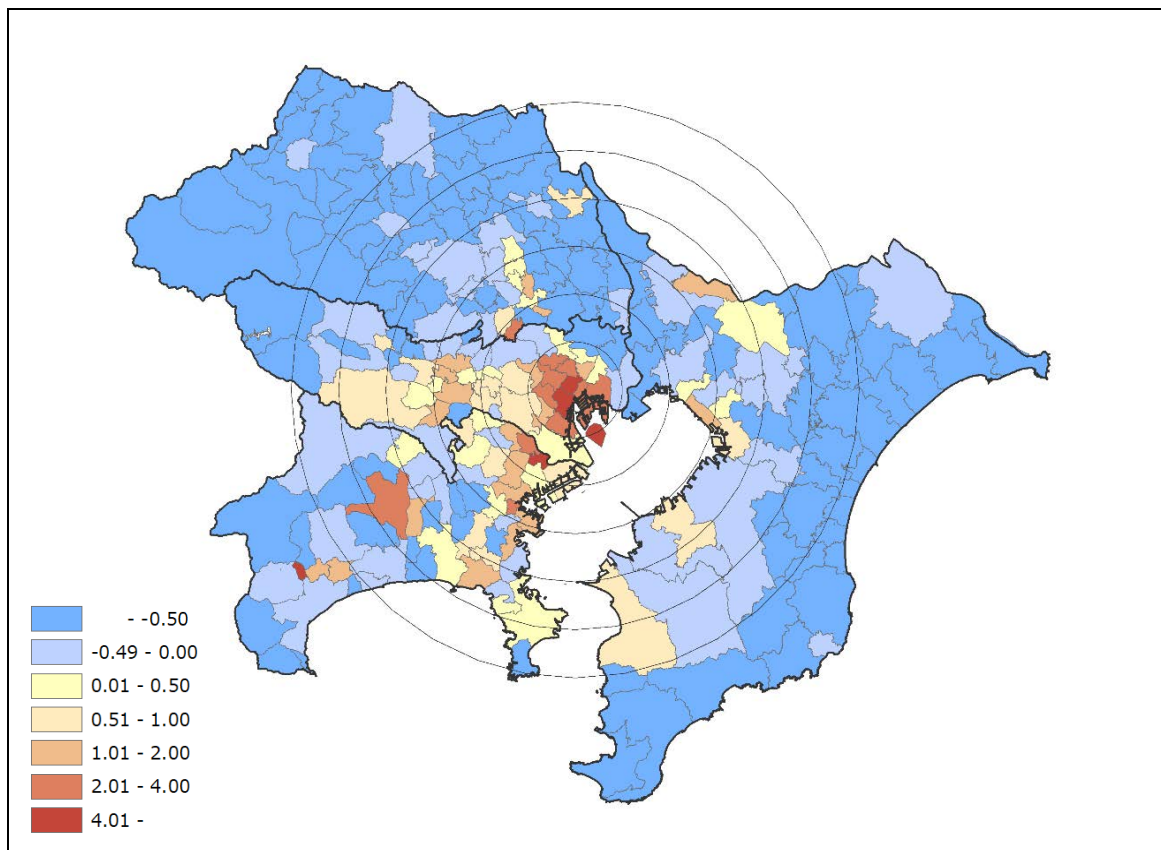


図 8 主成分得点 (東京圏)

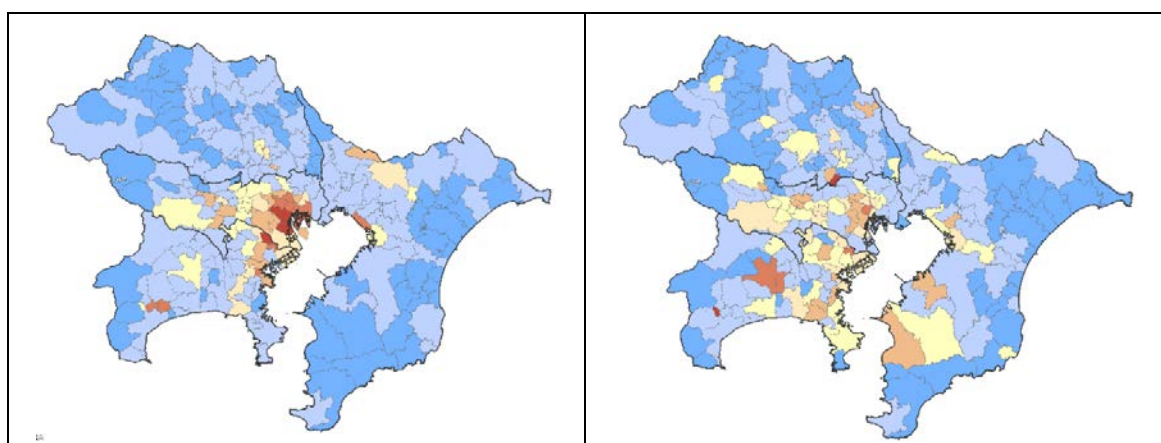


図 8-1 G 情報通信業従業員数割合 (東京圏)

図 8-2 学術研究，専門・技術サービス業従業員数割合 (東京圏)

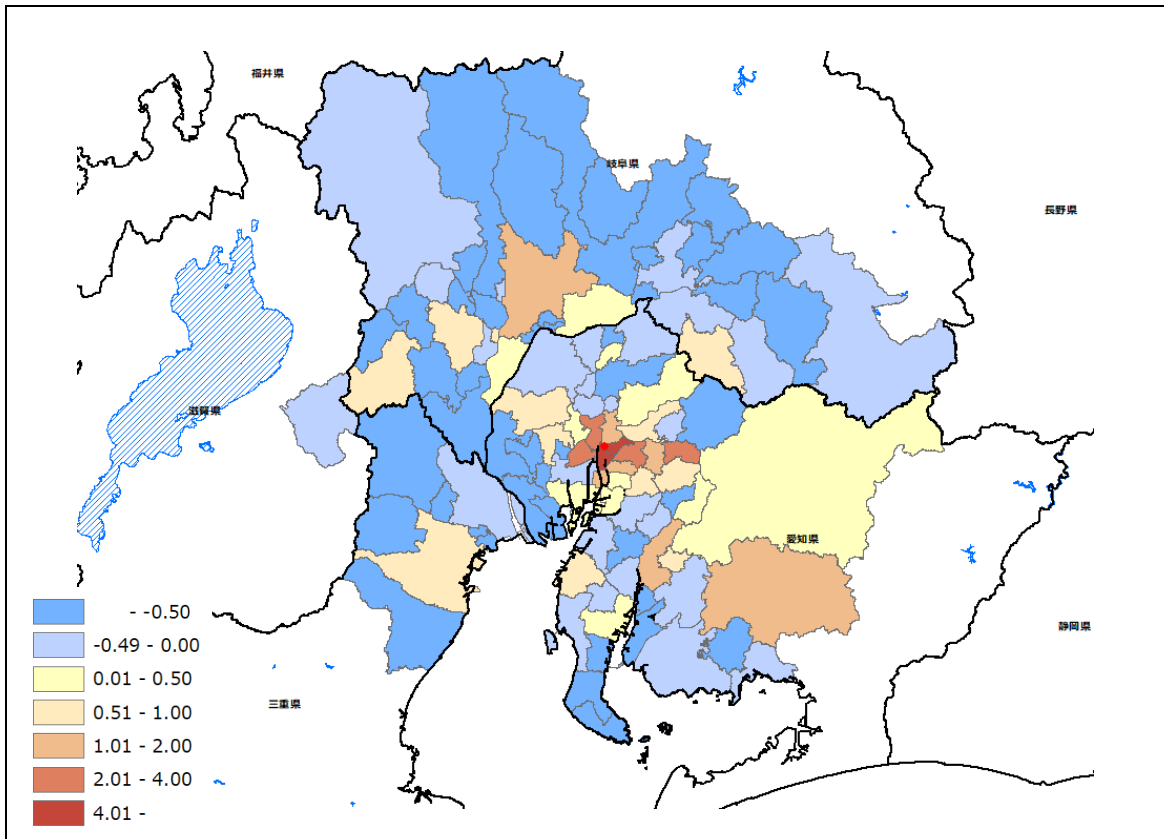


图9 主成分得点 (名古屋圏)

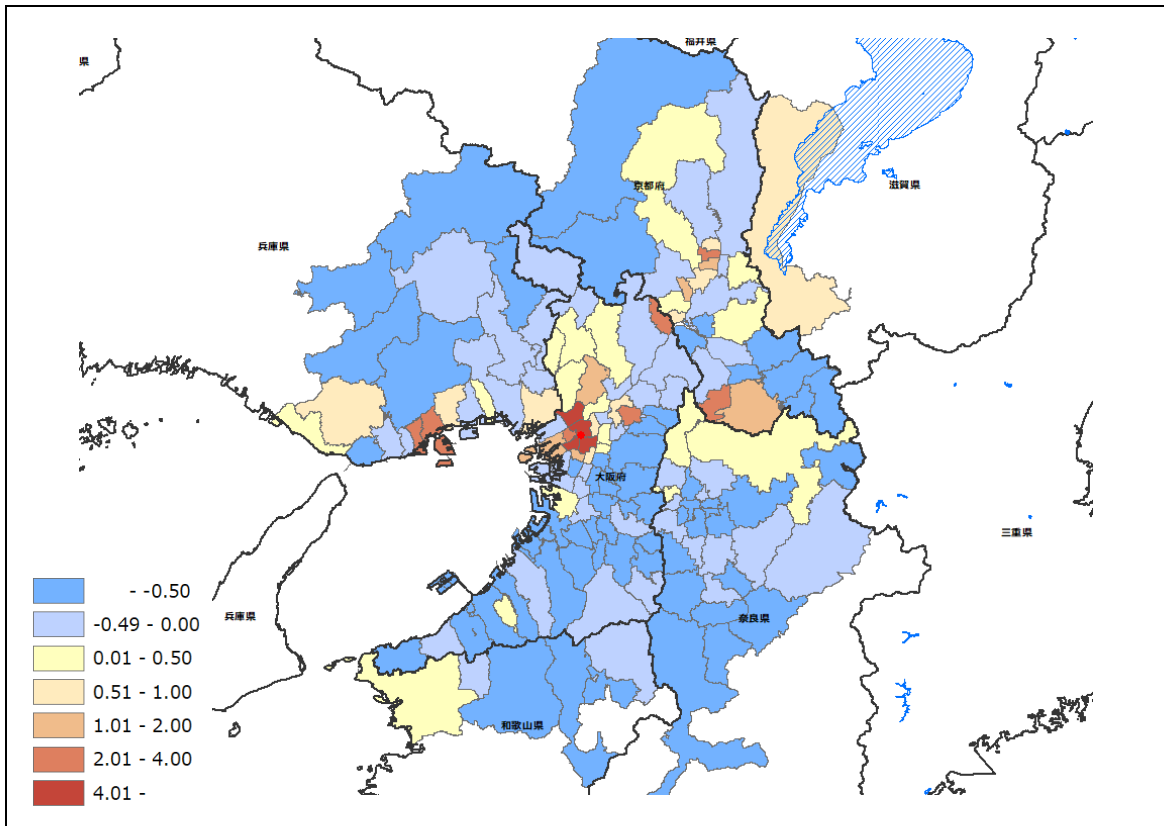


图10 主成分得点 (大阪圏)

図9、図10は名古屋圏、大阪圏の知識産業の主成分得点である。図9の名古屋圏の都心部の主成分得点は、名古屋市東区、中区が4.0以上と高く、名古屋市中村区、西区、千種区、長久手市も2.0以上と高い。図10の大阪圏では、大阪市淀川区、大阪市北区、大阪市西区、大阪市中区で主成分得点が4.0以上と高い。また、大阪市福島区、門真市、島本町（企業の研究所が立地）、神戸市中央区、神戸市兵庫区、京都市中京区、精華町（関西文化学術研究都市が立地）も2.0以上と高い。

2.1.5 地域メッシュ統計を利用した主成分分析

地域メッシュ統計を移動窓集計した結果を利用して主成分分析を行った。移動窓集計は地域範囲を重複して集計しているため、主成分分析の適用においては「ランダム抽出点」を用いてデータを抽出した。移動窓集計、ランダム抽出点について以下に説明する⁵。

(1) 移動窓集計

移動窓は分析地域全域の一部を覆う領域を表す。この領域（あるいは、「窓」）はある位置から他の位置に移動し、窓には円形か四角形が使われることが多い。多くのアプリケーションソフトウェアでは、窓は分析範囲全体を規則正しいステップで動き、各場所で処理（例えば窓内にある事象の平均の計算など）を実行する(Lloyd (2010))。

移動窓による集計の計算においては、下準備として地域メッシュ統計の表章地域単位である地域メッシュ区画単位で集計を行うのではなく、各地域メッシュ区画の図形重心点（以下「重心点」という）をGISで作成し、重心点に統計値を持たせた。統計値を属性値として持つ重心点を中心とする半径4,000mの円の範囲（窓）を分析地域単位として設定し、その範囲内にある重心点の統計データを集計した。図9に示すように分析地域単位（移動窓）を設け、さらにその範囲を固定させるのではなく、移動させ重ねながら統計データを集計している。この移動窓集計により場所を固定させずに重ねながら集計することによって、ゾーニングによる影響に捕らわれない分析結果を得られる。

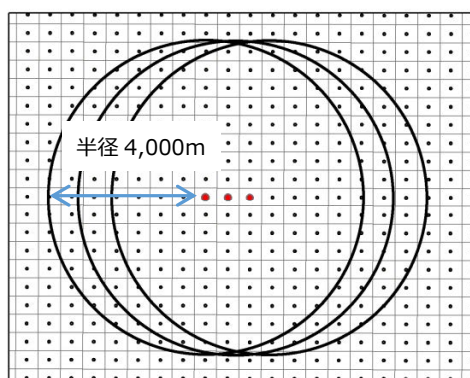


図9 移動窓のイメージ

⁵ 移動窓集計については、「I 分析の方法」により具体的に説明しているので参照されたい。

移動窓集計は、移動窓内の従業者数が空間的に一様に分布していると捉えて便宜的に移動窓の中心点に集計結果を集約している。しかし、実際には地域内で従業者数が一様ではない場合も多く、例えば集計範囲の外縁部にも従業者数が多いメッシュ区画が偏在し、中心点付近のメッシュ区画には従業者数が少ない場合がある。このことから、移動窓内の従業者数が空間的に一様に分布しているという前提は、中心点付近に従業者数が少ないにもかかわらず、中心点付近の従業者数として表現されるため、地域分析の手法としては、地理的な位置の精度の観点から問題がある。本研究ではこの問題点の改善策として、集計する従業者数に中心点からの距離による重みを付けたものについて集計した。

具体的には、移動窓の中心の図形重心点と移動窓内にあるその他の図形重心点との距離を測り、距離に応じて距離減衰関数（図 10 の実線（ 1σ ））による重み係数をかけて、中心点から遠い重心点の統計値を減じて集計している。

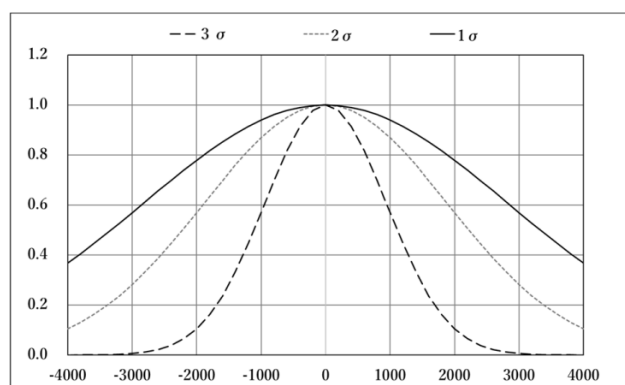


図 10 距離減衰関数

(2) ランダム抽出点

移動窓法による集計を行うと、図 9 のように隣接する分析地域単位ではほとんどの重心点が重複する。重複する範囲が大きいほど隣接する分析地域単位の集計値もほぼ等しくなるため、全ての重心点を用いて相関係数を計算すると相関係数の値は当然高くなる。主成分分析は、項目間の相関構造を考慮して、低い次元の合成変数（主成分）に変換し、データが有している情報をより解釈しやすくするための方法であるため、本研究ではこの影響を取り除いて東京圏における任意の分析地域単位別の主成分分析を行う。このため、分析地域単位ごとに空間的に互いに独立でランダムな点を作成した（以下、「ランダム抽出点」という）。

地域データは、集計する位置によってデータ値の傾向が異なるため、多くのパターンについて分析して東京圏における主成分分析を行うこととし、ランダム抽出点は 100 組作成した。ランダム抽出点の作成に当たっては、以下の①～③の処理を 15,000 回繰り返して行った。

- ① GIS の乱数発生関数により任意の重心点を検索し、ランダム抽出点であることを示

す符号を付与する。

- ② ランダム抽出点を中心として半径4,000mのバッファを作成する。
- ③ バッファに重なる重心点にランダム抽出点のバッファ内であることを示す符号を付与する。

2回目以降も①～③の処理を行うが、新しいバッファ内に既存の符号が存在しない場合にのみ新たに符号を付与する。図11-1、11-2、11-3にランダム抽出点の例を示す。

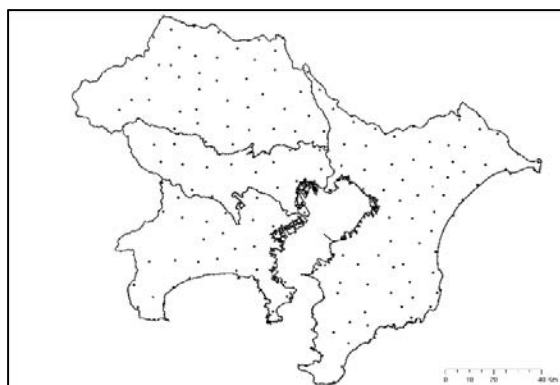


図 11-1 東京圏ランダム抽出点（半径4,000m）の例

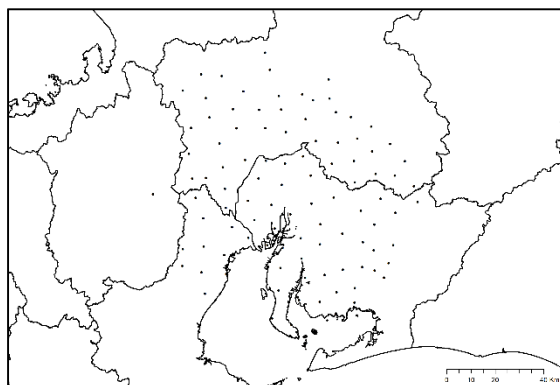


図 11-2 名古屋圏ランダム抽出点（半径4,000m）の例

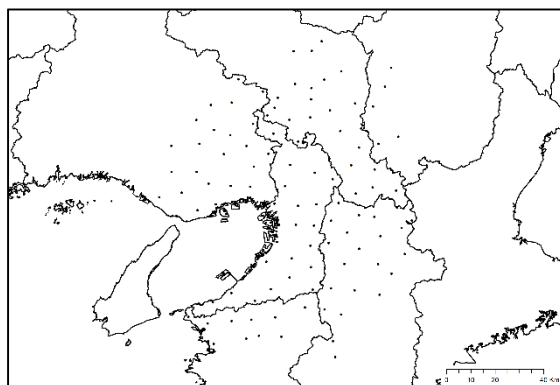


図 11-3 大阪圏ランダム抽出点（半径4,000m）の例

(3) 主成分分析

平成 28 年経済センサス - 活動調査の地域メッシュ統計 (2 分の 1 地域メッシュ) の移動窓集計結果を利用して、「G 情報通信業」、「L 学術研究, 専門・技術サービス業」についてランダム抽出点を利用して空間的に独立な分析地域単位を抽出して主成分分析を行った。ランダム抽出点が 100 組あるので、主成分分析を 100 回行う。

①基本統計量

表 7-1、7-2、7-3 に東京圏、名古屋圏、大阪圏における「G 情報通信業」、「L 学術研究, 専門・技術サービス業」従業者数の割合の平均と分散の基本統計量を示す。市区町村別集計の結果と比較して、両項目とも 100 組の平均は小さくなっている。とくに地域的に偏在して分布している「G 情報通信業」の従業者数の割合は、集計する地点数が多くなると従業者数が少ない地点も多くなるため、平均が小さくなると考えられる。

表 7-1 平均、分散の基本統計量 (東京圏)

	G 情報通信業		L 学術研究, 専門・技術サービス業	
	平均	分散	平均	分散
平均	0.97	1.36	2.21	0.65
最小	0.77	1.17	1.96	0.48
最大	1.16	1.54	2.54	0.85
データの個数	100	100	100	100

表 7-2 平均、分散の基本統計量 (名古屋圏)

	G 情報通信業		L 学術研究, 専門・技術サービス業	
	平均	分散	平均	分散
平均	0.33	1.26	1.70	0.76
最小	0.21	1.06	1.46	0.53
最大	0.44	1.49	2.06	0.96
データの個数	100	100	100	100

表 7-3 平均、分散の基本統計量 (大阪圏)

	G 情報通信業		L 学術研究, 専門・技術サービス業	
	平均	分散	平均	分散
平均	0.41	1.59	1.49	0.43
最小	0.30	1.45	1.31	0.30
最大	0.49	1.72	1.68	0.57
データの個数	100	100	100	100

②固有値と寄与率

次に地域メッシュ統計の移動窓集計結果においても標準化 (平均 0、分散 1) して主成分分析を行い、ランダム抽出点により分散共分散行列の固有値を第 1 主成分、第 2 主成分そ

それぞれ 100 組算出した。表 8-1、8-2、8-3 は東京圏、名古屋圏、大阪圏における固有値の基本統計量であり、ランダム抽出点による抽出の組み合わせによって固有値にばらつきがあることがわかる。

表 8-1 の東京圏における第 1 主成分の分散共分散行列の固有値の平均は 1.364 でその範囲は 1.167～1.536 となっている。100 組全てのランダム抽出点において 1 を超えており、第 1 主成分は「G 情報通信業」、「L 学術研究，専門・技術サービス業」の総合的指標として選択できる。同様に表 8-2 の名古屋圏における第 1 主成分の分散共分散行列の固有値の平均は 1.258 でその範囲は 1.058～1.491、表 8-3 の大阪圏における第 1 主成分の分散共分散行列の固有値の平均は 1.588 でその範囲は 1.449～1.716 となり、それぞれ 100 組のランダム抽出点全てにおいて 1 を超えており、第 1 主成分は「G 情報通信業」、「L 学術研究，専門・技術サービス業」の総合的指標として選択できる。

表 8-1 分散共分散行列の固有値の基本統計量（東京圏）

	第 1 主成分	第 2 主成分
平均	1.364	0.649
最小	1.167	0.477
最大	1.536	0.846
データの個数	100	100

表 8-2 分散共分散行列の固有値の基本統計量（名古屋圏）

	第 1 主成分	第 2 主成分
平均	1.258	0.762
最小	1.058	0.529
最大	1.491	0.961
データの個数	100	100

表 8-3 分散共分散行列の固有値の基本統計量（大阪圏）

	第 1 主成分	第 2 主成分
平均	1.588	0.429
最小	1.449	0.302
最大	1.716	0.569
データの個数	100	100

東京圏における分散共分散行列の固有値の寄与率についても 100 組の基本統計量をみると（表 9-1）、第 1 主成分の寄与率の平均は 67.8% となっており、70% に満たない。最小は 58.0%、最大は 76.3% でありばらつきがある。第 1 主成分の固有値の寄与率のヒストグラムをみると（図 12-1）、やや双峰性の分布を示しているが、0.70～0.73 に最頻値がある。本分析では寄与率が 70% 以上となっていれば、全体を代表する指標として選択できると考えるため、寄与率の基準からも第 1 主成分は「G 情報通信業」、「L 学術研究，専門・技術サービス業」を代表する総合的指標であるといえよう。

表 9-1 分散共分散行列の寄与率の基本統計量（東京圏）

	第1主成分	第2主成分
平均	0.678	0.322
最小	0.580	0.237
最大	0.763	0.420
データの個数	100	100

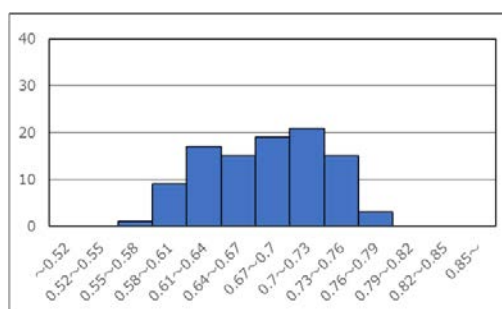


図 12-1 第1主成分の固有値の寄与率のヒストグラム（東京圏）

同様に名古屋圏における分散共分散行列の固有値の寄与率についても 100 組の基本統計量をみると（表 9-2）、第1主成分の寄与率の平均は 62.3%となっており、70%に満たない。図 12-2 の第1主成分の固有値の寄与率のヒストグラムをみると、0.58~0.61 に最頻値がある。本分析では寄与率が70%以上となっていれば、全体を代表する指標として選択できていることから、ランダム抽出点の位置によっては名古屋圏の第1主成分は「G情報通信業」、「L 学術研究、専門・技術サービス業」を代表する総合的指標であるとはいえない場合が多い。

表 9-2 分散共分散行列の寄与率の基本統計量（名古屋圏）

	第1主成分	第2主成分
平均	0.623	0.377
最小	0.524	0.262
最大	0.738	0.476
データの個数	100	100

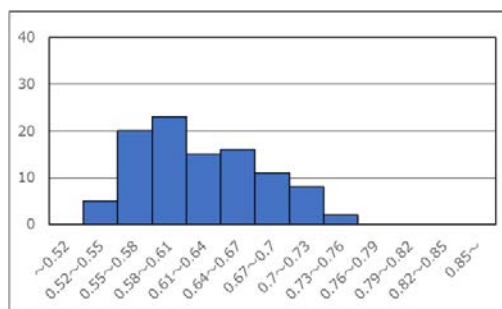


図 12-2 第1主成分の固有値の寄与率のヒストグラム（名古屋圏）

同様に大阪圏における分散共分散行列の固有値の寄与率についても 100 組の基本統計量をみると（表 9-3）、第 1 主成分の寄与率の平均は 78.7%となっており、70%以上となっている。また最小値が 0.718 となっていることから全てのランダム抽出点による第 1 主成分の固有値の寄与率は 70%を超えている。ヒストグラムをみると（図 12-3）、0.79~0.82 に最頻値がある。本分析では寄与率が 70%以上となっていれば、全体を代表する指標として選択できると考えるため、寄与率の基準からは第 1 主成分は「G情報通信業」、「L 学術研究、専門・技術サービス業」を代表する総合的指標であるといえる。

表 9-3 分散共分散行列の寄与率の基本統計量（大阪圏）

	第 1 主成分	第 2 主成分
平均	0.787	0.213
最小	0.718	0.150
最大	0.850	0.282
データの個数	100	100

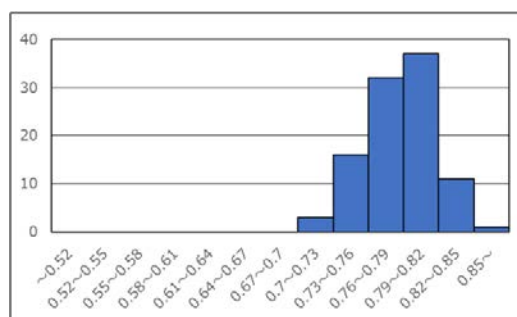


図 12-3 第 1 主成分の固有値の寄与率のヒストグラム（大阪圏）

市区町村別集計結果を利用した主成分分析では、東京圏、名古屋圏、大阪圏の寄与率は、71.1%、79.9%、77.7%となり、主成分 1 は知識産業の指標として採用できるとしたが、地域メッシュ統計を利用すると異なる結果となり、名古屋圏では地域メッシュ統計を集計した分析結果では寄与率 70%未滿となる場合が多かった。名古屋圏における市区町村数は 3 圏で最も少なく市区町村当たりの面積が大きいことから、市区町村別集計によって集積が成立しているとみなされ、主成分 1 の寄与率が高くなったと考えられる。

③主成分得点の地図化

ア) 東京圏

主成分得点についても各ランダム抽出点について 100 組算出できる。東京圏についてこの 100 組の主成分得点を標準化した後全て統合し重複するメッシュを削除すると 7,992 レコードとなる。このデータを利用して、主成分得点を地図化した（図 13-1）。特別区部、川崎市、横浜市において主成分得点が高くなっており、知識産業が集積している。郊外では鉄道路線沿いに主成分得点が高い地域が偏在しており、特に複数の鉄道路線が乗り入れる

地域において高い傾向がある。

図 14-1 に主成分得点が 2 以上と高い地域を拡大して示す。特別区の都心部を中心に 10 km 圏内に広く知識産業が集積している。10 km 圏内でも北東部の墨田区や荒川区は集積度が相対的に低い。その南側に川崎市高津区、川崎市中原区、川崎市幸区に知識産業の集積が分布している。さらに南下して沿岸部の横浜市神奈川区、横浜市保土ヶ谷区、横浜市西区、横浜市中区においても集積しており、東京都心から横浜において知識産業の集積が強い。さらに南下して鎌倉市にも知識産業が分布している。そのほか神奈川県の内陸では厚木市、伊勢原市において集積度が高い。東京都においては、特別区の集積と比較すると弱い。国分寺駅、府中本町駅、北八王子駅にも集積が高いメッシュが現れており、南多摩にも知識産業の集積が存在する。千葉県では千葉市美浜区にも知識産業の集積がある。

国分寺駅付近や府中本町駅付近など市区町村別の結果では現れない地域でもメッシュ別にみると知識産業の集積が現れることを確認でき、半径 4,000m の集計地域単位の面積は、東京圏の市区町村の面積の平均と同程度の水準であるが、移動窓集計を行うことによりゾーンニングに影響されない結果を把握できることが明らかになった。

イ) 名古屋圏

東京圏と同様に名古屋圏について主成分得点を地図化した (図 13-2)。名古屋市の都心部を中心に知識産業が分布している。郊外では鉄道路線が乗り入れる駅付近において集積している。

図 14-2 に主成分得点が 2 以上と高い地域を拡大して示す。名古屋市の集積は東京圏よりも狭い範囲となっている。その他、愛知県では刈谷市、知立市、豊田市に知識産業の集積が分布している。岐阜県では岐阜市、大垣市、瑞穂市に集積が分布している。

図 9 の市区町村別の結果と比較すると、地域メッシュ統計を利用した結果では岐阜県大垣市や愛知県豊田市において集積が現れている。市区町村合併により市域が山間地を含んで広がっている地域においては、地域メッシュ統計を用いて集積を地図化することにより、より実態に即した特徴を把握することができる。

ウ) 大阪圏

東京圏、名古屋圏と同様に大阪圏について主成分得点を地図化した (図 13-3)。大阪市、神戸市、京都市、大津市などの集積が顕著となっている。

図 14-3 に主成分得点が 2 以上と高い地域を拡大して示す。大阪圏の都心部は名古屋圏の都心部よりも集積の範囲が広いが東京圏の都心部よりも狭い集積となっている。また、大阪市の北に位置する豊中市、吹田市、箕面市にも集積が広がっている。大阪圏の集積は、東京圏のように特別区部、川崎市、横浜市のような連担した集積ではなく、地形の影響が大きいいためか、神戸市、京都市、精華町など地域的に偏在した集積となっている。

図 10 の市区町村別の結果では、大津市や京都市右京区のような広い市域の場合には知識産業従業者数の割合が平滑化されて主成分得点が小さくなってしまいが、地域メッシュ統計を利用して分析を行うと、市内の特定の地域の主成分得点を表現することができる。

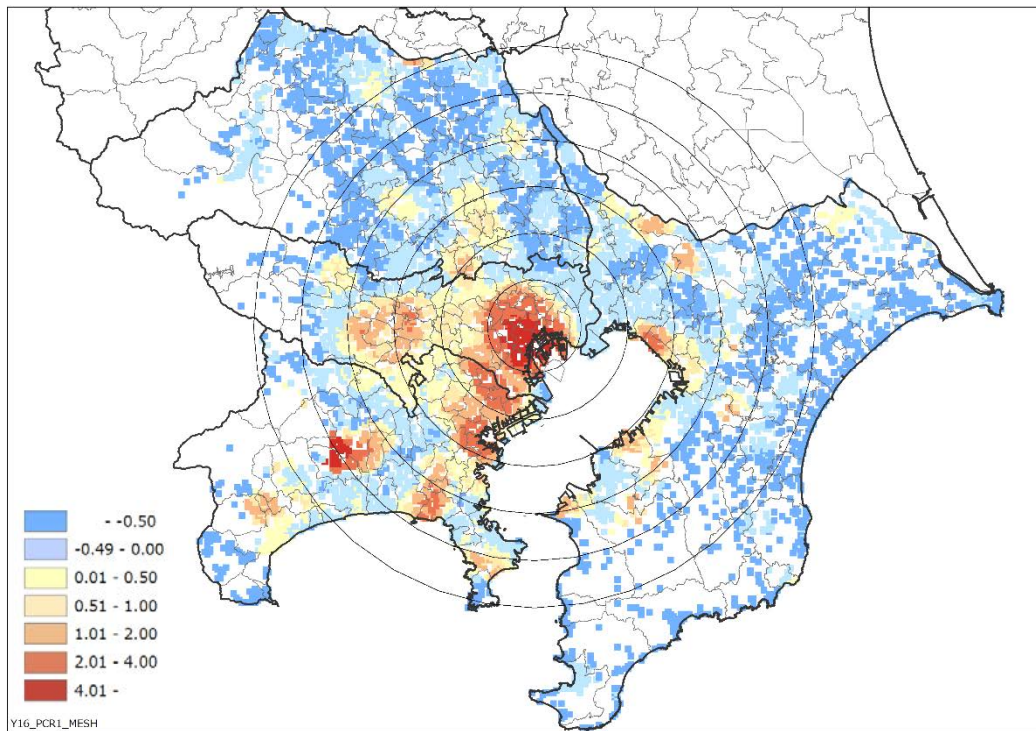


図 13-1 東京圏におけるメッシュ別主成分得点の地域分布 (2016 年)

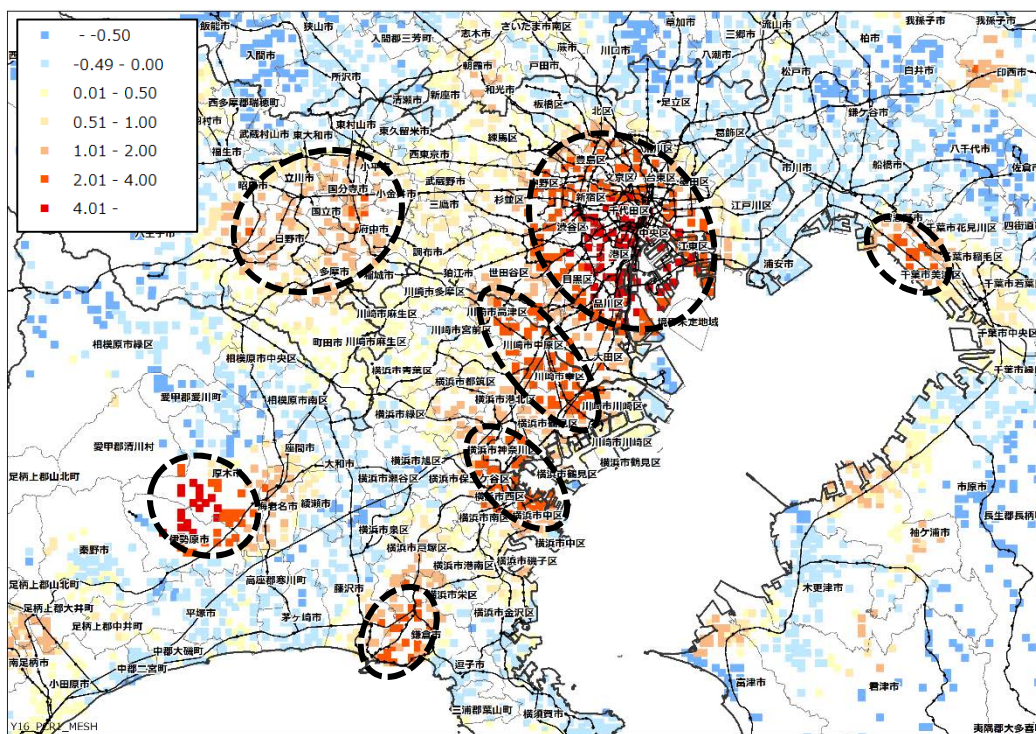


図 14-1 知識産業の主成分得点が 2 以上の地域 (東京圏)

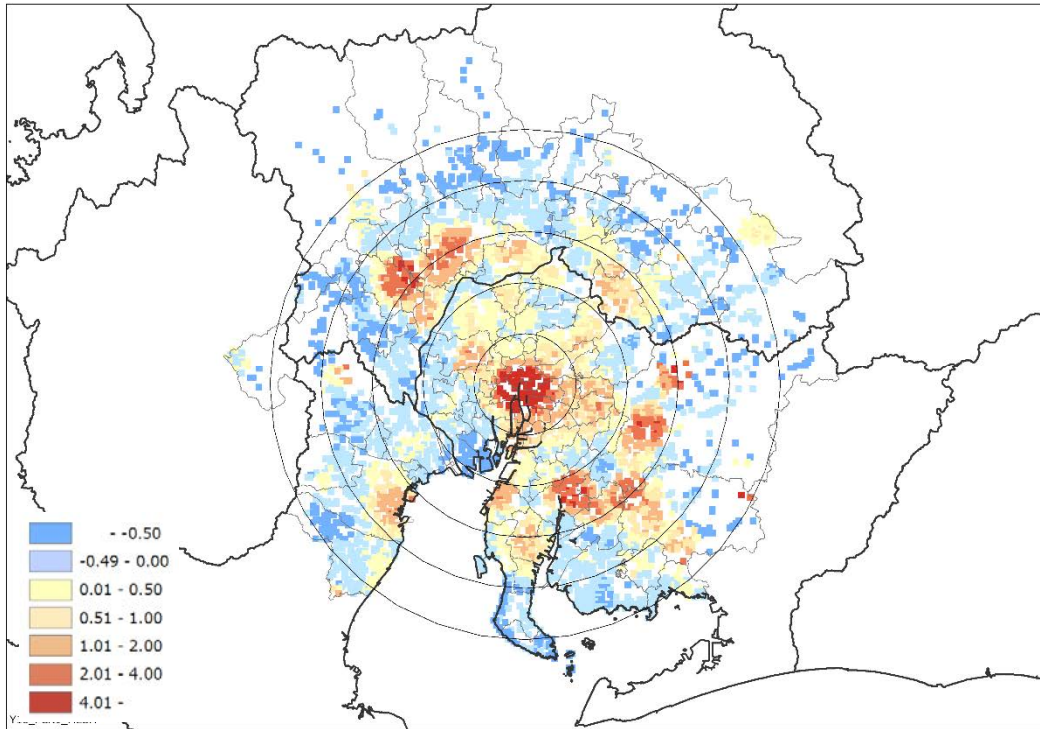


図 13-2 名古屋圏におけるメッシュ別主成分得点の地域分布（2016年）

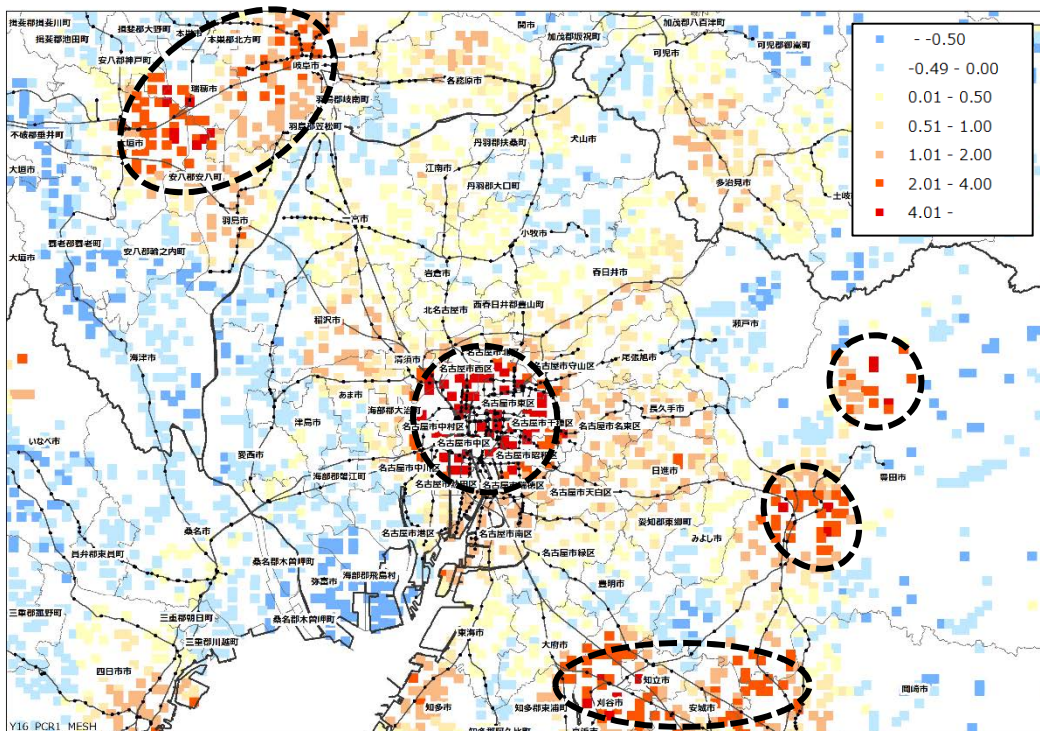


図 14-2 知識産業の主成分得点が2以上の地域（名古屋圏）

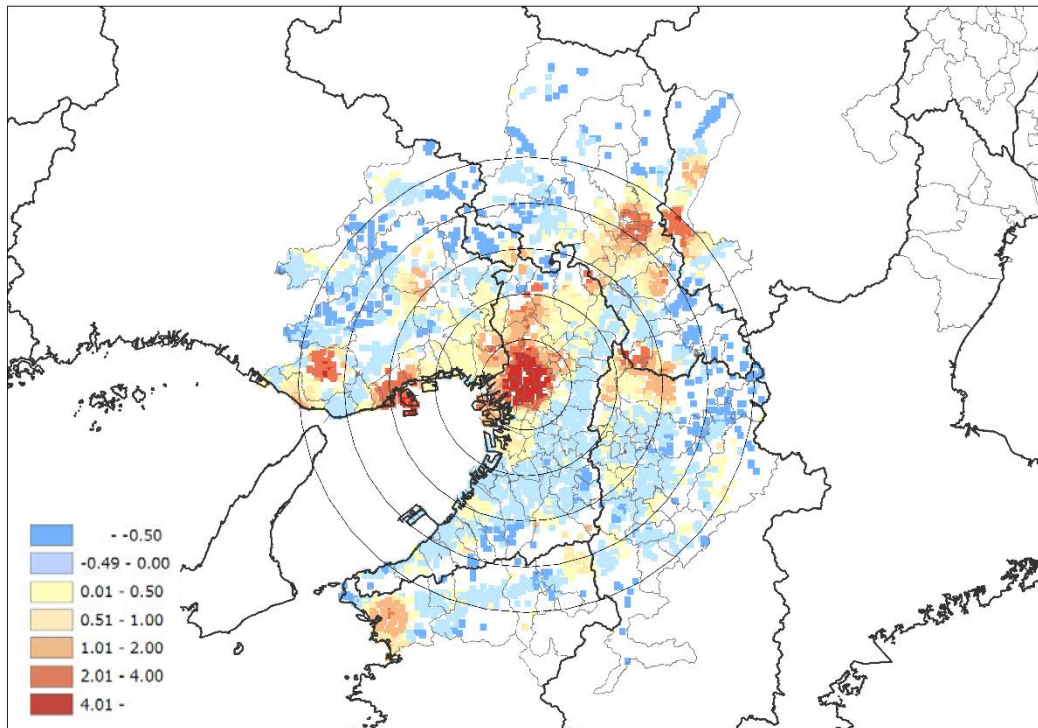


図 13-3 大阪圏におけるメッシュ別主成分得点の地域分布 (2016 年)

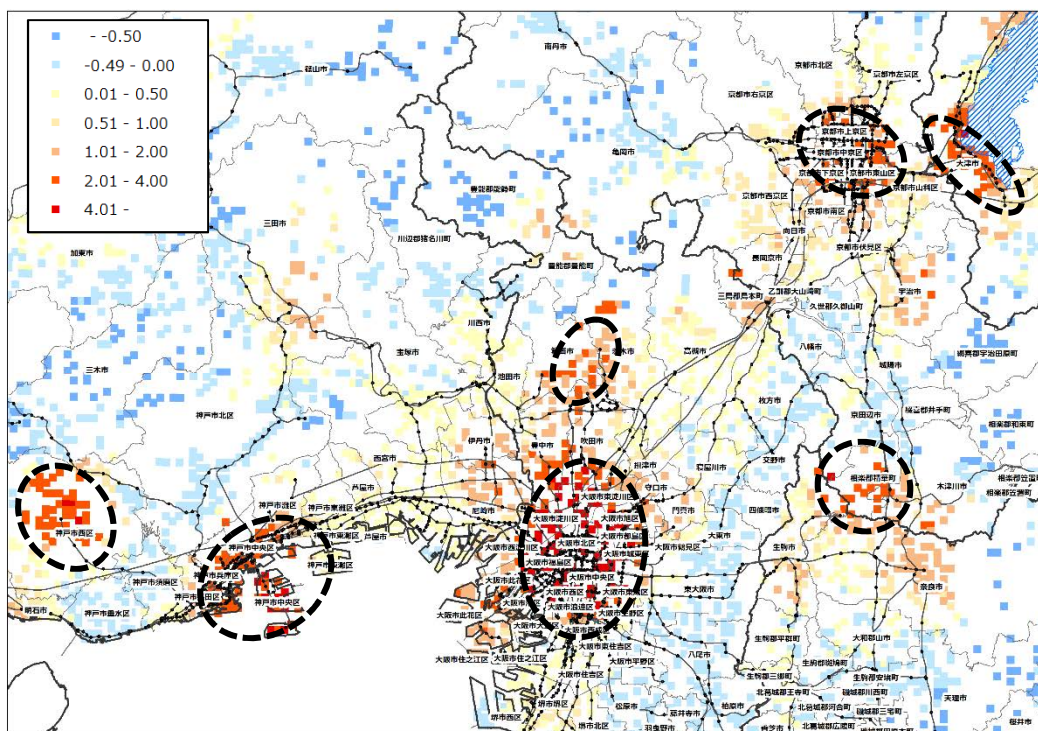


図 14-3 知識産業の主成分得点が 2 以上の地域 (大阪圏)

2.2 2012～16年の動向

東京圏について、平成24年経済センサス - 活動調査の地域メッシュ統計の移動窓集計結果を利用して前項と同様の分析を行い、平成28年の結果と比較する。

図15は平成24(2012)年の第1主成分の主成分得点を地図化したものである。図13-1の平成28年の第1主成分の主成分得点を地図化したものと比較すると、全体的な分布傾向は同じであるが、平成28年の地域分布は、多摩や川崎市北部において主成分得点が高い地域が広がっている。

図16は平成28年の第1主成分の主成分得点から平成24年の第1主成分の主成分得点の差を算出し⁶、地図化したものである。主成分得点の増加が明らかな地域に駅名の吹き出しを示す。

主成分得点が増加し、知識産業の集積が進んでいる地域を駅名で列挙すると、世田谷区の子玉川駅付近、横浜市の中山駅付近、鎌倉市の大船駅付近、海老名市の海老名駅付近、八王子市の八王子駅付近、国分寺市の国分寺駅付近、川越市の霞ヶ関駅付近、三郷市の三郷駅付近、千葉市の京成稲毛駅付近、蘇我駅付近などがある。

一方主成分得点が減少している地域もあり、印西市の千葉ニュータウン中央駅付近や伊勢原市の伊勢原駅付近では知識産業の集積が相対的に低下傾向にあると考えられる。そのほか、武蔵野市や三鷹市、稲毛市、川崎市麻生区、横浜市青葉区など特別区、多摩、横浜の中間地帯では主成分得点が減少している地域もあり興味深い。

知識産業の集積は都心から10km帯内の集積が顕著であるが、郊外でも複数路線が乗り入れている地域において集積が高まる傾向を示している。

⁶ 平成28年、平成24年の主成分得点はそれぞれ標準化(平均0、分散1)してから差を計算している。

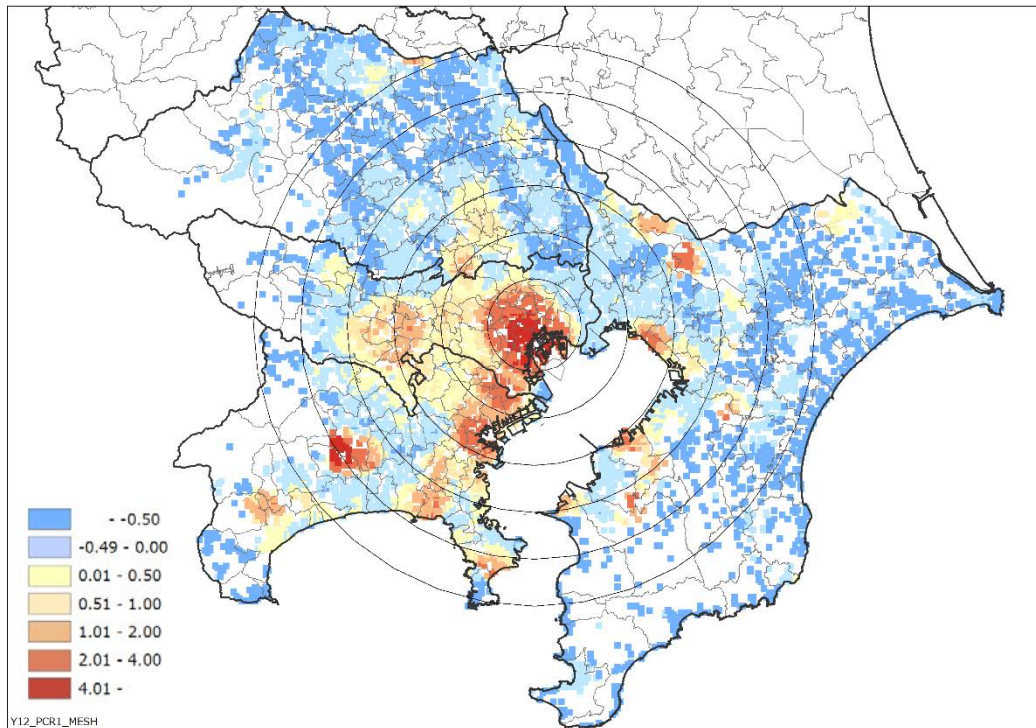


図 15 東京圏におけるメッシュ別主成分得点の地域分布 (2012 年)

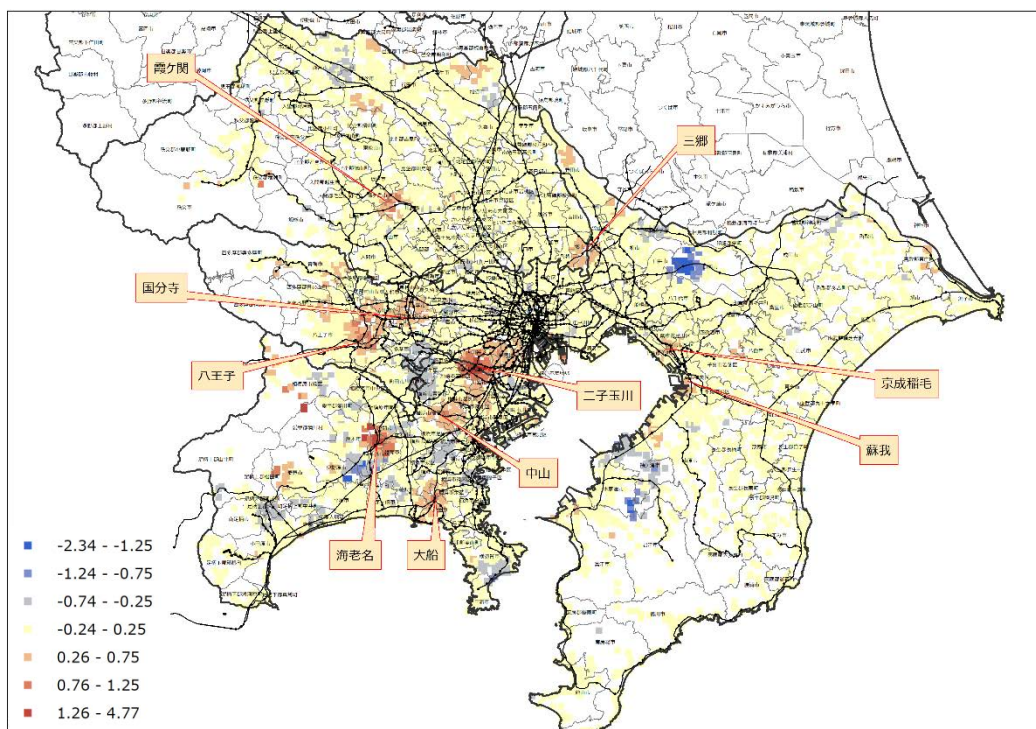


図 16 2012~16 年 第 1 主成分の主成分得点の増減の地域分布 (東京圏)

3 まとめと今後の課題

3.1 まとめ

本研究は、平成 24 年、平成 28 年の経済センサス - 活動調査に関する地域メッシュ統計の結果のうち、「G情報通信業」、「L 学術研究, 専門・技術サービス業」に着目し、主成分分析により知識産業の総合的指標を算出し、東京圏、名古屋圏、大阪圏を対象に地域分析を行ったものである。具体的には、半径 4,000m の円内に重心が含まれる 2 分の 1 地域メッシュの統計値を距離減衰関数による重みをかけて集計後、主成分分析を行った。知識産業の総合的指標として第 1 主成分を選択し、主成分得点を地図化して地域分布について把握した。

東京圏においては、都心から半径 10 km 帯に知識産業が集積しているほか、神奈川県横浜市、川崎市でも集積度が高いことを明らかにした。地域メッシュ統計データを利用することにより、郊外では国分寺駅など複数路線が乗り入れている駅の付近で知識産業が集積していることが確認でき、利便性が知識産業の集積に影響を与えていると考えられる。

名古屋圏においては、名古屋市の中心部に知識産業が集積しているほか、愛知県刈谷市、知立市、豊田市、岐阜県岐阜市、大垣市、瑞穂市にも集積している。地域メッシュ統計を集計して主成分分析を行うと、寄与率が 70% 未満となる場合が多く、名古屋圏では「G情報通信業」、「L 学術研究, 専門・技術サービス業」を合わせた知識産業の集積が東京圏、大阪圏と比較して弱いと考えることができる。

大阪圏では大阪市の中心部における集積のほか、兵庫県神戸市、京都府京都市、精華町、滋賀県大津市での集積がみられる。東京圏、名古屋圏とは異なり、大阪圏では 100 組全てのランダム抽出点において寄与率が 70% を超えており、「G情報通信業」、「L 学術研究, 専門・技術サービス業」を合わせた知識産業が地域的に強く集積していると考えられる。

併せて、市区町村別集計のデータを利用して同様の分析を行い、市区町村別集計では主成分得点が低く知識産業の集積が認められない場合でも、地域メッシュ統計を利用すると駅付近では主成分得点が高いなど集積が認められる場合があることを示し、主成分分析においても集計地域単位によって分析結果に違いがあり、ゾーニング問題が現れることを示した。

3.2 今後の課題

また、本稿では「G情報通信業」、「K 学術研究, 専門・技術サービス業」に着目し、一つの総合的指標を算出し分析を行ったが、他の産業大分類の項目も利用して多変量の主成分分析を行い、地域別の産業構造の動向について把握する必要がある。

また、平成 24 年経済センサス - 活動調査の結果を利用した動向の分析は東京圏のみについて行ったが、名古屋圏、東京圏についても同様の分析を行うことも今後の課題としたい。

<参考文献>

- ウィリアム・ペティ著、大内兵衛、松川七郎訳、『政治算術』、岩波文庫、1955年
- コーリン・クラーク著、大川一司、小原敬士、高橋長太郎、山田雄三約編、『経済進歩の諸条件』、勁草書房、1955年
- 田中豊、脇本和昌、『多変量統計解析法』、現代数学社、1983年
- 中谷友樹、「空間分析におけるスケール」、『地理情報科学 GIS スタンダード』、浅見泰司、矢野佳司、貞広幸雄、湯田ミノリ編、古今書院、2015年
- 永田靖、棟近雅彦共著、『多変量解析方入門』、サイエンス社、2001年
- フリッツ・マハループ著、高橋達男、木田宏監訳、『知識産業』、産業能率短期大学出版部、1969年
- 三橋規宏、内田茂男、池田吉紀、『新・日本経済入門』、日本経済新聞出版社、2015年
- Christopher D. Lloyd. *Spatial Data Analysis: an Introduction for GIS Users*, Oxford University Press, 2010.
- Stan Openshaw. The Modifiable Areal Unit Problem, 38, GeoBooks. 1984.