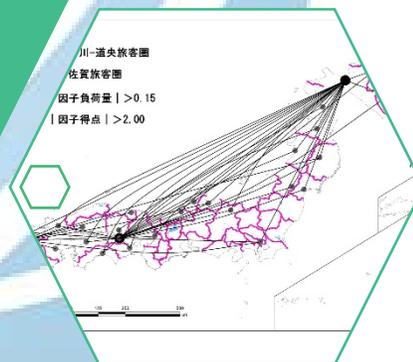
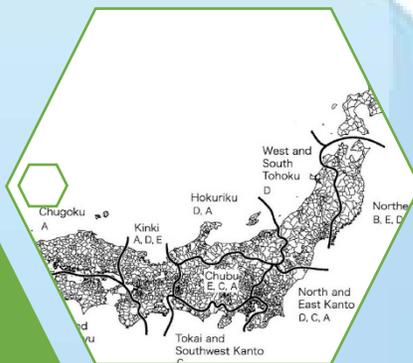
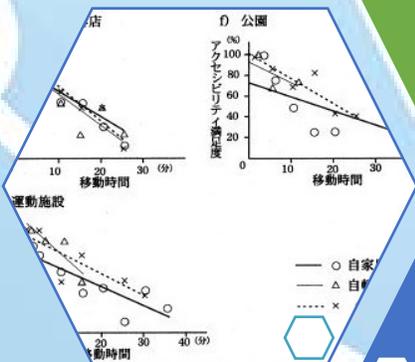


多変量解析 地域分析への応用

回帰
分析

因子
分析

クラスタ
分析



担当教員 村山 祐司

TA LIU Kai

筑波大学
University of Tsukuba



● 地域分析とは

地域分析[Regional Analysis]は、理論・定量的な手法やモデルを駆使して地域の諸特徴を探るとともに、地域的なパターンやその構造、秩序、法則、そして一般化を追究することを課題としている。さらに地域の形成・変容メカニズムを明らかにし、当該地域における将来予測や計画・政策に役立てることを目指している。

● 現在までの地域分析

最近では、地域に関する多種多様な情報がデジタル化されるとともに、これらのデジタルデータを処理するパーソナル・コンピュータの性能が飛躍的に向上し、だれでも手軽に地域分析が行える環境が整いつつある。さらに、地域分析を支援する地理情報システム(GIS)や統計ソフトウェアが安価であるいは無償で入手できるようになっている。ビッグデータ時代の到来とともに、地域分析はエリアマーケティングをはじめビジネスでも注目を集め、データマイニングを中心として新しい地域分析も芽生えつつある。

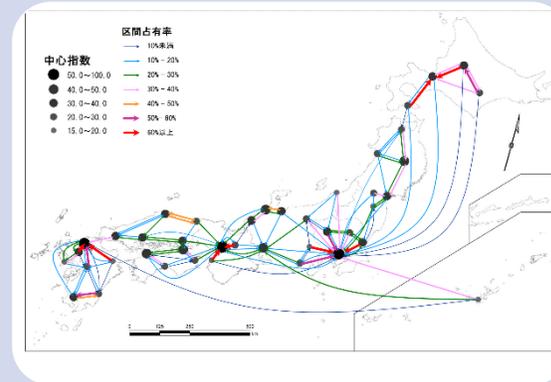
● 多変量解析とは

多変量解析は、複雑に関連しあう多変量データを統計的に解析することによって、データの構造をより低い次元に縮減したり、データの背後に潜む潜在的な構造を見出したりするための数学的手法である。

● 要旨

地域分析において多変量解析で最も利用頻度が高い三大手法(回帰分析、因子分析、クラスター分析)を紹介し、事例をまじえながら各手法の有効性について概説する。

データソース



ID	Modelo	Estadístico	Valor	Signif.	Modelo	Estadístico	Valor	Signif.
1	REGRESION LINEAL	R	.928	.000	Modelo lineal general	F	10.0	.000
2	REGRESION LINEAL	R	.928	.000	Modelos lineales generalizados	F	10.0	.000
3	REGRESION LINEAL	R	.928	.000	Modelos mixtos	F	10.0	.000
4	REGRESION LINEAL	R	.928	.000	Correlaciones	R	.928	.000
5	REGRESION LINEAL	R	.928	.000	Loglineal	Chi-Square	3.38	.064
6	REGRESION LINEAL	R	.928	.000	Regresion	F	4.46	.033
7	REGRESION LINEAL	R	.928	.000	Clasificar	Chi-Square	4.03	.043
8	REGRESION LINEAL	R	.928	.000	Reduccion de datos	Chi-Square	4.05	.043
9	REGRESION LINEAL	R	.928	.000	Escalas	Chi-Square	3.02	.083
10	REGRESION LINEAL	R	.928	.000	Pruebas no parametricas	Chi-Square	4.85	.028
11	REGRESION LINEAL	R	.928	.000	Series temporales	F	2.90	.093
12	REGRESION LINEAL	R	.928	.000	Superficies	F	3.73	.023
13	REGRESION LINEAL	R	.928	.000	Respuestas multiples	Chi-Square	4.88	.027
14	REGRESION LINEAL	R	.928	.000	Analisis de valores perdidos	Chi-Square	5.38	.019
15	REGRESION LINEAL	R	.928	.000	Muestras conexas	Chi-Square	5.93	.014
16	REGRESION LINEAL	R	.928	.000	Control de calidad	Chi-Square	4.16	.012
17	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	2.00	.020
18	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	5.00	.018
19	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	3.23	.073
20	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	5.58	.018
21	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	4.22	.024
22	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	5.96	.012
23	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	4.31	.023
24	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	3.80	.022
25	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	5.22	.015
26	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	5.00	.018
27	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	4.06	.020
28	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	4.97	.019
29	REGRESION LINEAL	R	.928	.000	Curva COX	Chi-Square	2.90	.093

フィールドワーク

場所(現地)を実際に訪れ、その対象を直接観察し、関係者には聞き取り調査やアンケート調査を行う。データの収集を行う。

リモートセンシング

地理的座標に関する事前情報を活用した画像解析などが可能となる。リモートセンシングデータは地理的な空間データ、あるいは時間座標も加えた時空間データとして扱うべきである。

地図資料

基盤地図情報
 国土数値情報
 ゼンリン地図情報
 CSIS統計データベース
 国勢調査地図データ
 アドレスマッチング

統計資料

国勢調査
 パーソントリップデータ
 物流データ
 事務所・企業統計
 住民基本台帳カード
 商業センサス

回帰分析とは

回帰分析は、因果関係を数学的に解析することを意図しており、変動の構造を定量化する手段である。ある変数の振る舞いをそれと関連する諸変数の変動を通じて明らかにする。原因と考える変数(説明変数)と、結果とする変数(被説明変数)との間に一方的な因果関係が存在すると仮定し、結果となる変数の変動は、1個あるいは複数個の説明変数によって規定されたとする。説明変数の変動が原因となって被説明変数が変動するとみなすので、説明変数は「独立変数」、被説明変数は「従属変数」とも呼ばれる。

回帰分析は、被説明変数 y とそれに影響を与える $x_1 x_2 x_3 \dots x_p$ の p 個の説明変数群との間に回帰式:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_px_p + e$$

切片
偏回帰係数(各変数要素の影響力)
残差項

を構築する。回帰分析は、理論値と観測値の相関係数を最大化するまたは残差を最小化する基準で未知の「偏回帰係数」 $a_1 a_2 a_3 \dots a_p$ を推定することにより、説明対象となる変数 y と原因となる変数 $x_1 x_2 x_3 \dots x_p$ の回帰関係を明らかにする。

決定係数 R^2 回帰式の説明力、すなわち回帰関数への当てはまりの程度である。

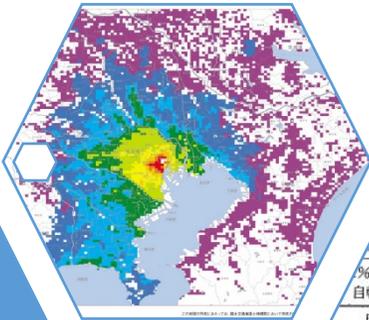
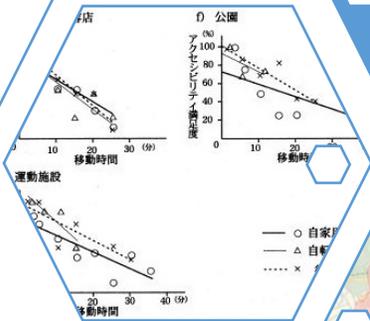
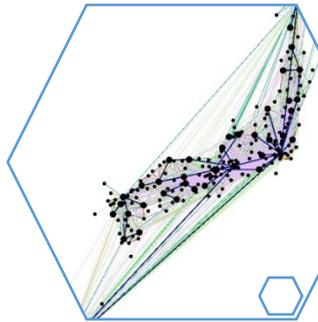
統計信頼度(confidence level)と有意水準(significance level)の検定

F検定、T検定

どんな領域に応用できるか?



人口移動
要因分析



地価分析



目的地	回帰係数	切片	決定係数
最寄品の買物先	4.95	107.08*	.96
買回品の買物先	1.82	86.34**	.623
医療機関	2.49*	86.00**	.739
金融機関	3.03*	97.36**	.936
美容・理容店	2.90*	92.70**	.724
公園	2.02	93.70	.39
運動施設	3.69	113.85**	

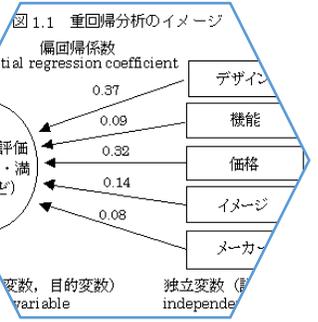
%水準で有意 *5%水準で有意
自転車

変数	回帰係数	標準誤差	t値	有意性
1	.07	.18	.38**	.24
2	.17	.12	.48	.10
3	.50	.53	-.02	-.10
4	.83**	.64*	.36	.10
5	1.20**	1.33**	-.29	.18
6	-.04	-.04	.00	.12
7	-.48*	.32	-.09	-.05
8	-.60	.80	.61	-.24
9	.21	.18	.03	.11
10	-.21	-.14	-.12	-.16
11	-.17	.33	-.38*	-.24
12	-.37*	-.18	-.06	.05
13	.78**	.64**	.17	.11
14	.64	1.19**	.25	.80*
15	-.08	-.05	-.03	-.27
16	-.15	-.78	.81**	-.49
17	1.75**	2.06**	1.96**	2.52**
18	.31	.24**	.89*	.71
19	-.25	.13	-.02	-.44
20	1.22**	.25	-.59*	.45
21	-.58*	-.41	-.59*	-.84**
22	.78	1.28**	.37	-.17
23	-.68	-.46	.46	.13
24	.63	-.41	1.35**	1.36**
25	1.41**	1.80**	1.03**	.39**
26	-.42**	-.46**	-.156**	-.316**
27	1.44**	.39**	.35**	.42**
28	1.44**	.39**	.35**	.42**

犯罪率
分析



選挙投票
結果の要因
分析



等々

消費者選
好分析

産業の
立地状
況

近接性
分析

●応用事例(1)—生活関連施設へのアクセシビリティ(近接性)の満足度により生活利便性の評価

表1 移動手段別にみたアクセシビリティ満足度に関する回帰分析の結果

Table 1 Result of regression analysis on accessibility satisfaction by transport mode

a) 自家用車

目的地	回帰係数	切片	決定係数
最寄品の買物先	2.92**	93.50**	.882
買回品の買物先	2.44**	91.46**	.860
医療機関	1.83*	85.05**	.749
金融機関	2.16**	88.34**	.960
美容院・理容店	2.42**	88.02**	.827
公園	1.25**	68.12**	.599
運動施設	1.82**	80.52**	.737

**1%水準で有意 *5%水準で有意

b) 自転車

目的地	回帰係数	切片	決定係数
最寄品の買物先	4.95	107.08*	.969
買回品の買物先	1.82	86.34**	.623
医療機関	2.49*	86.00**	.739
金融機関	3.03*	97.36**	.936
美容院・理容店	2.90*	92.70**	.724
公園	2.02	93.70	.350
運動施設	3.69	113.85**	.795

**1%水準で有意 *5%水準で有意

c) 徒歩

目的地	回帰係数	切片	決定係数
最寄品の買物先	2.43*	96.69**	.704
買回品の買物先	2.03**	100.26**	.826
医療機関	1.60*	94.03**	.728
金融機関	3.31**	104.16**	.969
美容院・理容店	3.33**	103.09**	.938
公園	2.42*	101.74**	.821
運動施設	2.08*	98.84**	.724

**1%水準で有意 *5%水準で有意

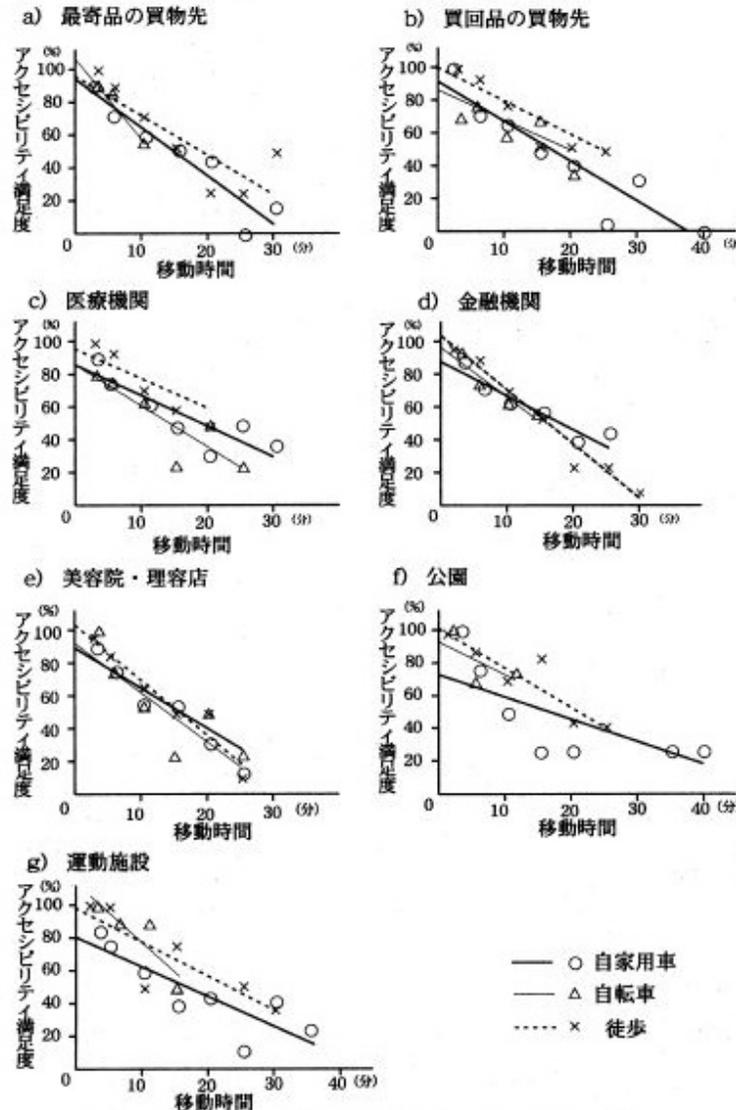


図3 移動手段別にみた各施設へのアクセシビリティ満足度の近似直線

○・△・×記号は各時間帯(5分)ごとの平均値
(1998年8月のアンケート調査により作成)

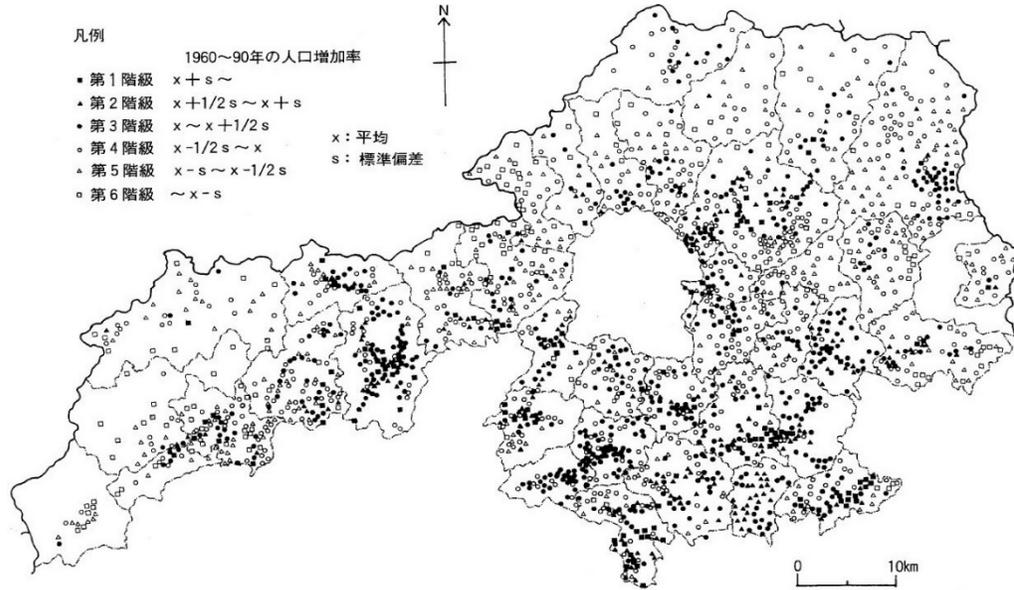
Fig. 3 Regression lines of accessibility satisfaction for facilities by transport mode

●結果解釈

表1は、単回帰分析の結果である。図1は、施設ごとにアクセシビリティ満足度をy軸、移動時間をx軸にとって描いた散布図である。自家用車利用に比べて自転車利用や徒歩利用の方が回帰係数は高い値を示し、回帰直線の傾きも急である。したがって、自家用車よりも自転車や徒歩を利用する方が、アクセシビリティ満足度に対する移動時間の距離減衰効果が高い傾向にある。この背景には、自家用車に比べて、徒歩や自転車による遠距離移動は施設利用者にとって負担になり、長距離の移動は行われないう現状がある。施設別に比較すると、医療機関や運動施設、公園などは他の施設に比べて回帰係数が小さく、回帰直線の傾きも緩やかである。一方、金融機関や美容院・理髪店は回帰係数が大きく、回帰直線の傾きも急である。このことから、利用者にとって前者を利用する際は移動時間を負担に感じないが、後者を利用する際は長距離の移動を避けることがうかがい知れる。

田中耕市2001. 個人属性別にみたアクセシビリティに基づく生活利便性評価: 福島県いわき市を事例として

●応用事例(2)―過疎化過程の地域的差異



第2図 人口増加率による階級区分の空間分布

第2表 全集落を対象としたステップワイズ重回帰分析の結果

変数	偏回帰係数	標準偏回帰係数	t 値	有意性
定数	0.799			
DID までの道路距離	-0.004	-0.172	-2.469	**
役場までの道路距離	-0.749	-0.435	-12.531	*
小学校までの道路距離	-0.056	-0.315	-5.248	**
農家率	-0.034	-0.281	-3.251	**
耕地率	0.046	0.335	5.933	**
旧役場集落との標高差	-0.035	-0.295	-3.551	**
重相関係数	0.589			
重決定係数	0.347			

注: * は 5%水準にて有意。** は 1%水準にて有意。

●結果解釈

図2は、過疎地域指定を受けた広島県の40市町村村内2,141集落(農家集落)における1960～1990まで30年間にける人口増加率を示したものである。作野(1994)は、このような分布パターンがいかなる要因によって生じたのかを、重回帰分析の適用により解明を試みた。具体的には、人口増加率を被説明変数、そしてDID(人口集中地区)、役場、小学校までそれぞれの道路距離、農家率、耕地率、旧役場集落との標高差の6指標を人口増加率に影響を与える説明変数とみなした。そして、ステップワイズ法を用いた重回帰分析を施し、6指標それぞれが人口増加率に与える影響を探った(表2)。

F検定の結果、得られた重回帰式は1%水準で有意であることが判明した。各説明変数の説明力は、標準変化回帰係数により示される。6つ説明変数の中では、「役場までの道路距離」の標準変化回帰係数が -0.435 と絶対値の最高であり、人口増加率の最も重要な説明要因である。役場、すなわち中心地からの距離が遠いほど、人口減少が進むことを意味している。2番目に重要な変数は、標準偏回帰係数0.335を有する「耕地率」である。符号がプラスであることから、農業集落における耕地率が高ければ、その集落の人口増加率は高くなることを意味する。ついで重要な変数は「小学校までの道路距離」である。以下、6変数はt検定の結果、5%水準ですべて標準偏回帰係数が有意であると判定された。

作野 広和1994. 広島県山間集落における過疎化過程の地域的差異

● 因子分析とは

因子分析は、相関のある変数(特性値)が有する情報を少数個の潜在的な「因子」に縮減する多変量解析の一手法である。その目的は、多変数のデータ行列をともに、それらのデータの背後に潜む共通因子を抽出し、それらを解釈することにより、データの構造を明らかにすることである。

因子分析においては、観測される p 個の変数 $x_1 x_2 x_3 \dots x_p$ が、それより少数の m 個の変数 $f_1 f_2 f_3 \dots f_m$ と x_j にのみ影響する変数 e_j との1次結合で表される仮定する。

$$x_j = \lambda_{j1}f_1 + \lambda_{j2}f_2 + \dots + \lambda_{jm}f_m + e_j \quad (j = 1, 2, 3, \dots, p)$$

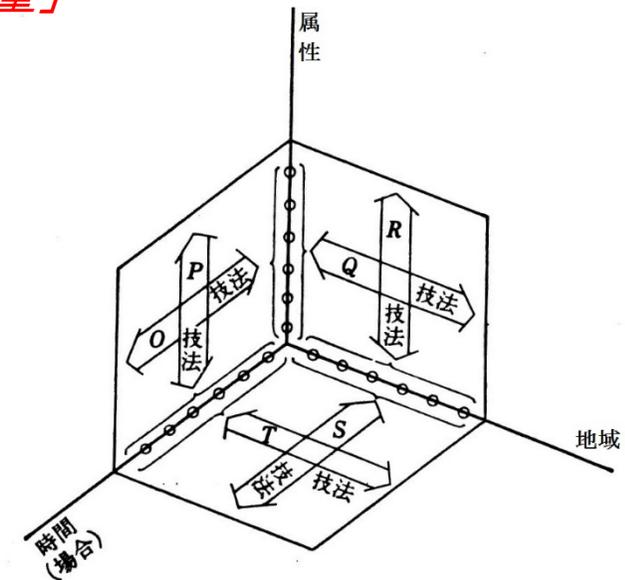
j 番目の変数 x_j の m 番目の因子 f_m への負荷「因子負荷量」

m 個の変数 f_m はすべて x_j に共通な潜在的要因であるので「共通因子」、それぞれの個体の共通因子の値は「因子得点」と呼ばれる。

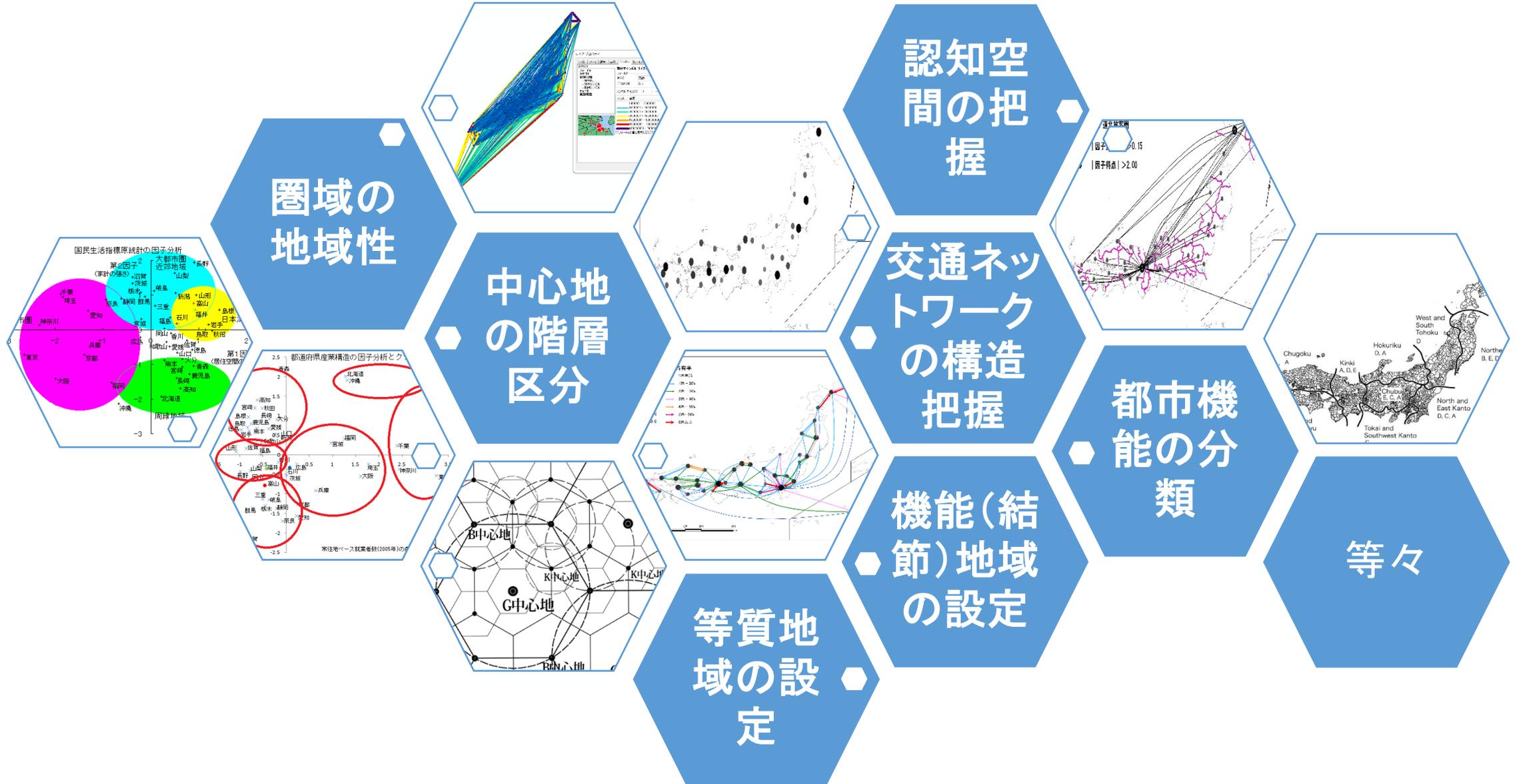
※SAS、SPSS、R等多くのソフトで因子分析を扱うことができる。

● 因子分析の技法と手順

地理行列データの作成 → 相関行列 → 因子分析を実施 →
 因子負荷量行列 → 因子得点行列 → 地図化



地理学にはどんな適用例があるか？



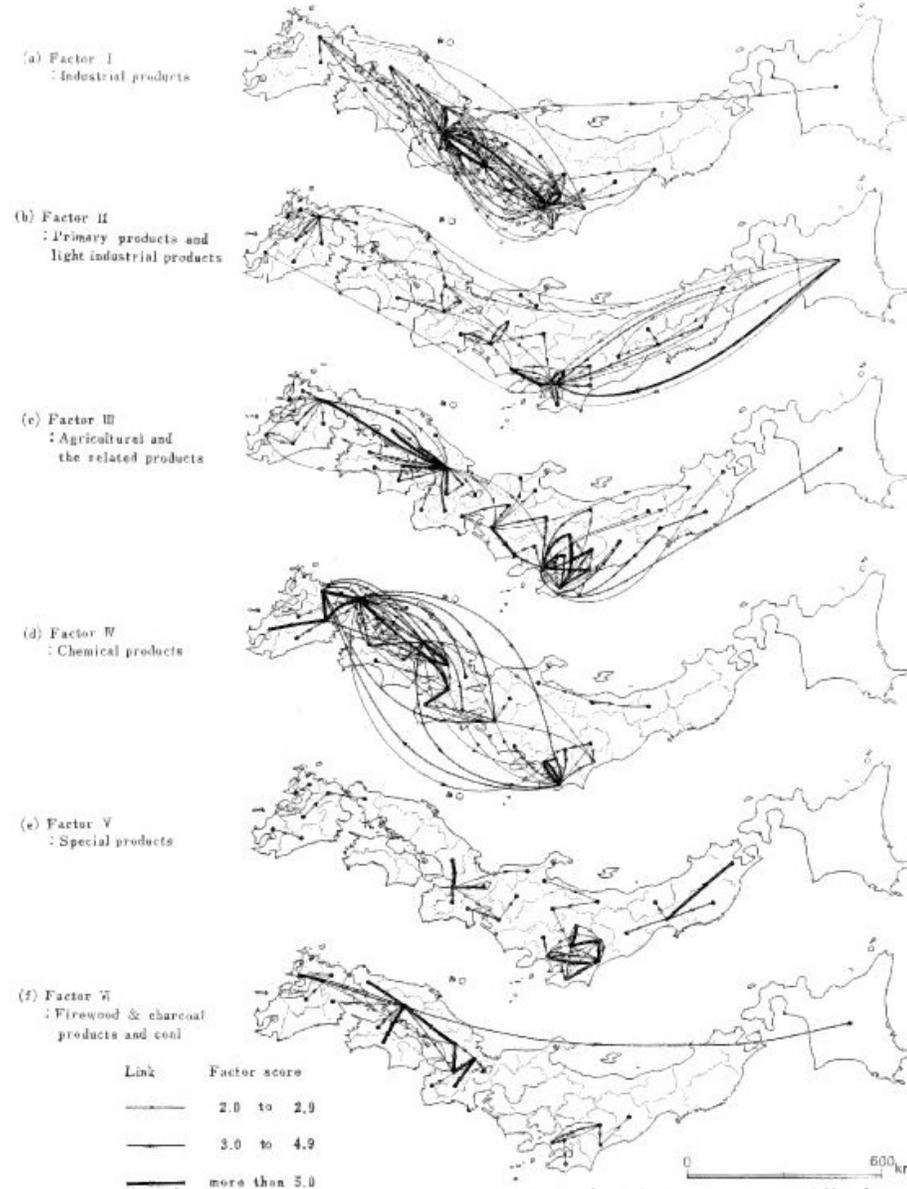
●応用事例(1)―貨物流動の地域構造

第3表 貨物流動の因子分析(1980年)
Table 3 Factor Analysis of Commodity Flows (1980)

Variables	Communality	Factor loadings					
		I	II	III	IV	V	VI
1 Grains	0.49			0.72			
2 Vegetables & fruits	0.20						
3 Livestock products	0.45		0.45				
4 Fishing products	0.38		0.53				
5 Other agricultural products	0.49			0.62			
6 Timber	0.43		0.43				
7 Firewood & charcoal products	0.15						0.55
8 Coals	0.11						(0.35)
9 Metallic minerals	0.13						
10 Non-metallic minerals for construction	0.37						
11 Non-metallic minerals for industry	0.24				0.40		
12 Metals	0.71	0.58			0.41		
13 Metal products	0.75	0.66					
14 Machine products	0.77	0.72					
15 Ceramic products	0.52				0.45		
16 Oil products	0.43			0.41	0.51		
17 Coal products	0.17						
18 Chemical medicines	0.63	0.40			0.66		
19 Chemical fertilizers	0.21						
20 Other chemical products	0.76	0.71			0.44		
21 Paper & pulp	0.57	0.49	0.51				
22 Textile products	0.50	0.63					
23 Foodstuffs	0.79	0.51	0.48	0.40			
24 Daily necessities	0.74	0.69	0.45				
25 Other manufacturing products	0.63	0.56	0.45				
26 Trash	0.32						
27 Feeds & manures for animals	0.51			0.63			
28 Wastes	0.19					0.45	
29 Transport containers	0.64	0.43				0.50	
30 Assorted goods	0.30					0.40	
Eigenvalues		11.56	1.61	1.26	1.21	1.14	1.04
Percent of total variance		38.5	5.4	4.2	4.0	3.8	3.5
Cumulative percent		38.5	43.9	48.1	52.2	55.9	59.4

Note: Only loadings exceeding ±0.4 are indicated.

第2図 貨物流動の地域的流動パターン(1980年)
Figure 2 Regional Commodity Flow Patterns (1980)



●結果解釈

第I因子
工業製品の流動

第II因子
第1次産品および軽工業品の流動

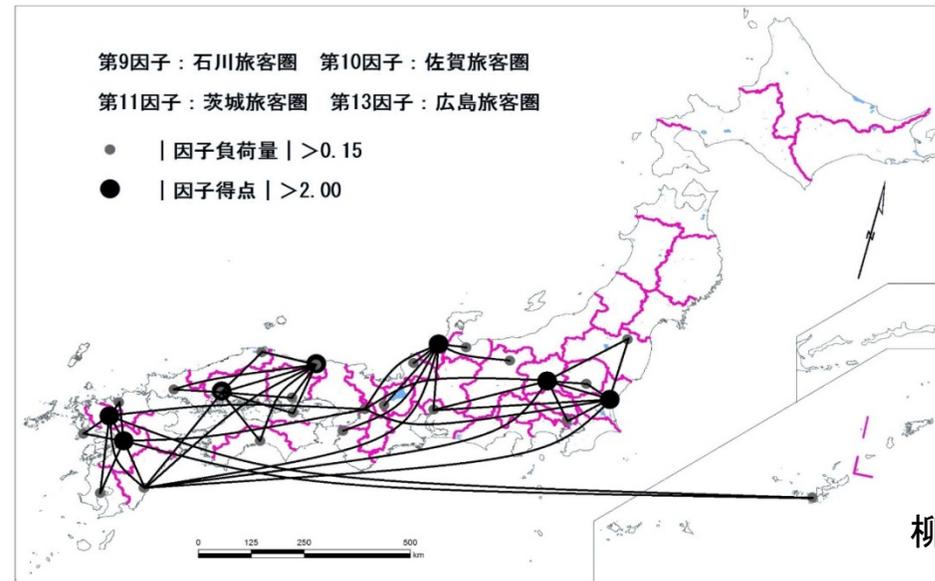
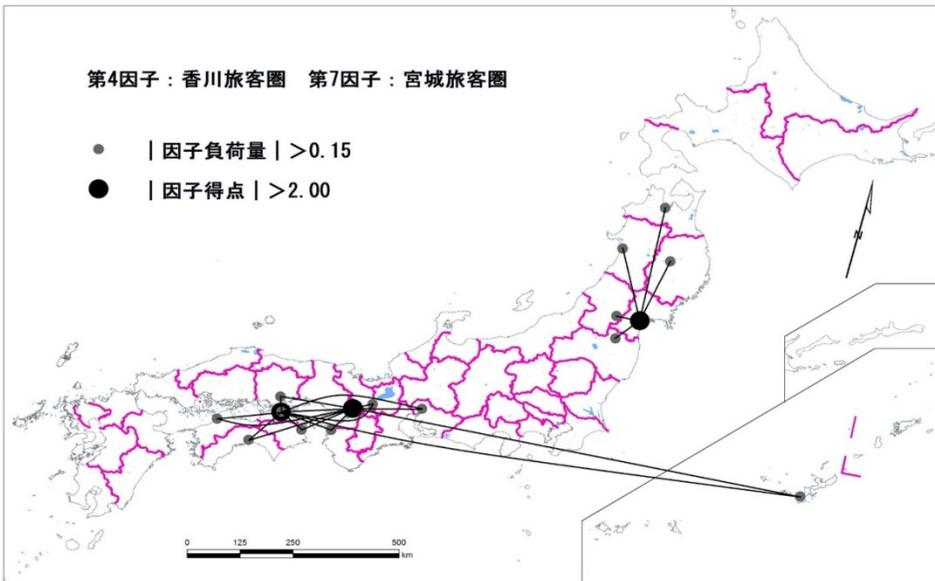
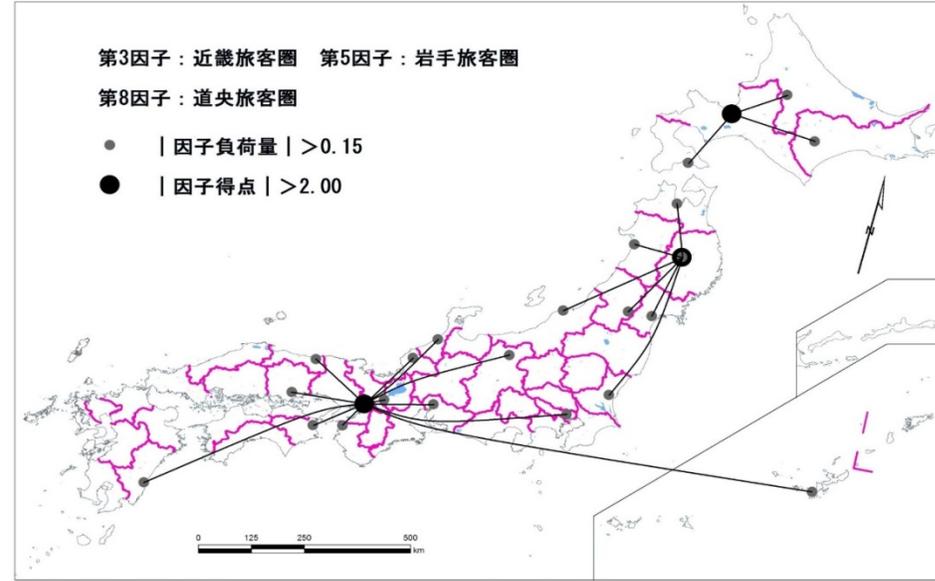
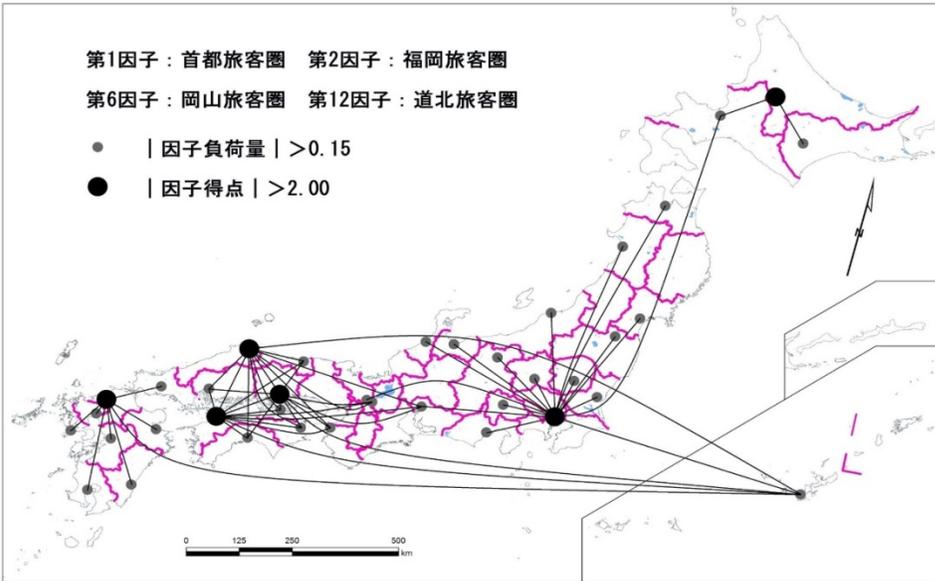
第III因子
農業品および関連品の流動

第IV因子
化学工業品の流動

第V因子
特殊品の流動

第VI因子
薪炭および石炭の流動

● 応用事例(2) — 交通ネットワークの結節構造

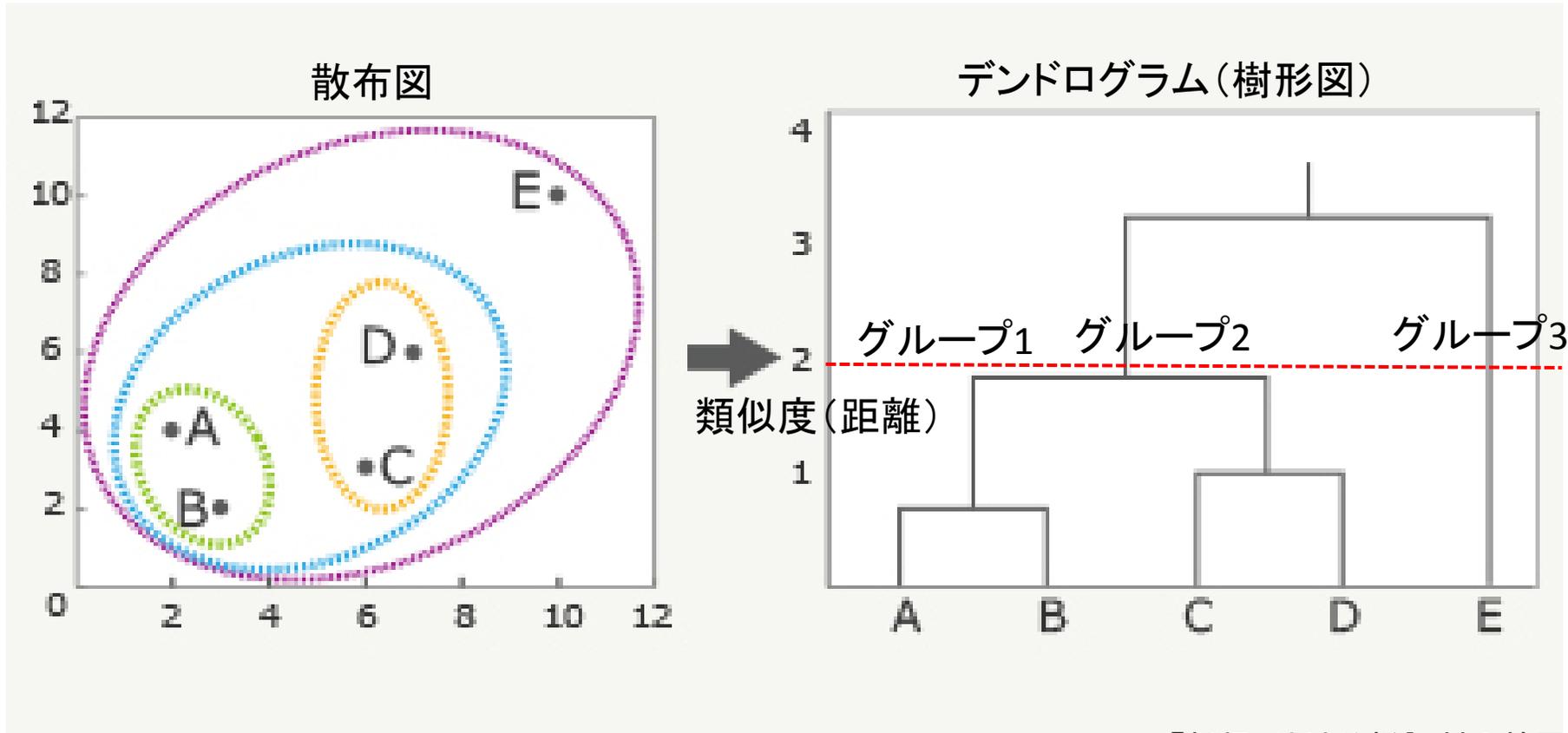


因子番号	分散説明量(%)	因子名(旅客圏)
1	16.88	首都旅客圏
2	9.36	福岡旅客圏
3	7.28	近畿旅客圏
4	6.74	香川旅客圏
5	5.79	岩手旅客圏
6	5.48	岡山旅客圏
7	5.44	宮城旅客圏
8	5.13	道央旅客圏
9	4.65	石川旅客圏
10	4.55	佐賀旅客圏
11	4.08	茨城旅客圏
12	3.72	道北旅客圏
13	3.47	広島旅客圏

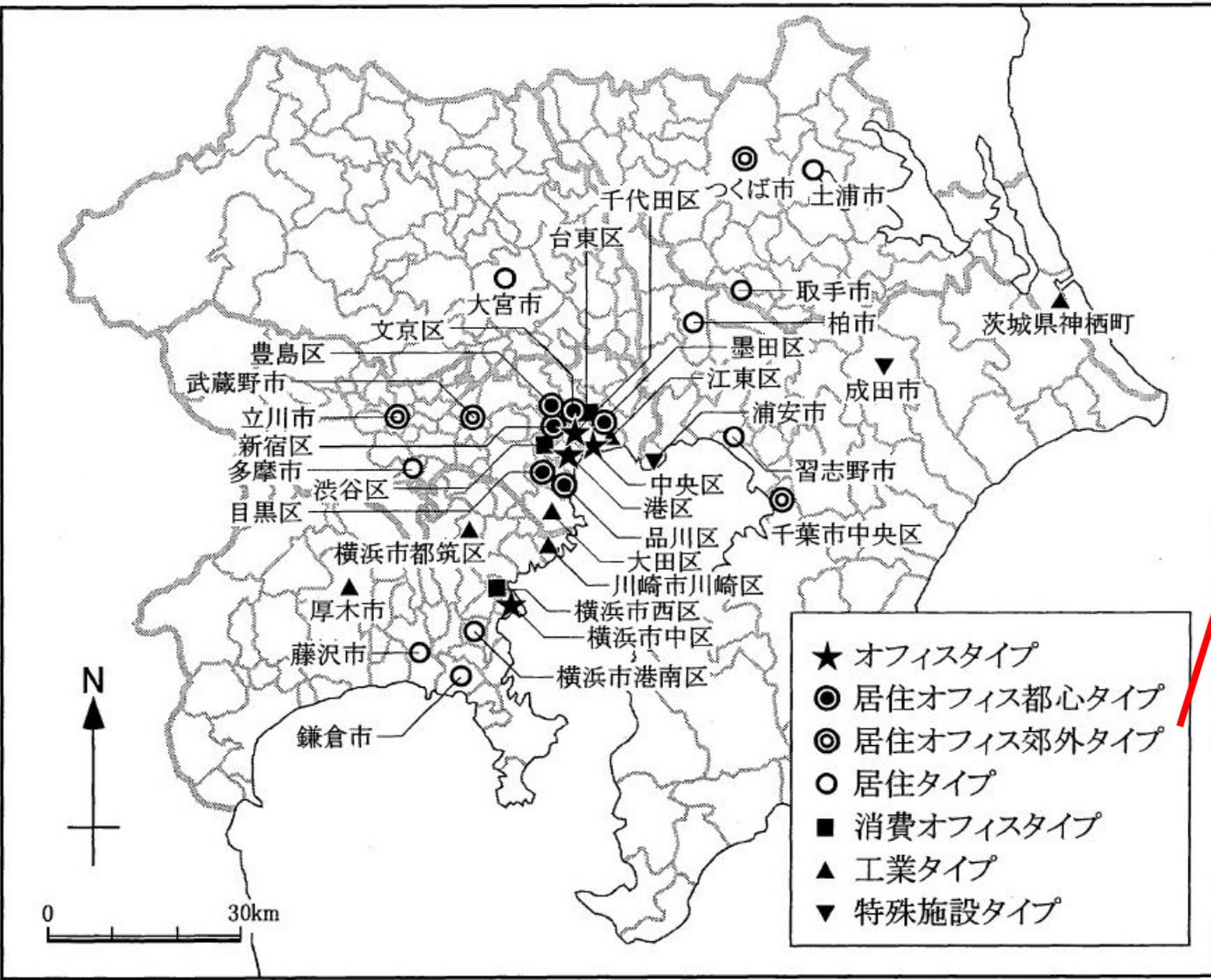
● クラスター分析とは

クラスター分析は、対象間になんらかの関係が存在し、その大きさが数値として与えられている時に、その対象間の関係を手がかりに、類似した対象を集め、全体をいくつかのグループ(クラスター 英:cluster)に分類することを目的とする手法である。

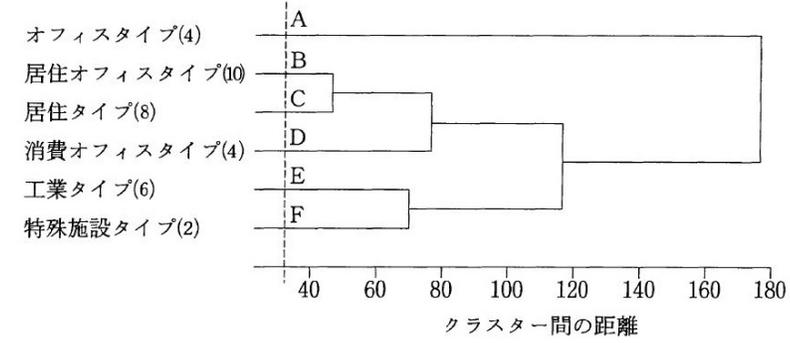
※SAS、SPSS、R等多くのソフトでクラスター分析を扱うことが出来る。



●応用事例(1)ーパーソンとリップ調査を用いた都市の類型区分



第5図 クラスタ分析ほかによりグループ化された中心核(1998年)
 (「第4回東京都市圏パーソントリップ調査」により作成)

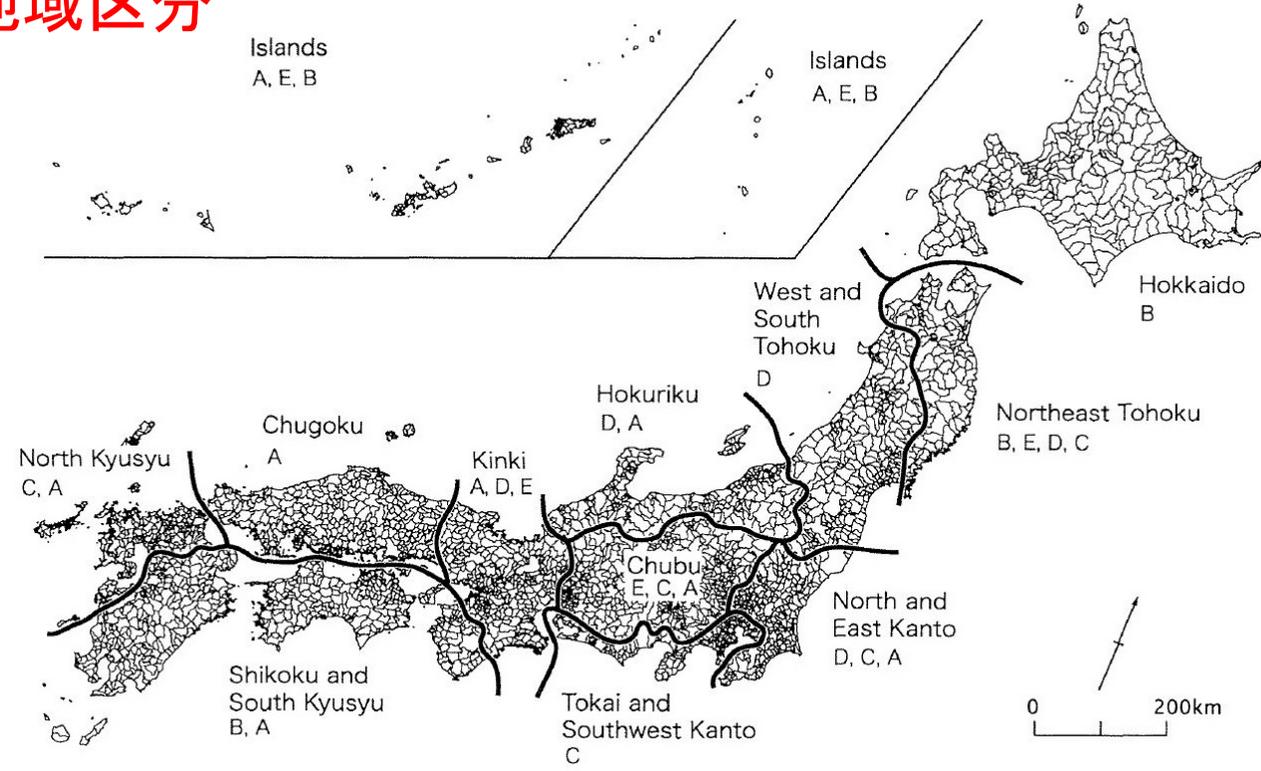
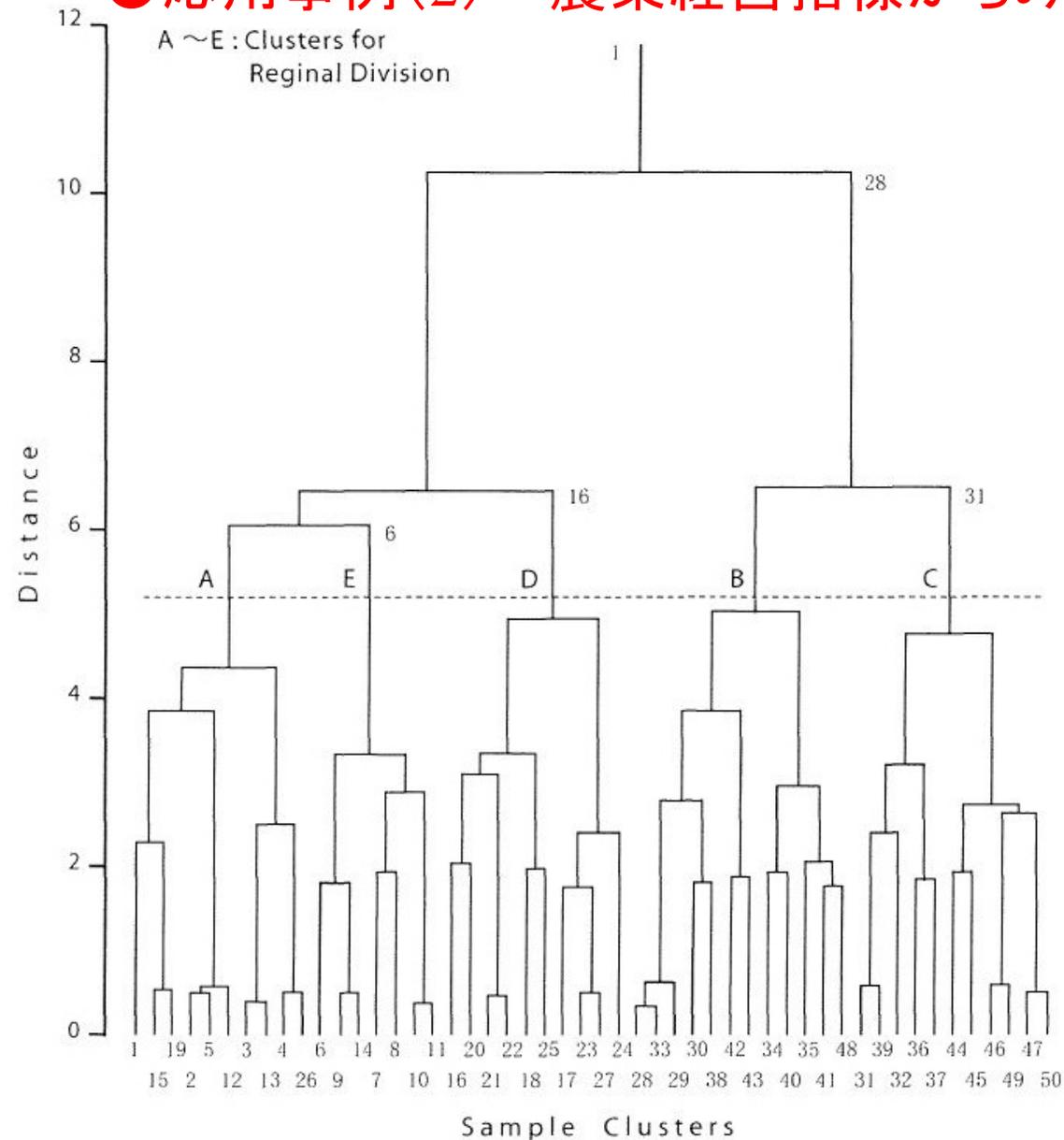


第4図 中心核に関するクラスター樹形図(第26段階以上)
 注) カッコ内は該当ゾーン数
 (「第4回東京都市圏パーソントリップ調査」により作成)

●結果解釈

駒木(2004)は、PT調査を用いて、東京都市圏における都市を6グループに類型化した。都市部においては、「オフィスタイプ」を中心都市、その周囲に「消費オフィスタイプ」、「居住オフィス都心タイプ」、「居住オフィス郊外タイプ」が隣接しており、すべてがオフィス機能を有していること、郊外部では「居住タイプ」、「居住オフィス郊外タイプ」に分類されており、全て居住機能を有していること、都心部と郊外部との関係においてオフィスと居住の対称性がみられること、郊外部でも郊外中心都市とそれ以外では特性に差異があること、などを指摘した。

● 応用事例(2) — 農業経営指標からみた地域区分



● 結果解釈

- Aタイプ: 農業就業人口率の高さ、稲作率の高さ、農業就業者の高齢化
- Bタイプ: 農業就業人口率の高さ、同居農業後継者の少なさ
- Cタイプ: 農業就業人口率の高さ、同居農業後継者の多さ、所有耕地面積の少なさ
- Dタイプ: 農業就業人口率の低さ、稲作率の高さ、農業就業者の高齢化
- Eタイプ: 農業就業人口率の低さ、稲作率の低さ

参考文献

- 駒木伸比古(2004): 通勤・消費行動からみた東京大都市圏の空間構造. 新地理, 52, 1-15.
- 作野 広和(1994): 広島県山間集落における過疎化過程の地域的差異. 人文地理, 46, 22-42.
- 田中耕市(2001): 個人属性別にみたアクセシビリティに基づく生活利便性評価: 福島県いわき市を事例として. 地理学評論, 74, 264-286.
- 仁平 尊明(2006): 農業経営に関する総合的な指標からみた日本の農業地域区分--多変量解析とGISの適用. 人文地理研究, 30, 69-89.
- 村山祐司(1984): 本邦における貨物流動の地域構造: 地理的場理論の枠組を用いて. 経済地理学年報, 30(2), 95-111
- 村山祐司・駒木伸比古(2013): 多変量解析で地域の特徴をさぐる. 村山祐司・駒木伸比古編『新版 地域分析』古今書院