

歯科疫学統計  
- 第7報 空間（地理）疫学の基礎 その1 -  
瀧口 徹

A review of oral epidemiological statistics  
- Part VII : The bases of Spatial (Geographical) Epidemiology 1<sup>st</sup> follow-up review -  
Toru Takiguchi

キーワード：空間（地理）疫学、地域集積性、地域（格）差、モラン統計量、局所空間統計量：LISA

はじめに

第4報から第6報<sup>1,2,3)</sup>でトレンド（傾向）分析を扱ってきた。今回扱った空間（地理）疫学も地理的分布の傾向をみるという意味で広義にはトレンド（傾向）分析である。しかし、その目的と手技はこれまで紹介してきたトレンド（傾向）分析とは大きく異なるので独立の項を起こすこととした。ただ第6報で紹介した統計R（statistical R environment）<sup>Web01)</sup>のspdep等の複数のパッケージが今回使用したソフトを構成する基本パーツであり、その意味では統計Rの有用性を追加紹介する意味も持っている。

1) 事例のプロフィール

著者居住の埼玉県の80区市町村の人口動態（出生率、死亡率、死因別年齢調整死亡率等）および医療関係職種（医師、歯科医師、薬剤師対人口比率）等をweb-siteの2006年埼玉県保健統計年報<sup>Web02)</sup>から得て事例とした。また地域集積性と

の差違が議論になると思われる地域格差に関して、南アジア7ヶ国の社会経済的指標の時代変化を変異係数（coefficient of variance：CV）で評価した研究<sup>4)</sup>、および26ヶ国間の経済格差およびその時代間比較をジニ係数（Gini coefficient）で評価した研究<sup>Web03)</sup>事例とした。また地域集積性の検証結果を疫学調査や保健政策に反映する際の有効手段となる局所空間統計量：local indicator of spatial association（LISA）の最新の応用例を紹介する。

2) 解説および解析手順

下記の手順で行う。

- (1) 地域集積性研究の意義
- (2) 解析ソフトウェア
- (3) 空間的自己相関指標（Moran's I, Geary's C）
- (4) 地域（格）差指標と地域集積性指標の使い分け
- (5) 局所空間統計量：LISAの有用性

【著者連絡先】

〒341-0003 埼玉県三郷市彦成3-86  
深井保健科学研究所  
主席研究員 瀧口 徹  
TEL&FAX：048-957-3315  
E-mail：taki8020@mth.biglobe.ne.jp

(1) 地域集積性研究の意義

疫学の歴史<sup>5)</sup>の中で、例えば近代疫学確立の規範となったJ Snow<sup>6)</sup>のコレラの伝播様式に関する研究において疾病発生の時間的（時系列的）視点と地理学的（空間的）な視点が重要<sup>5,6,7)</sup>であっ

た。しかし後者はマップを描きカテゴリーごとに濃淡あるいは色分けした分布から仮説を引き出すという主観的で統計的でない評価が主流で、疫学分野における扱いが補完的であり定量的な空間疫学の歴史は浅い。更に時系列傾向と空間的傾向を合わせた先行研究<sup>8)</sup>は極めて少ない。歯科領域においても学童う蝕や3歳児う蝕に関する時系列的な研究は1960年前後に山田、深田、片山、長田、高橋、相田ら多くの研究者<sup>9, 10, 11, 12, 13, 14, 15, 16, 17, 18)</sup>の研究があり地域（格）差および背景要因との関連性に言及している。しかし、いずれの先行研究においてもう蝕のマップ情報を採用し地域集積性に言及しているものの客観的な検証はなされていない。う蝕の地域（格）差は隣接地域（都道府県、区市町村等）と遠隔地で差は同じ程度が否か、もし異なっていたとしたらその原因は何かの視点で解析する必要があると考えられた。そこで著者らは先頃3歳児う蝕の地域集積性に関して47の都道府県単位<sup>19)</sup>と神奈川県内の58の区市町村単位<sup>20)</sup>の2つの研究を行いそれぞれの地域単位の解析で3歳児う蝕の地理分布に高度の地域集積性があることを確認した。

## (2) 解析ソフトウェア

ジニ係数（Gini coefficient）<sup>21)</sup>の計算は統計R（R statistical language <http://www.r-project.org>）

を用いた。また地域集積性の指標の計算および図示は統計Rのspdep等の7つのパッケージの簡便な利用のため筑波大学が開発しWeb上に公開している空間データ分析マシンSDAM<sup>Web-04, 24)</sup>を用いた。

## (3) 空間的自己相関指標（Moran's I, Geary's C）

a) 地域集積性研究は米国オハイオ州コロンバス市から始まった

空間的自己相関研究は1950年代にE.Moranによる米国オハイオ州コロンバス市犯罪（殺人）発生の地理的分布Web<sup>05)</sup>の分析から始まった。図1はそのコロンバス市の行政区分であるカウンティ単位でみて重大な犯罪（殺人および非過失性致死）発生の比率に地域集積性の有無を統計学的に判定した最初の事例である。ここでこのような図示法をコロプレス図（Choropleth Map）という。図1の左図が実測値でMoran's I = 0.486であり、この値は空間的自己相関（spatial autocorrelation）といい相関係数と同義である。すなわち地域集積性の相関係数は0.486であったという解釈になる。一方、右図は本検定の帰無仮説であるランダム分布の場合でMoran's I = -0.003である。一見したところ両図はさした違いが無いように見えることにこの検定の意義が大きい。すなわち我々の視覚はマップ情報から地域（格）差と地域集積性を正確

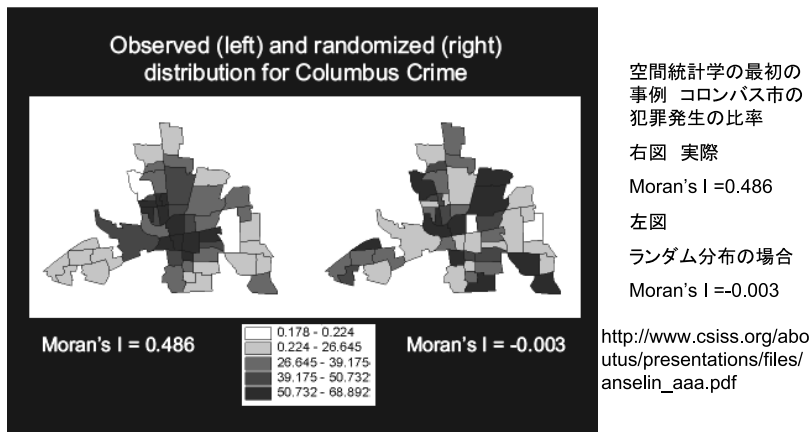


図1 空間的自己相関研究は1950年代に米国オハイオ州コロンバス市の殺人発生の分布の分析から始まった

に峻別できない。こうした解析結果は世間の安寧を著しく脅かす凶悪犯罪がランダムでなくある地域に偏っているということを証明し、地域の治安対策に貢献することになる。

b) ランダムな分布は地域集積性が無い

図2はExcel関数：randbetween()を用いて埼玉県の区市町村数の80にあわせて整数0,1,2,3,4を80個ランダムに発生(Choropleth Map)させ、順次機械的に区市町村に割り振った結果である。ランダムな数値は同値が続くことが頻繁にあるためマップ上も同一カテゴリーが近接かつ連続している区画が多く見られ、一見“地域集積性”があるかのような誤った印象を受けやすい。つまりコロ

ンバス市の例のように視覚のみに頼ると集積性を誤認しやすい。このMoranのI統計量に類似した統計量が直後に開発されたGearyのC統計量である。いずれの統計量も有意差検定は統計量が正規分布に近似する<sup>25)</sup>ことを利用して行われる。Moran's I = -0.065の解釈は相関係数と同義であり、この値は無相関であることを示している。すなわちランダムな分布は地域集積性(regional clustering)が無い。また0から+2に分布するGeary's C<sup>25)</sup>も有意差が無い。なお、MoranとGearyの統計量の違いはd)で後述する。

c) 背景因子(交絡因子)の調整で地域集積性は変化する

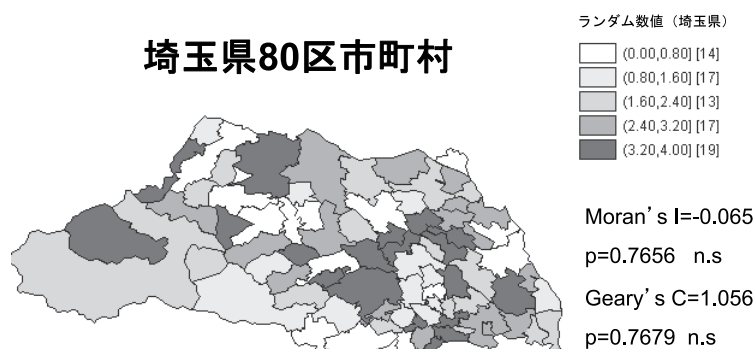


図2 ランダムな分布は地域集積性(regional clustering)が無い

- 1) Excel関数randbetween()を用いてランダムに整数0,1,2,3,4を80個発生させ埼玉県の80の区市町村に機械的に当てはめた場合のコロプレス図
- 2) ランダムな分布は同一の数値が連続することが多いのでマップ上も同一カテゴリーが近接かつ連続している区画が多く見られ、一見“地域集積性”があるかのような誤った印象を受けやすい。
- 3) Moran's I = -0.065は相関係数と同義で無相関である。0から+2に分布するGeary's Cも有意差が無い。

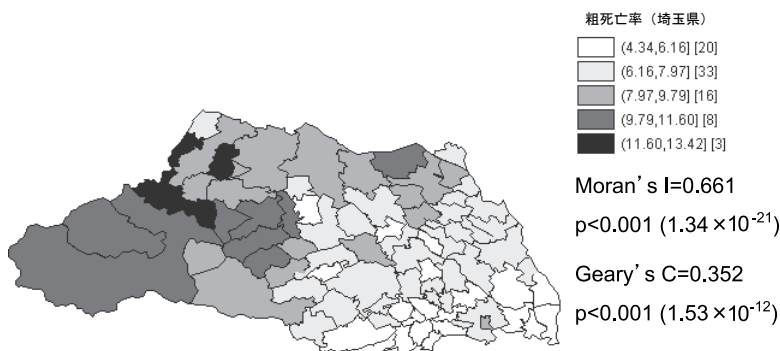


図3 粗死亡率の分布は高度の地域集積性が有る

図3は2006年の埼玉県の粗死亡率のコロプレス図である。左半分（西部）および上部（北部）に死亡率が高い市町村が集中していることがわかる。埼玉県においてこれらの地域は右側および下部（南部）の地域と比較して都市化度が低い。地域集積性はMoran's I=0.661  $p<0.001$  ( $1.34 \times 10^{-21}$ )で高度に有意でありGeary's Cも同様に高度に有意 ( $p<0.001$ ) で明らかな地域集積性がある。図4はモラン・プロット図<sup>24)</sup>を示す。X軸は実測値（この場合は粗死亡率）を示し、y軸は標本間距離である空間ラグ  $X_y$  を表している。この図に示される散布状況がMoran's I=0.661、すなわち空間

的自己相関係数=0.661の状態である。

図5は年齢調整した粗死亡率のコロプレス図である。図3において高度に有意であった地域集積性が消失していることがわかる。すなわち粗死亡率の地域集積性は住民の年齢差が交絡因子（背景要因）のいわば見かけの現象である。従って年齢調整をすると地域集積性の有意性が消失する。図3において左半分（西部）および上部（北部）は他地域と比較して高齢化率が高いことによって死亡率を高めていたということを示している。仮に年齢調整しても有意差がある場合は死亡率に関係する何らかの別の要因が残っていることを示して

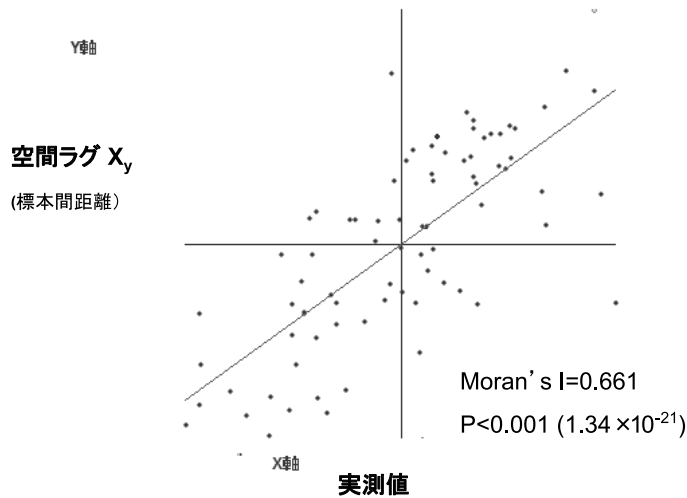


図4 粗死亡率のモラン・プロット（埼玉県）

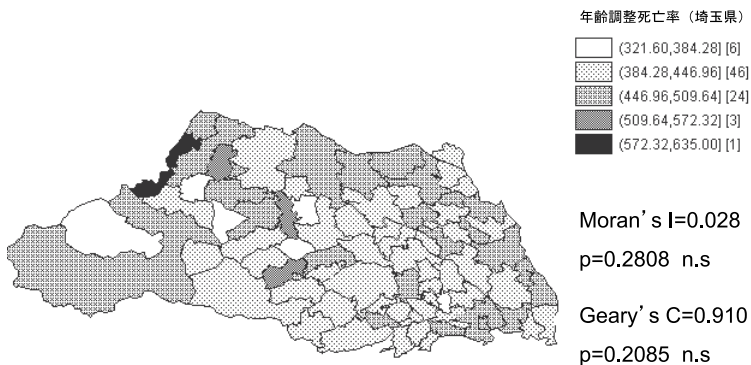


図5 年齢調整死亡率の分布は地域集積性が消失

いる。

表1は粗死亡率および年齢調整した死亡率、ガン、心疾患、脳血管障害、および糖尿病のMoran's IとGeary's C検定結果を示している。年齢調整死亡率を構成する3大死亡疾患および糖尿病いずれもが年齢調整すると地域集積性が消失する。但し、ガンのGeary's Cは有意水準に近く(p = 0.0851)、ガンの年齢調整死亡率の差が大きい区市町村の関係が一部にあることを示している。

d) Moran's IとGeary's C統計量の違い

表2にMoran's IとGeary's C統計量の計算

式<sup>Web06)</sup>を示す。両指標の計算式は近似している。しかし両指標の大きな違いは前者が分子の構成要素として任意の2点それぞれと全平均との差の積の総和を求めるのに対し、後者は任意の2点間の差の二乗の総和を求めている点である。両指標の統計的性質の詳細な比較がWalter SD<sup>26)</sup>によってなされている。それによれば①両指標は近似している、②その違いは複雑である、③Moran's Iは極端な値に反応性が高く、④一方Geary's Cは近接(neighbor)した領域の値の差に他法より敏感となる傾向がある、と報告している。図6、7は両指標の検定結果が異なる例である。区市町村の違

表1 粗死亡率および年齢調整死亡率(3大死亡率+糖尿病)の地域集積性(埼玉県)

	Moran's I		Geary's C	
	I	p	C	p
粗死亡率	0.66	0.0000	0.35	0.0000
死亡率	0.03	0.2807	0.91	0.2085
年齢調整				
ガン	0.02	0.3341	0.88	0.0851
心疾患	-0.06	0.7591	0.90	0.1607
脳血管障害	-0.06	0.7577	0.96	0.3357
糖尿病	-0.06	0.7543	0.97	0.3861

表2 Global Moran's IとGeary's Cの違い

出典: <http://www.css.cornell.edu/courses/620/lecture9.ppt>

$$I = \frac{N \sum_i \sum_j W_{i,j} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{i,j}) \sum_i (X_i - \bar{X})^2}$$

$$C = \frac{[(N-1) \sum_i \sum_j W_{ij} (X_i - X_j)^2]}{2(\sum_i \sum_j W_{ij} (X_i - \bar{X})^2)}$$

**N:** 対象エリア数(都道府県、市町村等)

**X<sub>i</sub>:** i番目のデータ値

**X<sub>j</sub>:** j番目のデータ値

**X:** 変数の平均値

**W<sub>ij</sub>:** 2点間を比較する場合の重み付け(隣接の有無、距離の逆数等)

いが地図上の南東部に偏在している図6は Moran's Iの③の性質を、一方全体的にばらついている図7は Geary's Iの④の性質を反映していると考えられる。また、Walter SD<sup>26,27)</sup> は両指標に rank adjacency statistic Dを加えた3つの指標の地域集積性の検出力の差や特徴についてシミュレーションを行った結果 Moran's Iが最も検出力が高く少数標本の場合や幾つか欠損値のあるデータにも頑健 (robust) と結論づけている。これらのことから地域集積性の解析に際しては Moran's Iを標準として、できれば特殊な集積性の検出のため Geary's Cも計算して Moran's Iと極端な差がある場合はその違いを考察する必要があるだろう。

(4) 地域（格）差指標と地域集積性指標の使い分け

a) 地域差と地域格差

地域差 (regional disparities, differences) と地域格差 (regional disparities, differences, inequities) は日本語の言葉としては同義語である。しかし前者は生態学的あるいは統計的な差、後者は保健医療福祉の視点からみた不公平 (inequity) という視角でみた差であり改善すべき政策課題<sup>28)</sup> のニュアンスがあるとされる。

b) 変異係数による地域（格）差検証

表3は変異係数<sup>28)</sup> 用いて南アジア7ヶ国の社会

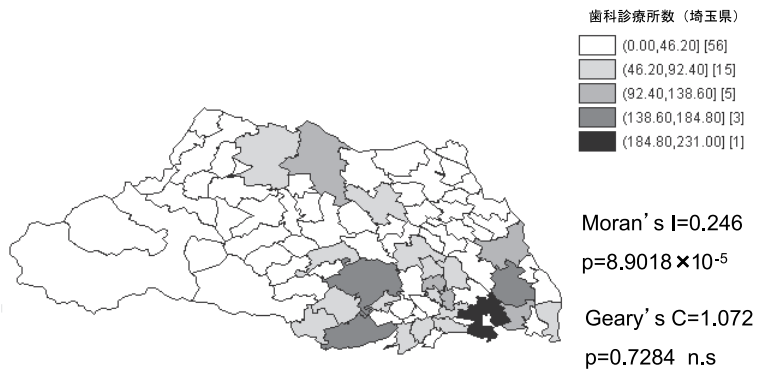


図6 Moran's Iと Geary's Cの有意性が著しく異なる場合 ー歯科診療所数(実数)の分布ー  
Moran's I: 高度に有意 Geary's C: 有意差無し

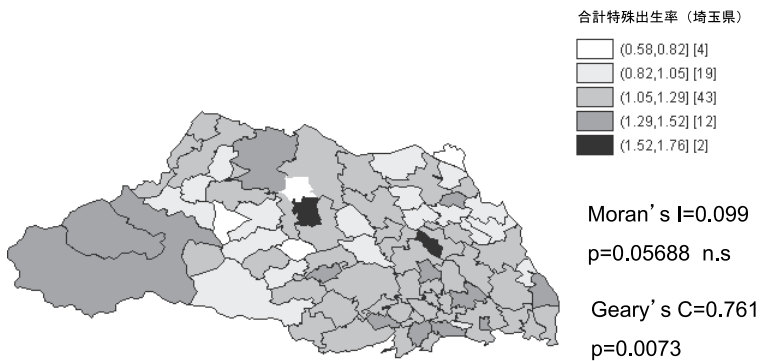


図7 Moran's Iと Geary's Cの有意性が著しく異なる場合 ー合計特殊出生率 (TPFR)ー  
Moran's I: 有意差無し Geary's C: 高度に有意



経済的指標の1971-72年から2001-2年までの時代変化を11の指標で検証<sup>4)</sup>したものである。灌漑地割合、電気使用量、識字率等6指標は地域（格）差が減少しているが乳児死亡率、交通機関の普及等は南アジア各国間で地域（格）差の減少がみられない。

c) ジニ係数による地域（格）差検証

図8にジニ係数<sup>21)</sup>の算出法と統計的意味を示す。ジニ係数はイタリアの統計学者Corrad Giniによって1936年にローレンツ曲線をもとに所得格差を表す指標として考案された。ジニ係数の計算は幾何的にLorenz曲線を用いる方法と平均差を用いる代数的な方法があるが本報告では後者の方法で

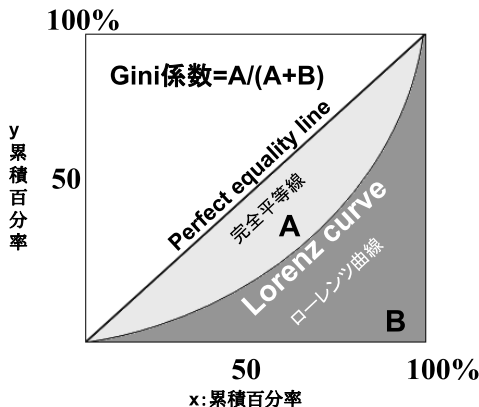
計算した。類似の統計量として分散と平均偏差がある。個々の値と平均値の差の二乗の総和を総数：nで割ると分散になり、個々の値と平均値の差の絶対値をnで割ると平均偏差となる。ジニ係数の場合は平均からの差ではなく個々の値の差（同一の値を除く）の総数をnで割った統計量である。0-1の範囲の値をとり数値が大きいほど格差が大きく、小さいほど格差が小さいことを示す。元々所得格差を捕らえる指標として開発されたが地域（格）差を捕らえる有用な指標としてもよく利用される。図9はジニ係数を用いて26ヶ国間の経済的不平等<sup>Web03)</sup>を1980、1990、2000年代の3時代間で比較したものである。26ヶ国間の比較で

表3 変異係数による南アジア7ヶ国の社会経済的指標の時代変化の検証

評価指標	変異係数:Coefficient of variance(CV)			
	1971-72	1981-82	1991-92	2001-02
1 IL: 総耕作地に占める灌漑地割合	0.995	0.763	0.643	0.618
2 PCE: 一人当たり電気使用量	1.050	0.987	0.975	0.861
3 PC: 港湾能力利用率	0.878	0.867	0.869	0.883
4 TL: 人口1000対電話線数	0.639	0.762	1.143	0.820
5 TF: 交通機関	1.173	1.139	0.982	1.177
6 FC: 耕地(ha)当たり肥料消費量	1.670	1.223	0.886	0.931
7 LR: 識字率	0.754	0.614	0.472	0.396
8 IMR: 乳児死亡率	0.743	0.799	0.877	0.712
9 BC: 銀行部門国内信用取引のGDP比率	1.023	0.597	0.513	0.365
10 AM: 耕地(ha)当たり肥料消費量	2.070	1.385	1.096	1.091
11 TC: 源泉徴収のGDPに対する比率	0.542	0.441	0.427	0.268

調査対象国: ネパール、バングラディシュ、ブータン、モルディブ、インド、パキスタン、スリランカ  
 トレンド: ↓: 調査4時点でCV値が連続的に減少している指標(差が減少してきている指標)

出典: Prabir De, Buddhadeb Ghosh: Effect of infrastructure on regional income in the era of globalization: new evidence from south Asia. Asia-Pacific Development Journal, vol 12,81-107,2005.



注1) Gini係数: ジニ係数はイタリアの統計学者Corrad Giniによって1936年にローレンツ曲線をもとに所得格差を表す指標として考案された。0-1範囲の値をとり数値が大きいほど格差が大きく、小さいほど格差が小さいことを示す。

注2) Gini係数の計算は幾何的にLorenz曲線を用いる方法と平均差を用いる代数的な方法がある。このうち本法では後者の方法で計算

Gini係数=  $\Delta / 2 \times \text{平均}$   
 $\Delta$ : 非重複平均差:  $\Sigma$  (全ての数値の差の絶対値) / (縦横nの数値行列の2つの数値の組み合わせ数 = n(n-1))

図8 ジニ係数とローレンツ曲線

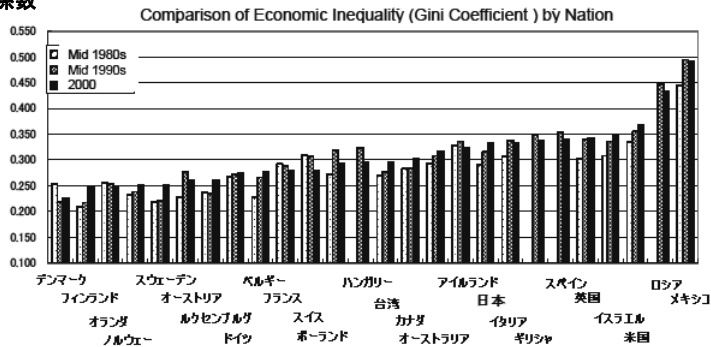
はメキシコ、ロシア、米国が三大地域(格)差国であること、一方デンマーク、フィンランド、オランダ等の北欧が最も地域(格)差が少ないこと、フィンランド、ベルギー、台湾、日本、米国等は地域(格)差が増加傾向にあること、オランダ、

フランス等は地域(格)差が減少していることが見て取れる。

d) 地域(格)差と地域集積性の関係

表4に埼玉県内の地域(格)差と地域集積性の関係を示す。比較対象とした13項目うち変異係数と

Gini 係数



[http://ocw.u-tokyo.ac.jp/wp-content/uploads/lecture-notes\\_eng/GF\\_12/12shirahase\\_eng.pdf](http://ocw.u-tokyo.ac.jp/wp-content/uploads/lecture-notes_eng/GF_12/12shirahase_eng.pdf)

図9 ジニ係数による経済的不平等の比較 - 26ヶ国間、時代間比較 - Sawako Shirahase, 2009

表4 地域(格)差と地域集積性の関係

地域差検出: 変異係数、ジニ係数 地域集積性検出: Moran統計量、ギアリー統計量

人口動態、死因別死亡率、医療関係者数の指標 <sup>1)</sup>	単位	地域(格)差				地域集積性	
		変異係数: CV		Gini係数		Moran統計量: I	Geary統計量: C
		CV値	ランク(降順)	Gini値	ランク(降順)		
1 人口	人	97.42	2	0.4773	1	***	n.s
2 人口粗出生率	人口千対	25.08	7	0.1414	7	***	***
3 人口粗死亡率	人口千対	23.75	9	0.1304	9	***	***
4 人口動態(補正)人口増加率 <sup>2)</sup>	人口千対	46.75	6	0.2674	5	***	***
5 合計特殊出生率(TPFR)	人	15.93	11	0.0870	11	n.s	n.s
6 年齢調整死亡率	人口10万対	10.36	13	0.0544	13	n.s	n.s
7 がん死亡率	人口10万対	10.51	12	0.0585	12	n.s	n.s (p=0.09)
8 心疾患死亡率	人口10万対	19.75	10	0.1079	10	n.s	n.s
9 脳血管障害死亡率	人口10万対	24.67	8	0.1312	8	n.s	n.s
10 糖尿病死亡率	人口10万対	68.47	4	0.3669	3	n.s	n.s
11 医療関係者	人口10万対	126.88	1	0.3890	2	n.s	n.s
12 医師数	人口10万対	67.35	5	0.2399	6	n.s	n.s
13 歯科医師数	人口10万対	74.49	3	0.3333	4	n.s	n.s

CVとGini係数間の単相関係数r (x:CV, y:Gini)= 0.9295

1) 出典: 2006年埼玉県保健統計年報【衛生統計年報】

<http://www.pref.saitama.lg.jp/A01/BP00/h1020/menu1.html>

2) Gini係数の計算は幾何的にLorenz曲線を用いる方法と平均差を用いる代数的な方法のうち本方法は後者の方法で計算 Gini係数=Σ/2\*平均 Δ: 非重複平均差(同一数値以外の全ての数値差の総和)/差の個数n(n-1)

3) ジニ係数は0-1に分布し1に近いほど格差が大きい

4) 人口増加率は出生率-死亡率で計算されるため負値を含みCV, Gini係数が異常値となる。これを回避するため最小値を0に補正した。

5) n.s: 有意差無し \*\*\*: p<0.001



ジニ係数のいずれかのランク（順位）が上位3位以内、すなわち地域（格）差が大きい指標は人口、糖尿病死亡率、医師数、および薬剤師数である。しかし地域集積性はMoran's Iのみが高度に有意で他は有意差が無い。一方地域集積性がMoran's IとGeary's Cいずれも有意である指標は粗出生率、粗死亡率、および（補正）人口増加率であるが、それぞれ対応する地域（格）差のランク（順位）は5~9位の範囲にあった。これらのことから統計学的には地域（格）差と地域集積性の関係は相関性がほとんど無いと言える。無論国語的な概念では地域（格）差の特殊な形態が地域集積性であるが地域（格）差を数量化する代表的な統計量の変異係数とジニ係数はデータ値の順番をどのように入れ替えても各統計量は全く同じである。一方、Moran's IやGeary's Cのような地域集積性を表す統計量順番を入れ替ると必ず結果に影響する。分散分析の加法定理のように全地域（格）差分散=ランダム分散+地域集積性分散+誤差 という入り子の関係があればわかりやすいがそういう関係は成り立たない。「ならば地域（格）差と地域集積性は統計的には全く別物ということではないのか」という問いに対し私見であるが「地域（格）差パターンの類型化」ということで両者は接近遭遇すると思われる。後述する局所空間統計量：LISAは地域（格）差と地域集積性の両方が混在した状況を類型化できる大変興味ある指標である。

#### (5) 局所空間統計量：LISAの有用性

a) LISAはglobal Moran's Iの有意性の地理的根拠を示す（冒険小説「宝島」のイメージ）

Luc Anselin (1995)<sup>29)</sup>はglobal Moran's Iだけではわからない地域集積性の存在位置を検出する指標として局所空間統計量：LISA<sup>29)</sup>を開発した。対象地図で地域集積性があるかないかを表す統計量global Moran's Iの結果を解釈する特殊なcluster mapが表示され近接した周囲の区分（市区町村等の区分）と対象領域の値が表5に示すように①high-high、②low-high、③high-low、④low-low、

⑤non-significantの5つに分類される。①はいわゆるhot spot、④はcold spotと言われる。②は周囲がほとんど（全て）ある疾患の有病率が高いのにそこだけ何故か低い組み合わせ、③は逆に周囲がほとんど（全て）低いのにそこだけ高い場合である。ここで②、③は外れ値（outlier）といい、有病率が高い区市町村に囲まれている有病率が低い区市町村、あるいは有病率が低い区市町村に囲まれている有病率が高い区市町村である。これらに注目し新たな調査の仮説を絞り込む手法をHot spot patterns分析<sup>27)</sup>という。global Moran's IとLISA統計量の関係はいわば少年冒険小説「宝島」に似ている。古文書（global Moran's I）はモラン島に宝があるということを示している。しかし、その情報だけでは宝に到達するのは至難の業である。そこでその宝の在処を示すのがLISA地図ということになる。トレジャーハンターが宝に到達するには何としてもこの地図を手に入れなければならない。

#### b) LISA分析例1 インドの人口動態

インドの2004年の国勢調査データを利用して593の地域区分（district）における人口動態5指標（人口密度、人口増加率、都市化度、出生率、死亡率）の地域集積性の検証<sup>Web07)</sup>をLISA法（local Moran's I）すなわち局所空間統計量解析（Hot(cold)spot patterns analysis）を用いて行った。1つの地域区分の人口は平均で170万人である。図10はそのうち出生率と死亡率のLISA分析を示す。

#### c) LISA分析例2 大気汚染と慢性虚血性心疾患（CIHD）

図11の例は2変量のLISA解析の例である。この研究ではこれまで解明されていなかった北米における人工衛星が捕らえた細かい地理メッシュごとの大気汚染（AOD）情報と慢性虚血性心疾患（CIHD）の因果関係<sup>30)</sup>の解明を目的としてAODとCIHDの関係と地理上の分布の重なりが解析された。大気汚染（AOD）データは喘息や気管支炎の発生と関連があるとされる空中浮遊微粒子（PM2.5）の探査衛星データ、一方CIHDデータは

表5 LISAによるホットスポット・パターン分析

1. hot spots(ホットスポット)	high-high: 高値に囲まれた高値エリア
LISA分析カテゴリー値が正	
2. cold spots(コールドスポット)	low-low: 低値に囲まれた低値エリア
LISA分析カテゴリー値が正	
3. outliers(外れ値)	LISA分析カテゴリー値が負
a) high-low: 低値に囲まれた高値エリア	
b) low-high: 高値に囲まれた低値エリア	
4. no significant (有意差なし)	

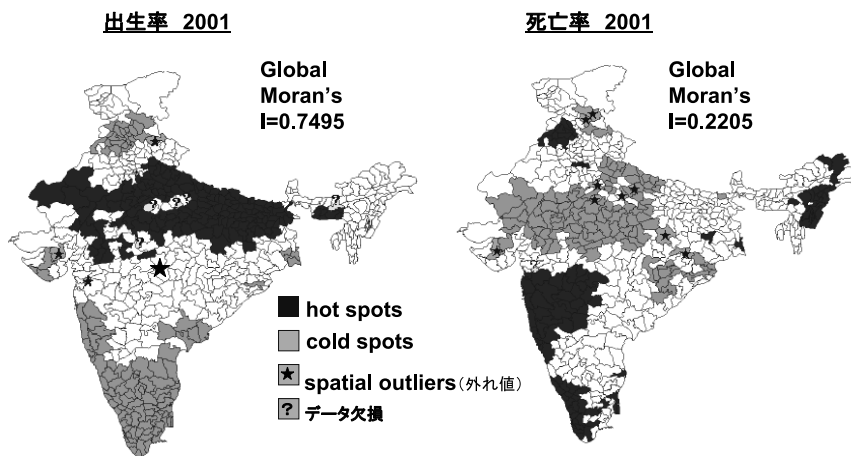


図10 LISA分析 例1ー インドにおける人口動態  
 出典：S. Oliveau : Spatial correlation and demography. Exploring India's demographic patterns, 2004.  
<http://iussp2005.princeton.edu/download.aspx?submissionId=51529>

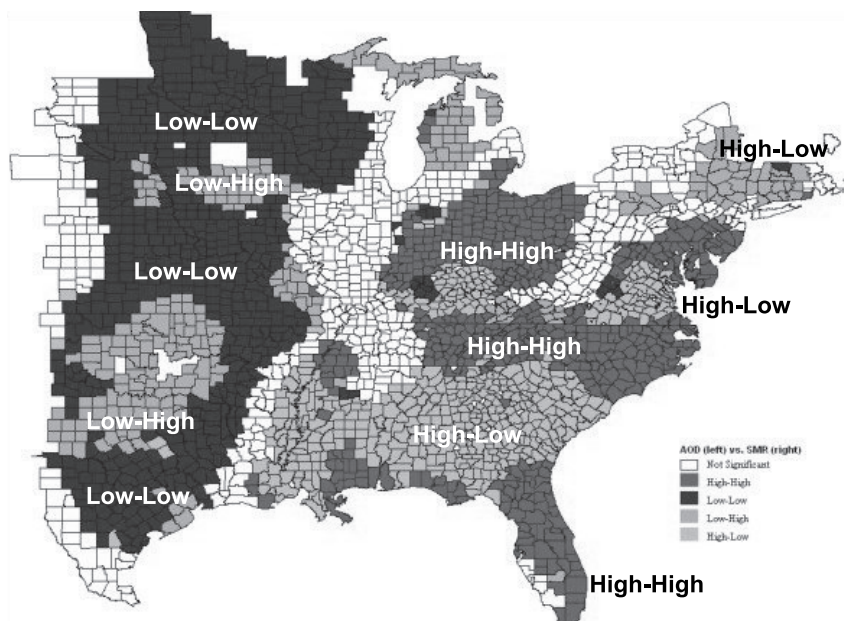
2003 - 2004年間の総数2306のカウンティにおいて人種と年齢で標準化したSMR (standardized mortality rate) が指標として用いられた。図11は2変量のMoran's I 散布図である。Moran's I 値は0.2673で高度に有意 ( $p < 0.001$ ) であり、大都市が集中する東部は高AODで高SMRが多く分布し、一方低AODで低SMRが西部 (ロッキー山脈の裾野付近) であった。またはずれ値 (outlier) である high-low エリアが南東部と北東部に、また low-high エリアが南西部と北西部にそれぞれ存在している。この例は個人単位の情報の集積と解析のみでは限界がある環境汚染と健康との関係を解析する上で強力な手段と言える。

d) LISA分析例3 人口自然増加率 (埼玉県)

図12は埼玉県の80区市町村の人口自然増加率のLISA分析結果を示す。表5示した判定基準に従いホットスポット、コールドスポットおよびはずれ値の区市町村を示している。

e) LISA分析の有用性を決める条件

開発から15年たったLISA分析結果が今後の疫学研究や保健医療施策にどう役立つかは未知数である。しかしこれまでの疫学的、統計的な解析に正確な地理情報が加わることはいままで主観的で曖昧だった情報の質が向上するわけであるから新たな知見の発見を予感させる。また先行研究から判断するに①対象疾患の重篤性 (致死性) あるい



global Moran's I = 0.2673 ( $p = 0.001$ )

**High-High:** 高大気汚染—高慢性虚血性心疾患  
**Low-Low:** 低大気汚染—低慢性虚血性心疾患  
**High-Low:** 高大気汚染—低慢性虚血性心疾患  
**Low-High:** 低大気汚染—高慢性虚血性心疾患

図11 LISA分析 例2— 人口衛星情報を用いた米国における微粒子大気汚染と慢性虚血性心疾患との関連  
 出典: Hu Z, K Rao KR: Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data. Environ Health 8 : 26-35, 2009.

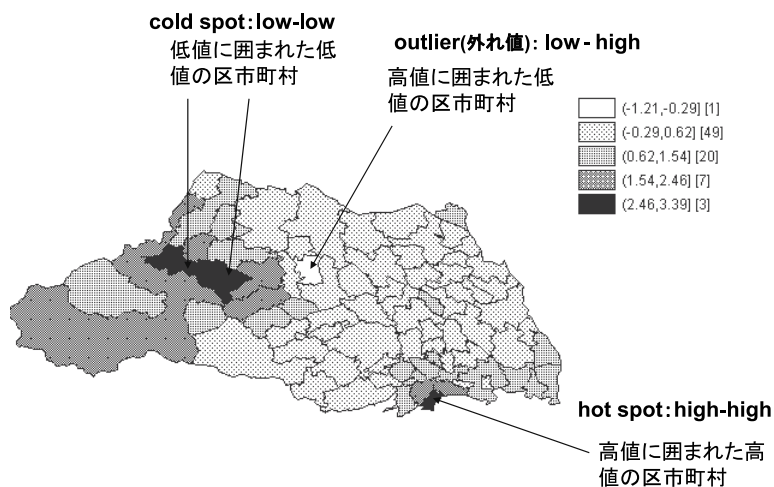


図12 LISA分析 例3— 人口自然増加率（出生数—死亡数）  
 注) hot spotsと cold spotsはいずれも Local Moran統計量:  $I_i$ は正值となり実際値の大小関係から hot pots、cold spotsを判定する。一方、はずれ値は負値となる。

は対象事象の重大性、②原因不明、③伝染性（伝播性）、④急性（緊急性）、⑤対策未確立、これら全ての条件または幾つかの条件に該当する場合、かつ正確な地理情報を欠損値が出来るだけ少ない状況でデータを入手できた場合のLISA解析は極めて有効な情報を提供するであろう。

#### おわりに

「歯科疫学統計」シリーズの第7報として新たなテーマである空間（地理）疫学、特に地域集積性研究の意義と代表的指標を紹介した。加えて地域（格）差との概念の違いを事例を用いて概説した。地域集積性の計算はかなり複雑な地図情報処理と統計処理を合体させたものであり、そのため一頃のパソコンでは計算が相当難しかった領域である。この意味で関連汎用ソフトの一般化にはまだまだ時間がかかると思われる。しかし、一昔前のようにスーパーコンピュータでしか計算できないというような特殊なテーマ（例えば、気象予測や銀河どうしの衝突シミュレーション等）は別として市販解析ソフトの価格あるいは開発されたシステムの利用可能な大学または研究所等の研究機関に帰属していないと利用できないという問題が壁になっている例が多いと思われる。幸い第6報で紹介した統計R、数値計算ソフトMaxima、本例の筑波大学提供の空間統計フリーソフト：SDAMのようにこうした壁を国際的な視点で撤廃したソフトの利用が可能な時代になりつつある。今後益々進化し利便性が高くなるインターネット環境下で操作性が飛躍的に向上した統計解析法の1つがこの空間疫学である。この流れを有効に利用する基礎として本報が役立つべきである。

謝辞：本総説を執筆するにあたり各段階でご指導いただいた歯科疫学研究会の顧問である重松逸造、蓑輪真澄、高江洲義矩、境 脩、伊藤学而、佐々木英忠の各先生に深謝申し上げます。さらに貴重なご助言をいただいた深井穫博同研究会会長をはじめとする幹

事の方々に感謝申し上げます。また空間分析ソフトSDAMを開発提供していただいている筑波大学に深謝します。

#### 参照 Web-site (2010/2/7 現在)

- Web01) Rのセットアップ+R入門  
<http://cwoweb2.bai.ne.jp/~jgb11101/files/cart/cart.html> (2009年12月1日アクセス)
- Web02) 埼玉県保健統計年報  
<http://www.prefsaitama.lg.jp/A01/BP00/b1020/menu1.html><sup>[Web01]</sup> (2009年12月1日アクセス)
- Web03) ジニ係数による経済的不平等の比較ー26ヶ国間、時代間比較ー  
[http://ocw.u-tokyo.ac.jp/wp-content/uploads/lecture-notes\\_eng/GF\\_12/12shirahase\\_eng.pdf](http://ocw.u-tokyo.ac.jp/wp-content/uploads/lecture-notes_eng/GF_12/12shirahase_eng.pdf) (2010年1月10日アクセス)
- Web04) 空間データ分析マシン「SDAM」ユーザガイド  
[http://giswin.geo.tsukuba.ac.jp/teacher/murayama/sdam/doc/SDA\\_userguide.pdf](http://giswin.geo.tsukuba.ac.jp/teacher/murayama/sdam/doc/SDA_userguide.pdf) (2009年12月1日アクセス)
- Web05) オハイオ州コロンバス市の殺人の地域集積性  
[http://www.csiss.org/aboutus/presentations/files/anselin\\_aaa.pdf](http://www.csiss.org/aboutus/presentations/files/anselin_aaa.pdf) (2009年10月10日アクセス)
- Web06) Moran's I統計量とGeary's C統計量の計算式  
<http://www.css.cornell.edu/courses/620/lecture9.ppt> (2009年4月10日アクセス)
- Web07) インドの人口動態LISA分析  
 Oliveau : Spatial correlation and demography. Exploring India's demographic patterns, 2004.  
<http://iussp2005.princeton.edu/download.aspx?submissionId=51529> (2009年10月1日アクセス)

#### 文 献

- 1) 瀧口 徹：歯科疫学統計 ー第4報 傾向（トレンド）分析の基礎ー, Health Science and Health Care, No 1 vol 6 : 53-67, 2006.
- 2) 瀧口 徹：歯科疫学統計 ー第5報 傾向（トレンド）分析の基礎 その2, Health Science and Health Care, No 1 vol 8 : 19-33, 2008.
- 3) 瀧口 徹：歯科疫学統計 ー第6報 傾向（トレンド）分析の基礎 その3, Health Science and Health



- Care, No 1 vol 8 : 19-33, 2008.
- 4) Prabir De, Buddhadeb Ghosh: Effect of infrastructure on regional income in the era of globalization: new evidence from south Asia. *Asia-Pacific Development Journal*, vol 12, 81-107, 2005.
  - 5) 柳川 洋編集：疫学マニュアル，南山堂 5版6刷，2000，2-7頁
  - 6) John M.Last編，重松逸造ら監訳：疫学事典，(財)日本公衆衛生協会，第1版，1987，56頁.
  - 7) 深井稜博，池主憲夫，川口陽子ほか：口腔保健推進ハンドブックー地域を支えるオーラルヘルスプロモーションー Part4 20 疫学手法と統計，医歯薬出版，第1版，2009，185-189頁.
  - 8) Fosgate GT, Carpenter TE, Chomel BB et al: Time-Space Clustering of Human Brucellosis, California, 1973-1992. *Emerg Infect Dis* 8 : 672-678, 2002.
  - 9) 山田文夫：食料消費量を中心とした我が国齲蝕罹患の地理病理学的考察. *口腔衛生会誌* 17 : 21-28, 1967.
  - 10) 深田英朗，山田 茂，石井欣一ほか：小児ムシバ，医歯薬出版，東京，第1版，1962，27-35頁.
  - 11) 片山 剛，氏家高志，長田公子ほか：3歳児歯科健康診査成績の時系列解析. 都道府県別にみた齲蝕有病者率の推移. *口腔衛生会誌* 36 : 609-614, 1986.
  - 12) 長田公子，片山 剛，氏家高志ほか：3歳児歯科健康診査成績の時系列解析. 2. 都道府県別にみた一人平均齲蝕数の推移. *口腔衛生会誌* 37 : 57-62, 1987.
  - 13) 高橋文恵，片山 剛，長田公子：3歳児歯科健康診査成績の時系列解析. 3. 3歳児歯科保健水準の地域較差と社会特性指標の関連性. *口腔衛生会誌* 39 : 264-273, 1989.
  - 14) 高橋文恵，片山 剛，長田公子ほか：3歳児歯科健康診査成績の時系列解析. 4. 3歳児歯科保健水準の時系列変化と社会変動の関連性. *口腔衛生会誌* 40 : 216-223, 1990.
  - 15) 相田 潤，安藤雄一，青山 旬ほか：経験的ベイズ推定値を用いた市町村別3歳児う蝕有病者率の地域比較および歯科保健水準との関連. *口腔衛生会誌* 54 : 566-576, 2005.
  - 16) Aida J, Ando Y, Aoyama H et al: An ecological study on the association of public dental health activities and sociodemographic characteristics with caries prevalence in Japanese 3-year-old children. *Caries Res* 40 : 466-472, 2006.
  - 17) Aida J, Ando Y, Oosaka M et al: Contributions of social context to inequality in dental caries: a multilevel analysis of Japanese 3-year-old children. *Community Dent Oral Epidemiol* 36 : 149-156, 2008.
  - 18) 相田 潤，森田 学，安藤雄一ほか：歯科疾患の地域差の検討. *J.Nat.Inst.Public Health* 57 : 93-98, 2008.
  - 19) 瀧口 徹，深井稜博，安藤雄一ら：全国都道府県3歳児 dmft 指数の25年間(1981-2006)の時系列変化の地域集積性. *口腔衛生会誌* ?? : ???-???, 2010.
  - 20) 平田幸夫，瀧口徹，山本龍生ら：神奈川県区市町村3歳児う蝕の25年間(1981-2006 間)の変化の地域集積性. *口腔衛生会誌* ?? : ???-???, 2010.
  - 21) 木村和範：ジニ係数の形成，北海道大学出版会，札幌，第1刷，2008，247-265頁
  - 22) Crawley MJ: *Statistics An introduction using R*. John Wiley & Sons, Ltd, West Sussex, England, 6th ed., 2009, pp. 103-154.
  - 23) Crawley MJ著，野間謙太郎，菊池泰樹訳：統計学：Rを用いた入門書，共立出版，東京，初版，2008，111-166頁.
  - 24) 村山祐司，駒木伸比古：空間データ分析マシン(SDAM)を活用した計量地理学の講義と実習. *人文地理学研究* 30 : 99-112, 2006.
  - 25) Stephen P. Kaluzny, Silvia C. Vega, Tamre P. Cardoso, et al: *S+ SPATIAL STATS - User's Manual for Windows and UNIX*, Insightful Corporation, Seattle, 1998, pp. 124-127.
  - 26) Walter SD : The analysis of regional patterns in health data. II. The power to detect environmental effects. *Am J Epidemiol* 136 : 742-759, 1992.
  - 27) Walter SD: Assessing spatial patterns in disease rates. *Stat Med* 19-20 : 1885-1894, 1993.
  - 28) 加賀美雅弘：病気の地域差を読む 地理学からのアプローチ，古今書院，東京，第1刷，2004，11-8，75頁.
  - 29) Byron Moldofsky, Justin Ngan, Angela Colantonio: *GIS and Cartography at the University of Toronto Technical Paper Series -Paper No.3*, Toronto Univ., Toronto, 1st ed., 2008, pp.25-25.
  - 30) Hu Z, K Rao KR: Particulate air pollution and chronic ischemic heart disease in the eastern United States: a county level ecological study using satellite aerosol data. *Environ Health* 8 : 26-35, 2009.

## A review of oral epidemiological statistics

ー Part VII : The bases of Spatial (Geographical) Epidemiology 1<sup>st</sup> follow-up review ー

Toru Takiguchi

(Fukai Institute of Health Science)

Key words : Spatial (Geographical) epidemiology, Regional clustering, Regional disparity, Moran' s I, local indicator of spatial association (LISA)

Abstract: In this review the significance and usefulness of three indicators (global Moran' s I, global Geary' s C and LISA: local indicator of spatial association) which are the statistical tools to verify and access regional clustering in terms of geographical epidemiology.

Contents of this review:

- (1) Significance and advantages of analyses of regional clustering
- (2) Soft wares for calculation of regional clustering
- (3) Indicators for spatial autocorrelation (Moran' s I and Geary' s C)
- (4) How to separate usage of indicators for regional disparity using coefficient of variance (CV) or Gini coefficient and indicators of regional clustering
- (5) Usage and availability of LISA

Health Science and Health Care 9 (1) : 19 – 32, 2009