

ARTICLES

歯科疫学統計
ー第8報 空間（地理）疫学の基礎 その2ー
ー地域差をとらえる指標の相互関係ー

瀧 口 徹

A review of oral epidemiological statistics
ー Part Ⅷ：The bases of Spatial（Geographical）Epidemiology 2nd follow-up review ー
Toru Takiguchi

キーワード：地域差、変異係数、ジニ係数、モラン統計量、局所空間統計量：LISA

はじめに

第7報¹⁾で扱った地域（格）差と地域集積性の関係が複雑で分かりにくいというご指摘を受けた。そこで今回同一の項目で各指標を比較することにより両者の特徴の違いをより明確にすることを試みた。用いた指標は地域差を表す指標として変異係数とジニ係数、地域集積性の代表的指標としてMoran's I指標（以下、モランのI）および局所空間統計量：LISA（Local Indicator(s) of Spatial Association, 以下 LISA）である。

1. 事例のプロフィール

日本の総人口の約1／3が集中し、経済、文化の中心である関東7都県の277市区町村（2005年10月1日現在）の部門別産業就業者割合、および勤務施設種別歯科医師数人口10万比（2008年12月31現在）の2種類、5項目の地域情報を地域差と地域集積性解析の共通の事例とした。

2. 解説および共通事例による解析

下記の手順で行う。

＜解説＞

- (1) 地理疫学と地域集積性研究の経緯
- (2) 解析ソフトウェア
- (3) 変異係数（変動係数）
- (4) ローレンツ曲線とジニ係数
- (5) 空間的自己相関指標（モランのIとLISA）

＜同一事例による比較検証＞

- (1) 事例対象と地図情報化
- (2) 変異係数による地域差検証
- (3) ローレンツ曲線、ジニ係数による地域差検証
- (4) 地域集積性による地域差検証

(1) 地理疫学と地域集積性研究の経緯

a) 英国ロンドン・ブロードストリートから始まった地理疫学

図1は米国CDC（疾病予防管理センター）提供の疫学・統計ソフトEpiInfo^{Web01)}の開始画面である。この画面はジョン・スノー（John.Snow）が作成したコレラ発生頻度のスポット地図（Spot Map）^{2,3)}をモチーフしている。疫学確立の祖

【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@mtb.biglobe.ne.jp



<http://wwwn.cdc.gov/epiinfo/>

図1 米国CDC疫学解析ソフト
- 1849年のブロード・ストリートにおけるコレラ発生をモチーフ-

ジョン・スノーがコレラ菌の発見されるおよそ30年前の1849年にロンドンのブロードストリートに発生したコレラを詳細に記述し、原因を突き止め、コレラ対策を成功させた事例は余りにも有名である。これら一連の過程が疫学方法論の基礎となった。この疫学手法確立の当初から地域集積性の確認は重要な要素であった。しかし数値化された他の統計処理と異なり地理情報は視覚的なアナログ情報で集中しているとか分散しているとか主観的な判断に解釈を委ねるため長い間補完的な地位に甘んじてきたといえる。

b) 米国オハイオ州コロンバス市から始まった地域集積性の定量化

ジョン・スノーのスポット地図から100年後、疫学ではなく犯罪防止の視点からパット・モラン (Pat Moran) によって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⁽²⁾ の定量的評価が行われた。この研究で開発された指標がモランのI (Moran's I)¹⁾ であり相関係数: r と同様の解釈ができるという統計的な利便性が図られ、ようやく地域集積性の客観的評価法が確立した。しかしながらモランのIは地域集積性を定量化できたもののいったいどの地域 (ピクセル) が強く関連しているかは依然としてデータ値を四分位や標準偏差で数段階にカテゴリー化し色分けして示したコロプレス図 (Choroplethic Map 段彩地図)^{2,3)} 等に頼るしかなかった。この問題を解決したのがLuc AnselinでありIモランのIだけではわからない地域集積性の地図上の存在位置を検出す

る指標として局所空間統計量: LISA (Local Indicator(s) of Spatial Association)^{4,5)} 開発した。このためモランのIを global Moran、リュック・アンセリン (Luc Anselin) が開発したLISAを local Moranと呼ぶ場合がある。LISAによって地図区分のどこに有意な地域集積性があるのかの統計的根拠を示したことにより利便性が飛躍的に向上した。このLISAの用法は後述する。

(2) 解析ソフトウェア

ローレンツ曲線 (Lorenz curve) とジニ係数 (Gini coefficient)⁶⁾ の計算はエクセル統計2010を用いた。モランのIとLISAの計算と関連図表についてはLISA開発者のリュック・アンセリンが提供しているOpen-GeoDaシステム (2003年公開)^{Web(3)} を利用した。

(3) 変異係数 (変動係数)

a) 相対的な標準偏差として単位の違いを超えて

表1に変異係数 (変動係数: CV: coefficient of variance (or variation)) の計算式および使用にあたっての注意点を示した。本指標は単位が異なる指標のバラツキ度の比較や同一単位でも桁が異なる集団間のバラツキの比較に力を発揮する。そのため相対標準偏差 (RSD) ととも呼ばれる。第7報で紹介した南アジア7ヶ国の社会経済的指標⁷⁾ のような使用例に加えて、典型的な使用例を挙げると、馬場園⁸⁾ は一人当たり老人医療費の健康保険間格差が外来医療費や入院医療費ではなく千人当たり入院件数の違いによることを明らかにしている。また広田ら⁹⁾ は50代女性の20項目に分類した食品群別摂取状況を調査し、動物性油脂類、種実類、酒類の違いが大きいことを明らかにしている。Carolina. P. C.ら¹⁰⁾ は歯科矯正臨床でのセファログラム (頭部X線規格写真) 計測項目のANB, FMIA等の再現性比較をセファログラム、2D-CT、3D-CT間で行った結果、3D-CTが最も再現性が高いことを示している。

b) 変異係数の統計検定

なお、同一項目についての変異係数が時系列的

表1 変異係数（変動係数）

CV(coefficient of variance)

CV=100×σ:標準偏差／μ:平均値

注1) 100を掛けない場合もある

注2) 単位が異なる指標のバラツキ度の比較に有効(例 身長、体重、最高血圧のいずれのバラツキが大きいかの比較)

注3) 同一単位でも桁が異なる集団間のバラツキの比較(例 象の体重とネズミの体重のいずれのバラツキが大きいか)

注4) 標準化、偏差値は集団における個々の位置(順位)を判定するが変異係数は集団の分布の裾野の広がり进行比较する

注5) 比尺度にしか使用できない。すなわち絶対零点の無い間隔尺度では使えない(例 東北・北海道と九州の年間気温の比較は温度(摂氏)が間隔尺度であるため変異係数では比較できない)

注6) 平均値が0に近い場合、変異係数は大きく変動し易い欠点がある

に繰り返し計測され、それ自身が平均値と分散(標準偏差)を持つような調査¹¹⁾ の場合は対応のある t 検定等の有意性検定がされることがある。また一般に計測単位が異なる項目間の変異係数の有意性検定は行われないが複数集団間の変異係数の一致度を Wald test を用いて検定する方法^{Web04)} が提案されている。

(4) ローレンツ曲線とジニ係数

a) 不公平性(inequity)の視覚化と指標化

表2に仮想データを用いてローレンツ曲線⁶⁾ 作成のための基礎表を示した。ここで10人の仮想所得は昇順にソートされ、完全平等点および相対累積所得が計算されている。この表に基づいて図2に示すローレンツ曲線が作成された。更に図2はジニ係数との関係^{2,3,6)} を示しており、ジニ係数=Aの面積×2となる。ジニ係数を計算するには2法あり、第1法は表3に示すローレンツ曲線下の面積を多数の台形に見なして積分のような方法で幾何学的に面積を求める方法^{2,3,6)} であり、第2法は表4に示すデータ値全ての平均差から代数的に求める方法⁶⁾ である。ここで分散(標準偏差)とジニ係数の違いは、分散が個々の値の平均値から

の差の二乗(平方和)の総和を例数:n(またはn-1)で割って計算されるのに対し、ジニ係数はデータ値総当たりの2個づつの組合せの差の絶対値の総和を例数:nで割って計算される。所得に関してジニ係数が0の場合は完全平等社会、1の場合は完全不平等社会であるとされる。また、ローレンツ曲線は元来ジニ係数を計算するためにあるのではなく所得分布などの不公平性(inequity)を視覚に訴えて表現する方法である。ローレンツ曲線もジニ係数も特定の個人の所得や特定の個人間の所得格差は問題とされない。すなわちマクロ的な所得対策の指針(目安)として機能する。また時に複数のローレンツ曲線が交叉する場合があり、この場合は解釈が難しくなる。ローレンツ曲線に近似したものにパレート線(指数)があるがパレート・モデルとジニ・モデルのいずれが所得の実態を正確に表しマクロ的な所得対策に結びつくかには学術論争があり両者の特徴の違いは木村⁶⁾ の著書に詳しい。

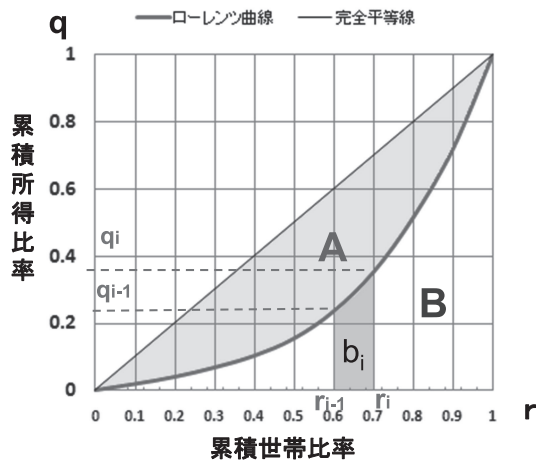
なお、ローレンツ曲線もジニ係数もそれらが捕らえている所得格差は平均値からの乖離度に基づく単なる個人差、世帯差ではなく不公平性である。なお、不公平性の概念はWHO¹²⁾ 等によって所得

表2 ローレンツ曲線作成の基礎表

No	所得	完全平等点	所得比率	相対累積所得
1	80	0.10	0.0186	0.0186
2	90	0.20	0.0210	0.0396
3	120	0.30	0.0280	0.0676
4	150	0.40	0.0350	0.1026
5	220	0.50	0.0513	0.1538
6	350	0.60	0.0816	0.2354
7	500	0.70	0.1166	0.3520
8	700	0.80	0.1632	0.5152
9	880	0.90	0.2051	0.7203
10	1200	1.00	0.2797	1.0000
Σ		4290		

注1) 所得は昇順にソートされている

注2) 完全平等点=No/対象者総数(均等間隔カテゴリ数)



計算法1:ローレンツ曲線からジニ係数を求める方法
 曲線下部の b_i を台形と見なし台形の面積から求める

$$B \text{ の面積} \div \sum_{i=1}^n b_i = \frac{1}{2} \sum_{i=1}^n (q_i + q_{i-1})(r_i - r_{i-1})$$

$$A \text{ の面積} \div \frac{1}{2} - \frac{1}{2} \sum_{i=1}^n (q_i + q_{i-1})(r_i - r_{i-1})$$

$$\text{Gini係数} = A \text{ の面積} \times 2$$

図2 ローレンツ曲線とジニ係数の関係

表3 ローレンツ曲線下の面積計算

所得	完全平等点	所得比率	相対累積所得	台形面積: b_i
80	0.10	0.0186	0.0186	0.0009
90	0.20	0.0210	0.0396	0.0029
120	0.30	0.0280	0.0676	0.0054
150	0.40	0.0350	0.1026	0.0085
220	0.50	0.0513	0.1538	0.0128
350	0.60	0.0816	0.2354	0.0195
500	0.70	0.1166	0.3520	0.0294
700	0.80	0.1632	0.5152	0.0434
880	0.90	0.2051	0.7203	0.0618
1200	1.00	0.2797	1.0000	0.0860
4290				
B:台形面積総和				0.2705
A:上弦部分面積				0.2295
ジニ係数($A \times 2$)				0.4590

表4 方法2：平均差からジニ係数を求める方法
ー個々の所得の全ての組合せの差の絶対値から求めるー

No	所得	80	90	120	150	220	350	500	700	880	1200	合 計
1	80	0	10	40	70	140	270	420	620	800	1120	4290
2	90	10	0	30	60	130	260	410	610	790	1110	3410
3	120	40	30	0	30	100	230	380	580	760	1080	3230
4	150	70	60	30	0	70	200	350	550	730	1050	3110
5	220	140	130	100	70	0	130	280	480	660	980	2970
6	350	270	260	230	200	130	0	150	350	530	850	2970
7	500	420	410	380	350	280	150	0	200	380	700	3270
8	700	620	610	580	550	480	350	200	0	180	500	4070
9	880	800	790	760	730	660	530	380	180	0	320	5150
10	1200	1120	1110	1080	1050	980	850	700	500	320	0	7710
合計		3490	3410	3230	3110	2970	2970	3270	4070	5150	7710	39380

M:所得平均(mean)=4290/10=429

MD:所得平均差(mean difference)=39380/10*10=393.8

ジニ係数(Gini coefficient)=MD/(2*M)=393.8/(2*429)=0.4590

に止まらず健康や寿命にも拡大して適用されるようになってきている。また地域差（regional differences）と地域格差（regional disparities, inequities）は日本語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、後者は単に疾病の差にとどまらず背景にある社会的不平等（Inequality）や不公平（inequity）をも視野に入れた差であり改善すべき政策課題¹³⁾のニュアンスがあるとされる。従ってローレンツ曲線、ジニ係数の本来の使い方からすれば地域格差の評価に用いるのが自然だと考えられる。

b) ジニ係数の統計検定

ジニ係数は不公平性を扱う社会学調査に多用されているにもかかわらずその統計的誤差はほとんど考慮されず、統計検定は通常行われない。その理由として石田ら¹⁴⁾によればジニ係数には中心極限定理を適用できないため正規分布を仮定した検定不可であり、かつ②サポートする統計ソフトパッケージが無い、ためとしている。その上でモンテカルロ法的一种であるノンパラメトリック（正規分布等の母数を想定しない）なブートストラップ（bootstrap）法によるジニ係数の有意差検定¹⁴⁾を提言している。

(5) モランの I（空間的自己相関指標）と LISA

a) モランの I 計算のパラメータ

表5にモランの I の計算式を示す。隣接するエリアとのデータ値の類似性は8方向で確認するクイーン法（チェスの駒の動きになぞらえて）が一般的で、接しているかないかの近接性（contiguity）を有り無しの0, 1パターンで評価する場合と距離の逆数等の距離を加味してウェイトを掛ける方法等がある。また解析の単位は市区町村のように地図の区分情報で行う場合や一定の距離で基盤の目状に区切られたメッシュで行う場合がある。解析目的と扱うデータ値によって条件を選定する。今回の事例は近接性を0, 1パターン、区分を市区町村で行う方法で行った。

LISAによるホットスポット・パターン分析^{4,5)}は有用で、ホット（コールド）スポット、外れ値を持つ疫学的意味を解釈することにより新たな疫学調査や保健医療対策の重要な指針を得ることができる。ホット（コールド）スポット、外れ値の評価は扱う項目によって大きく異なる。例えば感染症を対象とした場合は感染（+）エリアで囲まれたエリア、すなわちホットスポット（high-high）は最もリスクなエリアであるし、コールドスポット（low-low）はその逆である。一方、外れ

表5 Global Moran's I

空間的自己相関 (spatial autocorrelation) 指標

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

N: 対象エリア数(都道府県、市町村等)

Xi: i番目のデータ値

Xj: j番目のデータ値

X: 変数の平均値

Wij: 2点間を比較する場合の重み付け

バイナリー: 隣接していれば0, 離れていれば1

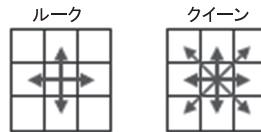
重心間距離の逆数: またはその2乗の逆数

時間距離: 交通手段を加味した時間的近接性による重み付

隣接関係: チェスの駒の動きになぞらえて

ルーク、ビショップ、クイーンという。

一般には8方向であるクイーンを用いる



モランのI指標は相関係数と同様に解釈できる

-1から+1の範囲をとり、空間的な位置関係が「似たもの同士が集まっている」場合は+1に近くなり、「似たもの同士が避け合っている」場合は-1に近づく。

検定法

1: 正規分布に近似したことを用いる検定

<http://giswin.geo.tsukuba.ac.jp/teacher/murayama/sdam/download.html>

空間データ分析マシン「SDAM」ユーザーズガイド p37-47

2: Permutation(組合せ法)による方法

<http://geodacenter.org/downloads/pdfs/geoda093.pdf>

空間データ分析ソフト「OpenGeoDa」ユーザーズガイド p90-92

値のうち高い値に囲まれた低値のエリア (low-high) は何らかの感染症伝搬が何らかの抑制要因が推定されるエリアである。その逆に低い値に囲まれた高値のエリア (high-low) は伝搬の初期を暗示させる。この手法を用いて新たな調査の仮説を絞り込む手法を Hot spot patterns 分析¹⁵⁾ という。global モランの 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。古文書はモラン島に宝があることを示している。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には LISA 地図が必須である。

3. 変異係数、ジニ係数、モランの I は地域差をどう捕らえているか

— 同一事例による比較検証 —

(1) 対象と地図情報化: 関東7都県277市区町村情報の shapefile 化

図3に解析対象地域である関東7都県277市区町村(2006年12月31日現在)を示す。地域集積性解析に用いる特殊な地図情報ファイルは shapefile^{Web05, 06)} と呼ばれ米国 ESRI 社が開発した地図情報処理のための GIS 標準データフォーマット形式の集合ファイルである。シェープファイルは3つのファイルからなり、解析したい属性情報を Dbase 形式で編集する。今回 ESRI ジャパンが提供している最新の全国市区町村 shapefile^{Web07)} から関東7都県277市区町村を抽出した。次に2005

表6 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) LISA分析カテゴリー値が正	high-high: 高値に囲まれた高値 エリア
2.cold spots(コールドスポット) LISA分析カテゴリー値が正	low-low: 低値に囲まれた低 値エリア
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)	

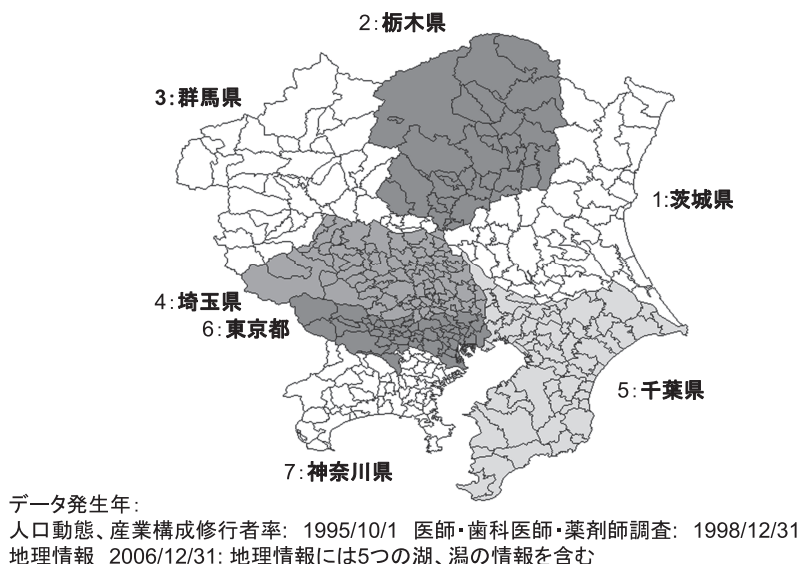


図3 解析対象地域：関東7都県の277市区町村

年の国勢調査^{Web08)}に基づく産業就業者割合、2008年厚生労働省 医師・歯科医師・薬剤師調査^{Web09)}と直近の2005年国勢調査から全国市区町村の就業施設別歯科医師数比を、国土地理協会提供の市区町村変遷情報（合併、分離等）^{Web10)}に基づきshapefileの2006年末の市区町村情報に適合させた。

(2) 変異係数による地域差検証

表7と図4に関東7都県全市区町村の関東7都県全市区町村の部門別産業別就業者割合（以下 産

業就業者割合）、勤務施設別歯科医師数人口10万比（以下 施設別歯科医師数比）の変異係数を示す。病院歯科勤務の歯科医師数の人口比（以下、病院歯科医師数比）の変異係数は952.93と明らかに抜きん出ている。事例の5つの指標のうち病院歯科医師数比は歯科大学付属病院勤務の歯科医師を含んでおり管内4都県の市区町村には9歯科大学（歯学部）が存在していることが影響している。ちなみに2008年12月31日現在、病院、行政、診療所、その他全ての歯科医師数人口比の最大値を示す自治体は歯科大学が集中し、居住人口が少な

表7 関東7都県全市区町村の産業別就業者割合、歯科医療機関分布の変異係数

	第1次産業	第2次産業	第3次産業	病院歯科	診療所歯科
n:市区町村数	377	377	377	377	377
mean: 平均	3.11	26.85	65.41	10.43	63.60
sd: 標準偏差	1.53	7.72	10.11	99.42	67.48
c.v: 変異係数	49.28	28.74	15.46	952.93	106.10

注1) 変異係数(変動係数):coefficient of variance (C.V)

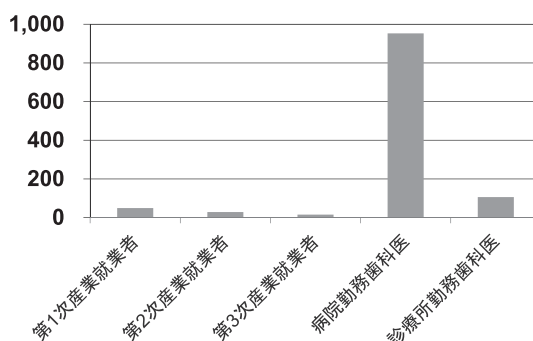
注2) $C.V = sd / mean \times 100$ 100を掛けない場合もあり

図4 関東7都県全市区町村の産業別就業者割合、歯科医療施設別歯科医師数人口10万比の変異係数

い千代田区(3,281人)が抜きん出て(全国一であり)、次が中央区(514人)、文京区(497人)、港区(371人)、坂戸市(331人)である。診療所勤務歯科医師数比(以下、診療所歯科医師数比)が106.10、産業別就業者割合では第1次産業が49.28と第2次、第3次と比較して大きい。総じてこれら5つの指標の中で病院歯科医師数比が最も相対的なバラツキが大きいことがわかる。このように変異係数は相対標準偏差であり、単位の異なる2種類の指標と単位が同じ指標間のバラツキの差を比較することができる。変異係数はこのような偏在の状況を比較する場合1つの指標で表す利便性の高い統計指標である。

(3) ローレンツ曲線とジニ係数による地域差検証

図5に産業就業者割合、図6に施設別歯科医師比のローレンツ曲線とジニ係数を示す。産業就業者割合については第1次産業のローレンツ曲線の上弦の弧の大きさが最大、ジニ係数も0.216と最大となり地域差が最も大きいことがわかる。これ

に対し図6に示す施設別歯科医師数比は診療所歯科医師数比が第1次産業の分布に近似しているが病院歯科医師数比がx軸：累積相対度数が右端にほとんど垂直に立ち上がっており、ジニ係数0.930と1に近く、ごく少数の自治体に存在していることが明白である。

(4) 地域集積性がある場合の地域差検証

a) 地域集積性の帰無仮説はランダム

図7に地域集積性の統計学的有意性の意味を示した。左図の関東7都県自治体マップは0, 1, 2, 3, 4の数値がランダムに配置された場合を示し、これに対し右のマップはに首都圏に4、その周辺に3を意図的に集中配置した場合で、いずれも仮想データである。図8にOpen-GeoDa^{Web03)}で計算したモランのIの値と検定結果を示している。ランダム配置の場合は図8, 9に示すようにモランの $I=0.0070$ ($p=0.3890$)、非ランダム配置の場合はモランの $I=0.5683$ ($p=0.0001$)となった。ここで用いた統計検定は組合せ検定(permutation)^{Web03)}で

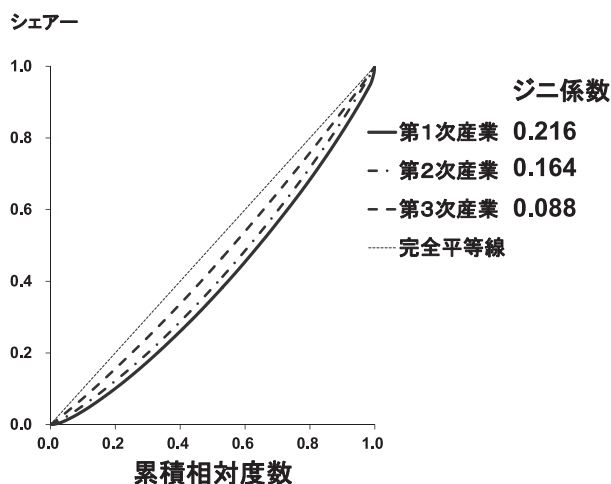


図5 関東7都県市区町村部門別産業就業者割合（データ：2005年国勢調査）

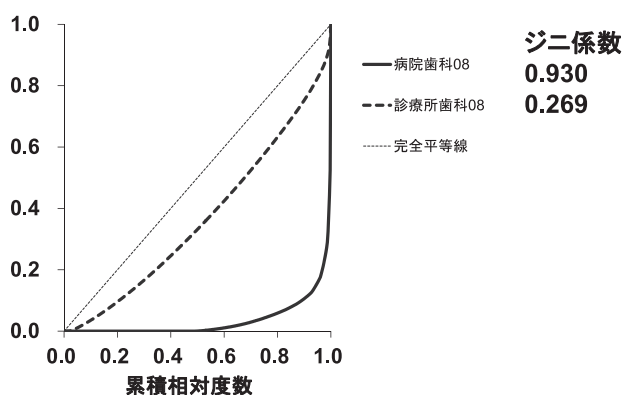


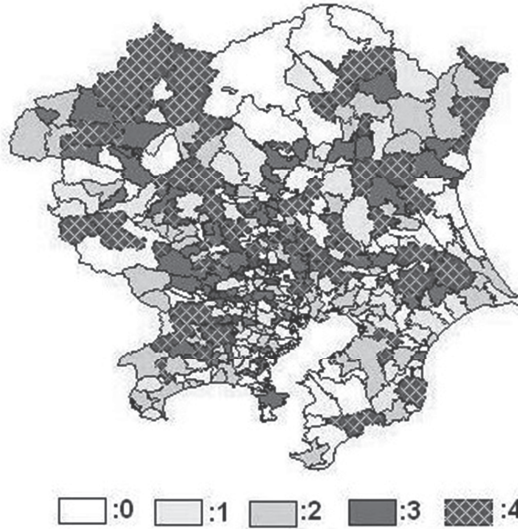
図6 関東7都県における勤務施設別歯科医師数10万比のローレンツ曲線とジニ係数
ー2008年医師・歯科医師・薬剤師調査ー

ある。一方、第7法で紹介したSDAMの場合は正規分布を仮定した検定である。組合せ検定はモンテカルロ法（Monte Carlo method）と違い非復元抽出（データ値の重複を許さない）方法で並べ替えを変えただけである。本法はFisherの直接確率法に近似しているが全ての組合せを行ったのではなくランダムに組合せを選択して指定回数だけ行った結果である。本事例では多数の理論的組み合わせからランダムに10,000回抽出して求めた。ここで、地域集積性を客観的、定量的に評価する手法（モランのIとLISA）の意義は図7の左図をみてどう解釈し対処するかにかかっている。例え

ば下記のような解釈と対処のうちどれが適切であろうか。

- ① カテゴリー4の分布はマップの中央付近で多数癒合しており地域集積性は高いようだ（と主観的に判断し、それを結論とする）
- ② カテゴリー1の分布がマップの周辺地区にたまっており地域集積性があるようだ（と主観的に判断し、それを結論とする）
- ③ いずれのカテゴリーもバラバラに存在しているようだ（と主観的に判断し、それを結論とする）
- ④ ①、②、③のいずれか判然としないので（あ

ランダム配置



非ランダム配置

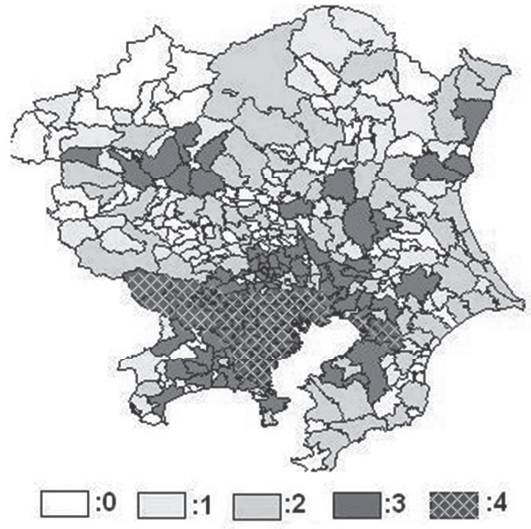
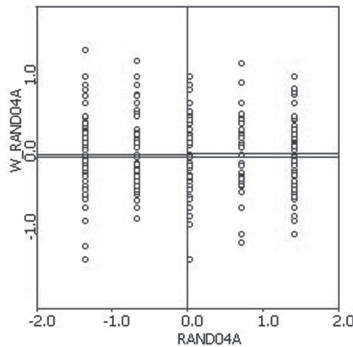


図7 ランダムと非ランダムのコロプレス図

注1) ランダム：277市区町村それぞれに0, 1, 2, 3, 4を無作為に配置 但し、5カ所の湖は0を配置
 注2) 非ランダム：島嶼を除く東京市区町村と近隣の市区町村に4を配置、その他の市に3を配置

ランダム配置

Moran's I=0.0070(p=0.3890)



非ランダム配置

Moran's I=0.5683 (p=0.0001)

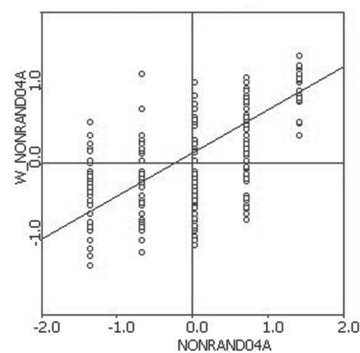


図8 ランダムと非ランダムの地域集積性 (uni-Moran's I)

るいは確信が持てないので) 地域集積性の解析を行う

このうち当然④が正解なのだが視覚的にわかりにくい、あるいは誤解しやすいのが地理情報の特徴である。この理由の一つとして複雑な模様をみると何らかのパターンを無意識に仮定してしまう

ためかも知れない。ランダムというのはバラバラだから同じカテゴリー繋がっていないというのは誤りでランダムの場合、各ピクセルの接している周囲のピクセルについて0, 1, 2, 3, 4が概ね同確率に存在するのだがそのようなことを判別するには人間の目が慣れていない。



図9 ランダム配置と非ランダム配置のpermutationによる有意差検定

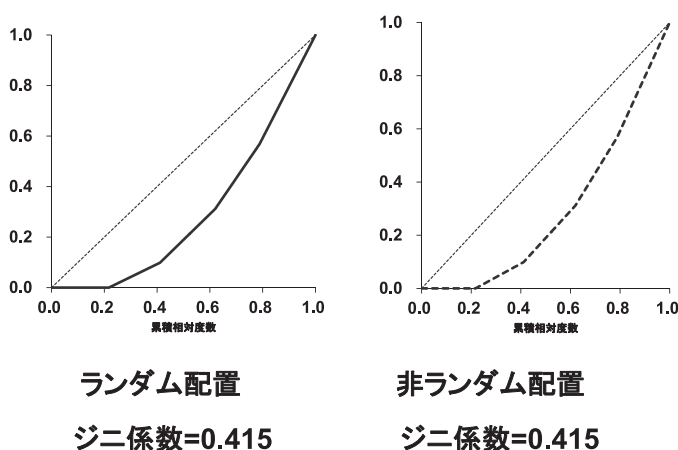


図10 ランダム配置と非ランダム配置によってローレンツ曲線とジニ係数は変わらない

b) 地域集積性の有無とローレンツ曲線・ジニ係数の関係

図10に上記c)の事例のローレンツ曲線とジニ係数を示す。ランダム配置と非ランダム配置でローレンツ曲線、ジニ係数とも全く同じである。すなわち同じ数値である場合は地図上の配置（並び方）が大きく異なってもローレンツ曲線、ジニ係数は変わらない。このことはローレンツ曲線やジニ係数で捕らえられる格差や不公平性は地域集積性と全く無関係ということを意味しない。なぜならローレンツ曲線が直線でジニ係数が0であればデータ値は全て同一値であるので地域集積性も無しとなる。しかし本事例で示したように一のローレンツ曲線やジニ係数が同一であっても地域集積性は固定されず様々な地理学的な分布が出現

する。このことは職種の違いや男女差に起因する所得の不公平性をマクロの視点で議論する際には地域集積性の差があっても余り問題にならないかも知れない。しかし感染症や環境汚染（大気、水、土壌）、犯罪の場合の地域集積性の有無は原因究明と対策上重大な意味を持つ。非感染症の場合も生活習慣、食習慣の背景要因との関係を明らかにする上で有効な情報を提供するであろう。

c) 部門別産業就業者割合の地域集積性

図11に第1次～3次産業就業者割合の地域集積性をLISA cluster mapで示す。図5に示したジニ係数では第1次、2次、3次の順でLISAがhigh-highとなる自治体で示される地域差が強いが地域集積性は第2次、第3次、第1次の順であった。特に第3次産業（サービス業）は首都圏に集中し、

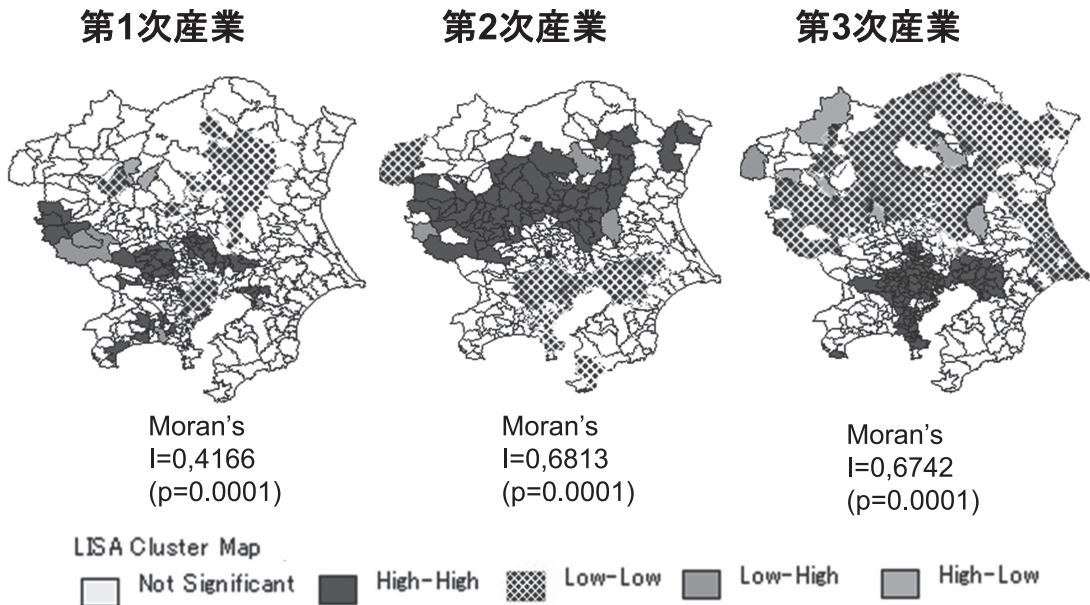


図11 関東7都県における部門別産業就業者割合のLISA cluster map - 2005年国勢調査 -

これに対し第2次産業（商工業）は首都圏の上部に横断的なベルト地帯を形成している。第1次産業（農業）は首都圏の近傍に分布している。これらは首都圏の機能とそれを支える周辺の産業構造を明確に示していると思われる。この事例のようにジニ係数で示される地域差情報と地域集積性の情報がそれぞれ大きく異なった情報を提示している場合がある。この理由はジニ係数が格差を示す情報であるが地理的情報は含まれていないのに対し、地域集積性の情報は地理的な情報から成り立ち隣接する周囲の自治体との類似性を判別しているためである。このように地域差を論議する場合は地域集積性の強さの情報処理が必要と考えられる。

d) 勤務施設別人口10万対歯科医師数比

図6に示したジニ係数では診療所歯科医師数比は第1次産業就業者割合に近似した地域差を示したが病院歯科医師数比は極めて高い地域差を示した。これに対し図12に勤務施設別歯科医師数人口10万比を病院歯科医師数比と診療所歯科医師数比（管理者を含む）に分けてLISA cluster mapで示

す。病院歯科医師数比についてはモランの $I=0.0756$ ($p=0.0015$)と有意であるが地域集積性は極めて低い。一方診療所歯科医師数比の場合はモランの $I=0.3577$ ($p=0.0001$)と地域集積性は高い。しかしながらLISA cluster mapは病院歯科がlow-lowの自治体が首都圏の外郭に点在し、診療所歯科医師数比の場合はベルト地帯を形成することなくhigh-high地域は首都圏のみに集中し、low-low地域は地図の片縁部分に散在していることがわかる。

e) 湖、潟の地理情報の扱い

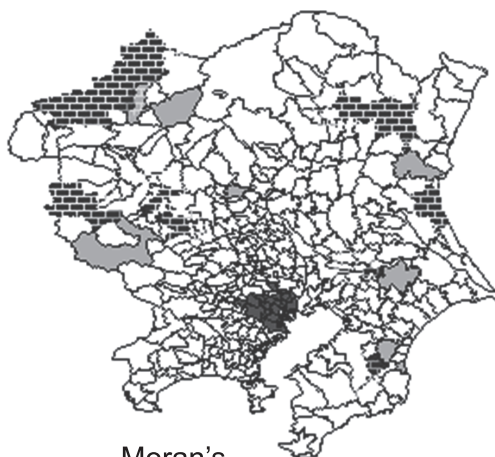
図13の左図は第2次産業就業者割合について関東7都県の5つの湖、潟（霞ヶ浦、芦ノ湖等）を含んだ場合のモランのIを示しており、湖、潟は最左端にy軸に平行に並んでいることがわかる。この理由はデータ値が0となっていることに起因している。一方、右図はこれら5つの湖、潟を除いた場合を示しており、この結果モランのIは0.5943から0.6813に上昇した。このように湖、潟は特定のデータ値を持ち得ない場合は地理情報ファイル（shapefile）から除去するのが望ましい。

病院歯科勤務歯科医師数 人口10万比



Moran's
 $I=0,0758$
($p=0.0015$)

診療所歯科勤務歯科医師数 人口10万比



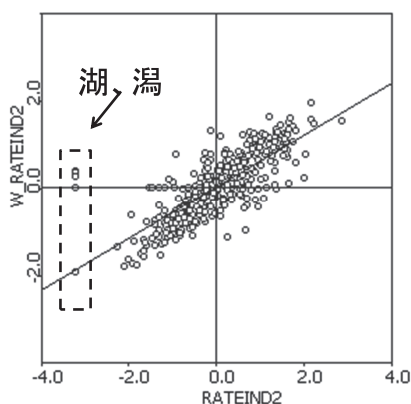
Moran's
 $I=0,3577$
($p=0.0001$)

LISA Cluster Map

Not Significant
 High-High
 Low-Low
 Low-High
 High-Low

図12 関東7都県における勤務施設別歯科医師数比 ー2008年医師・歯科医師・薬剤師調査ー

Moran' $I=0.5943$
 $p<0.0001$



Moran' $I=0.6813$
 $p<0.0001$

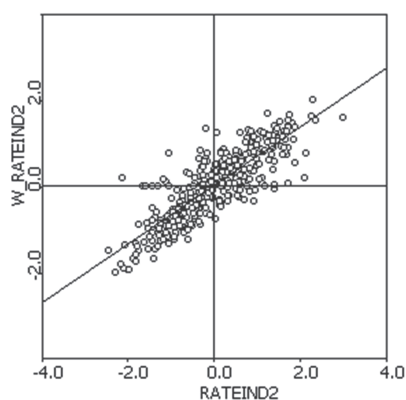


図13 Moran's Iと湖、潟の扱い 第二次産業就業者率

すなわち湖、潟を内海と考えて海と同じ扱いをすることに等しい。しかしながら例えば環境汚染、

特に水質汚染などの情報等の場合は湖、潟は重要な意味を持ち、特定のデータ値を持つので除去し

てはならない。

おわりに

「歯科疫学統計」シリーズの第7報¹⁾に続いて空間（地理）疫学を扱った。内容的には第7報との重なりが多いが地域差研究に地理学的かつ定量的な視点を加える場合の指標の適切使用を重点的に追加解説した。また今回 LISA を開発した Luc Anselin が公開している Open-GeoDa^{Web03)} システム（2003 年公開）の紹介も副次的な目的とした。ここ 10 年余の間の空間（地理）疫学研究の環境が急速に整ってきている。第7報で紹介した筑波大学開発の空間データ分析マシン SDAM（2004 年公開）^{Web11)}、第6報で紹介した統計 R^{Web12)} また米国 CDC の疫学・統計ソフト EpiInfo^{Web01)} それぞれで shapefile が扱える。いずれもフリーソフトとして公開されている Open-GeoDa, SDAM、統計 R、および EpiInfo をいわばネットサーフィンすることにより第7報、8報の事例のような解析が可能であることを紹介した。今後急速に進化し利便性が高くなるインターネット環境下で空間（地理）疫学へのチャレンジと関連の公開ソフトを有効に利用するための基礎として第7報および本報が役立てば幸いである。

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学会の顧問である重松逸造、蓑輪眞澄、高江洲義矩、境 脩、伊藤学而、佐々木英忠の各先生、ならびに空間疫学に関する共同研究を行っている神奈川歯科大学の平田幸夫、山本龍生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穂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。また分析ソフト SDAM を開発提供していただいている筑波大学、分析ソフト Open-GeoDa を開発提供いただいている米国 イリノイ大学 GeoDa center の Luc Anselin 先生に深謝します。さらに、貴重な全国市区町村マップの shapefile を提供い

ただいている ESRI Japan 社に深謝します。

参照 Web-site (2011/1/17 現在)

- Web01) 米国 CDC 疫学・統計ソフト EpiInfo のセットアップ+ユーザーガイド
<http://www.cdc.gov/epiinfo/>（2010 年 12 月 1 日アクセス）
- Web02) 米国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_aaap.pdf
（2009 年 10 月 10 日アクセス）
- Web03) Open-GeoDa セットアップ+ユーザーズガイド
<http://geodacenter.org/downloads/pdfs/geoda093.pdf>（2011 年 1 月 16 日アクセス）
- Web04) Test for Equality of Coefficient of Variation of K Populations
Dr. K Aruna Rao: Mangalore University Department of Statistics
Mangalagangothri 574 199 India, Fecib@blr. vnslnet.in
<http://isi.cbs.nl/iamamember/CD2/pdf/361.PDF>
（2011 年 1 月 16 日アクセス）
- Web05) Epi Info 日本語版 地図作製の手引き 岡山理科大学 山本英二
<http://zeus.mis.ous.ac.jp/EpiInfo/epi332jmap.pdf>
- Web06) shapefile の定義
<http://walrus.wr.usgs.gov/infobank/programs/html/definition/shapefile.html>（2011 年 1 月 16 日アクセス）
- Web07) shapefile ライブラリー（ESRI ジャパン公開全国市区町村界データ）
<http://www.okada.jp.org/RWiki/index.php?ShapeFile%A5%E9%A5%A4%A5%D6%A5%E9%A5%EA>
- Web08) e-stat: 政府統計の総合窓口 総務省国勢調査
<http://www.e-stat.go.jp/SG1/estat/GL02100104.do?tocd=00200521>（2011 年 1 月 16 日アクセス）
- Web09) 厚生労働省 医師・歯科医師・薬剤師調査
<http://www.mhlw.go.jp/toukei/list/33-20.html>
（2011 年 1 月 16 日アクセス）
- Web10) 市町村変更情報 国土地理協会 (JGDC)
<http://www.kokudo.or.jp/marge/change.html>
（2011 年 1 月 16 日アクセス）
- Web11) 空間データ分析マシン「SDAM」セットアップ+ユーザガイド
http://giswin.geo.tsukuba.ac.jp/teacher/murayama/sdam/doc/SDA_userguide.pdf

(2010年12月1日アクセス)

Web12) 統計Rのセットアップ+ユーザーガイド

<http://cwoweb2.baine.jp/~jgb11101/files/cart/cart.html>

(2010年12月1日アクセス)

文 献

- 1) 瀧口 徹：歯科疫学統計 ー第7報 空間（地理）疫学の基礎 その1ー, Health Science and Health Care, No 1 vol 9 : 19-33, 2009.
- 2) Miquel Porta : A DICTIONASRY of Epidemiology, 5th ed, OXFORD Univ press, New York, 2008, pp.40, 104, 234-235.
- 3) Miquel Porta編, 日本疫学会訳：疫学事典, (財)日本公衆衛生協会, 第5版, 2010, 50-51, 137-138, 303-304頁.
- 4) Anselin L: Local Indicators of Spacial Association -LISA, Geographical Analysis, 27 : 93-115, 1995.
- 5) Byron Moldofsky, Justin Ngan, Angela Colantonio: GIS and Cartography at the University of Toronto Technical Paper Series -Paper No. 3, Tronto Univ., Toronto, 1st ed., 2008, pp.25-25.
- 6) 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, 63-98, 99-103, 181-246頁.
- 7) 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.
- 8) 馬場園 明：老人医療費の健康保険組合間格差, J. Health.Sci, vol20 : 173-176, 1998.
- 9) 広田直子, 曾根良昭：炭水化物エネルギー比率の視点からみた中高齢者の栄養素摂取ならびに食品群別摂取状況, 生活科学研究誌, vol3 : 33-45, 2004.
- 10) Carolina Perez CouceiroI, Oswaldo de Vasconcellos Vilella: Dental Press J. Orthod. vol 15 (online article), 2D / 3D Cone-Beam CT images or conventional radiography : Which is more reliable?, 2010.
- 11) 佐藤佐和子, 吉田紀明：在宅パーキンソン病高齢者の歩行障害に対する聴覚的リズム刺激を用いた治療法（第2報）ーパーソナルテンポへの影響ー, vol 5, 211-220, 2009.
- 12) World Health Organization : The World Health Report 2002 Reducing Risks, Promoting Healthy Life 2002, WHO, Geneva, pp. 4-99.
- 13) 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, 37-51頁.
- 14) 石田 淳, 浜田 宏：仮想的機会調整による不平等分析ーブートストラップ法による機会調整前後のジニ係数の有意差検定ー, vol 20, 109-125, 2005.
- 15) Walter SD : Assessing spatial patterns in disease rates. Stat Med 19:20 : 1885-1894, 1993.

A review of oral epidemiological statistics

– Part VII : The bases of Spatial (Geographical) Epidemiology 2nd follow-up review –

Toru Takiguchi

(Fukai Institute of Health Science)

Key words : Regional disparity (inequity), Coefficient of variance (CV), Gini coefficient, Moran's I, Local indicators of spatial association (LISA)

Abstract: In this review, important differences of respective indices to verify and assess regional disparities /regional clustering are discussed following previous Part VII review.

Contents of this review:

<Tutorial overview>

- (1) Transition of analyses of regional clustering
- (2) Soft wares for calculation of regional clustering
- (3) Index 1: Coefficient of variance (CV)
- (4) Index 2&3: Lorenz curve and Gini coefficient
- (5) Index 4&5: Spatial autocorrelation statistics measures (Moran's I & LISA)

<Comparisons/Verifications of differences among Index 1-5 using common cases>

Each Index (1-5) was applied to geographic data of Kanto area composed of 277 municipalities in Japan using geographic files (spss data file, excel data file and shapefile).

Health Science and Health Care 10 (1) : 4 – 19, 2010