

GIS学会ハンズオン・セッション 空間統計

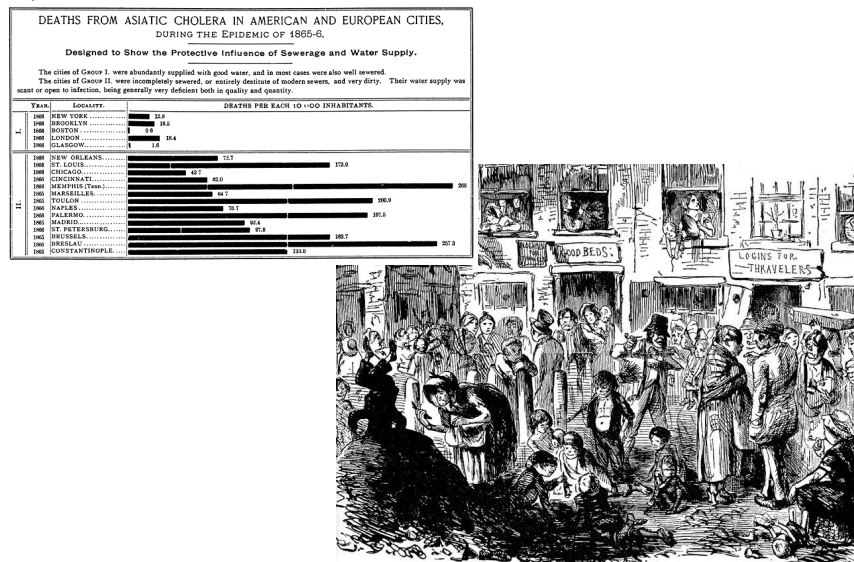
慶應義塾大学 古谷知之

本セッションの構成

- 講義(約50分)
- 演習(約50分)

内容

- 空間的自己相関(Moran's I)
- 内挿補間(Kriging)
- 空間計量経済モデル(Geographically weighted regression model)



上) http://www.sewerhistory.org/images/bm/bms9/1899_bms902.gif

下) <http://johnsnow.matrix.msu.edu/images/fullbanner6.jpg>

11.8.3-5

© T. Furutani, Keio SFC

3

John Snow's cholera map



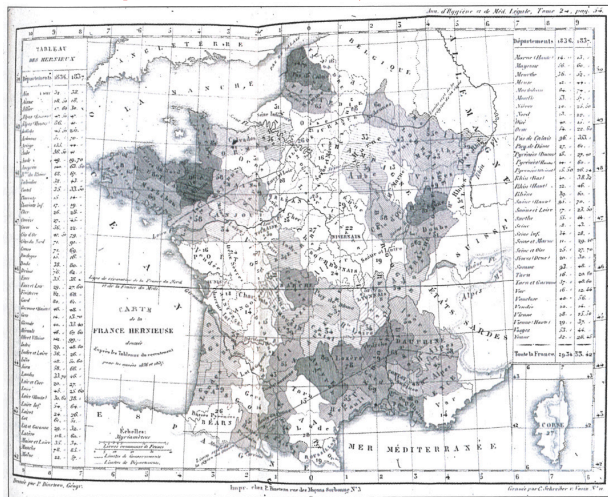
<http://upload.wikimedia.org/wikipedia/commons/2/27/Snow-cholera-map-1.jpg>

© T. Furutani, Keio SFC

11.8.3-5

4

ヘルニア罹患率(仏、1840年)



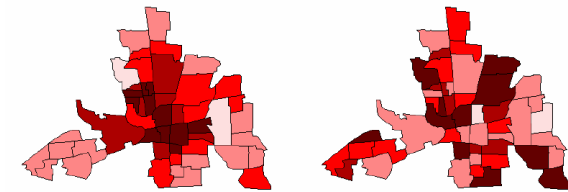
Original source: New York Academy of Medicine Library. Source: Koch (2005)
© T. Furutani, Keio SFC

11.8.3-5

5

空間的自己相関

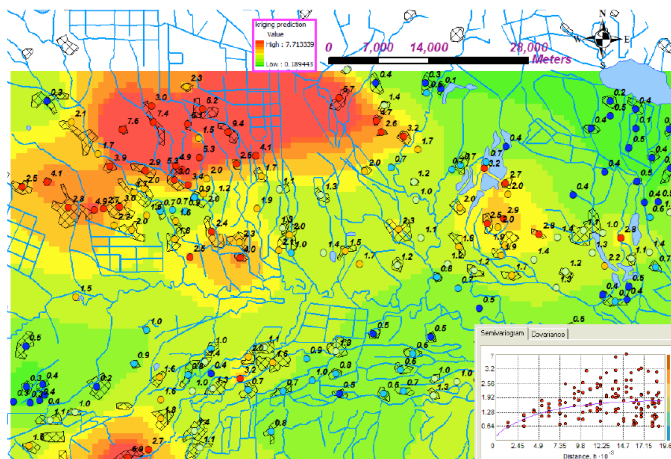
- 自己相関 (autocorrelation): 一つの変数に対する時間または空間で順序付けられた観測値間の相関
 - 時系列相関
- 空間的自己相関 (spatial autocorrelation)



同じような値がかたまっている＝
空間的自己相関

空間属性がランダム
に分布

チェルノブイリ原発事故によるCs137土壌汚染の空間補間分析例



Data courtesy of International Sakharov Environmental University, Minsk, Belarus; Institution of Radiation Safety, Belrad, Republic of Belarus.

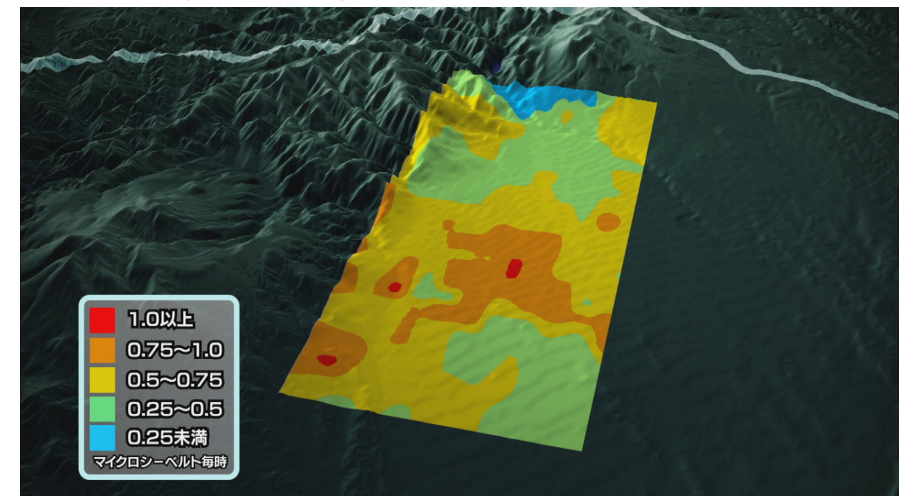
11/08/25

© T. Furutani, Keio SFC

Konstantine (2011)

7

放射線空間線量率の内挿補間



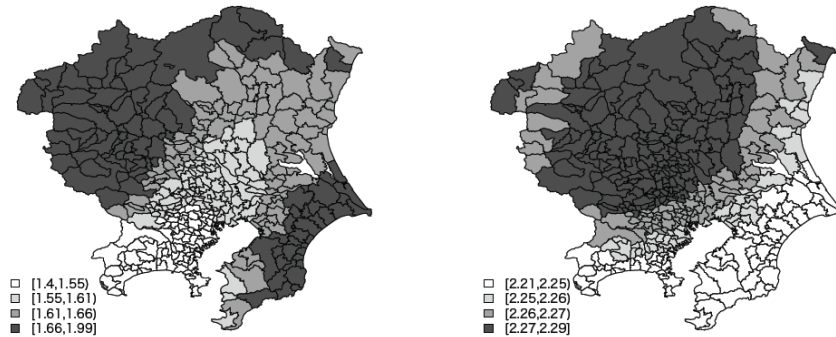
7月3日NHKスペシャル

11.8.3-5

© T. Furutani, Keio SFC

8

GWRの推定結果の例(地価モデル)



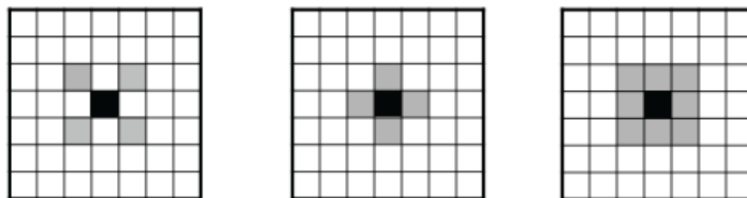
(a) 夜間人口密度

(b) 第三次産業従業人口密度

空間的な隣接関係や近さの表現

- ポリゴンデータの隣接関係→「空間隣接行列」
- ポイントデータやポリゴン中心点間の距離(直線距離、時間距離、など)→「距離行列」
- 隣接関係や距離関係をもとに空間オブジェクト間の地理的関係性を示す方法として「空間重み付け行列」などが用いられる

グリッドデータの隣接関係



(a) ルーク型

(b) エッジ型

(c) クイーン型

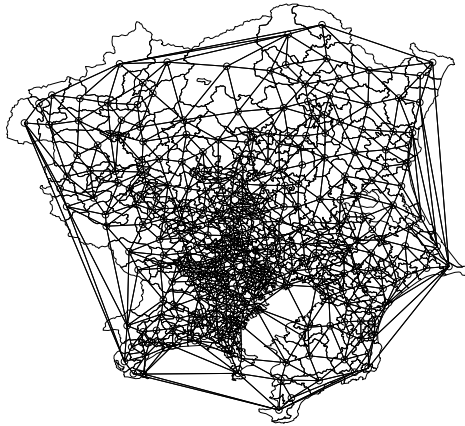
空間隣接行列 - Rook's case

neighborhood
空間オブジェクトの配置例

| | | |
|-------|-------|-------|
| x_1 | x_2 | x_3 |
| x_4 | x_5 | x_6 |
| x_7 | x_8 | x_9 |

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| x_1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| x_2 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| x_3 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| x_4 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| x_5 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| x_6 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| x_7 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| x_8 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| x_9 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

ドロネー三角網図



空間重み付け行列(1)

- 空間隣接行列の要素 c_{ij} に基づくもの

$$w_{ij} = \frac{c_{ij}}{\sum_{j=1}^N c_{ij}},$$

$$c_{ij} = \begin{cases} 1, & \text{if } i \text{ is linked to } j \\ 0, & \text{otherwise} \end{cases}$$

空間重み付け行列(2)

- 地点間の距離 d_{ij} の関数 $f(d_{ij})$ を用いた例

$$w_{ij} = f(d_{ij}), \text{ or } w_{ij} = \frac{w_{ij}}{\sum_{j=1}^N w_{ij}}, w_{ij} = \begin{cases} f(d_{ij}), & \text{if } i \text{ is linked to } j \\ 0, & \text{otherwise} \end{cases}$$

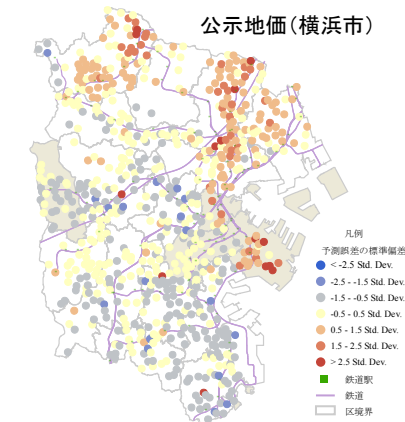
$f(d_{ij})$ の代表例:

1. $f(d_{ij}) = d_{ij}^{-\beta}$,
2. $f(d_{ij}) = \exp(-\beta d_{ij})$,

但し, $\beta > 0$, $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$, $w_{ii} = 0$.

空間的自己相関の例

地価モデル(OLS)の誤差項



空間的自己相関の計測

- 空間的自己相関を計測することで、**誤差項の空間的自己相関を伴った回帰モデル**を作成することができる
- どうやるか？
 - 空間的自己相関があるかどうかを検出
 - 空間的自己相関がある場合、どの範囲で相関関係が存在するのかが検出
 - 誤差項に空間的自己相関を考慮した

17

Moran's I

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

- ある地区属性の他の地区属性との共分散に基づく指標.
- 空間的に連続する点や面などのデータに適用.
- 相関係数と同様に、 $-1 \sim 1$ の間をとる
 - 同じような空間属性が配置されるほど正の相関 $=1$
 - 異なった空間属性が配置されるほど負の相関 $= -1$
 - 空間属性がランダムに分布するとき、0の値をとる

空間的自己相関の計測方法

- 空間的自己相関があるかどうかの計測
 - Moran's I (Moran coefficient)
 - Geary's C (Geary ratio)
- 空間的自己相関がどの範囲で存在するかの計測
 - セミバリオグラム (semi-variogram) モデル
 - 異方性モデル
- 空間的自己相関を考慮したモデル推定
 - 空間的自己相関モデル
 - 地理的加重回帰モデル

18

Moran's I

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

N : サンプル数, X_i, X_j : 任意の地区属性, \bar{X} : 地区属性の平均
 $W_{i,j}$: 二地点 $i-j$ 間の空間重み付け行列

- 接続行列 W_{ij} の設定方法のオプション
 - i と j が隣接している場合 $= 1$, その他 $= 0$
 - i, j 間の距離 d_{ij} の逆数 ($1/d_{ij}$)
 - より高次の隣接ペアも扱うことができる (第2回講義資料 p.17 参照)

Moran's I

- Moran統計量を用いて空間的自己相関が0であるという帰無仮説を検定するために、以下のZ統計量が用いられる。

$$Z(I) = \frac{I - E(I)}{S_E(I)}, I = \text{Moran's } I$$

$$S_E(I) = \sqrt{\frac{N^2 \sum_{ij} w_{ij}^2 + 3 \left(\sum_{ij} w_{ij} \right)^2 - N \sum_i \left(\sum_j w_{ij} \right)^2}{(N^2 - 1) \left(\sum_{ij} w_{ij} \right)^2}}$$

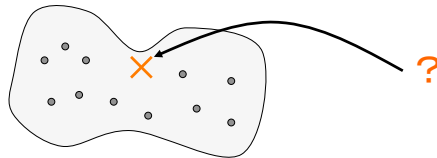
11.8.3-5

© T. Furutani, Keio SFC

21

空間の補間(空間内挿)

- 鉱脈を見つけるには？
- 幾つかの地点をボーリングし、サンプル地点で得られた鉱物含有量の分布と変化、標高などの結果を元に、鉱脈の空間的連続性を予測する。

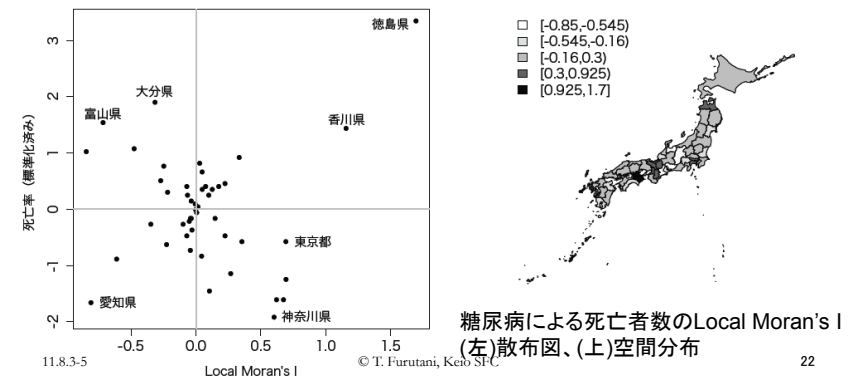


23

Local Moran's I

地区 i の Local Moran's I は次式から得られる。この指標は、Moran's I の地区 i に関する指標であり、ローカルな空間的自己相関を意味する。Moran's I 同様に、-1 から 1 の間の値をとる。

$$I_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2 / N} \sum_{j=1}^N w_{ij} (x_j - \bar{x})$$



11.8.3-5

© T. Furutani, Keio SFC

22

Spatial Interpolation of Radiation Risks

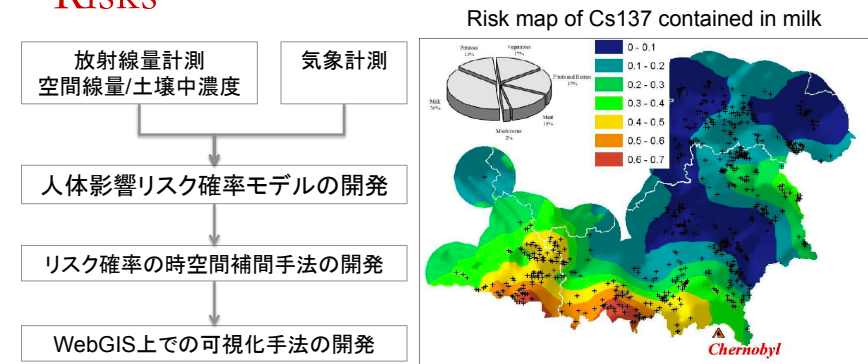


Figure 7. Probability that radiocesium milk contamination exceeded 75% of upper permissible level in southern Belarus in 1993. Disjunctive kriging was used to create this map.
<http://www.esri.com/library/whitepapers/pdfs/chernobyl-consequences.pdf>

© Scanning the Earth (Keio University and Safecast)

12/10/10

24

空間の補間(空間内挿)

- 限られたサンプリングデータから、より広範囲な空間の状況を把握するための技法.
- 代表的な手法
 - 離散型
 - TIN (Triangulated Irregular Network)
 - 逆距離加重法 (Inverse Distance Weighted: IDW)
 - 連続型
 - スプライン関数法
 - クリギング法

25

クリギング (Kriging) による空間予測

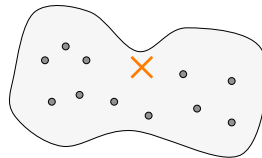
- 観測された空間属性に対する未知の空間属性の推定値 $Z^*(x_\alpha)$ を求めるのに、その**周辺**の**(n 個の)観測値 $Z(x_\alpha)$ の重みつき移動平均**を利用.
- $$Z^*(x_\alpha) = \sum_{\alpha=1}^n w_\alpha Z(x_\alpha), \quad \sum_{\alpha=1}^n w_\alpha = 1.$$
- 「重み w_α 」に観測値から得られた**セミバリオグラム**を利用.
 - 予測誤差を最小にするような「重み」の組み合わせを計算. **予測誤差** $\varepsilon = Z^*(x_\alpha) - \sum_{\alpha=1}^n w_\alpha Z(x_\alpha)$.

26

クリギング (Kriging)

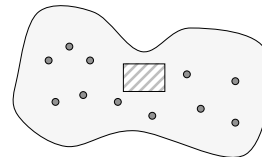
■ 通常クリギング

- ある地点における値を予測



■ 普遍クリギング

- ある領域における平均値を予測
- 通常クリギング + 地域トレンド



27

クリギングによる空間予測

3サンプル地点(1, 2, 3)から地点 p の予測例 (Davis, 1973)

$$\textcircled{1} \begin{cases} w_1\gamma(h_{11}) + w_2\gamma(h_{12}) + w_3\gamma(h_{13}) = \gamma(h_{1p}) \\ w_1\gamma(h_{21}) + w_2\gamma(h_{22}) + w_3\gamma(h_{23}) = \gamma(h_{2p}) \\ w_1\gamma(h_{31}) + w_2\gamma(h_{32}) + w_3\gamma(h_{33}) = \gamma(h_{3p}) \\ w_1 + w_2 + w_3 = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} w_1\gamma(h_{11}) + w_2\gamma(h_{12}) + w_3\gamma(h_{13}) + \lambda = \gamma(h_{1p}) \\ w_1\gamma(h_{21}) + w_2\gamma(h_{22}) + w_3\gamma(h_{23}) + \lambda = \gamma(h_{2p}) \\ w_1\gamma(h_{31}) + w_2\gamma(h_{32}) + w_3\gamma(h_{33}) + \lambda = \gamma(h_{3p}) \\ w_1 + w_2 + w_3 + 0 = 1 \end{cases}$$

λ : ラグランジュ乗数

28

クリギングによる空間予測

$$\begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \gamma(h_{13}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \gamma(h_{23}) & 1 \\ \gamma(h_{31}) & \gamma(h_{32}) & \gamma(h_{33}) & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(h_{1p}) \\ \gamma(h_{2p}) \\ \gamma(h_{3p}) \\ 1 \end{bmatrix}$$

これを解いて、 $w_{1..3}$ と λ を求める。

次に、①式に求めた値を代入し、予測地点 p の値を算出する。

併せて、予測誤差の標準偏差を次式により求める。

$$s_e^2 = \sqrt{w_1 \gamma(h_{1p}) + w_2 \gamma(h_{2p}) + w_3 \gamma(h_{3p})}$$

地点 p の予測値 $Z^*(x_p)$ は $\pm s_e^2$ の範囲で95%水準で統計的に信頼できる予測値となる。

29

セミバリオグラム

- 空間的自己相関が、どの距離にどれだけ関連を持つか、を測定する指標
- データ間の関連がサンプル地点間の相対的距離だけに依存する(方向や絶対位置は関係ない)と仮定。
- セミバリオグラム: 次式で定義されるセミバリエンスと2地点間の相対距離差 h との関係をプロットした図

$$\gamma(h) = \frac{1}{2N(h)} \sum (Z_i - Z_{i+h})^2$$

$h_{12} = Z_1 - Z_2$
 $h_{31} = Z_3 - Z_1$
 $h_{23} = Z_2 - Z_3$

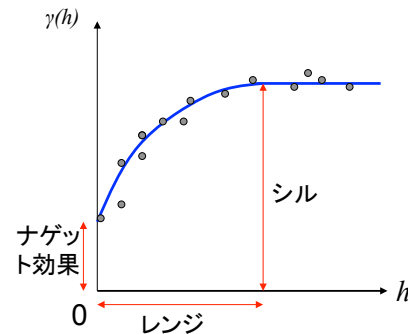
但し、 Z_i は地点 i の属性、 h は地点間の相対距離、 $N(h)$ はサンプル地点の組み合わせの数、を意味する。

30

セミバリオグラム

■ (セミ)バリオグラムのパラメータ

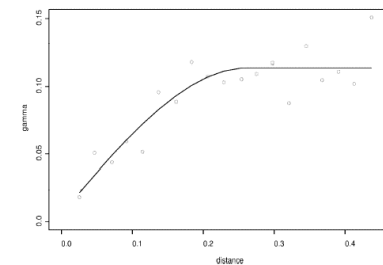
- ナゲット (nugget) 効果: $\gamma(h)$ の $h=0$ のときの値
- シル (sill) : $h \rightarrow \infty$ にしたときの $\gamma(h)$
- レンジ (range) : 自己相関がなくなる距離



重み付き最小二乗法 (weighted least squares) や最尤推定法などにより推定される。

31

pH値の分布とセミバリオグラム



猫本 (2002)

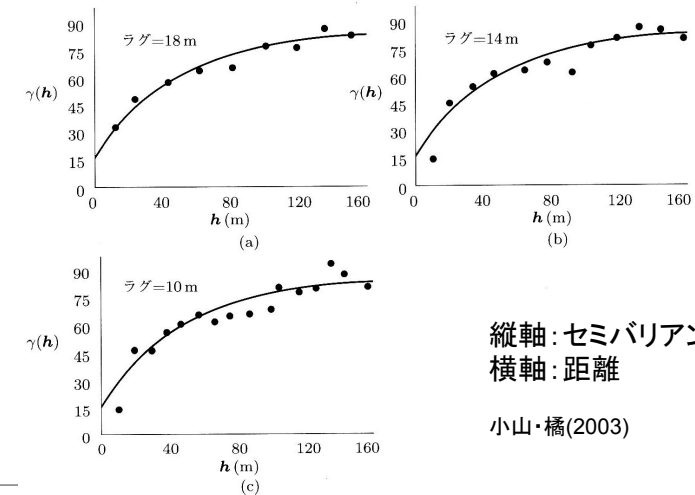
32

セミバリオグラム

- 考慮すべき点
- サンプル間の相対距離 h をどのように選ぶか
- サンプル間の間隔(=空間ラグ)をどのように選ぶか
- 空間ラグと距離 h の影響を試行錯誤的に検討するため、何種類の h に対して計算を行うか

33

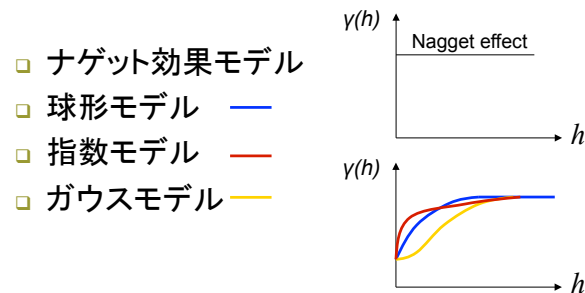
異なる空間ラグの影響



34

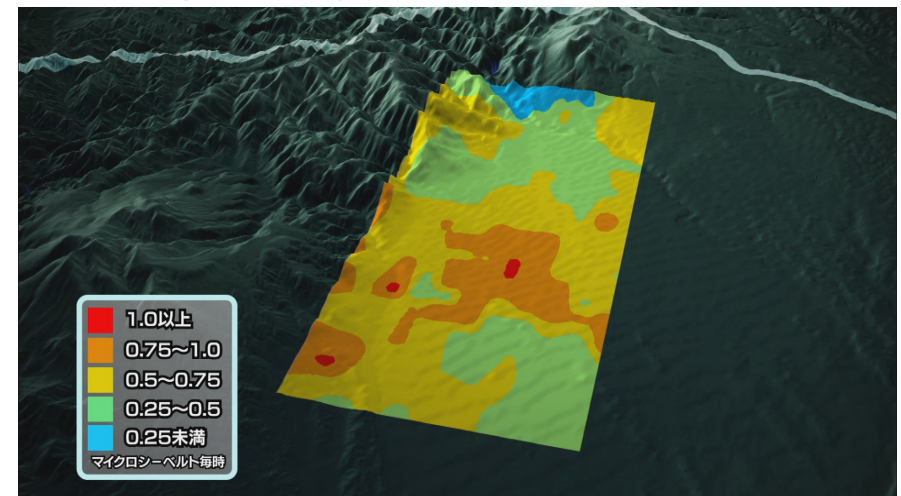
セミバリオグラム雲

- セミバリオグラムは、セミバリオグラムをプロットしたもの(セミバリオグラム雲)の形状によって様々な理論分布モデルを当てはめることができる。



35

放射線空間線量率の内挿補間



7月3日NHKスペシャル

11.8.3-5

© T. Furutani, Kcio SFC

36

回帰モデル誤差項の空間的自己相

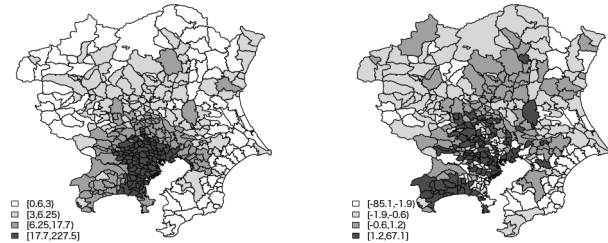


図 10.1 市区町村別地価データ分布 図 10.2 地価モデル (OLS) の誤差項分布

表 10.1 線形回帰モデルの推定結果 (最小二乗法)

| 変数 | 回帰係数 | t 値 |
|--------------------------|-------|-------|
| 定数項 | 2.55 | 4.50 |
| 夜間人口密度 | 1.68 | 15.85 |
| 第三次産業従業人口密度 | 2.25 | 28.65 |
| 自由度修正済み R ² 値 | 0.828 | |

1

37

空間的自己相関モデル

- 誤差項の空間的自己相関を考慮した回帰モデル
- 空間データを地域トレンドと空間的自己相関をもつ誤差項とに分離

$$Y = X\beta + u = \text{地域トレンド} + \text{空間的自己相関誤差}$$

$$= X\beta + (\rho Wu + \varepsilon) \quad \text{空間的自己相関をもつ誤差と持たない誤差とに分解}$$

$$= X\beta + \rho W(Y - X\beta) + \varepsilon$$

$$\Leftrightarrow Y = X\beta + (I - \rho W)^{-1} \varepsilon$$

11.8.3-5

© T. Furutani, Keio SFC

38

空間的自己相関モデル

- SACモデル(一般形)
 $y = \rho W_1 y + X\beta + u, u = \lambda W_2 u + \varepsilon, \varepsilon \sim N(0, \sigma^2 I_n)$
- SARモデル: $y = \rho W_1 y + X\beta + \varepsilon, \varepsilon \sim N(0, \delta^2 I_n)$
- SEMモデル: $y = X\beta + u, u = \lambda W_2 u + \varepsilon, \varepsilon \sim N(0, \delta^2 I_n)$
- Spatial Durbin モデル:
 $(I_n - \rho W)y = (I_n - \rho W)X\beta + \varepsilon$
 $y = \rho W y + X\beta + \rho W X\beta + \varepsilon$
 $\varepsilon \sim N(0, \delta^2 I_n)$

11.8.3-5

© T. Furutani, Keio SFC

39

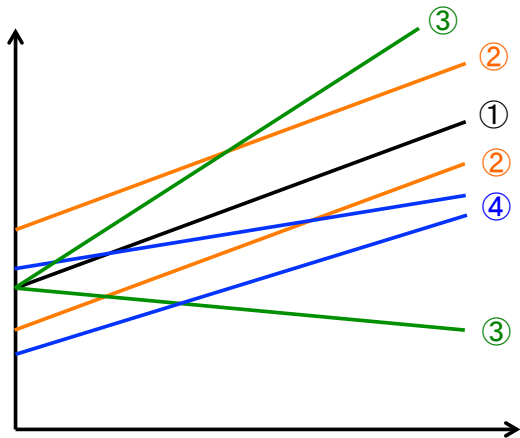
マルチレベル・モデル

- パラメータを個人・グループごとに推定する手法
→ 個人や地域の異質性を明示的に考慮

- 固定効果とランダム効果を持つ混合効果モデル

- (1) 定数項と回帰係数が変動しないモデル: $y_{ij} = \beta_0 + \beta_1 x_{ij1} + \dots + \beta_k x_{ijk} + \varepsilon_i$
- (2) 定数項が変動するモデル: $y_{ij} = \beta_{j0} + \beta_1 x_{ij1} + \dots + \beta_k x_{ijk} + \varepsilon_i$
- (3) 回帰係数が変動するモデル: $y_{ij} = \beta_0 + \beta_{j1} x_{ij1} + \dots + \beta_{jk} x_{ijk} + \varepsilon_i$
- (4) 定数項と回帰係数が変動するモデル: $y_{ij} = \beta_{j0} + \beta_{j1} x_{ij1} + \dots + \beta_{jk} x_{ijk} + \varepsilon_i$

マルチレベルモデル



11.8.3-5

© T. Furutani, Keio SFC

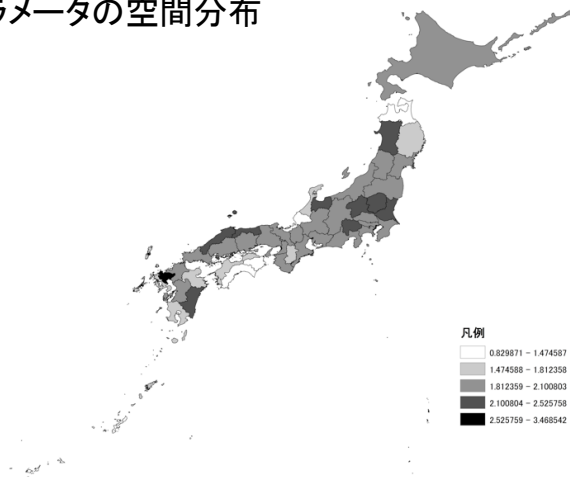
41

固定効果モデルとランダム効果モデル

- 固定効果モデル
 - グループや個人ごとに回帰係数のパラメータのばらつきを認めない
- ランダム効果モデル
 - グループや個人ごとに回帰係数のパラメータのばらつきを認めるモデル
- 混合効果モデル
 - 固定効果とランダム効果が混在しているモデル

マルチレベルモデルの推定例

- 交通発生モデルのアクセシビリティ指標に対するパラメータの空間分布



凡例
0.829871 ~ 1.474587
1.474588 ~ 1.812358
1.812359 ~ 2.100803
2.100804 ~ 2.525758
2.525759 ~ 3.468542

地理的加重回帰モデル

geographically weighted regression (GWR)

$$\begin{cases} W_i y = W_i X \beta + \varepsilon, \\ \varepsilon \sim N(0, \sigma^2 I_n) \end{cases}$$

- パラメータをサンプル毎に与えるマルチレベルモデル

11.8.3-5

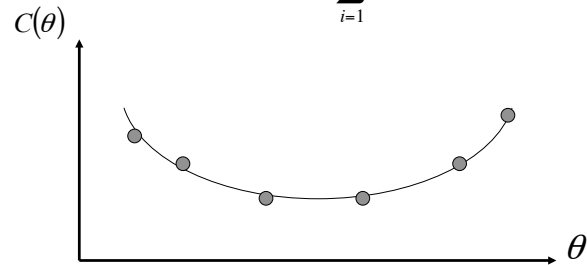
© T. Furutani, Keio SFC

44

バンド幅

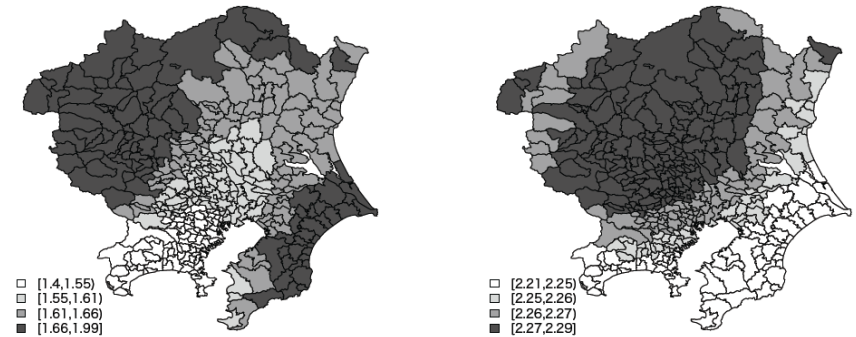
- 以下のcross validation scoreが最小になるような θ を用いる (Paez et al, 2002など)

$$C(\theta) = \sum_{i=1}^n (y_i - y_{\hat{i}})^2$$



45

GWRの推定結果の例(地価モデル)



(a) 夜間人口密度

(b) 第三次産業従業人口密度

11.8.3-5

© T. Furutani, Keio SFC

46