空間情報を用いた社会・経済分析(第4回)

統計数理研究所　村上大輔

dmuraka@ism.ac.jp

担当回（前半）

内容

- 第2回(4/21 月) ：空間データの処理・地図化
- 第3回(4/28 月) ：探索的空間データ解析
- 第4回(5/8, 木) ：空間計量経済モデルと応用

計算コード：

https://github.com/dmuraka/HIAS_class

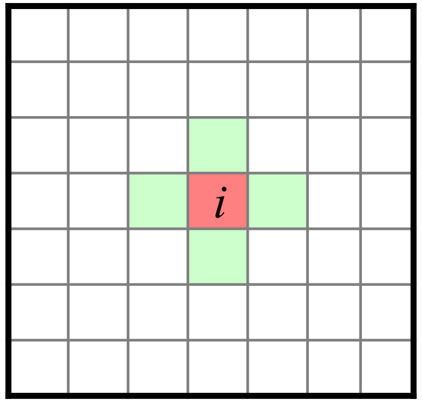
- 質問等は村上(dmuraka@ism.ac.jp)までご連絡ください

復習(1/2)：モランI統計量　≒自分と近所との**相関係数**

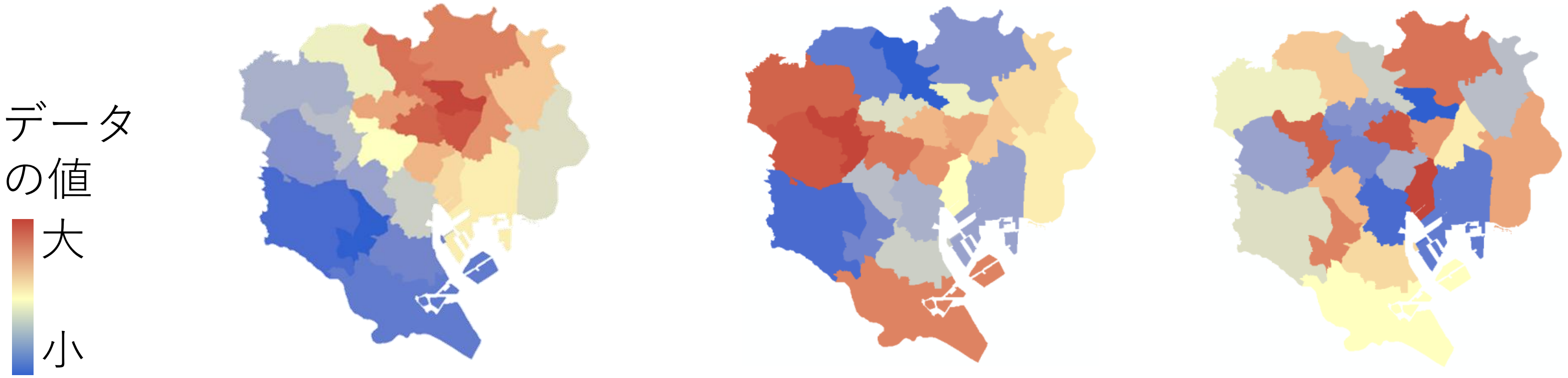
近隣は平均以上？

自分は平均以上？

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_j - \bar{y})(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$



y_i: 地域iの観測値　$\bar{y}$: 標本平均　n:標本サイズ　w_{ij}: 地域jの空間重み（例: 近隣1それ以外0）



I =0.78　　$I \approx 0.00$　　$I = -0.28$

復習(2/2)：ローカルモランI統計量とモラン散布図

モランI統計量

$$I \rightarrow \frac{1}{n}\sum_{i=1}^{n} I_i$$

地域毎に分解

ローカルモランI統計量

$$I_i = \frac{1}{m}(y_i - \bar{y})\sum_{j=1}^{J} w_{ij}(y_j - \bar{y})$$

↑

mは基準化項(重要でない)

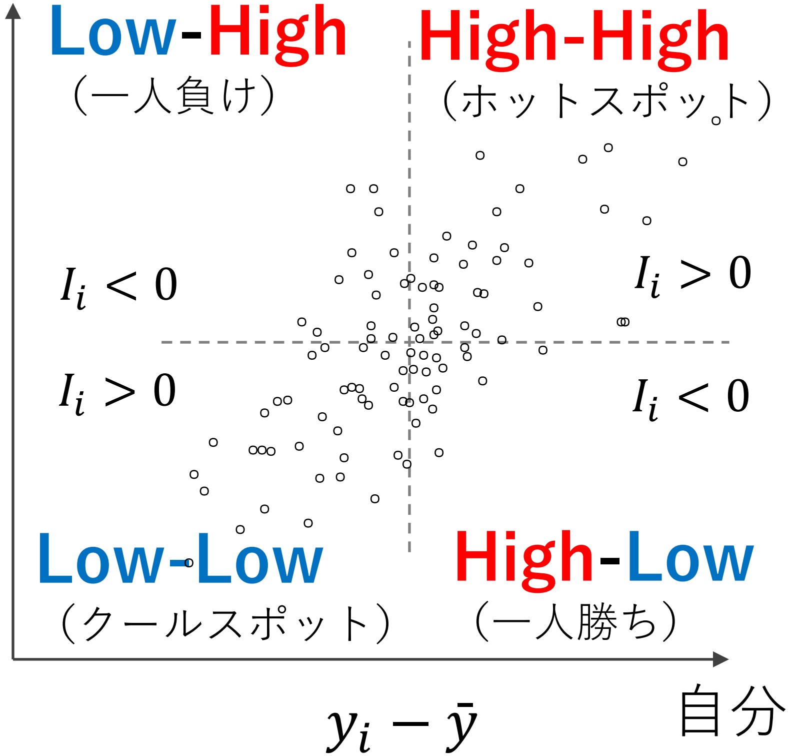
場合分け

モラン散布図

$\sum_{j=1}^{J} w_{ij}(y_j - \bar{y})$

行基準化($\sum_{j=1}^{J} w_{ij} = 1$; 後述)されている場合

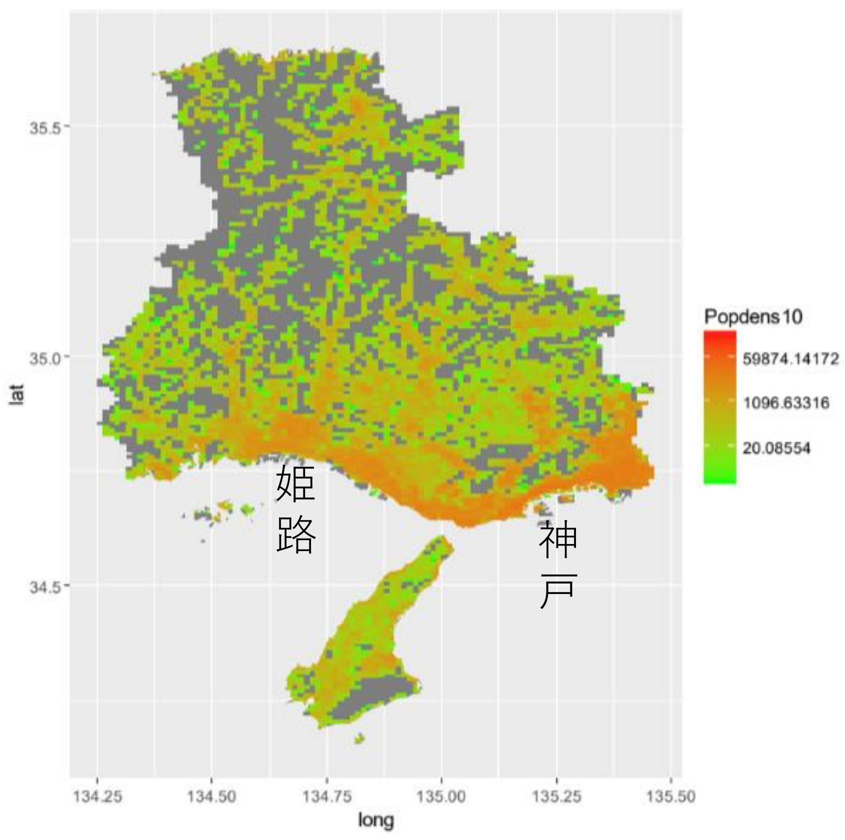
$\sum_{j=1}^{J} w_{ij} y_j - \bar{y}$



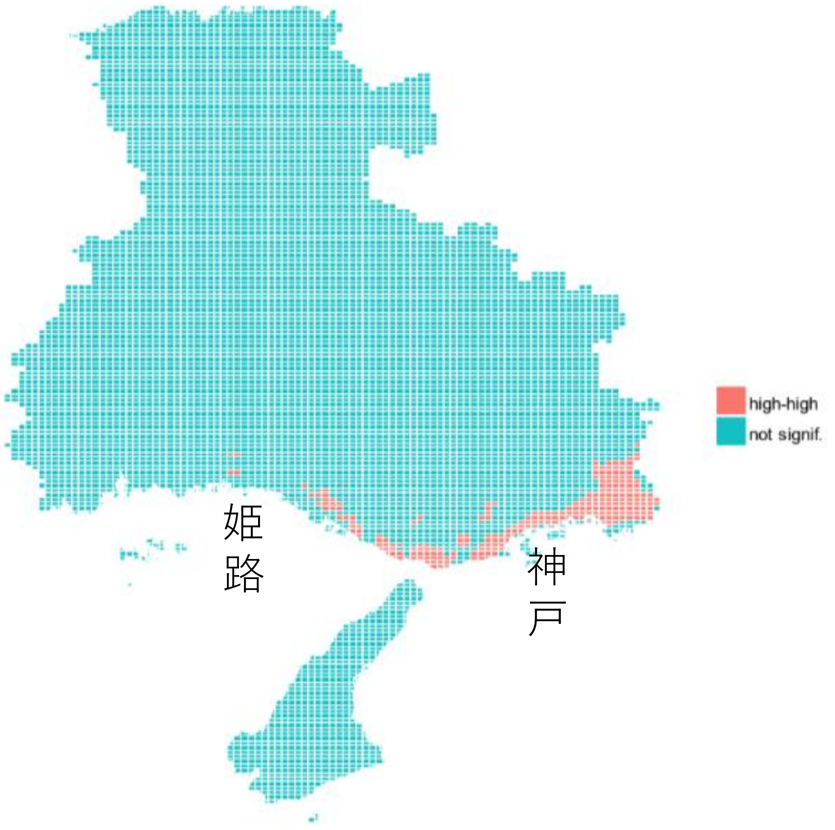
モランI統計量の特徴

瀬谷創准教授(神戸大)
よりお借りしました

- 標本平均$\bar{y}$との比較に基づくため、局所的な傾向に対する感度が低い



人口密度(兵庫県)



モラン散布図

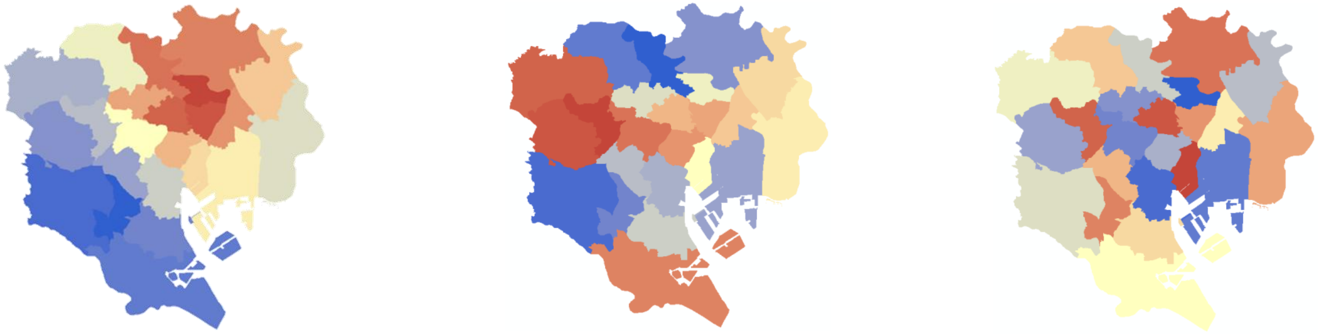
ギアリーC統計量

自分と近隣の差

$$C = \frac{n-1}{2\sum_i \sum_j w_{ij}} \frac{\overbrace{\sum_i \sum_j w_{ij}(y_i - y_j)^2}}{\sum_i (y_i - \bar{y})^2}$$

y_i: i番目の標本
$\bar{y}$: 標本平均
n:標本数
w_{ij}:空間重み(例: 近隣は1それ以外0)

正の空間相関 $C < 1$　　空間相関なし $C \approx 1$　　負の空間相関 $C > 1$



ローカル・ギアリーC統計量

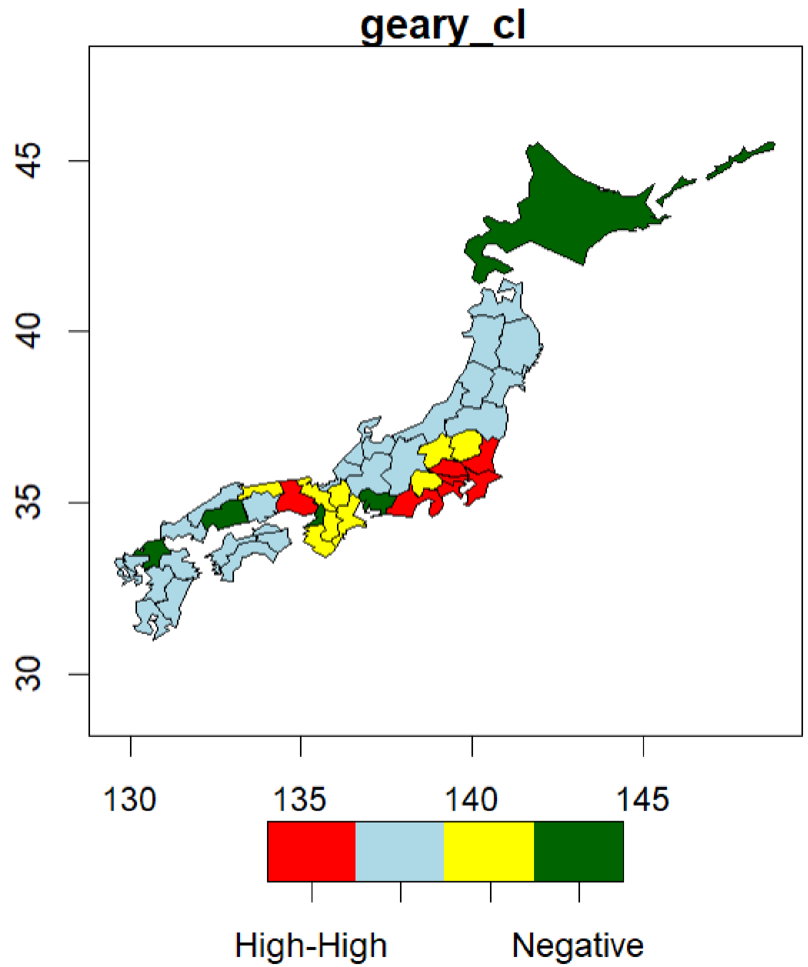
$$C_i = \sum_j w_{ij}(y_i - y_j)^2$$

ローカル・ギアリーC統計量に基づく分類図

```
> lgeary0  <- localC_perm(pop, listw=w)
> geary_cl <-attr(lgeary0,"cluster")
```
: ローカルGeary's C
: 分類結果

```
> pref$geary_cl<- geary_cl
> levels(geary_cl)
[1] "High-High"      "Low-Low"         "Other Positive" "Negative"
```
: 分類結果を都道府県別の新たな列(geary_cl)に追加
: 分類の確認

```
> plot(pref[,"geary_cl"],
+       axes=TRUE,
+       pal = c("red",
+               "light blue",
+               "yellow",
+               "dark green"),
+       key.pos=1,
+       key.length=0.5)
```
: High-Hiwh :赤
: Low-Low :薄青
: Other.Pos. :黄色
: Negative :濃緑

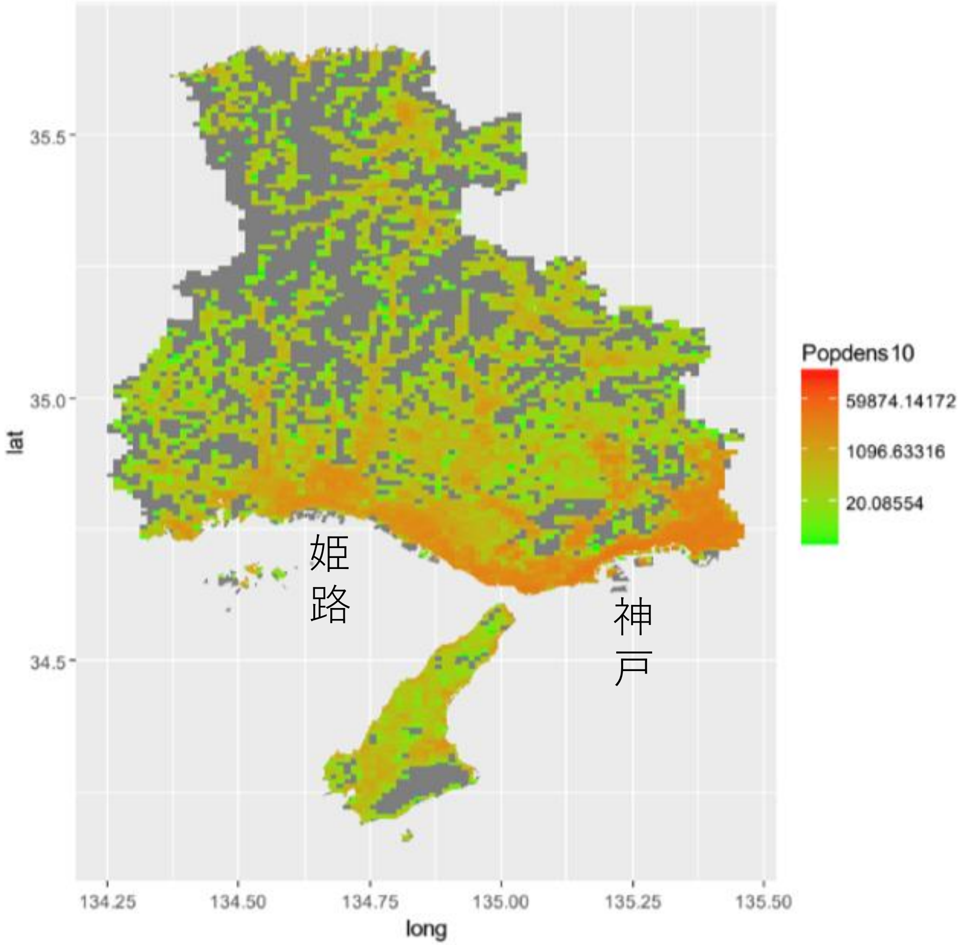


ギアリーCに基づく4分類

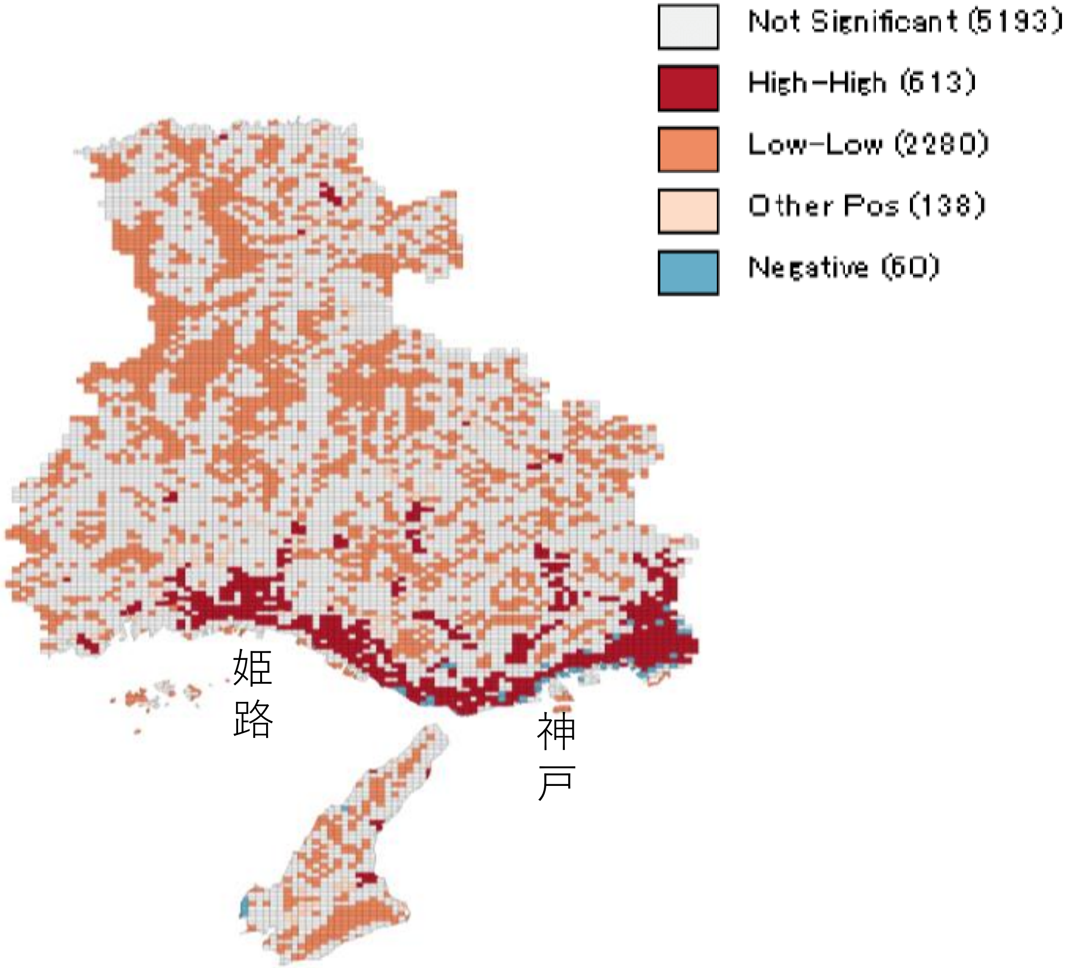
High-High　: 自分も隣も高い
Low-Low　: 自分も隣も低い
Other positive: 隣接と類似（上記以外）
C < 1
Negative　: 隣接と逆の傾向
C > 1

ギアリCの方が局所的傾向が捉えられる

※ただし、個人的経験としては解釈しにくい結果が得れる場合もあり、
　モランIの方が解釈性は優れている印象



人口密度(兵庫県)



ローカルギアリーCに基づく分類

前回分のまとめ

- 空間データの基本的な処理・作図方法を紹介

- spdepパッケージを用いた探索的な空間分析の方法を説明
  - 大域空間統計量
    - モランI , ギアリーC,…
  - 局所空間統計量
    - ローカルモランI, ローカルギアリーC, Getis/Ord G*, LOSH統計量,、…

  - データの特性を把握するために広く用いられている。

担当回（前半）

内容

- 第2回(4/21 月)　：空間データの処理・地図化
- 第3回(4/28 月)　：探索的空間データ解析
- 第4回(5/8, 木)　：空間計量経済モデルと応用

資料は以下にもアップしています：

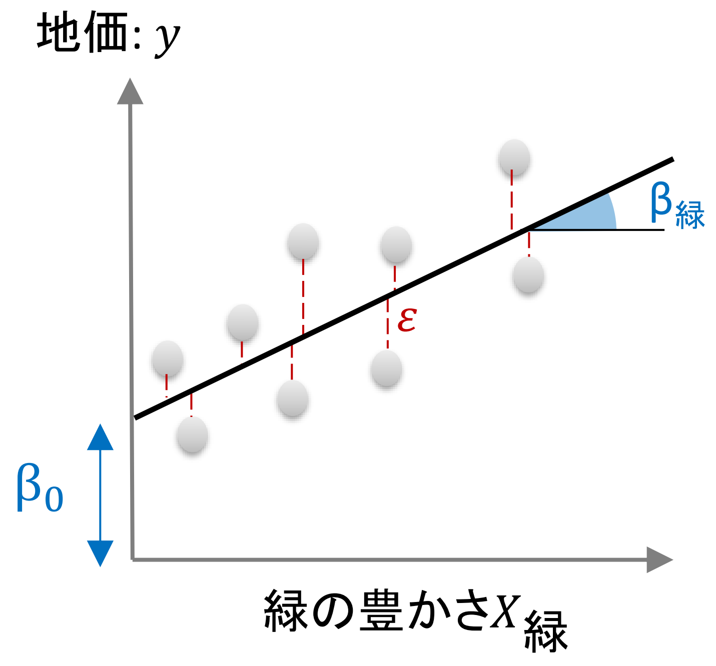
https://github.com/dmuraka/HIAS_class

- 質問等は村上(dmuraka@ism.ac.jp)までご連絡ください

線形回帰モデル

- 例えば住宅地価の要因分析に用いられる

$$y = \beta_0 + x_{駅}\beta_{駅} + x_{大学}\beta_{大学} + x_{緑}\beta_{緑} + \cdots + \varepsilon$$



説明変数

$x_{駅}$: 駅までの距離 — $\beta_{駅}$

$x_{大学}$:大学までの距離 — $\beta_{大学}$

$x_{緑}$:緑の豊かさ — $\beta_{緑}$

⋮

誤差項

ε:説明されない要因

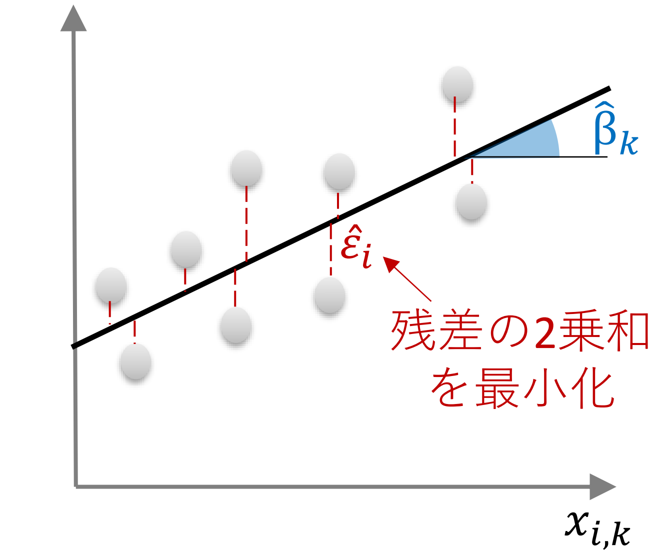
→ y: 地価

被説明変数

回帰係数$\beta_{駅}$, $\beta_{大学}$, $\beta_{緑}$の
推定が主な使用目的

線形回帰モデルの推定：まとめ

$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_k + \varepsilon_i \qquad E[\varepsilon_i] = 0 \quad V[\varepsilon_i] = \sigma^2 \quad Cov[\varepsilon_i, \varepsilon_j] = 0$$



β_kは通常最小2乗法(Ordinary least squares: OLS)で推定

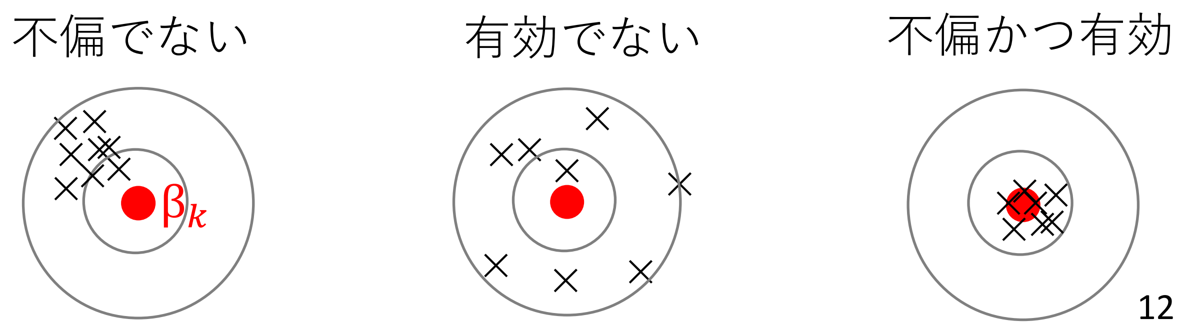
– 残差$\hat{\varepsilon}_i = y_i - \sum_{k=1}^{K} x_{i,k}\hat{\beta}_k$の2乗和の最小化により$\hat{\beta}_k$(OLS推定量)を与える

以下を満たす場合、OLS推定量は不偏性と有効性を満たす (Best Linear unbiased estimator: BLUE)

– 説明変数は非確率変数（自動的に満足）
– 説明変数と誤差項は直交 ($\sum_{i=1}^{N} x_{i,k}\varepsilon_i = 0$)
– 誤差項の期待値は0　: $E[\varepsilon_i] = 0$
– 誤差項の分散は一定　: $V[\varepsilon_i] = \sigma^2$
– 誤差項間の共分散(相関係数)は0

: $Cov[\varepsilon_i, \varepsilon_j] = 0$

不偏性: $E[\hat{\beta}_k] = \beta_k$
有効性: $\mathrm{Var}[\hat{\beta}_k] \leq \mathrm{Var}[\tilde{\beta}_k]$ ($\tilde{\beta}_k$は他の(線形)不偏推定量)



空間データに対して線形回帰モデルを使ってよいか

例: つくばエキスプレス沿線の住宅地価分析(Tsutsumi and Seya, 2008)

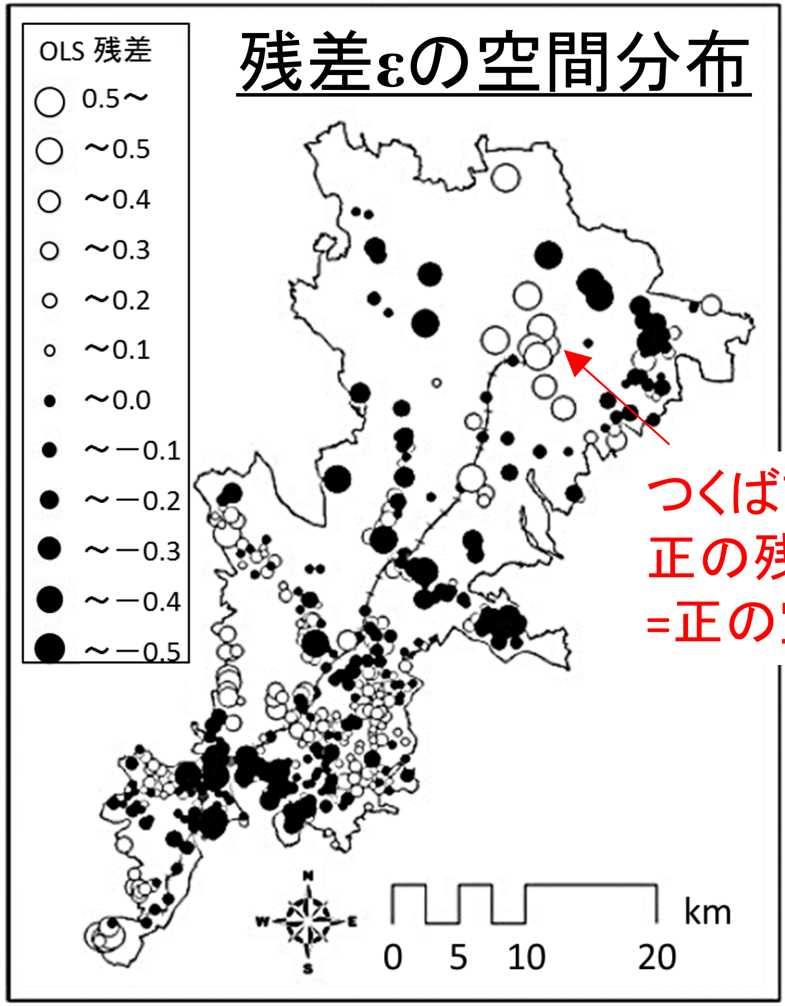
$$y_i = \beta_0 + \sum_{k=1}^{K} x_{i,k}\beta_k + \varepsilon_i$$

$$E[\varepsilon_i] = 0 \qquad Var[\varepsilon_i] = \sigma^2$$

$$Cov[\varepsilon_i, \varepsilon_j] = 0$$

残差(誤差項)には空間相関(近所と強い相関)がみられる場合が多い (例えば”つくばブランド”のような説明変数に加えられなかった重要変数が存在した場合)

→ 誤差項間の共分散がゼロにならない
→ OLSを使うべきではない



つくば市周辺に
正の残差が集中
=正の空間相関

Tsutsumi, M., & Seya, H. (2008). Measuring the impact of large-scale transportation projects on land price using spatial statistical models. *Papers in Regional Science*, *87*(3), 385-401.

空間データのための回帰モデル

空間相関を考慮した回帰モデルが空間計量経済学 (Spatial econometrics) で提案されてきた

- 誤差項の空間相関 (先ほどの例)
  - 空間エラーモデル(Spatial error model: SEM)
- 被説明変数の空間相関
  - 空間ラグモデル (Spatial lag model: SLM)
- 被説明変数と説明変数の空間相関
  - 空間ダービンモデル (Spatial Durbin model: SDM)

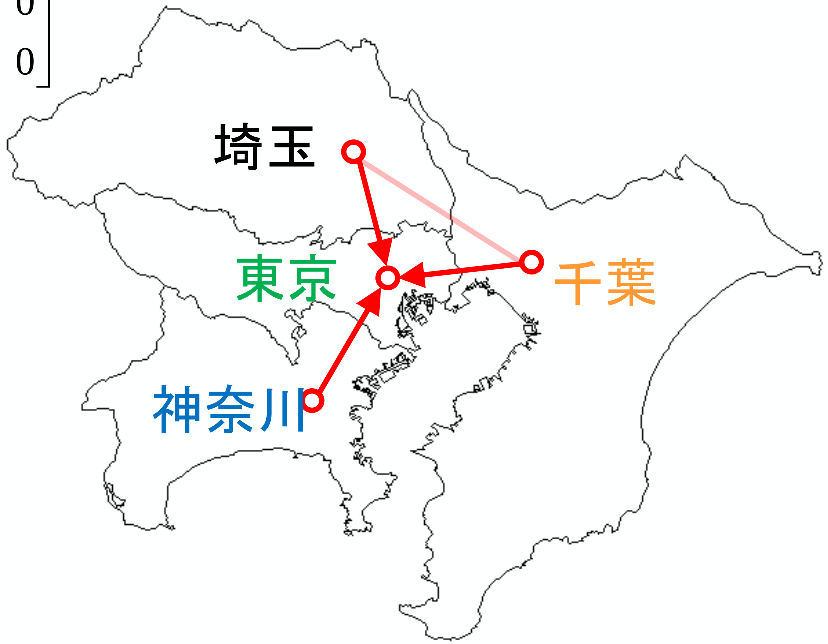
→各モデルでは、**空間重み行列**を用いて近所との関係を表す

空間重み行列

= 行基準化した近接行列

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

近接性



行基準化
(行和を1に揃える)

$$\begin{array}{cc} & \begin{array}{cccc} \text{東} & \text{千} & \text{神} & \text{埼} \end{array} \\ \begin{array}{c} \text{東京} \\ \text{千葉} \\ \text{神奈川} \\ \text{埼玉} \end{array} & \begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/2 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

y_i ：県iでの観測データ

$$\rho \sum_{j \neq i}^{n} w_{ij} y_j$$

：隣接ゾーンからの影響
(ρは影響の強さを表すパラメータ)

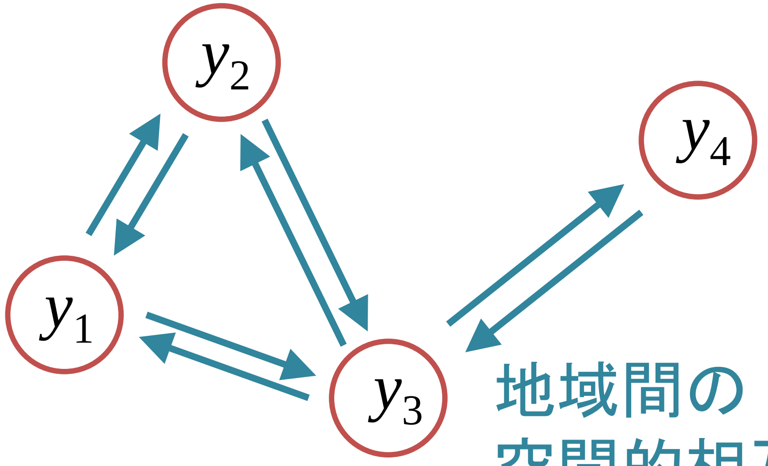
空間ラグモデル(SLM)

- 被説明変数の空間相関(波及)を考慮

$$y_i = \rho \sum_{i \neq j}^{n} w_{i,j} y_j + \sum_{k=1}^{K} x_{i,k} \beta_k + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

ρは空間相関の強さを表すパラメータ（$1 \geq \rho > 1/\omega_{min}$）
- **正**：**正**の空間相関
- **負**：**負**の空間相関（←注意が必要）

空間重み行列の最小固有値



地域間の
空間的相互作用の結果生じる

空間エラーモデル（SEM)

- 誤差項の空間相関を考慮

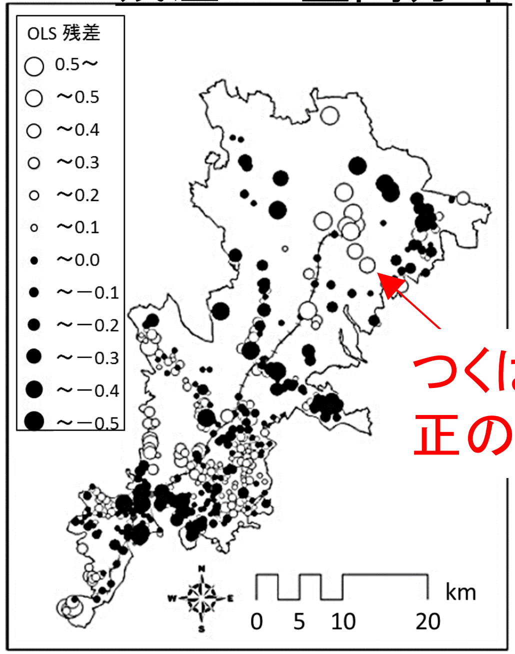
$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_k + u_i \quad u_i = \lambda \sum_{j \neq i}^{n} w_{i,j} u_j + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

λは空間相関パラメータ
（$1 \geq \rho > 1/\omega_{min}$）
- 正：正の空間相関
- 負：負の空間相関

残差(誤差項)には空間相関(近所と強い相関)がみられる場合が多い(例えば”つくばブランド”のような説明変数に加えられなかった重要変数が存在した場合)

→誤差項u_iの空間相関を
明示的にモデル化

残差εの空間分布



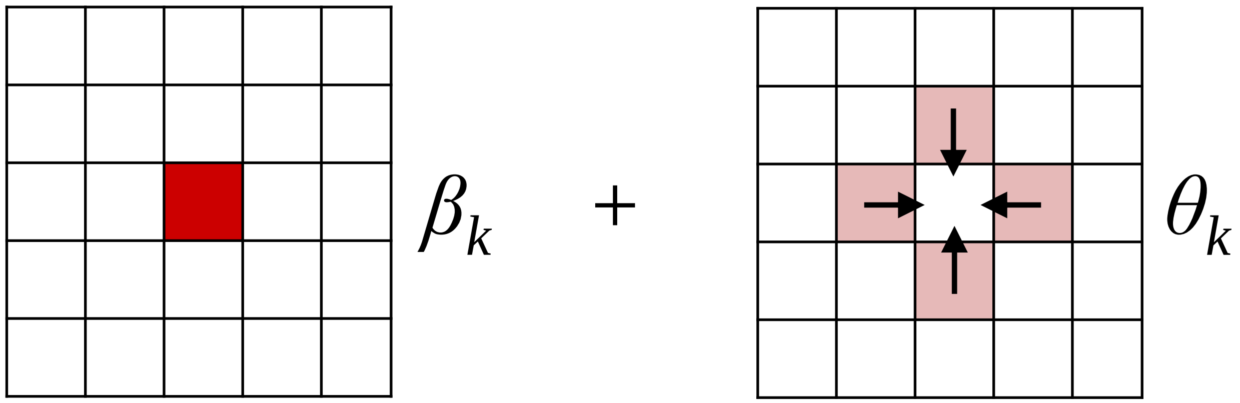
つくば市周辺に
正の残差が集中

空間ダービンモデル(SDM)

- 被説明変数の空間相関と、説明変数の空間波及を考慮

$$y_i = \rho \sum_{i \neq j}^{n} w_{i,j} y_j + \sum_{k=1}^{K} x_{i,k} \beta_k + \sum_{k=1}^{K} \left[\sum_{j \neq i} w_{i,j} x_{j,k} \right] \theta_k + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

- 説明変数の影響

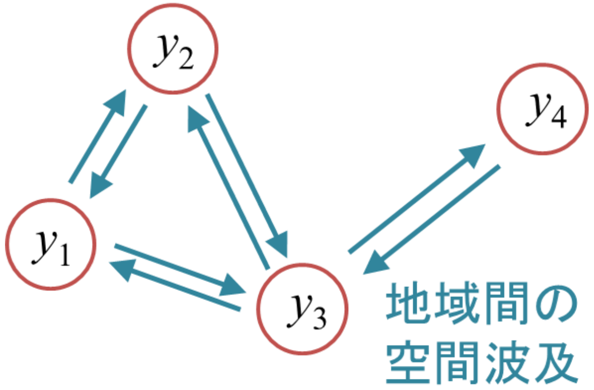


同地域の説明変数の影響　　近隣の説明変数の影響

主な空間計量経済モデル（行列標記）

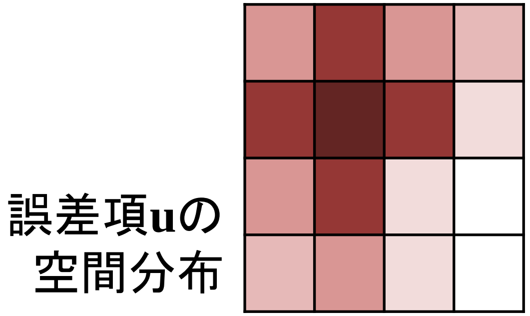
- 空間ラグモデル (SLM)

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$



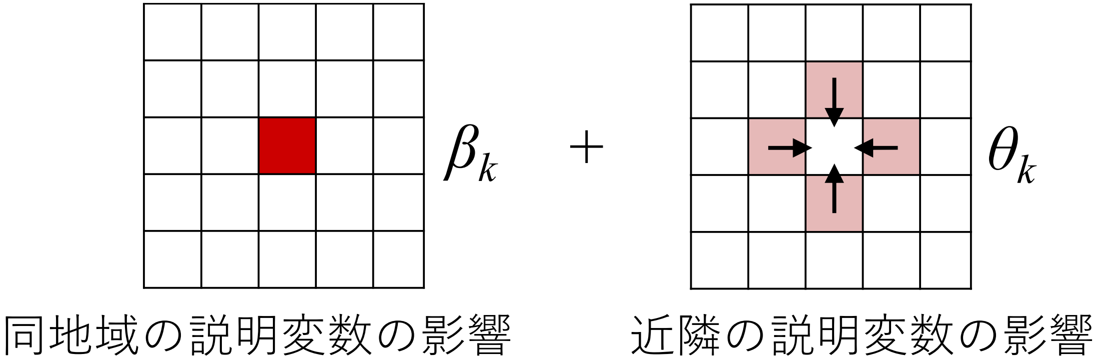
- 空間エラーモデル (SEM)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad \mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$$



- 空間ダービンモデル (SDM)

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

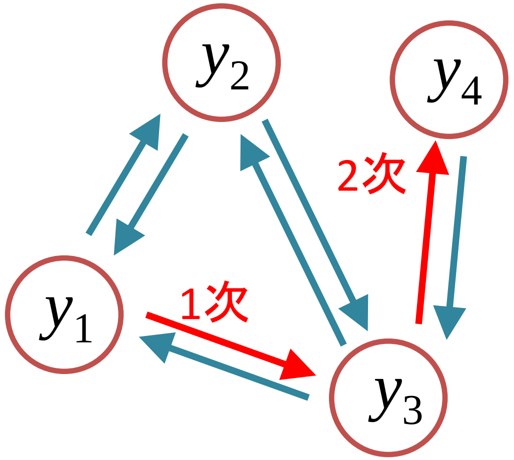


Spatial multiplier : $(\mathbf{I}-\rho\mathbf{W})^{-1}$

- SLMを以下のように展開

$$(\mathbf{I}-\rho\mathbf{W})\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$$

$$\mathbf{y}=\boxed{(\mathbf{I}-\rho\mathbf{W})^{-1}}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})$$



↓ レオンチェフ展開

$$=(\mathbf{I}+\rho\mathbf{W}+\rho^2\mathbf{W}^2+\cdots)(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})$$

自分　1次のつながり　2次のつながり　…　∞次のつながり

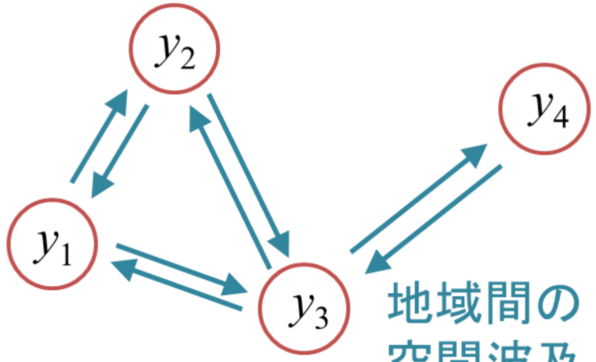
$(\mathbf{I}-\rho\mathbf{W})^{-1}$は大域的な空間波及を記述

基礎的なモデル（行列標記）

- **空間ラグモデル (SLM)**

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

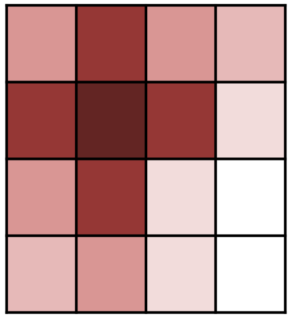
大域波及



- **空間エラーモデル (SEM)**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad \mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

大域波及

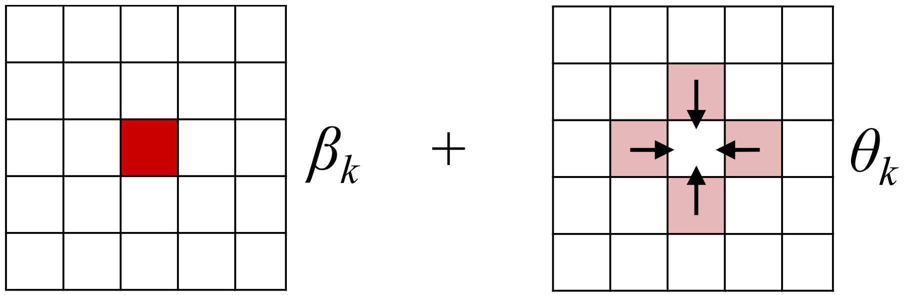


誤差項uの
空間分布

- **空間ダービンモデル (SDM)**

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta})$$

大域波及　　局所波及



同地域の説明変数の影響　　近隣の説明変数の影響

窃盗密度データ

- 地域：サンフランシスコ市の街区
- 時期：2019年

```
> d        <-read.csv("larceny_SanFran2019.csv")
> d[1:3,]
       GEOID        px       py         y        y0
1 6075010100 -122.4109 37.80772 294.07789 0.4552289
2 6075010200 -122.4221 37.80537 302.70323 0.6787068
3 6075010300 -122.4157 37.80164  85.19025 1.8519620
  pop_thou age_med   uni_rat  pov_rat  race_div
1    3.889    37.2 0.2929453 1.230307 0.6351828
2    4.167    39.8 0.4459940 1.139459 0.2835990
3    4.359    36.1 0.4451394 1.136044 0.4566414
```

記号	定義
px	街区の重心点の経度(WGS84)
py	街区の重心点の緯度(WGS84)
y	窃盗密度(面積あたり件数; 2019)
y0	前年の窃盗密度
pop_thou	人口密度（単位：千人）
age_med	居住者の年齢中央値
univ_rat	25歳以上の人口に占める大卒率
pov_rat	一定の所得水準を下回る居住者率
race_rat	人種多様性 $(1-\sum_{g=1}^{G} p_g$: p_gは人種gの比率)

- 以下の被説明変数、説明変数を仮定

被説明変数 ⟵ 説明変数 ⟶

```
> formula   <-log(y) ~ log(y0) + pop_thou+age_med+pov_rat+race_div
```

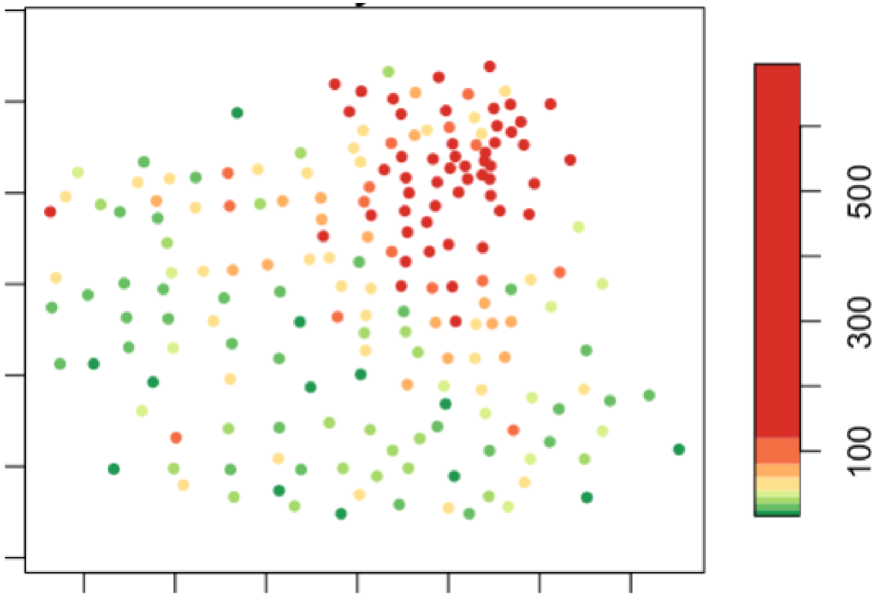
対象地域と空間重み行列

今回

- 重心点毎のポイントデータとみなし、
  4近傍を隣接とみななす

```
> coords     <-d[,c("px","py")]        #位置座標
> W_nn_nb    <-knn2nb( knearneigh(coords, k=4) ) #4近傍
> listw      <-nb2listw(W_nn_nb)       #空間重み行列
```

窃盗密度 (街区の重心点毎)

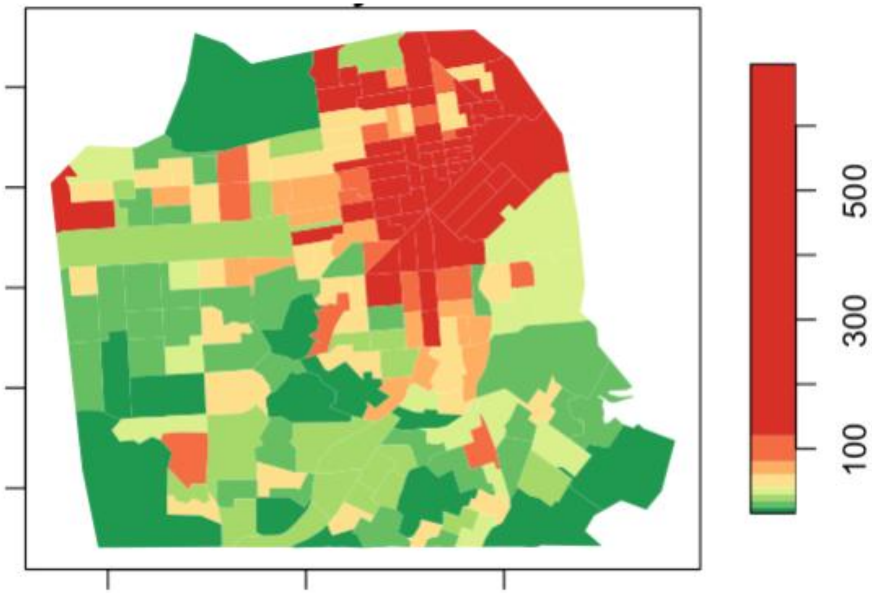


参考

- 境界または点を共有しているポリゴンを隣接とみなす場合
  (クイーン型)

```
> poly       <-st_read("sf_polygon.shp")#街区ごとのポリゴン
> W_poly_nb  <-poly2nb( poly )          #ポリゴンの隣接情報
> listw      <-nb2listw(W_poly_nb)      #空間重み行列
```

窃盗密度 (街区ポリゴン毎)



線形回帰モデル(LM)

- **ほとんどの説明変数が有意**
- **自由度調整済み決定係数 (Adjusted R2)は0.43で精度も悪くない**
- **残差のモランI統計量は0.326**
  - p値: 2.54×10^{-13}
    - → 残差に有意な正の空間相関
    - **→ 線形回帰を使うべきでない**

```
> mod0      <-lm(formula,data=d)
> summary(mod0)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.51012    0.84584   8.879 5.34e-16 ***
log(y0)      0.86874    0.08775   9.900  < 2e-16 ***
pop_thou     0.06936    0.03516   1.972  0.05002 .
age_med     -0.04548    0.01215  -3.745  0.00024 ***
pov_rat     -1.82692    0.42307  -4.318 2.54e-05 ***
race_div     0.28585    0.52539   0.544  0.58703
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9027 on 188 degrees of freedom
Multiple R-squared:  0.4452,    Adjusted R-squared:  0.4304
```

*が1つ以上ついていると5%水準で有意

```
> lm.morantest(mod0,listw=listw)

        Global Moran I for regression residuals

data:
model: lm(formula = formula, data = d)
weights: listw

Moran I statistic standard deviate = 7.2232, p-value = 2.54e-13
alternative hypothesis: greater
sample estimates:
Observed Moran I      Expectation         Variance
     0.326408169     -0.013452323      0.002213842
```

空間エラーモデル(SEM)の推定結果

- 5%水準で有意(p値<0.05)な効果の解釈
  - 前年に窃盗(y0)が多く、人口密度(pop_thou)が高いほど、窃盗が多い
- 残差の正の空間相関を捉えた
  - Lambda: 0.68523
- 線形回帰(LM)よりAIC良好

```
> sem    <-errorsarlm(formula,data=d, listw=listw, tol.solve=1.0e-20)
> summary(sem)
Coefficients: (asymptotic standard errors)
               Estimate Std. Error z value  Pr(>|z|)
(Intercept)    4.905926   0.884274  5.5480 2.890e-08
log(y0)        0.488558   0.083226  5.8702 4.352e-09
pop_thou       0.061331   0.028336  2.1644   0.03043
age_med       -0.020608   0.011460 -1.7982   0.07214
pov_rat       -0.300186   0.384462 -0.7808   0.43492
race_div      -0.134895   0.552440 -0.2442   0.80709

Lambda: 0.68523, LR test value: 58.574, p-value: 1.954e-14
Asymptotic standard error: 0.057043
    z-value: 12.012, p-value: < 2.22e-16
Wald statistic: 144.3, p-value: < 2.22e-16

Log likelihood: -223.0901 for error model
ML residual variance (sigma squared): 0.51279, (sigma: 0.71609)
Number of observations: 194
Number of parameters estimated: 8
AIC: 462.18, (AIC for lm: 518.75)
```

p値

空間ラグモデル(SLM)の推定結果

- 5%水準で有意(p値<0.05)な効果の解釈
  - 前年に窃盗(y0)が多く、人口密度(pop_thou)が高く、若い方が多い街区ほど、窃盗が多い
- 被説明変数の正の空間相関を捉えた
  - Rho: 0.59719
- 線形回帰やSEMよりAIC良好

```
> slm    <-lagsarlm(formula,data=d, listw=listw, tol.solve=1.0e-20)
> summary(slm)

Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value  Pr(>|z|)
(Intercept)  2.6709054  0.7709848  3.4643 0.0005317
log(y0)      0.5320988  0.0730056  7.2885 3.135e-13
pop_thou     0.0617820  0.0268199  2.3036 0.0212459
age_med     -0.0206024  0.0094791 -2.1735 0.0297463
pov_rat     -0.5183603  0.3372969 -1.5368 0.1243405
race_div     0.0474878  0.4006069  0.1185 0.9056400

Rho: 0.59719, LR test value: 81.215, p-value: < 2.22e-16
Asymptotic standard error: 0.055812
    z-value: 10.7, p-value: < 2.22e-16
Wald statistic: 114.49, p-value: < 2.22e-16

Log likelihood: -211.7698 for lag model
ML residual variance (sigma squared): 0.47375, (sigma: 0.6883)
Number of observations: 194
Number of parameters estimated: 8
AIC: 439.54, (AIC for lm: 518.75)
LM test for residual autocorrelation
test value: 8.9462, p-value: 0.0027805
```

p値

空間ダービンモデル(SDM)の推定結果

- 5%水準で有意(p値<0.05)な効果の解釈
  - 前年に窃盗(y0)が多く、人口密度(pop_thou)が高く、近隣街区の貧困率(lag.pov_rat)が低いほど、窃盗が多い
- 被説明変数の正の空間相関を捉えた
  - Rho: 0.47774
- AICはこれまでのモデルより良好

```
> sdm     <-lagsarlm(formula,data=d, listw=listw, tol.solve=1.0e-20,
+                     type="mixed")
> summary(sdm)
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value  Pr(>|z|)
(Intercept)   4.543438   1.201341  3.7820 0.0001556
log(y0)       0.407275   0.082824  4.9174 8.771e-07
pop_thou      0.055253   0.028126  1.9644 0.0494784
age_med      -0.012937   0.011142 -1.1611 0.2455841
pov_rat      -0.142349   0.383426 -0.3713 0.7104482
race_div      0.118588   0.543445  0.2182 0.8272610
lag.log(y0)   0.254531   0.155777  1.6339 0.1022698
lag.pop_thou -0.026918   0.058557 -0.4597 0.6457432
lag.age_med  -0.010449   0.017286 -0.6045 0.5455360
lag.pov_rat  -1.535436   0.636674 -2.4117 0.0158804
lag.race_div  0.454832   0.703629  0.6464 0.5180146

Rho: 0.47774, LR test value: 36.068, p-value: 1.906e-09
Asymptotic standard error: 0.074438
    z-value: 6.418, p-value: 1.3812e-10
Wald statistic: 41.19, p-value: 1.3812e-10

Log likelihood: -205.7897 for mixed model
ML residual variance (sigma squared): 0.46237, (sigma: 0.67998)
Number of observations: 194
Number of parameters estimated: 13
AIC: 437.58, (AIC for lm: 471.65)
LM test for residual autocorrelation
test value: 3.6339, p-value: 0.056614
```

p値

近隣からの影響

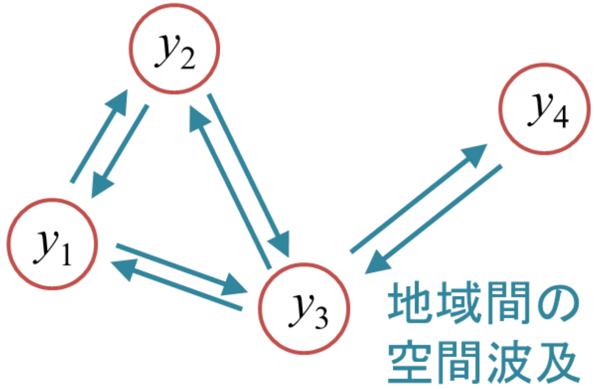
直接効果と間接効果

基礎的なモデル（行列標記）

- <u>空間ラグモデル (SLM)</u>

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

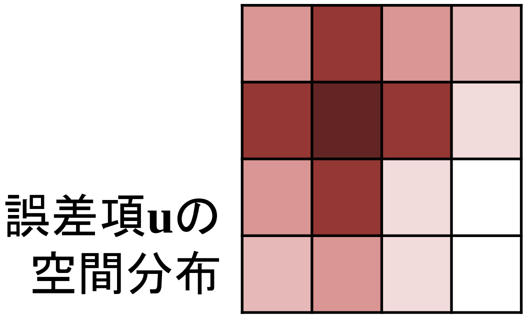
大域波及



- <u>空間エラーモデル (SEM)</u>

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad \mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

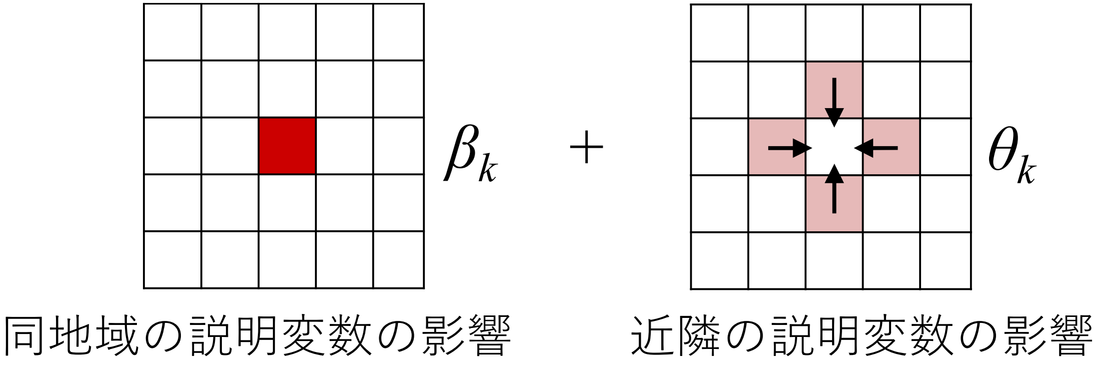
大域波及



- <u>空間ダービンモデル (SDM)</u>

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\theta})$$

大域波及　　局所波及

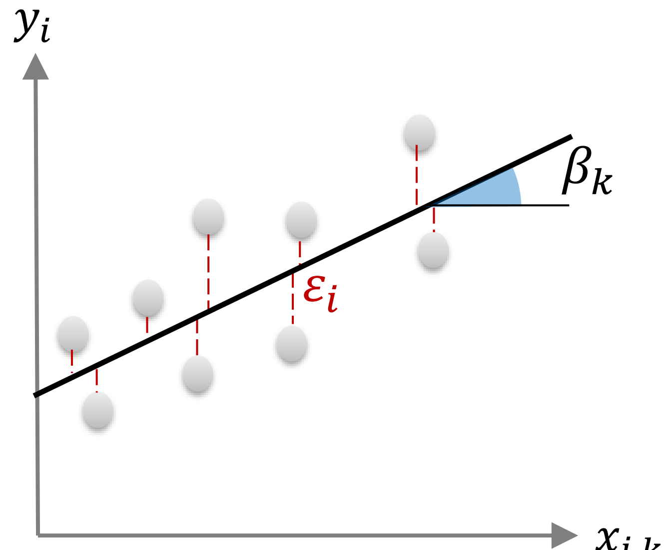


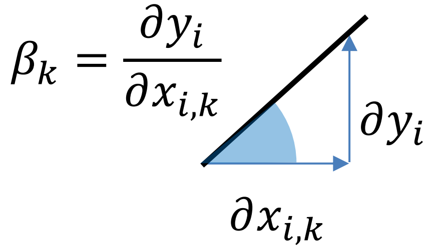
回帰係数の解釈

- 線形回帰モデル
  - β_kは$x_{i,k}$の限界効果

$x_{i,k}$で微分

$$y_i = \sum_{k=1}^{K} x_{i,k}\beta_k + \varepsilon_i \quad \Rightarrow \quad \frac{\partial y_i}{\partial x_{i,k}} = \beta_k$$





- 空間ラグモデル
  - β_kは$x_{i,k}$の限界効果ではない

$\mathbf{x}_k$で微分

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad \Rightarrow \quad \frac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \beta_k(\mathbf{I} - \rho\mathbf{W})^{-1}$$

$$\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_K]$$

直接効果と間接効果:空間ラグモデルの場合

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \mathbf{M}_k = \begin{bmatrix} \frac{\partial y_1}{\partial x_{1,k}} & \cdots & \frac{\partial y_1}{\partial x_{N,k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_{1,k}} & \cdots & \frac{\partial y_N}{\partial x_{N,k}} \end{bmatrix} = \beta_k (\mathbf{I} - \rho \mathbf{W})^{-1}$$

直接効果 -------------→

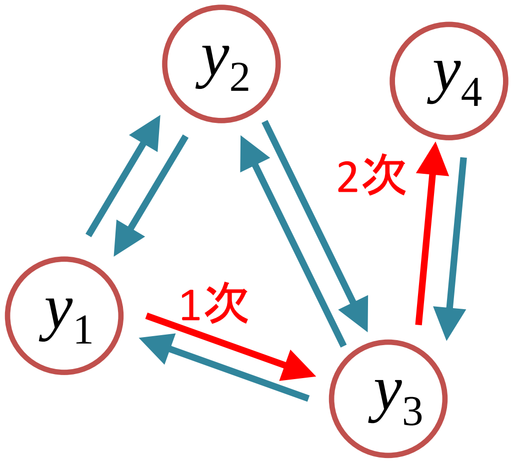
対角要素：自分への影響

$x_{i,k} \to y_i$

間接効果 ---------------→

非対角要素: 近隣からの影響

$x_{j,k} \to y_i$



要約統計量として以下が用いられる

- 直接: $\mathbf{M}_k$の対角要素の平均
- 間接: $\mathbf{M}_k$の非対角要素の平均

直接/間接効果: $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \mathbf{M}_k = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_{1,k}} & \cdots & \dfrac{\partial y_1}{\partial x_{N,k}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_N}{\partial x_{1,k}} & \cdots & \dfrac{\partial y_N}{\partial x_{N,k}} \end{bmatrix}$

- それぞれのモデルで$\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}_k} = \mathbf{M}_k$を求める以下になる

Model	$\mathbf{M}_k$	特徴
線形回帰	$\beta_k \mathbf{I}$	直接効果は β_k、間接効果は 0
SLM	$\beta_k(\mathbf{I} - \rho\mathbf{W})^{-1}$	直接効果と間接効果の比率は全説明変数で共通
SEM	$\beta_k \mathbf{I}$	直接効果は β_k、間接効果は 0
SDM	$(\mathbf{I} - \rho\mathbf{W})^{-1}(\beta_k \mathbf{I} + \theta_k \mathbf{W})$	直接効果と間接効果の比率は説明変数毎に推定

直接効果の評価：空間ダービンモデルの場合

平均値: $\frac{1}{N}\sum_{i=1}^{N}\frac{\partial y_i}{\partial x_{i,k}}$

```
> listw2  <- as(listw, "CsparseMatrix")          # 空間重み行列を疎行列へ（高速化）
> trMat   <- trW(listw2, type="mult")            # 行列のトレース評価のための関数
> ires_sdm<-impacts(sdm, tr=trMat, R=1000)       # 直接・間接効果の推定
> summary(ires_sdm)
```

前年に窃盗(y0)が多いほど
今年も窃盗は多い

```
========================================================
Simulation results ( variance matrix):
Direct:
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

               Mean      SD  Naive SE Time-series SE
log(y0)     0.46359 0.08451 0.0026724      0.0026724
pop_thou    0.05409 0.02923 0.0009244      0.0008844
age_med    -0.01574 0.01060 0.0003353      0.0003353
pov_rat    -0.36627 0.38391 0.0121402      0.0121402
race_div    0.19357 0.52117 0.0164807      0.0164807

2. Quantiles for each variable:

                2.5%      25%      50%       75%    97.5%
log(y0)     0.311254  0.40577  0.46152  0.520282 0.630314
pop_thou   -0.001346  0.03278  0.05371  0.074699 0.109647
age_med    -0.037086 -0.02280 -0.01592 -0.008794 0.004977
pov_rat    -1.133641 -0.61265 -0.36585 -0.116602 0.421050
race_div   -0.854432 -0.15715  0.17871  0.537493 1.250995
```

直接効果

パーセンタイル値

ゼロをまたいでいなければ5%水準で有意と解釈可

間接効果の評価：空間ダービンモデルの場合

平均値: $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial y_j}{\partial x_{i,k}}$

近隣で前年に窃盗(y0)が多いほど
窃盗は増える

近隣の貧困率(pov_rat)が高まるほど
窃盗は増える
- 近隣から貧しい人が集まって窃盗をする(?)

```
==================================================================
Indirect:

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
```

直接効果

	Mean	SD	Naive SE	Time-series SE
log(y0)	0.798502	0.22782	0.0072044	0.0072044
pop_thou	-0.006584	0.10241	0.0032384	0.0032384
age_med	-0.030050	0.02717	0.0008592	0.0008592
pov_rat	-2.890184	1.03605	0.0327628	0.0327628
race_div	0.930821	1.06701	0.0337417	0.0337417

```
2. Quantiles for each variable:
```

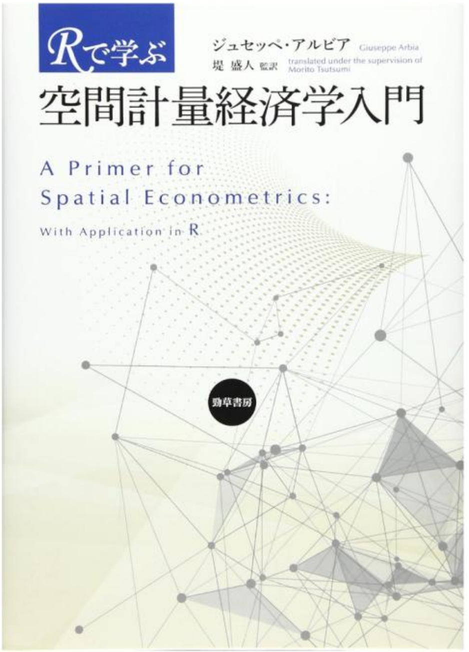
パーセンタイル値

	2.5%	25%	50%	75%	97.5%
log(y0)	0.36832	0.64466	0.79402	0.95238	1.25312
pop_thou	-0.20401	-0.07551	-0.00692	0.06136	0.20036
age_med	-0.08528	-0.04731	-0.02888	-0.01202	0.02161
pov_rat	-4.95957	-3.55508	-2.84934	-2.19634	-0.85181
race_div	-1.13842	0.20239	0.92272	1.68432	2.99673

ゼロをまたいでいなければ5%水準で有意と解釈可

まとめ

- 空間計量経済学について整理した
  - 空間相関を考慮した回帰分析手法
  - より詳しく知りたい方は、以下などが参考になると思います（図書館等にあるかもしれないです）



課題

- **今回/前回に紹介した手法を使って何らかの分析をしてください（締切：5月31日）**
  - 分析結果と考察について、ワードで2ページかそれ以上でレポートとしてまとめてください
    - ✓ 提出形式: doc, docx, またはpdf形式
  - どの変数を使っても構いません(※元々は面積あたり窃盗件数crimdenの分析用データ)
  - 特定の市区町村のみを分析対象にしてもかまいません
    - ✓ I_CODE5が市区町村コード。例えばdata[data$I_CODE5==13101,]とすると、千代田区(13101)のデータだけ抽出できます（ https://ecitizen.jp/Sac/13）

データ

- crime_tokyo.csv　　: 町丁目毎の窃盗件数ならびに関連データ
- crime_tokyo.geojson: 町丁目毎のポリゴン。CSVと同じ属性データが与えられている
- ReadMe.txt　　　　: 各変数の定義をまとめたメモ

I_CODE11	I_CODE5	lon	lat	area	crimden	popden	dpopden	agingrat	unemploy	stau_len	single_hh	u_grad
1.3101E+10	13101	139.7492	35.69851	0.03497	0.06	7800	93142.92	0.2637	0.013793	20.94578	0.481203	0.333333
1.3101E+10	13101	139.7505	35.69913	0.05309	18.83594	18477.44	100995.8	0.174	0.034749	11.45507	0.420502	0.235602
1.3101E+10	13101	139.7487	35.70158	0.12506	71.96546	3857.83	92124.59	0.147	0.025271	7.078522	0.311881	0.367292
1.3101E+10	13101	139.746	35.70099	0.04572	546.8066	7423.58	76964.71	0.2706	0.022936	14.05556	0.315789	0.334328
1.3101E+10	13101	139.7421	35.68723	0.21718	23.02238	13003.68	38936.92	0.2083	0.020799	10.23636	0.366612	0.434217

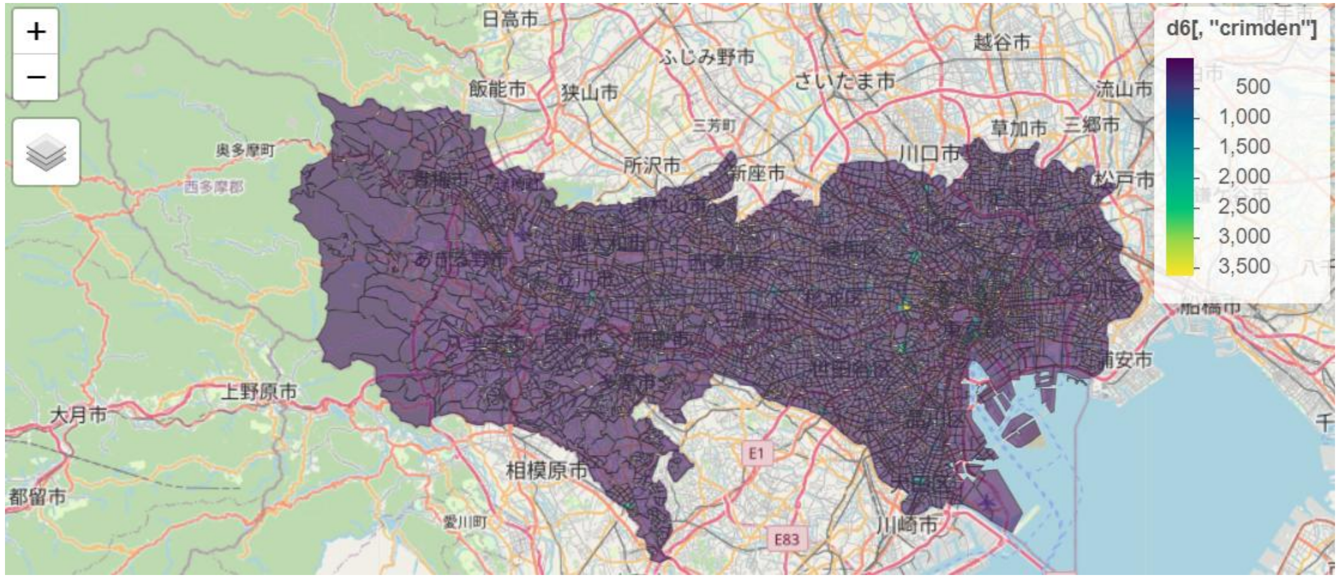
データの詳細(2つあり、どちらを使ってもOK)

出典: 国勢調査＋筑波大学社会工学コモンズ・データバンク 雨宮護・小地域時系列犯罪統計データベースの作成
（https://commons.sk.tsukuba.ac.jp/data）

crime_tokyo.csv: 町丁目毎の窃盗件数ならびに関連データ
- 地域：東京23区（奥多摩町、檜原村以外）
- 年度：2017年

変数定義

- I_CODE_5　市区町村コード（https://ecitizen.jp/Sac/13 参照）
- I_CODE11　町丁目コード
- lon　重心の経度
- lat　重心の緯度
- area　面積(km2)
- crimden　面積あたりの窃盗件数
- popden　昼間人口密度(人/km2)
- dpopden　夜間人口密度(人/km2)
- forpden　外国人の人口密度(人/km2)
- agingrat　65歳人口の比率
- unemploy　失業率
- stau_len　平均居住年数
- single_hh　単身世帯率
- u_grad　大卒率



町丁目毎の窃盗密度(crimden)

crime_tokyo.geojson: 町丁目毎のポリゴンデータ(GeoJSON形式)
- CSVと同じ変数を収録。町丁目の並び順はCSVと同じ