Review Pemodelan Regresi

Kuliah 1 | Regresi Spasial (STA1352)

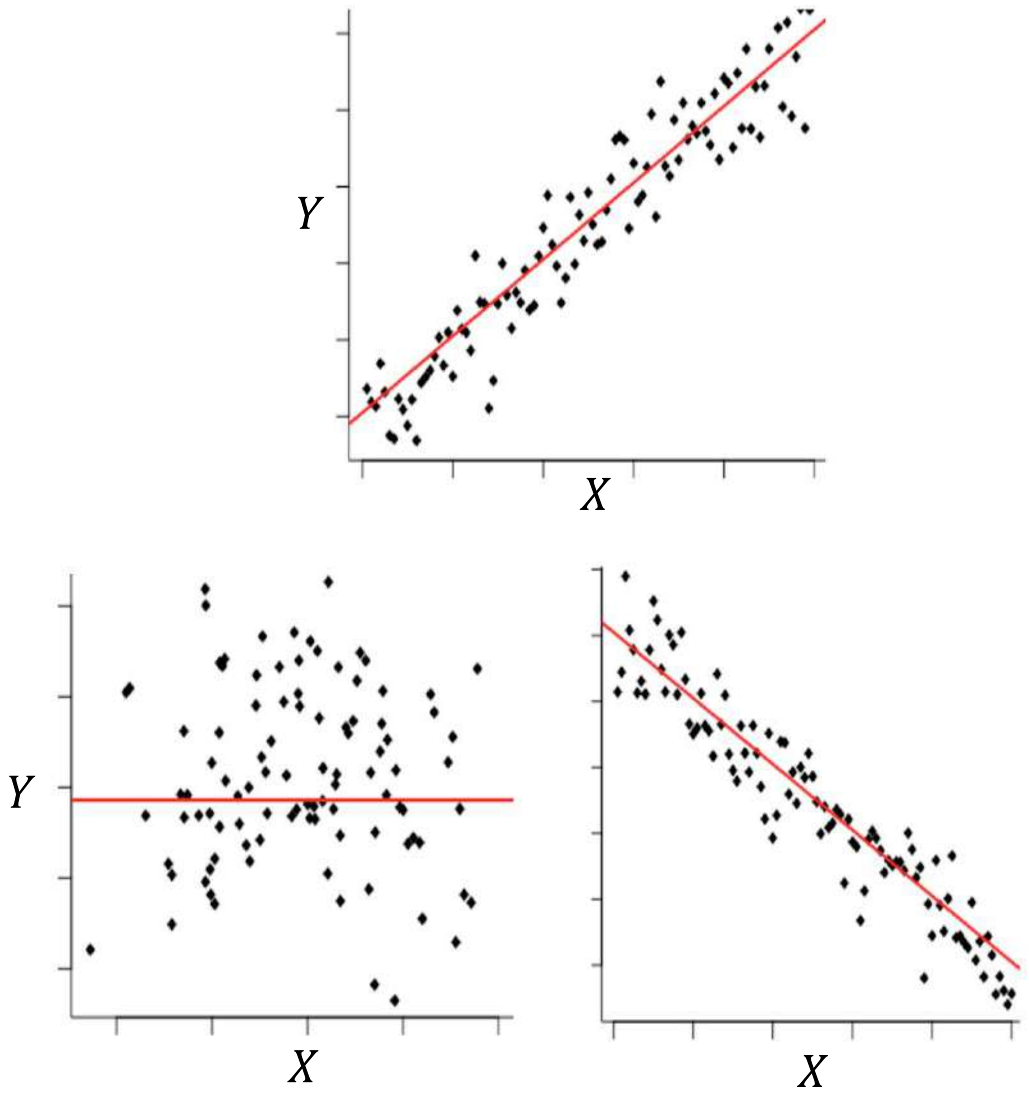
rahmaanisa@apps.ipb.ac.id

Outline

- Pemodelan regresi linear sederhana
- Penduga kuadrat terkecil
- Pemodelan regresi berganda
- Pengujian hipotesis dalam pemodelan regresi
- Kebaikan model regresi
- Asumsi dalam pemodelan regresi linear
- Beberapa isu dalam pemodelan regresi linear

Pemodelan Regresi

- Analisis Regresi digunakan untuk:
  - **Memprediksi** nilai dari peubah tak bebas (*dependent variable*) berdasarkan nilai setidaknya satu peubah bebas (*independent variable*)
  - Menjelaskan **dampak perubahan** peubah bebas terhadap peubah tak bebas
- Peubah tak bebas (Y) → peubah yang ingin dijelaskan
- Peubah bebas (X) → peubah yang digunakan untuk menjelaskan peubah tak bebas



Pemodelan Regresi Sederhana

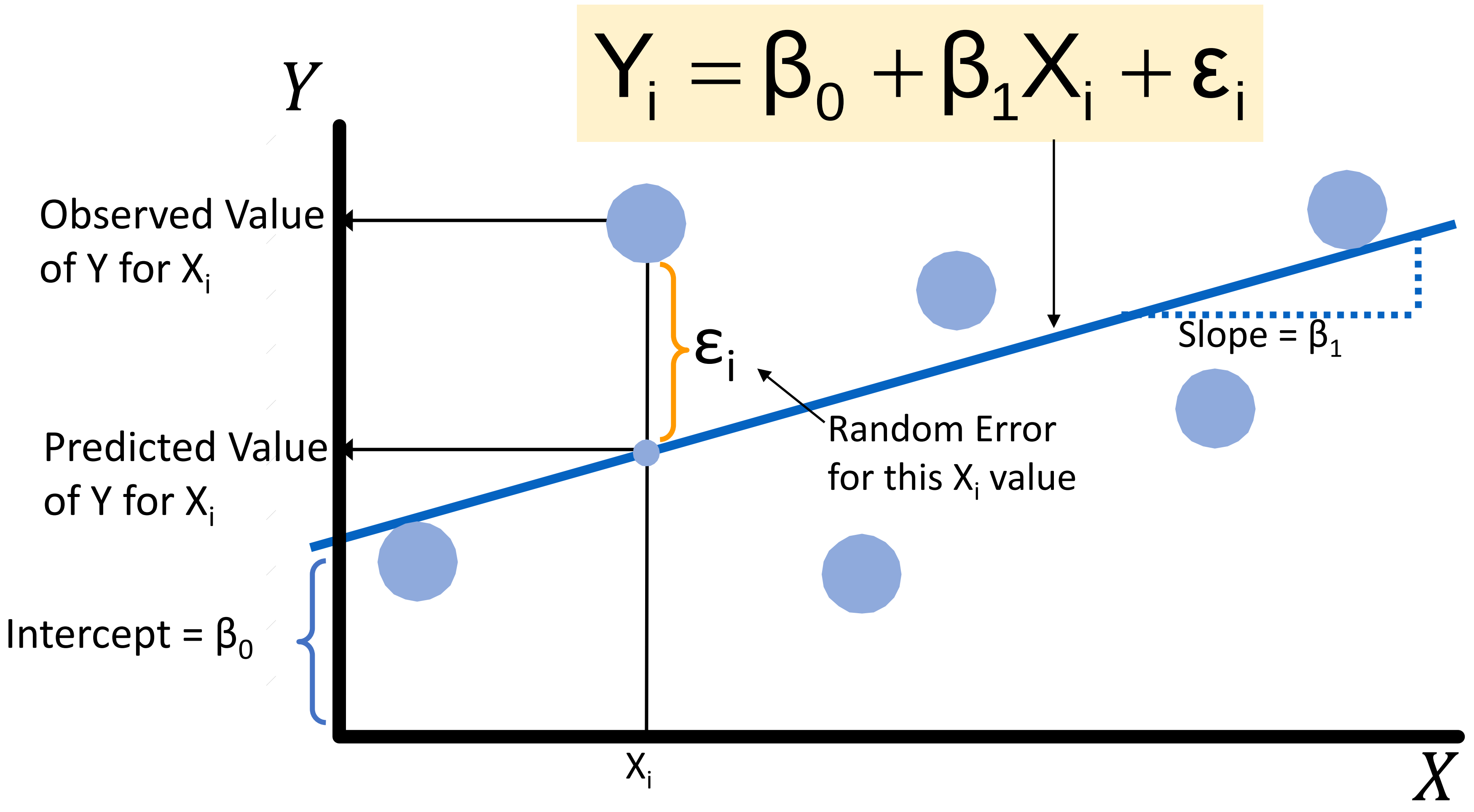
- Hubungan antara X dan Y dideskripsikan sebagai suatu fungsi linier
- Perubahan pada peubah Y diasumsikan disebabkan oleh perubahan pada peubah X
- Model persamaan regresi linier untuk **POPULASI** : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

  dengan β_0 dan β_1 adalah koefisien model populasi dan ε adalah *random error term*.

Dependent variable
Population Y intercept
Population Slope Coefficient
Independent Variable
Random Error term

$$Y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{Linear component}} + \underbrace{\varepsilon_i}_{\text{Random Error component}}$$

Pemodelan Regresi Sederhana



Pemodelan Regresi Sederhana

Persamaan regresi linier sederhana memberikan dugaan bagi garis regresi populasi

Estimated (or predicted) y value for observation i

Estimate of the regression intercept

Estimate of the regression slope

Value of x for observation i

$$\hat{y}_i = b_0 + b_1 x_i$$

The individual random error terms e_i have a mean of zero

$$e_i = (y_i - \hat{y}_i) = y_i - (b_0 + b_1 x_i)$$

Penduga Kuadrat Terkecil

- Salah satu cara memperoleh nilai b_0 dan b_1 adalah dengan meminimumkan jumlah kuadrat galat (JKG).
- Metode ini dikenal sebagai metode kuadrat terkecil (MKT)

$$\begin{aligned} \min \quad JKG &= \min \sum e_i^2 \\ &= \min \sum (y_i - \hat{y}_i)^2 \\ &= \min \sum [y_i - (b_0 + b_1 x_i)]^2 \end{aligned}$$

Penduga Kuadrat Terkecil

- Penduga bagi koefisien kemiringan garis β_1 ialah:

$$b_1 = \frac{\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}_{S_{XY}}}{\underbrace{\sum_{i=1}^{n}(x_i - \bar{x})^2}_{S_{XX}}} = \frac{S_{XY}}{S_{XX}} = \underbrace{r_{xy}}_{\text{Koefisien Korelasi Pearson}} \frac{s_Y}{s_X}$$

- Penduga bagi intersep β_0 ialah:

$$\boxed{b_0 = \bar{y} - b_1\bar{x}}$$

- Garis regresi selalu melalui titik $\bar{x}$, $\bar{y}$

Pemodelan Regresi Berganda

$$Y_i = \beta_0 + \underbrace{\beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki}}_{\text{Menggunakan banyak peubah penjelas}} + \varepsilon_i$$

Uji Hipotesis

Uji Simultan

- Menggunakan uji F dengan table ANOVA (*Analysis of Variance*)
- Hipotesis
  - H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$
  - H_1: minimum ada satu $\beta_j \neq 0$

Uji Parsial

- Menggunakan uji T
- Hipotesis untuk menguji intersep
  - H_0: $\beta_0 = 0$
  - H_1: $\beta_0 \neq 0$
- Hipotesis untuk menguji parameter lainnya
  - H_0: $\beta_j = 0$
  - H_1: $\beta_j \neq 0$

  dengan $j = 1, 2, \ldots, k$

Ilustrasi

Mathematics Achievement Test Scores and Final Calculus Grades for College Freshmen

Student	Mathematics Achievement Test Score	Final Calculus Grade
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

FIGURE 12.6(a)

MINITAB output for the data of Table 12.1

Regression Analysis: y versus x

```
The regression equation is
y = 40.8 + 0.766 x

Predictor        Coef     SE Coef        T       P
Constant       40.784       8.507     4.79   0.001
x              0.7656      0.1750     4.38   0.002

S = 8.70363       R-Sq = 70.5%     R-Sq(adj) = 66.8%

Analysis of Variance
Source           DF          SS          MS        F        P
Regression        1      1450.0      1450.0    19.14    0.002
Residual Error    8       606.0        75.8
Total             9      2056.0
```

Calculations for the Data in Table 12.1

	y_i	x_i	x_i^2	$x_i y_i$	y_i^2
	65	39	1521	2535	4225
	78	43	1849	3354	6084
	52	21	441	1092	2704
	82	64	4096	5248	6724
	92	57	3249	5244	8464
	89	47	2209	4183	7921
	73	28	784	2044	5329
	98	75	5625	7350	9604
	56	34	1156	1904	3136
	75	52	2704	3900	5625
Sum	760	460	23,634	36,854	59,816

Then

$$b = \frac{S_{xy}}{S_{xx}} = \frac{1894}{2474} = .76556 \quad \text{and} \quad a = \bar{y} - b\bar{x} = 76 - (.76556)(46) = 40.78424$$

The least-squares regression line is then

$$\hat{y} = a + bx = 40.78424 + .76556x$$

Prediction: x=50

→ $\hat{y} = a + b(50) = 40.78424 + (.76556)(50) = 79.06$

FIGURE 12.6(b)

MS Excel output for the data of Table 12.1

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.8398
R Square	0.7052
Adjusted R Square	0.6684
Standard Error	8.7036
Observations	10

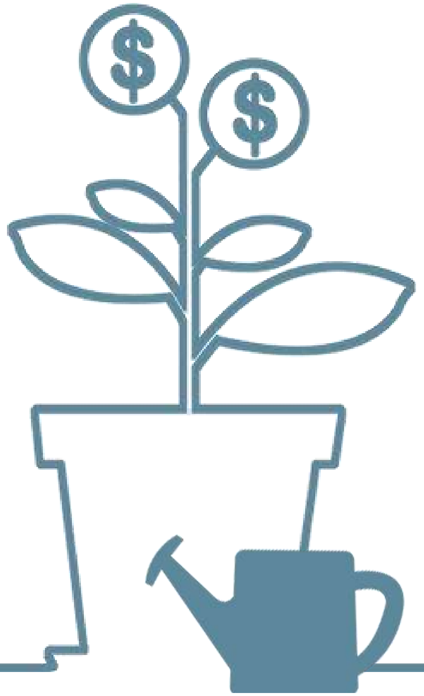
ANOVA							
	df	*SS*	*MS*	*F*	*Significance F*		
Regression	1	1449.974	1449.974	19.141	0.002		
Residual	8	606.026	75.753				
Total	9	2056					
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*
Intercept	40.784	8.507	4.794	0.001	21.167	60.401	21.167
Score	0.766	0.175	4.375	0.002	0.362	1.169	0.362

Measuring the Strength of the Relationship:
The Coefficient of Determination

The Coefficient of Determination or R-squares (R^2) can be calculated as follows:

$$R^2 = \frac{SSR}{Total\ SS}$$

This value represents the proportion of the total variation that is explained by the linear regression of y on x.



Note: for simple linear regression, $\mathrm{R{-}squares} = (r)^2$, where r is the correlation coefficient between x and y.

Ukuran Kebaikan Model

- Koefisien determinasi (R^2)
- *Root mean squared error (RMSE)*
- *Akaike information criterion (AIC)*
- *Bayesian information criterion (BIC)*

Assumptions for Regression Analysis

Basic assumption for using a linear regression model:

The population means of y at different value of x have a straight-line relationship with x, that is:

$$\mu_y = \underbrace{\alpha + \beta x}_{\text{Line of means}}$$

However, the regression model contains an error term ε. In order to use this probabilistic model for making inferences, we need to be more specific about the error term, that is by assuming that the value of ε satisfy these assumptions:

1) the errors are **independent**
2) the errors have a mean 0 and a **common variance** equal to σ^2
3) the errors have a **normal probability distribution**

Beberapa Isu dalam Pemodelan Regresi

- Multikolinieritas
- Pencilan
- Pelanggaran asumsi

Referensi

- Hoffmann, J. P. (2021). *Linear regression models: applications in R*. Crc Press.

- Sumber lain yang relevan.

Pengenalan Data Spasial

Kuliah 2

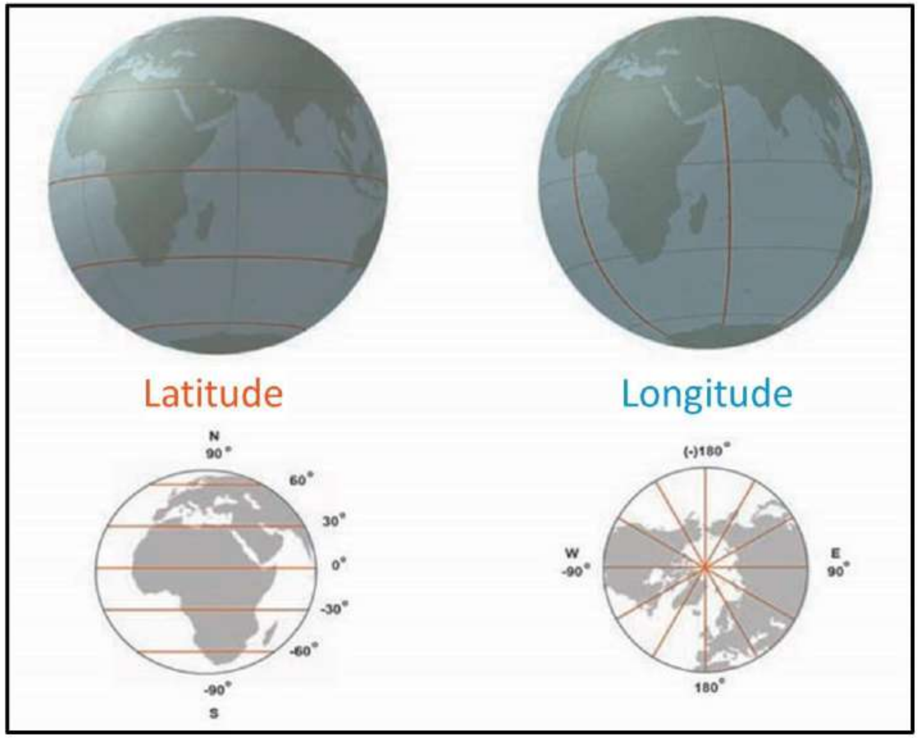
Regresi Spasial (STA1352)

rahmaanisa@apps.ipb.ac.id

Outline

- Definisi data spasial
- Sistem referensi koordinat
- Jenis-jenis data spasial

Apa itu Data Spasial?



- Data yang memiliki atribut lokasi
- Data spasial dapat memiliki dua atribut:
  - atribut lokasi
  - atribut peubah yang ingin diamati
- Lokasi: tidak hanya terkait wilayah/daerah, bisa juga lokasi sel di dalam tubuh
- Data geospasial → atribut spasial berupa lokasi suatu wilayah

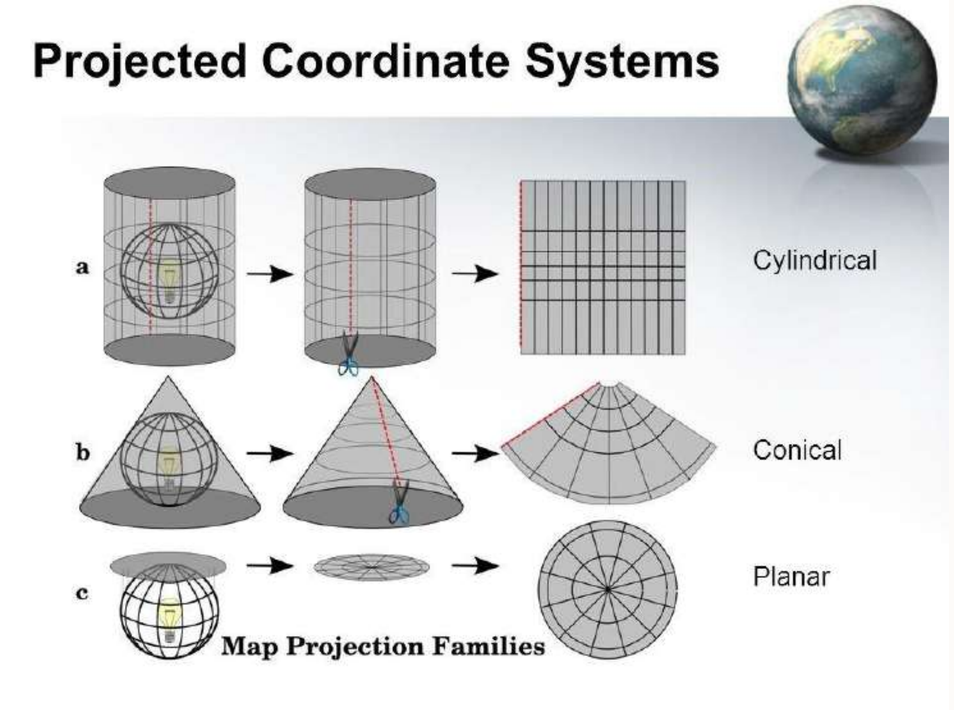
Data Spasial

- Data spasial dapat disimpan dalam beberapa format:
  - Atribut peubah : format tabular, seperti .csv, .xlsx, .dbf, …
  - Atribut spasial: shape files (.shp), GeoJSON, tiff, …
- Peta tematik relatif banyak tersedia dalam format shapefile.

Sistem Referensi Koordinat

Coordinate Reference System (CRS)

- CRS diperlukan untuk memetakan rupa bumi yang semula 3D menjadi suatu bidang 2D
- CRS mencakup beberapa komponen, yaitu:
  - Sistem koordinat
  - Satuan horizontal dan vertical
  - Datum
  - Informasi proyeksi
- Contoh CRS yang cukup banyak digunakan untuk system GPS dan pemetaan: **WGS84 (World Geodetic System 1984)**

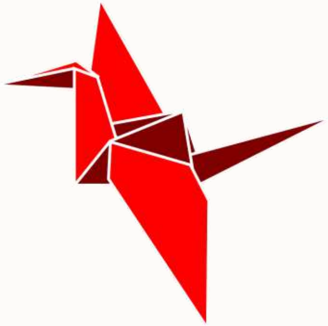


Referensi:

1. https://www.earthdatascience.org/courses/earth-analytics/spatial-data-r/intro-to-coordinate-reference-systems/#:~:text=In%20summary%20%2D%20a%20coordinate%20reference,on%20a%202%2Ddimensional%20surface.
2. https://mgimond.github.io/Spatial/chp09_0.html

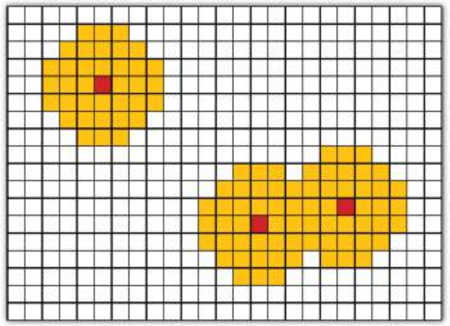
Jenis-jenis data spasial

Data Titik



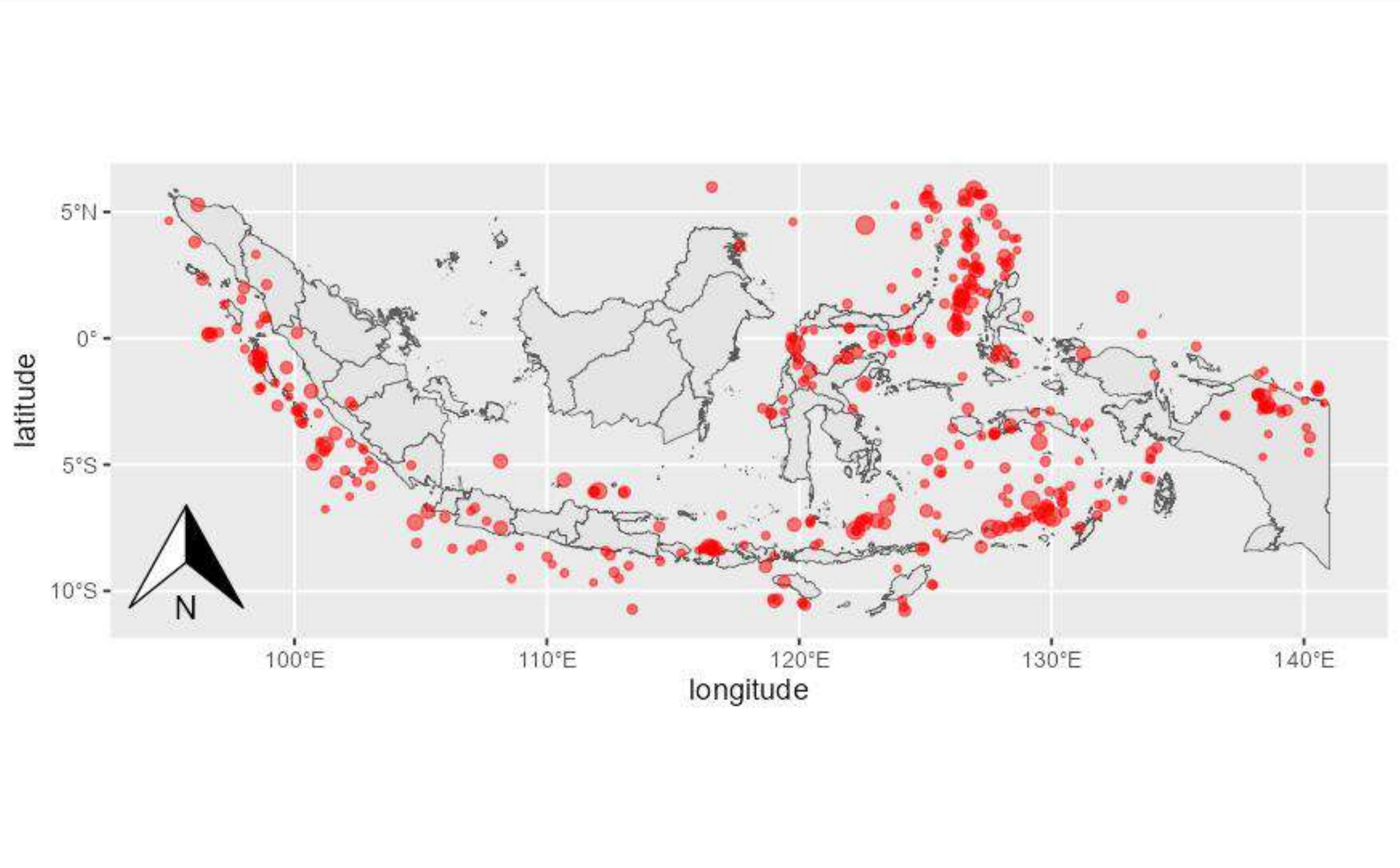
Data Poligon

Data Garis



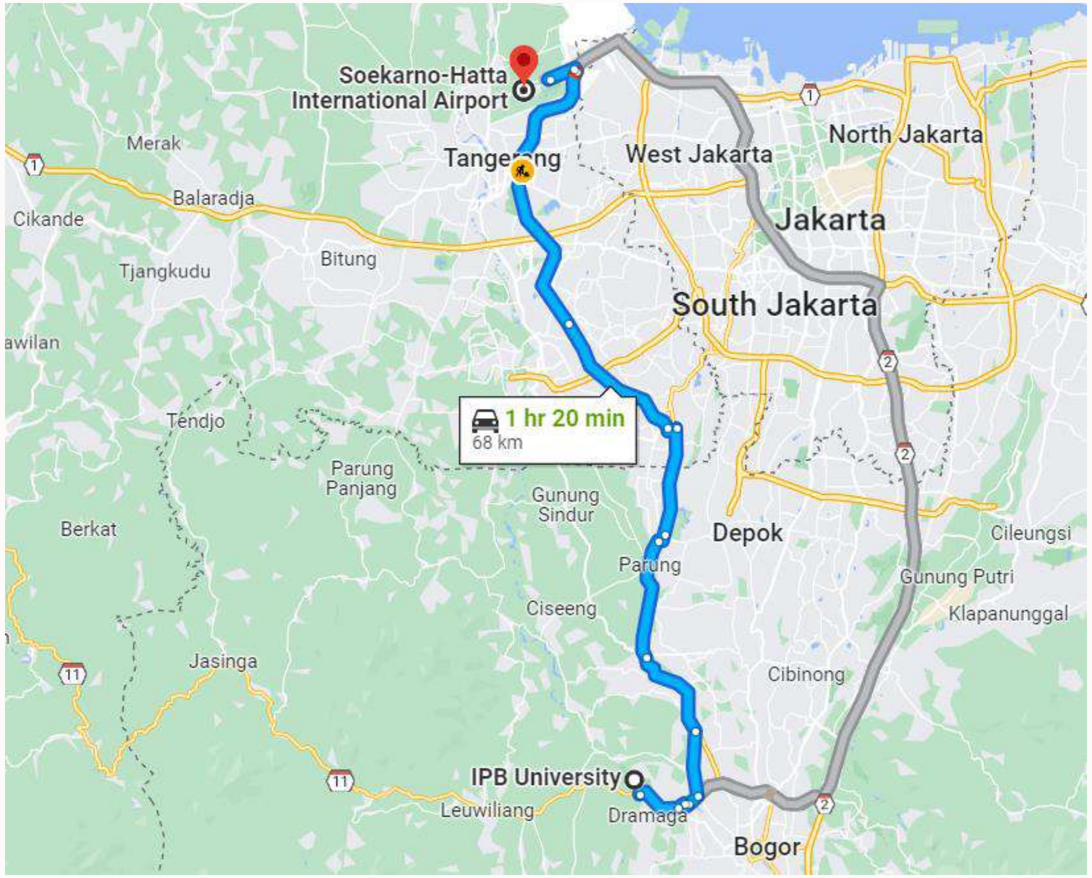
Data Grid/Raster

Data Titik

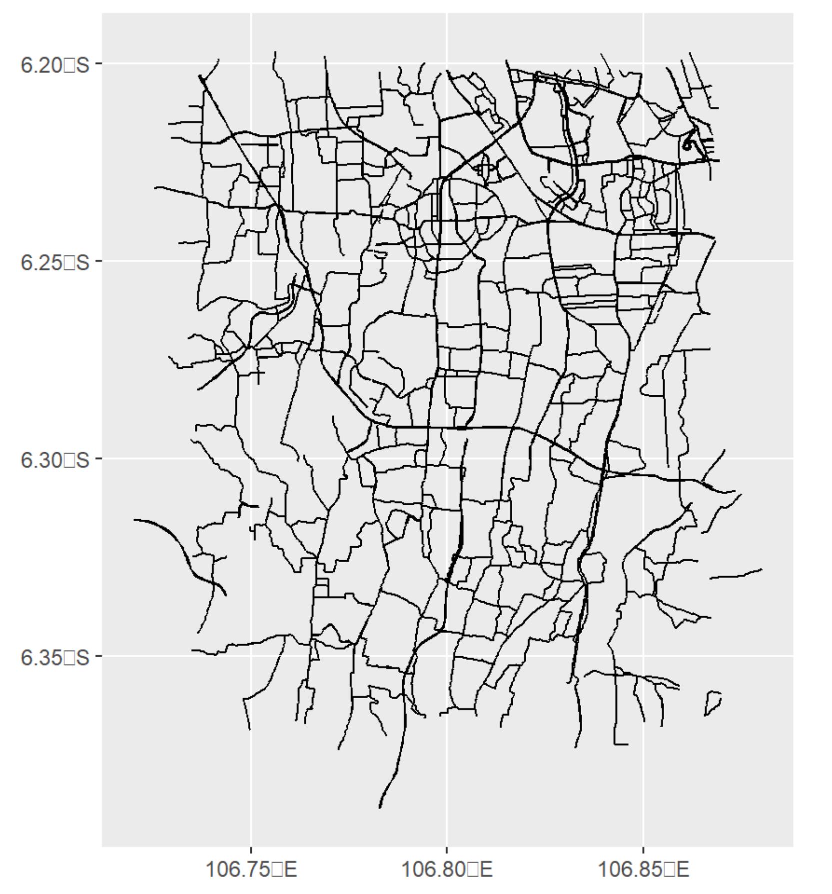


Peta titik epicentrum gempa bumi (> M 5.4) di Indonesia sejak 2015

Data Garis

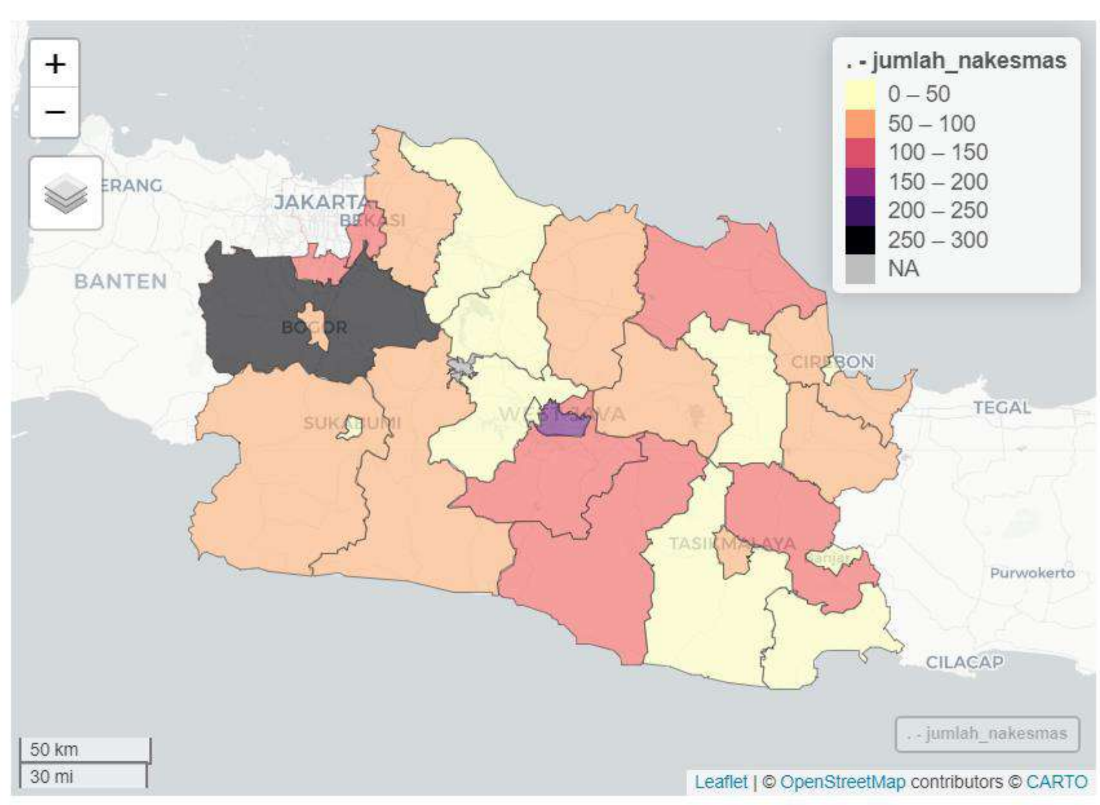


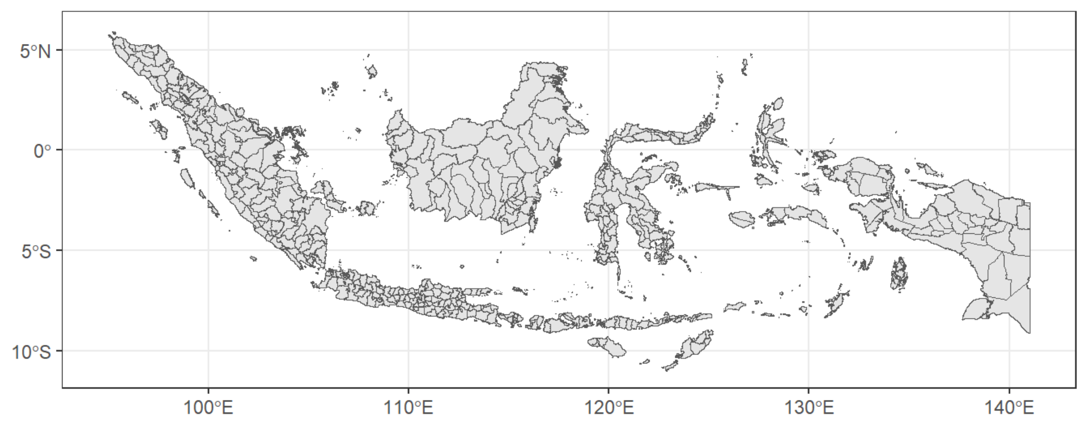
Rute perjalanan



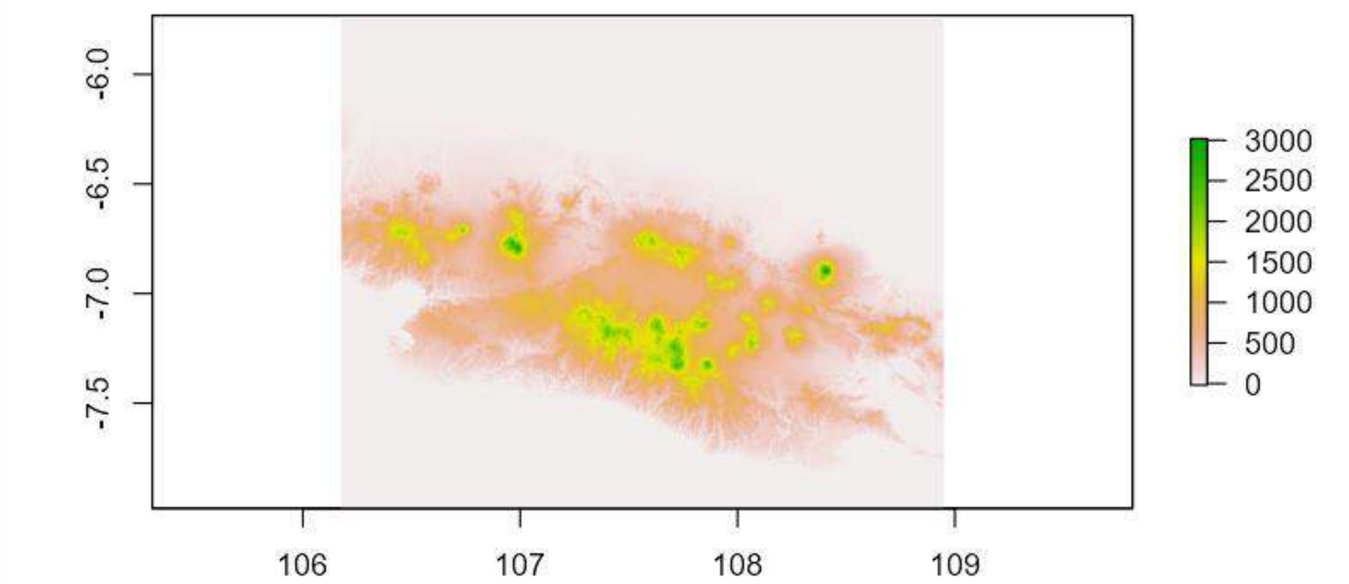
Jalan raya di Kawasan Jakarta Selatan

Data Poligon

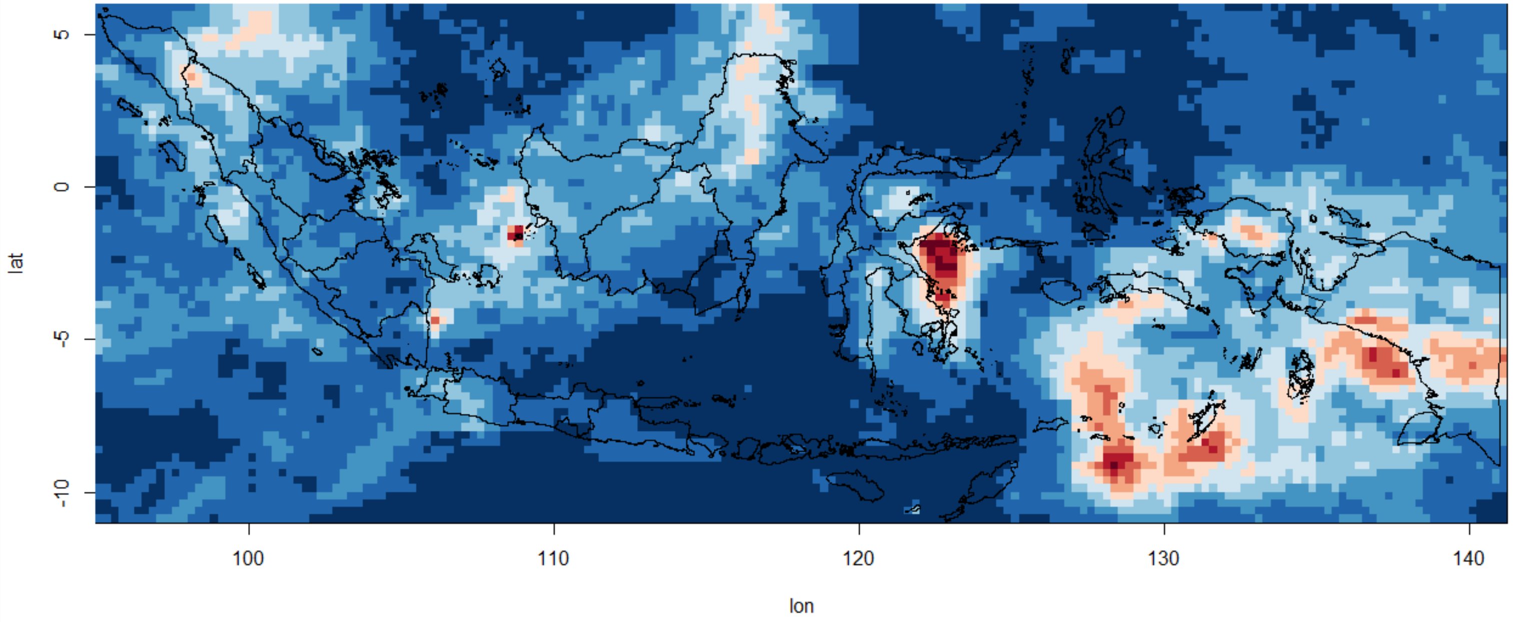




Data Grid / Raster



(1) DEMNAS Provinsi Jawa Barat



(2) Presipitasi Indonesia pada bulan Mei 2019

Bahan Diskusi

- Tentukan beberapa contoh data spasial lainnya, serta jelaskan bagaimana data tersebut dapat dimanfaatkan?

Pemodelan Proses Spasial

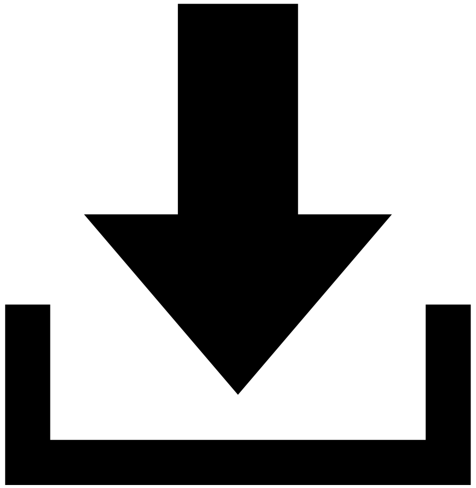
Menurut Cressie dan Moores (2021) [2] , suatu pemodelan proses spasial dapat dibedakan menjadi tiga jenis, yaitu:

1. point process (titik)
2. lattice process (bersifat diskret)
3. geostatistical process (bersifat kontinu)

[2] Cressie, N., & Moores, M. T. (2021). Spatial statistics. arXiv preprint arXiv:2105.07216. Retrieved from: https://arxiv.org/pdf/2105.07216.pdf

Beberapa sumber data

- USGS (https://www.usgs.gov/)
- Indonesia Geospatial Portal (https://tanahair.indonesia.go.id/)
- Open Street Map (https://www.openstreetmap.org/)
- NOAA
- etc



Analisis Data Area

Kuliah 3 | Regresi Spasial

rahmaanisa@apps.ipb.ac.id

Outline

- Review pemodelan proses spasial
- Analisis data spasial
- Lingkup analisis data area / *lattice*
- Tahapan analisis data area
- Ilustrasi beberapa kasus

Review pemodelan proses spasial

- *Point process*
- *Lattice process*
- *Geostatistical process*

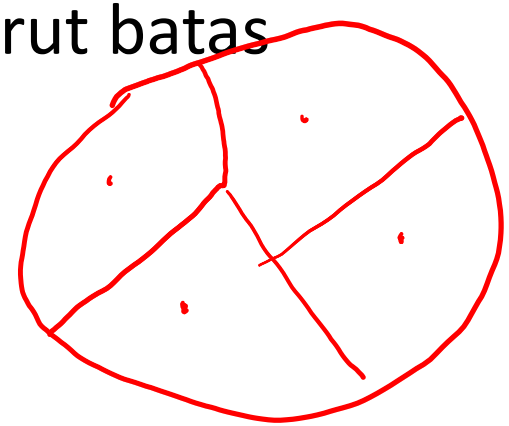
Analisis Data Spasial

- Konfigurasi titik dalam ruang *(spatial point pattern)*
  - Pola sebaran titik: acak, clustered, uniform
- Analisis data spasial kontinu
  - Kestationeran spasial
  - Interpolasi spasial
- Analisis data area

Lingkup Analisis Data Area / *Lattice*

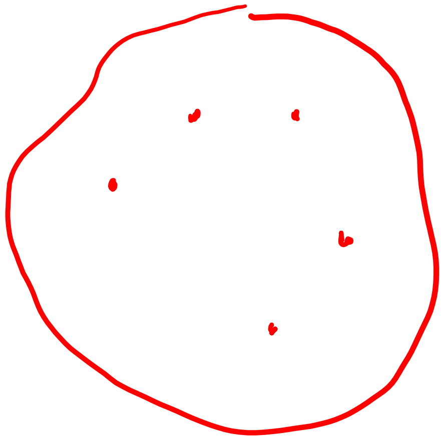
Data area / *lattice*

- Data bersifat diskret
- Merupakan data yang dikumpulkan pada tingkat area, atau hasil agregasi pada setiap unit area (sesuai level subpopulasi menurut batas wilayah tertentu)



Data kontinu / *geostatistical*

- Data bersifat kontinu
- Merupakan titik contoh dari suatu sebaran spasial kontinu

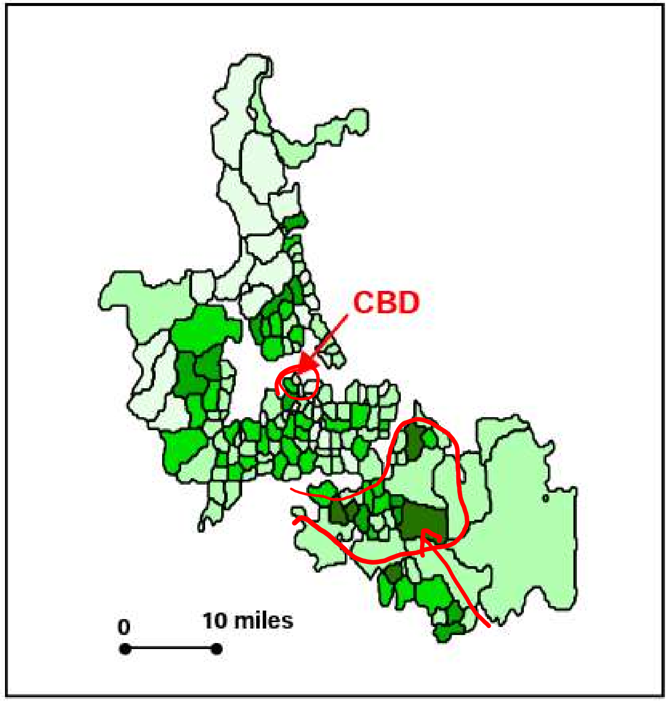


Lingkup Analisis Data Area / *Lattice*

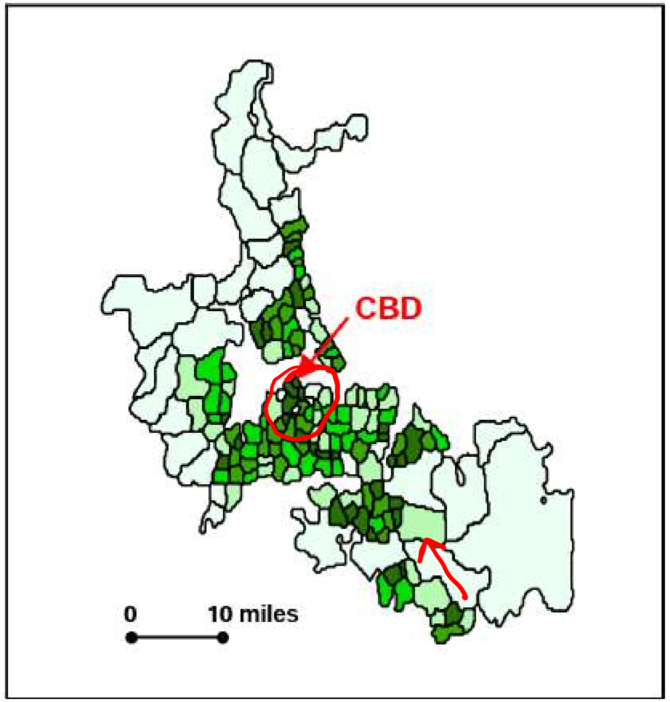
- Data direpresentasikan sebagai atribut dari suatu area poligonal yang bersifat *fixed*.
- Bentuk area umumnya tidak beraturan dan tidak seragam satu sama lain.
- Contohnya: blok sensus, daerah pemilihan, …
- Kita perlu memperhatikan apakah struktur spasial mempengaruhi struktur *feature space*.
- Penerapan analisis data area dapat mencakup beberapa bidang berikut:
    - *Spatial econometrics*
    - *Epidemiology*
    - *Sociology / demographics*
    - *Political science*
    - *Natural resources*

Lingkup Analisis Data Area / *Lattice*

- Data area seringkali direpresentasikan dalam bentuk peta *choropleth.*
- Perhatikan cara menampilkan data populasi berikut ini.



(a) *Raw population*



(b) *Population density*

Manakah yang lebih relevan dalam menggambarkan permasalahan kependudukan di area CBD ?

Sumber: Smith (2014)

Lingkup Analisis Data Area / *Lattice*

- Kajian terkait data area juga dapat mencakup analisis perbandingan pola spasial *(comparative pattern analysis)*

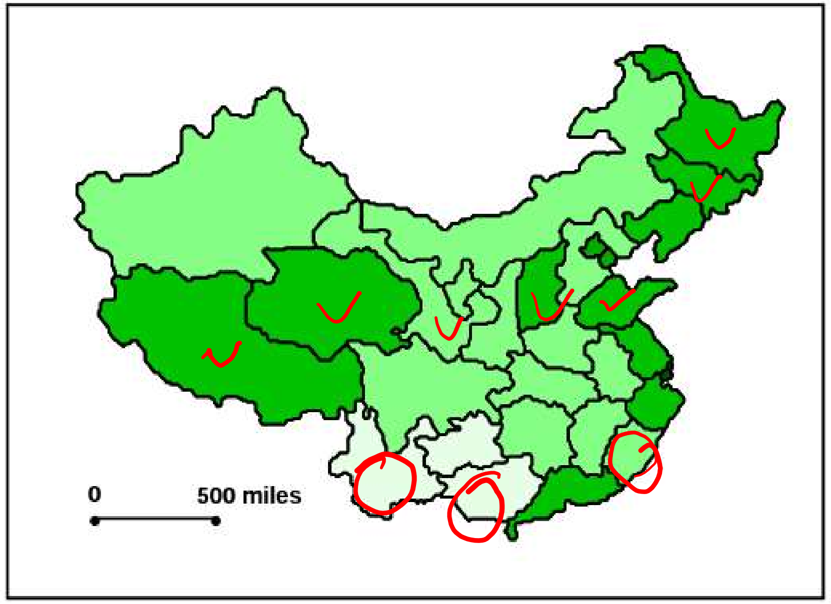


Figure 1.5. 1984 Per Capita GDP

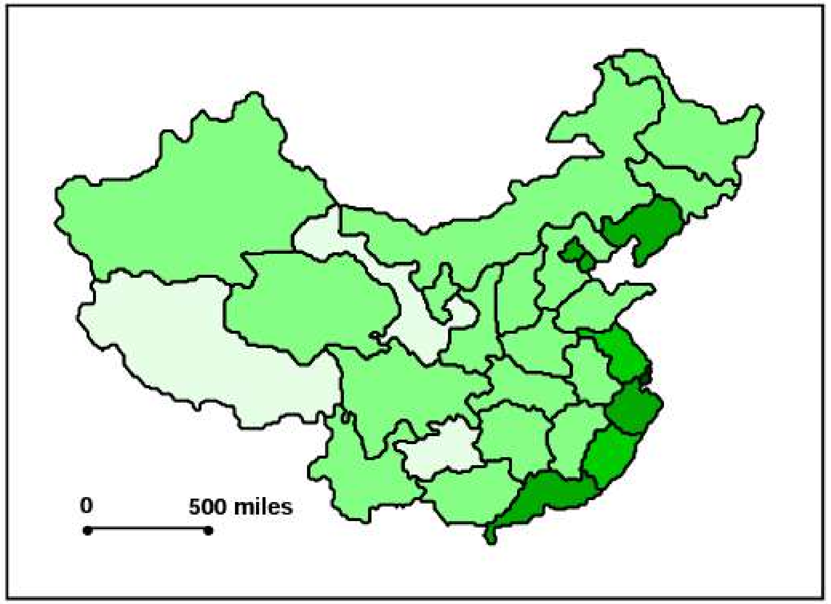


Figure 1.6. 1994 Per Capita GDP

Sumber: Smith (2014)

Lingkup Analisis Data Area / *Lattice*

- Lebih jauh lagi, pola hubungan pada data area dapat dianalisis dengan pemodelan regresi spasial.
- Contoh: data kematian akibat serangan jantung dan data “social deprivation” yang diukur dengan Jarman score.
- Warna gelap menunjukkan kematian dan social deprivation yang lebih parah.

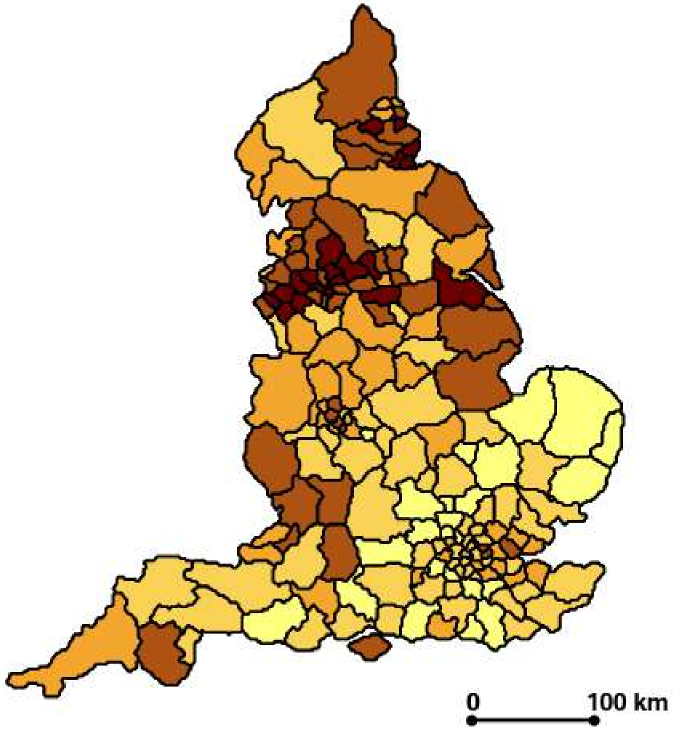


Figure 1.9. Myocardial Infarctions

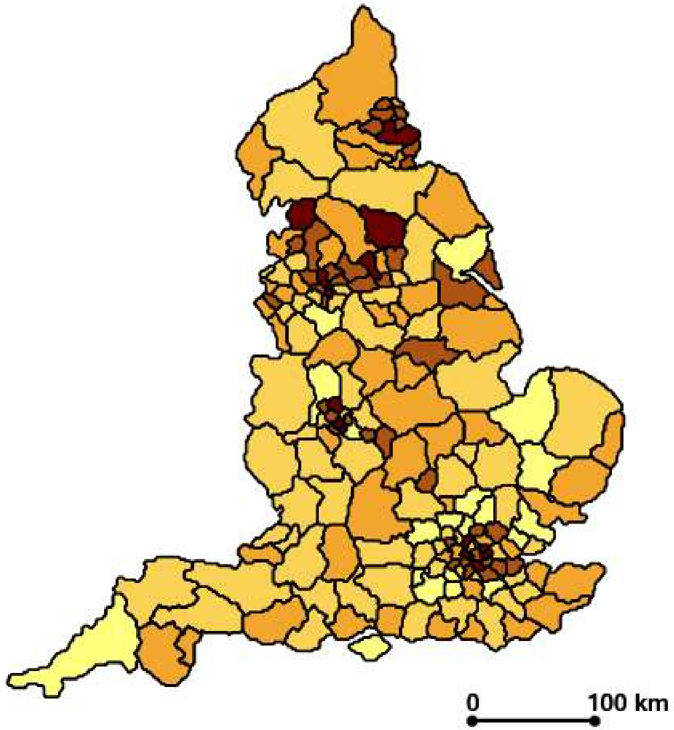
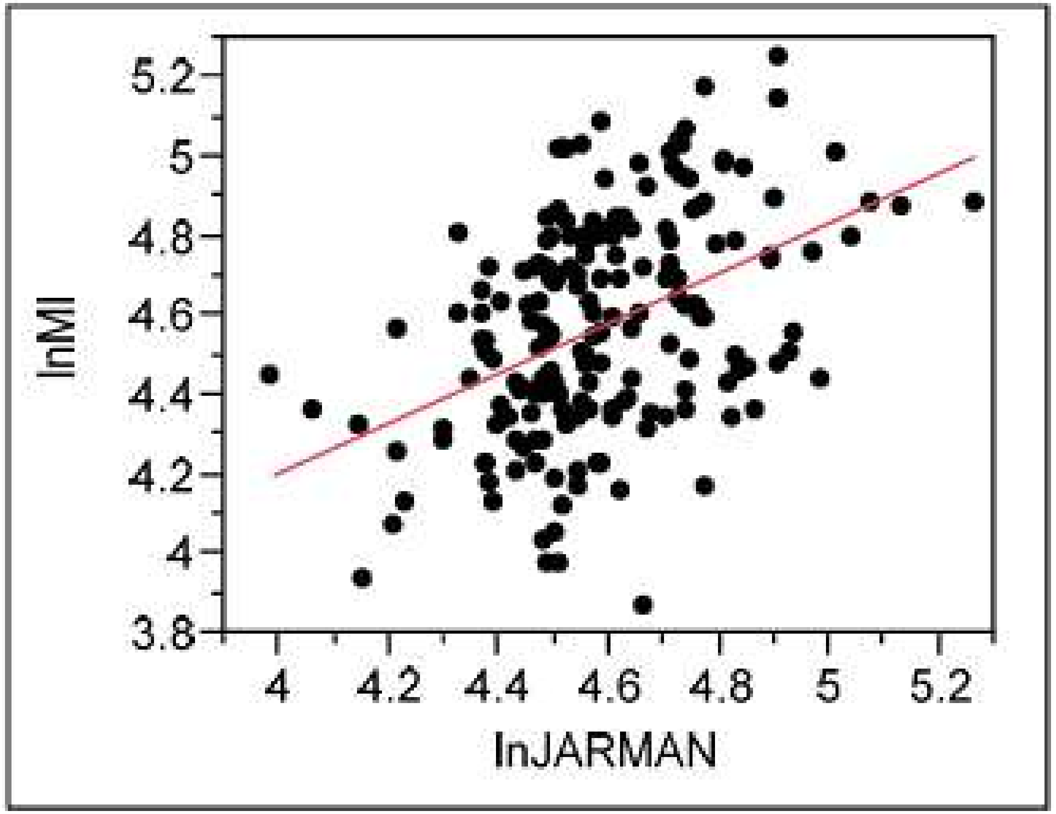


Figure 1.10. Jarman Scores

Lingkup Analisis Data Area / *Lattice*

- Terlihat kecenderungan pola hubungan positif antara log dari *Myocardial Infarction* dan Jarman *Score*



Analisis

→ eksplorasi → sebaran, perbandingan

→ pemodelan regresi spasial

→ penggerombolan (clustering)

Tahapan Analisis

- Untuk mengukur pola hubungan antar area, perlu ditentukan cara untuk mengukur hubungan ketetanggaan.
- Informasi tersebut dapat dianalisis dengan cara memasukkannya ke dalam suatu matriks pembobot spasial.
- Setelah itu, pemodelan lebih lanjut dapat dilakukan.

Contoh Kasus

No	Judul	url
1	Penerapan Model Durbin Spasial untuk Mengidentifikasi Faktor-Faktor yang Memengaruhi Diare di Jawa Timur	http://repository.ipb.ac.id/handle/123456789/99879
2	Penentuan Peubah-Peubah yang Mempengaruhi Persentase Penderita Tuberkulosis (TB) di Kota Bogor dengan Pendekatan Regresi Spasial	http://repository.ipb.ac.id/handle/123456789/58994
3	Penerapan Model Regresi Spasial Terhadap Kasus Pernikahan Dini di Provinsi Jawa Barat tahun 2016	http://repository.ipb.ac.id/handle/123456789/102689
4	Analisis Pola Penyebaran dan Faktor yang Mempengaruhi Kasus Demam Berdarah Dengue (DBD) di Provinsi Jawa Tengah	http://repository.ipb.ac.id/handle/123456789/87928
5	Identifikasi Faktor-Faktor yang Memengaruhi Kemiskinan di Jawa Tengah Menggunakan Analisis Regresi Spasial	http://repository.ipb.ac.id/handle/123456789/85827

Referensi

- Smith, T.E., (2014) *Notebook on Spatial Data Analysis* [online]. Retrieved from https://www.seas.upenn.edu/~tesmith/NOTEBOOK/index.html

PEMODELAN DATA AREA

Matriks Pembobot & Autokorelasi Spasial

Kuliah 4, 5, 6 | Regresi Spasial (STA1352)

Outline

- Ketetanggaan Spasial
- Pembobot Spasial
- Autokorelasi Spasial
  - Autokorelasi Global
  - Autokorelasi Lokal

KETETANGGAAN SPASIAL

Ketetanggaan Spasial

Secara umum, terdapat dua pendekatan untuk menentukan ketetanggaan antar dua wilayah, yaitu:

1) Kontiguitas
2) Jarak

Pendekatan Kontiguitas

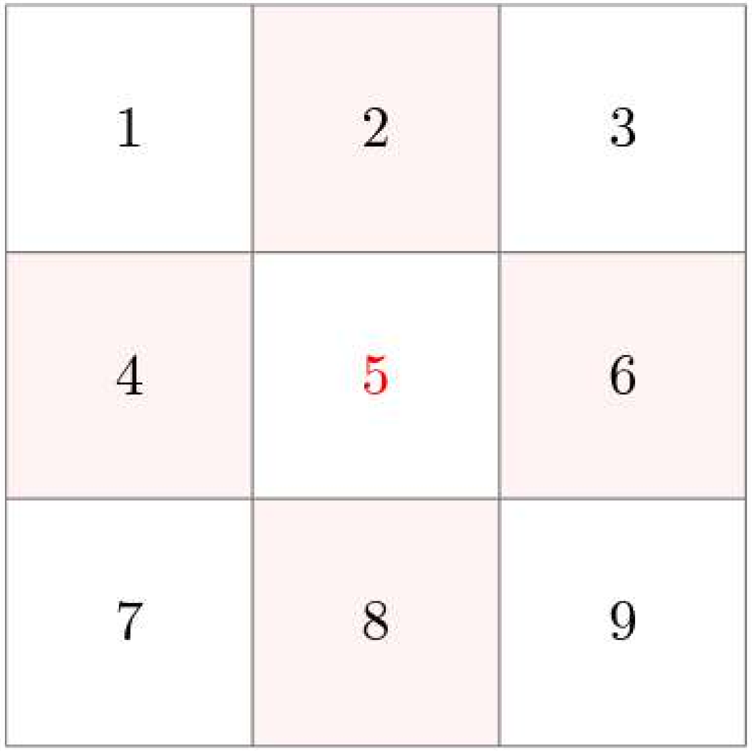
Penentuan daerah tetangga berdasarkan batas wilayah. Dua wilayah dikatakan bertetangga jika bersinggungan.

Terdapat beberapa cara, di antaranya:

1) Rook
2) Bishop
3) Queen

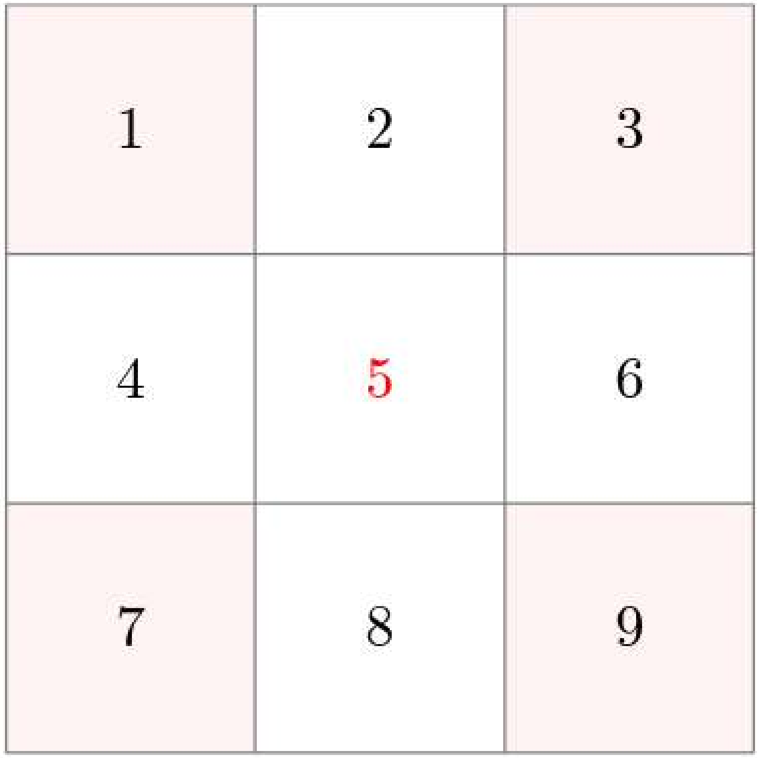
Rook Contiguity

- Bertetangga jika bersingungan sisi.
- Misalnya, gambar di samping menunjukkan wilayah 5 bertetangga dengan wilayah 2, 4, 6, dan 8.



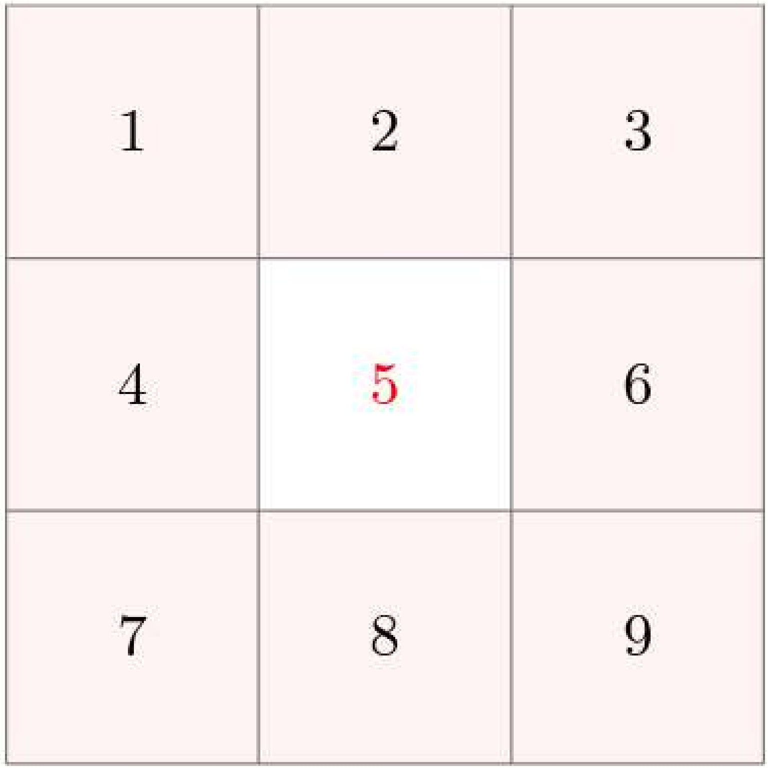
Bishop Contiguity

- Bertetangga jika bersinggungan sudut
- Tetangga dari wilayah 5 adalah 1, 3, 7, 9



Queen Contiguity

- Bertetangga jika bersinggungan sisi atau sudut
- Tetangga dari wilayah 5 adalah wilayah 1, 2, 3, 4, 6, 7, 8, dan 9



Matriks Pembobot Spasial

Definition ($\mathbf{W}$ Matrix)

Let n be the number of spatial units. The spatial weight matrix, $\mathbf{W}$, a $n \times n$ positive symmetric and **non-stochastic** matrix with element w_{ij} at location i, j. The values of w_{ij} or the weights for each pair of locations are assigned by some preset rules which defines the spatial relations among locations. By convention, $w_{ij} = 0$ for the diagonal elements.

The symmetry assumption can be dropped later.

$$\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

MATRIKS PEMBOBOT SPASIAL
ROOK CONTIGUITY

1	2	3
4	5	6
7	8	9

$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

MATRIKS PEMBOBOT SPASIAL

BISHOP CONTIGUITY

1	2	3
4	5	6
7	8	9

$$\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

MATRIKS PEMBOBOT SPASIAL
QUEEN CONTIGUITY

1	2	3
4	5	6
7	8	9

??

Jarak Spasial

- Jarak: mengukur seberapa jauh antar lokasi
- Terkadang juga perlu mempertimbangkan batas negara, pegunungan, atau batas lainnya.
- Jarak antara A ke B dan B ke A, belum tentu harus simetrik.
- Jarak juga bisa diukur dengan waktu, biaya, dll.

Ukuran Jarak

- Dapat diukur dengan jarak berikut:

Definition (Minkowski metric)

Let two point i and j, with respective coordinates (x_i, y_i) and (x_j, y_j):

$$d_{ij}^p = \left(|x_i - x_j|^p + |y_i - y_j|^p\right)^{1/p}$$

Definition (Euclidean metric)

Consider Minkowski metric and set $p = 2$, then

$$d_{ij}^e = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}.$$

Definition (Manhattan metric)

Consider Minkowski metric and set $p = 1$, then

$$d_{ij}^m = |x_i - x_j| + |y_i - y_j|\,.$$

Ukuran Jarak

- Bila kita ingin memperhitungkan kelengkungan bumi:

Definition (Great Circle Distance)

Let two point i and j, with respective coordinates (x_i, y_i) and (x_j, y_j):

$$d_{ij}^{cd} = r \times \arccos^{-1}\left[\cos|x_i - x_j| \cos y_i \cos y_j + \sin y_i \sin y_j\right] \tag{6}$$

where r is the Earth's radius. The arc distance is obtained in miles with $r = 3959$ and in kilometers with $r = 6371$.

Pendekatan Jarak

- Ketetanggaan ditentukan menurut perhitungan jarak antar wilayah.
- Beberapa pendekatan:
  a) K-tetangga terdekat (KNN)
  b) Radial
  c) Pangkat (*power*)
  d) Eksponensial

K-Tetangga Terdekat (KNN)

- k-nearest neighbors: We explicitly limit the number of neighbors.

$$w_{ij} = \begin{cases} 1 & \text{if centroid of } j \text{ is one of the } k \text{ nearest centroids to that of } i \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

- Threshold Distance (Distance Band Weights): In contrast to the k-nearest neighbors method, the threshold distance specifies that an region i is neighbor of j if the distance between them is less than a specified maximum distance:

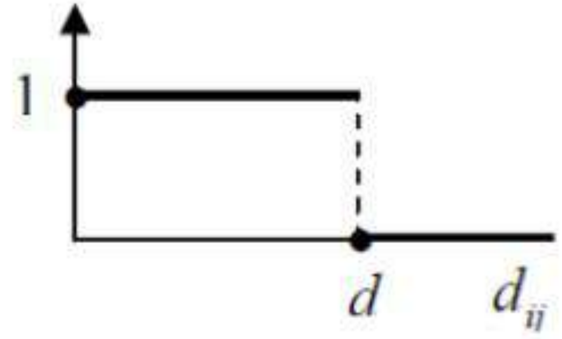
$$w_{ij} = \begin{cases} 1 & \text{if } 0 \leq d_{ij} \leq d_{max} \\ 0 & \text{if } d_{ij} > d_{max} \end{cases} \tag{10}$$

  To avoid isolates that would result from too stringent a critical distance, the distance must be chosen such that each location has at least one neighbor. Such a distance conforms to a max-min criterion, i.e., it is the largest of the nearest neighbor distances.

Pembobot Jarak Radial

1.2 Radial Distance Weights. If distance itself is an important criterion of spatial influence, and if d denotes a *threshold distance* (or *bandwidth*) beyond which there is no direct spatial influence between spatial units, then the corresponding *radial distance* weight matrix, W, has spatial weights of the form:

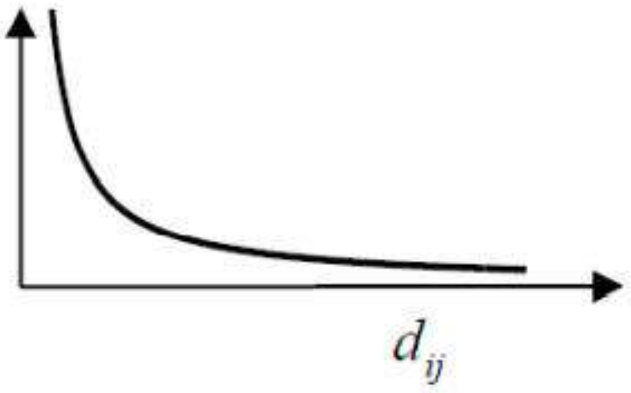
(2) $$\boxed{w_{ij} = \begin{cases} 1 & , \ 0 \le d_{ij} \le d \\ 0 & , \ d_{ij} > d \end{cases}}$$



Pembobot Jarak Pangkat

1.3 Power Distance Weights. Note that in radial distance weights there is assumed to be no diminishing effect in distance up to threshold *d*. If there are believed to be diminishing effects, then one standard approach is to assume that weights are a *negative power function* of distance of the form

(3) $$\boxed{w_{ij} = d_{ij}^{-\alpha}}$$

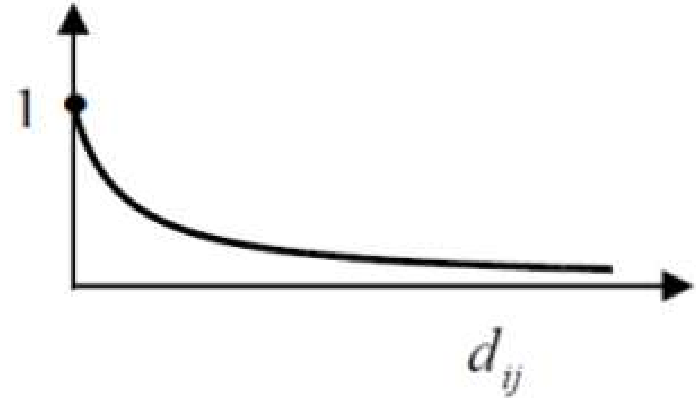


where α is any positive exponent, typically $\alpha = 1$ or $\alpha = 2$.

Pembobot Jarak Eksponensial

1.4 Exponential Distance Weights. An alternative to negative power functions are *negative exponential functions* of distance of the form:

$$(4) \quad \boxed{w_{ij} = \exp(-\alpha\, d_{ij})}$$

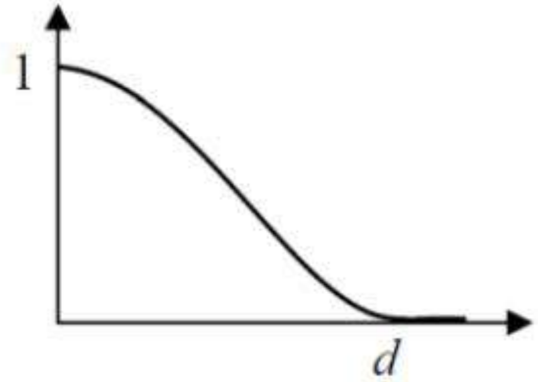


where α is any positive exponent.

Pembobot Jarak Pangkat Ganda

1.5 Double-Power Distance Weights. A somewhat more flexible family incorporates finite bandwidths with "bell shaped" taper functions. If d again denotes the maximum radius of influence (*bandwidth*) then the class of *double-power* distance weights is defined for each positive integer k by

$$(5) \qquad \boxed{w_{ij} = \begin{cases} \left[1-\left(d_{ij}/d\right)^{k}\right]^{k} & , 0 \le d_{ij} \le d \\ 0 & , d_{ij} > d \end{cases}}$$



where typical values of k are 2, 3 and 4. Note that w_{ij} falls continuously to zero as d_{ij} approaches d, and is defined to be zero beyond d. The graph shows the case of a *quadratic distance function* with $k = 2$ (see also [BG]. P.85).

Standarisasi Baris pada Matriks Pembobot Spasial

- $\mathbf{W}$'s are used to compute weighted averages in which more weight is placed on nearby observations than on distant observations.
- The elements of a row-standardized weights matrix equal

$$w_{ij}^{s} = \frac{w_{ij}}{\sum_j w_{ij}}.$$

  This ensures that all weights are between 0 and 1 and facilities the interpretation of operation with the weights matrix as an averaging of neighboring values.
- Under row-standardization, the element of each row sum to unity.
- The row-standardized weights matrix also ensures that the spatial parameter in many spatial stochastic processes are comparable between models.
- Under row-standardization the matrices are not longer symmetric!.

Spatial Lag

The spatial lag operator takes the form $\mathbf{y}_L = \mathbf{W}\mathbf{y}$ with dimension $n \times 1$, where each element is given by $\mathbf{y}_{Li} = \sum_j w_{ij} y_j$, i.e., a weighted average of the $\mathbf{y}$ values in the neighbor of i.
For example:

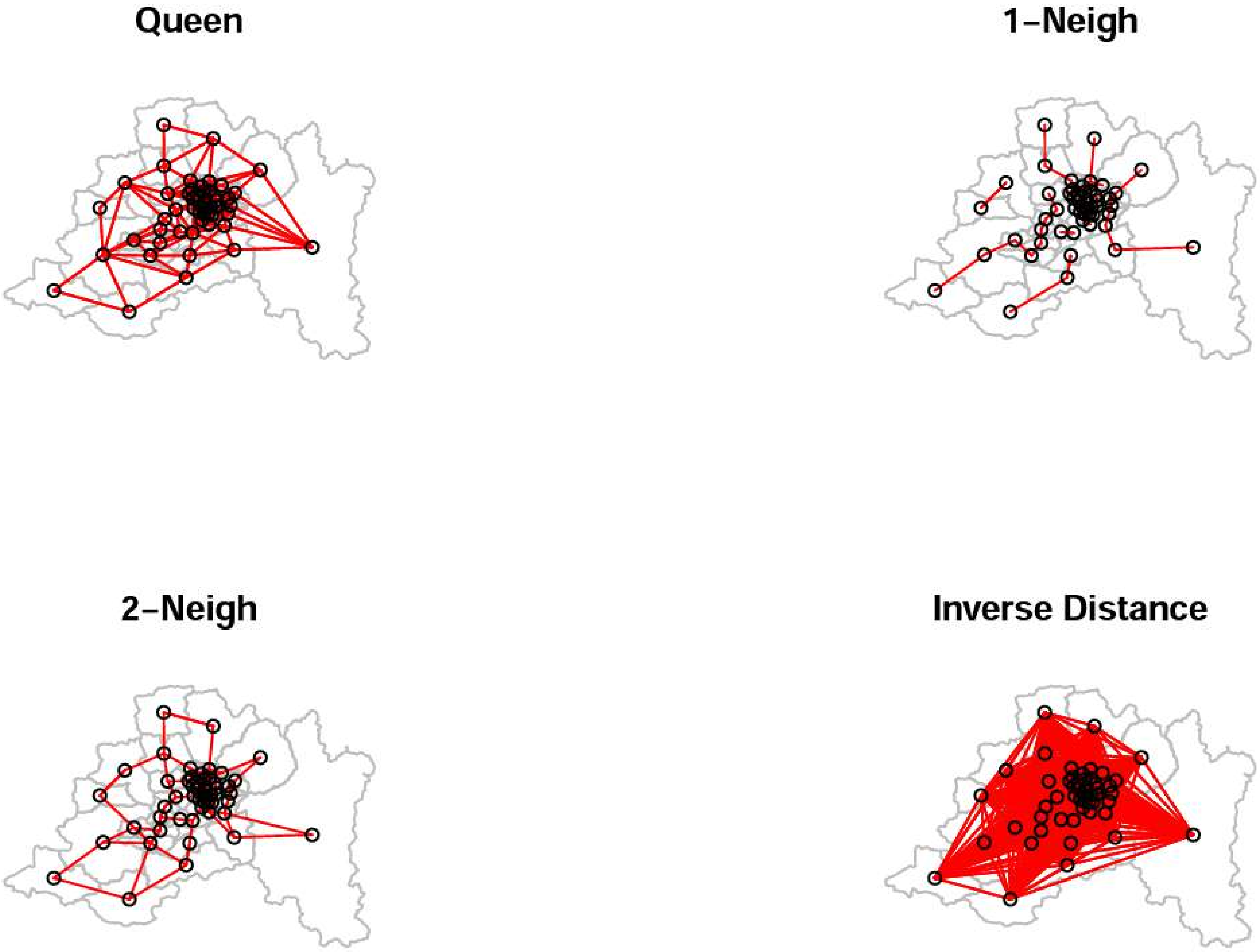
$$\mathbf{W}\mathbf{y} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 50 \\ 30 \end{pmatrix} = \begin{pmatrix} 50 \\ 10 + 30 \\ 50 \end{pmatrix} \tag{11}$$

Using a row-standardized weight matrix:

$$\mathbf{W}^s\mathbf{y} = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 50 \\ 30 \end{pmatrix} = \begin{pmatrix} 50 \\ 5 + 15 \\ 50 \end{pmatrix} \tag{12}$$

Therefore, when $\mathbf{W}$ is standardized, each element $(\mathbf{W}\mathbf{y})_i$ is interpreted as a weighted average of the y values for i's neighbors.

Ilustrasi



ILUSTRASI

- Contoh proses menentukan matriks pembobot spasial menggunakan R.

AUTOKORELASI SPASIAL

Autokorelasi Spasial

- Autocorrelation $\Longrightarrow$ the correlation of a variables with itself
  - Time series: the values of a variable at time t depends on the value of the same variable at time $t-1$.
  - Space: the correlation between the value of the variable at two different locations.

Definition (Spatial Autocorrelation)

- Correlation between the same attribute at two (or more) different locations.
- Coincidence of values similarity with location similarity.
- Under spatial dependency it is not possible to change the location of the values of certain variable without affecting the information in the sample.
- It can be positive and negative.

Autokorelasi Spasial

Definition (Positive Autocorrelation)

Observations with high (or low) values of a variable tend to be clustered in space.

Figure: Positive Autocorrelation

1	1	
1	1	

Autokorelasi Spasial

Definition (Negative Autocorrelation)

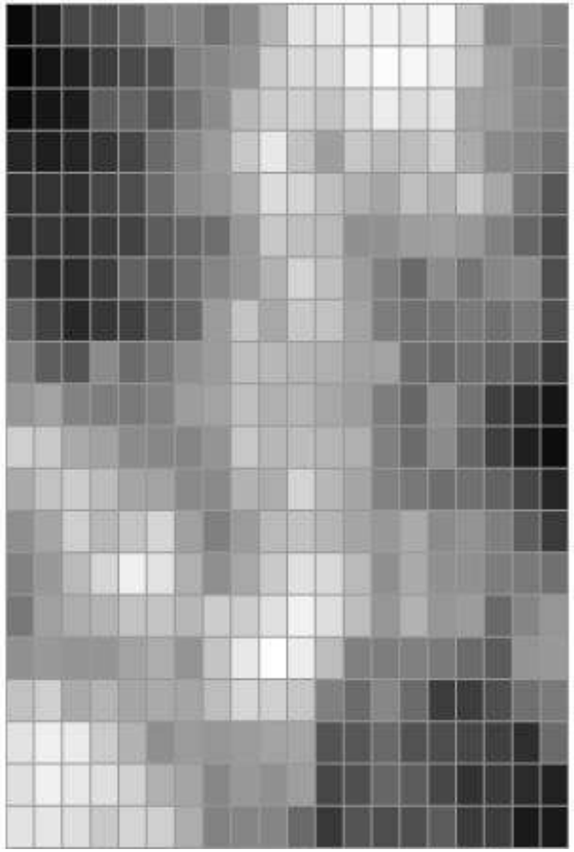
Locations tend to be surrounded by neighbors having very dissimilar values.

Figure: Negative Autocorrelation

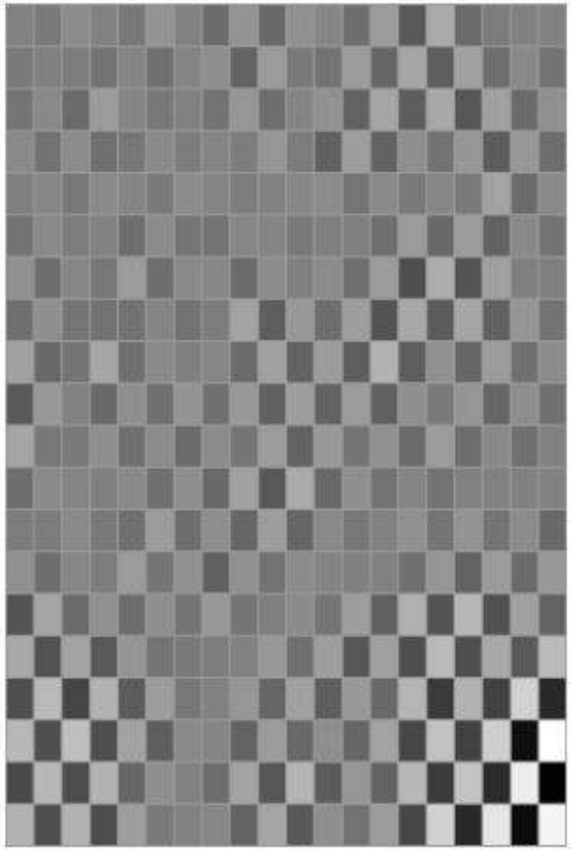
1		1
	1	
1		1

Autokorelasi Spasial

POSITIF



NEGATIF

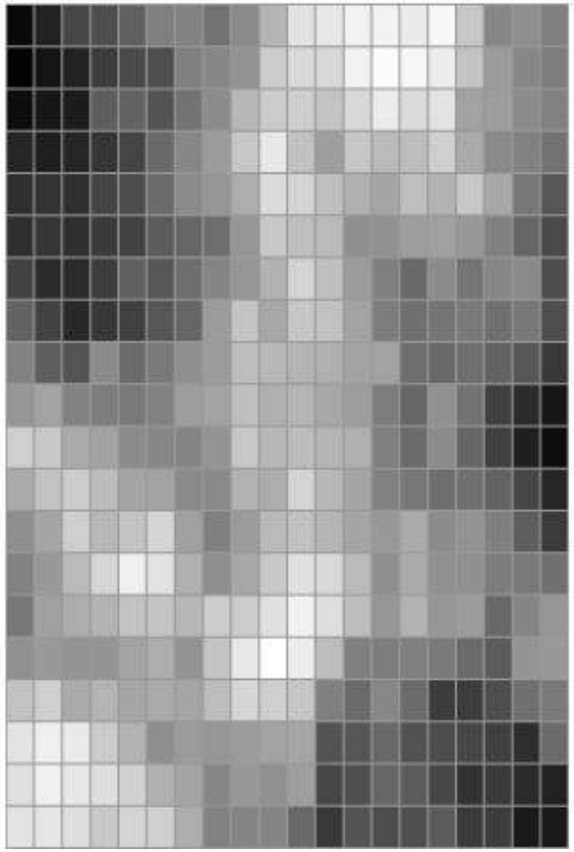


Definition (Spatial Randomness)

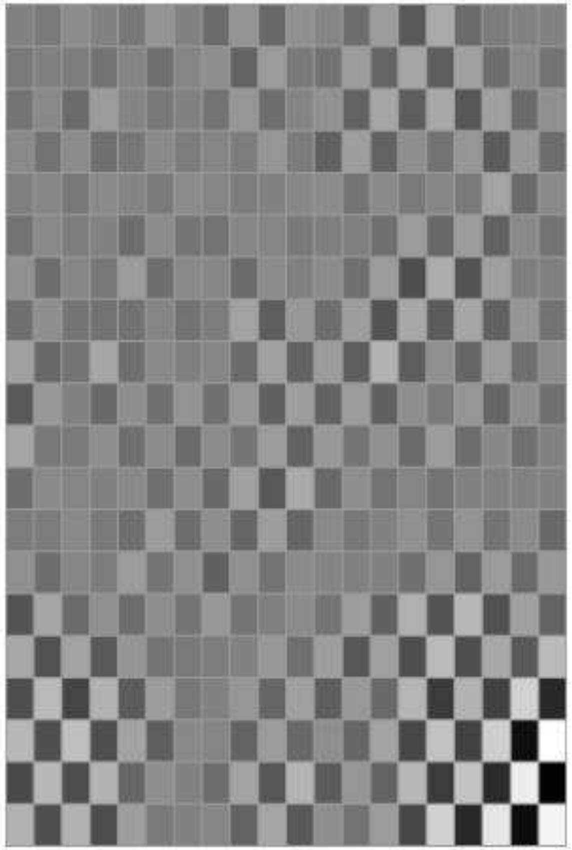
When none of the two situations occurs.

Autokorelasi Spasial

POSITIF



NEGATIF



Autokorelasi Spasial

Two main sources of spatial autocorrelation (Anselin, 1988):

- Measurement errors.
- Importance of Space.

The second source is of much more interest.

- The essence of regional sciences and new economic geography is that location and distance matter.
- What is observed at one point is determined by what happen elsewhere in the system.

Autokorelasi Spasial

Tobler's First Law of Geography

Everything depends on everything else, but closer things more so

- Important ideas:
  - Existence of Spatial Dependence.
  - Structure of Spatial Dependence
    - Distance decay.
    - Closeness = Similarities.

Autokorelasi Spasial

- Indicators of spatial association
  1. Global Autocorrelation
  2. Local Autocorrelation

Definition (Global Autocorrelation)

It is a measure of overall clustering in the data. It yields only one statistic to summarize the whole study area (Homogeneity).

1. Moran's I.
2. Gery's C.
3. Getis and Ord's $G(d)$

Definition (Local Autocorrelation)

A measure of spatial autocorrelation for each individual location.

- Local Indices for spatial Spatial Analysis (LISA)

Indeks Moran

This statistic is given by:

$$I = \frac{\sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} w_{ij} \left(x_i - \bar{x}\right) \left(x_j - \bar{x}\right)}{S_0 \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 / n} = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \bar{x}\right) \left(x_j - \bar{x}\right)}{S_0 \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \tag{15}$$

where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ and w_{ij} is an element of the spatial weight matrix that measures spatial distance or connectivity between regions i and j. In matrix form:

$$I = \frac{n}{S_0} \frac{\mathbf{z}'\mathbf{W}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \tag{16}$$

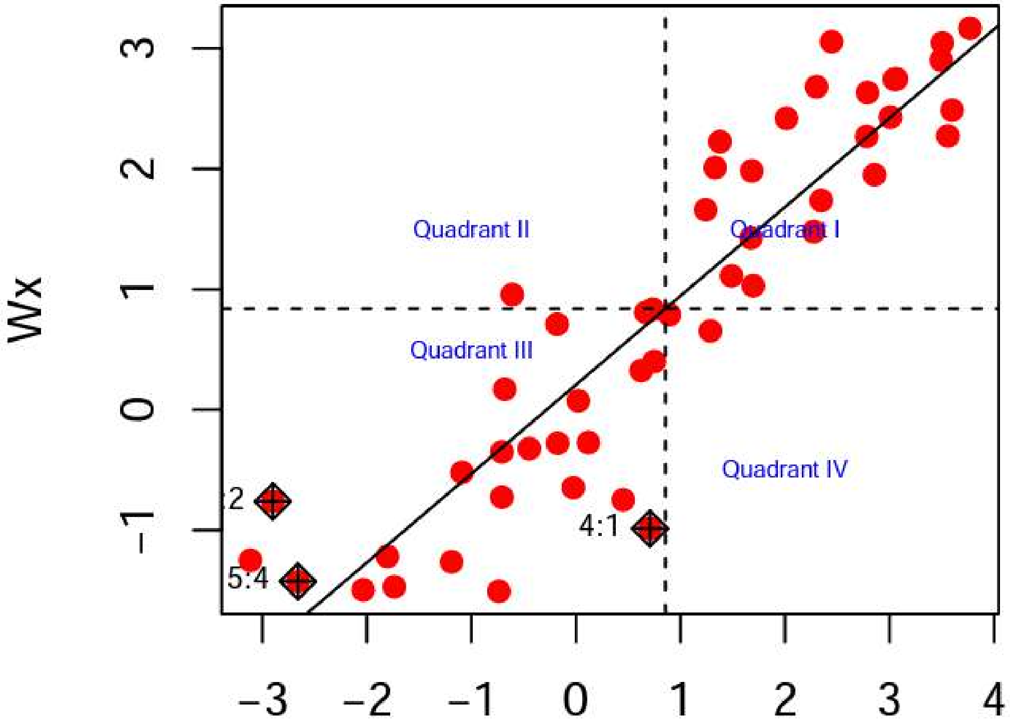
where $\mathbf{z} = \mathbf{x} - \bar{x}$. If the $\mathbf{W}$ matrix is row standardized, then:

$$I = \frac{\mathbf{z}'\mathbf{W}^s\mathbf{z}}{\mathbf{z}'\mathbf{z}} \tag{17}$$

because $S_0 = n$. Values range from -1 (perfect dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern.

Moran Scatterplot

- A very useful tool for understanding the Moran's I test



Indeks Moran

Note that:

$$\widehat{\beta}_{OLS} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

Therefore?

Remark

I is equivalent to the slope coefficient of a linear regression of the spatial lag $\mathbf{Wx}$ on the observation vector $\mathbf{x}$ measured in deviation from their means. It is, however, not equivalent to the slope of $\mathbf{x}$ on $\mathbf{Wx}$ which would be a more natural way.

Indeks Moran

- H_0: x is spatially independent, the observed x are assigned at random among locations. (I is close to zero)
- H_1: X is not spatially independent. (I is not zero)

Indeks Moran

- We are interested in the distribution of the following statistic:

$$T_I = \frac{I - \mathbb{E}(I)}{\sqrt{\mathrm{Var}(I)}} \tag{18}$$

- Three approaches to compute the variance of Moran's I:
  - Monte Carlo
  - Normality of x_i: It is assumed that the random variable x_i are the result of n independently drawings from a normal population.
  - Randomization of x_i: No matter what the underlying distribution of the population, we consider the observed values of x_i were repeatedly randomly permuted.

Indeks Moran

Theorem (Moran's I Under Normality)

Assume that $\{\mathbf{x}_i\} = \{x_1, x_2, ..., x_n\}$ *are independent and distributed as* $\mathrm{N}(\mu, \sigma^2)$, *but* μ *and* σ^2 *are unknown. Then:*

$$\mathbb{E}(I) = -\frac{1}{n-1} \tag{19}$$

and

$$\mathbb{E}\left(I^2\right) = \frac{n^2 S_1 - n S_2 + 3 S_0^2}{S_0^2 (n^2 - 1)} \tag{20}$$

where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$, $S_1 = \sum_{i=1}^{n} \sum_{j=1}^{n} (w_{ij} + w_{ji})^2 / 2$, $S_2 = \sum_{i=1}^{n} (w_{i.} + w_{.i})^2$, *where* $w_{i.} = \sum_{j=1}^{n} w_{ij}$ *and* $w_{i.} = \sum_{j=1}^{n} w_{ji}$ *Then:*

$$\mathrm{Var}(I) = \mathbb{E}\left(I^2\right) - \mathbb{E}(I)^2 \tag{21}$$

Indeks Moran

Theorem 17 gives the moments of Moran's I under randomization.

Theorem (Moran's I Under Randomization)

Under permutation, we have:

$$\mathbb{E}(I) = -\frac{1}{n-1} \tag{22}$$

and

$$\mathbb{E}\left(I^2\right) = \frac{n\left[\left(n^2-3n+3\right)S_1 - nS_2 + 3S_0^2\right] - b_2\left[\left(n^2-n\right)S_1 - 2nS_2 + 6S_0^2\right]}{(n-1)(n-2)(n-3)S_0^2} \tag{23}$$

where $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$, $S_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}+w_{ji})^2/2$, $S_2 = \sum_{i=1}^{n}(w_{i.}+w_{.i})^2$, *where* $w_{i.} = \sum_{j=1}^{n} w_{ij}$ *and* $w_{i.} = \sum_{j=1}^{n} w_{ji}$*. Then:*

$$\operatorname{Var}(I) = \mathbb{E}\left(I^2\right) - \mathbb{E}(I)^2 \tag{24}$$

It is important to note that the expected value of Moran's I under normality and randomization is the same.

Indeks Moran

- Normality and randomization? We can use a Monte Carlo simulation
  - To test a null hypothesis H_0 we specify a test statistic T such that large values of T are evidence against H_0.
    - H_0 : no spatial autocorrelation.
  - Let T have observed value t_{obs}. We generally want to calculate:

$$p = \Pr(T \geq t_{obs} | H_0) \tag{25}$$

  - We need the distribution of T when H_0 is true to evaluate this probability.

Indeks Moran

Theorem (Moran's' I Monte Carlo Test)

The procedure is the following:

1. *Rearrange the spatial data by shuffling their location and compute the Moran's I S times. This will create the distribution under* H_0.
2. *Let* $I_1^*, I_2^*, ..., I_S^*$ *be the Moran's I for each time. A consistent Monte Carlo p-value is then:*
$$\widehat{p} = \frac{1 + \sum_{s=1}^{S} 1(I_s^* \geq I_{obs})}{S+1} \tag{26}$$
3. *For tests at the* α *level or at* $100(1-\alpha)\%$ *confidence intervals, there are reasons for choosing* S *so that* $\alpha(S+1)$ *is an integer. For example, use* $S = 999$ *for confidence intervals and hypothesis tests when* $\alpha = 0.05$.

Indeks Moran

- Normality and randomization? We can use a Monte Carlo simulation
  - To test a null hypothesis H_0 we specify a test statistic T such that large values of T are evidence against H_0.
    - H_0 : no spatial autocorrelation.
  - Let T have observed value t_{obs}. We generally want to calculate:

$$p = \Pr(T \geq t_{obs} | H_0) \tag{25}$$

  - We need the distribution of T when H_0 is true to evaluate this probability.

Ilustrasi

- Silahkan klik tautan berikut:

https://rpubs.com/r_anisa/spatial-autocorrelation

Referensi

- Anselin L. 1988. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publisher.
- Bivand RS, Pebesma EJ, Rubio VG. 2008. Applied Spatial Data Analysis with R. New York (USA): Springer Science & Business Media, LLC.
- Sarrias, M. 2020. Lecture 1: Introduction to Spatial Econometric. (Lecture Slide) Universidad de Talca. Retrieved from https://www.msarrias.com/uploads/3/7/7/8/37783629/lecture1.pdf

Spatial Regression

6. Specification Spatial Heterogeneity

Luc Anselin

http://spatial.uchicago.edu

homogeneity and heterogeneity

spatial regimes

spatially varying coefficients

spatial random effects

Homogeneity and Heterogeneity

- Global Perspective

  single equilibrium - stationarity

  functional form fixed

  coefficients fixed

- Local Perspective

  multiple equilibria

  non-stationarity

  functional and/or parameter variability

- Extreme Homogeneity

  model same everywhere

  parameters same everywhere

  $y_i = x_i\beta + \varepsilon_i$

  β constant across i

  $\varepsilon_i \sim$ i.i.d. with $Var[\varepsilon_i] = \sigma^2$ for all i

- Extreme Heterogeneity

every observation is different

$y_i = x_i\beta_i + \varepsilon_i$

a different parameters β_i for each observation i

$\varepsilon_i \sim$ i.n.i.d. with $Var[\varepsilon_i] = \sigma_i^2$

possible different functional forms for i

- Incidental Parameter Problem

  number of unknown parameters increases with sample size

  no consistent estimation of individual parameters β_i, σ_i^2

- Solutions

  imposing structure

  discrete variation - finite subsets of the data

  continuous variation - parameter surface

  heterogeneity parameters

  fixed effects

  random effects

  spatial heterogeneity may be complicated by spatial autocorrelation

Spatial Regimes

Discrete Heterogeneity

- Spatial Regimes

  systematic discrete spatial subsets of the data

  different coefficient values in each subset

  corrects for heterogeneity, but does not explain

- Spatial Regime Specifications

    varying intercepts

        spatial ANOVA

        spatial fixed effects

    full spatial regimes

Varying Intercepts

- ANOVA - Difference in Means

approach is standard, regimes are spatial

$E[y_1] = \mu_1 \; \forall \; i \in R_1$ (R_1 is region 1)

$E[y_2] = \mu_2 \; \forall \; i \in R_2$ (R_2 is region 2)

$H_0: \mu_1 = \mu_2$

- Dummy Variable Regression - Variant 1

  no constant term

  indicator variable for each regime

  $y_i = \beta_1 d_{1i} + \beta_2 d_{2i} + \varepsilon_i$

  $d_{1(2)i} = 1 \ \forall \ i \in R_{1(2)}$, 0 elsewhere

  $H_0: \beta_1 = \beta_2$

- Dummy Variable Regression - Variant 2

constant term for overall mean

$y_i = \alpha + \beta d_i + \varepsilon_i$

$d_i = 1 \ \forall \ i \in R_1$, 0 elsewhere

H_0: $\beta = 0$, difference from reference mean α

- Spatial Fixed Effects

reference mean and difference by regime

$$y_{ij} = \alpha + \alpha_2 d_{i2} + \ldots + \alpha_G d_{iG} + \mathbf{x}_i' \beta + \varepsilon_{ij},$$

fixed effects multi-level specification

- Spatial Fixed Effects and Spatial Autocorrelation (Anselin and Arribas-Bel 2013)

  common misconception that spatial fixed effects “fix” spatial autocorrelation

  only in special case of group weights

  each observation has all other observations as neighbors

  so-called “Case” weights (Case 1992)

Full Spatial Regimes

- Spatial Regimes - Full Specification

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} & 0 \\ 0 & \mathbf{X_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix},$$

all coefficients (intercept, slope, variance) vary by regime
equivalent to separate regression by regime

Testing for Spatial Heterogeneity

- Test on Spatial Homogeneity

    null hypothesis

        equal intercepts, equal slopes

    alternative hypothesis

        different intercepts

        different slopes

        both

- Chow Test

test on structural stability

based on residual sum of squares in constrained (all coefficients equal - R) and unconstrained (coefficients different - U) regressions

classic form

$$C = \frac{e'_R e_R - e'_U e_U}{k} / \frac{e'_U e_U}{N - 2k} \sim F(k, N - 2k)$$

- General Test on Coefficient Stability

as a set of linear constraints on the coefficients in a pooled regression

can be readily extended to spatial models

$$\mathbf{R}\beta = 0,$$

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{bmatrix} = 0.$$

V = variance

$$(\mathbf{R}\hat{\beta})'[\mathbf{RVR}']^{-1}(\mathbf{R}\hat{\beta}) \sim \chi^2(G),$$

G = (J - 1)K
J regimes
K coefficients

Spatial Regimes with Spatial Dependence

- Spatial Lag and Spatial Error Models
    - allow varying coefficients by regime
    - fixed spatial coefficient
        - same spatial process throughout
    - varying spatial coefficient
        - different spatial process for each regime
        - difficult assumption - needs to be based on a strong foundation

- Spatial Regimes - Spatial Lag Model

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} = \rho \mathbf{W} \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} + \begin{bmatrix} \mathbf{X_1} & 0 \\ 0 & \mathbf{X_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.$$

fixed spatial autoregressive coefficient

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} = \begin{bmatrix} \rho_1 \mathbf{W_1} & 0 \\ 0 & \rho_2 \mathbf{W_2} \end{bmatrix} \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} + \begin{bmatrix} \mathbf{X_1} & 0 \\ 0 & \mathbf{X_2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix},$$

varying spatial autoregressive coefficient

- Spatial Regimes - Spatial Error Model

$$\begin{bmatrix} \epsilon_{\mathbf{1}} \\ \epsilon_{\mathbf{2}} \end{bmatrix} = \lambda \mathbf{W} \begin{bmatrix} \epsilon_{\mathbf{1}} \\ \epsilon_{\mathbf{2}} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}.$$

fixed spatial autoregressive coefficient

$$\begin{bmatrix} \epsilon_{\mathbf{1}} \\ \epsilon_{\mathbf{2}} \end{bmatrix} = \begin{bmatrix} \lambda_1 \mathbf{W_1} & 0 \\ 0 & \lambda_2 \mathbf{W_2} \end{bmatrix} \begin{bmatrix} \epsilon_{\mathbf{1}} \\ \epsilon_{\mathbf{2}} \end{bmatrix} + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix},$$

varying spatial autoregressive coefficient

- Spatial Weights Specification

    necessary to construct spatially lagged variables

        neighbors spill over across regimes

        neighbors constrained to be within each regime

            weights truncated, possible isolates

$$\mathbf{W_R} = \begin{bmatrix} \mathbf{W_1} & 0 & \dots & 0 \\ 0 & \mathbf{W_2} & \dots & 0 \\ & \dots & \dots & \\ 0 & \dots & 0 & \mathbf{W_J} \end{bmatrix},$$

- Spatial Chow Test

use general form of the test with V as coefficient variance matrix in pooled model

$$(\mathbf{R}\hat{\beta})'[\mathbf{RVR}']^{-1}(\mathbf{R}\hat{\beta}) \sim \chi^2(G),$$

Spatially Varying Coefficients

- Spatially Varying Coefficients

    systematic variation with covariates

        coefficient as a function of other variables (including as a trend surface)

        spatial expansion method

    local estimation over space

        coefficients obtained from a subset (kernel) of nearby data points

        geographically weighted regression (GWR)

Expansion Method

- Casetti's Expansion Method

    special case of varying coefficients

    each coefficient is a function of other covariates

        creates interaction effects

    similar in form to multi-level models

- Sequential Modeling Strategy

  initial model

  $$y_i = \alpha + x_i\beta_i + \varepsilon_i$$

  expansion equation

  $$\beta_i = \gamma_0 + z_{i1}\gamma_1 + z_{i2}\gamma_2$$

  final model

  $$y_i = \alpha + x_i\,(\gamma_0 + z_{i1}\gamma_1 + z_{i2}\gamma_2) + \varepsilon_i$$

  $$y_i = \alpha + x_i\gamma_0 + (z_{i1}x_i)\gamma_1 + (z_{i2}x_i)\gamma_2 + \varepsilon_i$$

- Implementation Issues
  - multicollinearity
    - t-test values unreliable
    - various fixes
      - principal components (orthogonal expansion)
  - danger of overfitting

- Random Expansion Model

random error in expansion equation

$$\beta_i = \gamma_0 + z_{i1}\gamma_1 + z_{i2}\gamma_2 + \psi_i$$

error term in final model is heteroskedastic

$$y_i = \alpha + x_i\,(\gamma_0 + z_{i1}\gamma_1 + z_{i2}\gamma_2 + \psi_i) + \varepsilon_i$$

$$\nu_i = x_i\psi_i + \varepsilon_i$$

$$Var[\nu_i] = x_i^2\,\sigma^2_\psi + \sigma^2_\varepsilon$$

similar to random coefficient model and multilevel models

Geographically Weighted Regression

- Geographically Weighted Regression

  local regression

  a different set of parameter values for each location

  parameter values obtained from a subset of observations using kernel regression

- Local Regression

non-parametric specification

simple bivariate regression

$y_i = m(x_i) + u_i$

functional form of m is unspecified

$m(x_i)$ yields the conditional expectation of $y \mid x$

- Local Average

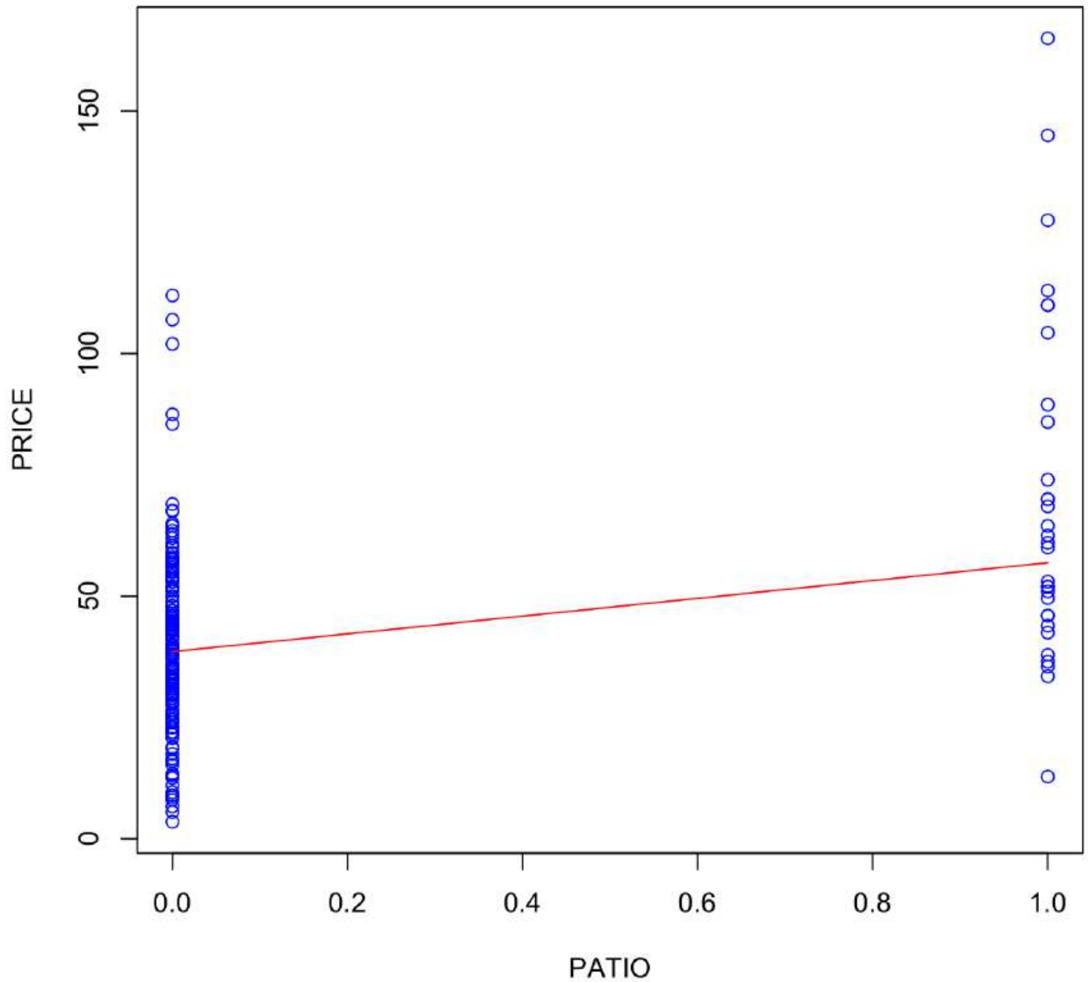
what is the expected value of y_i given x, $E[y_i \mid x]$

special case: for a given x_0 with multiple y_i

example: two values for PATIO dummy, house price

solution:

take $m(x_0)$ as the average of y_i for $x_0 = 0$ and $x_0 = 1$



local average

predictor of PRICE for two values of PATIO

- Locally Weighted Average

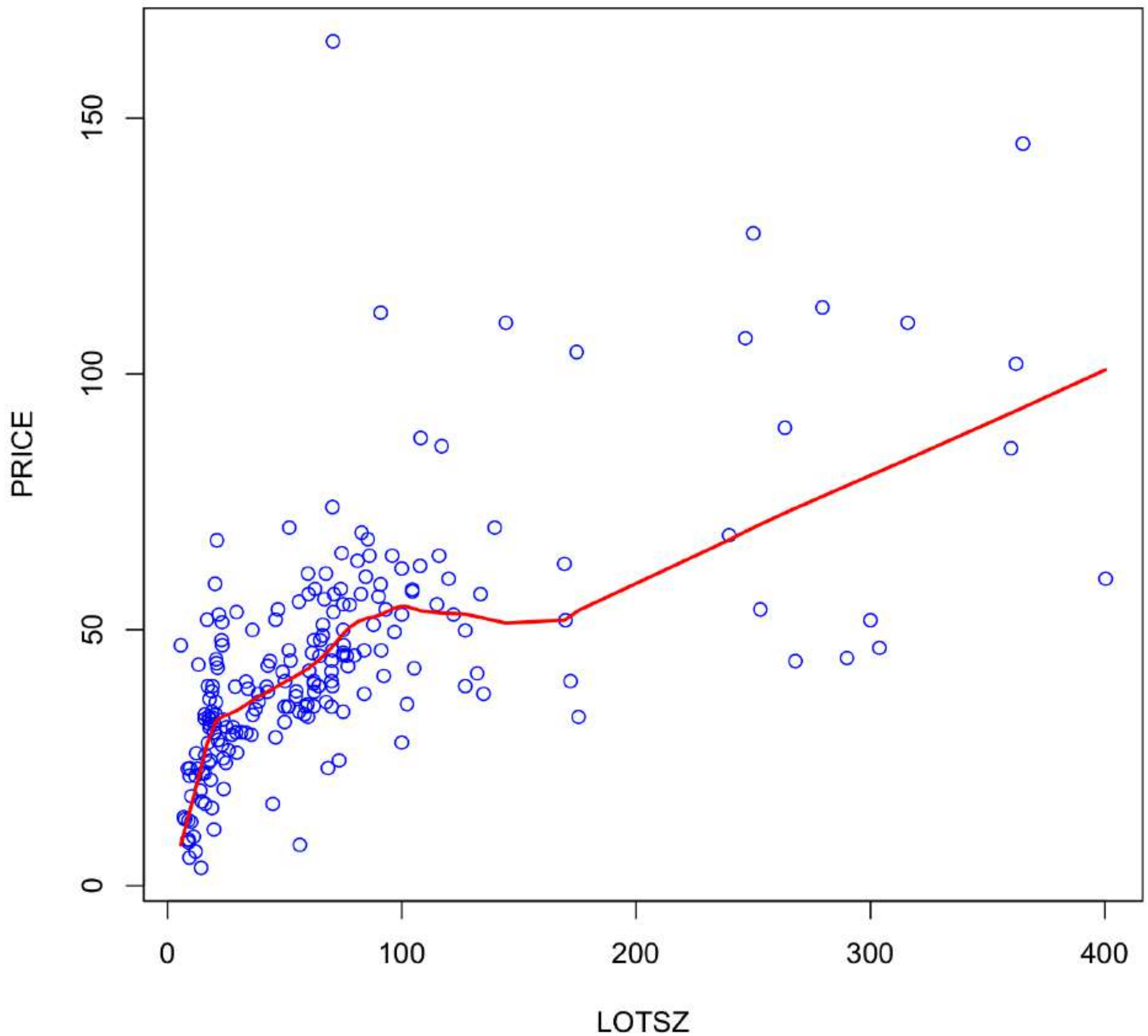
expand the estimate of $m(x_0)$ to include values of y_i observed for values of x "close" to x_0

compute a locally weighted average

weights sum to one

weights larger as x closer to x_0 (for h $\downarrow$)

$m(x_0) = \sum_i w_{i0,h}\, y_i$



locally weighted average (lowess) of PRICE for LOTSZ

Kernel Regression

special case of locally weighted average

use kernel function as the weights

$m(x_0) = \sum_i K\,[(x_i - x_0)/h\,]y_i$

K is kernel function

h is bandwidth s.t. $K = 0$ for $x_i - x_0 > h$

- Kernel Functions with Finite Bandwidth

Epanechnikov

$$K(z) = 1 - z^2$$

Bisquare

$$K(z) = (1 - z^2)^2$$

with $z = (x_i - x_0) / h$

- Gaussian Kernel

asymptotic bandwidth

specified in function of standard error or variance

$$K(z) = \exp(- z^2/2)$$

- GWR Estimation

  local estimation based on nearby locations

  not just y_i but x-y pairs at nearby locations

  kernel regression yields a different coefficient for each location

  specify kernel function and bandwidth

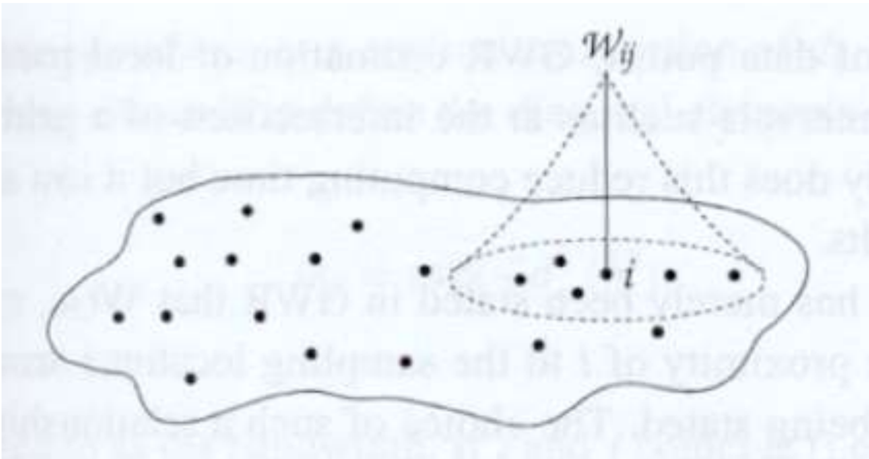
- GWR Kernel Regression

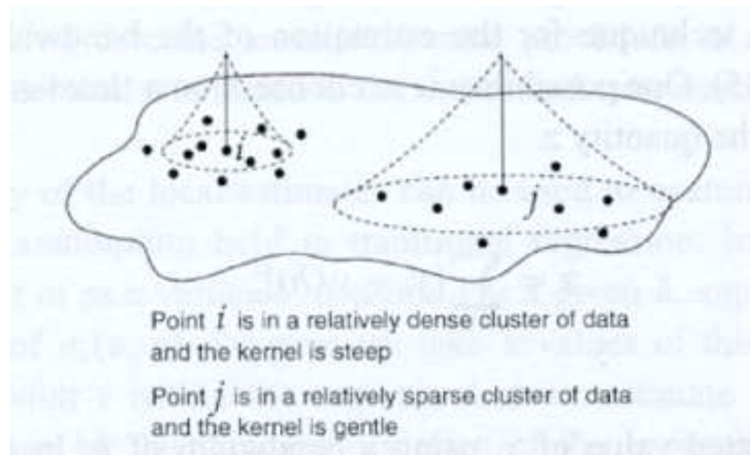
    location-specific kernel weights

        $W(u_i, v_i)$ diagonal elements are weights

        $b(u_i,v_i) = [X'W(u_i,v_i)X]^{-1}X'W(u_i,v_i)y$

    fixed kernel vs adaptive kernel





fixed bandwidth kernel

adaptive kernel

Source: Fotheringham et al (2002)

- GWR - Practical Issues

  choice of bandwidth

    use cross-validation

  parameter inference

    still several theoretical loose ends

  visualizing parameter heterogeneity

Spatial Random Effects

Random Coefficients

- Random Coefficient Regression

extreme heterogeneity, but variability in β_i driven by a random process - no space

$\beta_i = \beta + \psi_i$ with $E[\psi_i]=0$ and $Var[\psi_i]=\sigma^2$

heteroskedastic regression for mean effect

$y_i = \alpha + x_i\beta + \nu_i$, $var[\nu_i] = \sigma^2_\psi x_i^2 + \sigma^2_\varepsilon$

- Mixed Linear Models

  both fixed and random coefficients

  $y = X\beta + Z\psi + \varepsilon$

  Z a "design matrix", could be same as X

  ψ random coefficients with mean zero and variance Σ_ψ

  ε random error vector with variance Σ_ε

- Spatial Random Coefficients

introduce spatial dependence structure in random variation of coefficient

$\beta_i - \beta = \rho \, \Sigma_j \, w_{ij} \, (\beta_j - \beta) + \psi_i$ - SAR model

$\beta_i = \beta + \lambda \, \Sigma_j \, w_{ij} \, \psi_j + \psi_i$ - SMA model

complex covariance structures

Spatial Random Effects

- **Spatial Random Effects**

  $\beta_i = \beta + \psi_i$ with spatial effects introduced through random effect ψ_i

  typically a CAR process

  Bayesian hierarchical model - BYM model

  $\beta_i = \beta + \psi_i + \nu_i$

  spatial dependence in ψ_i, heterogeneity in ν_i

  not identified in Gaussian (linear regression) model

- Example - Poisson Regression

    spatial autocorrelation needs to be introduced indirectly

        auto-Poisson model only allows negative spatial autocorrelation

    random effects model

- Poisson Regression

  $P[Y = y] = e^{-\mu} \mu^y / y!$

  μ is the mean

  μ as a function of regressors to model heterogeneity

  $\mu_i = \exp(x_i'\beta)$ no error term

  random effects

  $\mu_i = \exp(x_i'\beta + \psi_i + \nu_i)$

  spatial effects through ψ_i, e.g, CAR model

  non-spatial heterogeneity through ν_i