Spatial Statistics

Geostatistical Data

Satyaki

June 29, 2022

Overview

1. Introduction
2. Formalization
3. Exploratory analysis
4. Semivariogram
5. Kriging

A study on brain cancer

- Navarre is a province in Spain.
- It is divided in 40 health districts.

A study on brain cancer

- Navarre is a province in Spain.
- It is divided in 40 health districts.
- A study on the number of deaths due to brain cancer for years 1988-1994 was done in 2006 (Ugarte et al. (2006)).

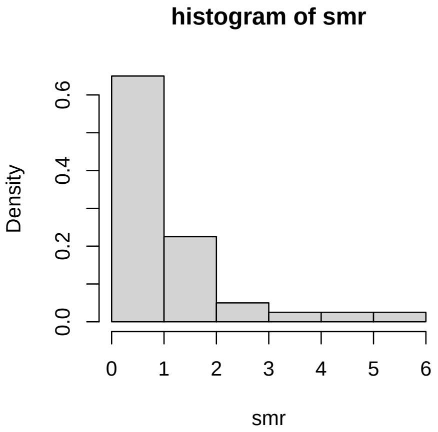
A study on brain cancer

- Navarre is a province in Spain.
- It is divided in 40 health districts.
- A study on the number of deaths due to brain cancer for years 1988-1994 was done in 2006 (Ugarte et al. (2006)).
- In each district, there is standardized morbidity ratio (SMR). This is calculated based on the observed data in a district.
- The data is given as Navarre.RData in the statistical software R.

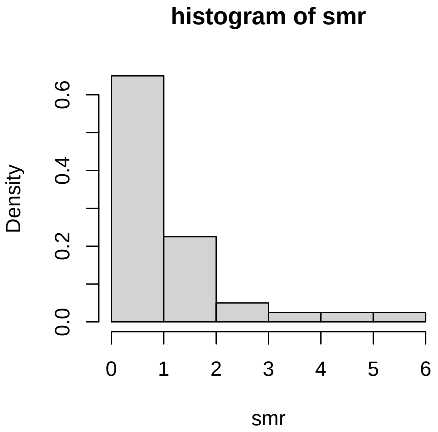
Table: Glimps of the data

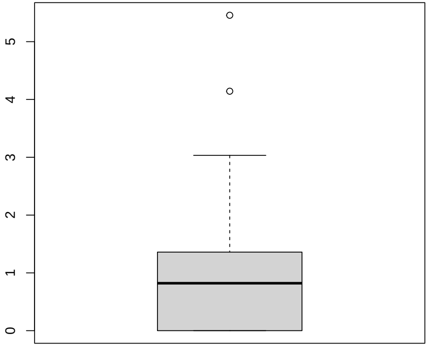
NAME	SMR
ALLO	1.8709074
ALSASUA	0.0000000
ARTAJONA	5.3980815
⋮	⋮

Visualizing the data

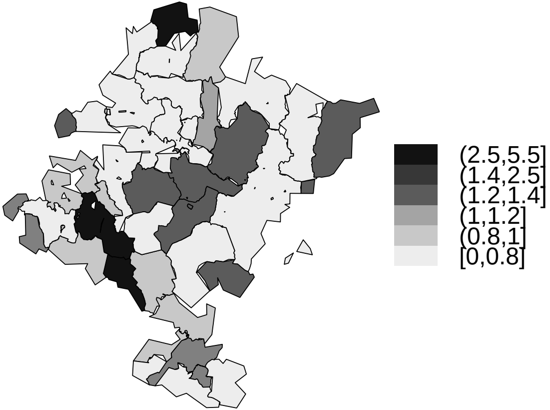


Visualizing the data





Spatial diagram



A study on percentage of population working outside

- It is a study on 56 counties of Scotland.

A study on percentage of population working outside

- It is a study on 56 counties of Scotland.
- The purpose of the study was to find out relationship between Lip cancer and the percentage of people working out in the sun (Clayton and Kaldor (1987)).

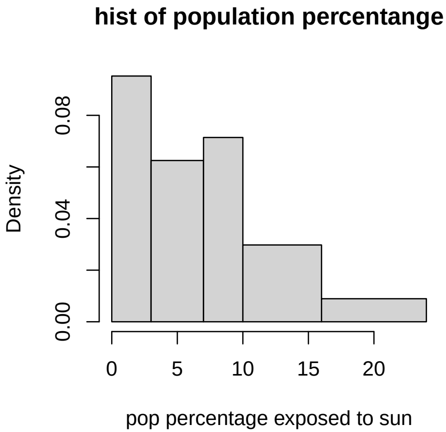
A study on percentage of population working outside

- It is a study on 56 counties of Scotland.
- The purpose of the study was to find out relationship between Lip cancer and the percentage of people working out in the sun (Clayton and Kaldor (1987)).
- We just have taken a part which is percentage of population working in the sun from the complete data set.

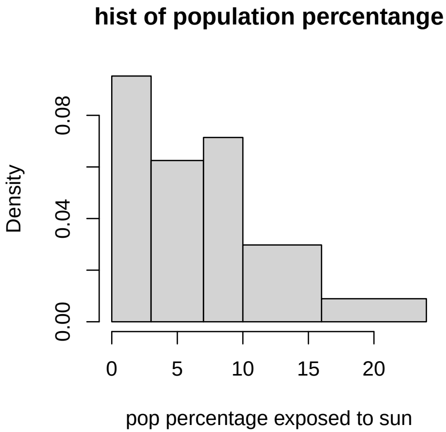
A study on percentage of population working outside

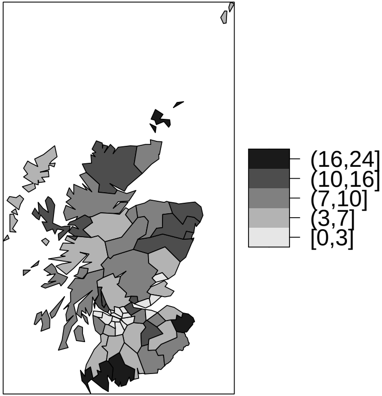
- It is a study on 56 counties of Scotland.
- The purpose of the study was to find out relationship between Lip cancer and the percentage of people working out in the sun (Clayton and Kaldor (1987)).
- We just have taken a part which is percentage of population working in the sun from the complete data set.
- The data is available as LipCancer.RData in the statistical software R.

Visualization



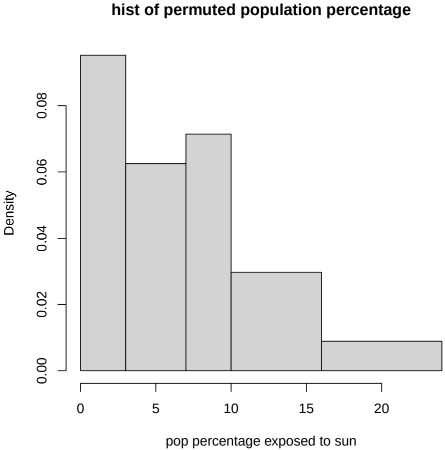
Visualization





Is there anything special about spatial statistics?

Is there anything special about spatial statistics?



Let us go back to Sir Fisher

- In his seminal work "The Design of Experiments, (1966)", spatial consideration was implicit.

Let us go back to Sir Fisher

- In his seminal work "The Design of Experiments, (1966)", spatial consideration was implicit.
- R.A. Fisher was employed at Rothamsted between 1919 and 1933.

Let us go back to Sir Fisher

- In his seminal work "The Design of Experiments, (1966)", spatial consideration was implicit.
- R.A. Fisher was employed at Rothamsted between 1919 and 1933.
- Majority of his time was engaged in developing of methodology for analysis of data arising from agricultural field trails.

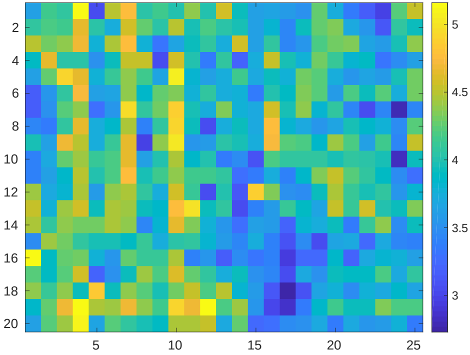
Let us go back to Sir Fisher

- In his seminal work "The Design of Experiments, (1966)", spatial consideration was implicit.
- R.A. Fisher was employed at Rothamsted between 1919 and 1933.
- Majority of his time was engaged in developing of methodology for analysis of data arising from agricultural field trails.
- A data set was collected analyzed by Mercer (1911) that dealt with the wheat production at Rothamsted.

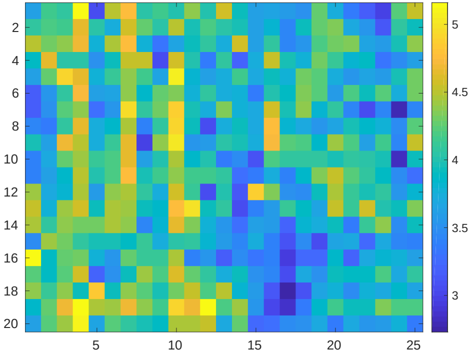
Let us go back to Sir Fisher

- In his seminal work “The Design of Experiments, (1966)”, spatial consideration was implicit.
- R.A. Fisher was employed at Rothamsted between 1919 and 1933.
- Majority of his time was engaged in developing of methodology for analysis of data arising from agricultural field trails.
- A data set was collected analyzed by Mercer (1911) that dealt with the wheat production at Rothamsted.
- Fisher encountered this type of data while working at Rothamsted.

Mercer's wheat data

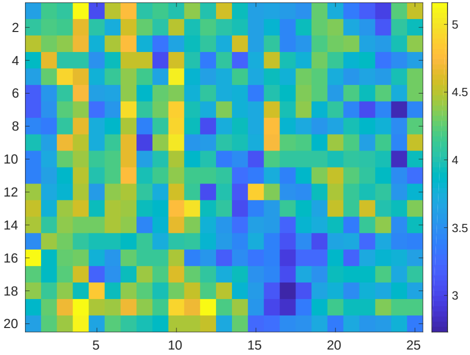


Mercer's wheat data



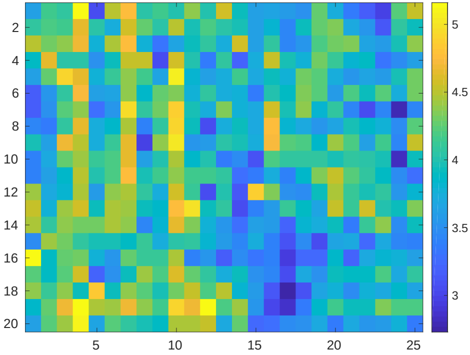
- Each square denotes a plot of size 3.30 m $\times$ 2.59 m.

Mercer's wheat data



- Each square denotes a plot of size 3.30 m $\times$ 2.59 m.
- Fisher commented: "*the widely verified fact that patches in close proximity are commonly more alike, . . ., than those which are farther apart*"

Mercer's wheat data



- Each square denotes a plot of size 3.30 m $\times$ 2.59 m.
- Fisher commented: "*the widely verified fact that patches in close proximity are commonly more alike, . . ., than those which are farther apart*"
- He proposed blocking to tackle spatial variation: a form of covariate adjustment under the implicit assumption that systematic spatial variation, if it exists at all, is piecewise constant within blocks.

Other than blocking

- Long time back, Papadikas (1937) had other idea about agricultural data analysis.

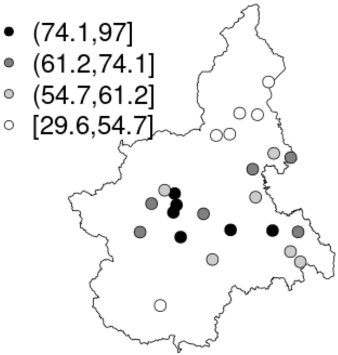
Other than blocking

- Long time back, Papadikas (1937) had other idea about agricultural data analysis.
- That essentially leads to *Markov Random Field*.
- In modern language, his proposal is similar to a conditional model for the distribution of the yield of the each plot, given the average yield, when the average is taken over "neighbouring" plots.

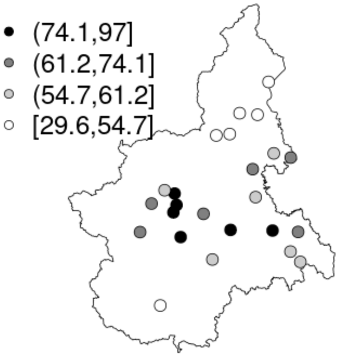
Other than blocking

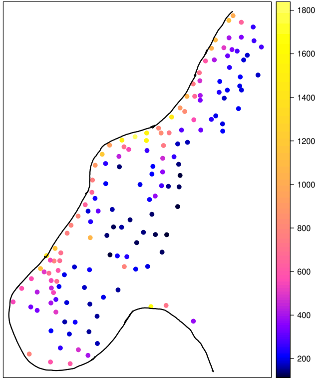
- Long time back, Papadikas (1937) had other idea about agricultural data analysis.
- That essentially leads to *Markov Random Field*.
- In modern language, his proposal is similar to a conditional model for the distribution of the yield of the each plot, given the average yield, when the average is taken over "neighbouring" plots.
- Thus, the moral: "**everything is related to everything else, but near things are more related than distant things**". Tobler (1979)

Different types of spatial data

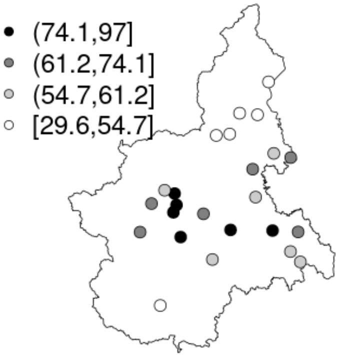


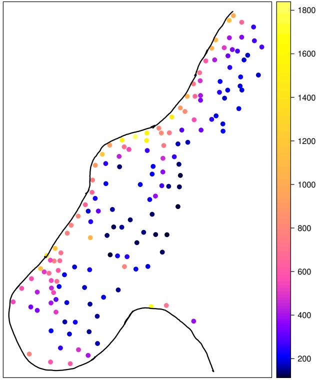
Different types of spatial data





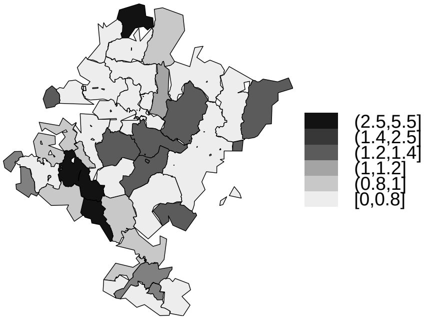
Different types of spatial data

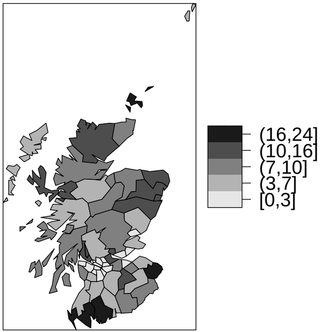




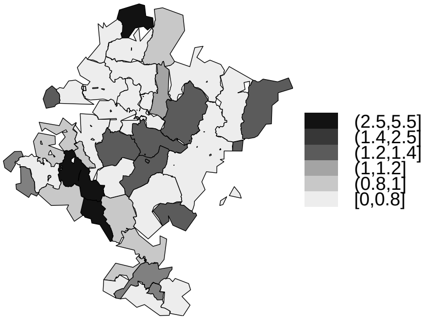
Geostatistical Data or point referenced data

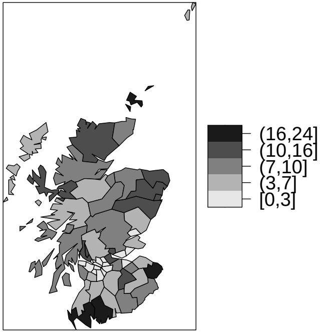
Second type of spatial data





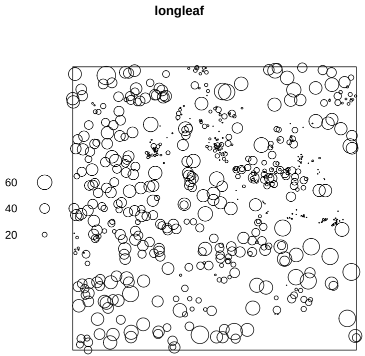
Second type of spatial data



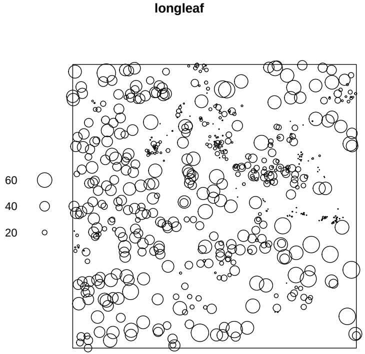


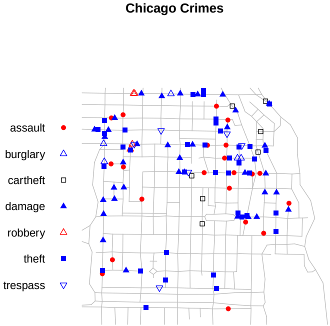
Areal data

Third type of spatial data



Third type of spatial data





Point Pattern Data

Formal definitions

Formal definitions

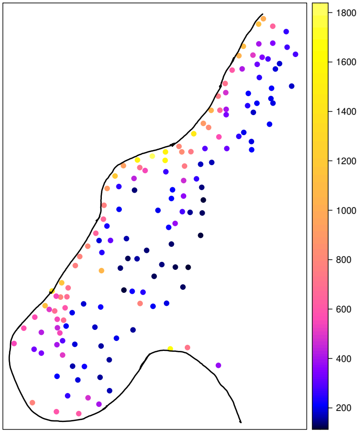
- The spatial observations comes from a spatial process $\{X_s, s \in \mathbb{S}\}$;

Formal definitions

- The spatial observations comes from a spatial process $\{X_s, s \in \mathbb{S}\}$;
- *Geostatistical Data* or **Point referenced data**: s varies continuously in a fixed subset $\mathbb{S}$ of $\mathbb{R}^2$, which contains a two dimensional rectangle of positive volume. X_s is observed at fixed sites $\{s_1, \ldots, s_n\} \subset \mathbb{S}$.

Formal definitions

- *Geostatistical Data* or **Point referenced data**: $s_1, s_2, \ldots, s_n$ are fixed and $X_{s_1}, X_{s_2}, \ldots, X_{s_n}$ are random variables.
- Goal: modeling, identification and separation of small and large scale variations, prediction (or kriging) at unobserved sites.

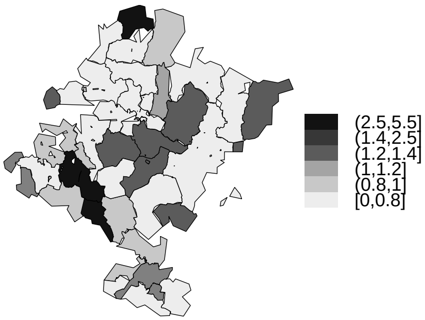


Formal definitions

- *Geostatistical Data* or **Point referenced data**: s varies continuously in a fixed subset $\mathbb{S}$, which contains a two dimensional rectangle of positive volume. X_s is observed at fixed sites $\{s_1, \ldots, s_n\} \subset \mathbb{S}$.
- *Lattice data* or **areal data**: $\mathbb{S}$ is fixed collection of countably many points.

Formal definitions

- *Lattice data* or **areal data**: Here $s_1, s_2, \ldots, s_n$ denote the blocks and the observation $X_{s_1}, X_{s_2}, \ldots, X_{s_n}$ are random observations.
- Goal: Constructing models, quantifying spatial correlations.

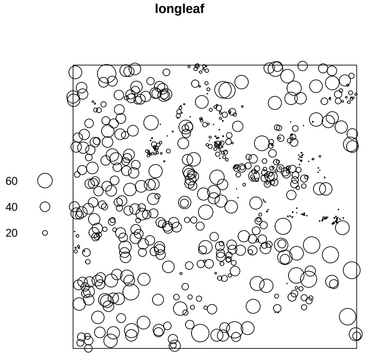


Formal definitions

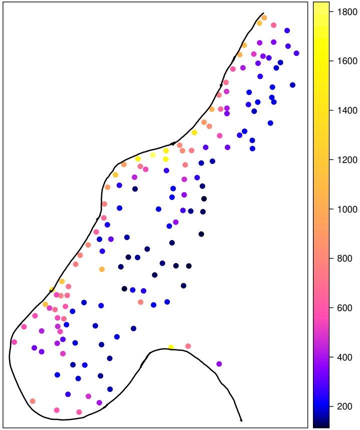
- The spatial observations comes from a spatial process $\{X_s, s \in \mathbb{S}\}$; Generally, $\mathbb{S} \subset \mathbb{R}^2$.
- *Geostatistical Data* or **Point referenced data**: s varies continuously in a fixed subset $\mathbb{S}$, which contains a two dimensional rectangle of positive volume. Z_s is observed at fixed sites $\{s_1, \ldots, s_n\} \subset \mathbb{S}$.
- *Lattice data* or **areal data**: $\mathbb{S}$ is fixed and partitioned into a finite number of blocks with clearly defined boundaries.
- **Point pattern data**: Here the observation sites $\{s_1, \ldots, s_n\}$ is random and the number of observation site n is also random. X_s can simply be equal to 1 indicating the occurrence of an event. Additionally, we may have some covariate information at these locations.

Formal definitions

- **Point pattern data**: Here the observation sites $\{s_1, \ldots, s_n\}$ is random and the number of observation site n is also random. Z_s can simply be equal to 1 indicating the occurrence of an event. Additionally, we may have some covariate information at these locations.
- Goal: Decide whether distribution of points are regular or completely random or aggregated?

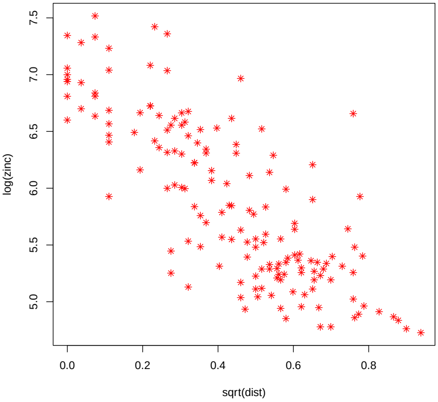


Exploratory analysis on point referenced data

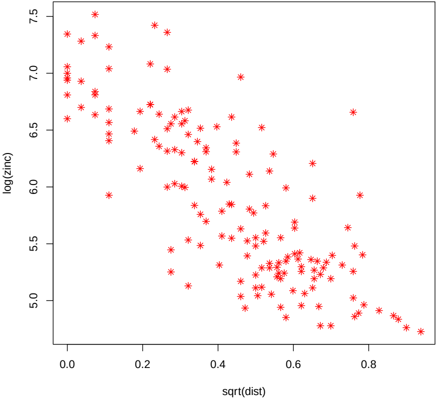


- Part of river Meuse in Netherlands
- Zinc concentration measurements
- Collected in a flood plain
- The concentration seems to be decreasing as distance increases from the river

Scatter plot and simple linear regression

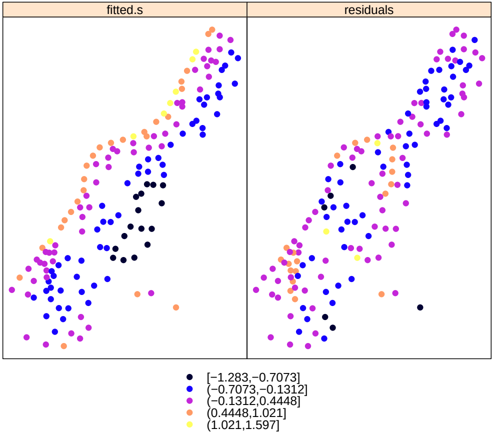


Scatter plot and simple linear regression

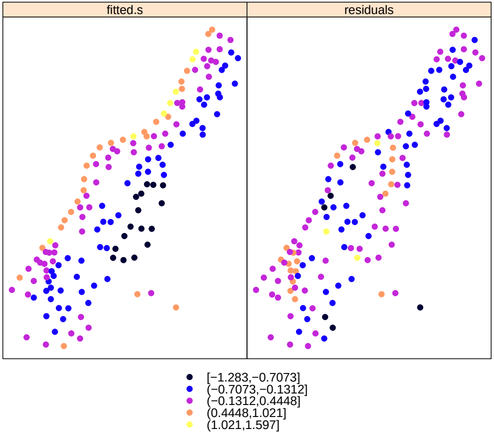


- Clearly correlated.
- A simple linear regression can be tested.
- $y = a_0 + a_1 x + \epsilon$

Fitted and the residuals

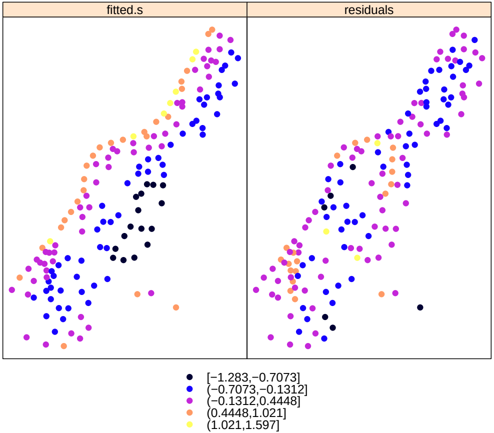


Fitted and the residuals



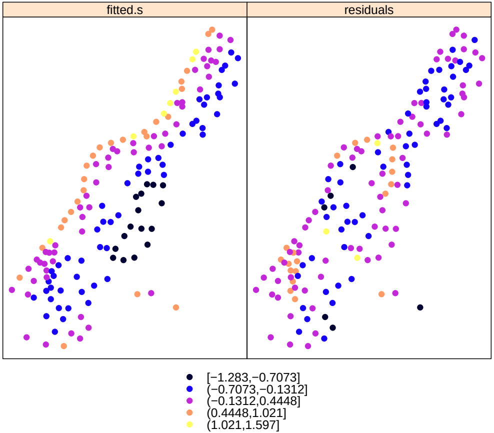
- A large part of variability is being taken care of.

Fitted and the residuals



- A large part of variability is being taken care of.
- However, the residuals do not seems to be spatially uncorrelated.

Fitted and the residuals



- A large part of variability is being taken care of.
- However, the residuals do not seems to be spatially uncorrelated.
- More analysis, taking the spatial structure into the account, required.

Covariogram

- Stationarity (weak stationarity) of the process $\{X(s), s \in \mathbb{S}\}$.

Covariogram

- Stationarity (weak stationarity) of the process $\{X(s), s \in \mathbb{S}\}$.
- Weak stationary essentially mean that the covariance relationship between the values of the process at any two locations can be summarized by a function $C(\mathbf{h})$, depending on the separation vector $\mathbf{h} = s_2 - s_1$.

- $E(X(s)) = \mu$, $\forall s \in \mathbb{S}$
- $\text{Cov}(X(s_1), X(s_2)) = C(s_2 - s_1) = C(\mathbf{h})$

Covariogram

- Stationarity (weak stationarity) of the process $\{X(s), s \in \mathbb{S}\}$.
- Weak stationary essentially mean that the covariance relationship between the values of the process at any two locations can be summarized by a function $C(\mathbf{h})$, depending on the separation vector $\mathbf{h} = s_2 - s_1$.

- $E(X(s)) = \mu$, $\forall s \in \mathbb{S}$
- $\text{Cov}(X(s_1), X(s_2)) = C(s_2 - s_1) = C(\mathbf{h})$
- If $C(\mathbf{h}) = \psi(||\mathbf{h}||)$, then the covariance function is called isotropic.

Covariance function is also known as **covariogram**.

Important Properties

- $|C(\mathbf{h})| \leq C(0)$.
- $C(\mathbf{h})$ is positive semidefinite.
- If $C(\mathbf{h})$ is continuous at the origin then it is continuous everywhere.

Intrinsic stationarity (Mathéron (1962))

- $E(X(s_i) - X(s_j)) = 0$ and
- $E(X(s_i) - X(s_j))^2 = 2\gamma(s_j - s_i) = 2\gamma(\mathbf{h})$

Intrinsic stationarity (Mathéron (1962))

- $E(X(s_i) - X(s_j)) = 0$ and
- $E(X(s_i) - X(s_j))^2 = 2\gamma(s_j - s_i) = 2\gamma(\mathbf{h})$
- Then the process is said to be intrinsic stationary.
- $\gamma(\mathbf{h})$ is called semivariogram.
- If the process is stationary, then $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$.
- If $\gamma(\mathbf{h}) = \phi(||\mathbf{h}||)$, then the semivariogram is called isotropic.

Important Properties

- $\gamma(\mathbf{h}) \geq 0$
- $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$
- $\gamma(\mathbf{h})$ is conditionally negative definite.

Reminder: γ(h) is property of difference

- Let $X(s) = Z(s) + V$, where $Z(s)$ iid normal variables and V is Cauchy.

Reminder: γ(h) is property of difference

- Let $X(s) = Z(s) + V$, where $Z(s)$ iid normal variables and V is Cauchy.
- Importantly, we may have situation where

Reminder: γ(h) is property of difference

- Let $X(s) = Z(s) + V$, where $Z(s)$ iid normal variables and V is Cauchy.
- Importantly, we may have situation where
  - the difference $X(s) - X(s')$ has a proper density

Reminder: γ(h) is property of difference

- Let $X(s) = Z(s) + V$, where $Z(s)$ iid normal variables and V is Cauchy.
- Importantly, we may have situation where
  - the difference $X(s) - X(s')$ has a proper density
  - even the $X(s)|X(s')$ and $X(s')|X(s)$ have proper density

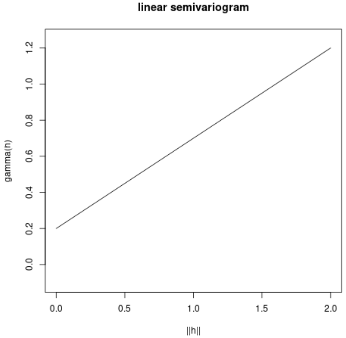
Reminder: γ(h) is property of difference

- Let $X(s) = Z(s) + V$, where $Z(s)$ iid normal variables and V is Cauchy.
- Importantly, we may have situation where
  - the difference $X(s) - X(s')$ has a proper density
  - even the $X(s)|X(s')$ and $X(s')|X(s)$ have proper density
  - but $(X(s), X(s'))$ does not have one.

Examples of $\gamma(h)$

- Linear:

$$\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 ||\mathbf{h}|| & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise.} \end{cases}$$

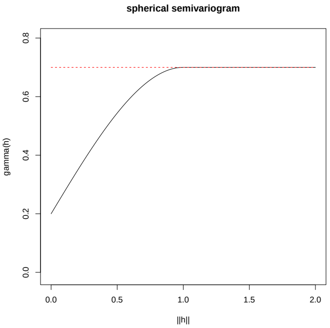


Examples of $\gamma(h)$

- Spherical: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 & \text{if } ||\mathbf{h}|| > 1/\phi \\ \tau^2 + \sigma^2 \left(\frac{3\phi||h||}{2} - \frac{1}{2}(\phi||h||)^3 \right) & \text{if } 0 < ||\mathbf{h}|| \leq 1/\phi \\ 0 & \text{otherwise} \end{cases}$

Examples of $\gamma(h)$

- Spherical: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 & \text{if } ||\mathbf{h}|| > 1/\phi \\ \tau^2 + \sigma^2 \left(\frac{3\phi||h||}{2} - \frac{1}{2}(\phi||h||)^3 \right) & \text{if } 0 < ||\mathbf{h}|| \leq 1/\phi \\ 0 & \text{otherwise} \end{cases}$

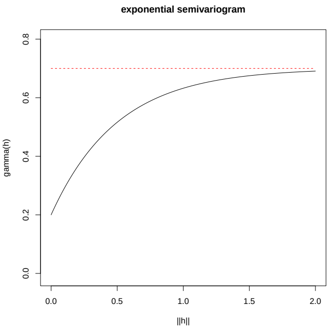


Examples of $\gamma(h)$

- Exponential: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi||h||)) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$

Examples of $\gamma(h)$

- Exponential: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi||h||)) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$

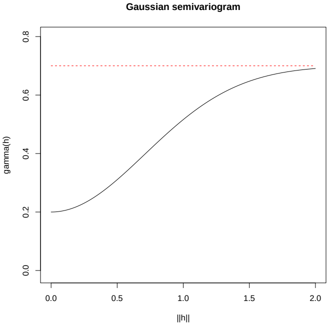


Examples of $\gamma(h)$

- Gaussian: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2||h||^2)) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$

Examples of $\gamma(h)$

- Gaussian: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 ||h||^2)) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$

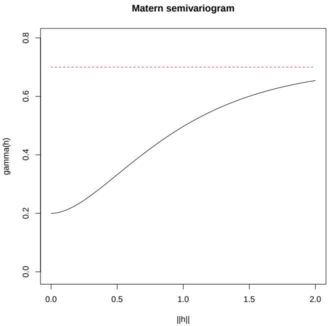


Examples of $\gamma(h)$

- Matérn: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{(2\sqrt{\nu}\phi||h||)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}\phi||\mathbf{h}||)\right) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$
- $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 \left(1 - (1 + \phi||h||) \exp(-\phi||h||)\right) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$

Examples of $\gamma(h)$

- Matérn: $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 \left(1 - \frac{(2\sqrt{\nu}\phi||h||)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}\phi||\mathbf{h}||)\right) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$
- $\gamma(||\mathbf{h}||) = \begin{cases} \tau^2 + \sigma^2 (1 - (1 + \phi||h||) \exp(-\phi||h||)) & \text{if } ||\mathbf{h}|| > 0 \\ 0 & \text{otherwise} \end{cases}$



Data again

Data again

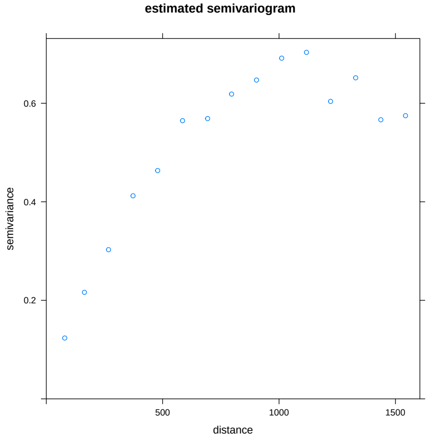
- $\widehat{\gamma}(d) = \frac{1}{2|(N(d))|} \sum_{(s_i,s_j)\in N(d)} [x(s_i) - x(s_j)]^2$
- $N(d)$: set of pairs of points such that $||s_i - s_j|| = d$
- $|N(d)|$: Cardinality of $N(d)$.

Data again

- $\widehat{\gamma}(d) = \frac{1}{2|(N(d))|} \sum_{(s_i, s_j) \in N(d)} [x(s_i) - x(s_j)]^2$
- $N(d)$: set of pairs of points such that $||s_i - s_j|| = d$
- $|N(d)|$: Cardinality of $N(d)$.
- Problem: The distances between the pairs can be all different. So it is of no use.
- Replace the $N(d)$ by $N(d_k) = \{(s_i, s_j) : d_{k-1} < ||s_i - s_j|| < d_k\}$, for $k = 1, \ldots, K$.
- $d_0 = 0$ and $d_{k-1} < d_k$

Data again

- $\widehat{\gamma}(d) = \dfrac{1}{2|(N(d))|} \sum_{(s_i,s_j)\in N(d)} [x(s_i) - x(s_j)]^2$
- $N(d)$: set of pairs of points such that $||s_i - s_j|| = d$
- $|N(d)|$: Cardinality of $N(d)$.
- Problem: The distances between the pairs can be all different. So it is of no use.
- Replace the $N(d)$ by $N(d_k) = \{(s_i, s_j) : d_{k-1} < ||s_i - s_j|| < d_k\}$, for $k = 1, \ldots, K$.
- $d_0 = 0$ and $d_{k-1} < d_k$



Semivariogram fitting

- Try with spherical semivariogram.

Semivariogram fitting

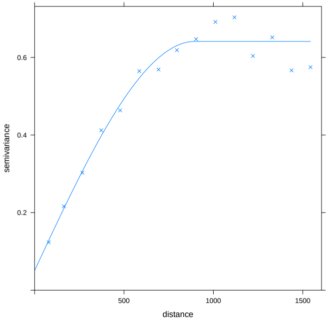
- Try with spherical semivariogram.
- Fit can be done using weighted least square technique: $\sum_{k=1}^{K} w_k \left(\hat{\gamma}(h_k) - \gamma(h_k)\right)^2$

Semivariogram fitting

- Try with spherical semivariogram.
- Fit can be done using weighted least square technique: $\sum_{k=1}^{K} w_k \left(\hat{\gamma}(h_k) - \gamma(h_k)\right)^2$
- Generally, weights $\propto$ the # of samples available in a particular distance interval, i.e., $|N(h_k)|$.

Semivariogram fitting

- Try with spherical semivariogram.
- Fit can be done using weighted least square technique: $\sum_{k=1}^{K} w_k \left(\hat{\gamma}(h_k) - \gamma(h_k)\right)^2$
- Generally, weights $\propto$ the $\#$ of samples available in a particular distance interval, i.e., $|N(h_k)|$.



Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging:

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging: ordinary kriging, simple kriging, universal kriging etc.

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging: ordinary kriging, simple kriging, universal kriging etc.
- Ordinary kriging: $X(s_0)$ is predicted using a linear combination of elements of $\mathbf{X}$:

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging: ordinary kriging, simple kriging, universal kriging etc.
- Ordinary kriging: $X(s_0)$ is predicted using a linear combination of elements of $\mathbf{X}$: $\sum_i \ell_i X(s_i)$.

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging: ordinary kriging, simple kriging, universal kriging etc.
- Ordinary kriging: $X(s_0)$ is predicted using a linear combination of elements of $\mathbf{X}$: $\sum_i \ell_i X(s_i)$.
- How to choose ℓ's?
- Two requirements:

Spatial prediction: kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- There are different kinds of kriging: ordinary kriging, simple kriging, universal kriging etc.
- Ordinary kriging: $X(s_0)$ is predicted using a linear combination of elements of $\mathbf{X}$: $\sum_i \ell_i X(s_i)$.
- How to choose ℓ's?
- Two requirements: $E(X(s_0) - \sum_i \ell_i X(s_i)) = 0$ and $E(X(s_0) - \sum_i \ell_i X(s_i))^2$ gets minimized w.r.t. ℓ's.
- Defining $a_0 = 1, a_i = -\ell_i, \forall i = 1, \ldots, n$, the above criteria become $E(\sum_{i=0}^n a_i X(s_i))^2$ with $\sum_{i=0}^n a_i = 0$.

Ordinary kriging contd.

- Expansion of $E(\sum_{i=0}^{n} a_i X(s_i))^2$ with the assumption of intrinsic stationarity, and with $\sum_i a_i = 0$, leads to $-\sum_i \sum_j a_i a_j \gamma(s_i - s_j)$.
- Now defining $\gamma_{ij} = \gamma(s_i - s_j)$, for $i, j \in \{1, \ldots, n\}$ and $\gamma_{0j} = \gamma(s_0 - s_j)$, differentiating $E(X(s_0) - \sum_i \ell_i X(s_i))^2$ w.r.t. ℓ, equating to zero will lead to the solution for ℓ's.
- However, the so obtained ℓs are dependent on the unknown γ and hence to be estimated from the data.

Universal kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.

Universal kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- Assume that $X(s)$ is related to p-dimensional covariates $z(s_1), z(s_2), \ldots, z(s_n)$ by the model: $X = Z\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(0, \Sigma)$.

Universal kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- Assume that $X(s)$ is related to p-dimensional covariates $z(s_1), z(s_2), \ldots, z(s_n)$ by the model: $X = Z\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(0, \Sigma)$.
- $\Sigma = \sigma^2 H(\phi) + \tau^2 I_n,\ H(\phi)_{i,j} = \rho(\phi; d_{i,j})$.
- $d_{i,j} = ||s_i - s_j||$, τ^2 can be 0 if we assume there is not nugget effect.

Universal kriging

- Task: Given observations of $\mathbf{X} = (X(s_1), \ldots, X(s_n))^T$, predict $X(s_0)$ at a location s_0, where it is not observed.
- Assume that $X(s)$ is related to p-dimensional covariates $z(s_1), z(s_2), \ldots, z(s_n)$ by the model: $X = Z\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(0, \Sigma)$.
- $\Sigma = \sigma^2 H(\phi) + \tau^2 I_n,\ H(\phi)_{i,j} = \rho(\phi; d_{i,j})$.
- $d_{i,j} = ||s_i - s_j||$, τ^2 can be 0 if we assume there is not nugget effect.
- Then the predicted value of $X(\cdot)$ at s_0, will be

$$E(X(s_0)|\mathbf{X} = \mathbf{x}) = \mathbf{z}_0^T \beta + \boldsymbol{\gamma}^T \Sigma^{-1}(\mathbf{x} - Z\boldsymbol{\beta}),\ \text{with}$$

- $\text{Var}(X(s_0)|\mathbf{X} = \mathbf{x}) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T \Sigma^{-1} \boldsymbol{\gamma}$.
- γ: a vector containing the covariance between $X(s_0)$ and the other X's.

Kriging contd

- $\boldsymbol{\beta}$, γ and Σ have to be estimated.

Kriging contd

- $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Σ have to be estimated.
- $\widehat{\boldsymbol{\beta}} = \left(Z^T\widehat{\Sigma}^{-1}Z\right)^{-1} Z^T\widehat{\Sigma}^{-1}\boldsymbol{x}$, $\widehat{\Sigma} = \hat{\tau}^2 I_n + \hat{\sigma}^2 H(\hat{\phi})$.

Kriging contd

- $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Σ have to be estimated.
- $\widehat{\boldsymbol{\beta}} = \left(Z^T\widehat{\Sigma}^{-1}Z\right)^{-1} Z^T\widehat{\Sigma}^{-1}\boldsymbol{x}$, $\widehat{\Sigma} = \hat{\tau^2} I_n + \hat{\sigma}^2 H(\hat{\phi})$.
- $\widehat{X(s_0)} = \boldsymbol{z}_0^T\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\gamma}}^T\widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$
- Prediction error variance $\sigma^2(s_0) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}^T(X^T\Sigma^{-1}X)^{-1}\boldsymbol{\delta}$,

Kriging contd

- $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Σ have to be estimated.
- $\widehat{\boldsymbol{\beta}} = \left(Z^T\widehat{\Sigma}^{-1}Z\right)^{-1} Z^T\widehat{\Sigma}^{-1}\boldsymbol{x}$, $\widehat{\Sigma} = \hat{\tau^2} I_n + \hat{\sigma}^2 H(\hat{\phi})$.
- $\widehat{X(s_0)} = \boldsymbol{z}_0^T\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\gamma}}^T\widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$
- Prediction error variance $\sigma^2(s_0) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}^T(X^T\Sigma^{-1}X)^{-1}\boldsymbol{\delta}$,
- $\boldsymbol{\delta} = z(s_0)^T - Z^T\Sigma^{-1}\boldsymbol{\gamma}$

Kriging contd

- $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Σ have to be estimated.
- $\widehat{\boldsymbol{\beta}} = \left(Z^T\widehat{\Sigma}^{-1}Z\right)^{-1} Z^T\widehat{\Sigma}^{-1}\boldsymbol{x}$, $\widehat{\Sigma} = \hat{\tau^2} I_n + \hat{\sigma}^2 H(\hat{\phi})$.
- $\widehat{X(s_0)} = \boldsymbol{z}_0^T\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\gamma}}^T\widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$
- Prediction error variance $\sigma^2(s_0) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}^T(X^T\Sigma^{-1}X)^{-1}\boldsymbol{\delta}$,
- $\boldsymbol{\delta} = z(s_0)^T - Z^T\Sigma^{-1}\boldsymbol{\gamma}$

Question: What if $z(\cdot)$ is not known at s_0?

Kriging contd

- $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and Σ have to be estimated.
- $\widehat{\boldsymbol{\beta}} = \left(Z^T\widehat{\Sigma}^{-1}Z\right)^{-1} Z^T\widehat{\Sigma}^{-1}\boldsymbol{x}$, $\widehat{\Sigma} = \hat{\tau}^2 I_n + \hat{\sigma}^2 H(\hat{\phi})$.
- $\widehat{X(s_0)} = \boldsymbol{z}_0^T\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\gamma}}^T\widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$
- Prediction error variance $\sigma^2(s_0) = \sigma^2 + \tau^2 - \boldsymbol{\gamma}^T\Sigma^{-1}\boldsymbol{\gamma} + \boldsymbol{\delta}^T(X^T\Sigma^{-1}X)^{-1}\boldsymbol{\delta}$,
- $\boldsymbol{\delta} = z(s_0)^T - Z^T\Sigma^{-1}\boldsymbol{\gamma}$

Question: What if $z(\cdot)$ is not known at s_0? ☹

Kriging contd.

$\widehat{X(s_0)} = \boldsymbol{z}_0^T \widehat{\boldsymbol{\beta}} + \widehat{\gamma}^T \widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$ can also be written as $\boldsymbol{\lambda}^T \boldsymbol{x}$, where

- $\boldsymbol{\lambda} = \widehat{\Sigma}^{-1}\widehat{\gamma} + \widehat{\Sigma}^{-1} Z (Z^T \widehat{\Sigma}^{-1} Z)^{-1} \left(z(s_0) - Z^T \widehat{\Sigma}^{-1} \widehat{\gamma} \right)$

Kriging contd.

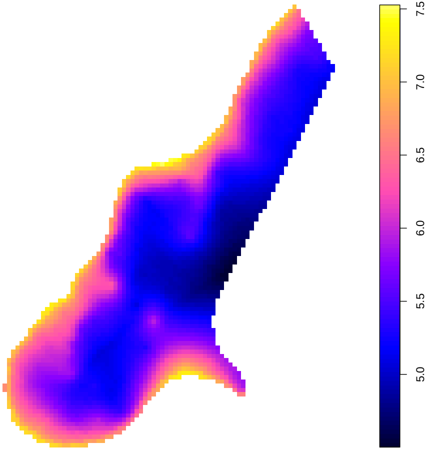
$\widehat{X(s_0)} = \boldsymbol{z}_0^T \widehat{\boldsymbol{\beta}} + \widehat{\gamma}^T \widehat{\Sigma}^{-1}(\mathbf{x} - Z\widehat{\boldsymbol{\beta}})$ can also be written as $\boldsymbol{\lambda}^T \boldsymbol{x}$, where

- $\boldsymbol{\lambda} = \widehat{\Sigma}^{-1}\widehat{\gamma} + \widehat{\Sigma}^{-1}Z(Z^T\widehat{\Sigma}^{-1}Z)^{-1}\left(z(s_0) - Z^T\widehat{\Sigma}^{-1}\widehat{\gamma}\right)$
- Multiplying λ with Z^T from the left, we get

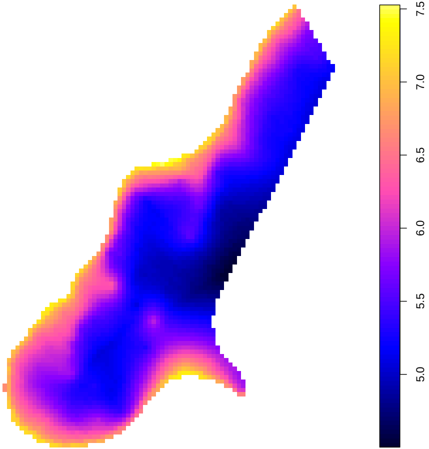
$$z(s_0) = Z^T\boldsymbol{\lambda}$$

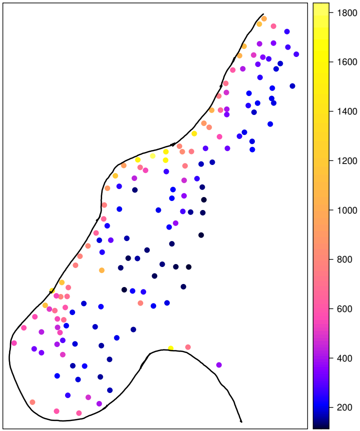
- Iterative procedure will lead us estimates of $z(s_0)$ and $X(s_0)$.

Meuse data kriging



Meuse data kriging





References

S. Banerjee, B.P. Carlin and A. E. Gelfand (2015)
Hierarchical Modeling and Analysis for Spatial Data

N. A. C. Cressie (1993)
Statistics for Spatial Data

R. S. Bivand, E. P. V. Gómez-Rubio (2013)
Applied Spatial Data Analysis with R


THANK YOU!