

Iowa Social Science Research Center
2020-21 Workshop series

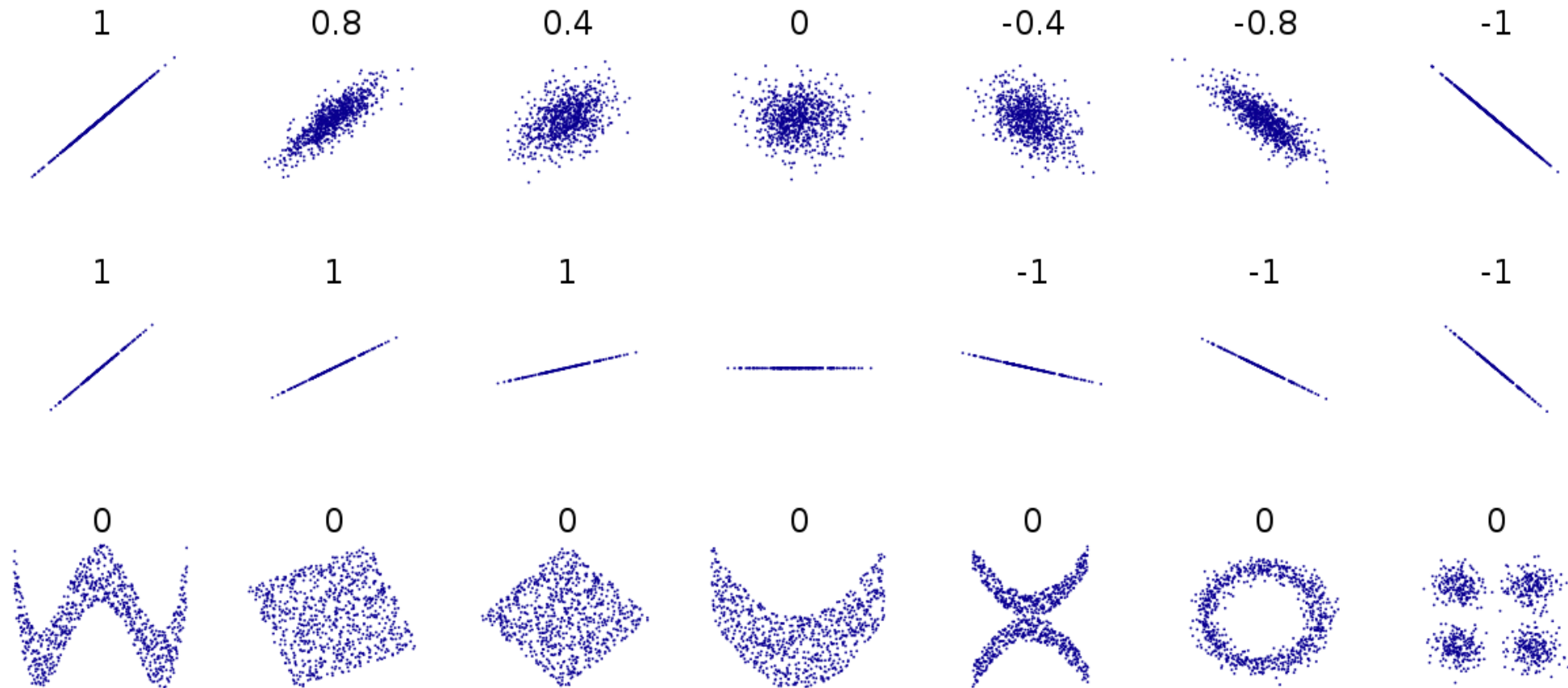


Advanced Spatial analysis with R

Seungwon Kim

Geographical and Sustainability Sciences

Correlation



Regression

- A set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables
- Goals
 - Identify the statistically significant predictors
 - Education, Epidemiology,...
 - Prediction

Ordinary Least Squares Regression (OLS)

- y_i is the **response variable**
 - a.k.a. Dependent variable
- x_i is the **explanatory variable** or **predictor variable**
 - a.k.a. Independent variable
- We are using X to explain some or most of the variability of y , in particular, the non-random part of y 's variability

Steps for linear regression

- Response variable
 - Type of variable
 - Continuous: Simple or multiple regression
 - Count data: Poisson regression or negative binomial regression
 - Binary: Logistic/Probit regression
 - Independent?
 - Study design
 - Ex) School, classroom, students
- Linear relationship between response and predictor(s)
 - Scatter plot

Steps for linear regression (cont.)

- Predictor
 - Check multicollinearity
 - Correlation matrix
 - Variance Inflation Factor (VIF)
- Error
 - Normality
 - Q-Q plot
 - Shapiro-Wilk test
 - H_0 : The population is normally distributed
 - Heteroscedasticity
 - Plot (studentized) residuals vs. predicted y
 - Breusch-Pagan test
 - H_0 : Homoskedasticity
 - Autocorrelation
 - Plot
 - Durbin-Watson test

The first law of geography

- Waldo Tobler
 - “Everything is related to everything else, but near things are more related than distant things.”
- Spatial dependence
 - The co-variation of properties within geographic space
 - Weather map (Kriging)
 - Spatial autocorrelation

Back to regression analysis

- Response variable

- Type of variable

- Continuous: Simple or multiple regression

- Count data: Poisson regression or negative binomial regression

- Binary: Logistic regression

- Independent? (Spatially independent?)

- Study design

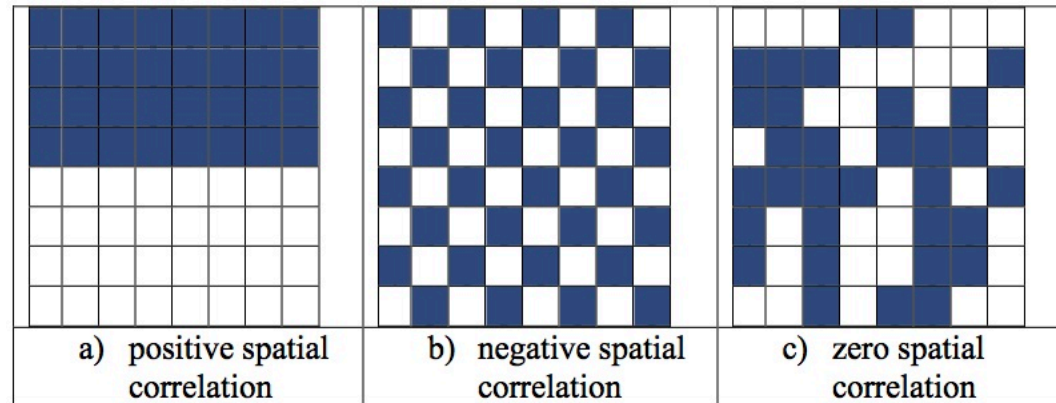
- Ex) School, classroom, students

Back to regression analysis

- Error
 - Normality
 - Q-Q plot
 - Shapiro-Wilk test
 - Heteroscedasticity
 - Plot (studentized) residuals vs. predicted y
 - Breusch-Pagan test
 - Autocorrelation (Spatial autocorrelation)
 - Plot
 - Burbin-Watson test

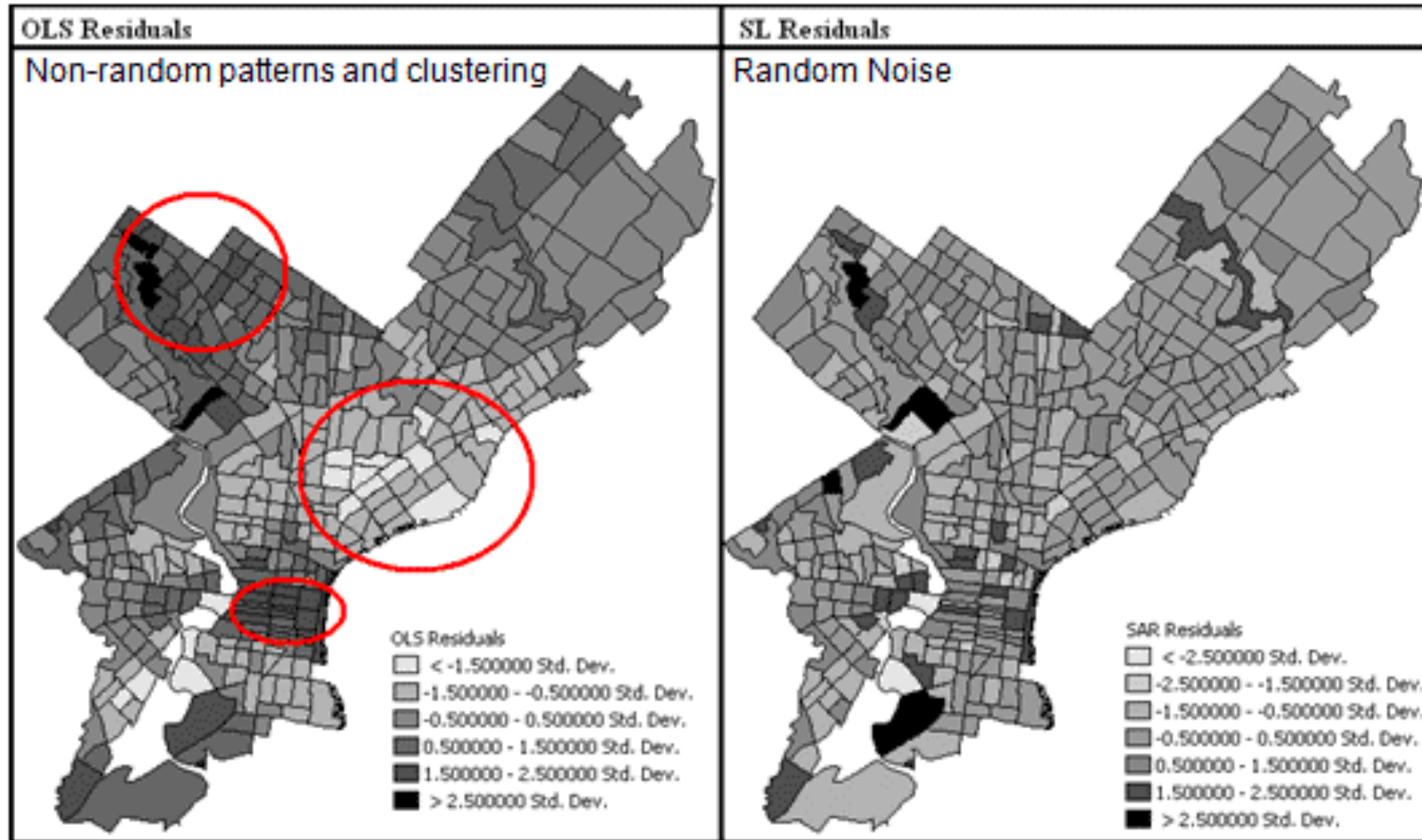
Check Spatial dependence

- Moran's I
 - A measure of spatial autocorrelation



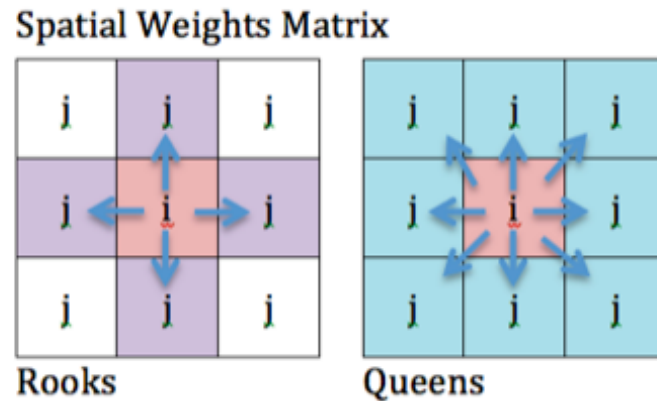
- A summary statistic with p-value from bootstrapping

Check Spatial dependence – residual map



Spatial Weight matrix

- The spatial relationships that exist among the features in your dataset
- NxN matrix
- Two ways to determine neighbors
 - Rooks case
 - Queens case



Spatial regression models

- Spatial Lag model

- Assumes that dependencies exist directly among the levels of the response variable
- Used when we know the structure of spatial dependence
- $Y = \rho WY + X\beta + \varepsilon$

- Spatial Error model

- Error term has dependence
- Used when structure of dependence is unknown
- $Y = X\beta + \lambda W\varepsilon + \xi$

Data Structures

- Vectors

- A 1-dimensional object that consists of indexed elements of the same data type
- Numeric / Character / Date / Logical / Factors
- `X <- c(1, 2, 3, 4, ...)`
- `X[i]`

- Matrix

- A 2-dimensional object that consists of indexed elements of the same data type
- `X[i, j]`

Data Structures

- Data frame
 - A 2-dimensional object that consists of indexed elements
 - Elements within a given column are of the same type, but types may differ between columns.
 - `x[i, j]`
 - `x[i,] ## Extract a row`
 - `x[, j] ## Extract a column`

R Studio

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains an R script named `spatial_analysis_with_r.R`. The script includes comments and code for installing and loading the `sf` and `spdep` packages, and for creating a spatial weights matrix using `spdep`.
- Environment:** Shows the Global Environment with a list of objects: `ia_shp` (99 obs. of 10 variables), `m1` (Large gg (9 elements, 3.7 Mb)), `m2` (Large gg (9 elements, 608 Kb)), `p1.1` (List of 9), `p2.1` (List of 9), `p3` (List of 9), `p4.1` (List of 9), `pie1` (2 obs. of 8 variables), `pie2.1` (5 obs. of 7 variables), `pie3` (2 obs. of 7 variables), `pie4.1` (2 obs. of 7 variables), and `shp` (3070 obs. of 10 variables).
- Files:** Shows a file explorer with a list of files and folders, including `.RData` (1.6 MB, Nov 9, 2019, 11:49 PM) and `.Rhistory` (19.1 KB, Nov 10, 2019, 12:06 AM).
- Console:** Displays the output of the R script, showing the installation and loading of the `sf` and `spdep` packages, and the creation of the spatial weights matrix.

```
1 #####
2
3
4 ## Packages
5 ### sf
6 ## A package to read/write shapefile in R
7 if(!require(sf)){
8   install.packages("sf")
9   library(sf)
10 }else{
11   library(sf)
12 }
13 ### spdep
14 # A collection of functions to create spatial weights matrix objects from polygon 'contiguities'
15 # and statistical tests for spatial 'autocorrelation'
16 if(!require(spdep)){
17   install.packages("spdep") ## A package to read/write shapefile in R
18   library(spdep)
19 }else{
20   library(spdep)
21 }
22 ### ggplot2
23 # Mapping
24 if(!require(ggplot2)){
25   install.packages("ggplot2") ## A package to read/write shapefile in R
26 }else{
27   library(ggplot2)
28 }
29
30 # Create spatial weights matrix
31 w = spdep::poly2nb(ia_shp)
32 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
33
34 # Create spatial weights matrix
35 w = spdep::poly2nb(ia_shp)
36 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
37
38 # Create spatial weights matrix
39 w = spdep::poly2nb(ia_shp)
40 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
41
42 # Create spatial weights matrix
43 w = spdep::poly2nb(ia_shp)
44 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
45
46 # Create spatial weights matrix
47 w = spdep::poly2nb(ia_shp)
48 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
49
50 # Create spatial weights matrix
51 w = spdep::poly2nb(ia_shp)
52 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
53
54 # Create spatial weights matrix
55 w = spdep::poly2nb(ia_shp)
56 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
57
58 # Create spatial weights matrix
59 w = spdep::poly2nb(ia_shp)
60 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
61
62 # Create spatial weights matrix
63 w = spdep::poly2nb(ia_shp)
64 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
65
66 # Create spatial weights matrix
67 w = spdep::poly2nb(ia_shp)
68 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
69
70 # Create spatial weights matrix
71 w = spdep::poly2nb(ia_shp)
72 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
73
74 # Create spatial weights matrix
75 w = spdep::poly2nb(ia_shp)
76 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
77
78 # Create spatial weights matrix
79 w = spdep::poly2nb(ia_shp)
80 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
81
82 # Create spatial weights matrix
83 w = spdep::poly2nb(ia_shp)
84 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
85
86 # Create spatial weights matrix
87 w = spdep::poly2nb(ia_shp)
88 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
89
90 # Create spatial weights matrix
91 w = spdep::poly2nb(ia_shp)
92 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
93
94 # Create spatial weights matrix
95 w = spdep::poly2nb(ia_shp)
96 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
97
98 # Create spatial weights matrix
99 w = spdep::poly2nb(ia_shp)
100 w = spdep::nb2listw(w, style="B", zero.policy=TRUE)
```