

# Modul Praktikum - Pelatihan Analisis Spasial

Anik Djuraidah, Rahma Anisa, Gerry Alfa Dito

17 dan 19 Maret 2021



# Contents



# Chapter 1

## Pengantar Analisis Spasial dengan R Software

### 1.1 Pengenalan R Software

Apabila Anda belum memiliki R dan RStudio pada perangkat PC atau laptop Anda, silahkan ikut langkah berikut:

1. Download base R sesuai dengan *operating system* Anda pada laman <https://cran.r-project.org>.
2. Install software tersebut pada system Anda.
3. Silahkan download versi desktop untuk RStudio pada system Anda pada laman berikut: <https://www.rstudio.com/products/RStudio/>.
4. Install RStudio pada system Anda.

Setelah menginstall R dan Rstudio, selanjutnya anda akan menginstall package di R yang akan dibutuhkan dalam pelatihan ini. Langkah-langkah menginstallnya dapat dilihat pada video berikut

Daftar Package yang dibutuhkan dalam pelatihan ini adalah sebagai berikut

```
install.packages("tidyverse")
install.packages("leaflet")
install.packages("sp")
install.packages("spData")
install.packages("spdep")
install.packages("rgdal")
```

## 6CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

```
install.packages("sf")
install.packages("GWmodel")
install.packages("spgwr")
install.packages("gwrr")
install.packages("plm")
install.packages("splm")
```

Kemudian, berikut adalah daftar file yang harus didownload

1. Data Atm Pontianak
2. Data SPBU Pontianak
3. Data SPBU Pontianak Excel
4. P9 Dataset 1
5. Pulau Jawa
6. Metro Manila
7. Peta Jawa
8. Peta Pontianak
9. ShowGIS SHP

Setelah mendownload file-file diatas silahkan taruh file-file tersebut dalam satu folder dan file yang masih dalam bentuk **zip** silahkan extract terlebih dahulu. Kemudian silahkan jalankan sintaks dibawah ini

```
setwd("D:/Research (eksternal dept)/pelatihan spasial (adj)/modul")
```

**Tolong sesuaikan directori-nya dengan folder yang memuat file-file yang sudah didownload sebelumnya**

### 1.1.1 Cara untuk assignment di R

```
A <- 5
A = 5
5 -> A
assign("A",5)
A
```

```
## [1] 5
```

Sintaks diatas berarti nilai 5 disimpan dalam Objek yang bernama A.

Beberapa aturan penamaan objek dalam R:

- a. Case-Sensitive (Huruf A dengan a merupakan 2 objek berbeda).
- b. Diawali huruf (A-Z atau a-z), misal A,dat1, a1.
- c. Usahakan untuk tidak menggunakan nama fungsi yang sudah ada di R, seperti plot, lm, data, boxplot.

### 1.1.2 Objek Data Standar di R

R memiliki 6 cara untuk menyimpan data dalam bentuk objek, yaitu `vector`, `factor`, `matrix`, `array`, `data.frame`, dan `list`.

#### 1. Vector

Vector dalam R berisi elemen-elemen yang memiliki tipe data yang sama.

```
A <- c(5, 3, 5, 1)
x <- 1:10
n <- c("a", "b", "c")
y <- seq(1, 10, by = 2)
```

Kita bisa mengakses elemen-elemen vektor dengan cara berikut

```
A[3]
```

```
## [1] 5
```

```
x[1:4]
```

```
## [1] 1 2 3 4
```

```
n[c(1,3)]
```

```
## [1] "a" "c"
```

```
y[-(1:2)]
```

```
## [1] 5 7 9
```

#### 2. Factor

Mirip dengan vector tapi mempunyai urutan (ordinal)

## 8CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

```
x <- c("SD", "SMP", "SMA", "S1")
sekolah <- factor(x)
sekolah

## [1] SD SMP SMA S1
## Levels: S1 SD SMA SMP

sekolah <- factor(x, levels = x)
sekolah

## [1] SD SMP SMA S1
## Levels: SD SMP SMA S1
```

Kita bisa mengakses elemen-elemen `factor` dengan cara berikut

```
sekolah[1]

## [1] SD
## Levels: SD SMP SMA S1

sekolah[1:2]

## [1] SD SMP
## Levels: SD SMP SMA S1

sekolah[c(1,4)]
```

```
## [1] SD S1
## Levels: SD SMP SMA S1
```

### 3. Matriks

Matrix dalam R memiliki dua dimensi yaitu baris dan kolom. Matrix hanya bisa digunakan untuk menyimpan 1 jenis tipe data saja, misalnya `numeric`.

```
M <- matrix(c(1, 0, 0, 0, 1, 0, 0, 0, 1),
            nrow = 3, ncol = 3, byrow = TRUE)
M

##      [,1] [,2] [,3]
## [1,]     1     0     0
## [2,]     0     1     0
## [3,]     0     0     1
```

Kita bisa mengakses elemen-elemen ‘matriks dengan cara berikut

```
M[3]
```

```
## [1] 0
```

```
M[1, 1]
```

```
## [1] 1
```

```
M[1, ]
```

```
## [1] 1 0 0
```

```
M[1, c(1,3)]
```

```
## [1] 1 0
```

```
M[, c(1,3)]
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    0
## [3,]    0    1
```

#### 4. Array

Array merupakan perluasan dari matrix dalam hal dimensi. Jika matrix hanya memiliki dua dimensi, array bisa terdiri dari banyak dimensi. Fungsi yang digunakan untuk mendefinisikan array di R adalah fungsi **array**.

```
BB <- array(1:12, dim = c(3,2,2)) #array berdimensi 3
BB
```

```
## , , 1
##
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
##
```

## 10CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

```
## , , 2
##
##      [,1] [,2]
## [1,]    7   10
## [2,]    8   11
## [3,]    9   12

CC <- array(1:24,c(4,2,2,2)) # array berdimensi 4
CC
```

```
## , , 1, 1
##
##      [,1] [,2]
## [1,]    1    5
## [2,]    2    6
## [3,]    3    7
## [4,]    4    8
##
## , , 2, 1
##
##      [,1] [,2]
## [1,]    9   13
## [2,]   10   14
## [3,]   11   15
## [4,]   12   16
##
## , , 1, 2
##
##      [,1] [,2]
## [1,]   17   21
## [2,]   18   22
## [3,]   19   23
## [4,]   20   24
##
## , , 2, 2
##
##      [,1] [,2]
## [1,]    1    5
## [2,]    2    6
## [3,]    3    7
## [4,]    4    8
```

Kita bisa mengakses elemen-elemen ‘matriks dengan cara berikut

```
BB[3]
```

```
## [1] 3
```

```
BB[1, 1,2]
```

```
## [1] 7
```

```
CC[1,1,2,2]
```

```
## [1] 1
```

## 5. Data Frame

Objek seperti tabel di Excel atau database SQL yang terdiri dari baris dan kolom. Tipe data kolom dapat berbeda-beda antar kolom.

```
x <- 10:1
y <- -4:5
q <- c("Hockey", "Football", "Baseball", "Curling", "Rugby", "Lacrosse", "Basketball", "Tennis", "Cricket", "Soccer")
theDF <- data.frame(x, y, q)
theDF

##      x   y      q
## 1 10 -4    Hockey
## 2  9 -3   Football
## 3  8 -2   Baseball
## 4  7 -1    Curling
## 5  6  0     Rugby
## 6  5  1   Lacrosse
## 7  4  2 Basketball
## 8  3  3     Tennis
## 9  2  4    Cricket
## 10 1  5    Soccer

theDF <- data.frame(first=x,second=y,sports=q)
theDF

##    first second      sports
## 1      10     -4    Hockey
## 2       9     -3   Football
## 3       8     -2   Baseball
```

## 12CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

```
## 4      7     -1    Curling
## 5      6      0    Rugby
## 6      5      1   Lacrosse
## 7      4      2 Basketball
## 8      3      3    Tennis
## 9      2      4   Cricket
## 10     1      5   Soccer
```

Kita bisa mengakses elemen-elemen `data.frame` dengan cara berikut

```
theDF[1,3] #baris 1 kolom 3
```

```
## [1] "Hockey"
```

```
theDF[c(1,5),7:10],] #baris 1,5,7,8,9,10
```

```
##   first second   sports
## 1     10     -4   Hockey
## 5      6      0    Rugby
## 7      4      2 Basketball
## 8      3      3    Tennis
## 9      2      4   Cricket
## 10     1      5   Soccer
```

```
theDF[,c(1,3)] #kolom 1 dan 3
```

```
##   first   sports
## 1     10   Hockey
## 2      9 Football
## 3      8 Baseball
## 4      7  Curling
## 5      6    Rugby
## 6      5 Lacrosse
## 7      4 Basketball
## 8      3    Tennis
## 9      2   Cricket
## 10     1   Soccer
```

```
theDF$sports #mengakses kolom sports
```

```
## [1] "Hockey"      "Football"     "Baseball"     "Curling"      "Rugby"
## [6] "Lacrosse"    "Basketball"   "Tennis"       "Cricket"      "Soccer"
```

```
theDF[,c("first","sports")] #mengakses kolom first dan sports
```

```
##   first      sports
## 1     10      Hockey
## 2      9    Football
## 3      8   Baseball
## 4      7    Curling
## 5      6     Rugby
## 6      5 Lacrosse
## 7      4 Basketball
## 8      3     Tennis
## 9      2    Cricket
## 10     1     Soccer
```

### 1.1.3 Mengakses bantuan di R

Untuk mengakses bantuan (help) dari suatu fungsi di R bisa menggunakan fungsi `help` atau operator tanda tanya `?`

```
help(lm)
?plot
```

## 1.2 Pengenalan Data Spasial

Data spasial tidak hanya berisi baris dan kolom, namun objek geometrik, seperti titik, garis, ataupun poligon. Beberapa jenis data spasial adalah data titik, data kontinu, dan data area yang akan dibawah berikut ini. Sebelum membahas jenis data tersebut satu per satu, terlebih dulu akan dibahas struktur data spasial serta cara mengimpor data spasial ke dalam R.

### 1.2.1 Struktur Data Spasial

Tipe data spasial yang paling umum digunakan adalah *shapefile*, adapun tipe lain yang juga cukup populer adalah KML (*Keyhole Markup Language*). Data *shapefile* sebenarnya terdiri dari beberapa file dengan beberapa extension, di antaranya adalah `.shp`, `.shx`, dan `.dbf`. Beberapa package yang umum digunakan untuk bekerja dengan data spasial adalah `sp` dan `rgdal`.

### 1.2.2 Ilustrasi Data Cholera

Sebagai ilustrasi, akan digunakan data yang tersedia pada laman [http://rtwilson.com/downloads/SnowGIS\\_SHP.zip](http://rtwilson.com/downloads/SnowGIS_SHP.zip). Pastikan Anda mengekstrak folder data tersebut pada direktori yang Anda inginkan. Selanjutnya, package `rgdal` akan digunakan untuk membaca data `SnowGIS` tersebut.

```
library(rgdal)
```

Untuk melihat file apa saja yang ada di dalam folder shapefile tersebut, kita dapat menggunakan fungsi `list.files()` dan tuliskan direktori Anda masing-masing, ini dikenal sebagai `dsn`.

```
dsn<-paste("SnowGIS_SHP/SnowGIS_SHP")
list.files(dsn)

##  [1] "Cholera_Deaths.dbf"          "Cholera_Deaths.prj"
##  [3] "Cholera_Deaths.sbn"          "Cholera_Deaths.sbx"
##  [5] "Cholera_Deaths.shp"          "Cholera_Deaths.shx"
##  [7] "OSMap.tfw"                  "OSMap.tif"
##  [9] "OSMap_Grayscale.tfw"         "OSMap_Grayscale.tif"
## [11] "OSMap_Grayscale.tif.aux.xml" "OSMap_Grayscale.tif.ovr"
## [13] "Pumps.dbf"                  "Pumps.prj"
## [15] "Pumps.sbx"                  "Pumps.shp"
## [17] "Pumps.shx"                  "README.txt"
## [19] "SnowMap.tfw"                "SnowMap.tif"
## [21] "SnowMap.tif.aux.xml"        "SnowMap.tif.ovr"
```

Terlihat pada output di atas bahwa folder tersebut memuat beberapa shapefile, di antaranya terdapat 6 file dengan nama `Cholera_Deaths` dan 5 file bernama `Pumps`. Kedua set data tersebut dikenal sebagai `layer`.

```
ogrListLayers(dsn)
```

```
## [1] "Cholera_Deaths" "Pumps"
## attr(,"driver")
## [1] "ESRI Shapefile"
## attr(,"nlayers")
## [1] 2
```

Kita dapat menggunakan fungsi `ogrInfo()` untuk mengetahui informasi mengenai layer tersebut.

```
ogrInfo(dsn, layer = "Cholera_Deaths")

## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Spasial\SnowCov.shp"
## Driver: ESRI Shapefile; number of rows: 250
## Feature type: wkbPoint with 2 dimensions
## Extent: (529160.3 180857.9) - (529655.9 181306.2)
## CRS: +proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000 +y_0=-100000 +ellps=airy +units=m
## LDID: 87
## Number of fields: 2
##      name type length typeName
## 1   Id    0     6  Integer
## 2 Count  0     4  Integer
```

Fungsi `readOGR()` dapat digunakan untuk membaca data shapefile.

```
CholeraDeaths <- readOGR(dsn, layer = "Cholera_Deaths")

## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Spasial\SnowCov.shp"
## with 250 features
## It has 2 fields

summary(CholeraDeaths)

## Object of class SpatialPointsDataFrame
## Coordinates:
##             min     max
## coords.x1 529160.3 529655.9
## coords.x2 180857.9 181306.2
## Is projected: TRUE
## proj4string :
## [+proj=tmerc +lat_0=49 +lon_0=-2 +k=0.9996012717 +x_0=400000
## +y_0=-100000 +ellps=airy +units=m +no_defs]
## Number of points: 250
## Data attributes:
##      Id     Count
## Min. :0 Min. : 1.000
## 1st Qu.:0 1st Qu.: 1.000
## Median :0 Median : 1.000
## Mean   :0 Mean  : 1.956
## 3rd Qu.:0 3rd Qu.: 2.000
## Max.  :0 Max. :15.000
```

## 16 CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

Selanjutnya kita dapat memeriksa `class` dari data `CholeraDeaths` tersebut.

```
class(CholeraDeaths)

## [1] "SpatialPointsDataFrame"
## attr(,"package")
## [1] "sp"
```

Data tersebut merupakan `SpatialPointsDataFrame` yang termasuk S4 class, maka untuk mengakses data slot perlu digunakan notasi `@`.

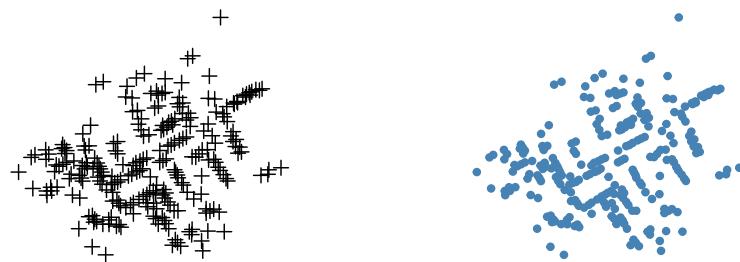
```
str(CholeraDeaths@data)

## 'data.frame':    250 obs. of  2 variables:
## $ Id   : int  0 0 0 0 0 0 0 0 0 ...
## $ Count: int  3 2 1 1 4 2 2 2 3 2 ...
```

### 1.2.3 Visualisasi Data Cholera

Fungsi `plot()` dapat digunakan untuk membuat grafik paling sederhana dari data `CholeraDeaths`.

```
par(mfrow=c(1,2))
plot(CholeraDeaths)
plot(CholeraDeaths, pch=20, col="steelblue")
```



Perhatikan bahwa plot di atas hanya menunjukkan sebaran titik spasial, tanpa memberikan informasi yang jelas tentang lokasi data tersebut. Jika kita memiliki peta dalam bentuk data polygon, kita dapat mengimpor data tersebut dengan cara yang sama (seandainya datanya berupa shapefile), kemudian kita plot peta baru kemudian plot titik seperti di atas.

Alternatif lainnya jika kita tidak ingin menggunakan peta polygon dari shapefile, kita dapat menggunakan beberapa package yang tersedia di R software, seperti `ggmap`, `OpenStreetMap`, `leaflet`, atau yang lain. Namun perhatikan bahwa untuk bisa menggunakan package `OpenStreetMap`, Anda harus memastikan bahwa jika Anda menggunakan R 64-bit maka Java yang terinstall di PC Anda juga harus sesuai, yaitu 64-bit.

Berikut ini akan ditunjukkan salah satu cara menampilkan peta dengan memanfaatkan package `leaflet`.

```
library(leaflet)
map <- leaflet() %>% setView(lng = -0.13659, lat = 51.51328, zoom = 12)
map %>% addTiles()
```

Sebelum kedua peta dan data titik digabungkan. Pastikan terlebih dahulu apakah koordinat yang digunakan menggunakan skala yang sama.

```
head(coordinates(CholeraDeaths))

##      coords.x1 coords.x2
## [1,] 529308.7 181031.4
## [2,] 529312.2 181025.2
## [3,] 529314.4 181020.3
## [4,] 529317.4 181014.3
## [5,] 529320.7 181007.9
## [6,] 529336.7 181006.0
```

Seperti terlihat di atas, koordinat pada data `CholeraDeaths` diukur pada skala yang berbeda dengan peta yang diambil dari package `leaflet`. Terdapat beberapa macam *coordinate reference system (CRS)*, beberapa di antaranya yang cukup populer adalah suatu set EPSG (*European Petroleum Survey Group*) berikut:

- **EPSG:4326** juga dikenal sebagai WGS84, ukuran standard yang digunakan pada sistem GPS dan *Google Earth*.
- **EPSG:3857** digunakan pada Google Maps, Open Street Maps, dsb.
- **EPSG:27700** juga dikenal sebagai OSGB 1936, atau *British National Grid: United Kingdom Ordnance Survey*.

```

cholera_latlong <- CholeraDeaths %>%
  spTransform(CRS("+init=epsg:4326"))
leaflet(data = CholeraDeaths) %>%
  addTiles() %>%
  addMarkers(cholera_latlong@coords[,1], cholera_latlong@coords[,2])

```

Dapat dilihat di atas, bahwa setelah koordinatnya disamakan, kita dapat menampilkan data CholeraDeaths pada peta yang diperoleh dari Open Street Map melalui package leaflet.

## 1.3 Tipe Data Spasial

### 1.3.1 Tipe Data titik

Data spasial dapat berupa titik pengamatan pada lokasi tertentu, yang umumnya menyimpan koordinat lokasi *longitude* dan *latitude*. Data jenis ini hanya memiliki nilai pada titik tertentu saja, misalnya data kejadian kecelakaan, data rumah sakit, data kejadian kriminal, dan lain-lain. Sebagai ilustrasi, akan diperlihatkan data atm dan SPBU berikut ini.

```

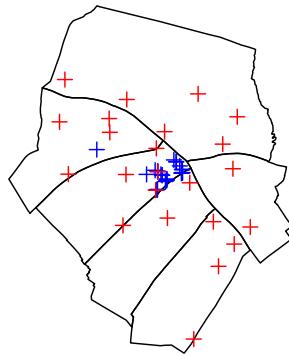
library(rgdal)
petapontianak=readOGR(dsn="Peta Pontianak", layer="Pontianak_kec")

## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Spa
## with 6 features
## It has 1 fields

dataATM=read.csv("DATA ATM PONTIANAK.csv" , header=T)
dataSPBU=read.csv("Data SPBU Pontianak.csv", sep=";",header=T)
plot(petapontianak)
points(dataATM$lon, dataATM$lat, col="blue", pch=3)
points(dataSPBU$lon, dataSPBU$lat, col="red", pch=3)
title("Lokasi ATM dan SPBU di Pontianak")

```

### Lokasi ATM dan SPBU di Pontianak



Data yang digunakan adalah berupa titik yang menunjukkan lokasi ATM dan SPBU. Lokasi ATM ditunjukkan dengan warna biru, sedangkan lokasi SPBU ditunjukkan dengan warna merah

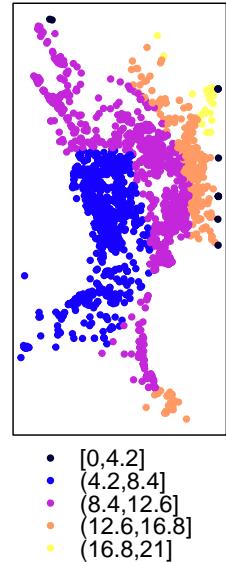
#### 1.3.2 Tipe Data Kontinu

Tipe data ini merupakan pengamatan yang memiliki nilai tidak hanya pada titik yang tersampel saja, namun nilai pengamatan sebenarnya kontinu untuk semua area. Artinya, di luar dari titik yang tersampel pun memiliki nilai untuk peubah yang diamati tersebut. Misalnya polusi udara, temperatur, kelembapan udara, presipitasi, dan sebagainya. Sebagai ilustrasi, berikut ini adalah contoh data presipitasi di daerah Metro Manila.

```
library(sp)
library(gstat)

## Warning: package 'gstat' was built under R version 4.0.3

metromanila=read.csv("metromanila.csv")
coordinates(metromanila)<-c("lon","lat")
spplot(metromanila,"precipitation", asp = 1,
       cex=0.5, pch = 19, main="Angka Presipitasi di Metro Manila")
```

**Angka Presipitasi di Metro Manila**

### 1.3.3 Tipe Data Area

Pada tipe data ini, pengamatan dilakukan pada level area. Area dapat mengacu pada sistem administrasi misalnya desa, kelurahan, kecamatan, kota, bahkan negara. Berikut ini adalah contoh data area yang diukur pada level kabupaten/kota. Untuk selanjutnya, pengamatan pada berbagai peubah dapat ditambahkan ke dalam data agar dapat diolah pada tahap analisis berikutnya.

```
library(spdep)
library(rgdal)
library(raster)
petajawa<- readOGR(dsn = "Jawamap", layer="jawa")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Sp
## with 119 features
## It has 5 fields

plot(petajawa)
```



## 1.4 Matriks Bobot dan Autokorelasi Spasial

Pada analisis data spasial, informasi ketergantungan antar lokasi dapat diukur dengan autokorelasi spasial. Untuk dapat menghitung nilai autokorelasi tersebut, terdapat beberapa tahap yang perlu dilakukan, yaitu:

- (1) menentukan kriteria kebertetanggaan antar lokasi pengamatan
- (2) menyusun matriks bobot spasial
- (3) matriks bobot spasial selanjutnya dapat dimanfaatkan baik untuk mengukur autokorelasi spasial maupun untuk menyusun pemodelan spasial.

### 1.4.1 Kriteria Kebertetanggaan

Ilustrasi yang akan digunakan pada bagian ini adalah data yang tersedia di dalam R. Data tersebut dapat dipanggil dengan fungsi berikut.

```
library(raster)
p <- shapefile(system.file("external/lux.shp", package="raster"))
p <- p[p$NAME_1=="Diekirch", ]
```

## 22CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

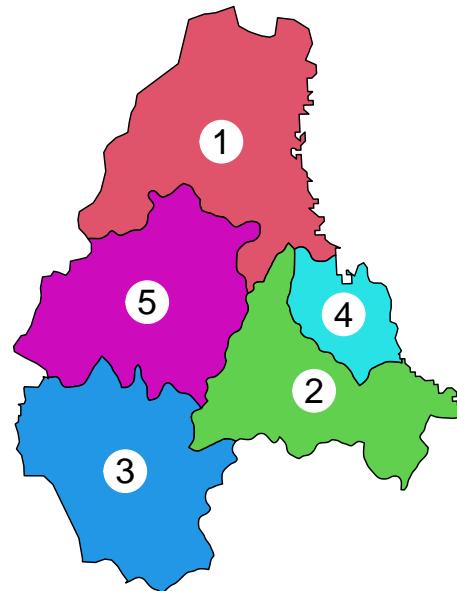
Selanjutnya kita akan tentukan sembarang nilai yang akan disimpan pada setiap lokasi untuk mengilustrasikan nilai peubah yang diamati.

```
p$value <- c(10, 6, 4, 11, 6)
data.frame(p)
```

```
##   ID_1  NAME_1 ID_2  NAME_2 AREA value
## 0    1 Diekirch    1 Clervaux  312    10
## 1    1 Diekirch    2 Diekirch  218     6
## 2    1 Diekirch    3 Redange  259     4
## 3    1 Diekirch    4 Vianden   76    11
## 4    1 Diekirch    5 Wiltz    263     6
```

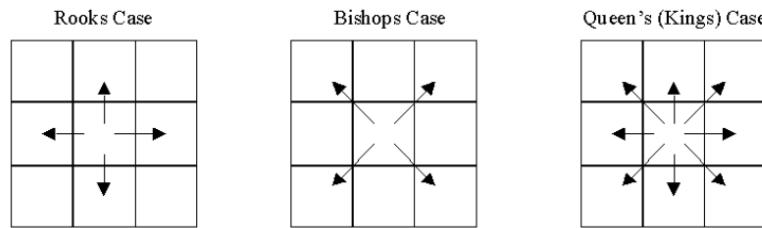
Berikut adalah visualisasi dari data yang telah kita persiapkan.

```
par(mai=c(0,0,0,0))
plot(p, col=2:7)
coords <- coordinates(p)
points(coords, cex=6, pch=20, col='white')
text(p, 'ID_2', cex=1.5)
```



### 1.4.1.1 Contiguity Based

Kriteria yang umum digunakan pada ketetanggaan berbasis *contiguity* adalah *queen contiguity*, *rook contiguity*, dan *bishop contiguity*, seperti yang terlihat pada ilustrasi berikut ini.

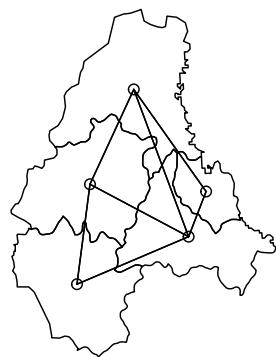


Program berikut ini dapat digunakan untuk dapat memperoleh matriks bobot berdasarkan kriteria *queen contiguity*.

```
library(spdep)
w <- poly2nb(p)

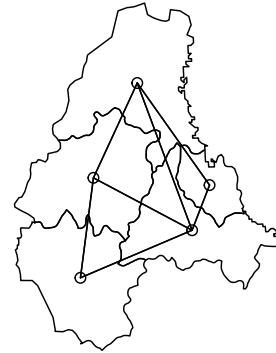
#lebih lengkap dapat dituliskan seperti berikut ini:
w <- poly2nb(p, queen=TRUE)

plot(p)
plot(w, coords, add=T)
```



Jika yang ingin digunakan adalah kriteria *rook contiguity*, maka kita dapat mengganti argumen pada program sebelumnya menjadi `queen=FALSE`.

```
w.rook <- poly2nb(p, queen=FALSE)
coords<-coordinates(p)
plot(p)
plot(w, coords, add=T)
```



Perhatikan bahwa pada kasus ini, kedua kriteria memperlihatkan hasil yang sama. Hal ini terjadi karena semua area pada ilustrasi ini bersinggungan sudut.

#### 1.4.1.2 Distance Based

Kriteria kebertetanggaan dapat pula ditentukan berdasarkan jarak antar lokasi, beberapa pendekatan jarak yang dapat digunakan adalah *k* tetangga terdekat (*k nearest neighbours (KNN)*), *radial distance*, *power distance*, dan *exponential distance*. Ilustrasi KNN dan *radial distance* dapat dilihat pada subbab berikutnya dalam modul ini.

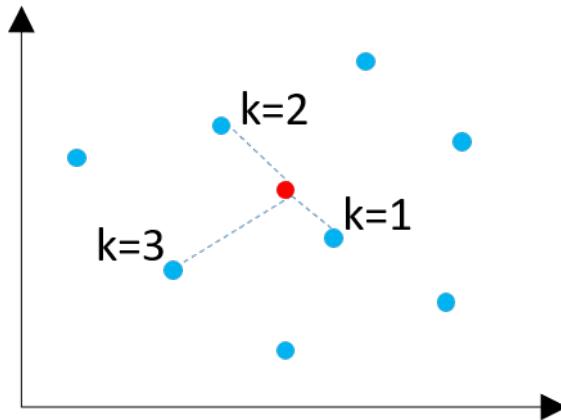
Pendekatan *power distance* dan *exponential distance* tidak diberikan ilustrasi pada modul ini, namun berikut adalah penjelasan singkat mengenai keduanya. Apabila bobot antara lokasi ke-*i* dan lokasi ke-*j* dinotasikan dengan  $w_{ij}$ , dan jarak antara kedua lokasi tersebut dinotasikan dengan  $d_{ij}$ , formula untuk memperoleh bobot jarak dengan pendekatan *power distance* adalah:

$$w_{ij} = d_{ij}^{-\alpha}$$

Sedangkan bobot jarak berdasarkan *exponential distance* dapat diperoleh dengan formula:

$$w_{ij} = e^{-\alpha d_{ij}}$$

**1.4.1.2.1 K-Nearest Neighbours** Pada pendekatan ini, kita mendefinisikan lokasi yang merupakan tetangga dari lokasi ke- $i$  adalah sejumlah  $k$  lokasi yang memiliki jarak terdekat dengan lokasi  $i$ .



Berikut ini adalah ilustrasi untuk memperoleh ketetanggaan dengan pendekatan KNN menggunakan program R.

```

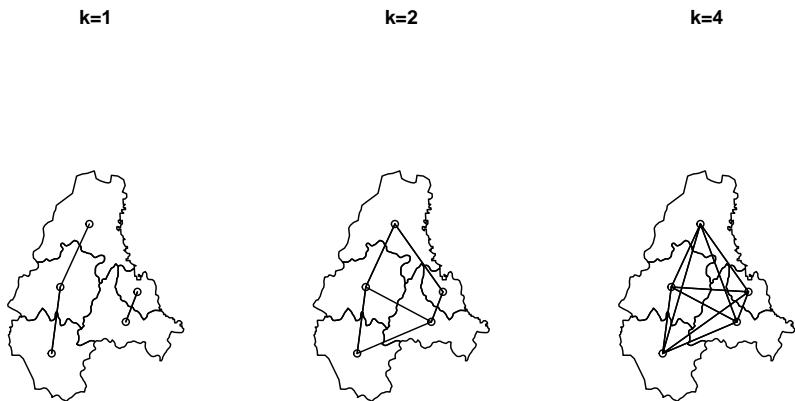
coords<-coordinates(p)
IDs<-row.names(as(p, "data.frame"))
p_kn1<-knn2nb(knearneigh(coords, k=1), row.names=IDs)
p_kn2<-knn2nb(knearneigh(coords, k=2), row.names=IDs)
p_kn4<-knn2nb(knearneigh(coords, k=4), row.names=IDs)

par(mfrow=c(1,3))
plot(p, main = "k=1")
plot(p_kn1, coords, add=T)

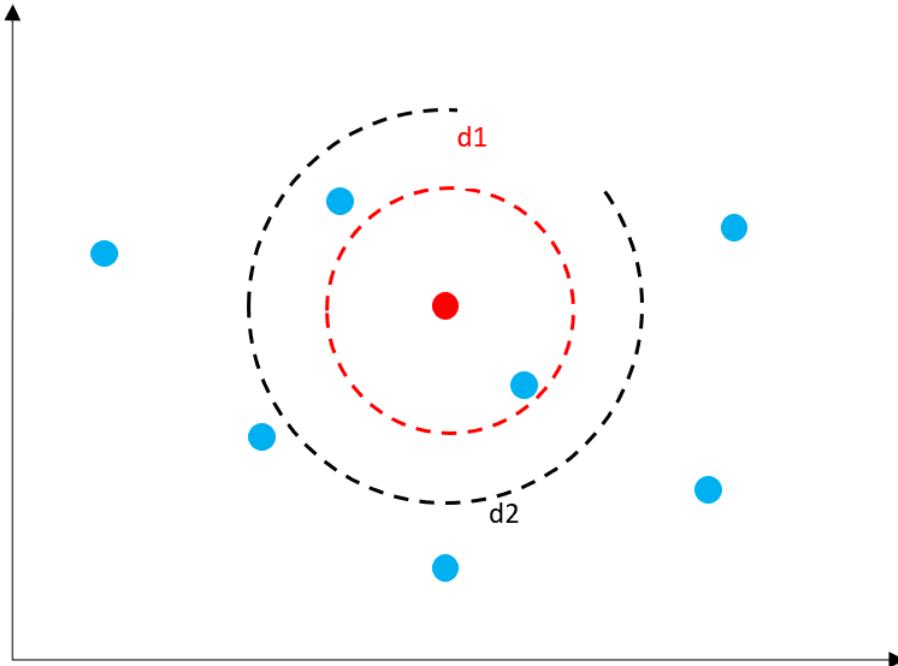
plot(p, main = "k=2")
plot(p_kn2, coords, add=T)

plot(p, main = "k=4")
plot(p_kn4, coords, add=T)

```



**1.4.1.2.2 *Radial Distance*** Pada pendekatan ini,kita mendefinisikan lokasi yang merupakan tetangga dari lokasi ke- $i$  adalah sejumlah  $k$  lokasi yang berada batas batas jarak (radius) antara  $d_1$  dan  $d_2$ , diukur dari lokasi  $i$ .



Berikut ini adalah ilustrasi untuk memperoleh ketetanggaan dengan pendekatan KNN menggunakan program R.

```

dist<-unlist(nbdists(p_kn1, coords))
summary(dist)

##      Min. 1st Qu. Median      Mean 3rd Qu.      Max.
## 0.07316 0.07316 0.14159 0.11832 0.14159 0.16213

sort(dist)

## [1] 0.07315711 0.07315711 0.14158521 0.14158521 0.16213194

max_k1<-max(dist)
p_kd1<-dnearneigh(coords, d1=0, d2=0.75*max_k1, row.names=IDs)
p_kd2<-dnearneigh(coords, d1=0, d2=1*max_k1, row.names=IDs)
p_kd3<-dnearneigh(coords, d1=0, d2=1.5*max_k1, row.names=IDs)

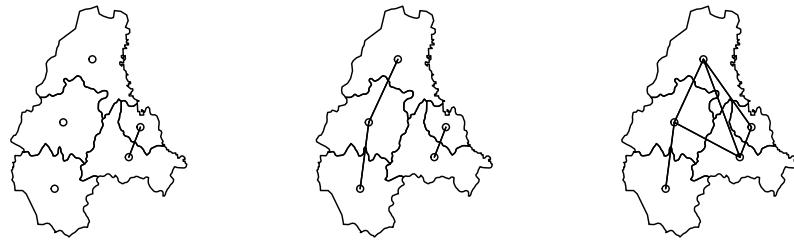
par(mfrow=c(1,3))
plot(p, main = "Distance=0.75*max_k1")
plot(p_kd1,coords, add=T)

```

```
plot(p, main = "Distance=0.75*max_k1")
plot(p_kd2,coords, add=T)

plot(p, main = "Distance=1*max_k1")
plot(p_kd3,coords, add=T)
```

Distance=0.75\*max\_k1                  Distance=1\*max\_k1                  Distance=1.5\*max\_k1



#### 1.4.2 Matriks Pembobot Spasial

Matriks pembobot spasial dapat berisi elemen biner (1 atau 0) untuk menunjukkan ketetanggaan setiap lokasi, atau berupa matriks yang sudah distandardisasi. Umumnya, matriks pembobot spasial merupakan matriks yang terstandarisasi baris (*row standardized*).

```
nb2mat(w, style="B") #matriks pembobot biner

##   [,1] [,2] [,3] [,4] [,5]
## 0    0    1    0    1    1
## 1    1    0    1    1    1
## 2    0    1    0    0    1
## 3    1    1    0    0    0
## 4    1    1    1    0    0
## attr(,"call")
## nb2mat(neighbours = w, style = "B")
```

```
nb2mat(w) # matriks pembobot row standardized
```

```
##      [,1]      [,2]      [,3]      [,4]      [,5]
## 0 0.0000000 0.3333333 0.0000000 0.3333333 0.3333333
## 1 0.2500000 0.0000000 0.2500000 0.2500000 0.2500000
## 2 0.0000000 0.5000000 0.0000000 0.0000000 0.5000000
## 3 0.5000000 0.5000000 0.0000000 0.0000000 0.0000000
## 4 0.3333333 0.3333333 0.3333333 0.0000000 0.0000000
## attr(,"call")
## nb2mat(neighbours = w)
```

### 1.4.3 Autokorelasi Spasial

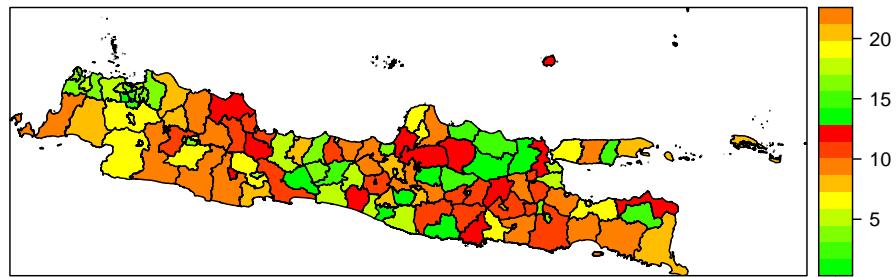
#### 1.4.3.1 Indeks Moran Global

Sebagai ilustrasi, akan digunakan data persentase kemiskinan kabupaten/kota di Pulau Jawa. Berikut adalah syntax untuk membaca peta Pulau Jawa dan data persentase kemiskinan Kabupaten/Kota di Pulau Jawa

```
library(rgdal)
library(spdep)
library(sp)
petajawa=readOGR(dsn="Jawamap", layer="jawa")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Spasial\Jawamap.shp"
## with 119 features
## It has 5 fields
```

```
datajawa=read.csv("Pulau Jawa.csv", header=T, sep=";")
petajawa$Kemiskinan<- datajawa$Kemiskinan
k=9
colfunc <- colorRampPalette(c("green", "yellow", "red"))
color <- colfunc(k)
spplot(petajawa, "Kemiskinan", col.regions=color)
```



Seandainya kita akan menggunakan kriteria queen contiguity, maka dapat dilakukan dengan syntax berikut untuk mendefinisikan matriks pembobotnya

```
w.queen <- poly2nb(petajawa)
w.queen

## Neighbour list object:
## Number of regions: 119
## Number of nonzero links: 522
## Percentage nonzero weights: 3.68618
## Average number of links: 4.386555
## 1 region with no links:
## 0
```

Terlihat pada output tersebut, bahwa terdapat 1 wilayah yang tidak memiliki tetangga, sehingga untuk syntax selanjutnya perlu ditambahkan `zero.policy=T`.

```
wqueen <- nb2listw(w.queen, zero.policy=T)
```

Seandainya kita ingin menguji autokorelasi menggunakan pendekatan indeks moran, maka kita dapat menggunakan fungsi `moran.test()`.

```
I1 <- moran.test(petajawa$Kemiskinan, wqueen, zero.policy=T, alternative="greater")
#alternative hypothesis could be either of "two.sided", "greater", or "less"
I1
##
## Moran I test under randomisation
##
## data: petajawa$Kemiskinan
## weights: wqueen n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = 7.7638, p-value = 4.12e-15
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.517060772     -0.008547009     0.004583226
```

Berdasarkan output di atas, diperoleh nilai p-value yang sangat kecil, artinya kita dapat menolak hipotesis nol yang menyatakan bahwa tidak terdapat autokorelasi. Artinya kita dapat menyimpulkan bahwa terdapat cukup bukti untuk menyatakan bahwa terdapat autokorelasi pada taraf nyata 5%.

Uji moran dapat pula dilakukan dengan melibatkan simulasi monte carlo.

```
set.seed(123)
MC<- moran.mc(petajawa$Kemiskinan, wqueen, nsim=99, zero.policy=T, alternative="greater")
# View results (including p-value)
MC
##
## Monte-Carlo simulation of Moran I
##
## data: petajawa$Kemiskinan
## weights: wqueen
## number of simulations + 1: 100
##
## statistic = 0.51706, observed rank = 100, p-value = 0.01
## alternative hypothesis: greater
```

#### 1.4.3.2 Indeks Moran Lokal

Pendekatan ini termasuk ke dalam *Local Indicators for Spatial Association (LISA)*, yang mengidentifikasi autokorelasi pada tingkat lokal.

32 CHAPTER 1. PENGANTAR ANALISIS SPASIAL DENGAN R SOFTWARE

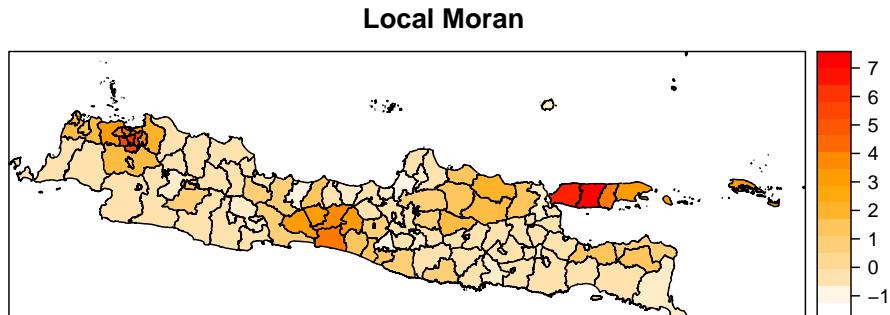
```
oid <- order(petajawa$Kemiskinan)
resI <- localmoran(petajawa$Kemiskinan, wqueen)

## Warning in lag.listw(listw, z, zero.policy = zero.policy, NAOK = NAOK): NAs in
## lagged values

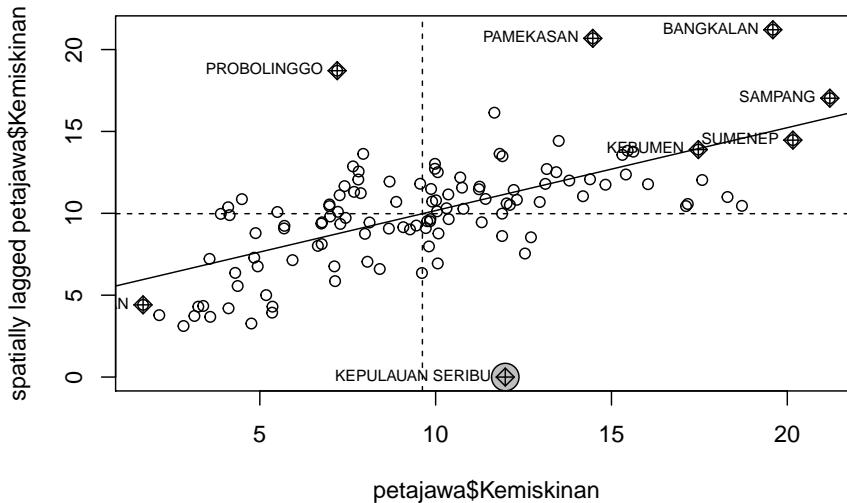
head(resI)

##          Ii      E.Ii   Var.Ii     Z.Ii    Pr(z > 0)
## 0        NA 0.000000000 0.0000000      NA       NA
## 1 2.514160 -0.008474576 0.1571389 6.363741 9.844878e-11
## 2 2.172630 -0.008474576 0.1571389 5.502177 1.875645e-08
## 3 2.040414 -0.008474576 0.2398473 4.183607 1.434598e-05
## 4 1.872452 -0.008474576 0.1902223 4.312619 8.066594e-06
## 5 1.382831 -0.008474576 0.1902223 3.190008 7.113445e-04

petajawa$z.li <- resI[,4]
petajawa$pvalue <- resI[,5]
lm.palette <- colorRampPalette(c("white","orange", "red"), space = "rgb")
spplot(petajawa, zcol="z.li", col.regions=lm.palette(20), main="Local Moran")
```



```
moran.plot(petajawa$Kemiskinan, wqueen, zero.policy=T, labels=petajawa$KABKOT)
```



## 1.5 Sumber Pustaka

Agafonkin, V. (n.d.). Leaflet for R - Markers. [rstudio.github.io](https://rstudio.github.io/leaflet/markers.html). Retrieved from <https://rstudio.github.io/leaflet/markers.html>

Baumer, B.S., Kaplan, D.T., Horton, N.J. 2017. Modern Data Science with R. CRC Press.

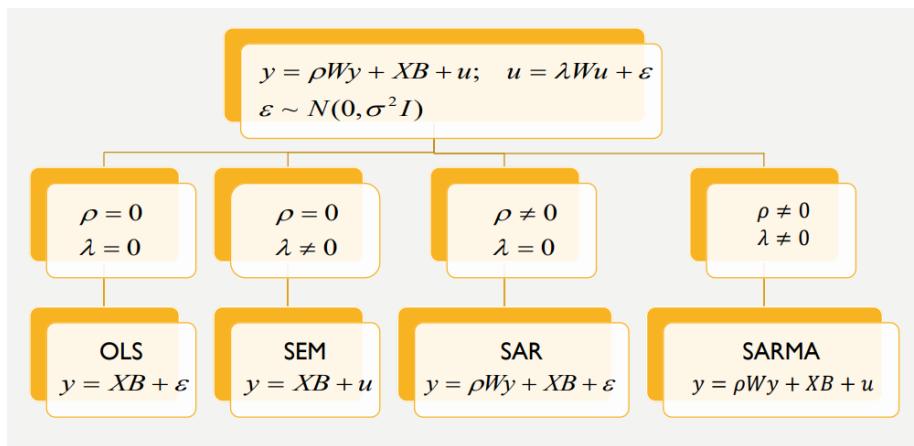
UQ SLC Digital Team. (2020, April 16). Creating maps using R. Language Technology and Data Analysis Laboratory (LADAL). Retrieved from <https://slcladal.github.io/maps.html>



# Chapter 2

## Pemodelan Dependensi Spasial

### 2.1 Model Spasial Global



Secara umum, tahapan pemodelan regresi spasial adalah sebagai berikut:

- (1) Eksplorasi Data
- (2) Regresi Klasik & Uji Asumsi
- (3) Matriks Pembobot Spasial
- (4) Uji Lagrange Multiplier
- (5) Regresi Spasial & Uji Asumsi

## (6) Kebaikan Model

Sebagai ilustrasi untuk menjelaskan tahapan tersebut, kita akan menggunakan data kemiskinan dan kependudukan di Pulau Jawa berikut ini:

1. Data polygon (peta Pulau Jawa, dengan extension .shp)
2. Data frame (data persentase kemiskinan, PDRB, pendidikan Angka Melek Huruf, Pengeluaran perkapita, Ruta Penerima Raskin, Penduduk berusia 15-64, Harapan Lama Sekolah, dan Rata-Rata Lama Sekolah, diperoleh dari BPS)

Seperti yang telah dijelaskan pada modul sebelumnya, impor data dapat dilakukan dengan program berikut ini.

```
datajawa = datajawa=read.csv("Pulau Jawa.csv", header=T, sep=";")  
head(datajawa)  
  
##      Provinsi      Kabupaten.kota Longitude Latitude Kemiskinan     PDRB  
## 1 DKI Jakarta    Kepulauan Seribu 106.5072 -5.7985    11.98 338.80  
## 2                      Kota Jakarta Selatan 106.8079 -6.2627    2.83 261.58  
## 3                      Kota Jakarta Timur 106.8995 -6.2249    3.14 155.53  
## 4                      Kota Jakarta Pusat 106.8279 -6.1801    3.59 692.24  
## 5                      Kota Jakarta Barat 106.7633 -6.1675    3.39 168.68  
## 6                      Kota Jakarta Utara 106.8926 -6.1555    5.35 271.85  
##      Pendidikan      AMH Pengeluaran.Perkapita Ruta.Penerima.Raskin Penduduk.15.64  
## 1      5.29 100.00          68.80      55.34    86.2882  
## 2      7.96 100.00          46.63       0.00    86.8031  
## 3      2.91 100.00          50.39       0.00    86.9343  
## 4      3.62 100.00          50.87       0.00    86.2032  
## 5      25.02 100.00         53.29       0.00    87.4070  
## 6      18.59 98.98          51.12       0.00    87.3943  
##      HLS Rata.Rata.Lama.Sekolah  
## 1 12.48                  8.46  
## 2 13.31                  11.57  
## 3 13.43                  11.64  
## 4 13.23                  11.24  
## 5 12.78                  10.38  
## 6 12.61                  10.69
```

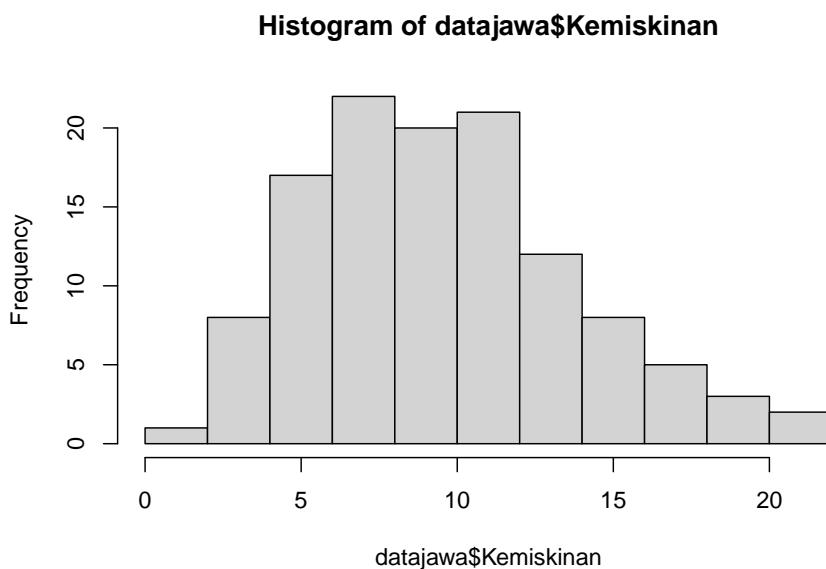
Pada ilustrasi ini, pemodelan dilakukan untuk mengkaji peubah respon **persentase penduduk miskin tahun 2018 di pulau Jawa (Y)** dengan menggunakan peubah **persentase Pendidikan yang ditamatkan di Bawah SD Tahun 2018(X)**.

```
library(spdep)
library(rgdal)
library(raster)

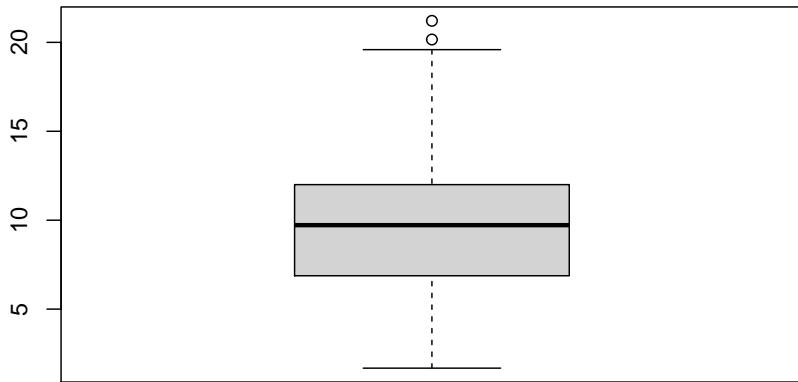
petajawa<-readOGR(dsn="directory tempat folder utk file .shp", layer="nama file shp")
```

## 2.2 Eksplorasi Data

```
hist(datajawa$Kemiskinan)
```



```
bxplt<-boxplot(datajawa$Kemiskinan, outlier.tagging=T)
```



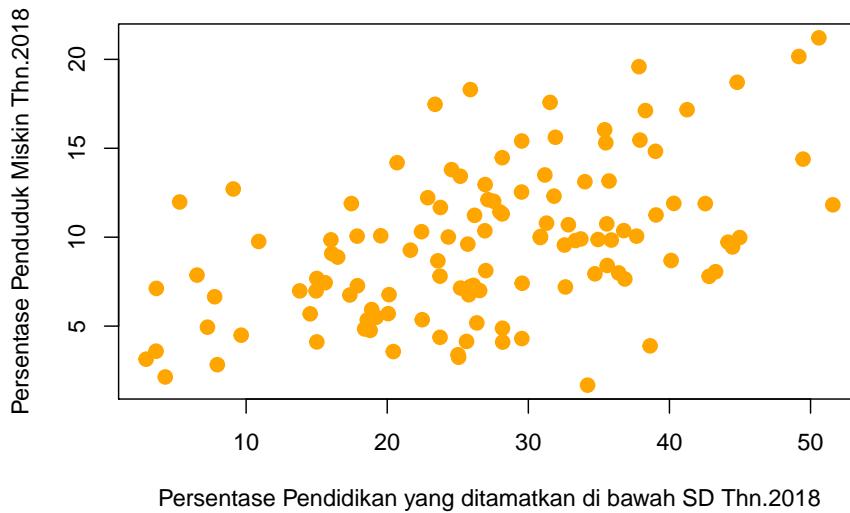
```
bxplt$out
```

```
## [1] 21.21 20.16
```

```
subset(datajawa, datajawa$Kemiskinan>bxplt$stats[5])
```

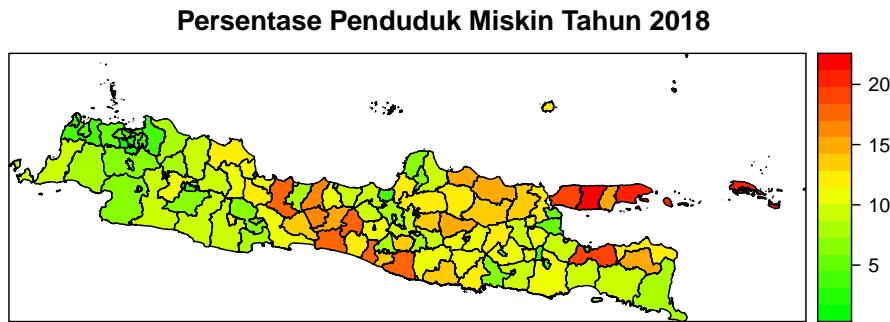
```
##      Provinsi Kabupaten.kota Longitude Latitude Kemiskinan PDRB Pendidikan
## 100          Sampang    113.2394   -7.0402     21.21 19.85    50.59
## 102          Sumenep    113.9061   -6.9254     20.16 30.17    49.17
##          AMH Pengeluaran.Perkapita Ruta.Penerima.Raskin Penduduk.15.64 HLS
## 100 88.94                  69.82           67.89    66.5778 11.76
## 102 85.90                  64.05           78.71    71.6376 13.07
##          Rata.Rata.Lama.Sekolah
## 100          4.36
## 102          5.23
```

```
plot(datajawa$Pendidikan, datajawa$Kemiskinan,
      xlab="Percentase Pendidikan yang ditamatkan di bawah SD Thn.2018",
      ylab="Percentase Penduduk Miskin Thn.2018",
      pch=20, col="orange", cex=2)
```



Plot tersebut memperlihatkan adanya pola hubungan linear positif antara persentase pendidikan yang ditamatkan di bawah SD terhadap persentase penduduk miskin di Pulau Jawa pada tahun 2018.

```
petajawa$Kemiskinan<- datajawa$Kemiskinan
k=16
colfunc <- colorRampPalette(c("green", "yellow","red"))
color <- colfunc(k)
spplot(petajawa, "Kemiskinan", col.regions=color, main="Persentase Penduduk Miskin Tahun 2018")
```



Berdasarkan plot di atas, dapat dilihat adanya kecenderungan pola bergerombol pada data persentase kemiskinan di kabupaten/kota di Pulau Jawa. Hal ini tampak dari gradasi warna yang cenderung mengumpul, seperti pada warna hijau, merah dan oranye.

### 2.2.1 Identifikasi Autokorelasi pada Data

```
w<-poly2nb(petajawa)
ww<-nb2listw(w, zero.policy=T)
moran(datajawa$Kemiskinan, ww, n=length(ww$neighbours),
      S0=Szero(ww), zero.policy=T)
```

```
## $I
## [1] 0.5214426
##
## $K
## [1] 2.853193
```

```
moran.test(datajawa$Kemiskinan, ww,randomisation=T,
            alternative="greater", zero.policy=T)
```

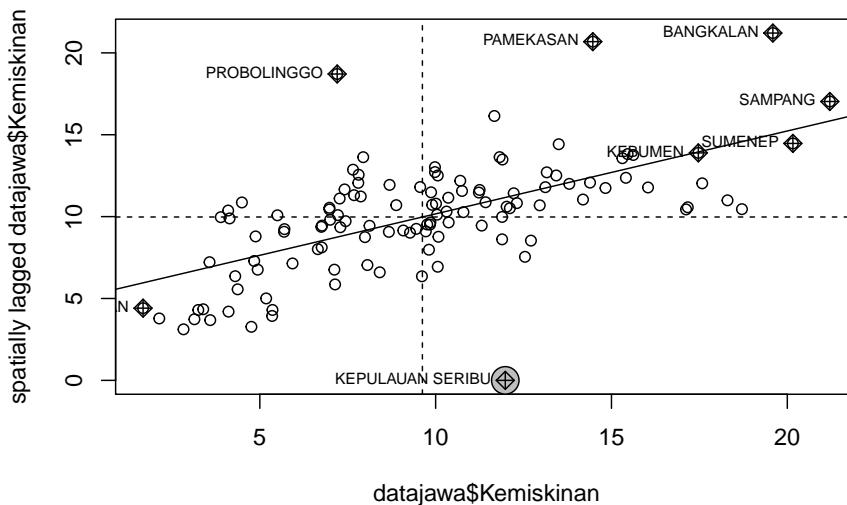
```
##
```

```

## Moran I test under randomisation
##
## data: datajawa$Kemiskinan
## weights: ww n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = 7.7638, p-value = 4.12e-15
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.517060772     -0.008547009     0.004583226

moran.plot(datajawa$Kemiskinan, ww, labels=petajawa$KABKOT, zero.policy=T)

```



## 2.3 Pemodelan Regresi Klasik

Pemodelan regresi dapat dilakukan menggunakan fungsi `lm` berikut.

```

reg.klasik = lm(Kemiskinan~Pendidikan, data = datajawa)
err.regklasik<-residuals(reg.klasik)
summary(reg.klasik)

```

```

## 
## Call:
## lm(formula = Kemiskinan ~ Pendidikan, data = datajawa)
##
## Residuals:
##       Min     1Q   Median     3Q    Max
## -9.3117 -2.6659 -0.3763  2.2998  8.8844
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.50549   0.89314   5.045 1.68e-06 ***
## Pendidikan  0.18965   0.03066   6.186 9.37e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.671 on 117 degrees of freedom
## Multiple R-squared:  0.2464, Adjusted R-squared:  0.24
## F-statistic: 38.26 on 1 and 117 DF,  p-value: 9.373e-09

cor(datajawa$Kemiskinan, fitted(reg.klasik))^-2

## [1] 0.2464408

```

## 2.4 Diagnostik Model

### 2.4.1 Kenormalan Sisaan

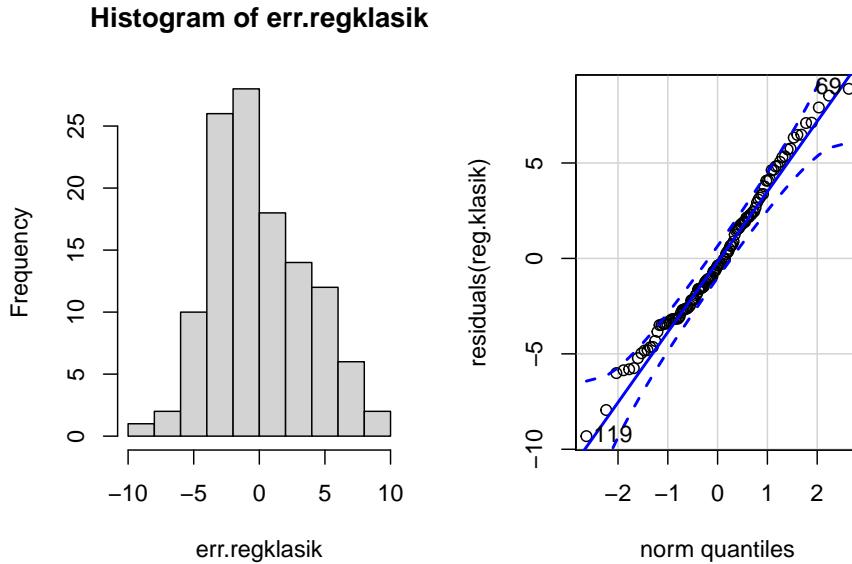
```

library(nortest)
library(car)
library(DescTools)
library(lmtest)
shapiro.test(err.regklasik)

##
## Shapiro-Wilk normality test
##
## data: err.regklasik
## W = 0.98607, p-value = 0.2608

par(mfrow=c(1,2))
hist(err.regklasik)
car::qqPlot(residuals(reg.klasik))

```



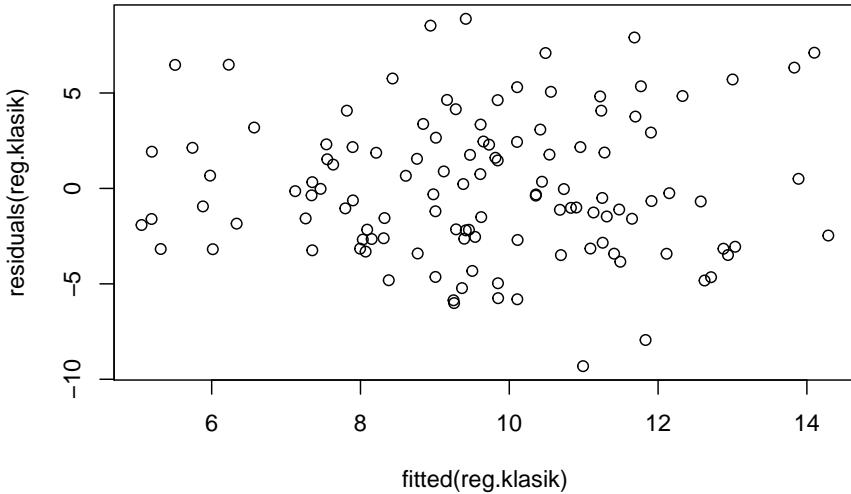
```
## [1] 119 69
```

H0: galat model menyebar normal

H1: galat model tidak menyebar normal

#### 2.4.2 Kehomogenan Ragam Sisaan

```
plot(fitted(reg.klasik), residuals(reg.klasik))
```



```
bptest(reg.klasik)

##
## studentized Breusch-Pagan test
##
## data: reg.klasik
## BP = 3.979, df = 1, p-value = 0.04607
```

H0: ragam galat homogen

H1: ragam galat tidak homogen

### 2.4.3 Kebebasan Sisaan

Uji kebebasan sisaan pada data spasial dapat dilakukan dengan uji moran menggunakan fungsi berikut.

```
w<-poly2nb(petajawa)
ww<-nb2listw(w, zero.policy =T)
lm.morantest(reg.klasik, ww, alternative="two.sided", zero.policy = T)

##
```

```

## Global Moran I for regression residuals
##
## data:
## model: lm(formula = Kemiskinan ~ Pendidikan, data = datajawa)
## weights: ww
##
## Moran I statistic standard deviate = 5.3932, p-value = 6.922e-08
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.351747300     -0.011502703     0.004536512

```

Selain menggunakan fungsi `lm.morantest`, uji moran dapat dilakukan menggunakan fungsi `moran.test` seperti yang dibahas pada modul pertemuan sebelumnya. Perbedaannya adalah pada fungsi pertama, input yang digunakan adalah objek `lm`, sedangkan pada fungsi kedua, yang digunakan sebagai input adalah data sisaan model.

```

moran.test(err.regklasik, ww,randomisation=F, alternative="two.sided", zero.policy=T)

##
## Moran I test under normality
##
## data: err.regklasik
## weights: ww n reduced by no-neighbour observations
##
##
## Moran I statistic standard deviate = 5.3242, p-value = 1.014e-07
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic      Expectation      Variance
##      0.351747300     -0.008547009     0.004579345

```

Terlihat pada output bahwa hasil kedua tes menunjukkan kesimpulan yang sama, yaitu tolak  $H_0$  yang menyatakan bahwa tidak terdapat autokorelasi pada sisaan model regresi klasik pada taraf nyata 5%. Oleh karenanya, untuk mencari model yang lebih baik, kita dapat melakukan uji LM (lagrange multiplier) untuk mengidentifikasi model dependensi spasial yang dapat digunakan pada kasus ini.

## 2.5 Uji Lagrange Multiplier

```

LM<-lm.LMtests(reg.klasik, nb2listw(w, style="W", zero.policy=T),
                 test=c("LMerr", "LMLag", "RLMerr", "RLMLag", "SARMA"), zero.policy=T)
summary(LM)

## Lagrange multiplier diagnostics for spatial dependence
## data:
## model: lm(formula = Kemiskinan ~ Pendidikan, data = datajawa)
## weights: nb2listw(w, style = "W", zero.policy = T)
##
##          statistic parameter   p.value
## LMerr     26.4886      1 2.651e-07 ***
## LMLag     27.2100      1 1.825e-07 ***
## RLMerr     1.4217      1    0.2331
## RLMLag     2.1431      1    0.1432
## SARMA     28.6317      2 6.063e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Output memperlihatkan bahwa hasil uji model SEM dan SAR sama-sama signifikan pada taraf 5%. Selanjutnya, hasil uji robust keduanya ternyata sama-sama tidak signifikan. Berdasarkan skema tersebut, kita dapat mencoba kandidat model SARMA atau GSM. Namun demikian, ada pula pendapat yang menyarankan agar kita mengambil kandidat model dengan p-value terkecil, pada kasus ini p-value terkecil juga terdapat pada model SARMA atau GSM.

Mohon diingat bahwa pada ilustrasi yang kita lakukan saat ini, kita hanya menggunakan satu peubah bebas sehingga kita tidak perlu mengkhawatirkan masalah multikolinieritas. Pada saat Anda memiliki lebih dari satu peubah bebas, pastikan Anda juga memperhatikan multikolinieritas pada model. Periksaan dapat dilakukan dengan fungsi `vif()` pada package `car`.

## 2.6 Pemodelan Regresi Spasial

Pada modul ini, untuk kepentingan pembelajaran, kita akan mencoba ketiga model, SEM, SAR, dan SARMA, meskipun pada prakteknya, Anda hanya perlu memodelkan yang menurut Anda terbaik saja.

```
w<-poly2nb(petajawa)
ww<-nb2listw(w, zero.policy=T)
```

### 2.6.1 Model SEM

```

library(spatialreg)

sem<-errorsarlm(Kemiskinan~Pendidikan,data=datajawa,ww, zero.policy=T)
summary(sem)

##
## Call:errorsarlm(formula = Kemiskinan ~ Pendidikan, data = datajawa,
##     listw = ww, zero.policy = T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.04950 -1.86484 -0.43797  1.87675  8.14992
##
## Type: error
## Regions with no neighbours included:
## 0
## Coefficients: (asymptotic standard errors)
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 6.359121  0.958495  6.6345 3.256e-11
## Pendidikan  0.114932  0.029332  3.9183 8.918e-05
##
## Lambda: 0.57487, LR test value: 29.606, p-value: 5.2937e-08
## Asymptotic standard error: 0.083992
## z-value: 6.8443, p-value: 7.6856e-12
## Wald statistic: 46.844, p-value: 7.6857e-12
##
## Log likelihood: -307.7876 for error model
## ML residual variance (sigma squared): 9.4425, (sigma: 3.0729)
## Number of observations: 119
## Number of parameters estimated: 4
## AIC: 623.58, (AIC for lm: 651.18)

pseudoR2.sem<-cor(datajawa$Kemiskinan, fitted(sem))^2
pseudoR2.sem

## [1] 0.4786585

```

Output di atas menunjukkan bahwa koefisien Lambda signifikan pada taraf nyata 5% ( p-value = 5.2937e-08 ). AIC model SEM adalah sebesar 623.58, dengan pseudo- $R^2 = 0.4787$  Selanjutnya kita akan coba memeriksa sisaan model SEM ini.

```

library(nortest)
err.sem<-residuals(sem)
shapiro.test(err.sem)

## 
##  Shapiro-Wilk normality test
##
## data: err.sem
## W = 0.98234, p-value = 0.1208

bptest.sarlm(sem)

## 
## studentized Breusch-Pagan test
##
## data:
## BP = 4.1883, df = 1, p-value = 0.0407

moran.test(err.sem, ww, alternative="two.sided", zero.policy=T)

## 
## Moran I test under randomisation
##
## data: err.sem
## weights: ww n reduced by no-neighbour observations
##
## 
## Moran I statistic standard deviate = -0.53535, p-value = 0.5924
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic      Expectation      Variance
##          -0.044773446     -0.008547009     0.004579087

```

Terlihat pada output di atas bahwa sisaan memenuhi asumsi kenormalan dan asumsi kebebasan.

### 2.6.2 Model SAR

```

sar<-lagsarlm(Kemiskinan~Pendidikan,data=datajawa,ww, zero.policy=T)
summary(sar)

```

```

## 
## Call:lagsarlm(formula = Kemiskinan ~ Pendidikan, data = datajawa,
##     listw = ww, zero.policy = T)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.17361 -1.97250 -0.56646  1.82449  9.57815
##
## Type: lag
## Regions with no neighbours included:
## 0
## Coefficients: (asymptotic standard errors)
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.779095  0.992235  1.7930  0.07297
## Pendidikan  0.117724  0.027717  4.2473 2.164e-05
##
## Rho: 0.46783, LR test value: 25.624, p-value: 4.1491e-07
## Asymptotic standard error: 0.087013
##      z-value: 5.3766, p-value: 7.5901e-08
## Wald statistic: 28.908, p-value: 7.5901e-08
##
## Log likelihood: -309.7788 for lag model
## ML residual variance (sigma squared): 10.098, (sigma: 3.1777)
## Number of observations: 119
## Number of parameters estimated: 4
## AIC: 627.56, (AIC for lm: 651.18)
## LM test for residual autocorrelation
## test value: 0.014268, p-value: 0.90492

pseudoR2.sar<-cor(datajawa$Kemiskinan, fitted(sar))^2
pseudoR2.sar

## [1] 0.4292466

```

Output di atas memperlihatkan bahwa koefisien Rho pada model SAR signifikan, dengan nilai AIC sebesar 627.56 dengan pseudo- $R^2 = 0.4292$ . Selain itu, terlihat pula hasil uji autokorelasi pada sisaan model memperlihatkan nilai p-value sebesar 0.90492, artinya tidak terdapat autokorelasi pada sisaan.

```

err.sar<-residuals(sar)
shapiro.test(err.sar)

##
## Shapiro-Wilk normality test

```

```

##  

## data: err.sar  

## W = 0.98005, p-value = 0.07421

bptest.sarlm(sar)

##  

## studentized Breusch-Pagan test
##  

## data:  

## BP = 1.6725, df = 1, p-value = 0.1959

moran.test(err.sar, ww, alternative="two.sided", zero.policy=T)

##  

## Moran I test under randomisation
##  

## data: err.sar
## weights: ww n reduced by no-neighbour observations
##  

##  

## Moran I statistic standard deviate = 0.18303, p-value = 0.8548
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic      Expectation      Variance
##          0.003823646     -0.008547009     0.004567919

```

Berdasarkan output di atas, pada taraf 5% dapat disimpulkan bahwa sisakan model memenuhi asumsi kenormalan, kehomogenan ragam, dan kebebasan.

### 2.6.3 Model GSM/SARMA

```

gsm<-sacsarlm(Kemiskinan~Pendidikan,data=datajawa,ww, zero.policy=T)
summary(gsm)

##  

## Call:sacsarlm(formula = Kemiskinan ~ Pendidikan, data = datajawa,
##     listw = ww, zero.policy = T)
##  

## Residuals:
##      Min       1Q   Median       3Q      Max

```

```

## -6.20932 -1.69855 -0.38001  1.70634  7.26252
##
## Type: sac
## Coefficients: (asymptotic standard errors)
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept) 10.338070   1.750226  5.9067 3.49e-09
## Pendidikan   0.099813   0.027137  3.6781 0.000235
##
## Rho: -0.43058
## Asymptotic standard error: 0.14723
##      z-value: -2.9246, p-value: 0.0034494
## Lambda: 0.79882
## Asymptotic standard error: 0.069513
##      z-value: 11.492, p-value: < 2.22e-16
##
## LR test value: 34.317, p-value: 3.5325e-08
##
## Log likelihood: -305.432 for sac model
## ML residual variance (sigma squared): 7.7401, (sigma: 2.7821)
## Number of observations: 119
## Number of parameters estimated: 5
## AIC: 620.86, (AIC for lm: 651.18)

pseudoR2.gsm<-cor(datajawa$Kemiskinan, fitted(gsm))^2
pseudoR2.gsm

## [1] 0.5933811

```

Output di atas memperlihatkan bahwa kedua koefisien dependensi spasial, Rho dan Lambda signifikan. AIC model SARMA adalah sebesar 620.86 dengan pseudo- $R^2 = 0.5934$

```

err.gsm<-residuals(gsm)
shapiro.test(err.gsm)

##
## Shapiro-Wilk normality test
##
## data: err.gsm
## W = 0.98527, p-value = 0.2219

bptest.sarlm(gsm)

```

```

## 
## studentized Breusch-Pagan test
## 
## data:
## BP = 2.3236, df = 1, p-value = 0.1274

moran.test(err.gsm, ww, alternative="two.sided", zero.policy=T)

## 
## Moran I test under randomisation
## 
## data: err.gsm
## weights: ww n reduced by no-neighbour observations
## 
## 
## Moran I statistic standard deviate = -0.17988, p-value = 0.8572
## alternative hypothesis: two.sided
## sample estimates:
## Moran I statistic      Expectation      Variance
##       -0.020720135     -0.008547009     0.004579838

```

Berdasarkan output di atas, terlihat bahwa sisaan model SARMA telah memenuhi asumsi kenormalan, kehomogenan ragam, dan kebebasan.

#### 2.6.4 Penentuan Model Terbaik

Akhirnya, kita akan coba merangkum hasil pemodelan yang telah dilakukan sepanjang ilustrasi pada modul ini.

Rincian	OLS	SEM	SAR	SARMA
AIC	651.180	623.580	627.560	620.860
pseudo-R2	0.246	0.479	0.429	0.593
p-value of Rho	NA	NA	0.000	0.000
p-value of Lambda	NA	0.000	NA	0.003
Kenormalan(p-value)	0.261	0.121	0.074	0.222
Homoskedastisitas (p-value)	0.046	0.041	0.196	0.127
Kebebasan Sisaan(p-value)	0.000	0.592	0.855	0.857

Ilustrasi pada kasus ini memperlihatkan bahwa ternyata GSM merupakan model terbaik berdasarkan nilai AIC dan pseudo- $R^2$ . Hal ini ternyata konsisten dengan p-value nya yang juga terkecil pada uji-LM.

## 2.7 Efek Marginal

Efek marginal atau limpahan (*spill-over*) adalah besarnya dampak perubahan pada peubah dependen pada wilayah-*i*, akibat perubahan prediktor di wilayah-*j*.

Efek marginal terdapat pada model dependensi spasial SAR, GSM, SDM, SDEM, dan SLX. Efek ini dapat dibedakan menjadi tiga, yaitu efek langsung (*direct effect*), efek tidak langsung (*indirect effect*), dan efek total (*total effect*).

```
impacts(gsm, listw=ww)

## Impact measures (sac, exact):
##             Direct    Indirect      Total
## Pendidikan 0.1038916 -0.03386839 0.07002321
```

Output di atas menunjukkan bahwa pertambahan satu satuan pada pendidikan di suatu wilayah akan diikuti oleh peningkatan kemiskinan di wilayah tersebut sebesar rata-rata 0.1938916 (pengaruh langsung), sedangkan pada wilayah tetangganya akan mengalami penurunan kemiskinan rata-rata sebesar 0.03386839 (pengaruh tak langsung).

## 2.8 Latihan

Sebagai latihan, silahkan lakukan pemodelan menggunakan data Kemiskinan di Pulau Jawa dengan peubah bebas ruta penerima raskin dan persentase pendidikan yang ditamatkan di bawah SD.

- periksa multikolineritas antar peubah bebas yang digunakan berdasarkan VIF
- eksplorasi autokorelasi spasial pada model menggunakan jarak `W_dist`
- lakukan pemodelan yang menurut Anda paling tepat, interpretasikan.

## 2.9 Sumber Pustaka

Guliyev, H. (2020). Determining the spatial effects of COVID-19 using the spatial panel data model. *Spatial Statistics*, 100443. doi: 10.1016/j.spasta.2020.100443. Retrieved from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7139267/>

Sarmiento-Barbieri, I. (2016, April 24). An introduction to spatial econometrics in R. Spatial Econometric Workshop, University of Illinois. Retrieved from: [https://www.econ.uiuc.edu/~lab/workshop/Spatial\\_in\\_R.html#modeling-spatial-dependence](https://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html#modeling-spatial-dependence)

Zhukov, Y. M. (2010, January 19). Applied Spatial Statistics in R, Section 6, Spatial Regression [PDF slides.]. IQSS, Harvard University. Retrieved from: <http://www.people.fas.harvard.edu/~zhukov/Spatial6.pdf>

## Chapter 3

# Geographically Weighted Regression (GWR)

Suatu pemodelan dapat bersifat global maupun lokal. Regresi linier klasik merupakan salah satu model global. Dikatakan global karena terdapat satu model yang berlaku umum untuk semua pengamatan.

Suatu model lokal bersifat lebih fleksibel, yang dalam konteks spasial, artinya setiap daerah/lokasi dapat memiliki model masing-masing.

Geographically Weighted Regression (GWR) merupakan salah satu model yang bersifat lokal. Beberapa keuntungan dengan menggunakan model ini, diantaranya adalah kita dapat:

- menduga galat baku lokal
- menghitung ukuran leverage lokal
- melakukan pengujian terhadap signifikansi keragaman spasial pada penduga parameter lokal
- menguji apakah model lokal lebih baik daripada model global

Terdapat salah satu *stand-alone software* untuk melakukan GWR, yaitu software GWR yang dapat diakses melalui <http://ncg.nuim.ie/ncg/GWR/>. Selain itu, pada R software, terdapat beberapa package yang dapat digunakan untuk membangun model GWR, yaitu:

- GWmodel
- spgwr

- gwrr

Pada modul ini akan dibahas pemodelan GWR menggunakan package `spgwr`.

### 3.1 Eksplorasi Data

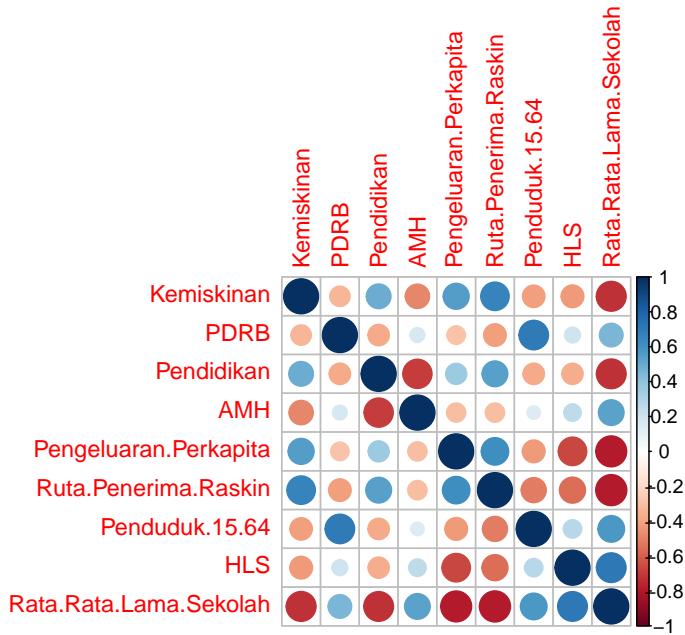
```
library(rgdal)
petajawa=readOGR(dsn="Jawamap", layer="jawa")

## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Sp
## with 119 features
## It has 5 fields

datajawa=read.csv("Pulau Jawa.csv", header=T, sep=";")
petajawa$Kemiskinan<- datajawa$Kemiskinan
```

Plot berikut ini dapat dimanfaatkan untuk mengeksplorasi hubungan antar peubah.

```
corr<-cor(datajawa[,-(1:4)])
corrplot::corrplot(corr, is.corr=T)
```



Sebagai ilustrasi, peubah rata-rata lama sekolah akan digunakan untuk memodelkan kemiskinan di pulau Jawa. Sebagai langkah awal, kita akan terlebih dulu memodelkannya menggunakan regresi linear.

```
fit.lm<-lm(Kemiskinan~Rata.Rata.Lama.Sekolah, data=datajawa)
summary(fit.lm)

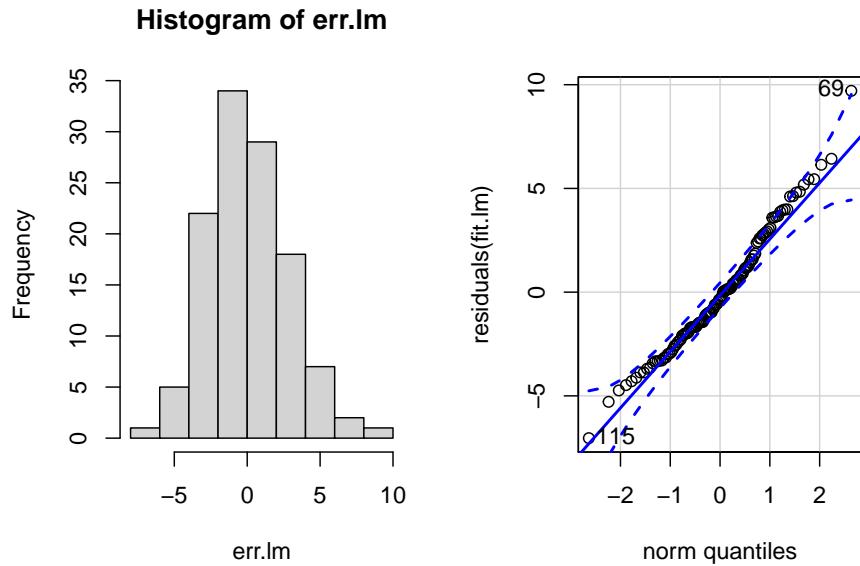
##
## Call:
## lm(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data = datajawa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -7.0384 -1.9865 -0.3389  1.6722  9.7042 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             24.7340    1.3417  18.44 <2e-16 ***
## Rata.Rata.Lama.Sekolah -1.8657    0.1624 -11.49 <2e-16 ***
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.898 on 117 degrees of freedom
## Multiple R-squared:  0.5302, Adjusted R-squared:  0.5262 
## F-statistic: 132 on 1 and 117 DF,  p-value: < 2.2e-16
```

Selanjutnya, diagnostik model dilakukan untuk memeriksa pemenuhan asumsi pada model regresi linear.

```
err.lm<-residuals(fit.lm)
shapiro.test(err.lm)

##
## Shapiro-Wilk normality test
##
## data: err.lm
## W = 0.98029, p-value = 0.07818

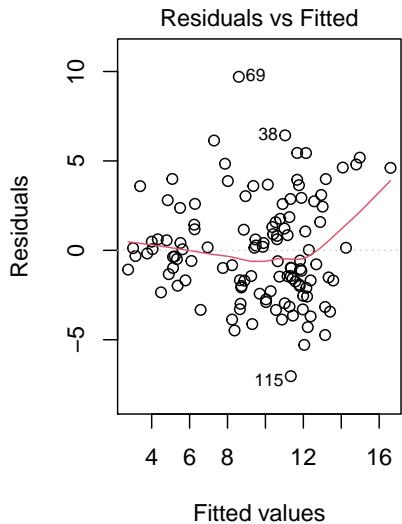
par(mfrow=c(1,2))
hist(err.lm)
car::qqPlot(residuals(fit.lm))
```



```
## [1] 69 115
```

```
plot(fit.lm,which=1)
lmtest::bttest(fit.lm)
```

```
##
## studentized Breusch-Pagan test
##
## data: fit.lm
## BP = 4.7982, df = 1, p-value = 0.02849
```



Terlihat pada output di atas bahwa sisaan model cenderung memiliki ragam yang tidak konstan. Selanjutnya juga akan diperiksa kebebasan sisaan menggunakan uji moran.

```
library(spdep)
w<-poly2nb(petajawa)
ww<-nb2listw(w, zero.policy=T)

lm.morantest(fit.lm, ww, alternative="two.sided", zero.policy = T)

##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data =
## datajawa)
## weights: ww
##
## Moran I statistic standard deviate = 5.9602, p-value = 2.519e-09
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##          0.38842759     -0.01284003     0.00453256
```

Terlihat pada output bahwa terdapat autokorelasi spasial pada sisaan model

## 60 CHAPTER 3. GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

regresi linier. Dengan mempertimbangkan bahwa sisiran model memiliki ragam yang tidak homogen, serta memiliki autokorelasi spasial, kita selanjutnya dapat mencoba memodelkannya dengan model regresi terboboti geografis atau GWR. Namun pada modul pembelajaran ini, kami akan perlihatkan pula bahwa pada pemodelan regresi spasial pun ternyata tetap menghasilkan sisiran dengan ragam yang heterogen.

```
LM<-lm.LMtests(fit.lm, ww,test=c("LMerr", "LMlag","RLMerr","RLMlag","SARMA"), zero.policy=0)
summary(LM)

## Lagrange multiplier diagnostics for spatial dependence
## data:
## model: lm(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data =
## datajawa)
## weights: ww
##
##          statistic parameter   p.value
## LMerr    32.3011683      1 1.320e-08 ***
## LMlag    18.2380122      1 1.949e-05 ***
## RLMerr   14.0653693      1 0.0001766 ***
## RLMlag   0.0022132      1 0.9624773
## SARMA    32.3033815      2 9.670e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

library(spatialreg)
sem<-errorsarlm(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,ww, zero.policy=T)
sar<-lagsarlm(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,ww, zero.policy=T)
gsm<-sacsarlm(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,ww, zero.policy=T)

bptest.sarlm(sem)

##
## studentized Breusch-Pagan test
##
## data:
## BP = 5.6364, df = 1, p-value = 0.01759

bptest.sarlm(sar)

##
## studentized Breusch-Pagan test
##
## data:
## BP = 4.0536, df = 1, p-value = 0.04408
```

```
bptest.sarlm(gsm)

##
## studentized Breusch-Pagan test
##
## data:
## BP = 3.8557, df = 1, p-value = 0.04958
```

## 3.2 Basic GWR

Kita dapat menggunakan fungsi `gwr` pada package `spgwr` untuk menyusun model GWR pada R software, seperti pada program berikut ini.

```
library(spgwr)

## Warning: package 'spgwr' was built under R version 4.0.3

## NOTE: This package does not constitute approval of GWR
## as a method of spatial analysis; see example(gwr)

coordinates(datajawa)<-c("Longitude","Latitude")
gwr20 <- gwr(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,bandwidth=20)
gwr20

## Call:
## gwr(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data = datajawa,
##      bandwidth = 20)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 20
## Summary of GWR coefficient estimates at data points:
##                               Min. 1st Qu. Median 3rd Qu. Max. Global
## X.Intercept.          24.6607 24.6971 24.7366 24.7620 24.8026 24.7340
## Rata.Rata.Lama.Sekolah -1.8708 -1.8673 -1.8651 -1.8617 -1.8585 -1.8657
```

Kita dapat pula mengganti bandwidth sesuai dengan yang diinginkan. Selanjutnya kita akan bandingkan perbedaan akibat penentuan bandwidth yang berbeda-beda tersebut.

```
gwr3 <- gwr(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa, bandwidth=3)
gwr3
```

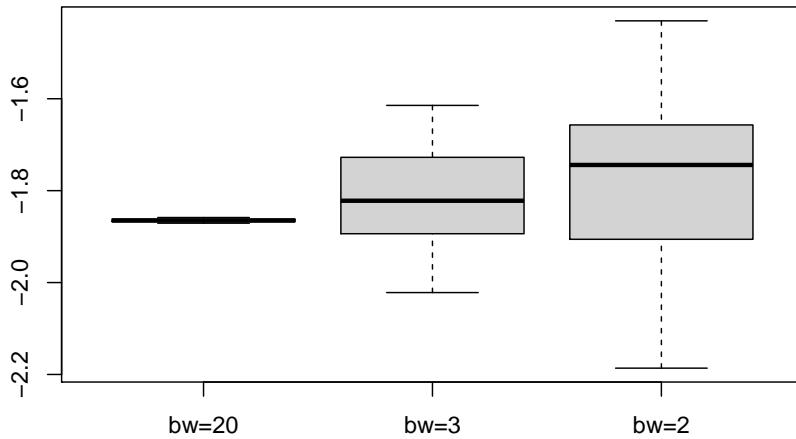
62 CHAPTER 3. GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

```
## Call:  
## gwr(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data = datajawa,  
##       bandwidth = 3)  
## Kernel function: gwr.Gauss  
## Fixed bandwidth: 3  
## Summary of GWR coefficient estimates at data points:  
##                                     Min. 1st Qu. Median 3rd Qu. Max. Global  
## X.Intercept.          21.8567 23.3823 24.6838 25.4191 26.4774 24.7340  
## Rata.Rata.Lama.Sekolah -2.0218 -1.8937 -1.8219 -1.7275 -1.6145 -1.8657
```

```
gwr2 <- gwr(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,bandwidth=2)  
gwr2
```

```
## Call:  
## gwr(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data = datajawa,  
##       bandwidth = 2)  
## Kernel function: gwr.Gauss  
## Fixed bandwidth: 2  
## Summary of GWR coefficient estimates at data points:  
##                                     Min. 1st Qu. Median 3rd Qu. Max. Global  
## X.Intercept.          19.5961 22.5941 24.3392 25.7094 27.7499 24.7340  
## Rata.Rata.Lama.Sekolah -2.1862 -1.9058 -1.7441 -1.6570 -1.4305 -1.8657
```

```
betabw20 <- gwr20$SDF$Rata.Rata.Lama.Sekolah  
betabw3 <- gwr3$SDF$Rata.Rata.Lama.Sekolah  
betabw2 <- gwr2$SDF$Rata.Rata.Lama.Sekolah  
boxplot(betabw20, betabw3, betabw2, names=c("bw=20", "bw=3", "bw=2"))
```



Output di atas memperlihatkan bahwa bandwidth yang lebih besar cenderung menghasilkan penduga koefisien model dengan rentang nilai yang lebih sempit. Sebaliknya, jika bandwidth yang digunakan lebih kecil, maka penduga koefisien model cenderung memiliki rentang nilai yang lebih lebar.

### 3.3 Menentukan Bandwidth Optimal

Penentuan bandwidth yang optimal dapat ditentukan berdasarkan kriteria AIC atau CV. Selain itu, kita juga dapat memilih fungsi pembobot kernel yang ingin digunakan pada pemodelan GWR.

```
bw1 <- gwr.sel(Kemiskinan~Rata.Rata.Lama.Sekolah, data=datajawa) # default method is CV

## Bandwidth: 3.450257 CV score: 946.7623
## Bandwidth: 5.577059 CV score: 985.8033
## Bandwidth: 2.135821 CV score: 877.1735
## Bandwidth: 1.323454 CV score: 806.2682
## Bandwidth: 0.8213845 CV score: 752.5299
## Bandwidth: 0.5110882 CV score: 697.962
## Bandwidth: 0.3193146 CV score: 699.4435
## Bandwidth: 0.4257648 CV score: 686.988
## Bandwidth: 0.4182171 CV score: 686.7128
## Bandwidth: 0.4036248 CV score: 686.6801
```

64 CHAPTER 3. GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

```
## Bandwidth: 0.3714212 CV score: 689.1402
## Bandwidth: 0.4101954 CV score: 686.6098
## Bandwidth: 0.4102361 CV score: 686.6098
## Bandwidth: 0.4101547 CV score: 686.6098
## Bandwidth: 0.4076605 CV score: 686.62
## Bandwidth: 0.410114 CV score: 686.6098
## Bandwidth: 0.4101547 CV score: 686.6098

gwr01 <- gwr(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,bandwidth=bw1)

bw2 <- gwr.sel(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,method="aic")

## Bandwidth: 3.450257 AIC: 587.2351
## Bandwidth: 5.577059 AIC: 591.8688
## Bandwidth: 2.135821 AIC: 578.5693
## Bandwidth: 1.323454 AIC: 569.3341
## Bandwidth: 0.8213845 AIC: 562.5431
## Bandwidth: 0.5110882 AIC: 558.7068
## Bandwidth: 0.3193146 AIC: 575.363
## Bandwidth: 0.6296109 AIC: 559.226
## Bandwidth: 0.5177358 AIC: 558.6444
## Bandwidth: 0.5525482 AIC: 558.5357
## Bandwidth: 0.5819835 AIC: 558.6792
## Bandwidth: 0.547683 AIC: 558.531
## Bandwidth: 0.5465411 AIC: 558.5307
## Bandwidth: 0.5462381 AIC: 558.5307
## Bandwidth: 0.5462788 AIC: 558.5307
## Bandwidth: 0.5461974 AIC: 558.5307
## Bandwidth: 0.5462381 AIC: 558.5307

gwr02 <- gwr(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,bandwidth=bw2)

bwbs1 <- gwr.sel(Kemiskinan~Rata.Rata.Lama.Sekolah,data=datajawa,gweight=gwr.bisquare)

## Bandwidth: 3.450257 CV score: 816.2604
## Bandwidth: 5.577059 CV score: 893.215
## Bandwidth: 2.135821 CV score: 765.5706
## Bandwidth: 1.323454 CV score: 696.4003
## Bandwidth: 0.8213845 CV score: 722.8235
## Bandwidth: 1.323495 CV score: 696.395
## Bandwidth: 1.633776 CV score: 719.824
## Bandwidth: 1.421531 CV score: 697.7054
## Bandwidth: 1.36793 CV score: 694.5978
## Bandwidth: 1.355794 CV score: 694.4734
```

```

## Bandwidth: 1.358597 CV score: 694.4715
## Bandwidth: 1.357479 CV score: 694.4699
## Bandwidth: 1.357438 CV score: 694.4699
## Bandwidth: 1.357397 CV score: 694.4699
## Bandwidth: 1.357438 CV score: 694.4699

gwr03 <- gwr(Kemiskinan~Rata.Rata.Sekolah,data=datajawa,gweight=gwr.bisquare,
               bandwidth=bwbs1)

bwbs2 <- gwr.sel(Kemiskinan~Rata.Rata.Sekolah,data=datajawa,
                  gweight=gwr.bisquare,method="aic")

## Bandwidth: 3.450257 AIC: 570.608
## Bandwidth: 5.577059 AIC: 580.6548
## Bandwidth: 2.135821 AIC: 564.1267
## Bandwidth: 1.323454 AIC: 558.0575
## Bandwidth: 0.8213845 AIC: 575.6082
## Bandwidth: 1.633751 AIC: 558.9865
## Bandwidth: 1.310745 AIC: 558.1617
## Bandwidth: 1.435388 AIC: 557.6479
## Bandwidth: 1.429706 AIC: 557.6416
## Bandwidth: 1.420138 AIC: 557.6375
## Bandwidth: 1.420321 AIC: 557.6375
## Bandwidth: 1.420399 AIC: 557.6375
## Bandwidth: 1.42044 AIC: 557.6375
## Bandwidth: 1.420399 AIC: 557.6375

gwr04 <- gwr(Kemiskinan~Rata.Rata.Sekolah,data=datajawa,gweight=gwr.bisquare,
               bandwidth=bwbs2)

```

### 3.4 Menentukan Model Terbaik

Penentuan model terbaik dapat ditentukan berdasarkan beberapa kriteria tertentu. Fungsi `gwr` memungkinkan kita untuk mengevaluasi model berdasarkan AIC dan global quasi- $R^2$ , dengan terlebih dulu menambahkan argumen `hatmatrix=TRUE`.

```

gwr01 <- gwr(Kemiskinan~Rata.Rata.Sekolah,data=datajawa,
               hatmatrix=T, bandwidth=bw1)
gwr01

## Call:

```

## 66 CHAPTER 3. GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

```
## gwr(formula = Kemiskinan ~ Rata.Rata.Lama.Sekolah, data = datajawa,
##       bandwidth = bw1, hatmatrix = T)
## Kernel function: gwr.Gauss
## Fixed bandwidth: 0.4101547
## Summary of GWR coefficient estimates at data points:
##           Min. 1st Qu. Median 3rd Qu. Max. Global
## X.Intercept. 12.68449 18.33784 23.15972 26.20076 38.10046 24.7340
## Rata.Rata.Lama.Sekolah -4.20789 -2.05946 -1.55904 -1.22414 -0.46292 -1.8657
## Number of data points: 119
## Effective number of parameters (residual: 2traceS - traceS'S): 34.33575
## Effective degrees of freedom (residual: 2traceS - traces'S): 84.66425
## Sigma (residual: 2traceS - traceS'S): 2.291766
## Effective number of parameters (model: traceS): 25.10699
## Effective degrees of freedom (model: traceS): 93.89301
## Sigma (model: traceS): 2.176225
## Sigma (ML): 1.933067
## AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 562.1913
## AIC (GWR p. 96, eq. 4.22): 519.682
## Residual sum of squares: 444.673
## Quasi-global R2: 0.7874434
```

Fungsi `gwr` memberikan nilai AIC dengan tiga pendekatan, sehingga kita memperoleh AICc, AICb, dan AICh.

```
gwr01$results$AICc
```

```
## [1] 569.1585
```

```
gwr01$results$AICb
```

```
## [1] 562.1913
```

```
gwr01$results$AICh
```

```
## [1] 519.682
```

Model terbaik adalah yang memiliki nilai AIC terkecil (bisa juga negatif) dan nilai global quasi- $R^2$  yang terbesar. Namun demikian, kriteria tersebut tidak memberikan informasi inferensia apapun terkait signifikansi model GWR.

Beberapa pendekatan uji dapat dilakukan untuk menguji  $H_0$  yang menyatakan bahwa model GWR tidak lebih baik daripada model OLS (regresi linier klasik), seperti yang dapat dilihat pada program-program berikut ini.

```
BFC02.gwr.test(gwr01)
```

```
##  
## Brunsdon, Fotheringham & Charlton (2002, pp. 91-2) ANOVA  
##  
## data: gwr01  
## F = 2.2103, df1 = 117.000, df2 = 84.664, p-value = 7.714e-05  
## alternative hypothesis: greater  
## sample estimates:  
## SS OLS residuals SS GWR residuals  
## 982.8726 444.6730
```

```
BFC99.gwr.test(gwr01)
```

```
##  
## Brunsdon, Fotheringham & Charlton (1999) ANOVA  
##  
## data: gwr01  
## F = 3.169, df1 = 87.18, df2 = 96.67, p-value = 2.996e-08  
## alternative hypothesis: greater  
## sample estimates:  
## SS GWR improvement SS GWR residuals  
## 538.1996 444.6730
```

```
LMZ.F1GWR.test(gwr01)
```

```
##  
## Leung et al. (2000) F(1) test  
##  
## data: gwr01  
## F = 0.62521, df1 = 96.67, df2 = 117.00, p-value = 0.00875  
## alternative hypothesis: less  
## sample estimates:  
## SS OLS residuals SS GWR residuals  
## 982.8726 444.6730
```

```
LMZ.F2GWR.test(gwr01)
```

```
##  
## Leung et al. (2000) F(2) test  
##  
## data: gwr01
```

## 68 CHAPTER 3. GEOGRAPHICALLY WEIGHTED REGRESSION (GWR)

```
## F = 1.9813, df1 = 47.916, df2 = 117.000, p-value = 0.001549
## alternative hypothesis: greater
## sample estimates:
##   SS OLS residuals SS GWR improvement
##                 982.8726           538.1996
```

Berdasarkan output di atas, seluruh uji menunjukkan nilai  $p$ -value yang lebih kecil daripada taraf nyata 0.05, artinya  $H_0$  dapat ditolak, dan kita dapat menyimpulkan bahwa model GWR lebih baik daripada OLS, pada taraf nyata 5%.

```
LMZ.F3GWR.test(gwr01)
```

```
##
## Leung et al. (2000) F(3) test
##
##                               F statistic Numerator d.f. Denominator d.f.      Pr(>)
## (Intercept)                2.6985        46.6539          96.67 1.988e-05
## Rata.Rata.Lama.Sekolah    2.1191        25.3140          96.67  0.004783
##
## (Intercept)             ***
## Rata.Rata.Lama.Sekolah **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 3.5 Menginterpretasikan Hasil Pemodelan GWR

Koefisien model GWR bersifat lokal, sehingga nilai penduga koefisien akan diperoleh pada setiap titik pengamatan. Oleh karenanya, interpretasi model GWR seringkali dilakukan dengan membuat visualisasi dalam bentuk peta, baru kemudian menginterpretasikannya.

```
str(gwr01$SDF)
```

```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
##   ..@ data      :'data.frame': 119 obs. of 12 variables:
##     ...$ sum.w            : num [1:119] 7.49 14.3 13.86 13.78 13.72 ...
##     ...$ (Intercept)       : num [1:119] 17.1 15.5 15.8 15.8 15.7 ...
##     ...$ Rata.Rata.Lama.Sekolah : num [1:119] -1.21 -1.08 -1.11 -1.1 -1.09 ...
##     ...$ (Intercept)_se     : num [1:119] 3.52 2.76 2.79 2.83 2.85 ...
##     ...$ Rata.Rata.Lama.Sekolah_se: num [1:119] 0.35 0.287 0.29 0.294 0.295 ...
##     ...$ gwr.e              : num [1:119] 5.168 -0.198 0.191 0.206 -0.93
```

```

## ...$ pred : num [1:119] 6.81 3.03 2.95 3.38 4.32 ...
## ...$ pred.se : num [1:119] 0.761 0.775 0.794 0.708 0.554 ...
## ...$ localR2 : num [1:119] 0.485 0.633 0.642 0.616 0.602 ...
## ...$ (Intercept)_se_EDF : num [1:119] 3.71 2.91 2.94 2.98 3 ...
## ...$ Rata.Rata.Lama.Sekolah_se_EDF: num [1:119] 0.369 0.303 0.306 0.309 0.311 ...
## ...$ pred.se : num [1:119] 0.801 0.816 0.836 0.746 0.583 ...
## ..@ coords.nrs : num(0)
## ..@ coords : num [1:119, 1:2] 107 107 107 107 107 ...
## ...- attr(*, "dimnames")=List of 2
## ... . : chr [1:119] "1" "2" "3" "4" ...
## ... . : chr [1:2] "Longitude" "Latitude"
## ..@ bbox : num [1:2, 1:2] 105.69 -8.24 114.37 -5.8
## ...- attr(*, "dimnames")=List of 2
## ... . : chr [1:2] "Longitude" "Latitude"
## ... . : chr [1:2] "min" "max"
## ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
## ... .@ projargs: chr NA

```

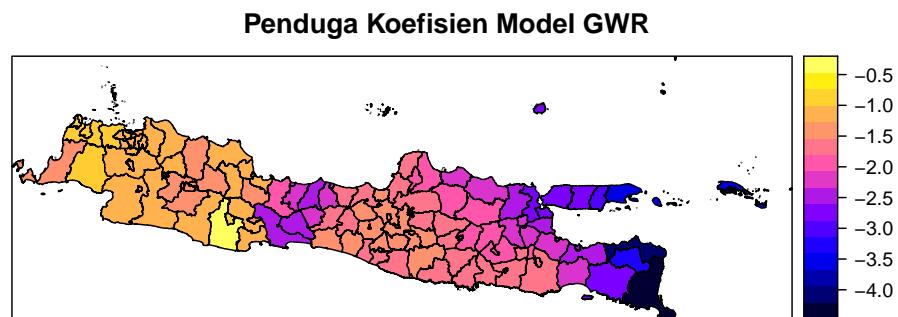
Beberapa nilai dapat diperoleh langsung dari output fungsi `gwr`, yaitu penduga koefisien, hasil prediksi model, serta *local-R<sup>2</sup>*. Selanjutnya, interpretasi dapat disesuaikan dengan konteks penelitian.

```

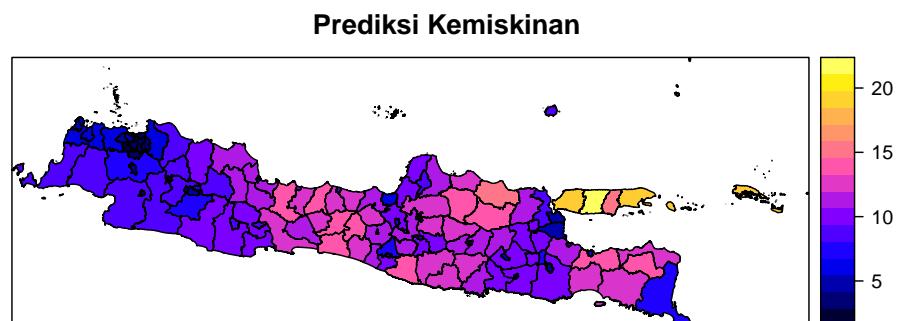
petajawa$beta<-gwr01$SDF$Rata.Rata.Sekolah
petajawa$pred<-gwr01$SDF$pred
petajawa$localR2<-gwr01$SDF$localR2

spplot(petajawa, "beta", main="Penduga Koefisien Model GWR")

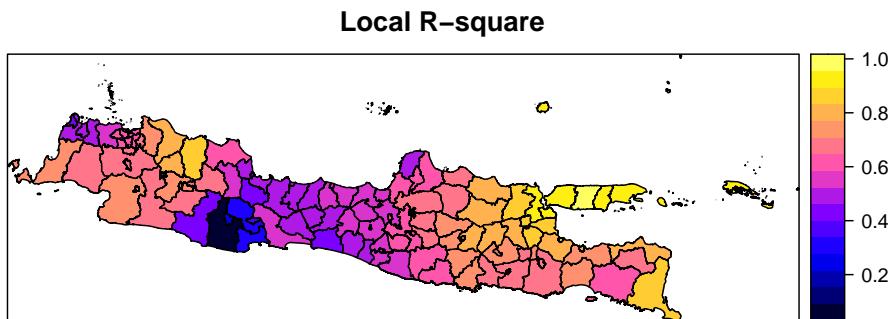
```



```
spplot(petajawa, "pred", main="Prediksi Kemiskinan")
```



```
spplot(petajawa, "localR2", main="Local R-square")
```



### 3.6 Sumber Pustaka

Brazil, N. (n.d.). Geographically weighted regression. CRD 230: Spatial Methods in Community Research. [https://crd230.github.io/gwr.html#ordinary\\_least\\_squares\\_regression](https://crd230.github.io/gwr.html#ordinary_least_squares_regression)

Brunsdon, C. 2015. Geographically Weighted Regression. [https://rstudio-pubs-static.s3.amazonaws.com/176883\\_06a3fa1fc77444be85e94dcd97ba9a34.html](https://rstudio-pubs-static.s3.amazonaws.com/176883_06a3fa1fc77444be85e94dcd97ba9a34.html)

Dennett, A. (2014, November 17). An introduction to geographically weighted regression in R. [https://rstudio-pubs-static.s3.amazonaws.com/44975\\_0342ec49f925426fa16ebcdc28210118.html](https://rstudio-pubs-static.s3.amazonaws.com/44975_0342ec49f925426fa16ebcdc28210118.html)



## Chapter 4

# Regresi Data Panel Spasial

Pemodelan ini merupakan salah satu alternatif yang dapat dilakukan ketika data yang kita kaji selain merupakan data spasial, juga merupakan data deret waktu.

Tahapan yang dapat dilakukan dalam melakukan pemodelan regresi data panel adalah sebagai berikut:

- (1) melakukan eksplorasi hubungan antar peubah, sebaran data, dsb.
- (2) menganalisa data panel dengan melakukan pengujian berikut:
  - Uji Breusch-Pagan untuk menguji adanya efek waktu, individu atau keduanya
  - Uji Chow untuk menguji signifikansi antara model gabungan dan model pengaruh tetap
  - Uji Hausman untuk menguji signifikansi antara model pengaruh acak dengan model pengaruh tetap
- (3) melakukan pemodelan data panel sesuai dengan indikasi yang diperoleh pada tahap (1) s.d (3)
- (4) mengevaluasi kebaikan model

Sedangkan tahapan dalam pemodelan data panel spasial adalah: (Rakhmi, 2015)

- (1) Eksplorasi Data
- (2) Uji ketergantungan spasial, dapat menggunakan uji CD Pesaran dengan hipotesis sebagai berikut:

$$H_0 : \rho_{ij} = \rho_{ji} = \text{korelasi}(\varepsilon_{it}, \varepsilon_{jt}) = 0, \text{ untuk } i \neq j$$

$$H_1 : \rho_{ij} = \rho_{ji} \neq 0, \text{ untuk beberapa } i \neq j$$

- (3) Bila perlu, menguji autokorelasi spasial pada peubah penjelas, dapat menggunakan uji Moran.
- (4) Sesuai dengan indikasi pada tahapan sebelumnya, kita dapat melakukan pemodelan data panel spasial, dengan menggunakan matriks pembobot spasial yang sesuai.
- (5) Mengevaluasi kebaikan model yang disusun pada tahap (4).
- (6) Menginterpretasikan model.

## 4.1 Ilustrasi: Kemiskinan di Pulau Jawa

Sebagai ilustrasi, modul ini akan membahas studi kasus yang diteliti oleh Agusti (2015). Pada ilustrasi ini, kita akan mengamati persentasi kemiskinan di kabupaten dan kota di Pulau Jawa pada tahun 2008-2012, dengan rincian peubah sebagai berikut:

- $Y$  = Persentase Kemiskinan di kabupaten/kota di Pulau Jawa
- $X_1$  = Indeks Pembangunan Manusia (IPM)
- $X_2$  = Tingkat Pengangguran Terbuka (TPT)
- $X_3$  = Pertumbuhan Ekonomi (PDRB)

Pertama-tama, kita impor data peta tematik Jawa Tengah dan data pengamatan ke dalam R.

```
data<-read.csv("data panel Jawa.csv", header = T)
head(data)
```

	Id	Provinsi	KabKot	Tahun	Kemiskinan	IPM	PDRB	TPT
## 1	3101	DKI JAKARTA	Kep. Seribu	2008	13.56	70.14	3.45	11.67
## 2	3101	DKI JAKARTA	Kep. Seribu	2009	12.66	70.50	3.48	11.57
## 3	3101	DKI JAKARTA	Kep. Seribu	2010	13.01	70.82	4.06	9.89
## 4	3101	DKI JAKARTA	Kep. Seribu	2011	11.53	71.16	5.34	11.38
## 5	3101	DKI JAKARTA	Kep. Seribu	2012	11.62	71.45	5.63	13.97
## 6	3171	DKI JAKARTA	Kota Jakarta Selatan	2008	3.41	79.00	152.15	11.96

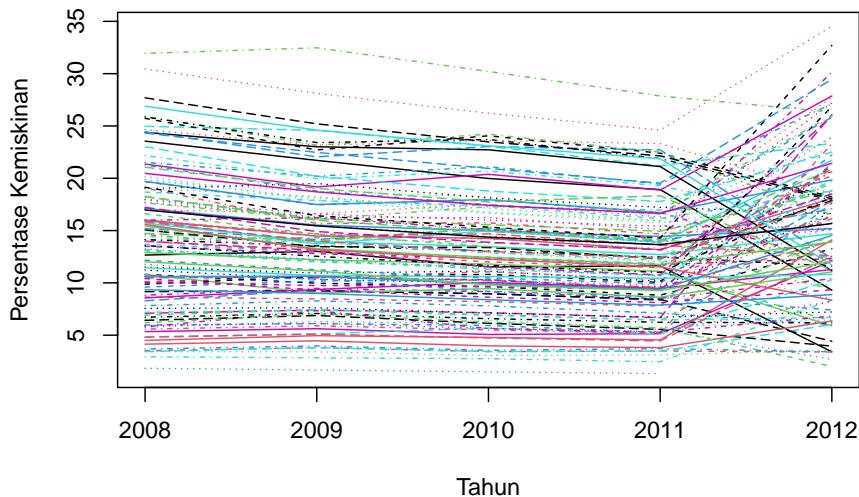
## 4.2 Eksplorasi Data

```

y<-matrix(NA, nrow=5, ncol=119) # 5 tahun, 118 kota/kabupaten
y[,1]<-2008:2012
a=2
for (i in 1:118)
{
  y[,i+1]<-data[a:(a+4),5]
  a=a+5
}

palette("default")
y<-as.data.frame(y)
matplot(y[,1], y[,-1], type="l", pch=19, ylab="Percentase Kemiskinan", xlab="Tahun")

```

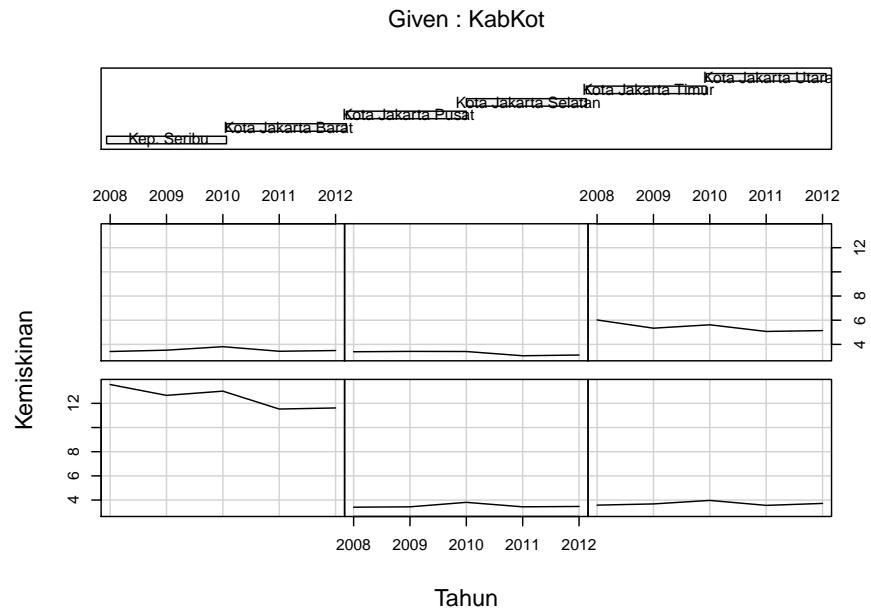


```

#install.packages("foreign")
library(foreign)

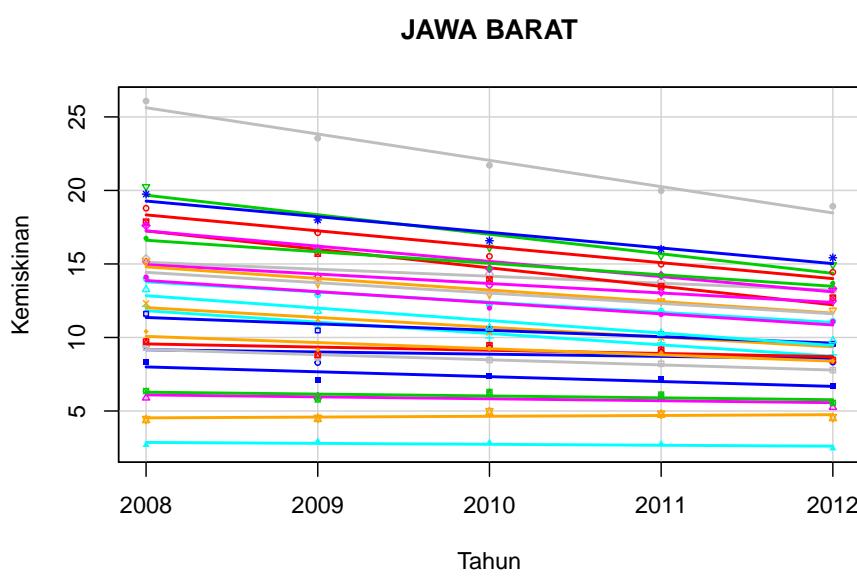
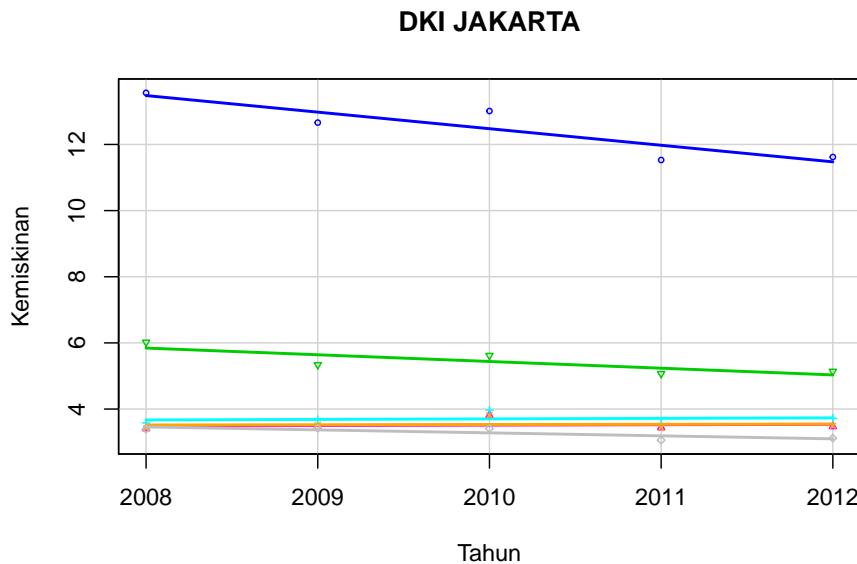
data1<-subset(data,data$Provinsi=="DKI JAKARTA")
coplot(Kemiskinan~Tahun|KabKot, type="l", data=data1)

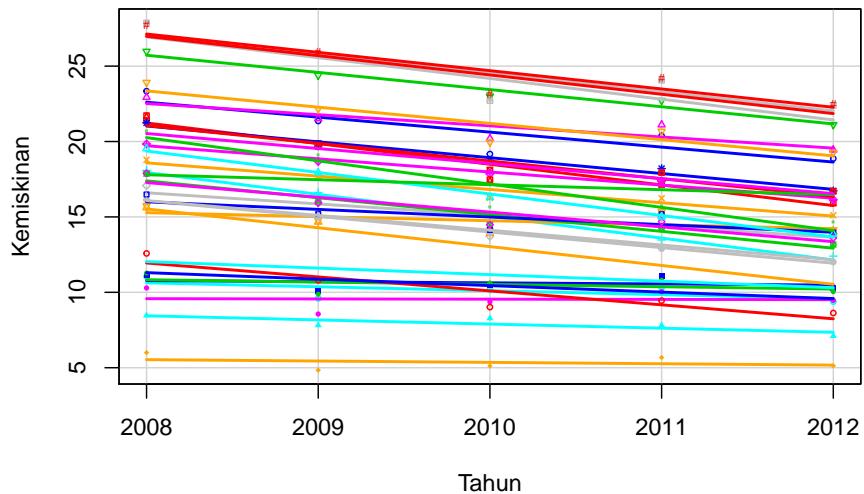
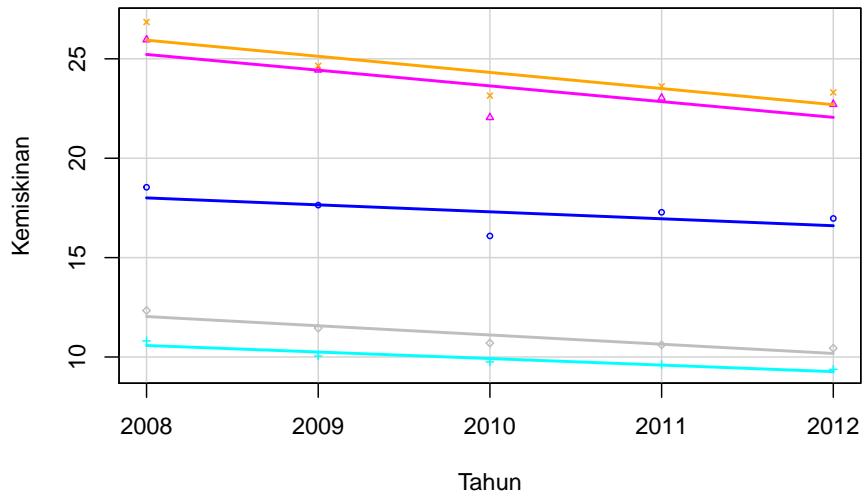
```

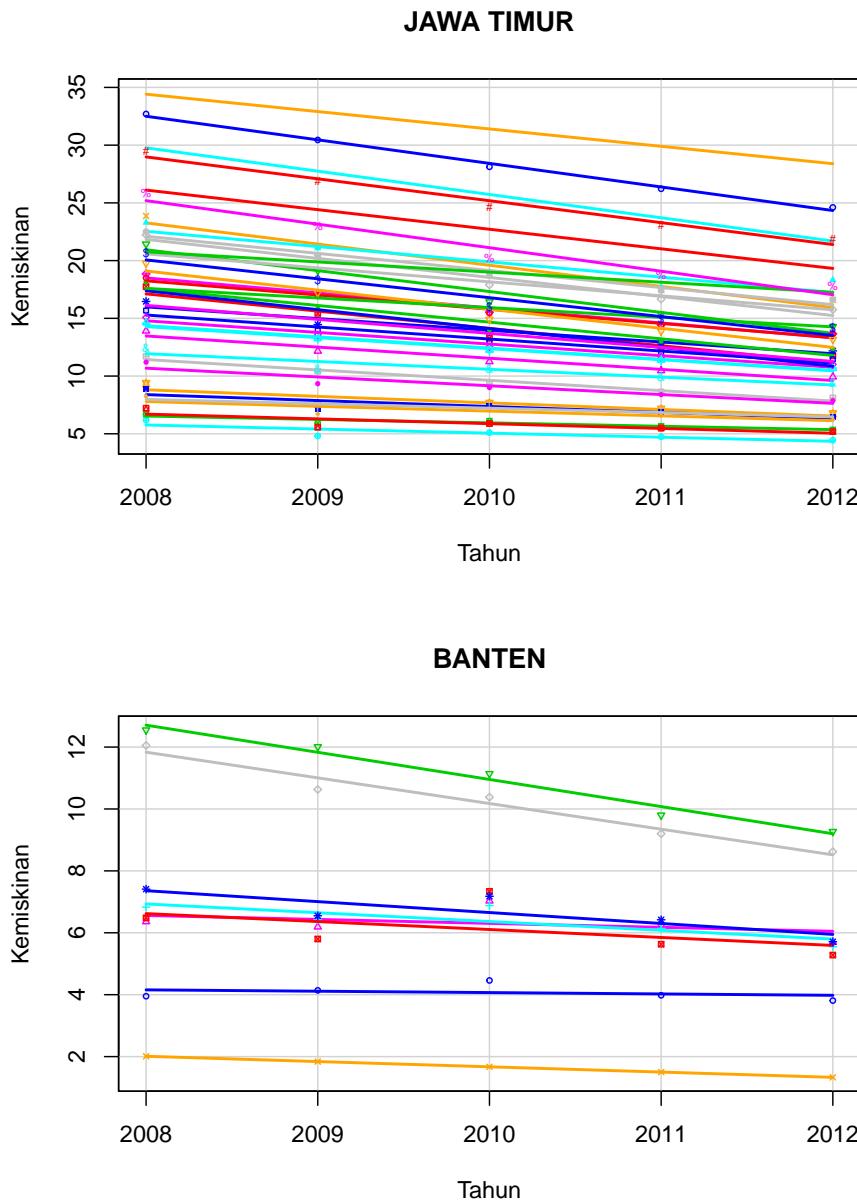


```

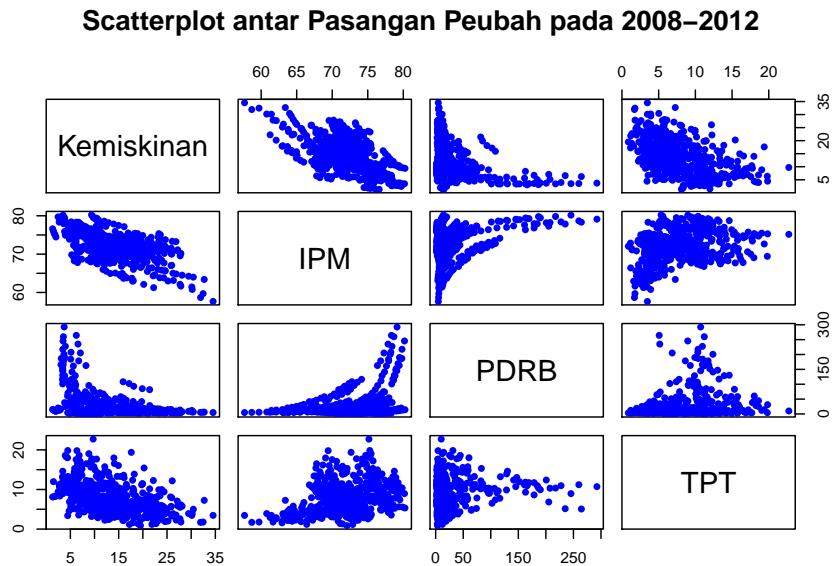
prov<-unique(data$Provinsi)
for (i in 1:length(prov))
{
  data1<-subset(data,data$Provinsi==prov[i])
  car::scatterplot(Kemiskinan~Tahun|KabKot, boxplots=F, smooth=F, legend=F,
                   grid=T,data=data1, reset.par=T, cex=0.5, main=prov[i])
}
  
```



**JAWA TENGAH****DI YOGYAKARTA**



```
plot(data[,5:8], pch=16, col="blue", main="Scatterplot antar Pasangan Peubah pada 2008-2012")
```

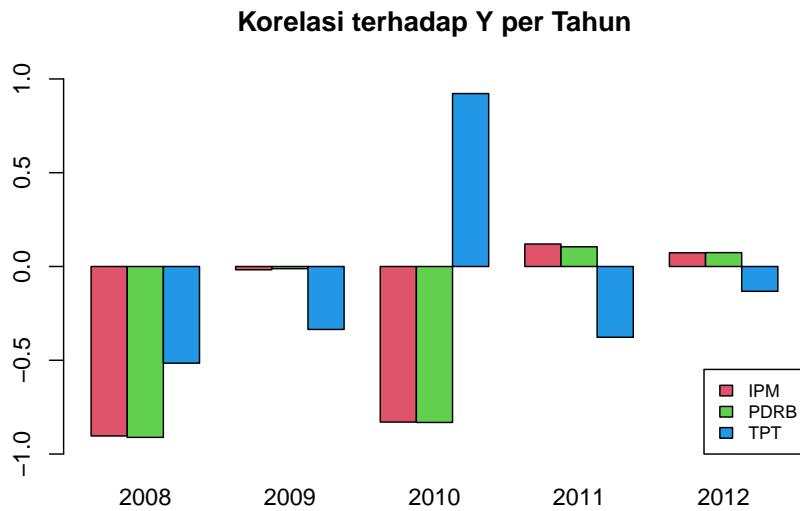


Scatterplot memperlihatkan bahwa IPM dan PDRB cenderung memiliki hubungan yang berbanding terbalik dengan persentase kemiskinan. Sedangkan pola hubungan antara persentase kemiskinan dengan peubah TPT memiliki ragam yang cukup besar sehingga agak sulit untuk mengidentifikasinya.

```
a=1; korelasipertahun<-NULL;
for (i in 1:5)
{
  korelasipertahun<-cbind(korelasipertahun,korelasipertahun)
  korelasipertahun<-cor(data[a:(a+4),5:8])[-1,1]
  a=a+5
}
colnames(korelasipertahun)<-2008:2012

barplot(korelasipertahun , beside=T, ylim=c(-1,1), col=2:4,main="Korelasi terhadap Y p
legend("bottomright",c("IPM", "PDRB", "TPT"),
fill = 2:4, cex=0.75)
```

#### 4.3. IDENTIFIKASI KERAGAMAN ANTAR INDIVIDU DAN ANTAR WAKTU81

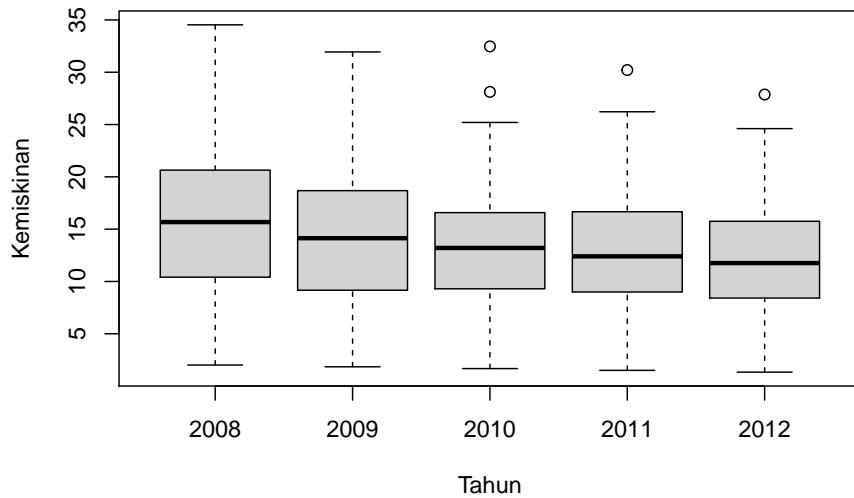


```
korelasipertahun
```

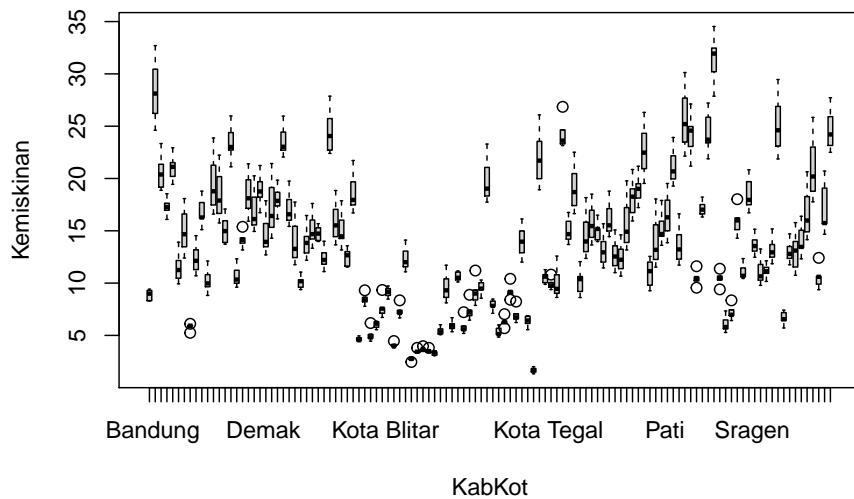
```
##          2008      2009      2010      2011      2012
## IPM -0.9035923 -0.01758226 -0.8295863  0.1198465  0.07306023
## PDRB -0.9110926 -0.01161981 -0.8315061  0.1050094  0.07401256
## TPT  -0.5155555 -0.33560415  0.9220472 -0.3772710 -0.13234028
```

### 4.3 Identifikasi Keragaman Antar Individu dan Antar Waktu

```
boxplot(Kemiskinan~Tahun, data=data)
```



```
boxplot(Kemiskinan~KabKot, data=data)
```



## 4.4 Pemodelan Data Panel

### 4.4.1 Model Gabungan

```

a<-scale(data[,5:8])
data1<-cbind(data[,c(3,4)],a)
head(data1)

##          KabKot Tahun Kemiskinan      IPM      PDRB      TPT
## 1      Kep. Seribu 2008 -0.01086532 -0.5558484 -0.5182081 1.0499174
## 2      Kep. Seribu 2009 -0.15332169 -0.4615811 -0.5174904 1.0239017
## 3      Kep. Seribu 2010 -0.09792199 -0.3777880 -0.5036157 0.5868385
## 4      Kep. Seribu 2011 -0.33218358 -0.2887578 -0.4729957 0.9744719
## 5      Kep. Seribu 2012 -0.31793795 -0.2128203 -0.4660583 1.6482777
## 6 Kota Jakarta Selatan 2008 -1.61745665  1.7641735  3.0389792 1.1253628

library(plm)

## Warning: package 'plm' was built under R version 4.0.4

## 
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
## 
##     between, lag, lead

model.pool<-plm(Kemiskinan~IPM+PDRB+TPT, data=data1, model="pooling")
summary(model.pool)

## Pooling Model
## 
## Call:
## plm(formula = Kemiskinan ~ IPM + PDRB + TPT, data = data1, model = "pooling")
## 
## Balanced Panel: n = 118, T = 5, N = 590
## 
## Residuals:
##     Min.   1st Qu.   Median   3rd Qu.   Max. 
## -1.541302 -0.465061 -0.078632  0.375658  2.393066
## 
```

```

## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 6.9268e-16 2.7939e-02 0.0000      1
## IPM         -5.2289e-01 3.1199e-02 -16.7600 < 2.2e-16 ***
## PDRB        -1.4241e-01 3.0670e-02 -4.6431 4.239e-06 ***
## TPT         -2.8226e-01 2.9722e-02 -9.4965 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 589
## Residual Sum of Squares: 269.87
## R-Squared: 0.54181
## Adj. R-Squared: 0.53947
## F-statistic: 230.985 on 3 and 586 DF, p-value: < 2.22e-16

```

Model di atas adalah model gabungan, atau sama dengan model OLS (*ordinary least squares*). Selanjutnya, kita perlu mengidentifikasi apakah terdapat pengaruh individu atau waktu, atau keduanya, serta yang manakah di antara model gabungan, tetap , atau acak yang paling sesuai untuk digunakan.

```
plmtest(model.pool,data=data1,type="bp",effect = "time")
```

```

##
## Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced
## panels
##
## data: Kemiskinan ~ IPM + PDRB + TPT
## chisq = 89.555, df = 1, p-value < 2.2e-16
## alternative hypothesis: significant effects

```

```
plmtest(model.pool,data=data1,type="bp",effect = "individual")
```

```

##
## Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels
##
## data: Kemiskinan ~ IPM + PDRB + TPT
## chisq = 893.81, df = 1, p-value < 2.2e-16
## alternative hypothesis: significant effects

```

```
#install.packages("plm")
library(plm)

plmtest(model.pool,type="bp",effect = "twoways")
```

```
##
## Lagrange Multiplier Test - two-ways effects (Breusch-Pagan) for
## balanced panels
##
## data: Kemiskinan ~ IPM + PDRB + TPT
## chisq = 983.37, df = 2, p-value < 2.2e-16
## alternative hypothesis: significant effects
```

Dari hasil di atas, dapat disimpulkan bahwa terdapat pengaruh signifikan yang diakibatkan oleh waktu dan individu.

#### 4.4.2 Uji Chow

Untuk mengidentifikasi salah satu di antara model gabungan dan model tetap yang lebih sesuai untuk digunakan, terlebih dulu dibangun model dengan pengaruh tetap. Terdapat empat kemungkinan model yang dapat diidentifikasi melalui uji Chow, yaitu:

- (1) jika kita menolak  $H_0$  pada pengujian pengaruh spesifik individu dan waktu, maka model yang sesuai adalah model pengaruh tetap komponen sisaan dua arah
- (2) jika kita menolak  $H_0$  pada pengujian pengaruh spesifik individu, maka model yang sesuai adalah model pengaruh tetap individu
- (3) jika kita menolak  $H_0$  pada pengujian pengaruh spesifik waktu, maka model yang sesuai adalah model pengaruh tetap waktu
- (4) jika kita gagal menolak  $H_0$  pada pengujian pengaruh spesifik individu dan waktu, maka model yang sesuai adalah model gabungan.

Pada ilustrasi ini, sebelumnya telah ditunjukkan bahwa terdapat pengaruh individu dan pengaruh waktu terhadap kemiskinan, sehingga pada tahap ini akan diuji model dengan komponen sisaan dua arah.

```
model.fixed<-plm(Kemiskinan~IPM+PDRB+TPT, data=data1, model="within", effect="twoways")
summary(model.fixed)

## Twoways effects Within Model
##
## Call:
## plm(formula = Kemiskinan ~ IPM + PDRB + TPT, data = data1, effect = "twoways",
##       model = "within")
##
```

```

## Balanced Panel: n = 118, T = 5, N = 590
##
## Residuals:
##      Min.    1st Qu.     Median    3rd Qu.     Max.
## -0.398939 -0.069804  0.001862  0.066238  0.361103
##
## Coefficients:
##             Estimate Std. Error t-value Pr(>|t|)
## IPM     -1.139669   0.100627 -11.3257 < 2.2e-16 ***
## PDRB    0.137278   0.029365   4.6749 3.859e-06 ***
## TPT     -0.058057   0.014710  -3.9469 9.142e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    10.232
## Residual Sum of Squares: 7.0096
## R-Squared:              0.31495
## Adj. R-Squared:          0.13227
## F-statistic: 71.2597 on 3 and 465 DF, p-value: < 2.22e-16

```

Setelah menyusun model dengan pengaruh tetap, selanjutnya kita dapat menggunakan uji Chow untuk mengidentifikasi kesesuaian model antara model gabungan atau model dengan pengaruh tetap.

```
pFtest(model.fixed, model.pool)
```

```

##
## F test for twoways effects
##
## data: Kemiskinan ~ IPM + PDRB + TPT
## F = 144.11, df1 = 121, df2 = 465, p-value < 2.2e-16
## alternative hypothesis: significant effects

```

Berdasarkan uji Chow pada output di atas,  $p\text{-value} < \alpha = 0.05$ , sehingga dapat disimpulkan bahwa model yang sesuai adalah model pengaruh tetap komponen sisaan dua arah. Selanjutnya, kita masih harus mengidentifikasi lebih lanjut apakah model pengaruh acak lebih sesuai untuk digunakan dibandingkan dengan model pengaruh tetap.

#### 4.4.3 Uji Hausman

Pada tahap ini kita akan mengidentifikasi lebih lanjut mengenai kesuaian model antara model dengan pengaruh tetap atau pengaruh acak.

```

model.random<-plm(Kemiskinan~IPM+PDRB+TPT, data=data1, model="random", effect="twoways")
summary(model.random)

## Twoways effects Random Effect Model
##      (Swamy-Arora's transformation)
##
## Call:
## plm(formula = Kemiskinan ~ IPM + PDRB + TPT, data = data1, effect = "twoways",
##       model = "random")
##
## Balanced Panel: n = 118, T = 5, N = 590
##
## Effects:
##           var   std.dev share
## idiosyncratic 0.0150743 0.1227776 0.035
## individual    0.4111332 0.6411967 0.964
## time         0.0001843 0.0135741 0.000
## theta: 0.9147 (id) 0.3601 (time) 0.3597 (total)
##
## Residuals:
##      Min. 1st Qu. Median 3rd Qu. Max.
## -0.454852 -0.083503 -0.010059  0.071157  0.533129
##
## Coefficients:
##             Estimate Std. Error z-value Pr(>|z|)
## (Intercept) 1.4693e-15 6.5106e-02 0.0000 1.000000
## IPM        -1.0194e+00 3.9214e-02 -25.9955 < 2.2e-16 ***
## PDRB        9.0152e-02 2.8350e-02   3.1800  0.001473 **
## TPT        -4.1548e-02 1.4364e-02  -2.8925  0.003822 **
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 22.996
## Residual Sum of Squares: 10.558
## R-Squared: 0.5409
## Adj. R-Squared: 0.53855
## Chisq: 690.408 on 3 DF, p-value: < 2.22e-16

```

Hipotesis yang diuji pada uji Hausman adalah:

$$H_0 : E(u_{it}|x_{k,it}) = 0 \text{ (model pengaruh acak)}$$

$$H_1 : E(u_{it}|x_{k,it}) \neq 0 \text{ (model pengaruh tetap)}$$

```
phtest(model.fixed, model.random)

##
## Hausman Test
##
## data: Kemiskinan ~ IPM + PDRB + TPT
## chisq = 57.774, df = 3, p-value = 1.756e-12
## alternative hypothesis: one model is inconsistent
```

Terlihat pada output bahwa kita dapat menolak hipotesis nol pada taraf nyata 0.05, artinya model yang lebih tepat digunakan adalah model dengan pengaruh tetap.

```
summary(model.fixed)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = Kemiskinan ~ IPM + PDRB + TPT, data = data1, effect = "twoways",
##       model = "within")
##
## Balanced Panel: n = 118, T = 5, N = 590
##
## Residuals:
##      Min.    1st Qu.     Median    3rd Qu.     Max.
## -0.398939 -0.069804  0.001862  0.066238  0.361103
##
## Coefficients:
##             Estimate Std. Error t-value Pr(>|t|)
## IPM -1.139669   0.100627 -11.3257 < 2.2e-16 ***
## PDRB 0.137278   0.029365   4.6749 3.859e-06 ***
## TPT -0.058057   0.014710  -3.9469 9.142e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 10.232
## Residual Sum of Squares: 7.0096
## R-Squared: 0.31495
## Adj. R-Squared: 0.13227
## F-statistic: 71.2597 on 3 and 465 DF, p-value: < 2.22e-16
```

Model dapat dituliskan sebagai berikut:

$$y_{it} = -1.139669x_{1it} + 0.137278x_{2it} - 0.058057x_{3it} + \hat{\mu}_i + \varepsilon_{it}$$

## 4.5 Pemodelan Data Panel Spasial

Peta persebaran jumlah penduduk miskin di kabupaten/kota Provinsi Jawa Tengah pada tahun 2015 terlihat pada gambar berikut.

```

library(rgdal)
library(spdep)
library(raster)
library(sp)
library(spatialreg)

petajawa<-readOGR(dsn='Jawamap',layer='jawa')

## OGR data source with driver: ESRI Shapefile
## Source: "D:\Research (eksternal dept)\pelatihan spasial (adj)\modeul 2\Pelatihan-Spasial\Jawa"
## with 119 features
## It has 5 fields

petajawa@data[24,] #dihilangkan dari peta karena tidak ada dalam data pengamatan

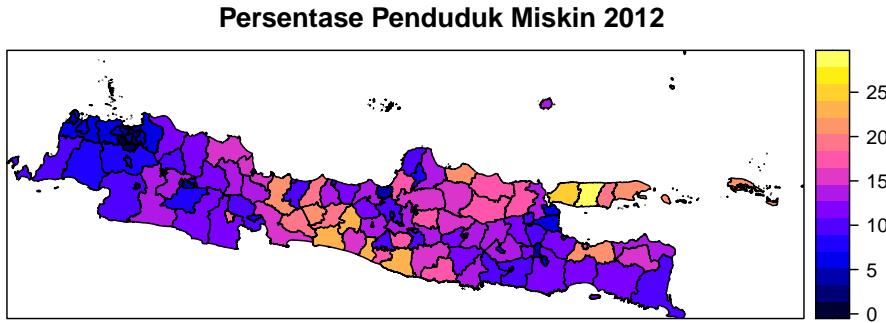
##      PROVNO KABKOTNO    PROVINSI      KABKOT ID2013
## 23          32          18 JAWA BARAT PANGANDARAN  3218

petajawa<-petajawa[-24,]

data2012<-subset(data, data$Tahun==2012)

petajawa$Kemiskinan<-data2012$Kemiskinan
spplot(petajawa, "Kemiskinan", main="Persentase Penduduk Miskin 2012")

```



#### 4.5.1 Uji Ketergantungan Spasial pada Data Panel

Seperti yang telah dijelaskan pada bagian awal modul ini, kita dapat menggunakan uji *cross-sectional dependency* (CD) Pesaran untuk mengidentifikasi ketergantungan spasial pada data panel.

```
library(splm)
```

## Warning: package 'splm' was built under R version 4.0.4

```
w<-as.matrix(1/dist(coordinates(petajawa)))
ww<-w
for(i in 1:118)
{
ww[i,]<-ww[i,]/apply(w,1,sum)[i]
}
rowSums(ww)
```

```

## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## 20 21 22 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

```

##  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
##  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80
##  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
##  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99  100
##  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118
##  1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1   1

rwtest(model.fixed, w=ww)

##
## Randomized W test for spatial correlation of order 1
##
## data: formula
## p-value < 2.2e-16
## alternative hypothesis: twosided

```

Terlihat pada output di atas bahwa p-value  $< \alpha = 0.05$  maka kita dapat menyimpulkan bahwa hipotesis nol ditolak, artinya pada data panel kemiskinan ini terdapat ketergantungan spasial pada peubah responnya. Hal ini mengindikasikan bahwa kemiskinan di suatu kab/kota, selain dipengaruhi oleh peubah penjelas yang dimilikinya, juga dipengaruhi oleh kemiskinan di kab/kota lain.

### 4.5.2 Uji Spesifikasi Model

Menurut Salima *et. al*, spesifikasi model pada pemodelan data panel spasial adalah sebagai berikut.

#### 4.5.3 Standard model: model dengan pengaruh spesifik individu

$$y_{it} = x_{it}\beta + z_i\alpha + \varepsilon_{it}$$

dimana  $z_i\alpha$  merupakan pengaruh spesifik individu.

##### Model gabungan

$$y_{it} = x_{it}\beta + \alpha + \varepsilon_{it}$$

Model ini tidak memodelkan heterogenitas individu, dan pada kasus ini penduga  $\beta$  dan  $\alpha$  diperoleh dengan Ordinary Least Squares (OLS).

##### Model Pengaruh Tetap

$$y_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}$$

### Model Pengaruh Acak

$$y_{it} = x_{it}\beta + \alpha + u_{it} \quad u_{it} = \alpha_i + \varepsilon_{it}$$

Model ini biasanya diduga dengan Generalized Least Squares (GLS).

#### 4.5.4 Efek Spasial pada Model Data Panel

##### Efek spasial pada model gabungan

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha + u_{it}$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \varepsilon_{it}$$

##### Efek spasial pada model pengaruh tetap

- (1) Spatial autoregressive model (SAR)

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \alpha_i + u_{it}$$

- (2) Spatial error model (SEM)

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \varepsilon_{it}$$

- (3) Durbin spatial model (DSM)

$$y_{it} = x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha_i + u_{it}$$

- (4) Durbin spatial error model (SDEM)

$$y_{it} = x_{it}\beta + \sum_{i \neq j} w_{ij} x_{jt}\theta + \alpha_i + u_{it}$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \varepsilon_{it}$$

- (5) Spatial autoregressive lag and error model (SARAR)

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \alpha_i + u_{it}$$

$$u_{it} = \lambda \sum_{i \neq j} m_{ij} u_{jt} + \varepsilon_{it}$$

dimana  $w_{ij}$  dan  $m_{ij}$  merupakan pembobot spasial.

##### Spatial Error Model-Random Effect

$$y_{it} = \rho \sum_{i \neq j} w_{ij} y_{jt} + x_{it}\beta + \alpha + u_{it}$$

$$u_{it} = \lambda \sum_{i \neq j} w_{ij} u_{jt} + \nu_{it}$$