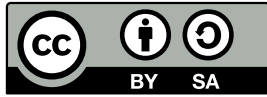





ANÁLISIS DE OPORTUNIDADES ENERGÉTICAS CON FUENTES. ALTERNATIVAS EN EL DEPARTAMENTO DE NARIÑO - ALTERNAR


INTERPOLACIÓN Y GEOESTADÍSTICA

OMAR ERNESTO CABRERA ROSERO
INGENIERO DE SISTEMAS
UNIVERSIDAD DE NARIÑO
2015





You are free:

 **to Share** – to copy, distribute and transmit this work

 **to Remix** – to adapt this work

Under the following conditions:

 **Attribution** – You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work)

 **Share Alike** – If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.

Subject to conditions outlined in the license.

This work is licensed under the *Creative Commons Attribution-ShareAlike 3.0 Unported* License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-sa/3.0/>

or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

CONTENIDO

INTRODUCCIÓN	4
1. Tutorial	5
1.1. Instalación	5
1.2. Análisis exploratorio de datos	5
1.3. Métodos de interpolación no geoestadísticos	6
1.3.1. Interpolación ponderada de distancia inversa	6
1.3.2. Regresión lineal	7
1.4. Estimación de correlación espacial: Variograma	9
1.4.1. Análisis exploratorio de variogramas	9
1.5. Ejemplo de Kriging Ordinario	9
1.6. Para tener en cuenta	11

INTRODUCCIÓN

Datos geoestadísticos son datos que podrían en principio ser medidos en cualquier lugar, pero que normalmente vienen como mediciones en un número limitado de ubicaciones de observación: piensa en la calidad de oro en un yacimiento o partículas en muestras de aires. El patrón de ubicaciones de observación general, no es de interés primordial, ya que a menudo resulta de consideraciones que van desde limitaciones económicas y físicas a ser 'representativos' o variedades de muestreo aleatorios. El interés es por lo general en la inferencia de los aspectos de la variable que no tiene medidas como mapas de los valores estimados, probabilidades de excedencia o estimaciones de los agregados más regiones indicadas o inferencia del proceso que generaron los datos. Otros problemas son la supervisión de la red de optimización: ¿Dónde deben ubicarse las nuevas observaciones o cuales ubicaciones de observación deben ser retirados de manera que el valor operacional de la vigilancia la red se maximiza.

Problemas espaciales típicas donde se utilizan geoestadística son los siguientes:

- La estimación de las calidades de unidades de minerales explotables, basado en datos de perforación.
- La interpolación de variables ambientales de la muestra o de la red de monitoreo de datos (por ejemplo, la calidad del aire, la contaminación del suelo, la cabeza del agua subterránea, conductividad hidráulica).
- La interpolación de variables físicas o químicas a partir de datos de la muestra.
- La estimación de promedios espaciales a partir de datos continuos, espacialmente correlacionadas

1 Tutorial

1.1. Instalación

Para la realización de este tutorial, se utilizaron los paquetes **sp**, **gstat**, **geoR**. los cuales se los instala de la siguiente manera:

```
install.packages("sp","gstat", "geoR", "geoRglm", "RColorBrewer")
```

1.2. Análisis exploratorio de datos

El análisis exploratorio de datos comienza con el trazado de mapas con una variable medida, los valores observados los podemos expresar utilizando el color o tamaño del símbolo.

```
> library(lattice)
> library(sp)
> data(meuse)
> coordinates(meuse) <- c("x", "y")
> spplot(meuse, "zinc", do.log = T, colorkey = TRUE)
> bubble(meuse, "zinc", do.log = T, key.space = "bottom")
```

La estructura evidente aquí es que la concentración de zinc es mayor cerca de la orilla del río Meuse. En caso de una tendencia espacial evidente, tales como la relación entre la parte superior de concentración de zinc del suelo y la distancia al río, también se puede trazar mapas con valores ajustados y residuales como lo muestra la figura 1.1.

Los análisis exploratorios se los puede analizar aún más en el contexto de modelos geoestadísticos; aunque primero vamos a referirnos a enfoques de interpolación simples y no geoestadísticos.

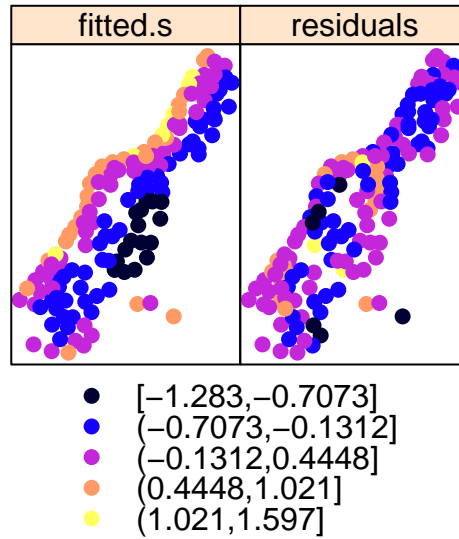


Figura 1.1: Mapa trazado con valores ajustados y residuales

1.3. Métodos de interpolación no geoestadísticos

Por lo general, la interpolación se realiza en una rejilla regular. Para el conjunto de datos Meuse, las coordenadas de puntos en una cuadrícula regular ya están definidos en el `meuse.grid` data.frame, y se convierten en un `SpatialPixelsDataFrame` así:

```
> data(meuse.grid)
> coordinates(meuse.grid) <- c("x", "y")
> meuse.grid <- as(meuse.grid, "SpatialPixelsDataFrame")
```

Alternativamente, podríamos interpolar los puntos individuales, conjuntos de puntos distribuidos irregularmente, o para un promedio de más áreas cuadradas o irregulares.

1.3.1. Interpolación ponderada de distancia inversa Interpolación ponderada basada en la distancia inversa (IDW) calcula un promedio ponderado, donde los coeficientes de ponderación para las observaciones son calculados de acuerdo a su distancia a la ubicación de interpolación.

La potencia de la distancia inversa determina el rango en cual punto(s) más cercano se prefieren sobre los puntos más distantes; para valores grandes IDW converge a la inter-

polación de un vecino más cercano. Puede ajustarse, por ejemplo usando la validación cruzada. IDW también se puede utilizar dentro de los alrededores de búsqueda local.

Debido a que el paquete **spatstat** también ofrece una función llamada `idw`, que desambigua los dos, debería **spatstat** ser cargado en el espacio de trabajo de la sesión, llamando a la función `gstat` usando el operador “::” para elegir el `idw` deseado:

```
> data(meuse)
> data(meuse.grid)
> coordinates(meuse) <- c("x", "y")
> coordinates(meuse.grid) <- c("x", "y")
> idw.out <- gstat::idw(zinc ~ 1, meuse, meuse.grid, idp = 2.5)
[inverse distance weighted interpolation]
> as.data.frame(idw.out)[1:5, ]
```

La variable de salida se llama `var1.pred`, y los valores de `var1.var` son NA porque la distancia inversa no proporciona varianzas de predicción de error.

Resultados de interpolación de distancia inversa generalmente en mapas que son muy similares a los mapas kriged¹ cuando se utiliza un variograma sin o con una pequeña nugget². En contraste con kriging, para considerar unicamente distancias a la ubicación de predicciones ignora la configuración espacial de las observaciones; esto puede dar lugar a efectos no deseados si los lugares de observación están fuertemente agrupadas. Otra diferencia es que los pesos están garantizados para estar entre 0 y 1, lo que resulta en valores interpolados, nunca fuera del rango de valores observados.

1.3.2. Regresión lineal Para la predicción espacial utilizando modelos lineales simples, podemos utilizar `lm` así:

```
> zn.lm <- lm(log(zinc) ~ sqrt(dist), meuse)
> meuse.grid$pred <- predict(zn.lm, meuse.grid)
> meuse.grid$se.fit <- predict(zn.lm, meuse.grid,
+ se.fit = TRUE)$se.fit
```

¹Proceso de regresión Gaussiana

²La altura del salto de la semivariogram en la discontinuidad en el origen

Alternativamente, el método predecir utilizado aquí puede proporcionar los intervalos de predicción o de confianza para un nivel de confianza dado. Alternativamente, podemos usar la función Krige en gstat para esto.

```
> meuse.lm <- krige(log(zinc) ~ sqrt(dist), meuse, meuse.grid)
[ordinary or weighted least squares prediction]
```

Esto en este caso no Krige ningún variograma es especificado, pero utiliza lineal regresión.

Utilizado de esta forma, el resultado es idéntico al de lm. Sin embargo, también se puede utilizar para predecir con modelos de regresión que estén colocados dentro de los alrededores locales en torno a una ubicación de predicción o proporcionan valores de las medias pronosticados para las áreas espaciales. La varianza devuelta es la varianza de predicción de error en la predicción de puntos o la varianza de estimación de error cuando se utiliza para bloques.

Se obtiene una forma especial de regresión lineal cuando se utilizan polinomios de coordenadas espaciales para predictores, por ejemplo para un polinomio de segundo orden

```
> meuse.tr2 <- krige(log(zinc) ~ 1, meuse, meuse.grid,
+ degree = 2)
[ordinary or weighted least squares prediction]
```

Esta forma se denomina análisis de superficie de tendencia. Es posible usar lm para el análisis de superficie de tendencia, por ejemplo, para la tendencia de segundo orden con una fórmula que utiliza I para tratar poderes y productos "tal cual":

```
> lm(log(zinc) ~ I(x^2) + I(y^2) + I(x * y) + x + y, meuse)
```

o la forma corta

```
> lm(log(zinc) ~ poly(x, y, degree = 2), meuse)
```

En la primera forma "lm" no estandariza las coordenadas, que a menudo se produce un gran número cuando se inicia. La segunda forma estandarizar coordenadas de tal manera

que no se puede utilizar en una subsecuencia “predict” llamada con diferentes rangos de coordenadas. Ajuste de superficie de tendencia es muy sensible a las observaciones de la periferia. Otro lugar para buscar análisis de superficie de tendencia es la función `surf.ls` en el paquete `espaial`.

1.4. Estimación de correlación espacial: Variograma

En geoestadística la correlación espacial se modela por el variograma en lugar de un correlograma o covariogram, en gran parte por razones históricas. Aquí, la palabra variograma se utiliza como sinónimo de semivariogram. el variograma traza semivarianza como una función de la distancia.

1.4.1. Análisis exploratorio de variogramas Una forma sencilla de reconocer que la correlación espacial está presente o no es para hacer gráficos de dispersión de pares $Z(s_{(i)})$ y $Z(s_{(j)})$, agrupados en función de su distancia de separación $h_{ij} = \|s_i - s_j\|$.

```
> data(meuse)
> coordinates(meuse) <- c("x", "y")
> hscat(log(zinc) ~ 1, meuse, (0:9) * 100)
```

Donde la franja de texto indica distancia de clases, y las muestras de correlaciones se muestran en cada panel.

1.5. Ejemplo de Kriging Ordinario

Para este ejemplo se utilizarán datos de temperatura que fueron obtenidos de las imágenes satelitales de Landsat 7 en el mes de junio de 2014, se toma 171 datos distribuidos alrededor del departamento de Nariño (Colombia).

Primero se carga las librerías requeridas para este ejemplo así:

```
> library(rgdal)
> library(maptools)
> library(gstat)
> library(sp)
```

```
> library(geoR)
```

Ahora se lee los datos:

```
> data<-read.table("temp.csv",sep=" ",header=T)
```

Para hacer un breve análisis exploratorio, en el cual se puede mostrar un trazo de X respecto a Y, los datos respecto a X, los datos respecto a Y, y el histograma de los datos y la densidad de los puntos.

```
> geoTemp <- as.geodata(data[,1:2],3)
> plot(geoTemp)
```

Se convierte los datos en un SpatialPointsDataFrame

```
> coordinates(data)=~X+Y
```

Se proyecta al sistema de coordenadas 3857

```
> proj4string(data)=CRS("+init=epsg:3857")
```

Se carga una grilla, que tiene puntos regulares con un espaciado de mil metros, dentro del departamento de Nariño, se lo convierte en SpatialPointsDataFrame y se proyecta al sistema de coordenadas 3857

```
> study.grid <-read.table("grid1000.csv",sep=" ",header=T)
> coordinates(study.grid)=~X+Y
> proj4string(study.grid)=CRS("+init=epsg:3857")
```

Se ajusta el variograma y se realiza el primer trazo

```
> plot(variogram(Temp~1,data))
```

Se crea un modelo del variograma

```

> study.grid<- as(study.grid, "SpatialPixelsDataFrame")
> mod<-vgm( psill=var(data$Temp), model="Sph",
+ range=sqrt(areaSpatialGrid(study.grid))/4, nugget=0)

```

Se ajusta el variograma con el modelo REML

```

> fit_reml<-fit.variogram.reml(Temp~1, data, model=mod)
> plot(variogram(Temp~1, data), fit_reml, main="REML Model")

```

Ahora se ajusta el variograma al modelo cuadrado minimo ordinario

```

> fit_ols<-fit.variogram(variogram(Temp~1, data), model=mod,
+ fit.method=6)
> plot(variogram(Temp~1, data), fit_ols, main="OLS Model")

```

Por último se realiza el kriging y se construye el mapa

```

> map<-krige(Temp~1, data, model=fit_ols, newdata=study.grid)
> spplot(map, "var1.pred", col.regions=terrain.colors(50))

```

1.6. Para tener en cuenta

Los primeros pasos de este manual estan basados en el libro [1], capítulo 8. Para continuar con los ejemplos se debe revisar la documentación del libro, además se puede descargar el código fuente en el repositorio del libro en <http://www.asdar-book.org/data2ed.php?chapter=7>.

Bibliografía

- [1] R. Bivand, E. Pebesma, and V. Gómez-Rubio, *Applied Spatial Data Analysis with R*, ser. Use R! Springer, 2013.