

Uso de R para pronóstico de cosecha

Contenido

Introducción
Geoestadística

Análisis descriptivo

Semivariograma

Pronóstico



El **objetivo** de esta presentación es mostrar el uso del paquete `geoR` para realizar un estudio geoestadístico.

Dicho estudio consiste en modelar la variabilidad espacial y realizar interpolación entre los puntos observados de manera tal de poder realizar un mapa.



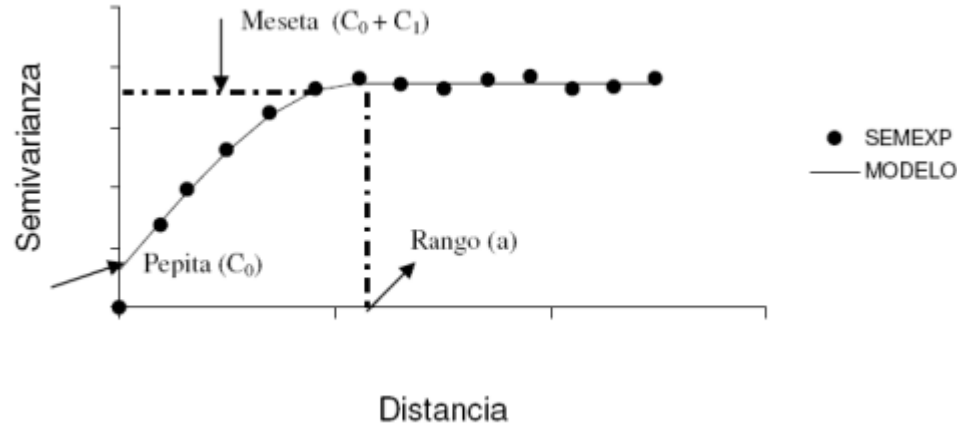
Variable regionalizada

Se denomina variable regionalizada a una variable que puede ser expresada como una función espacial numérica que varía de un lugar a otro con una continuidad aparente y cuya variación se puede representar con una función denominada semivariograma y cartografiar por medio del Krigado.

$$Z(x_i) = \mu + \varepsilon_i$$

Variación espacial (Variograma)

$$V(y) = \frac{1}{2} E[(Y_i - Y_j)^2]$$



La intercepción se denomina C_0 , la meseta ($C_0 + C_1$), corresponde a la varianza del proceso de observación Y . Luego la forma en que un variograma se incrementa desde la intercepción hasta la meseta lo determina la función de correlación $\rho(u)$. El rango en términos prácticos corresponde a la distancia a partir de la cual dos observaciones son independientes. El rango se interpreta como la zona de influencia.



Análisis de datos georeferenciados

Con datos *georeferenciados* y asumiendo cierto grado de correlación espacial se puede encontrar el modelo de interpolación de los valores intermedios faltantes de manera tal de generar una grilla de puntos para hacer un mapa, por ejemplo de producción.

La base de datos contiene información de 1 lote con datos de cosecha (kgxárbol) de naranja Valencia late en el año 2011.

Los pasos a seguir para un análisis geoestadístico son:

1. Lectura de datos
2. Preparación del archivo geoestadístico.
3. Análisis descriptivo de los datos.
4. Cálculo empírico de la función que cuantifica la variación espacial (Variograma).
5. Estimación de parámetros.
6. Ajuste de modelo de variación espacial.
7. Interpolación.
8. Validación cruzada



Lectura de datos

En primer lugar voy a trabajar con el **RStudio** que es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R

Lectura los datos de un archivo Excel, en el cuadrante superior derecho
Import Dataset → From Excel → Browse..(buscar el archivo) → Import

Sería equivalente a escribir la siguiente sentencia en R

```
>Naranja_2011 <- read.table("D:/Curso _ Charla/R ladies/Naranja2011.csv",  
header=TRUE, sep=",", dec=",")
```

Luego se debe cargar el paquete geoR

Con R Studio simplemente se debe tildar la opción, buscar en el cuadrante inferior derecho Package geoR

Sería equivalente a escribir la siguiente sentencia en R

```
require(geoR) o library(geoR)
```




Construcción del archivo geoestadístico

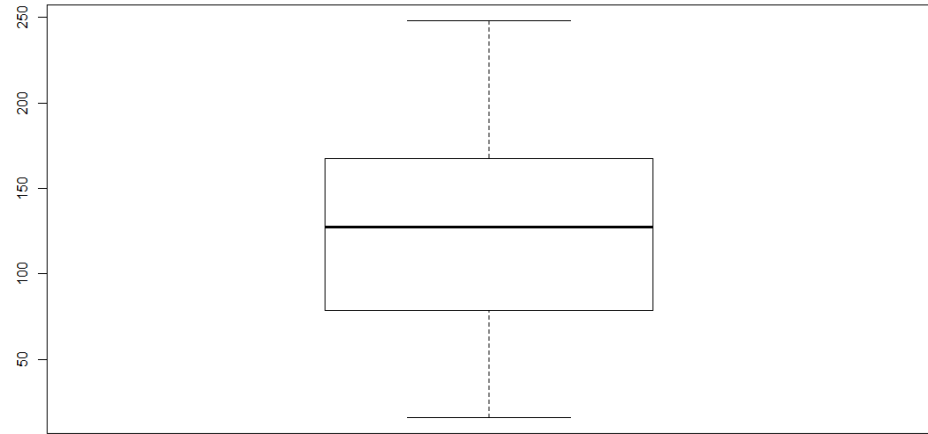
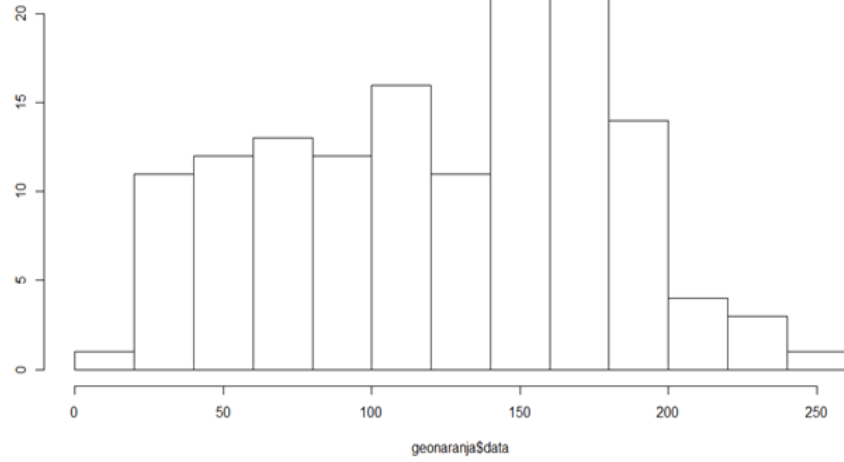
```
## Armando un archivo formato "geodata" (para uso paquete geoR)  
geonaranja <- as.geodata(Naranja_2011, coords=1:2, data.col=7)
```

Descripción de los datos

```
# Algunos estadísticos resumen  
summary(geonaranja$data)  
hist(geonaranja$data)  
boxplot(geonaranja$data)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.30	78.61	127.38	123.42	167.69	247.90

Histogram of geonaranja\$data





```
## Inspeccionando los elementos da lista.  
# Valores mínimos y máximos de las coordenadas  
>summary(geonaranja)
```

```
Number of data points: 143 Coordinates  
summary      Lat Lon      min  1  1 max 130 115 Distance summary  
min          max    2.236068 168.240899 Data summary  
Min.  1st Qu.  Median      Mean  3rd Qu.      Max.  
16.3000  78.6100 127.3800 123.4224 167.6900 247.9000
```

Tambien se obtiene lo mismo con las siguientes sentencias

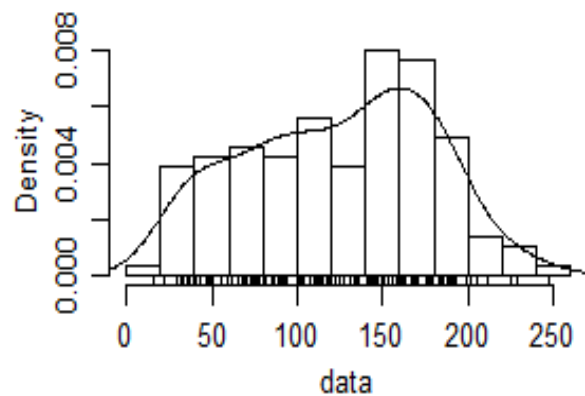
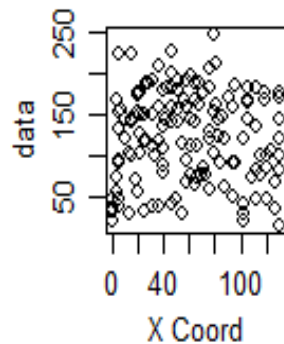
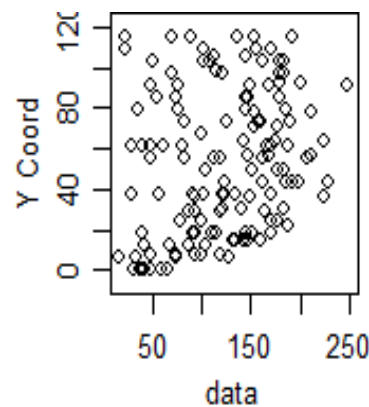
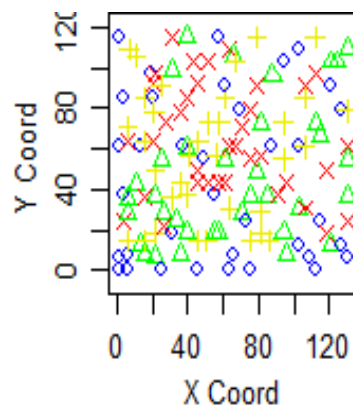
```
>apply(geonaranja$coords,2,range)  
>range(dist(geonaranja$coords))
```

```
Lat Lon[1,] 1
1[2,] 130 115
```

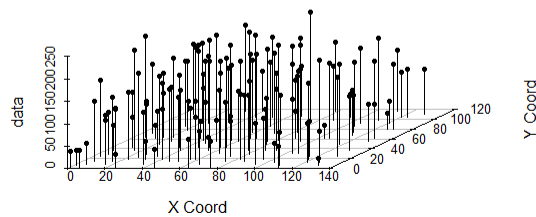
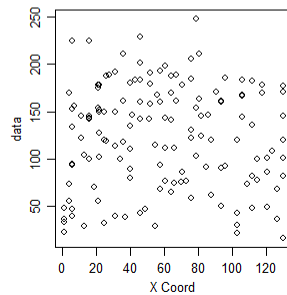
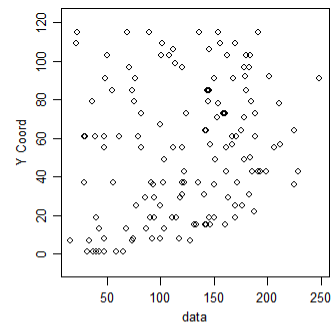
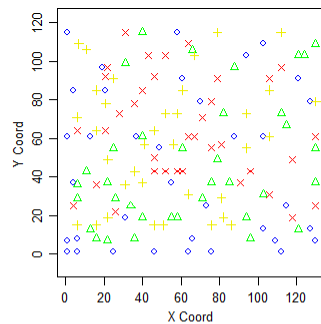
```
➤ range(dist(geona
  coords))
```

```
➤ [1] 2.236068
    168.240899
```

```
plot(geonaranja)
```

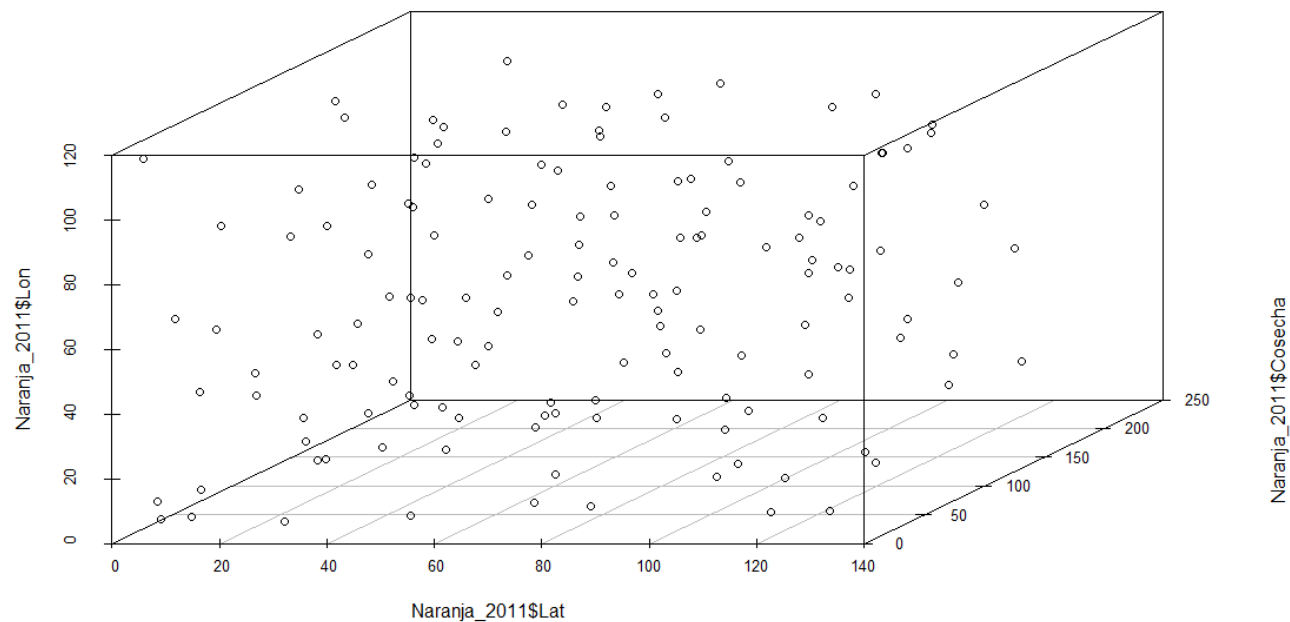


```
library(scatterplot3d)  
plot(geonaranja,scatter3d=TRUE)
```

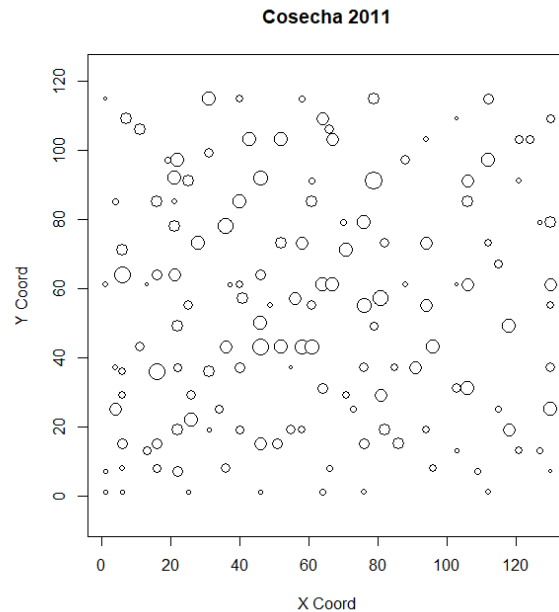


```
scatterplot3d(Naranja_2011$Lat,  
Naranja_2011$Cosecha,Naranja_2011$Lon,  
grid=TRUE)
```

```
axis=TRUE,
```



```
points.geodata(geonaranja, data=geonaranja$data, data.col=2,  
pt.divide="data.proportional",main="Cosecha 2011",  
cex.max=2.5)
```

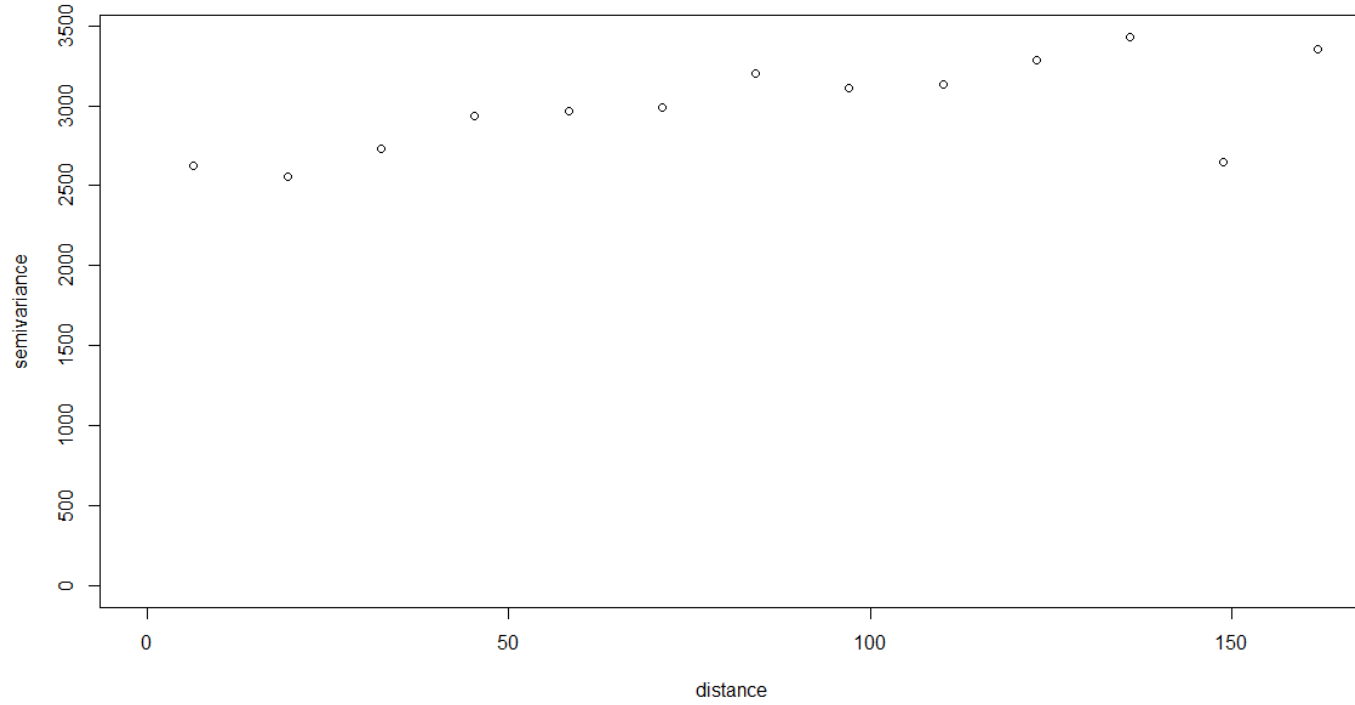


Variograma empírico

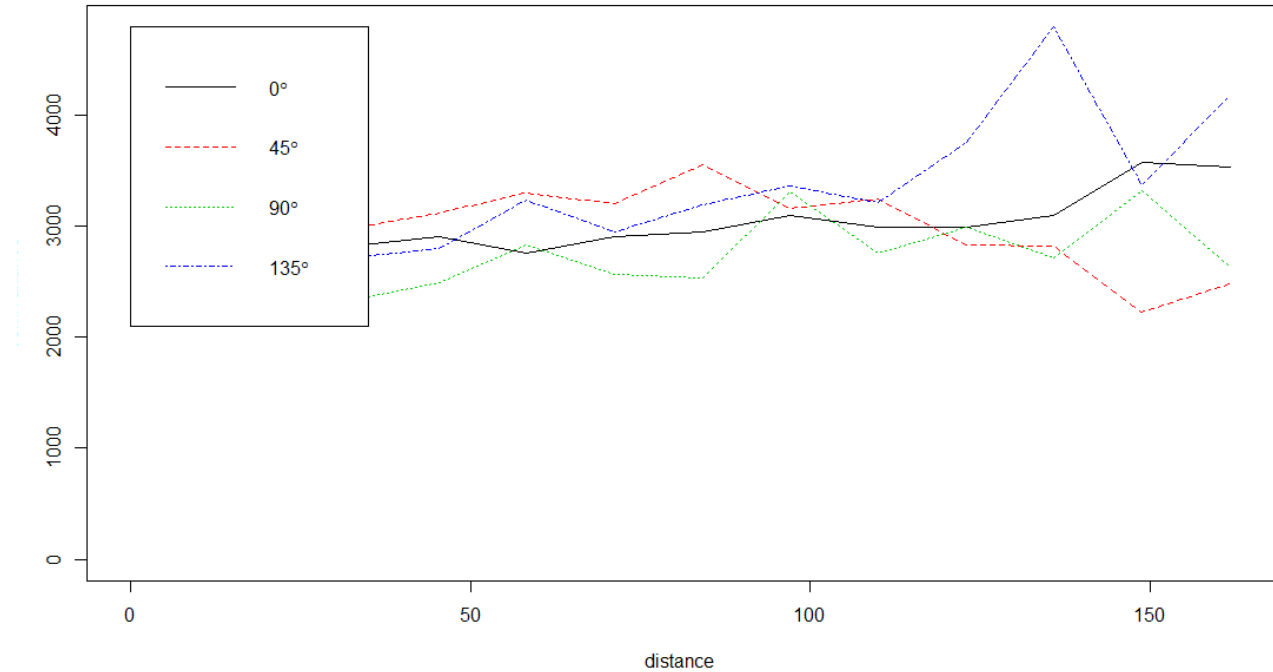
```
>varionaran <- variog(geonaranja, coords = geonaranja$coords, data  
= geonaranja$data, trend = "cte", lambda = 1, estimator.type =  
c("classical", "modulus"),      pairs.min = 2,      direction =  
"omnidirectional")  
>varionaran  
>plot(varionaran)
```

```
$u [1] 6.470804 19.412411 32.354019 45.295627 58.237234 71.178842 84.120449  
[8] 97.062057 110.003665 122.945272 135.886880 148.828487 161.770095
```

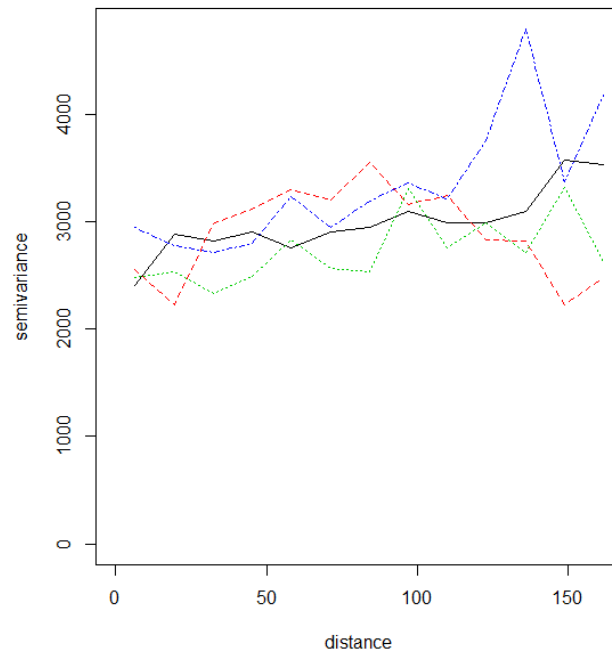
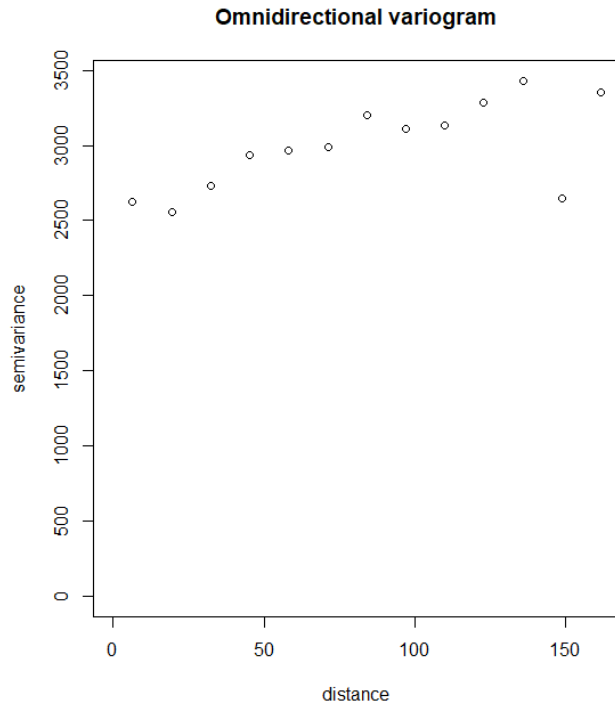
```
$v [1] 2622.149 2554.830 2728.269 2934.127 2969.178 2990.106 3203.589 3107.665  
[9] 3136.111 3283.076 3427.565 2645.481 3356.379
```

```
# Variación en distintas direcciones  
vario.4 <- variog4(geonaranja)  
plot(vario.4)
```



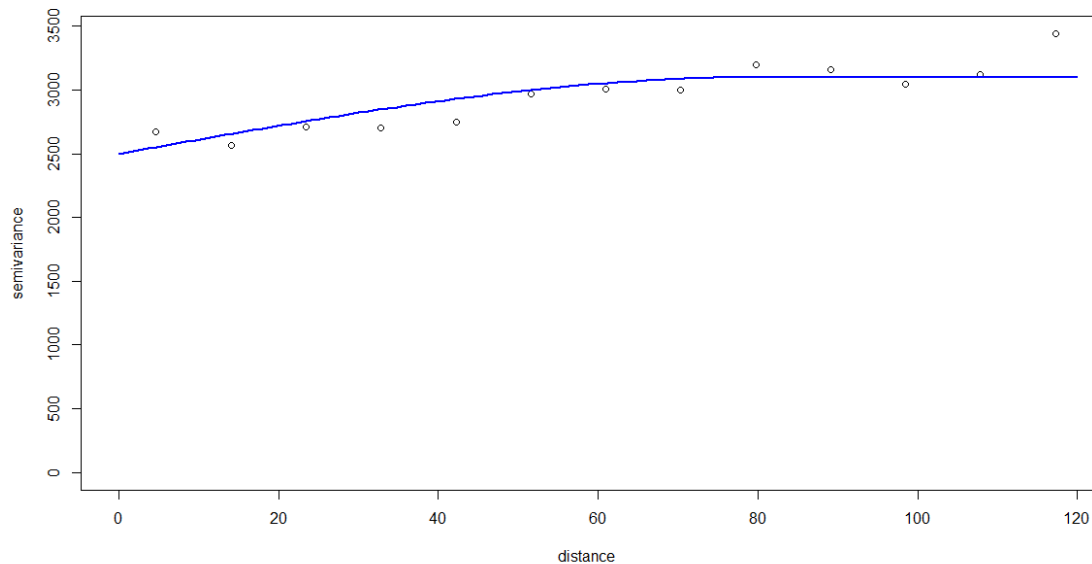
```
par(mfrow=c(1,2))
plot(varionaran, main = "Omnidirectional variogram")
plot(vario.4, type="l")
```



Ajuste de un modelo al variograma



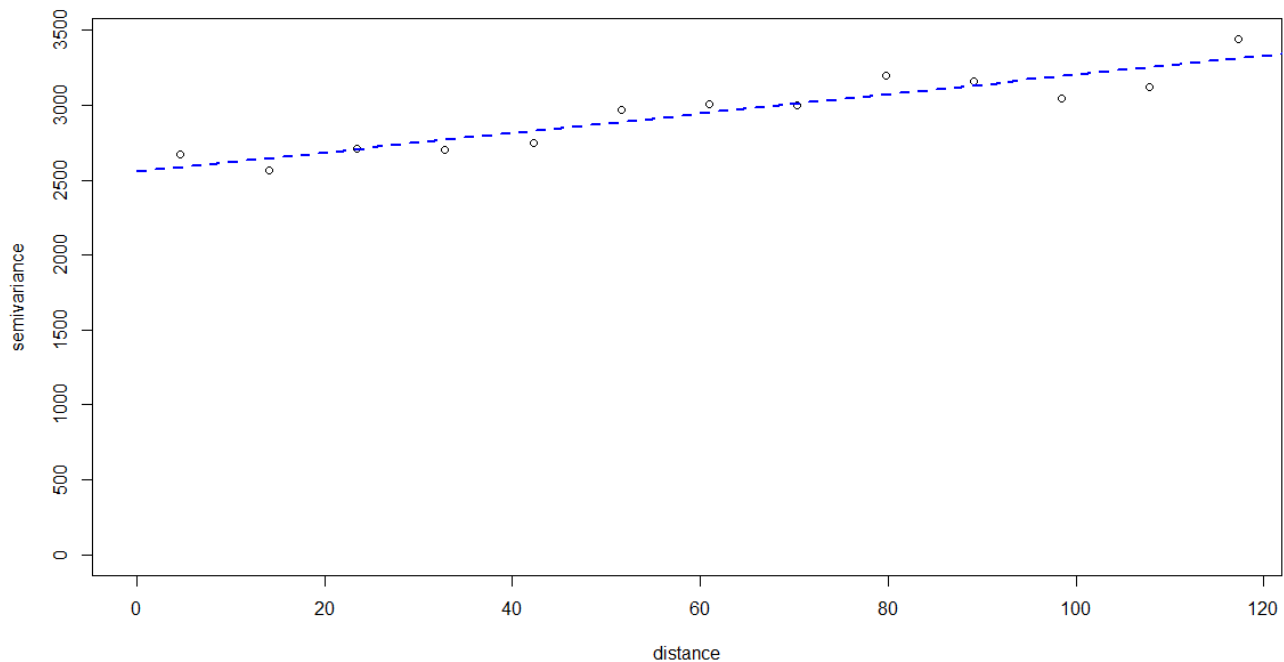
```
>ajuste<-variofit(varionaran, cov.model="sph", ini=c(600,80), nug=2500, max.dist=168)  
>ajuste  
  
>plot(varionaran)  
>lines(ajuste, lty=1, col="blue",lwd=1)
```



```
naran.ols <- variofit(naran.vario, cov.model = "linear",  
ini=c(100,150), wei="equal", nug=2000, wei="equal", min="optim")  
naran.ols
```



```
plot(varionaran)  
lines(naran.ols, lty=1, col="blue",lwd=1)
```



- `loci <- matrix(c(130,130,130,50,80,26), ncol=2)`
- `> loci`
 `[,1] [,2] [,1] 130 50 [,2] 130 80 [,3] 130 26`
- `points(geonaranja)> text(loci, lab=1:3, col="red")>`
- `naran.kc <- krige.conv(geonaranja, loc=loci, krige=krige.control(obj=ajuste))`
- `krige.conv`: model with constant mean
`krige.conv`: Kriging performed using global neighbourhood
- `naran.kc[1:2]`
- `$predict[1]` 116.8479 108.5893 110.0497
- `$krige.var[1]` 2736.820 2719.412 2724.666

Bibliografía

- HENAO, R.G.(2002) *Introducción a la Geoestadística* - Teoría y Aplicación. Universidad Nacional de Colombia Sede Bogotá Facultad de Ciencias Departamento de Estadística.
- ARMSTRONG, M.(1998) Basic Linear Geostatistics. Springer –Verlag. Berlin 153 pp.
- CLARK, I. (1979) Practical Geostatistics. Applied Science Pub. London 169 pp.
- CRESSIE, N, (1991) Statistics for Spatial Data. John Wiley & Sons. New York. 900 pp.
- JOURNEL, A.G. & HUIJBREGTS, C.H (1978) Mining Geostatistics Academic Press London 600 pp.
- MATHERON, G. (1963) Principles of Geostatistics. Econ. Geology 58: 1246-1266
- WEBSTER, R. & OLIVER, M.A. (2000) Geostatistics for Environmental Sciences. John Wiley & Sons. 271 pp.
- SCHABENBERGER, O.; GOTWAY, C.A.(2005) Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC, Taylor & Francis Group. 488 p.



Gracias