



MÁSTER EN MATEMÁTICA COMPUTACIONAL

TRABAJO FINAL DE MÁSTER

Una aplicación de interpolación Kriging para
el estudio de la evolución de la pandemia
Covid-19 en España y en la Comunidad
Valenciana

Autora:
Lucía MASERO BRAVO

Tutores académicos:
Vicente MARTINEZ GARCÍA
Pablo GREGORI HUERTA

Fecha de lectura: Septiembre de 2020
Curso académico 2019/2020

Dedicatorias

En primer lugar, quiero agradecer a mis tutores Vicente y Pablo por el apoyo, dedicación y conocimientos aportados. Gracias por ser siempre tan cercanos y hacer que me gusten todavía más las matemáticas.

Gracias a mis padres por el respaldo prestado tanto en el ámbito educativo como personal. Gracias por estar siempre a mi lado.

Resumen

La geoestadística es una área de la estadística que se centra en estudiar el comportamiento de un conjunto de datos sobre la superficie terrestre. El documento que se presenta a continuación detalla una investigación realizada sobre los datos del COVID-19, ésta se realiza tanto a nivel nacional como de Comunidad Valenciana.

El objetivo del presente trabajo se basa en estudiar distintos métodos de Kriging tanto a nivel teórico como práctico y su aplicación al caso concreto de representar geográficamente la evolución de la pandemia COVID-19. Por la parte teórica se verán los siguientes tipos de Kriging: Simple, Ordinario, con modelo de tendencia o universal, por bloques, factorial, Cokriging y Krigings no lineales (lognormal o logarítmico, multi-Gaussiano, de rango, indicatriz y disyuntivo). Por otra parte, se verá de forma práctica el Kriging Simple con variogramas implementados mediante 3 métodos distintos, los cuales son esférico, Gaussiano y exponencial. Además se harán predicciones de los casos de contagiados de COVID-19 en zonas donde no se dispone de datos.

La principal conclusión que se ha obtenido es que el Kriging es un método muy potente de predicción que se puede aplicar a muchos campos de estudio. Además, se han obtenido mejores resultados en las predicciones realizadas mediante el Kriging Simple con modelo Esférico que con los modelos Exponencial y Gaussiano.

Palabras clave

Krigeado, análisis geoestadístico, método de interpolación

Keywords

Kriging, geoestadistical analysis, interpolation technique

Índice general

I	Introducción y conceptos previos	7
1.	Introducción	9
2.	Conceptos previos	11
2.1.	Campo aleatorio y variograma	11
2.2.	Características de un modelo geoestadístico	13
2.2.1.	Isotropía y anisotropía	13
2.2.2.	Condiciones de positividad	16
2.2.3.	Estimación de variogramas	16
2.2.4.	Estimación de los parámetros del variograma	19
II	Teoría del Kriging	23
3.	Kriging	25
3.1.	Kriging Simple	26
3.2.	Kriging Ordinario	27
3.3.	Kriging con modelo de tendencia o universal	28
3.3.1.	Kriging con factor externo	29
3.4.	Kriging en Bloques	30
3.5.	Kriging Factorial	31
3.6.	Cokriging	31
3.7.	Krigings no lineales	32
3.8.	Validación cruzada	32
III	Aplicación: Análisis geoestadístico espacial sobre los casos de COVID-19 en la España peninsular y en la Comunidad Valenciana	33
4.	Aplicación de la interpolación Kriging a la expansión geográfica de la pandemia COVID-19	35
4.1.	Introducción	35

4.2.	Descripción de los datos utilizados	35
4.2.1.	Zona peninsular española	35
4.2.2.	Zona de la Comunidad Valenciana	42
4.2.3.	Los datos en su ubicación espacial	46
5.	Resultados	47
5.1.	Variogramas	48
5.2.	Predicciones	52
6.	Conclusiones	59
A.	Aplicación del Kriging Simple a los casos de Covid-19 en España	65
A.1.	Kriging para los datos del 3 de junio de 2020	65
A.1.1.	Los datos en su ubicación espacial	67
A.1.2.	Variograma	68
A.1.3.	Interpolación mediante Kriging simple	69
A.1.4.	Probamos distintos tipos de modelos	71
A.2.	Kriging para los datos del 18 de abril de 2020	76
A.2.1.	Los datos en su ubicación espacial	78
A.2.2.	Variograma	79
A.2.3.	Vamos a interpolar mediante Kriging Simple	81
A.2.4.	Probamos distintos tipos de modelos	82
B.	Aplicación del Kriging Simple a los casos de Covid-19 en la Comunidad Valenciana	89
B.1.	Kriging para los datos del 3 de junio de 2020	90
B.1.1.	Los datos en su ubicación espacial	91
B.1.2.	Variograma	93
B.1.3.	Interpolación mediante Kriging Simple	94
B.1.4.	Probamos distintos tipos de modelos	96

Parte I

Introducción y conceptos previos

Capítulo 1

Introducción

Como un gran físico dijo una vez “Saber dónde encontrar la información y cómo usarla. Ese es el secreto del éxito”, Albert Einstein. Por tanto, no es difícil pensar en la gran cantidad de hechos y sucesos que se cuantifican, evalúan y tratan a partir de grandes cantidades de información extraídas a priori o in situ: la duración de un vuelo comercial, el mejor tenista de la década o la evolución de un gran incendio. Incluso se pueden llegar a recabar datos de superficies tan grandes como la evolución de la capa de ozono o la dispersión de un nuevo patógeno.

Un tema actual y que tenemos todos en mente es el reciente caso de COVID-19 (Orthocoronavirinae). Resulta evidente que conocer sus cifras de contagios es clave para prever su expansión y poder actuar en consecuencia. Grandes ciudades ofrecieron esta información con notoria celeridad, pero, ¿cómo tratar aquellas zonas alejadas si no podemos hallar información? ¿Las ignoramos y concluimos que no poseen valor estratégico alguno? Evidentemente no.

Todo enclave, grande o pequeño, puede albergar información necesaria para el suceso, e ignorarla supondría una negligencia. Es, pues, imperativo tratar de obtener mediciones cercanas a la realidad en la medida de lo posible, hecho que podremos lograr mediante el Kriging.

El Kriging es un método ideado por el ingeniero de minas Daniel Gerhardus Krige en 1951 y posteriormente formulado por el matemático francés e Ingeniero Civil de minas Georges François Paul Marie Matheron en 1963. Se trata de un método de interpolación geoestadística la cual consiste en estimar valores desconocidos utilizando variogramas y una serie de valores conocidos. El enfoque se basa en el uso de un campo aleatorio y una colección de supuestos, como la estacionariedad y la ergodicidad espacial, para reducir la información necesaria a un llamado variograma que puede estimarse a partir de las mediciones disponibles [1] .

A lo largo del tiempo, este método ha sido utilizado en distintas áreas como pueden ser la minería (para determinar la concentración del mineral [2] o el grado de contaminación que se tiene en una área minera [3]), la astronomía [4] o la meteorología [5].

En la actualidad, el Kriging se ha empleado para muchos estudios de la distribución espacial epidemiológica como pueden ser el virus de la gripe Aviar (H7N9) [6] o el virus Ébola (Ebolavirus) [7].

En el presente documento se detalla un estudio realizado sobre la distribución del COVID-19 en toda la superficie peninsular de España. Para ello, se ha utilizado el método de Kriging Simple con modelo Esférico, Exponencial y Gaussiano.

Este documento está dividido en los diferentes capítulos que se detallan a continuación:

- Capítulo 1: Introducción y contexto del proyecto.
- Capítulo 2: detalla todos los conceptos previos necesarios para una adecuada comprensión del método Kriging.
- Capítulo 3: detalla la teoría en la que se basan los distintos tipos de Kriging.
- Capítulo 4: detalla un caso práctico de Kriging Simple sobre un tema de actualidad (contagios de COVID-19,) tanto a nivel nacional como de comunidad.
- Capítulos 5 y 6: presentan los resultados y conclusiones de la aplicación práctica realizada en el capítulo 4.
- Anexo A: se presenta el código empleado para realizar las predicciones de España.
- Anexo B: se presenta el código empleado para realizar las predicciones de la Comunidad Valenciana.

Capítulo 2

Conceptos previos

A lo largo de este capítulo se introducirán todos los términos necesarios para una mejor comprensión del presente trabajo, además de conocer más de cerca algunos conceptos propios del campo de la geoestadística.

Primero de todo tenemos que tener claro que a la hora de modelizar un problema estadístico es imprescindible determinar el variograma o covarianza experimental, analizar los variogramas que se puedan aplicar a nuestros datos. Además, nos puede ayudar conocer información adicional de nuestros datos como las características físicas del área de estudio o realizar un madograma. Todos los conceptos necesarios para realizar una correcta modelización se explicarán a continuación.

2.1. Campo aleatorio y variograma

Nuestro objeto de estudio es un proceso estocástico $Z(u), u \in D$ donde D es un subconjunto continuo de R^d (espacio euclídeo con d dimensiones). Estos procesos se llaman ‘Campo aleatorio’ o ‘procesos geoestadísticos’.

El valor medio de Z en la localización u se define como:

$$\mu(u) = E\{Z(u)\}, u \in D, \quad (2.1)$$

Asumiremos que la varianza de $Z(u)$ existe para todo $u \in D$.

Según la naturaleza de $Z(u)$ nos encontraremos ante un tipo de proceso u otro:

- Z es **Gaussiano** si para cualquier natural $k > 0$, u_1, u_2, \dots, u_k , el vector $(Z(u_1), Z(u_2), \dots, Z(u_k))$ sigue una distribución normal multivariante.

- Z es **estrictamente estacionario** si para cualquier natural $k > 0$, u_1, u_2, \dots, u_k , la distribución $(Z(u_1), Z(u_2), \dots, Z(u_k))$ es la misma que la del vector $(Z(u_1 + h), Z(u_2 + h), \dots, Z(u_k + h))$ para cualquier $h \in R^d$ siempre que $u_1, \dots, u_k, u_1 + h, \dots, u_k + h$ se encuentren dentro del dominio de D .
- Z es **estacionario de segundo orden** si para cualquier natural $k > 0$ y u_1, u_2, \dots, u_k , si $\mu(u_1) = \mu(u_2) = \dots = \mu(u_{k-1}) = \mu(u_k)$ y $cov(Z(u_i), Z(u_j))$ depende de $u_i - u_j$ para todos los $u_i, u_j \in D$.

Si todas las varianzas son finitas, un proceso estrictamente estacionario también será estacionario de segundo orden [8].

Un variograma es una herramienta eficaz para describir el comportamiento de procesos aleatorios espaciales tanto estacionarios como no estacionarios. Permite medir la disimilitud de los pares de puntos de la muestra de datos, los cuales están separados por un ‘salto’. Estos ‘saltos’ se analizan por sus diferencias cuadráticas medias. El grado de disimilitud entre las variables $Z(u_i)$ y $Z(u_j)$ se define con la función variograma:

$$2\gamma(h) = \text{Var}\{Z(u_i) - Z(u_j)\} \quad (2.2)$$

donde $h = u_i - u_j$ y la función $\gamma(h)$ se llama semivariograma [9].

Cuanto menor diferencias haya entre un par de observaciones, menor será la varianza entre ambas, y por tanto menor será el valor del semivariograma ($\gamma(h)$) [10].

Además, el máximo valor que puede alcanzar el semivariograma se denomina *sill*. Este valor es la varianza de la variable observada Z .

El *sill* también determina el rango del sistema, definiendo el límite de dependencia espacial. Los datos de la muestra que se encuentran separados por distancias inferiores al rango se encuentran correlacionadas espacialmente, en cambio las ubicaciones más alejadas que el rango no lo están.

El efecto nugget define la aleatoriedad a corto plazo o ruido en la variable regionalizada. El valor del efecto nugget es la intersección de la función variograma a una distancia de retraso de casi cero, dado que, a una distancia de retraso casi cero el variograma tiene valor cero. Este efecto mide la variabilidad entre ubicaciones muy próximas [11] [12].

2.2. Características de un modelo geoestadístico

2.2.1. Isotropía y anisotropía

Isotropía

Supongamos que tenemos un proceso intrínsecamente estacionario con semivariograma $\gamma(h)$ con $h \in R^d$. Diremos que el proceso es isotrópico si $\gamma(h) = \gamma_0(\|h\|)$. Es decir, si el semivariograma $\gamma(h)$ depende de h sólo a través de su longitud $\|h\|$.

Un proceso se dirá que es homogéneo si es isotrópico e intrínsecamente estacionario.

A continuación se expondrán algunos ejemplos de semivariogramas isotrópicos:

- El **modelo del efecto Nugget** es la estructura más sencilla de semivariograma. Éste no tiene estructura espacial.

$$\gamma(h) = \begin{cases} 0 & \text{si } \|h\| = 0 \\ 1 & \text{en otro caso} \end{cases}$$

- El **modelo Lineal**, es no acotado en función de dos constantes positivas. Este modelo no es estacionario dado que $\gamma(\|h\|)$ tiende a infinito cuando las distancias son grandes.

$$\gamma(h) = \begin{cases} 0 & \text{si } \|h\| = 0 \\ c_0 + c_1\|h\| & \text{si } \|h\| > 0 \end{cases}$$

- El **modelo Esférico** viene definido por:

$$\gamma(h) = \begin{cases} c_0 + c_1 \cdot \left[1,5 \frac{\|h\|}{a} - 0,5 \left(\frac{\|h\|}{a} \right)^3 \right] & \text{si } \|h\| \leq a \\ c_0 + c_1 & \text{si } \|h\| \geq a \end{cases},$$

siendo a el rango actual, c_1 una varianza a priori o *sill* y c_0 el efecto nugget.

- El **modelo Exponencial** se define como:

$$\gamma(h) = c_0 + c_1 \cdot \left[1 - \exp \left(- \frac{3\|h\|}{a} \right) \right],$$

siendo a el rango efectivo, c_1 una varianza a priori o *sill* y c_0 el efecto nugget.

- El **modelo Gaussiano** viene dado por la siguiente expresión:

$$\gamma(h) = c_0 + c_1 \cdot \left[1 - \exp\left(-\frac{(3\|h\|)^2}{a^2}\right) \right],$$

siendo a el rango efectivo, c_1 una varianza a priori o *sill* y c_0 el efecto nugget.

- El **modelo Potencial** viene definido por:

$$\gamma(h) = c_0 + c_1 \cdot \|h\|^\omega,$$

siendo ω un valor en $]0, 2[$, c_1 una pendiente positiva y c_0 el efecto nugget.

- El **modelo del efecto Hole (o sinusoidal)** se utiliza a la hora de definir componentes cíclicas subyacentes.

$$\gamma(h) = c_0 + c_1 \cdot \left[1, 0 - \cos\left(\frac{\|h\|}{a} \cdot \pi\right) \right],$$

siendo c_1 una varianza a priori y c_0 el efecto nugget. Tenemos que considerar que este modelo sólo se puede aplicar a una dirección para que el variograma esté definido correctamente.

- La **familia Mátern**. Para definir esta clase lo haremos mediante la función de covarianzas:

$$C_0(h) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)} \left(\frac{2\sqrt{\theta_2}\|h\|}{\theta_1}\right)^{\theta_2} k_{\theta_2}\left(\frac{2\sqrt{\theta_2}\|h\|}{\theta_1}\right),$$

siendo θ_1 el parámetro que define la dependencia espacial y θ_2 el parámetro que define la forma ambos positivos.

Además, la función $\Gamma(\cdot)$ representa la función Gamma y k_{θ_2} es una modificación de la función de Bessel de tercera clase con orden θ_2 .

Un aspecto a destacar de la familia Mátern es que cuando θ_2 tiende a 1 entonces obtendremos la forma Exponencial. Por otra parte, si se realiza el límite $\theta_2 \rightarrow \infty$ se obtendrá la forma Gaussiana.

Anisotropía geométrica

Los procesos anisotrópicos son una generalización de los procesos isotrópicos. Hay distintos tipos de anisotropía, la geométrica es el más simple y se define de la siguiente manera:

$$\gamma(h) = \gamma_0(\|Ah\|) \quad (2.3)$$

siendo γ_0 un semivariograma isotrópico y A una matriz $d \times d$ que representa una transformación lineal de \mathbb{R}^d .

Anisotropía zonal

Sean Z_1, Z_2, \dots, Z_p , p procesos independientes e intrínsecamente estacionarios, entonces:

$$Z = Z_1 + Z_2 + \dots + Z_p$$

es intrínsecamente estacionario dado el semivariograma que se define a continuación:

$$\gamma(h) = \gamma_1(h) + \dots + \gamma_p(h)$$

Siendo $\gamma_1, \dots, \gamma_p$ los semivariogramas de Z_1, \dots, Z_p respectivamente. Entonces,

$$\gamma(h) = \sum_{i=1}^p \gamma_0(A_i h) \quad (2.4)$$

con γ_0 un semivariograma isotrópico y A_i con $1 \leq i \leq p$ matrices.

2.2.2. Condiciones de positividad

Condición de positividad

En geoestadística no se puede definir la covarianza espacial o el semivariograma de forma arbitraria. Es necesario que la covarianza espacial cumpla la condición de positividad, para ello, se debe cumplir la siguiente condición:

$$\sum_{i=1}^p \sum_{j=1}^p a_i a_j C(u_i, u_j) \geq 0 \quad (2.5)$$

siendo $p \in \mathbb{N}$, u_1, \dots, u_p localizaciones arbitrarias, $C(u_i, u_j) = \text{cov}\{Z(u_i), Z(u_j)\}$ con $1 \leq i, j \leq p$ y a_1, \dots, a_p coeficientes reales arbitrarios.

Condición de no positividad condicional

Siendo $\gamma(\cdot)$ el semivariograma de un proceso estacionario de segundo orden y siendo a_1, \dots, a_n constantes con $\sum_{i=1}^p a_i = 0$ entonces se cumple que:

$$\sum_{i=1}^p \sum_{j=1}^p a_i a_j \gamma(u_i - u_j) \leq 0 \quad (2.6)$$

La no positividad condicional es una condición necesaria para que $\gamma(\cdot)$ sea un semivariograma intrínsecamente estacionario [13].

2.2.3. Estimación de variogramas

Sea un proceso observado $\{Z(u_i) : u_i \in D\}$ siendo $\{u_i\}$ un número finito de localizaciones. A continuación veremos distintos estimadores de variogramas y medidas de variabilidad espacial.

Método de los momentos

Este método se utiliza cuando las localizaciones de los puntos de muestreo se encuentran definidos dentro de una malla rectangular y se define como:

$$2\hat{\gamma}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} \{Z(u_i) - Z(u_j)\}^2, \quad (2.7)$$

donde $N(h)$ denota todos aquellos pares (u_i, u_j) donde $h = u_i - u_j$ y $\#N(h)$ el cardinal de $N(h)$.

Si las localizaciones de los puntos no se encuentran definidas dentro de una malla rectangular se podrá aplicar el variograma definido anteriormente pero definiendo $N(h)$ de la siguiente forma:

$$N(h) = \{(u_i, u_j) : u_i - u_j \in T(h)\}$$

Siendo $T(h)$ alguna vecindad o región de tolerancia sobre h .

Variograma cruzado

El variograma cruzado es una variación del método de los momentos, éste es de gran utilidad cuando se dispone de dos atributos Z e Y , medidos sobre las mismas localizaciones. En esta situación el variograma se define de la siguiente manera:

$$2\hat{\gamma}_{ZY}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} (Z(u_i) - Z(u_j))(Y(u_i) - Y(u_j)) \quad (2.8)$$

Covarianza muestral

La covarianza muestral es una variación del método de los momentos y se define como:

$$\widehat{C}(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} \{Z(u_i)Z(u_j)\} - m^2 \quad (2.9)$$

siendo $m = \frac{1}{\#N(h)} \sum_i Z(u_i)$ y $\#N(h)$ el número total de puntos.

Si $\widehat{\sigma}^2 = \frac{1}{\#N(h)} \sum_i Z(\mathbf{u}_i)^2 - m^2$ nos permitirá estandarizar la covarianza para definir el correlograma de la siguiente manera:

$$\widehat{\rho}(h) = \frac{\widehat{C}(h)}{\widehat{\sigma}^2}$$

Madograma

El madograma es un variograma muy similar al método de los momentos, éste viene definido por la siguiente expresión:

$$2\widehat{\gamma}_M(h) = \frac{1}{\#N(h)} \sum_{(u_i, u_j) \in N(h)} |Z(u_i) - Z(u_j)| \quad (2.10)$$

A la hora de estimar el variograma debemos de tener en cuenta dos normas básicas [14]:

- El variograma empírico solamente será posible definirlo si el número de pares es superior a 30.
- Sea D la distancia máxima de las localizaciones muestreadas, entonces la distancia de la fiabilidad para un variograma experimental será $h < \frac{D}{2}$.

2.2.4. Estimación de los parámetros del variograma

Algunos variogramas carecen de la propiedad de no positividad. Por ello se reemplaza el semivariograma empírico por algún modelo paramétrico. A continuación se expondrán tres métodos para estimar los diversos parámetros que componen el variograma.

Estimación por mínimos cuadrados

Supongamos que h es un conjunto finito de valores y estimamos el semivariograma $\gamma(h)$, ahora queremos ajustar el modelo mediante una función paramétrica $\gamma(h; \theta)$ siendo θ un número finito de parámetros. Usualmente θ es un vector compuesto por el rango, *sill* y el efecto nugget, entre otros parámetros. Supongamos que se ha utilizado el método de los momentos para definir el semivariograma $\hat{\gamma}$ siendo $\hat{\gamma}$ un vector que contienen los valores estimados y $\gamma(\theta)$ el vector que contiene los valores derivados por el modelo sobre h .

A continuación se exponen tres métodos de estimación por mínimos cuadrados no lineales, en todas ellas se considerará el parámetro θ como el valor que minimiza. Estos métodos están ordenados de manera creciente según la eficiencia que tienen y de manera decreciente según su simplicidad.

- Mínimos cuadrados ordinarios:

$$\{\hat{\gamma} - \gamma(\theta)\}'\{\hat{\gamma} - \gamma(\theta)\}$$

- Mínimos cuadrados ponderados:

$$\{\hat{\gamma} - \gamma(\theta)\}'W(\theta)^{-1}\{\hat{\gamma} - \gamma(\theta)\}$$

Siendo $W(\theta)$ una matriz diagonal siendo los elementos de la diagonal las varianzas de $\hat{\gamma}$.

- Mínimos cuadrados generalizados:

$$\{\hat{\gamma} - \gamma(\theta)\}'V(\theta)^{-1}\{\hat{\gamma} - \gamma(\theta)\}$$

Siendo $V(\theta)$ la matriz de covarianzas de $\hat{\gamma}$.

Estimación máximo verosímil

Si utilizamos un proceso Gaussiano para el muestreo, entonces podemos calcular la verosimilitud y maximizarla fácilmente [15] [16].

Consideramos $Z \sim N(X\beta, \Sigma)$ como el proceso espacial con el vector de observaciones Z de n dimensiones, X la matriz (de rango completo) de covariables con dimensión $n \times q$ con $q < n$ y β un vector de dimensión q de parámetros desconocidos y Σ la matriz de covarianzas de la muestra, en la práctica $\Sigma = \alpha V(\theta)$ con α un parámetro de escala desconocido y $V(\theta)$ la matriz de covarianzas estandarizadas.

Entonces, la función de densidad será de la forma:

$$(2\pi)^{-n/2} \det(\Sigma)^{-1/2} \exp \left\{ -\frac{1}{2} (Z - X\beta)' \Sigma^{-1} (Z - X\beta) \right\} \quad (2.11)$$

Siendo la log-verosimilitud:

$$\ell(\beta, \alpha, \theta) = \frac{n}{2} \log(2\pi) + \frac{n}{2} \log \alpha + \frac{1}{2} \log \det(V(\theta)) + \frac{1}{2\alpha} (Z - X\beta)' V(\theta)^{-1} (Z - X\beta)$$

Este método presenta ciertas dificultades dado que las propiedades de consistencia y normalidad solo se cumplen bajo ciertas circunstancias. Además, en algunos casos se puede producir multicolinealidad de la superficie de verosimilitud.

Estimación máximo verosímil restringida [17]

La estimación máximo verosímil restringida es la más utilizada en geoestadística y se basa en factorizar la verosimilitud de dos partes independientes (sin saber la media de una de estas dos partes), además no se pierde información usando parte de la verosimilitud respecto a usar verosimilitud completa. Esto hace que se obtengan estimadores con poca desviación incluso con muestras muy pequeñas.

Sean Y_1, \dots, Y_n n variables aleatorias independientes con una distribución normal $N(\mu, \sigma^2)$ con μ y σ^2 desconocidos, siendo sus estimadores máximos verosímiles definidos como:

$$\hat{\mu} = \bar{Y} = \frac{1}{n} \sum_i Y_i$$

y

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2$$

Ahora, trabajaremos con el vector de variables aleatorias independientes anterior, restándole a cada componente \bar{Y} , definiéndose como:

$$(Y_1 - \bar{Y}, Y_2 - \bar{Y}, \dots, Y_n - \bar{Y})$$

el vector anterior no depende de μ . De esta forma $\hat{\sigma}^2$ será un estimador insesgado con:

$$\sigma^2 = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2$$

Estos cambios también pueden aplicarse al modelo presentado de estimador máximo verosímil con $Z \sim N(X\beta, \Sigma)$.

Sea $W = A'Z$ un vector con A una matriz formada por $n-q$ columnas linealmente independientes y $A'X = 0$, entonces:

$$W \sim N(0, A'\Sigma A)$$

El logaritmo negativo de la función de verosimilitud definida por:

$$\ell_W(\alpha, \theta) = \frac{n-q}{2} \log(2\pi) + \frac{n-q}{2} \log \alpha + \frac{1}{2} \log |A'V(\theta)A| + \frac{1}{2\alpha} W' (A'V(\theta)A)^{-1} W \quad (2.12)$$

Si A satisface $AA' = I - X(X'X)^{-1}X'$, $A'A = I$, entonces se puede simplificar el logaritmo negativo de la función de verosimilitud quedando de la siguiente manera:

$$\begin{aligned} \ell_W(\alpha, \theta) = & \frac{n-q}{2} \log(2\pi) + \frac{n-q}{2} \log \alpha - \frac{1}{2} \log |X'X| + \frac{1}{2} \log |X'V(\theta)^{-1}X| \\ & + \frac{1}{2} \log \det(V(\theta)) + \frac{1}{2\alpha} G^2(\theta), \end{aligned} \quad (2.13)$$

siendo $G^2 = (Z - X\hat{\beta})'V^{-1}(Z - X\hat{\beta})$ la suma de cuadrados generalizados de los residuos y $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Z$.

Parte II

Teoría del Kriging

Capítulo 3

Kriging

El presente capítulo se centrará en el método Kriging, este método es muy importante en estudios geoestadísticos dado que nos permite predecir e interpolar datos espacialmente, mediante algoritmos de regresión por mínimos cuadrados.

La idea principal en la que se fundamenta el método es en predecir un valor $Z(u_0)$ a partir de n observaciones espaciales $Z(u_1), \dots, Z(u_n)$ con $u_0 \notin \{u_1, \dots, u_n\}$.

Aunque tengamos distintos métodos de Kriging, todos ellos vienen de variar el método de regresión lineal [18] [19] [20]. La predicción del valor $Z(u_0)$ se define como:

$$Z^*(u_0) - m(u_0) = \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}(u_0)[Z(u_{\alpha}) - m(u_{\alpha})] \quad (3.1)$$

con λ_{α} el peso de cada componente del sumatorio, $n(u_0)$ el número de observaciones utilizadas para predecir u_0 , los valores esperados de $Z(u_0)$ y $Z(u_{\alpha})$ son $m(u_0)$ y $m(u_{\alpha})$ respectivamente.

Otra cosa que caracteriza este método es que para predecir una localización, sea u_0 , sólo utilizaremos los datos de las u_{α} localizaciones vecinas. Además, con este método se pretende minimizar al máximo la varianza de error $\sigma_E^2(u) = \text{Var}[Z^*(u) - Z(u)]$, donde $Z^*(u)$ es el valor estimado de la localización u y $\sigma_E^2(u)$ es insesgado ($\text{E}[Z^*(u) - Z(u)] = 0$).

Habitualmente, la variable aleatoria $Z(u)$ se define como la suma de una componente residual $R(u)$ y una componente $m(u)$ determinista que define la tendencia, es decir, $Z(u) = R(u) + m(u)$.

La componente residual $R(u)$ se define como una variable aleatoria estacionaria de media cero y

covarianza $C_R(h)$, siendo $E\{R(u)\} = 0$ y $\text{Cov}\{R(u), R(u+h)\} = E\{R(u) \cdot R(u+h)\} = C_R(h)$.

Por tanto, para ver el valor esperado de una variable aleatoria Z en la localización u tenemos que calcular el valor de la tendencia en u , es decir, $E\{Z(u)\} = m(u)$.

A continuación se presentarán los tres tipos de Kriging lineales: Simple, Ordinario y Universal. Además, se presentarán los no lineales: Lognormal, Multi-Gaussiano, de Rango, Indicatriz y Disyuntivo.

3.1. Kriging Simple

Supongamos que tenemos una variable de tendencia $m(u)$, ésta es conocida y constante en toda la región de estudio A . Esto permite expresar la estimación como una combinación lineal de $n(u_0) + 1$ elementos siendo $n(u_0)$ el número de observaciones utilizadas para predecir u_0 . Además este estimador será insesgado ($E\{Z_{SK}^*(u_0) - Z(u_0)\} = m - m = 0$):

$$Z_{SK}^*(u_0) = \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{SK}(u_0) [Z(u_{\alpha}) - m] + m \quad (3.2)$$

La estimación del error es $Z_{SK}^*(u_0) - Z(u_0) = R_{SK}^*(u_0) - R(u_0)$, donde el error asociado a la estimación en este tipo de Kriging será $\sigma_{SK}^2 = Q(\lambda_{\alpha}^{SK}(u_0))$ con $\alpha = 1, \dots, n(u_0)$.

El sistema de ecuaciones obtenido para el Kriging Simple es:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \end{bmatrix} \implies$$

$$\sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{SK}(u_0) C(u_{\alpha} - u_{\beta}) = C(u_{\alpha} - u_0) \quad (3.3)$$

donde $\alpha = 1, \dots, n(u_0)$ y $C_{ij} = C(u_i - u_j)$.

La varianza que se tiene en el método de Kriging simple es $\sigma_{SK}^2(u_0) = C(0) - \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{SK}(u_0) C_{\alpha 0}$ y el peso de la media m es $\lambda_m^{SK}(u_0) = 1 - \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{SK}(u_0)$.

3.2. Kriging Ordinario

Supongamos que se quiere obtener el valor de $Z(u_0)$ donde no tenemos constancia de su valor a partir de las observaciones $Z(u_1), \dots, Z(u_n)$. El Kriging Ordinario nos permite obtener $Z(u_0)$ teniendo en cuenta las fluctuaciones locales de tendencia o media que se presenten. Para ello se limita el dominio de estacionariedad de la media a la vecindad local $W(u) : m(u^*)$. Este estimador se define como:

$$m_{OK}^*(u_0) = \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{OK}(u_0) Z(u_{\alpha}) \quad (3.4)$$

donde $\sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{OK}(u_0) = 1$ y $n(u_0)$ el número de observaciones utilizadas para predecir u_0 .

Este tipo de Kriging esta formado por un sistema de $n(u_0) + 1$ ecuaciones lineales, es decir:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}$$

Con la misma covarianza residual (en la práctica) que la de Z , por tanto:

$$\left\{ \begin{array}{l} \sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{OK}(u_0) C(u_{\alpha} - u_{\beta}) + L_{OK}(u_0) = C(u_{\alpha} - u_{\beta}) \quad \alpha, \beta = 1, \dots, n(u_0) \\ \sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{OK}(u_0) = 1 \end{array} \right.$$

siendo $L_{OK}(u_0)$ el parámetro de Lagrange que se tiene en cuenta a la hora de construir los pesos ($\lambda_{\alpha}^{OK}(u_0)$).

La varianza del Kriging Ordinario viene por:

$$\sigma_{OK}^2(u_0) = C(0) - \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{OK}(u_0) C(u_{\alpha} - u_0) - L_{OK}(u_0)$$

En particular, este método nos permite estimar la media del atributo localmente.

3.3. Kriging con modelo de tendencia o universal

Supongamos que la media local no es conocida y varía suavemente en cada punto de la región $W(u)$ [21] [22].

La componente de tendencia viene definida por una combinación lineal de K funciones, como se muestra a continuación:

$$m(u^*) = \sum_{k=0}^K a_k(u^*) f_k(u^*) \quad (3.5)$$

siendo $a_k(u^*)$ una constante que desconocemos con $u^* \in W(u)$ y $f_0(u^*) = 1$.

Consideramos que Z se define como $Z(u) = m(u) + R(u) = \sum_{k=0}^K a_k f_k(u) + R(u)$, siendo $f_k(u)$ funciones conocidas que dependen de las coordenadas y a_k (desconocidas). Además la componente residual es aleatoria estacionaria con media cero y covarianza $C_R(h)$.

Una vez visto lo anterior ya podemos definir el estimador Kriging de Tendencia o Universal, siendo éste:

$$\begin{aligned} Z_{KT}^*(u_0) &= \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{KT}(u_0) Z(u_{\alpha}) \\ \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{KT}(u_0) f_k(u_{\alpha}) &= f_k(u_0) \quad k = 0, \dots, K \end{aligned} \quad (3.6)$$

El sistema de ecuaciones del estimador en el que nos encontramos es el siguiente:

$$\begin{cases} \sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{KT}(u_0) C_{R\alpha\beta} + \sum_{k=0}^K L_k^{KT}(u_0) f_k(u_0) = C_{R\alpha\beta} \\ \sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{KT}(u_0) f_k(u_{\beta}) = f_k(u_0) \end{cases}$$

con $\alpha, \beta = 1, \dots, n(u_0)$, $k = 0, \dots, K$ y $\sum_{\beta=1}^{n(u_0)} \lambda_{\beta}^{KT}(u_0) = 1$.

En forma matricial el sistema quedaría de la siguiente manera:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & f_{11} & \dots & f_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} & f_{21} & \dots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & f_{1n} & \dots & f_{pn} \\ f_{11} & f_{12} & \dots & f_{1n} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{p1} & f_{p2} & \dots & f_{pn} & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ L_{k_1}^{KT} \\ \dots \\ L_{k_n}^{KT} \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ \dots \\ C_{n0} \\ f_{10} \\ \dots \\ f_{p0} \end{bmatrix}$$

La varianza asociada al estimador de Kriging de Tendencia es:

$$\sigma_{KT}^2(u_0) = C_R(0) - \sum_{\alpha=1}^{n(u_0)} \lambda_{\alpha}^{KT}(u_0) C_R(u_{\alpha} - u_0) - \sum_{k=0}^K L_k^{KT}(u_0) f_k(u_0)$$

Finalmente, se tiene que destacar que si $K = 0$, el Kriging de Tendencia y el Ordinario coinciden.

3.3.1. Kriging con factor externo

Una variante del Kriging de tendencia es el Kriging con factor externo o external drift. El Kriging con factor externo es simple y efectivo a la hora de incorporar un efecto de una segunda variable Y en la estimación de Z . La variable Y será válida si varía suavemente en toda la región de estudio. Además debe conocerse en todos los puntos u_{α} donde se conoce Z y donde se tenga que estimar Z . La tendencia vendrá definida por:

$$E[Z(u)] = m(u) = a_0 + a_1 Y(u)$$

El estimador Z en este caso se definirá como:

$$Z_{KFE}^*(u_0) = \sum_{k=1}^n w_k^{KFE}(u_0) \cdot Z(u_k), \quad (3.7)$$

siendo w_k^{KFE} los pesos en cada posición del área de estudio.

3.4. Kriging en Bloques

El Kriging en Bloques pretende realizar una estimación sobre una región o bloque con tamaño y forma concretos. Para llevar a cabo este método consideramos V_k como el bloque o región de estudio con área A_k y el punto de estudio u_k . Nuestro estimador será:

$$Z_{V_k}^* = \sum_{i=1}^n \lambda_i Z(u_i) \quad (3.8)$$

Además las covarianzas con respecto al bloque V_k se expresan como:

$$C_{V_k,i} = \frac{1}{A_k} \int_{V_k} C(u - u_i) dx$$

Siendo el sistema del Kriging por Bloques de la siguiente manera:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & 0 & \dots & C_{2n} & 1 \\ \dots & \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{V_k,1} \\ C_{V_k,2} \\ \dots \\ C_{v_k,n} \\ 1 \end{bmatrix}$$

siendo $C_{V_i,n} = \frac{1}{|A|} \sum_{j \in V_k} C_{ij}$

La varianza vendrá definida por:

$$\sigma_V^2(u_0) = \mu - \sum_{\alpha=1}^{n(u_0)} \lambda_\alpha^V V(u_0) C(u_\alpha - V(u_0)) - L_V(u_0),$$

donde $\mu = C(V(u_0), V(u_0))$ y $L_V(\Delta)$ es el multiplicador de Lagrange.

3.5. Kriging Factorial

Muchas veces nos sería de ayuda poder considerar la descomposición de $Z(u)$ en varios componentes independientes, es decir, $Z(u) = Z_0(u) + Z_1(u) + \dots + Z_{NC}(u)$.

La covarianza de Z vendrá definida por:

$$C_Z(h) = \sum_{i=0}^{NC} C_i(h)$$

Y el sistema de ecuaciones del kriging Factorial será de la forma:

$$\sum_{\beta=1}^{n(u_0)} d_{\beta}(u_0) C_{Z\beta\alpha} + L_F(u_0) \tag{3.9}$$

con $\sum_{\beta=1}^{n(u_0)} d_{\beta}(u_0) = 1$.

3.6. Cokriging

El Cokriging se utiliza cuando, al realizar la regresión, participan varios atributos. Sean Z e Y dos variables definidas en la misma localización [23]. Para estimar la variable principal Z en el punto u_0 realizaremos el siguiente cálculo:

$$Z_{COK}^*(u_0) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(u_0) Z(u_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}^*(u_0) Y(u_{\alpha_2}^*) \tag{3.10}$$

siendo necesarias para la estimación las covarianzas de Z e Y y las covarianzas cruzadas de $Z - Y$ e $Y - Z$, siendo éstas $C_{ZY}(h) = \text{Cov}\{Z(u), Y(u+h)\}$ y $C_{YZ}(h) = \text{Cov}\{Y(u), Z(u+h)\}$ respectivamente.

3.7. Krigings no lineales

Debido a la transformación de los atributos originales, podemos encontrarnos ante métodos de Kriging no lineales. Algunos de ellos se expondrán a continuación:

- El **Kriging Lognormal o logarítmico** es una técnica de estimación diseñada para trabajar con distribuciones de datos muy sesgadas. Este tipo de Kriging aprovecha una transformación logarítmica que reduce la variación de datos.
- El **Kriging Multi-Gaussiano** es una generalización del Kriging Lognormal y consiste en transformar Z en una distribución normal estándar.
- El **Kriging de Rango**, como su nombre indica, se aplica sobre una variable de rango para transformar los datos.
- El **Kriging Indicatriz** tiene una media constante y no conocida. Además la predicción es una variable binaria.
- El **Kriging Disyuntivo** es un Kriging aplicado sobre una transformación polinómica.

3.8. Validación cruzada

Una vez hayamos hecho la modelización no podemos suponer que ésta tendrá precisión. Necesitamos una garantía de que nuestro modelo hace buenas predicciones. En geoestadística la Validación Cruzada es la más utilizada para comprobar la efectividad del modelo. Ésta nos permite comparar los valores que se han estimado con los valores reales. Para realizar esta comparación, los datos los separamos en dos grupos: test y training. El training nos permitirá hacer las predicciones de los datos del grupo test para al fin poder comparar las predicciones con los datos del grupo test.

Cuando realizamos Validación cruzada para seleccionar el modelo de semivariograma que se utilizará finalmente, tendremos en cuenta las siguientes condiciones:

- Si realizamos un remuestreo del semivariograma éste no influirá ni en los pesos ni en Krigings.
- Al realizar la validación cruzada no pueden ser utilizados el *sill* relativo y el comportamiento del semivariograma a la vez.
- Si el modelo no es el adecuado no sabremos que parámetros se deben cambiar.

Parte III

Aplicación: Análisis geoestadístico espacial sobre los casos de COVID-19 en la España peninsular y en la Comunidad Valenciana

Capítulo 4

Aplicación de la interpolación Kriging a la expansión geográfica de la pandemia COVID-19

4.1. Introducción

El presente trabajo tiene como objetivo estudiar la predicción de la incidencia del COVID-19 tanto en la zona peninsular Española como en la Comunidad Valenciana. Para ello, se ha utilizado el método Kriging partiendo de las coordenadas que limitan estos dos territorios y los casos de coronavirus activos en distintos puntos de la zona de estudio. El código utilizado para realizar la aplicación se ve detallado en los Anexos A y B.

4.2. Descripción de los datos utilizados

Como ya se ha comentado previamente las predicciones se hicieron sobre dos zonas distintas, la zona peninsular española y la zona de la Comunidad Valenciana.

4.2.1. Zona peninsular española

Contorno peninsular

Con el fin de poder limitar la zona de predicción se han utilizado 324528 pares de coordenadas (Longitud (o coordenada x) y Latitud (o coordenada y)) y así dibujar con precisión el límite de la península [27].

Una vez se tienen estos datos se crea un grid del interior de España donde haremos las predicciones

del número de casos. El grid creado para esta situación se muestra a continuación en la figura 4.1.

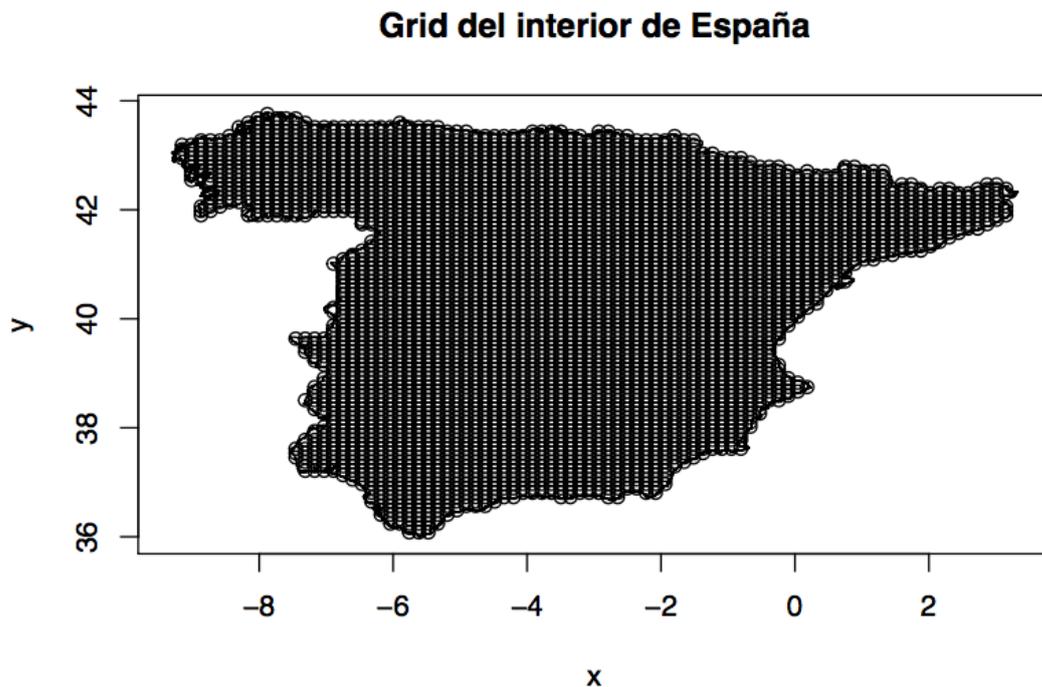


Figura 4.1: grid generado con las coordenadas del contorno de España

Datos de COVID-19 en el territorio español

Para poder hacer las predicciones se han recogido datos de todas las provincias españolas de la península ibérica (ver [24]), situándolos en longitud y latitud de las capitales de provincia, por tanto en total teníamos 47 muestras. La simplificación que se ha realizado con sólo un dato por provincia puede no ser del todo fiable pero no se disponen de datos más precisos.

En esta área de estudio se han realizado dos predicciones distintas. La primera de ellas es en plena pandemia del virus, es decir, el 18 de abril de 2020. La otra fecha sería cuando todo el territorio nacional estaba en proceso del desconfinamiento en las fases 1 y 2, es decir, el 3 de junio. Las tablas que recogen los datos se muestran a continuación en las tablas 4.1 y 4.2, teniendo a nuestra disposición los siguientes datos:

- Nombre de la provincia.
- Fecha de recogida de las muestras.

- Casos activos de COVID-19 en el momento de la recogida de los datos.
- Longitud de la capital de la provincia.
- Latitud de la capital de la provincia.

Casos de COVID-19 el 18 de abril del 2020

Provincia	Fecha	CASOS	Longitud	Latitud
A	2020-04-18	73	38.3452	-0.481006
AB	2020-04-18	35	38.99765	-1.86007
AL	2020-04-18	7	36.84016	-2.467922
AV	2020-04-18	15	40.65642	-4.700323
B	2020-04-18	373	41.38792	2.169919
BA	2020-04-18	28	38.8786	-6.970284
BI	2020-04-18	92	43.25696	-2.923441
BU	2020-04-18	17	42.34087	-3.699731
C	2020-04-18	42	43.35087	-8.395835
CA	2020-04-18	17	36.53969	-6.262657
CC	2020-04-18	53	39.47618	-6.37076
CO	2020-04-18	17	37.88473	-4.779152
CR	2020-04-18	40	38.9861	-3.927263
CS	2020-04-18	16	39.98598	-0.0376709
CU	2020-04-18	4	40.07183	-2.134005
GI	2020-04-18	49	41.9818	2.8237
GR	2020-04-18	28	37.17649	-3.597929
GU	2020-04-18	4	40.62981	-3.166493
H	2020-04-18	0	37.2571	-6.949555
HU	2020-04-18	1	42.1401	-0.408898
J	2020-04-18	10	37.76574	-3.789518
L	2020-04-18	11	41.61415	0.6257825
LE	2020-04-18	6	42.59988	-5.571752
LO	2020-04-18	17	42.46577	-2.449995
LU	2020-04-18	13	43.01208	-7.555851
M	2020-04-18	467	40.41669	-3.700346
MA	2020-04-18	47	36.72965	-4.420019
MU	2020-04-18	5	37.98344	-1.12989
NA	2020-04-18	65	42.81721	-1.646767
O	2020-04-18	9	43.36026	-5.844759
OR	2020-04-18	41	42.34001	-7.864641
P	2020-04-18	22	42.01246	-4.531175
PO	2020-04-18	11	42.43362	-8.648053

Provincia	Fecha	CASOS	Longitud	Latitud
S	2020-04-18	13	43.45096	-3.807934
SA	2020-04-18	27	40.96497	-5.663047
SE	2020-04-18	9	37.38264	-5.996295
SG	2020-04-18	23	40.94943	-4.119209
SO	2020-04-18	30	41.7636	-2.464921
SS	2020-04-18	10	37.65493	-4.824157
T	2020-04-18	16	41.11866	1.24533
TE	2020-04-18	2	40.34411	-1.10691
TO	2020-04-18	14	39.85678	-4.024476
V	2020-04-18	108	39.47024	-0.3768049
VA	2020-04-18	25	41.65295	-4.728388
VI	2020-04-18	6	42.84641	-2.667893
Z	2020-04-18	54	41.65629	-0.8765379
ZA	2020-04-18	6	41.50368	-5.743778

Cuadro 4.1: casos de COVID-19 en la zona peninsular Española el 18 de abril del 2020.

Casos de COVID-19 el 3 de junio del 2020

Provincia	Fecha	CASOS	Longitud	Latitud
Alicante	2020-06-03	1	38.3452	-0.481006
Albacete	2020-06-03	1	38.99765	-1.86007
Almeria	2020-06-03	2	36.84016	-2.467922
Avila	2020-06-03	1	40.65642	-4.700323
Alava	2020-06-03	2	42.84641	-2.667893
Barcelona	2020-06-03	79	41.38792	2.169919
Badajoz	2020-06-03	3	38.8786	-6.970284
Vizcaya	2020-06-03	9	43.25696	-2.923441
Burgos	2020-06-03	3	42.34087	-3.699731
La-Coruña	2020-06-03	1	43.35087	-8.395835
Cadiz	2020-06-03	2	36.53969	-6.262657
Caceres	2020-06-03	5	39.47618	-6.37076
Cordoba	2020-06-03	1	37.88473	-4.779152
Ciudad-Real	2020-06-03	12	38.9861	-3.927263
Castellon	2020-06-03	3	39.98598	-0.0376709
Cuenca	2020-06-03	3	40.07183	-2.134005
Gerona	2020-06-03	14	41.9818	2.8237
Granada	2020-06-03	1	37.17649	-3.597929
Guadalajara	2020-06-03	2	40.62981	-3.166493

Provincia	Fecha	CASOS	Longitud	Latitud
Huelva	2020-06-03	0	37.2571	-6.949555
Huesca	2020-06-03	3	42.1401	-0.408898
Jaen	2020-06-03	1	37.76574	-3.789518
Lerida	2020-06-03	17	41.61415	0.6257825
Leon	2020-06-03	3	42.59988	-5.571752
La-Rioja	2020-06-03	2	42.46577	-2.449995
Lugo	2020-06-03	0	43.01208	-7.555851
Madrid	2020-06-03	74	40.41669	-3.700346
Malaga	2020-06-03	3	36.72965	-4.420019
Murcia	2020-06-03	0	37.98344	-1.12989
Navarra	2020-06-03	5	42.81721	-1.646767
Asturias	2020-06-03	0	43.36026	-5.844759
Orense	2020-06-03	0	42.34001	-7.864641
Palencia	2020-06-03	3	42.01246	-4.531175
Pontevedra	2020-06-03	0	42.43362	-8.648053
Cantabria	2020-06-03	0	43.45096	-3.807934
Salamanca	2020-06-03	4	40.96497	-5.663047
Sevilla	2020-06-03	1	37.38264	-5.996295
Segovia	2020-06-03	2	40.94943	-4.119209
Soria	2020-06-03	3	41.7636	-2.464921
Guipuzcoa	2020-06-03	1	37.65493	-4.824157
Tarragona	2020-06-03	2	41.11866	1.24533
Teruel	2020-06-03	0	40.34411	-1.10691
Toledo	2020-06-03	2	39.85678	-4.024476
Valencia	2020-06-03	5	39.47024	-0.3768049
Valladolid	2020-06-03	5	41.65295	-4.728388
Zaragoza	2020-06-03	9	41.65629	-0.8765379
Zamora	2020-06-03	2	41.50368	-5.743778

Cuadro 4.2: casos de COVID-19 en la zona peninsular Española el 3 de junio del 2020.

Procedemos a un estudio de la variable dependiente, es decir, el número de casos de COVID-19. La distribución que presentan los datos se encuentra resumida a continuación en las tablas 4.3 y 4.4 para los días 18 de abril y 3 de junio respectivamente. Además, también se muestra en los histogramas dicha distribución 4.2 y 4.3.

Minímo	Primer cuantil	Mediana	Media	Tercer cuantil	Máximo
0.00	9.50	17.00	42.09	40.50	467.00

Cuadro 4.3: resumen estadístico de los datos de España el 18 de abril de 2020.

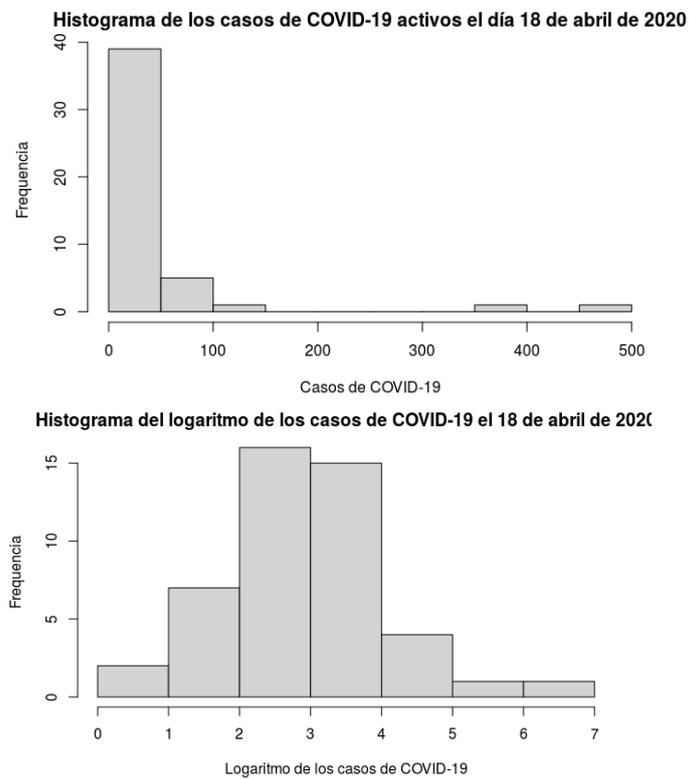


Figura 4.2: histograma de los casos de COVID-19 por provincias españolas el día 18 de abril de 2020.

Minímo	Primer cuantil	Mediana	Media	Tercer cuantil	Máximo
0.000	1.000	2.000	6.213	3.500	79.000

Cuadro 4.4: resumen estadístico de los datos de España el 3 de junio de 2020.

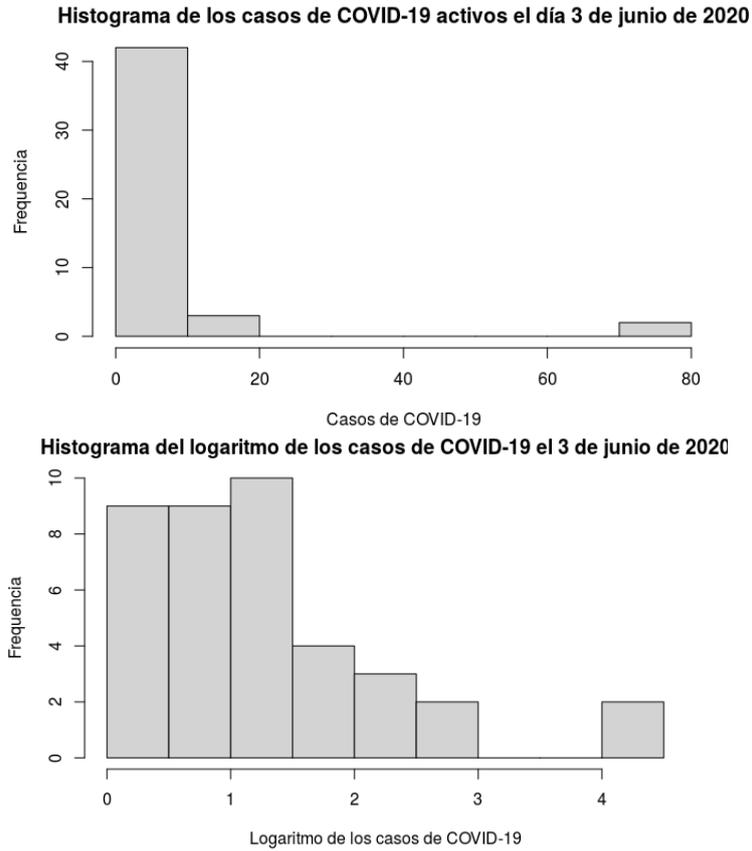


Figura 4.3: histograma de los casos de COVID-19 por provincias españolas el día 3 de junio de 2020.

4.2.2. Zona de la Comunidad Valenciana

Contorno de la Comunidad Valenciana

Para poder limitar la zona de predicción se han utilizado 2478 pares coordenadas (Longitud (o coordenada x) y Latitud (o coordenada y)) [26].

A partir de los datos de la tabla anterior se crea un grid que abarca todo el territorio Comunidad Valenciana, este grid se muestra a continuación en la figura 4.4.

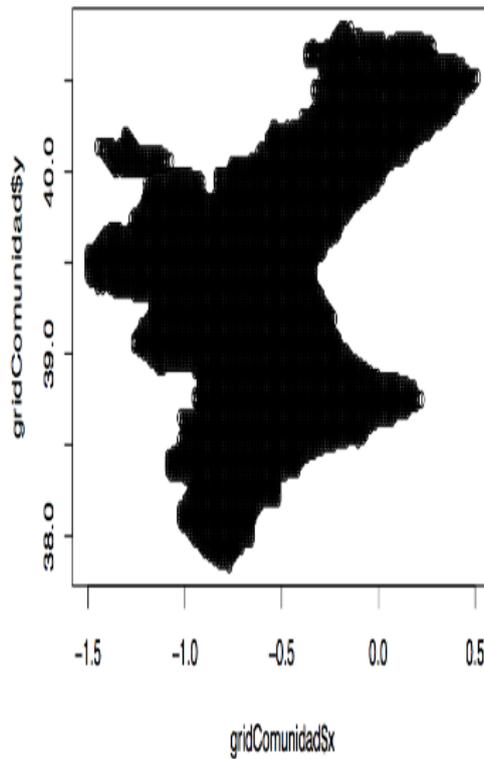


Figura 4.4: grid generado con las coordenadas del contorno de la Comunidad Valenciana.

Casos de COVID-19 Comunidad Valenciana

En el caso de la Comunidad Valenciana se ha utilizado una muestra formada por los datos recogidos de 542 localizaciones situadas en la superficie de estudio para el día 3 de junio de 2020 (ver [25]). Parte de la tabla que recoge los datos se muestran a continuación en la Tabla 4.5, teniendo a nuestra disposición los siguientes datos:

- Nombre de la población de donde tenemos los datos.
- Casos activos COVID-19 el 3 de junio de 2020.
- Casos de COVID-19 que han aparecido en los últimos 14 días.
- Fallecidos totales producidos por el COVID-19.
- Longitud de la capital de la provincia.
- Latitud de la capital de la provincia.

Población	Casos totales	Casos en últimos 14 días	Fallecidos	Longitud	Latitud
Ademuz	0	0	0	40.0750612319406	-1.2234098552
Ador	1	0	1	38.9333957743436	-0.2458111831
Agost	4	0	0	38.4485810	-0.6175769
Agres	0	0	0	38.7873758033688	-0.5123594218
Agullent	11	0	3	38.8423077903515	-0.5532777310
Aielo de Malferit	1	0	0	38.8867045449057	-0.6015268780
Aielo de Rugat	0	0	0	38.880984747761	-0.34485308315
Aigiües	0	0	0	38.346041	-0.484756
Aín	0	0	0	39.8921613977122	-0.3433565894
Alacant	853	11	87	38.4091508997595	-0.5127874934
Alaquàs	41	4	9	39.4563421919474	-0.4609107971
Albaida	2	0	0	38.853984813475	-0.52591297732
Albal	11	0	0	39.3917339552034	-0.4073643189
Albalat de la Ribera	4	0	0	39.2184835926986	-0.3738811106
Albalat dels Sorells	3	0	0	39.5459869546034	-0.3406259630
Albalat dels Tarongers	2	0	0	39.6828820828349	-0.3558499871
Albatera	4	0	1	38.2184703129668	-0.8969930557
Alberic	10	1	0	39.1187931934469	-0.5186253222
Albocàsser	3	0	1	40.3953996990405	0.03150439007
Alborache	40	0	9	39.3822012598204	-0.7565571727
Alboraia	33	1	4	39.5045293512052	-0.3380163501
Albuixech	25	0	8	39.5441492155589	-0.3217307842

Población	Casos totales	Casos en últimos 14 días	Fallecidos	Longitud	Latitud
Alcalà de Xivert	11	1	0	40.2579760662217	0.26890921572
Alcalalí	1	0	0	38.7482408881678	-0.0331180914
Alcàntera de Xúquer	1	0	0	39.0659424640052	-0.5555060036
Alcàsser	12	0	4	39.3779782184498	-0.4471027584
Alcocer de Planes	0	0	0	38.7885800351272	-0.4034330080
Alcoi	249	4	81	38.346041	-0.484756
Alcoleja	0	0	0	38.6699158246792	-0.3238175629
Alcublas	0	0	0	39.8062934441634	-0.6891239483
Alcudia de Veo	3	0	1	39.9236993117592	-0.3481725588
Aldaia	44	0	2	39.4646917930316	-0.4616832733
Alfajar	36	0	8	39.4097187632624	-0.3689823355
Alfajara	1	0	0	38.780631739482	-0.57339933020

Cuadro 4.5: casos de COVID-19 en la Comunidad Valenciana el 3 de junio del 2020.

Estudiando la variable dependiente, al igual que antes, ésta es el número de casos de COVID-19 en la Comunidad Valenciana. La distribución estadística que presentan los datos se encuentra resumida a continuación en la Tabla 4.6. En la figura 4.5 se muestra un histograma que explica el número de casos de contagios en función de la frecuencia.

Minímo	Primer cuantil	Mediana	Media	Tercer cuantil	Máximo
0.00	0.00	2.00	16.89	9.00	853.00

Cuadro 4.6: resumen estadístico de los datos de España el 3 de junio de 2020.

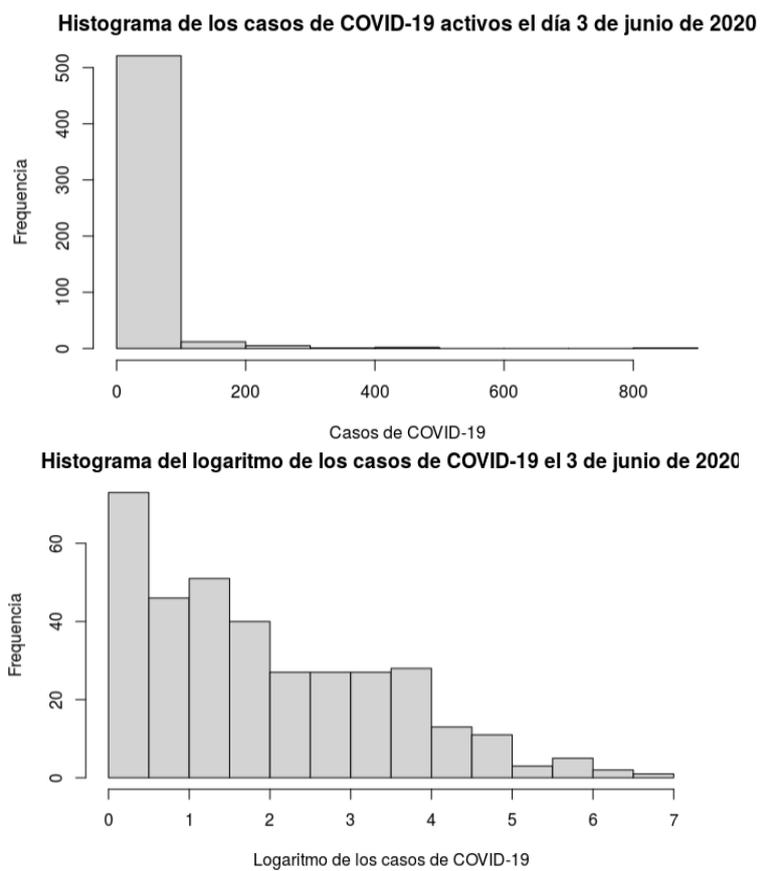


Figura 4.5: histograma de los casos de COVID-19 en la Comunidad Valenciana el día 3 de junio de 2020.

4.2.3. Los datos en su ubicación espacial

Una vez se dispone de toda la información representamos el mapa de España (Figura 4.6) y de la Comunidad Valenciana (Figura 4.7) incluyendo las localizaciones de los datos con puntos de diámetro proporcional al número de casos:

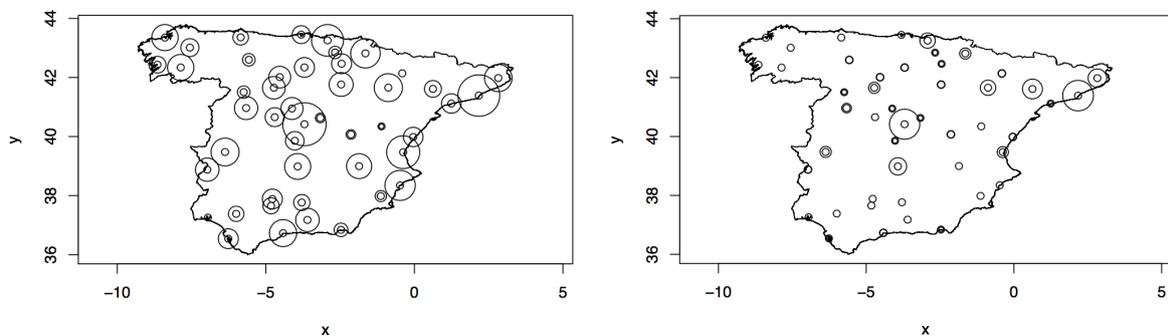


Figura 4.6: casos de COVID-19 en su ubicación espacial en el territorio Español (a la izquierda se encuentra la representación del día 18 de abril de 2020 y a la derecha la representación del día 3 de junio de 2020).

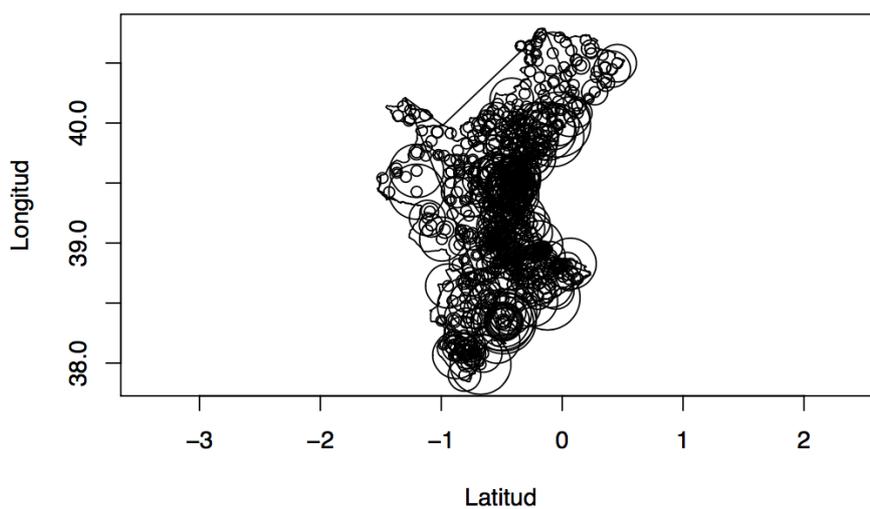


Figura 4.7: casos de COVID-19 en su ubicación espacial en el territorio de la Comunidad Valenciana para el día 3 de junio de 2020.

Capítulo 5

Resultados

Vamos a predecir la distribución del número de casos de COVID-19 tanto en el territorio español como en la Comunidad Valenciana. Para ello se hará uso del Kriging Simple con distintos modelos de variograma. Los métodos utilizados serán: Esférico, Gaussiano y Exponencial.

Previamente a ver cómo se distribuyen los datos en el mapa, se realizarán los variogramas pertinentes.

Los variogramas son una herramienta muy efectiva en los estudios de predicción espacial, dado que nos muestran gráficamente la precisión que está teniendo nuestro modelo.

En el momento de calcular el variograma, el principal objetivo es que éste se ajuste a nuestros datos para que describa la estructura espacial de nuestras observaciones. Al visualizar la distribución de los datos podemos ajustar distintos parámetros, los cuales son:

- Cutoff: es un parámetro de la función *variogram* del paquete de R *gstat* y se utiliza para modificar la distancia espacial máxima tomada en cuenta entre dos observaciones.
- Width: es un parámetro de la función *variogram* del paquete de R *gstat* y se utiliza para modificar el intervalo de distancia sobre el cual se calcula la semi-varianza.
- Range: Es la distancia que tienen pares de puntos los cuales no tienen dependencia espacial. Ver Figura 5.1.
- Nugget: Teóricamente, este variograma se define como nulo en el origen, pero en casos prácticos siempre existe una pequeña diferencia entre el origen y el valor que toma el variograma, este valor hay que tenerlo en cuenta. Si obtenemos un valor de nugget negativo,

lo consideraremos como 0. Ver Figura 5.1.

- Psill: Es la distancia en la que el modelo alcanza el rango menos el valor del nugget. Ver Figura 5.1.

En caso que dejemos estos valores como los predefinidos por R en algunas situaciones no obtendremos variogramas adecuados.

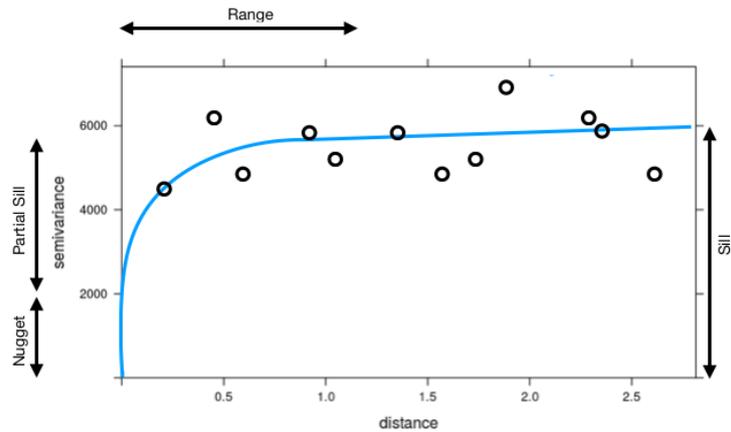


Figura 5.1: esquema de los parámetros del variograma.

Una vez tenemos el variograma aplicaremos el método de Kriging y representaremos el número de casos y la varianza de las predicciones en un mapa.

5.1. Variogramas

Casos de COVID-19 en el territorio Español

Los resultados de los variogramas para los casos de COVID-19 en el territorio español obtenidos para los días 18 de abril y 3 de junio de 2020 se muestran a continuación en las figuras 5.2 y 5.3 respectivamente.

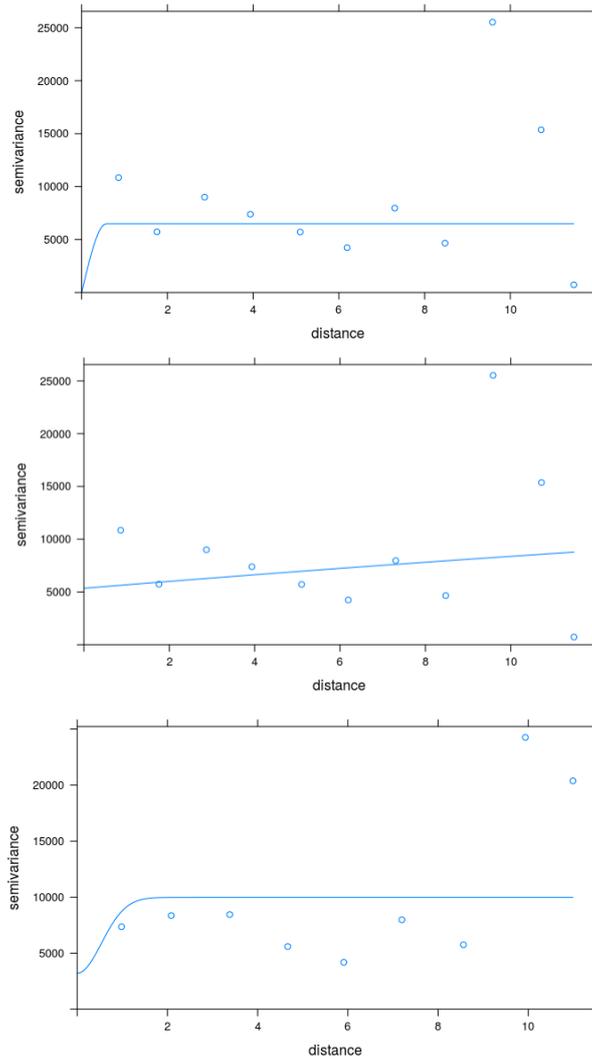


Figura 5.2: variogramas obtenidos para los casos de COVID-19 para el día 18 de abril, siendo el primer variograma generado por el modelo Esférico, el segundo generado por el modelo Exponencial y el tercero por el modelo Gaussiano.

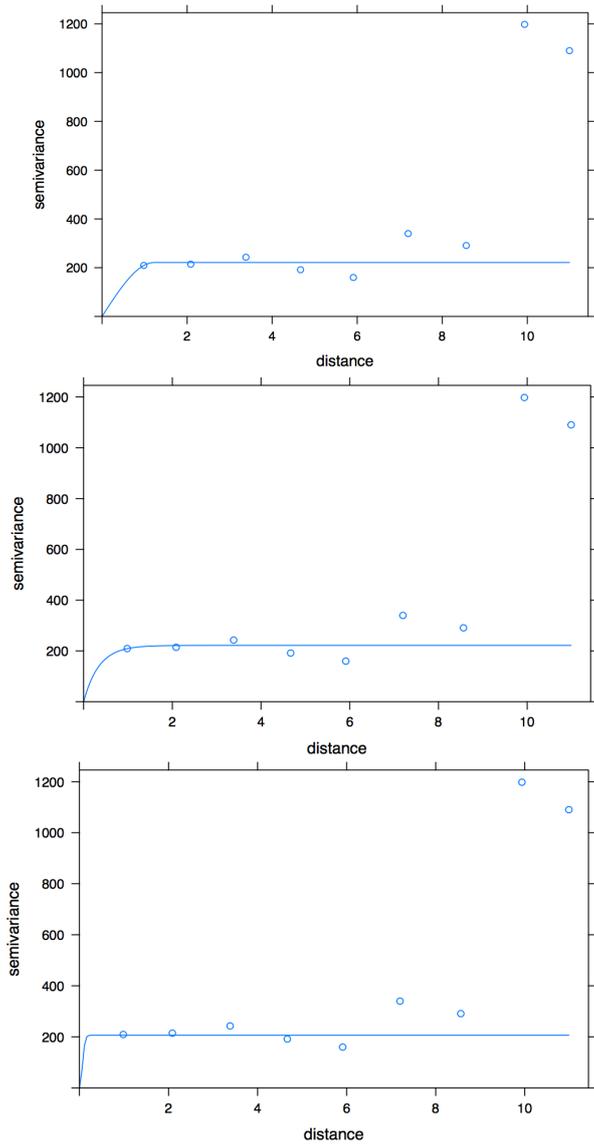


Figura 5.3: variogramas obtenidos para los casos de COVID-19 para el día 3 de junio siendo el primer variograma generado por el modelo Esférico, el segundo generado por el modelo Exponencial y el tercero por el modelo Gaussiano.

Casos de COVID-19 en la Comunidad Valenciana

Por otra parte, los variogramas obtenidos para los casos de COVID-19 en la Comunidad Valenciana el 3 de junio de 2020 se presentan a continuación en la figura 5.4.

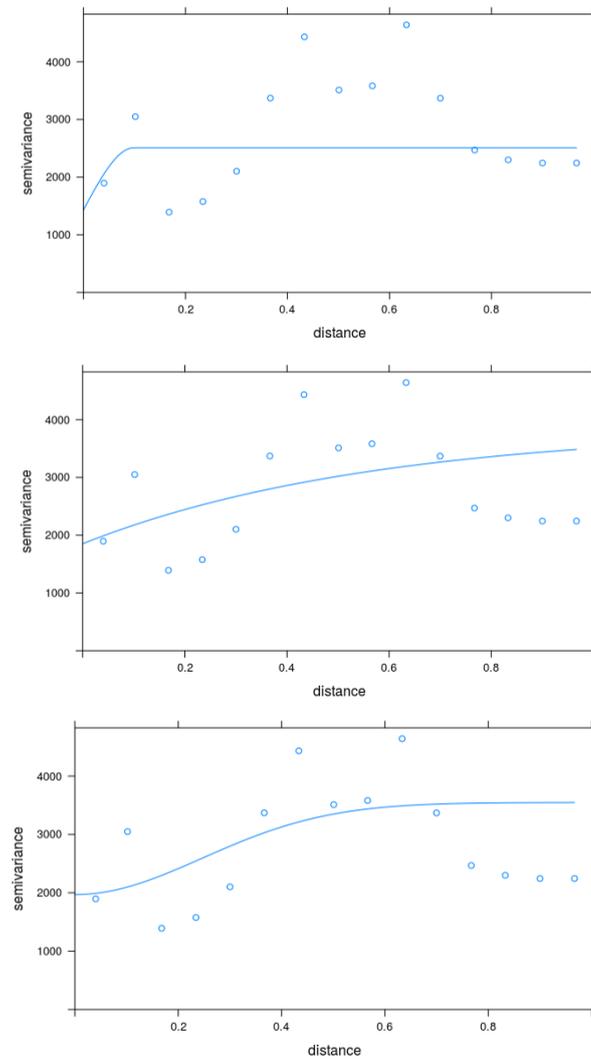


Figura 5.4: variogramas obtenidos para los casos de COVID-19 para el día 3 de junio. Siendo el primer variograma generado por el modelo Esférico, el segundo generado por el modelo Exponencial y el tercero por el modelo Gaussiano.

5.2. Predicciones

Casos de COVID-19 en el territorio español

Las predicciones y la varianza de las mismas, obtenidas a partir de los variogramas anteriores se muestran a continuación en las figuras 5.5, 5.6 y 5.7 para los datos obtenidos del día 18 de abril de 2020. Por otra parte, en las figuras 5.8, 5.9 y 5.10 se muestran para los casos de COVID-19 del día 3 de junio de 2020.

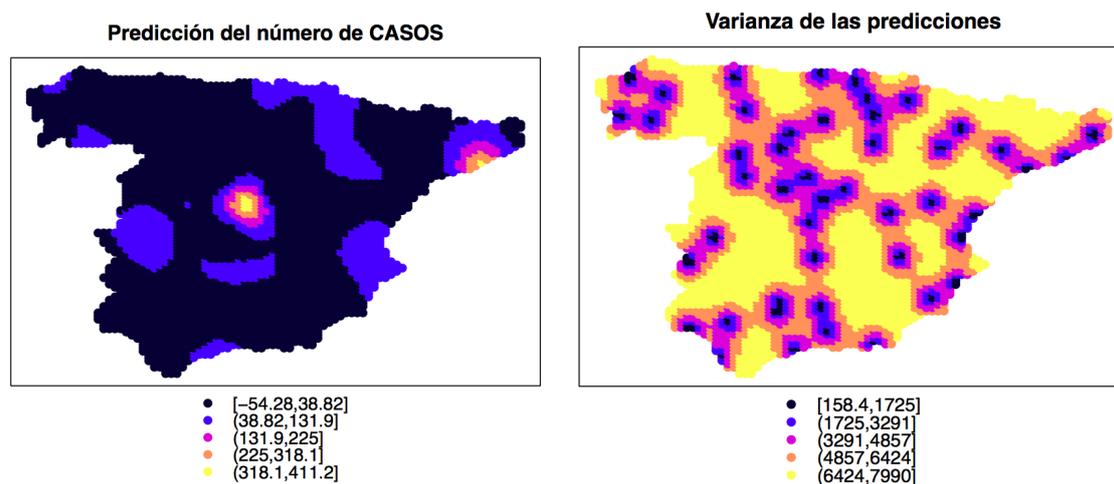


Figura 5.5: predicciones y varianzas obtenidas a partir del variograma Esférico para los casos de COVID-19 del día 18 de abril de 2020.

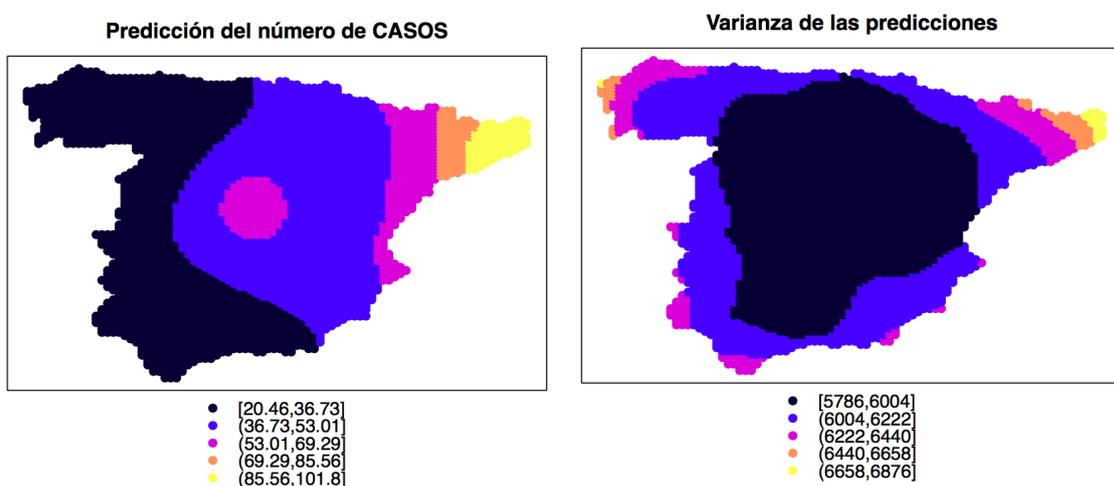


Figura 5.6: predicciones y varianzas obtenidas a partir del variograma Exponencial para los casos de COVID-19 del día 18 de abril de 2020.

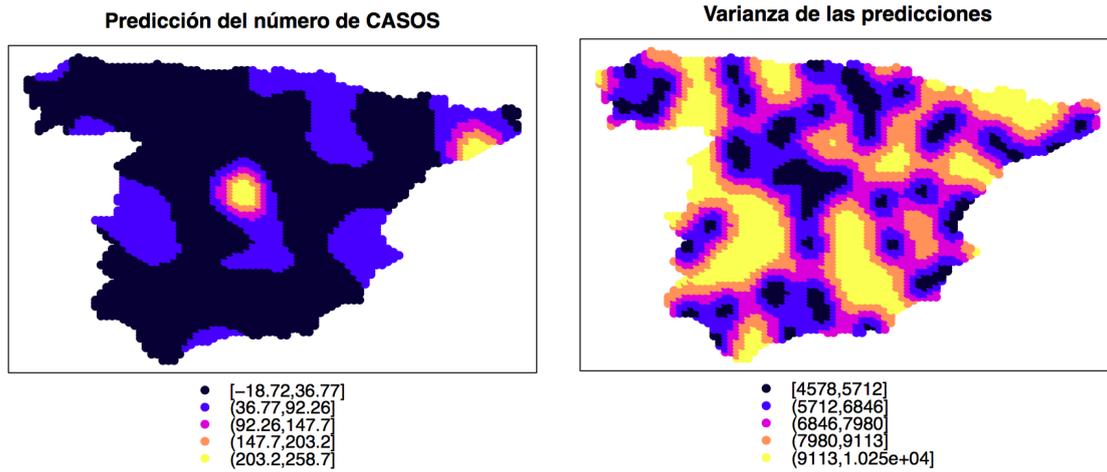


Figura 5.7: predicciones y varianzas obtenidas a partir del variograma Gaussiano para los casos de COVID-19 del día 18 de abril de 2020.

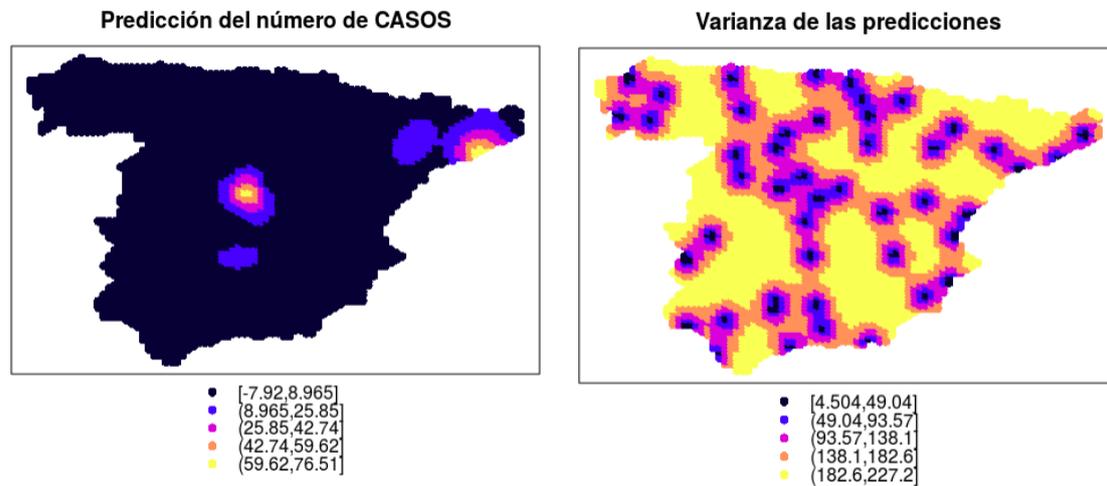


Figura 5.8: predicciones y varianzas obtenidas a partir del variograma Esférico para los casos de COVID-19 del día 3 de junio de 2020.

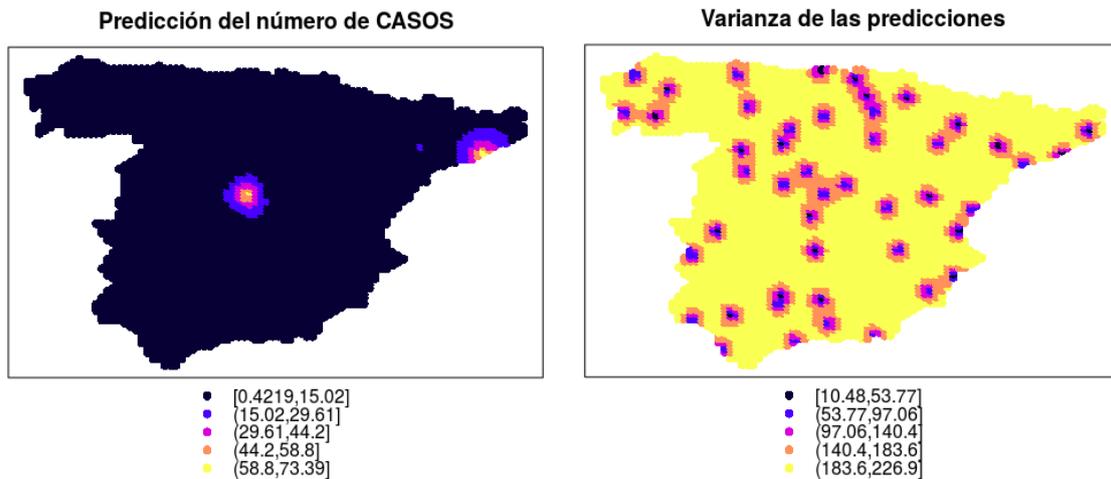


Figura 5.9: predicciones y varianzas obtenidas a partir del variograma Exponencial para los casos de COVID-19 del día 3 de junio de 2020.

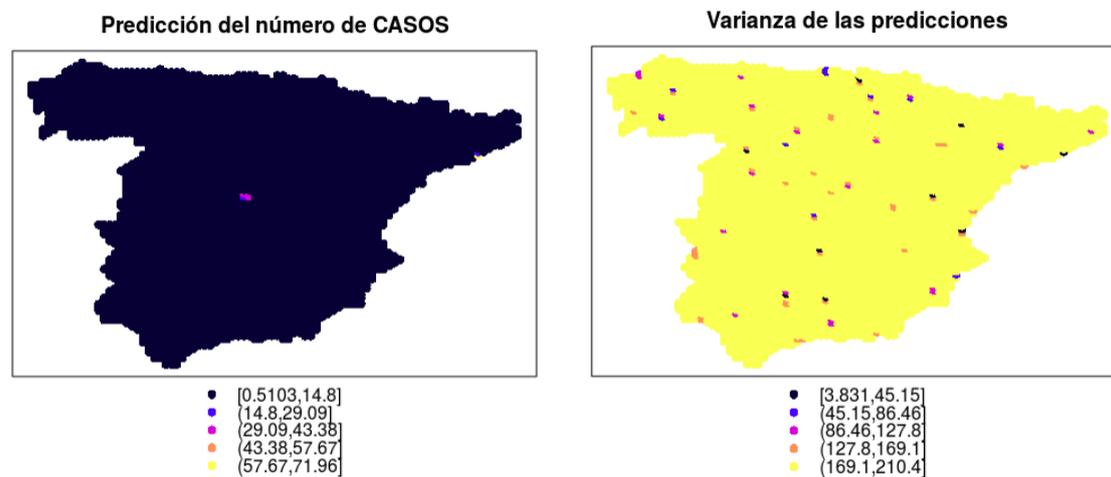


Figura 5.10: predicciones y varianzas obtenidas a partir del variograma Gaussiano para los casos de COVID-19 del día 3 de junio de 2020.

Casos de COVID-19 en el territorio Comunidad Valenciana

Las predicciones y la varianza de las mismas, obtenidas a partir de los variogramas anteriores se muestran a continuación en las figuras 5.11, 5.12 y 5.13 para los datos obtenidos el día 3 de junio de 2020.

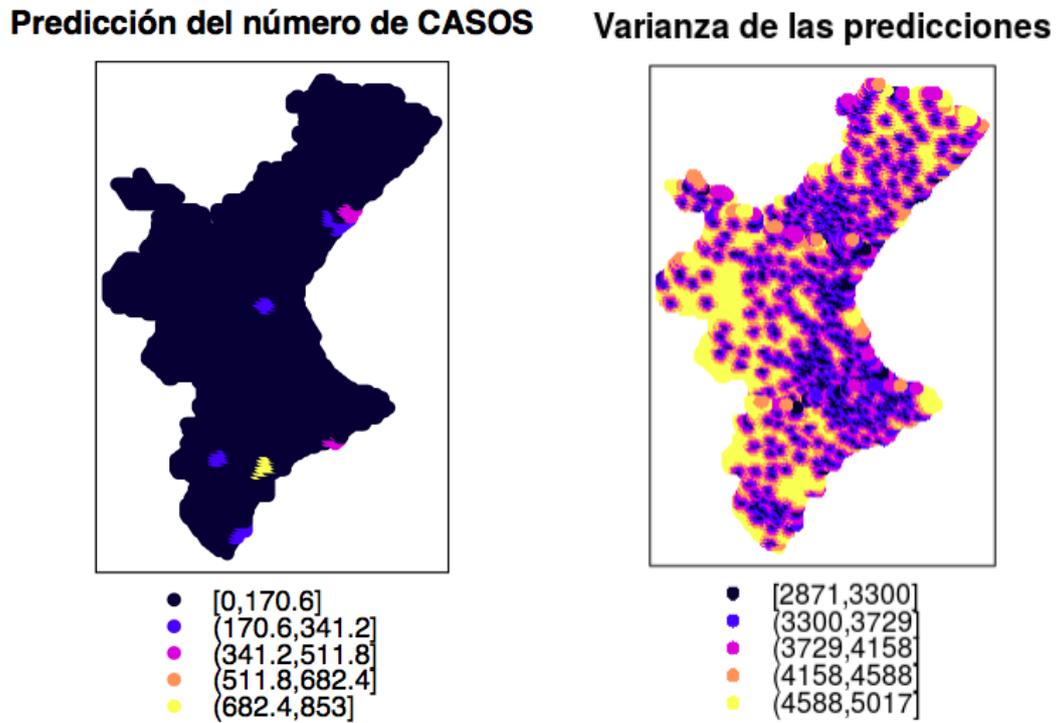
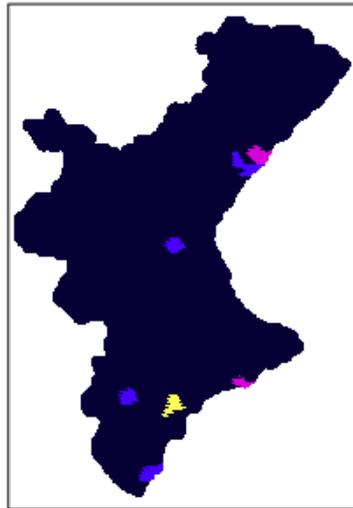


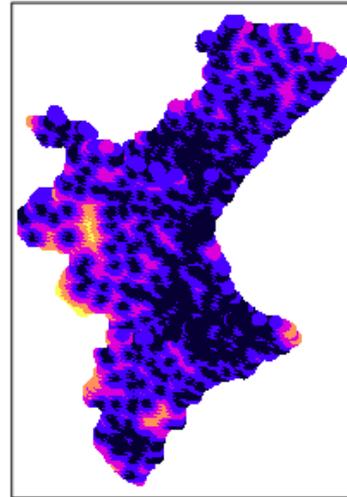
Figura 5.11: predicciones y varianzas obtenidas a partir del variograma Esférico para los casos de COVID-19 del día 3 de junio de 2020.

Predicción del número de CASOS



- [0,170.6]
- (170.6,341.2]
- (341.2,511.8]
- (511.8,682.4]
- (682.4,853]

Varianza de las predicciones



- [3705,3916]
- (3916,4126]
- (4126,4337]
- (4337,4547]
- (4547,4757]

Figura 5.12: predicciones y varianzas obtenidas a partir del variograma Exponencial para los casos de COVID-19 del día 3 de junio de 2020.

Predicción del número de CASOS Varianza de las predicciones

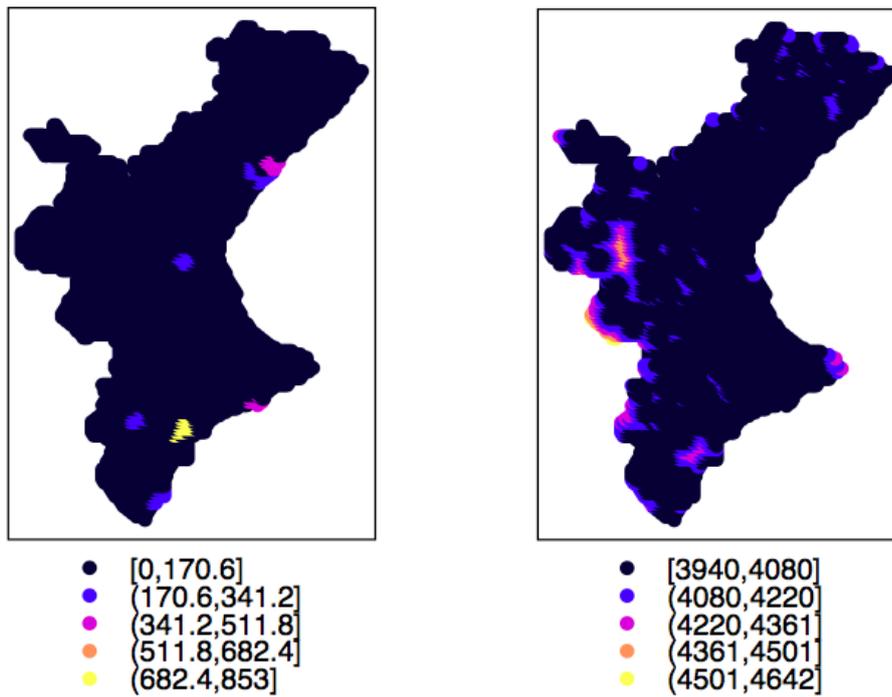


Figura 5.13: predicciones y varianzas obtenidas a partir del variograma Gaussiano para los casos de COVID-19 del día 3 de junio de 2020.

Capítulo 6

Conclusiones

En el presente estudio se ha utilizado el Kriging Simple con distintos modelos de variograma para predecir los casos de COVID-19 en situaciones en las que no disponemos de los datos. El uso adecuado de esta herramienta nos permite recabar una gran cantidad de información como puede ser la varianza en las predicciones y el número de casos en todo el territorio de estudio. Sin mencionar que podemos ver mediante un variograma qué método se ajusta mejor a los datos y así tener predicciones más fiables.

Para las predicciones hechas sobre el mapa de España para los datos del día 18 de abril de 2020, se ve que el modelo que más se ajusta a su variograma es el modelo Esférico. Los modelos Exponencial y Gaussiano no tienen un ajuste tan bueno a su variograma. En cambio, en los mapas de las predicciones vemos que los modelos Esférico y Gaussiano son bastantes similares, ambos nos muestran que las zonas de Madrid y Cataluña tendrán mayor número de casos de COVID-19 que el resto de la Península. En cambio, si se utiliza el modelo Exponencial, las predicciones nos dirán que la parte Central y Este de la península concentrarán mayor número de casos.

Por otra parte, para las predicciones hechas sobre el mapa de España para los datos del día 3 de junio de 2020, se ve que los tres Kriging que se han realizado previamente tenían variogramas muy similares, por lo tanto, desde el punto de vista del variograma los tres modelos serían igual de válidos. En los mapas obtenidos, los modelos Esférico y Exponencial son sensibles a la variabilidad de la muestra, y por tanto, las zonas de Barcelona y Madrid son las que presentan predicciones mayores. Por otra parte, el mapa obtenido mediante el modelo Gaussiano, no tiene sensibilidad ante cambios en la muestra y sólo predice más casos en la ciudad de Madrid.

En el caso de las predicciones hechas para el mapa de la Comunidad Valenciana, los variogramas obtenidos muestran que los modelos que más se ajustan al variograma empírico serían el Exponencial y Gaussiano. Sin embargo, en los mapas de las predicciones vemos que los tres modelos utilizados presentan las mismas predicciones.

Hemos visto que el método Kriging ha sido bastante efectivo a la hora de ver cómo se distribuyen los casos de COVID-19. Además, hemos comparado distintas técnicas para hacer el Kriging viendo que el modelo Esférico, en la mayoría de veces, hace mejores predicciones que los métodos Exponencial y Gaussiano. Asimismo, hemos visto que la elección del variograma depende de la forma de la curva. Los variogramas Esféricos y Exponenciales son lineales en el origen. El variograma Exponencial alcanza el umbral asintóticamente. El variograma Gaussiano es parabólico en el origen y llega al umbral asintóticamente.

Me gustaría señalar que el trabajo aquí realizado trata con datos de actualidad, los cuales pueden ser utilizados para saber donde delimitan las zonas con más casos de contagiados, y así disponer de más información a la hora de tomar decisiones de gran importancia, como pueden ser el abastecimiento de material sanitario, o saber de donde proviene el último foco de contagio y limitar su expansión.

Finalmente, este trabajo abre puertas a investigaciones futuras en el campo de la geoestadística, dejando abierta como líneas de investigación futura estudios de Kriging Multivariante, Cokriging o métodos de Kriging espaciotemporales. De esta manera se obtendrían unos resultados más precisos que con el método de Kriging Simple.

Referencias

- [1] Loquin, K., Dubois, D., (2010). Kriging with Ill-known variogram and data. *In Proceedings of the 4th international conference on Scalable uncertainty management (SUM'10)*. Springer-Verlag, Berlin, Heidelberg, 219–235. https://doi.org/10.1007/978-3-642-15951-0_23
- [2] Ortiz, J., Emery, X. (2006). Geostatistical estimation of mineral resources with soft geological boundaries: A comparative study. *Journal of the Southern African Institute of Mining and Metallurgy*. 106.
- [3] Antunes, M., Albuquerque, MTD. (2012). Using indicator kriging for the evaluation of arsenic potential contamination in an abandoned mining area (Portugal). *Science of The Total Environment*. 442. 545–552. [10.1016/j.scitotenv.2012.10.010](https://doi.org/10.1016/j.scitotenv.2012.10.010).
- [4] Reguzzoni, M., Sansó, F., Venuti, G. (2005). The theory of general kriging, with applications to the determination of a local geoid. *Geophysical Journal International*. 162. 303–314, <https://doi.org/10.1111/j.1365-246X.2005.02662.x>
- [5] Cantet, P. (2017). Mapping the mean monthly precipitation of a small island using kriging with external drifts. *Theor Appl Climatol*. 127, 31–44 . <https://doi.org/10.1007/s00704-015-1610-z>
- [6] Wu, H., Wang, X., Xue, M., Xue, M., Wu, C., Lu, Q., Ding, Z., Xv, X., Lin, J. (2017). Spatial characteristics and the epidemiology of human infections with avian influenza A(H7N9) virus in five waves from 2013 to 2017 in Zhejiang Province, China. *PLOS ONE*. 12. [10.1371/journal.pone.0180763](https://doi.org/10.1371/journal.pone.0180763)
- [7] Forna, A., Dorigatti, I., Nouvellet, P., Donnelly, C. (2020). Spatiotemporal variability in case fatality ratios for 2013–2016 Ebola epidemic in West Africa. *International Journal of Infectious Diseases*. 93. [10.1016/j.ijid.2020.01.046](https://doi.org/10.1016/j.ijid.2020.01.046).
- [8] Juan Verdoy, P. (2005). Técnicas geoestadísticas espectrales. Análisis de la estacionariedad e independencia. *Tesis doctoral Universitat Jaume I*.

- [9] Orshansky, M., Nassif, S., Boning, D. (2008). Design for manufacturability and statistical design: A constructive approach. *Design for Manufacturability and Statistical Design: A Constructive Approach*. 1-316. 10.1007/978-0-387-69011-7.
- [10] Ovalles, F., (1993). Variabilidad interna de unidades de fertilidad en suelos de la depresión del lago de Valencia. *Agronomía Tropical* 41 (1-2), 5-21.
- [11] Paz-González, A., Vieira, S.R., Taboada, M., (2000). The effect of cultivation on the spatial variability of selected properties of an umbric horizon. *Geoderma* 97, 273-292.
- [12] Camana, F., Deutsch, C. V., (2019). The Nugget Effect. *J.L. Deutsch (Ed.), Geostatistics Lessons*. <http://www.geostatisticslessons.com/lessons/nuggeteffect>
- [13] Schoenberg, I.J. (1938). Metric Spaces and Positive Definite Functions. *Transactions of the American Mathematical Society* Vol. 44, No. 3, pp. 522-536. 10.2307/1989894
- [14] Journel, A. G., Huijbregts, Ch. J. (2003). Mining Geostatistics. *Academic press, New York*.
- [15] Kitanidis, P.K. (1983). Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research*. <https://doi.org/10.1029/WR019i004p00909>
- [16] Mardia, K., Marshall, R. (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrika*, 71(1), 135-146. doi:10.2307/2336405
- [17] Patterson, H., Thompson, R. (1971). Recovery of Inter-Block Information when Block Sizes are Unequal. *Biometrika*, 58(3), 545-554. doi:10.2307/2334389
- [18] Journel, A.G. (1983). Non-parametric estimation of spatial distributions. *Mathematical Geology*, 15, 445-468
- [19] Yakowitz, S.J., Szidarovsk, F. (1985). A comparison of kriging with nonparametric regression methods. *Journal of Multivariate Analysis*, 16,21-53. [https://doi.org/10.1016/0047-259X\(85\)90050-8](https://doi.org/10.1016/0047-259X(85)90050-8)
- [20] Laslett, G.M. (1994). Kriging and Splines: An Empirical Comparison of Their Predictive Performance in Some Applications. *Journal of American Statistical Association*. 89, 391-400. 10.2307/2290837
- [21] Stein, A. y Corsten, L. (1991). Universal kriging and cokriging as a regression procedure. *Biometrics*, 47, 575-587. 10.2307/2532147

- [22] Deutsch, C.V. (1991). The relationship between universal kriging, kriging with an external drift, and cokriging. *Stanford Center of Reservoir Forecasting (SCRF)*, 4
- [23] Goovaerts, P. (1998). Ordinary cokriging revisited. *Mathematical Geology*, 30, 21-42.
<https://doi.org/10.1023/A:1021757104135>
- [24] <https://cnecovid.isciii.es/covid19/documentación-y-datos>
- [25] <https://icvgva.maps.arcgis.com/apps/opdashboard/index.html/3a3115ad642a4516b0928f21e395b32d>
- [26] <https://public.opendatasoft.com/explore/dataset/provincias-espanolas/table/?sort=provincia>
- [27] https://gadm.org/download_country_v3.html

Anexo A

Aplicación del Kriging Simple a los casos de Covid-19 en España

A continuación se aplicará el Kriging Simple a los casos de COVID-19 existentes los días 18 de abril y 3 de junio de 2020. El número de casos se han agrupado por provincias.

```
# Cargamos los paquetes necesarios
install.packages("sp")
install.packages("gstat")
install.packages("readr")
install.packages("kridge")
install.packages("kriging")
library(sp)
library(gstat)
library(readr)
```

Los datos se han obtenido de [<https://cnecovid.isciii.es/covid19/#documentación-y-datos>] para ambas fechas.

A.1. Kriging para los datos del 3 de junio de 2020

Los primeros datos son:

```
# cargamos el archivo donde se encuentran los datos de casos de Covid19:
puntos <- read_table2("casosProv.csv")
```

```
## Parsed with column specification:
```

```
## cols(
##   provincia = col_character(),
##   fecha = col_date(format = ""),
##   CASOS = col_double(),
##   Longitud = col_double(),
##   Latitud = col_double()
## )

## Warning: 24 parsing failures.
## row col expected actual file
## 2 -- 5 columns 6 columns 'casosProv.csv'
## 5 -- 5 columns 6 columns 'casosProv.csv'
## 6 -- 5 columns 6 columns 'casosProv.csv'
## 7 -- 5 columns 6 columns 'casosProv.csv'
## 9 -- 5 columns 6 columns 'casosProv.csv'
## ... ..
## See problems(...) for more details.
```

```
head(puntos)
```

```
## # A tibble: 6 x 5
##   provincia fecha      CASOS Longitud Latitud
##   <chr>      <date>    <dbl>   <dbl>   <dbl>
## 1 Alicante 2020-06-03     1    38.3   -0.481
## 2 Albacete 2020-06-03     1    39.0   -1.86
## 3 Almeria   2020-06-03     2    36.8   -2.47
## 4 Avila     2020-06-03     1    40.7   -4.70
## 5 Barcelona 2020-06-03    79    41.4    2.17
## 6 Badajoz   2020-06-03     3    38.9   -6.97
```

La variable dependiente: número de casos

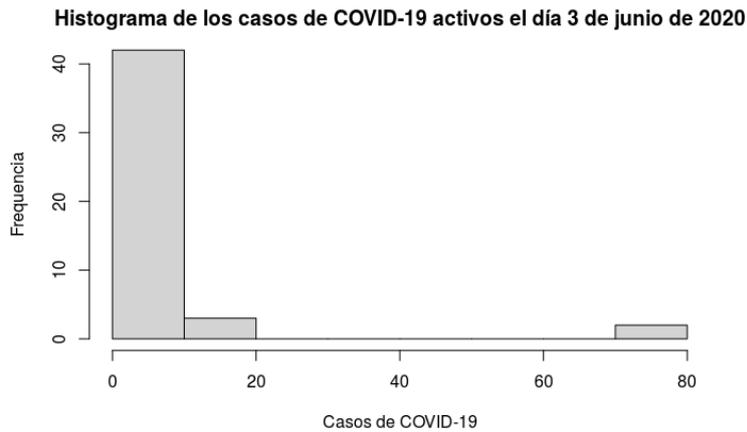
Veamos su distribución resumida en estadísticos, y en un histograma:

```
summary(puntos$CASOS)
```

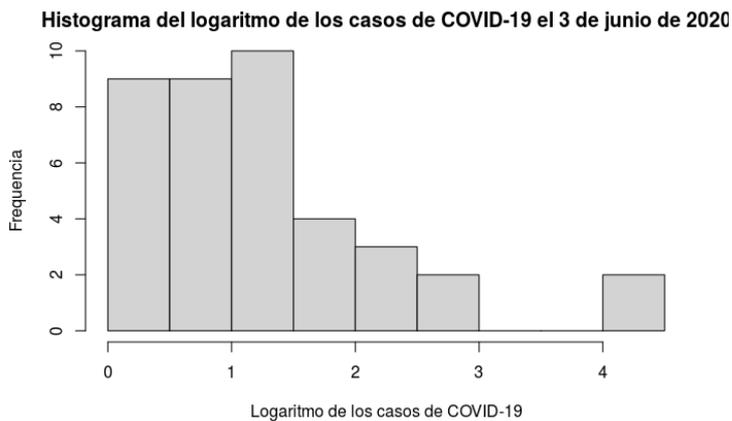
```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.000  1.000   2.000   6.213  3.500  79.000
```

```
Casos= puntos$CASOS
```

```
hist(Casos,
     main = "Histograma de los casos de COVID-19 activos el día 3 de junio de 2020",
     xlab = 'Casos de COVID-19',
     ylab = 'Frecuencia')
```



```
hist(log(Casos),
     main = "Histograma del logaritmo de los casos de COVID-19 el"
         "3 de junio de 2020",
     xlab = 'Logaritmo de los casos de COVID-19',
     ylab = 'Frecuencia')
```



En este caso no es necesario aplicar transformaciones logarítmicas de los datos dado que no hay diferencias significativas entre datos de distintas provincias

A.1.1. Los datos en su ubicación espacial

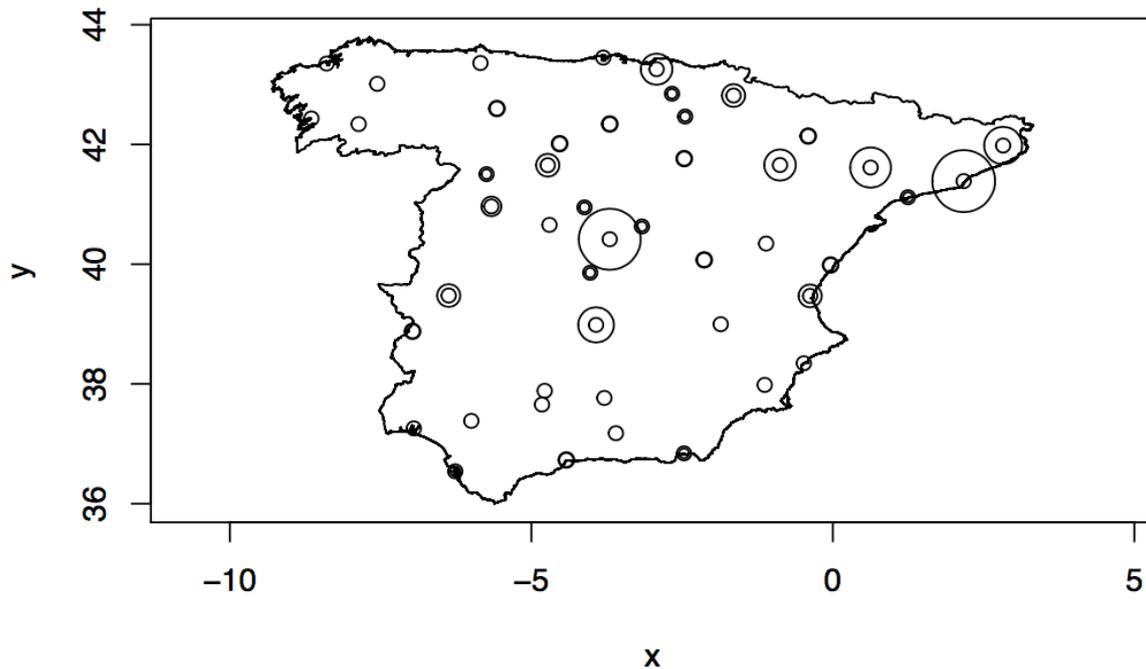
Representamos el mapa de España y las localizaciones de los datos, con puntos de diámetro proporcional al número de casos, que es nuestra variable dependiente:

```

gadm36_ESP_0_sp<-readRDS("gadm36_ESP_0_sp.rds")
peninsula <- gadm36_ESP_0_sp@polygons[[1]]@Polygons[[178]]@coords

#print(peninsula)
plot(peninsula, type='l', asp=1)
points(puntos[c('Latitud', 'Longitud')])
points(puntos[c('Latitud', 'Longitud')], cex=log(puntos$CASOS))

```



```

#print(peninsula)
#print(peninsula[1,])

```

Como se puede ver en el mapa los datos estan centrados en las capitales de provincia, dado que es donde se concentra la mayor parte de la población.

A.1.2. Variograma

Primero pasamos de hoja de datos “normal” a un objeto espacial para poder trabajar estas técnicas.

```
class(puntos)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"        "data.frame"
```

```
coordinates(puntos) <- ~ Latitud + Longitud
class(puntos)
```

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

```
str(puntos)
```

```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
## ..@ data      : tibble [47 x 3] (S3: tbl_df/tbl/data.frame)
## .. ..$ provincia: chr [1:47] "Alicante" "Albacete" "Almeria" "Avila" ...
## .. ..$ fecha    : Date[1:47], format: "2020-06-03" "2020-06-03" ...
## .. ..$ CASOS    : num [1:47] 1 1 2 1 79 3 9 3 1 2 ...
## .. ..- attr(*, "problems")= tibble [24 x 5] (S3: tbl_df/tbl/data.frame)
## .. .. ..$ row    : int [1:24] 2 5 6 7 9 10 11 15 17 24 ...
## .. .. ..$ col    : chr [1:24] NA NA NA NA ...
## .. .. ..$ expected: chr [1:24] "5 columns" "5 columns" "5 columns" "5 columns" ...
## .. .. ..$ actual  : chr [1:24] "6 columns" "6 columns" "6 columns" "6 columns" ...
## .. .. ..$ file    : chr [1:24] "'casosProv.csv'" "'casosProv.csv'" "'casosProv.csv'" "'c
## ..@ coords.nrs : int [1:2] 5 4
## ..@ coords     : num [1:47, 1:2] -0.481 -1.86 -2.468 -4.7 2.17 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:47] "1" "2" "3" "4" ...
## .. .. ..$ : chr [1:2] "Latitud" "Longitud"
## ..@ bbox       : num [1:2, 1:2] -8.65 36.54 2.82 43.45
## .. ..- attr(*, "dimnames")=List of 2
## .. .. ..$ : chr [1:2] "Latitud" "Longitud"
## .. .. ..$ : chr [1:2] "min" "max"
## ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
## .. .. ..@ projargs: chr NA
```

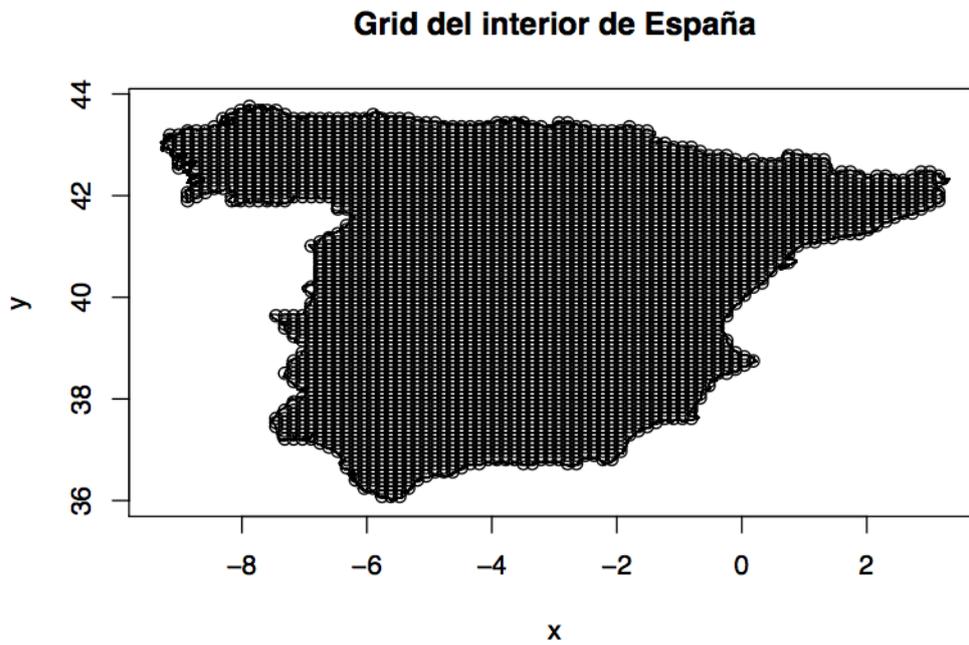
A.1.3. Interpolación mediante Kriging simple

Creamos un grid del interior de España donde haremos las predicciones del número de casos. Lógicamente habría que tener en cuenta la densidad de población, que por ahora supondremos uniformemente distribuida.

```
# crear un grid de puntos dentro del mapa de España
tam = 100
# esto crea la malla en un rectángulo que contiene España
penin.grid = expand.grid(x=seq(fr=-10, to=4, len=tam), y=seq(fr=36, to=44, len=tam))
# esto da TRUE o FALSE si cada punto de la malla está dentro o fuera
```

```
dentro = point.in.polygon(penin.grid$x, penin.grid$y, peninsula[,1], peninsula[,2])
penin.grid2 = penin.grid[dentro>0,] # esto elige los de dentro solo

# visualizarlo en el mapa
plot(peninsula, type='l', main='Grid del interior de España')
points(penin.grid2)
```



A.1.4. Probamos distintos tipos de modelos

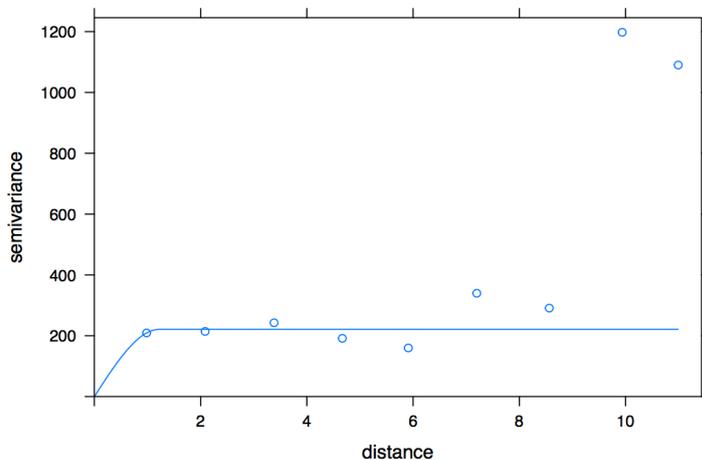
Esférico

Vemos la variabilidad de los datos en el espacio con el variograma:

```
vario = variogram(CASOS~1, puntos,cutoff=20)
vario.fit <- fit.variogram(vario, model=vgm(NA,"Sph",NA),fit.sills=NA, fit.ranges=TRUE)
print(vario.fit)
```

```
## model psill range
## 1 Sph 221.1759 1.223572
```

```
plot(vario, vario.fit)
```



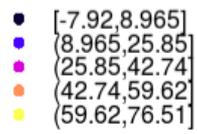
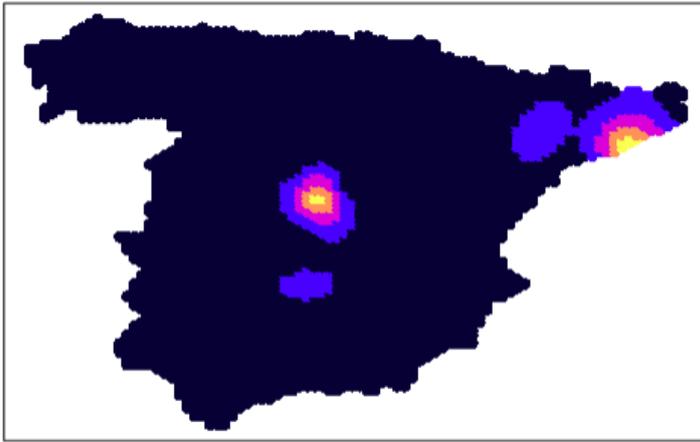
Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```
peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el Kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)
```

```
## [using ordinary kriging]
```

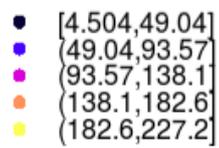
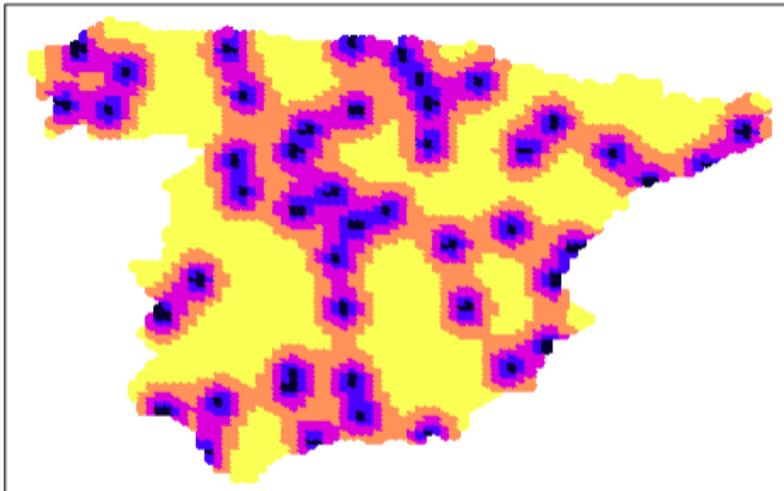
```
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
splot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

Predicción del número de CASOS



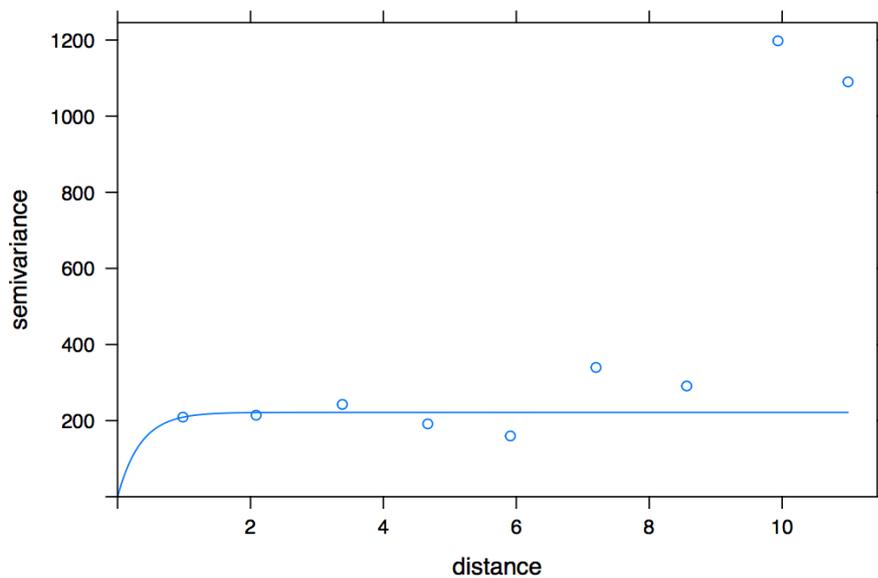
```
splot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



Exponencial

```
vario = variogram(CASOS~1, puntos,cutoff=20)
vario.fit <- fit.variogram(vario, model=vgm(NA,"Exp",NA),fit.sills=TRUE, fit.ranges=TRUE)
plot(vario, vario.fit)
```

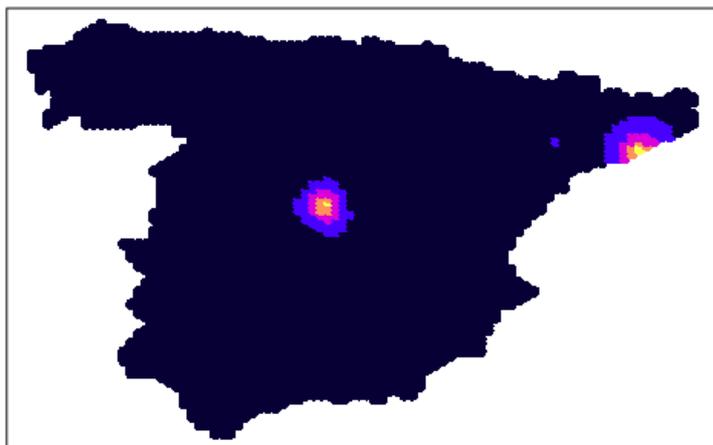


Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el kriging.

```
peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)

## [using ordinary kriging]
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
splot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

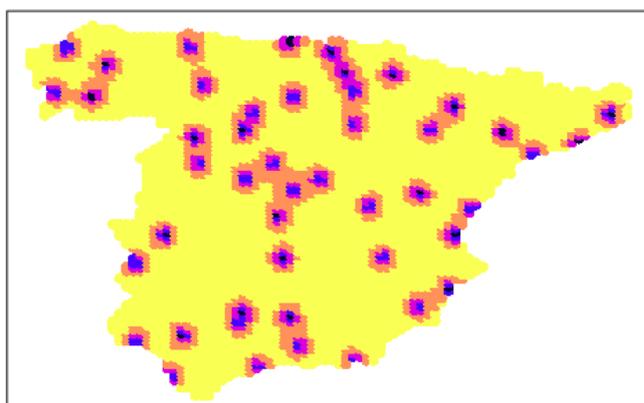
Predicción del número de CASOS



- [0.4219,15.02]
- (15.02,29.61]
- (29.61,44.2]
- (44.2,58.8]
- (58.8,73.39]

```
spplot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



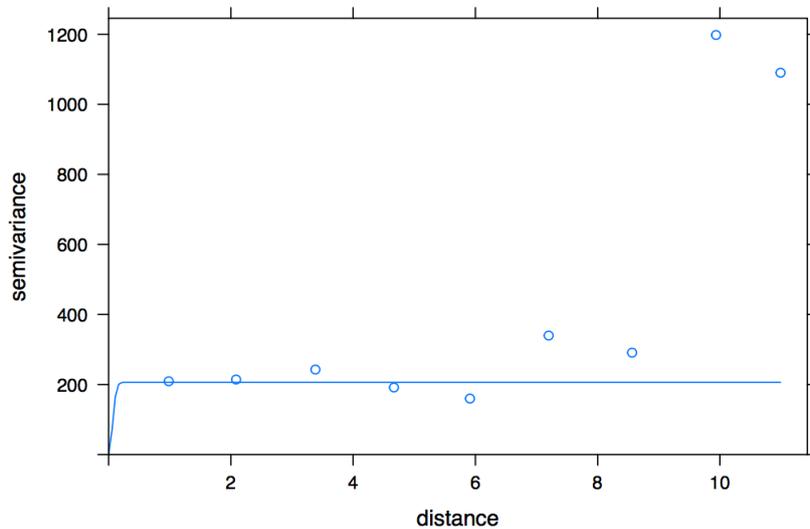
- [10.48,53.77]
- (53.77,97.06]
- (97.06,140.4]
- (140.4,183.6]
- (183.6,226.9]

Gaussiano:

```
vario = variogram(CASOS~1, puntos,cutoff=20)  
vario.fit <- fit.variogram(vario, model=vgm(NA,"Gau",NA),fit.sills=TRUE, fit.ranges=TRUE)
```

```
## Warning in fit.variogram(vario, model = vgm(NA, "Gau", NA), fit.sills = TRUE, :
## No convergence after 200 iterations: try different initial values?
```

```
plot(vario, vario.fit)
```



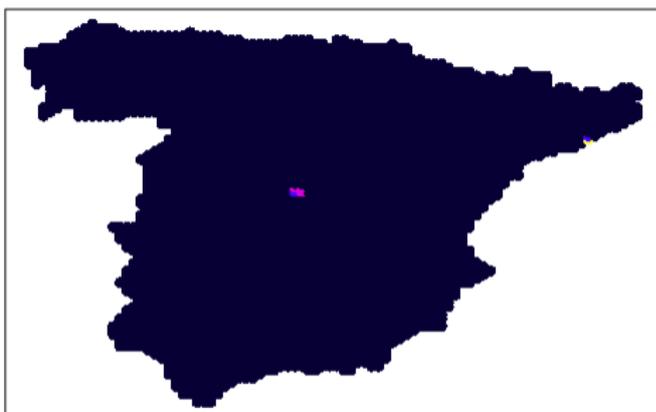
Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el kriging.

```
peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)
```

```
## [using ordinary kriging]
```

```
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
spplot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

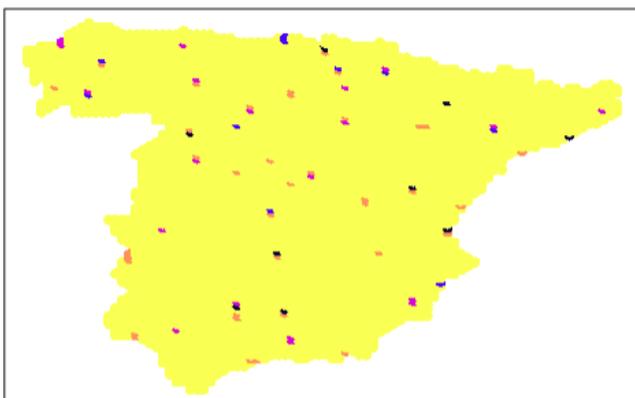
Predicción del número de CASOS



● [0.5103,14.8]
● (14.8,29.09]
● (29.09,43.38]
● (43.38,57.67]
● (57.67,71.96]

```
spplot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



● [3.831,45.15]
● (45.15,86.46]
● (86.46,127.8]
● (127.8,169.1]
● (169.1,210.4]

A.2. Kriging para los datos del 18 de abril de 2020

Los primeros datos son:

```
# cargamos el archivo donde se encuentran los datos de casos de Covid19:  
puntos <- read_table2("provincias18abr.csv")
```

```
## Parsed with column specification:
## cols(
##   provincia = col_character(),
##   fecha = col_date(format = ""),
##   CASOS = col_double(),
##   Longitud = col_double(),
##   Latitud = col_double()
## )

## Warning: 1 parsing failure.
## row col expected actual file
## 15 -- 5 columns 6 columns 'provincias18abr.csv'
```

```
head(puntos)
```

```
## # A tibble: 6 x 5
##   provincia fecha      CASOS Longitud Latitud
##   <chr>      <date>    <dbl>   <dbl>   <dbl>
## 1 A          2020-04-18    73     38.3   -0.481
## 2 AB         2020-04-18    35     39.0   -1.86
## 3 AL         2020-04-18     7     36.8   -2.47
## 4 AV         2020-04-18    15     40.7   -4.70
## 5 B          2020-04-18   373     41.4    2.17
## 6 BA         2020-04-18    28     38.9   -6.97
```

La variable dependiente: número de casos

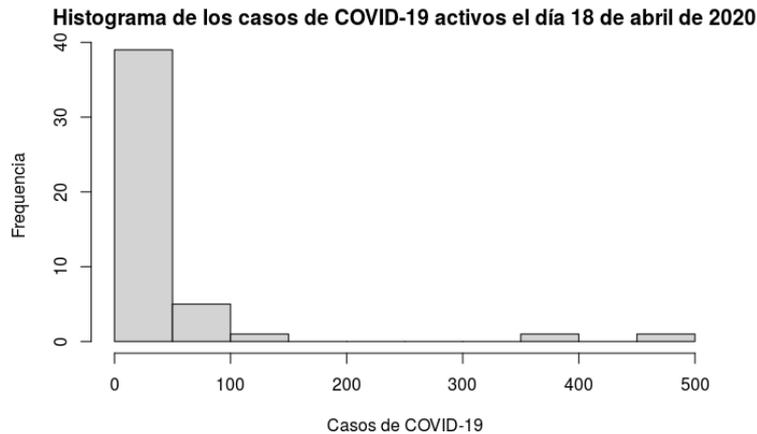
Veamos su distribución resumida en estadísticos, y en un histograma:

```
summary(puntos$CASOS)
```

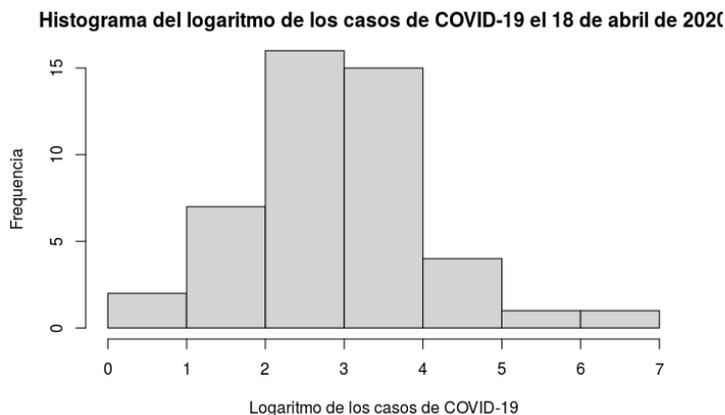
```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.00   9.50   17.00   42.09  40.50  467.00
```

```
Casos= puntos$CASOS
```

```
hist(Casos,
     main = "Histograma de los casos de COVID-19 activos el día 18 de abril de 2020",
     xlab = 'Casos de COVID-19',
     ylab = 'Frecuencia')
```



```
hist(log(Casos),
     main = "Histograma del logaritmo de los casos de COVID-19 el 18 de abril de 2020",
     xlab = 'Logaritmo de los casos de COVID-19',
     ylab = 'Frecuencia')
```



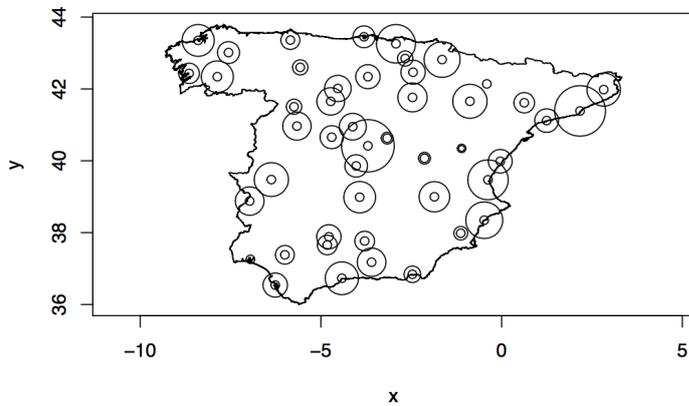
Igual que ocurría anteriormente, en este caso no es necesario aplicar transformaciones logarítmicas de los datos dado que no hay diferencias significativas entre datos de distintas provincias.

A.2.1. Los datos en su ubicación espacial

Representamos el mapa de España y las localizaciones de los datos, con puntos de diámetro proporcional al número de casos, que es nuestra variable dependiente:

```
gadm36_ESP_0_sp<-readRDS("gadm36_ESP_0_sp.rds")
peninsula <- gadm36_ESP_0_sp@polygons[[1]]@Polygons[[178]]@coords
plot(peninsula, type='l', asp=1)
points(puntos[c('Latitud', 'Longitud')])
```

```
points(puntos[c('Latitud', 'Longitud')], cex=log(puntos$CASOS))
```



A.2.2. Variograma

Primero pasamos de hoja de datos “normal” a un objeto espacial para poder trabajar estas técnicas.

```
class(puntos)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"        "data.frame"
```

```
coordinates(puntos) <- ~ Latitud + Longitud
class(puntos)
```

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

```
str(puntos)
```

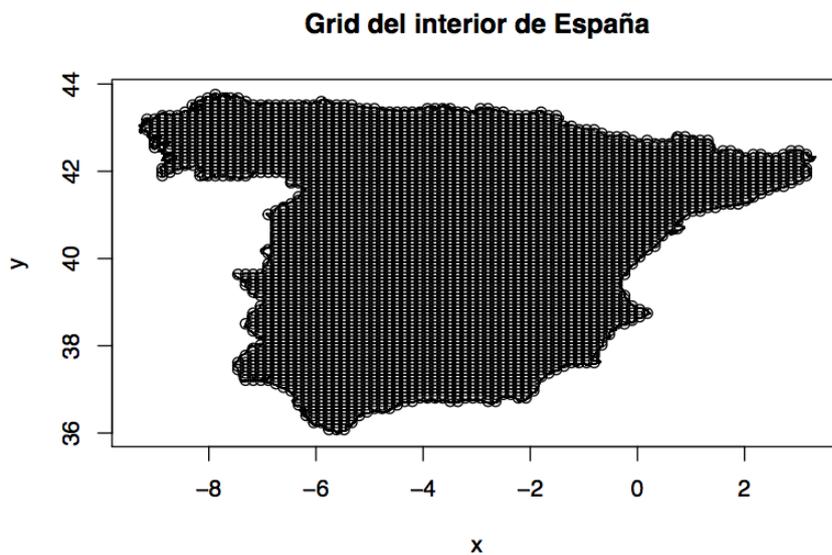
```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots
## ..@ data      : tibble [47 x 3] (S3: tbl_df/tbl/data.frame)
## .. ..$ provincia: chr [1:47] "A" "AB" "AL" "AV" ...
## .. ..$ fecha    : Date[1:47], format: "2020-04-18" "2020-04-18" ...
## .. ..$ CASOS    : num [1:47] 73 35 7 15 373 28 92 17 42 17 ...
## .. ..- attr(*, "problems")= tibble [1 x 5] (S3: tbl_df/tbl/data.frame)
## .. .. ..$ row    : int 15
## .. .. ..$ col    : chr NA
## .. .. ..$ expected: chr "5 columns"
## .. .. ..$ actual  : chr "6 columns"
```

```
## .. .. .$ file      : chr "'provincias18abr.csv'"
## ..@ coords.nrs     : int [1:2] 5 4
## ..@ coords        : num [1:47, 1:2] -0.481 -1.86 -2.468 -4.7 2.17 ...
## .. ..- attr(*, "dimnames")=List of 2
## .. .. .$          : chr [1:47] "1" "2" "3" "4" ...
## .. .. .$          : chr [1:2] "Latitud" "Longitud"
## ..@ bbox          : num [1:2, 1:2] -8.65 36.54 2.82 43.45
## .. ..- attr(*, "dimnames")=List of 2
## .. .. .$          : chr [1:2] "Latitud" "Longitud"
## .. .. .$          : chr [1:2] "min" "max"
## ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot
## .. .. ..@ projargs: chr NA
```

A.2.3. Vamos a interpolar mediante Kriging Simple

Creamos un grid del interior de España donde haremos las predicciones del número de casos. Lógicamente habría que tener en cuenta la densidad de población, que por ahora supondremos uniformemente distribuida.

```
# crear un grid de puntos dentro del mapa de España
tam = 100
penin.grid = expand.grid(x=seq(fr=-10, to=4, len=tam), y=seq(fr=36, to=44, len=tam)) # esto da
dentro = point.in.polygon(penin.grid$x, penin.grid$y, peninsula[,1], peninsula[,2]) # esto da
penin.grid2 = penin.grid[dentro>0,] # esto elige los de dentro solo
# visualizarlo en el mapa
plot(peninsula, type='l', main='Grid del interior de España')
points(penin.grid2)
```

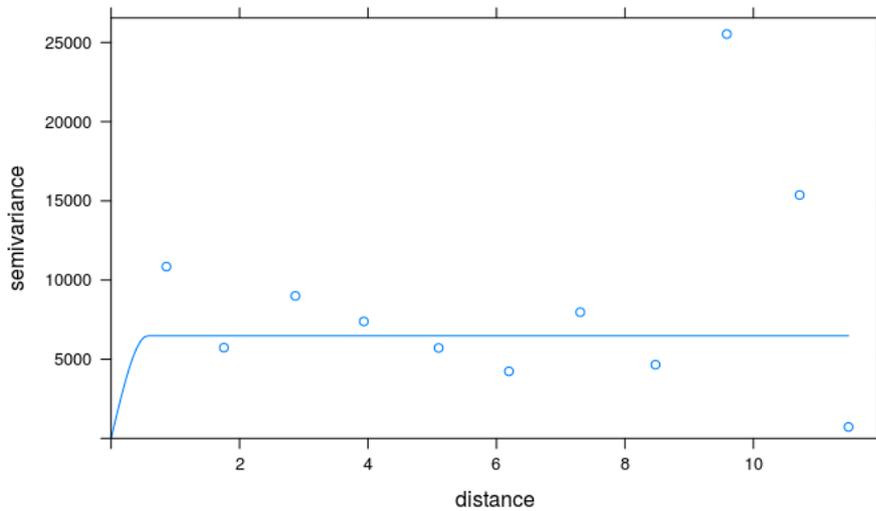


A.2.4. Probamos distintos tipos de modelos

Esférico

Vemos la variabilidad de los datos en el espacio con el variograma:

```
vario = variogram(CASOS~1, puntos,cutoff=17)
vario.fit <- fit.variogram(vario, model=vgm(NA,"Sph",NA),fit.sills=TRUE, fit.ranges=TRUE)
plot(vario, vario.fit)
```

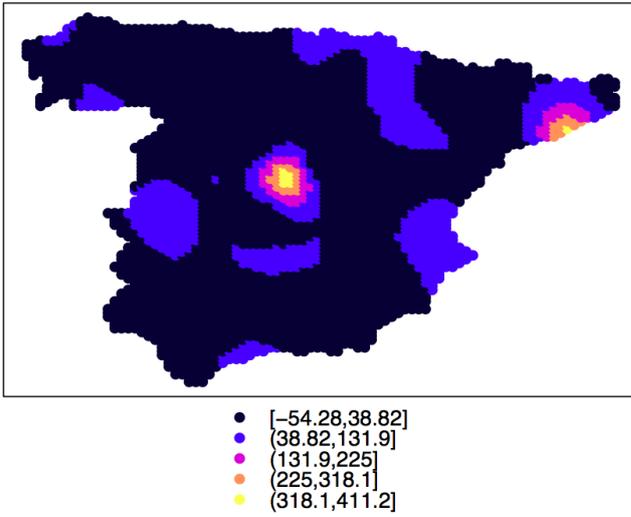


Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```
peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el Kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)

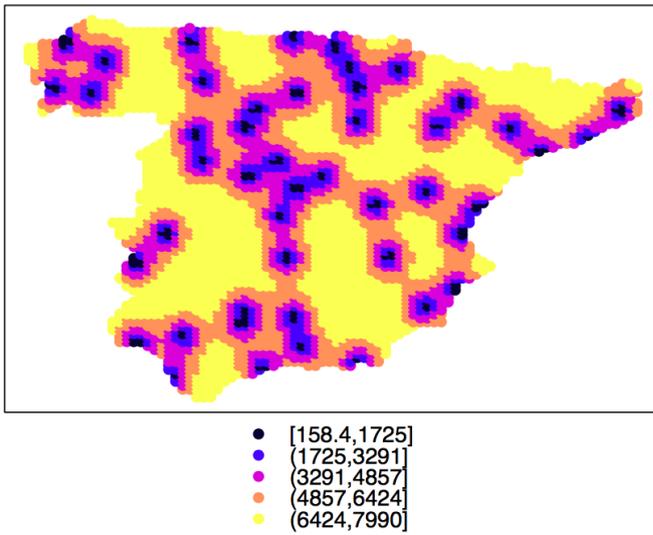
## [using ordinary kriging]
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
spplot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

Predicción del número de CASOS



```
spplot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



Exponencial

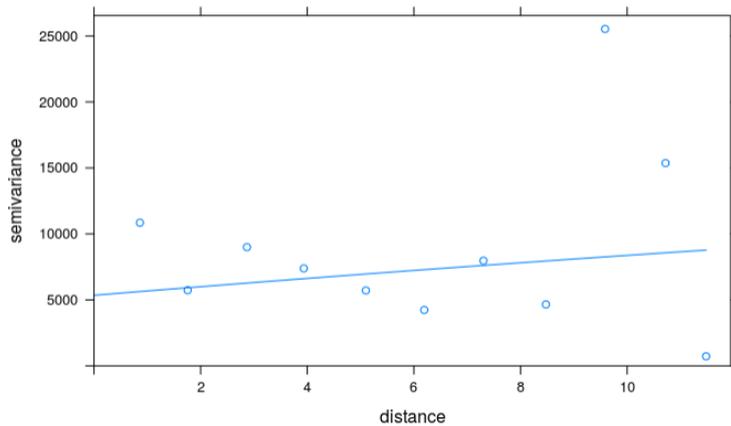
```

vario = variogram(CASOS~1, puntos,cutoff=17)
vario.fit <- fit.variogram(vario, model=vgm(psill = 5000, model = "Exp", range = 10, nugget =

## Warning in fit.variogram(vario, model = vgm(psill = 5000, model = "Exp", : No
## convergence after 200 iterations: try different initial values?

plot(vario,vario.fit)

```



Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```

peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)

```

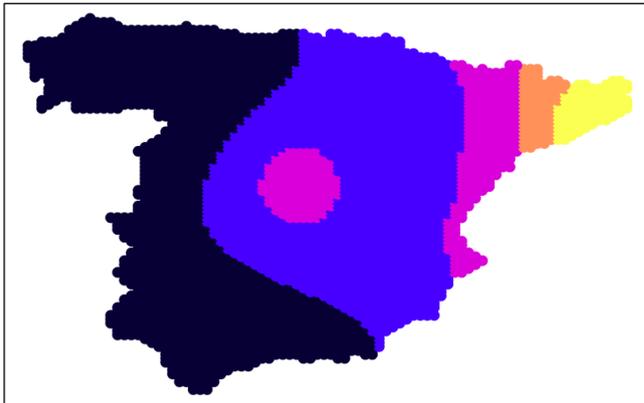
```
## [using ordinary kriging]
```

```

# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
spplot(res[c('var1.pred')], main='Predicción del número de CASOS')

```

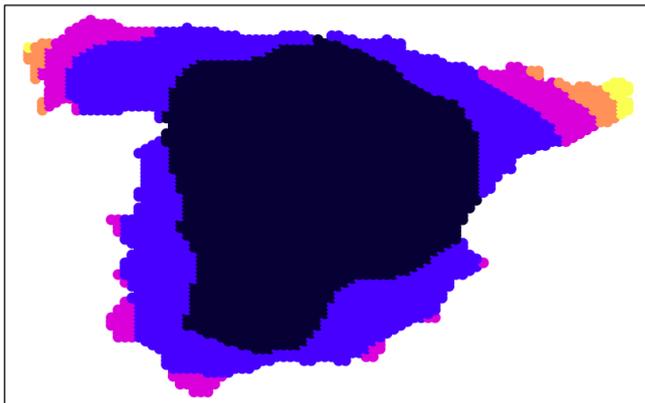
Predicción del número de CASOS



- [20.46,36.73]
- (36.73,53.01]
- (53.01,69.29]
- (69.29,85.56]
- (85.56,101.8]

```
splot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



- [5786,6004]
- (6004,6222]
- (6222,6440]
- (6440,6658]
- (6658,6876]

Gaussiano:

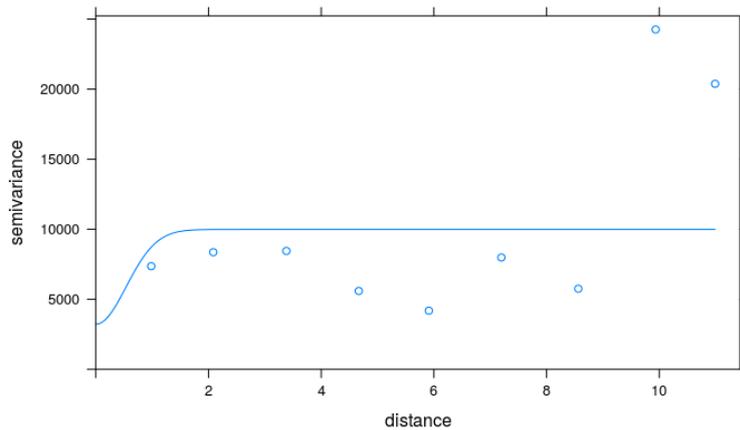
```

vario = variogram(CASOS~1, puntos,cutoff=20)
vario.fit <- fit.variogram(vario, model=vgm(psill = 5000, model = "Gau", range = 1, nugget =

## Warning in fit.variogram(vario, model = vgm(psill = 5000, model = "Gau", : No
## convergence after 200 iterations: try different initial values?

plot(vario,vario.fit)

```



Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

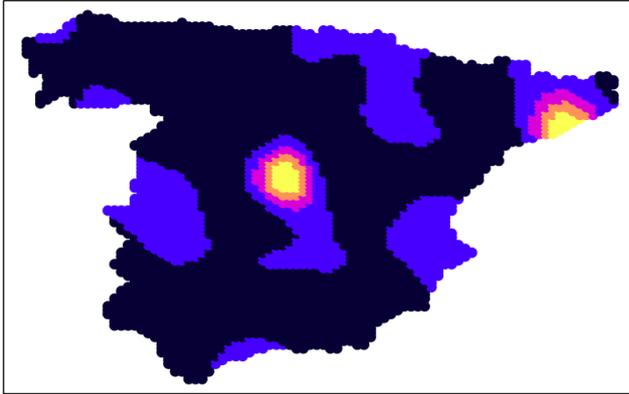
```

peningrid = as.data.frame(penin.grid2) # hace falta que sea data frame
names(peningrid) = c('Longitud', 'Latitud')
coordinates(peningrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=CASOS ~ 1, locations=puntos, newdata=peningrid, model=vario.fit)

## [using ordinary kriging]
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
spplot(res[c('var1.pred')], main='Predicción del número de CASOS')

```

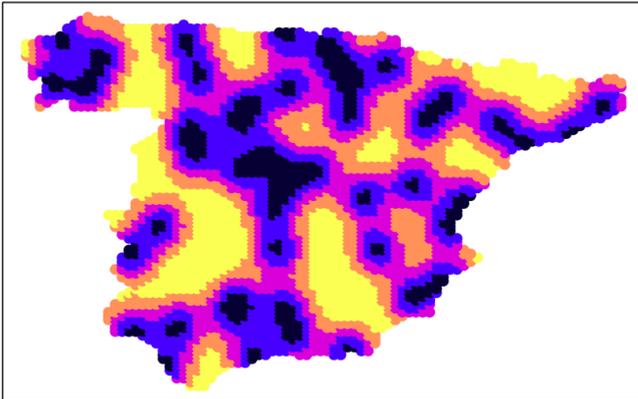
Predicción del número de CASOS



- [-18.72,36.77]
- (36.77,92.26]
- (92.26,147.7]
- (147.7,203.2]
- (203.2,258.7]

```
spplot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



- [4578,5712]
- (5712,6846]
- (6846,7980]
- (7980,9113]
- (9113,1.025e+04]

Anexo B

Aplicación del Kriging Simple a los casos de Covid-19 en la Comunidad Valenciana

A continuación se aplicará el Kriging Simple a los casos de covid existentes el día 3 de junio de 2020. El número de casos se han agrupado por provincias.

```
# Cargamos los paquetes necesarios  
#install.packages("sp")  
#install.packages("gstat")  
#install.packages("readr")  
#install.packages("kridge")  
#install.packages("kriging")  
library(sp)  
library(gstat)  
library(readr)
```

Los datos se han obtenido de [<https://icvgva.maps.arcgis.com/apps/opsdashboard/index.html#/3a3115ad642a4516b0928f21e395b32d>].

B.1. Kriging para los datos del 3 de junio de 2020

Los primeros datos son:

```
# cargamos el archivo donde se encuentran los datos de casos de Covid19:
puntos <- read_csv("casosPueblosCVal.csv")
```

```
## Parsed with column specification:
## cols(
##   Poblacion = col_character(),
##   Casos_totales = col_double(),
##   Casos_en_ultimos_14dias = col_double(),
##   Fallecidos = col_double(),
##   Longitud = col_double(),
##   Latitud = col_double()
## )
```

```
head(puntos)
```

```
## # A tibble: 6 x 6
##   Poblacion      Casos_totales Casos_en_ultimos_14~ Fallecidos Longitud Latitud
##   <chr>          <dbl>          <dbl>          <dbl>    <dbl>  <dbl>
## 1 Ademuz          0              0              0      40.1  -1.22
## 2 Ador            1              0              1      38.9  -0.246
## 3 Agost           4              0              0      38.4  -0.618
## 4 Agres           0              0              0      38.8  -0.512
## 5 Agullent       11              0              3      38.8  -0.553
## 6 Aiello de Malfe~ 1              0              0      38.9  -0.602
```

La variable dependiente: número de casos

Veamos su distribución resumida en estadísticos, y en un histograma:

```
summary(puntos$Casos_totales)
```

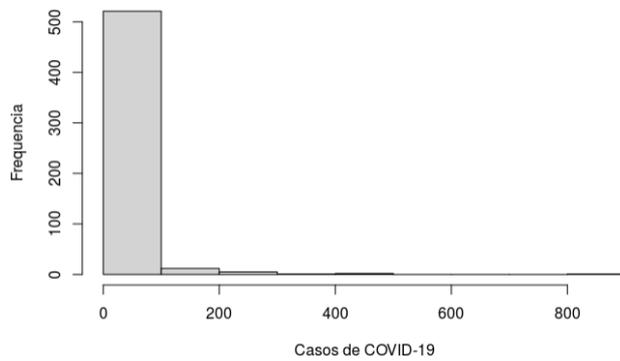
```
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.00   0.00    2.00   16.89   9.00   853.00
```

```
Casos= puntos$Casos_totales)
```

```
Casos= puntos$CASOS
```

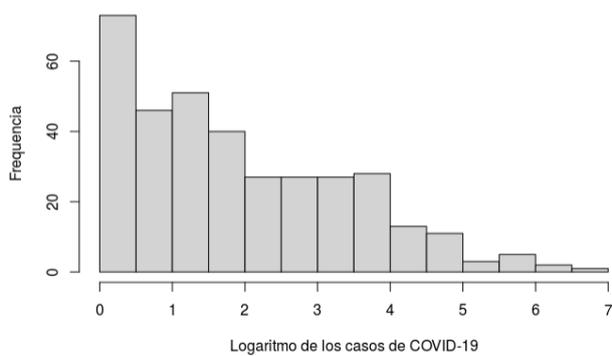
```
hist(Casos,
      main = "Histograma de los casos de COVID-19 activos el día 3 de junio de 2020",
      xlab = 'Casos de COVID-19',
      ylab = 'Frecuencia')
```

Histograma de los casos de COVID-19 activos el día 3 de junio de 2020



```
hist(log(Casos),  
     main="Histograma del logaritmo de los casos de COVID-19 el"  
     3 de junio de 2020",  
     xlab='Logaritmo de los casos de COVID-19',  
     ylab='Frecuencia')
```

Histograma del logaritmo de los casos de COVID-19 el 3 de junio de 2020



En este caso no es necesario aplicar transformaciones logarítmicas de los datos dado que no hay diferencias significativas entre datos de distintas poblaciones.

B.1.1. Los datos en su ubicación espacial

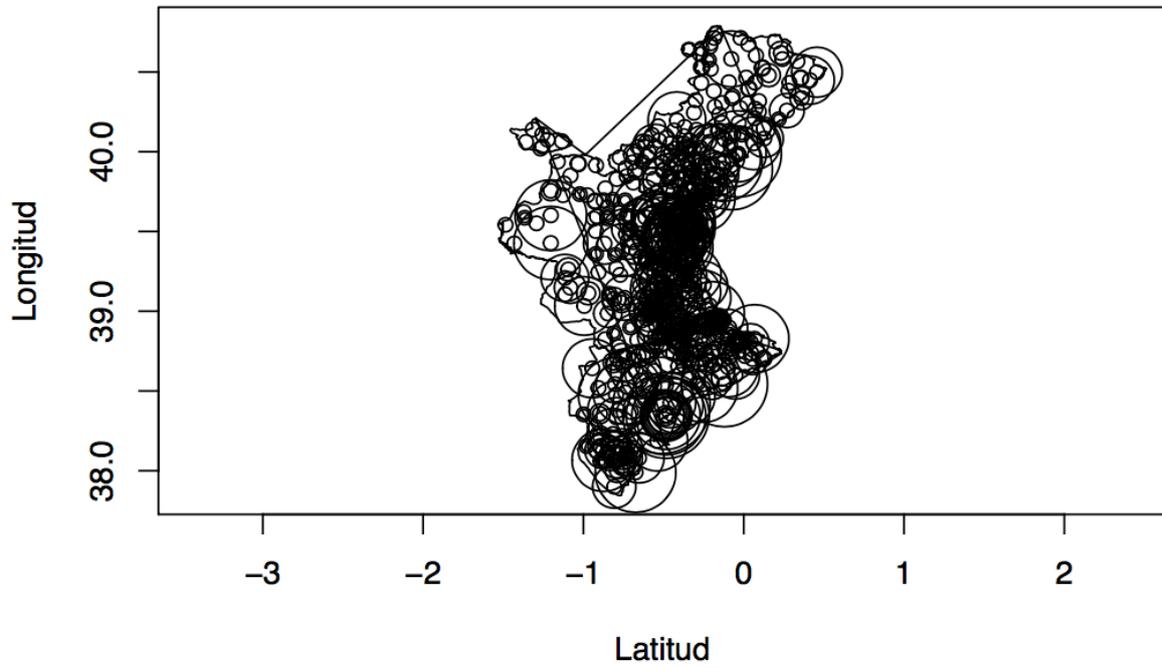
Representamos el mapa de la Comunidad Valenciana y las localizaciones de los datos, con puntos de diámetro proporcional al número de casos, que es nuestra variable dependiente:

```
cv<-read_csv("Comunidad_contorno2.csv")
```

```
## Parsed with column specification:  
## cols(  
##   Latitud = col_double(),  
##   Longitud = col_double()
```

```
## )
```

```
plot(cv, type='l', asp=1)  
points(puntos[c('Latitud', 'Longitud')])  
points(puntos[c('Latitud', 'Longitud')], cex=log(puntos$Casos_totales))
```



B.1.2. Variograma

Primero pasamos de hoja de datos “normal” a un objeto espacial para poder trabajar estas técnicas.

```
class(puntos)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"        "data.frame"
```

```
coordinates(puntos) <- ~ Latitud + Longitud  
class(puntos)
```

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

```
str(puntos)
```

```
## Formal class 'SpatialPointsDataFrame' [package "sp"] with 5 slots  
## ..@ data      : tibble [542 x 4] (S3: tbl_df/tbl/data.frame)  
## .. ..$ Poblacion      : chr [1:542] "Ademuz" "Ador" "Agost" "Agres" ...  
## .. ..$ Casos_totales  : num [1:542] 0 1 4 0 11 1 0 0 0 853 ...  
## .. ..$ Casos_en_ultimos_14dias: num [1:542] 0 0 0 0 0 0 0 0 0 11 ...  
## .. ..$ Fallecidos    : num [1:542] 0 1 0 0 3 0 0 0 0 87 ...  
## ..@ coords.nrs : int [1:2] 6 5  
## ..@ coords     : num [1:542, 1:2] -1.223 -0.246 -0.618 -0.512 -0.553 ...  
## .. ..- attr(*, "dimnames")=List of 2  
## .. .. ..$ : chr [1:542] "1" "2" "3" "4" ...  
## .. .. ..$ : chr [1:2] "Latitud" "Longitud"  
## ..@ bbox      : num [1:2, 1:2] -1.49 37.9 5.74 40.74  
## .. ..- attr(*, "dimnames")=List of 2  
## .. .. ..$ : chr [1:2] "Latitud" "Longitud"  
## .. .. ..$ : chr [1:2] "min" "max"  
## ..@ proj4string:Formal class 'CRS' [package "sp"] with 1 slot  
## .. .. ..@ projargs: chr NA
```

B.1.3. Interpolación mediante Kriging Simple

Creamos un grid del interior de la Comunidad Valenciana donde haremos las predicciones del número de casos. Lógicamente habría que tener en cuenta la densidad de población, que por ahora supondremos uniformemente distribuida.

Para la realización de este mapa se divide la Comunidad Valenciana en 4 zonas: el rincón de Ademuz y las provincias de Valencia, Alicante y Castellón. Una vez tenemos los grids de las 4 zonas los unimos.

```
# crear un grid de puntos dentro del mapa
tam = 200
# esto crea la malla en un rectángulo que contiene la Comunidad Valenciana:
com.grid = expand.grid(x=seq(fr=-1.5, to=1, len=tam), y=seq(fr=37, to=41, len=tam))
ademuz = cv[1605:1713,]
valencia = cv[765:1600,]
alicante = cv[1768:2489,]

## Warning: The `i` argument of ``[.tbl_df()]` must lie in [0, rows] if positive, as of tibble
## Use `NA` as row index to obtain a row full of `NA` values.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.

castellon = cv[0:763,]

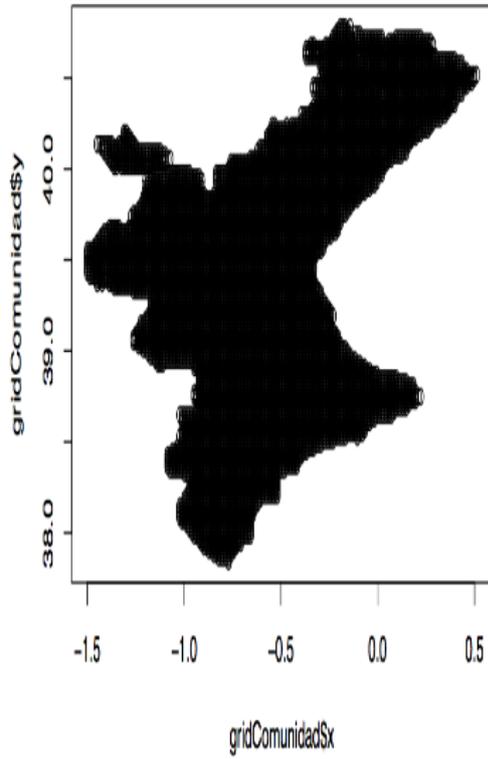
dentro_ademuz = point.in.polygon(com.grid$x, com.grid$y,
as.matrix(ademuz[,1]), as.matrix(ademuz[,2]))
dentro_valencia = point.in.polygon(com.grid$x, com.grid$y,
as.matrix(valencia[,1]), as.matrix(valencia[,2]))
dentro_alicante = point.in.polygon(com.grid$x, com.grid$y,
as.matrix(alicante[,1]), as.matrix(alicante[,2]))
dentro_castellon = point.in.polygon(com.grid$x, com.grid$y,
as.matrix(castellon[,1]), as.matrix(castellon[,2]))

com.gridAdemuz = com.grid[dentro_ademuz>0,]
com.gridValencia = com.grid[dentro_valencia>0,]
com.gridAlicante = com.grid[dentro_alicante>0,]
com.gridCastellon = com.grid[dentro_castellon>0,]

compx = c(com.gridAdemuz$x, com.gridCastellon$x, com.gridValencia$x, com.gridAlicante$x)
compy = c(com.gridAdemuz$y, com.gridCastellon$y, com.gridValencia$y, com.gridAlicante$y)

gridComunidad=list(x=compx, y=compy)
```

```
# visualizarlo en el mapa  
plot(gridComunidad)
```



B.1.4. Probamos distintos tipos de modelos

Esférico

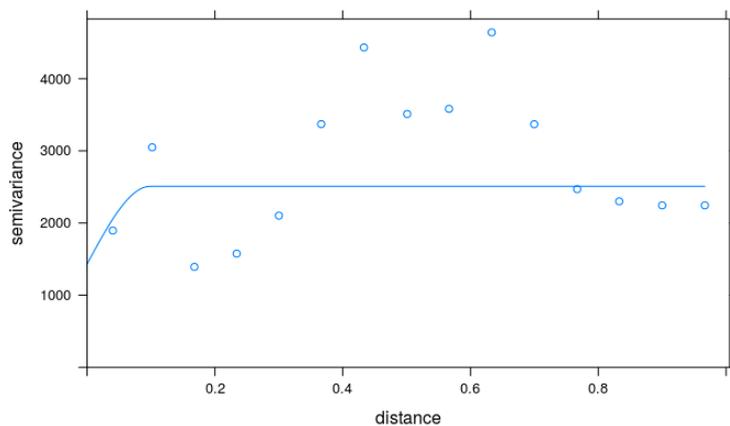
Vemos la variabilidad de los datos en el espacio con el variograma:

```
vario = variogram(Casos_totales~1, puntos,cutoff=1)

vario.fit <- fit.variogram(vario, model=vgm(psill = 4000, model = "Sph", range = 1, nugget =

## Warning in fit.variogram(vario, model = vgm(psill = 4000, model = "Sph", :
## singular model in variogram fit

plot(vario,vario.fit)
```



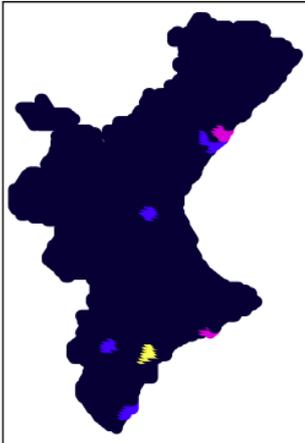
Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```
cvgrid = as.data.frame(gridComunidad) # hace falta que sea data frame
names(cvgrid) = c('Longitud', 'Latitud')
coordinates(cvgrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res<-krige((Casos_totales) ~ 1, locations=puntos, newdata=cvgrid, vario.fit, nmax = 1)

## [using ordinary kriging]

par(mfrow=c(1,2))
splot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

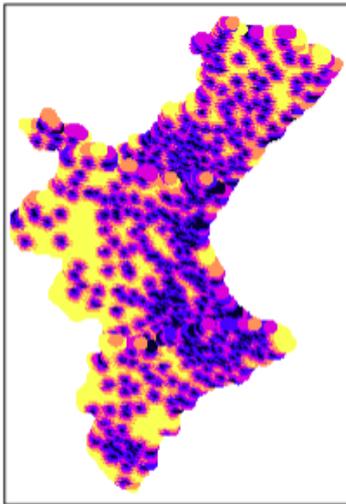
Predicción del número de CASOS



- [0,170.6]
- (170.6,341.2]
- (341.2,511.8]
- (511.8,682.4]
- (682.4,853]

```
splot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



- [2871,3300]
- (3300,3729]
- (3729,4158]
- (4158,4588]
- (4588,5017]

Exponencial

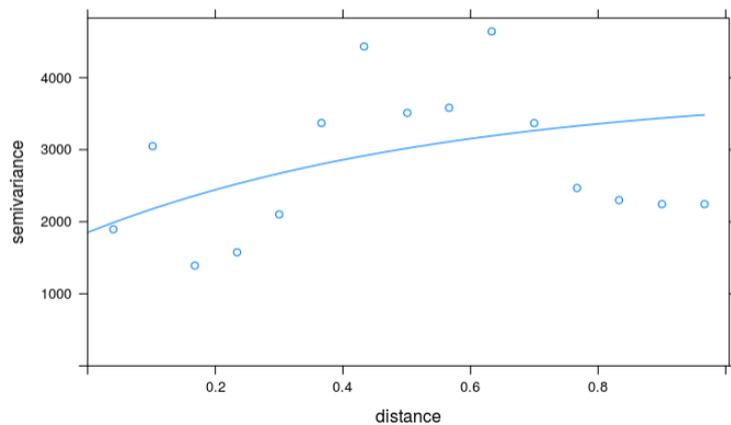
Vemos la variabilidad de los datos en el espacio con el variograma:

```
vario = variogram(Casos_totales~1, puntos,cutoff=1)

vario.fit <- fit.variogram(vario, model=vgm(psill = 6000, model = "Exp", range = 0.3, nugget

## Warning in fit.variogram(vario, model = vgm(psill = 6000, model = "Exp", : No
## convergence after 200 iterations: try different initial values?

plot(vario,vario.fit)
```



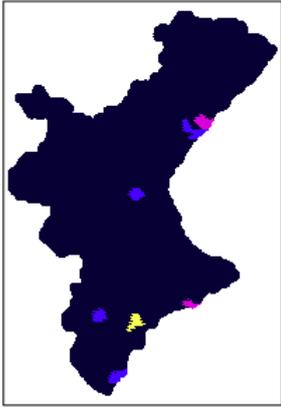
Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```
cvgrid = as.data.frame(gridComunidad) # hace falta que sea data frame
names(cvgrid) = c('Longitud', 'Latitud')
coordinates(cvgrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=Casos_totales ~ 1, locations=puntos, newdata=cvgrid, model=vario.fit,nma

## [using ordinary kriging]

# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
splot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

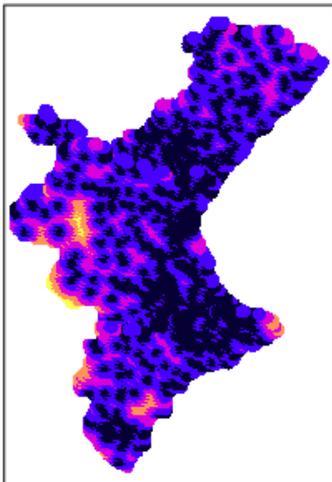
Predicción del número de CASOS



- [0,170.6]
- (170.6,341.2]
- (341.2,511.8]
- (511.8,682.4]
- (682.4,853]

```
spplot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



- [3705,3916]
- (3916,4126]
- (4126,4337]
- (4337,4547]
- (4547,4757]

Gaussiano:

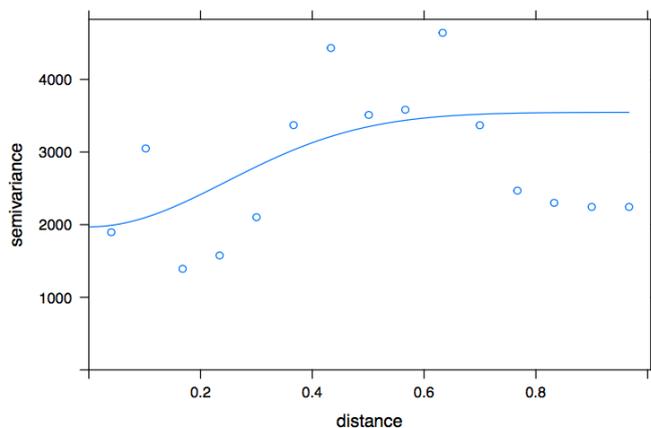
Vemos la variabilidad de los datos en el espacio con el variograma:

```
vario = variogram(Casos_totales~1, puntos,cutoff=1)

vario.fit <- fit.variogram(vario, model=vgm(psill = 1500, model = "Gau", range = 1, nugget =

## Warning in fit.variogram(vario, model = vgm(psill = 1500, model = "Gau", : No
## convergence after 200 iterations: try different initial values?

plot(vario,vario.fit)
```

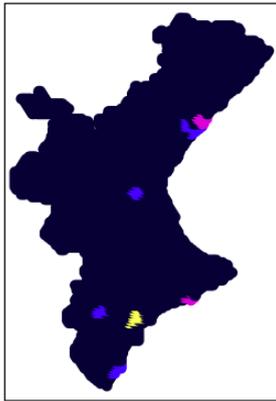


Ahora manipulamos el grid para que sea un objeto que se pueda trabajar con el Kriging.

```
cvgrid = as.data.frame(gridComunidad) # hace falta que sea data frame
names(cvgrid) = c('Longitud', 'Latitud')
coordinates(cvgrid) = ~ Longitud + Latitud # hace falta que sea spatial...
# hacer el kriging
res <- krige(formula=Casos_totales ~ 1, locations=puntos, newdata=cvgrid, model=vario.fit,nma

## [using ordinary kriging]
# dibujar las predicciones y sus varianzas
par(mfrow=c(1,2))
splot(res[c('var1.pred')], main='Predicción del número de CASOS')
```

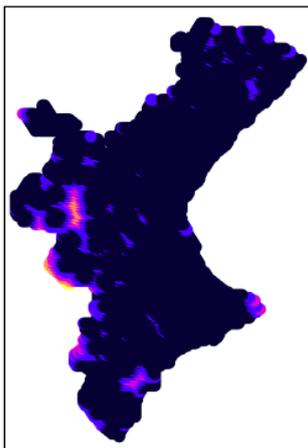
Predicción del número de CASOS



- [0,170.6]
- (170.6,341.2]
- (341.2,511.8]
- (511.8,682.4]
- (682.4,853]

```
splot(res[c('var1.var')], main='Varianza de las predicciones')
```

Varianza de las predicciones



- [3940,4080]
- (4080,4220]
- (4220,4361]
- (4361,4501]
- (4501,4642]