Introducción a la Geoestadística y su Aplicación.

José Luis Moreno López

1 de diciembre de 2009

AGRADECIMIENTO

La elaboración de esta tesis no ha sido un esfuerzo solo mío, sino que en ella, está involucrado el apoyo brindado de diversas personas y dependencias y que gracias a ellos fué posible su realización.

En primer lugar agradezco a Dios por la vida que me ha dado y que me ha permitido dar un paso más en este camino que comenzó desde el momento de mi nacimiento.

Agradezco el respaldo que he tenido por parte de mi familia, a mis padres que me brindaron la oportunidad de iniciar una etapa que ha llegado a su fin y a mis hermanos por el apoyo brindado para que la culminación de la carrera sea de lo mas placentera.

Otro tipo de respaldo me la han brindado las personas que he conocido a lo largo de estos 7 años: amigos y profesores, quienes de una u otra forma me han acompañado y ayudado para mi formación tanto profesional como personal.

Quiero mencionar, en este apartado, a la Universidad Autónoma Chapingo por permitir que jóvenes de escasos recursos económicos inicien una carrera profesional.

Hago mención especial al director de esta tesis, el Dr. Bulmaro Juárez Hernández por compartirme su visión para hacer un trabajo de tesis de un tema que era totalmente desconocido para mi, pero que me llamó la atención desde el primer momento que me lo comentó y por su tiempo brindado en todo momento que acudí a él.

Mis mas sinceros agradecimientos al Comité de Becas del Comité Mexicano de Ciencia y Tecnología (COMECyT), entidad perteneciente al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico proporcionado y que ayudó a que la culminación de este trabajo de tesis no se viera interrumpida. Agradezco también a los responsables de las oficinas de Sanidad Vegetal, ubicadas en la Cd. de México por haberme proporciondao la base de datos, ya que con ello fué posible realizar la parte práctica de este trabajo.

DEDICATORIA

A mis padres, Martha y Raúl por su apoyo brindado durante todo el proceso de mi formación profesional, por su confianza que aún sigue vigente y que haré todo lo posible para su continuación, por su amor incondicional y por todos los momentos de oración que dedicaron para que mi salud y bienestar me acompañaran en todo momento. A mis hermanos, René, Mary, Rosy, Nelvy y Lupita, por su apoyo incondicional. Para todos ellos, con cariño.

Índice general

Índice de figuras

Índice de cuadros

RESUMEN

En este trabajo se realiza una revisión y estudio de la teoría Geoestadística, así como su aplicación para determinar la distribución histórica de la mosca de la fruta en la zona media del Estado de San Luis Potosí. Antes de la revisión de los conceptos y técnicas que utiliza la Geoestadística para su análisis, se hace una revisión de los conceptos básicos de Estadística: Descripción univariada y Bivariada, ya que una parte importante de un estudio Geoestadístico es el análisis exploratorio de los datos. Otra parte del estudio Geoestadístico es analizar el conjunto de datos para determinar las características de variabilidad y correlación espacial del fenómeno que se está estudiando, esto para tener conocimiento de cómo cambia la variable de estudio de una localización a otra.

El siguiente paso, es modelar la correlación espacial y la información proporcionada se utiliza para predecir el valor de la variable que se está midiendo en lugares donde no se tiene información. La técnica que utiliza la Geoestadística para predecir en lugares no muestreados es el kriging. Mediante esta técnica se crearon mapas de densidad de la mosca de la fruta para observar el comportamiento de su distribución a lo largo de 42 semanas en una zona media del estado de San Luis Potosí donde existe cultivos de cítricos, principalmente de naranja.

SUMMARY

This work tries about the study of the Geostatistic theory, as well as their application to determine the historical distribution the fruit fly in the half area of the State of San Luis Potosi. Before the study the concepts and technical the Geostatistic uses for its analysis, a revision is made about basic concepts of Statistical: univariate and Bivariate description, since an important part a study geoestadistics is the exploratory analysis the data. Another part the study geoestadistics is to analyze the data set to determine variability characteristics and space correlation about the phenomenon study, it is important to have knowledge like it changes the variable study a localization to another. To model the space correlation is the following step, and the proportionate information is used to predict the value the variable that is being measured in places where one doesn't have information, or, where one didn't take sample. The kriging is the technique that Geoestatistics uses to predict in places sampled no. By means of this technique maps were created density the fly to observe the behavior its distribution along 42 weeks in an area where it exists cultivation orange mainly.

Capítulo 1

INTRODUCCIÓN

Este trabajo se centra en la presentación teórica y la implementación metodológica del análisis de la variabilidad espacial, para obtener mapas de incidencia de una plaga que ataca a la fruta, desde un enfoque geoestadístico. Así como también la implementación de una metodología en el software R para su análisis.

Geoestadística es un término concebido por G. Matheron (Matheron, 1968, citado por Maximiano, 2007) que sirve para definir a la ciencia aplicada basada en el estudio de variables distribuidas espacialmente. Surgió a partir de los años sesenta, especialmente con el propósito de predecir valores de las variables en sitios no muestreados. Como antecedentes suelen citarse trabajos de Sichel (1947; 1949) y Krige (1951) (ambos autores citados por Giraldo, 2002). Sichel observó la naturaleza asimétrica de la distribución del contenido de oro en las minas sudafricanas, la equiparó a una distribución de probabilidad lognormal y desarrolló las fórmulas básicas para esta distribución. Ello permitió una primera estimación de las reservas, pero bajo el supuesto de que las mediciones eran independientes, una clara contradicción con la experiencia de que existen "zonas" más ricas que otras. Una primera aproximación a la solución del problema de estimación fue dada por el geólogo G. Krige que propuso una variante del método de medias móviles, el cual puede considerarse como el equivalente al krigeado simple que, como se verá más adelante, es uno de los métodos de estimación lineal en el espacio con mayores cualidades teóricas. La formulación rigurosa y la solución al problema de predicción o estimación vinieron de la mano de Matheron (1962) en la escuela de minas de París. En los años sucesivos la teoría se fue depurando, ampliando su campo de validez y reduciendo las hipótesis necesarias (Samper y Carrera, 1990, citado por Giraldo).

Geoestadística se define formalmente como el estudio de las variables numéricas que se encuentran distribuidas de manera dependiente en una determinada porción del espacio

(Chauvet, 1994, citado por Maximiano, 2007). Es decir, cada valor observado perteneciente a una distribución, se encuentra asociado (está en función) a una posición espacial. Por consiguiente, el cambio en los valores de la variable, dependerá de su localización.

Aunque la aplicación de la herramienta geoestadística es relativamente reciente, son innumerables los ejemplos en los que se ha utilizado esta técnica con el ánimo de predecir fenómenos espaciales (Robertson, 1987; Cressie y Majure, 1995; Diggle et al., 1995, autores citados por Giraldo).

La geoestadística es sólo una de las áreas del análisis de datos espaciales, utiliza funciones para modelar la variación espacial, y estas funciones son utilizadas posteriormente por la técnica conocida como kriging para interpolar en el espacio el valor de la variable en sitios no muestreados. La fortaleza de la geostadística radica en que el uso del kriging para la interpolación es considerada una estimación muy robusta ya que se basa en la función continua que explica el comportamiento de la variable en distintas direcciones del espacio, y que en contraste con otros métodos de interpolación (como por ejemplo interpolar un punto usando los valores de los puntos que le rodean ponderados por la distancia que los separa) permite asociar la variabilidad de la estimación. Mediante este trabajo se pretende integrar un campo en el cual podría ser de mucha ayuda el uso de la geoestadística aplicando sus herramientas a los diversos sistemas agrícolas, por ejemplo estudiar el comportamiento de alguna plaga para predecir el área de mayor ataque y así mejorar su control, el estudio de los niveles de algún contaminante en diferentes sitios de una parcela, o simplemente observar la distribución de los rendimientos de cosecha de una variedad de maíz en una área donde involucre cientos de hectáreas.

OBJETIVOS

Objetivo general

Presentación de los conceptos básicos de la geoestadística y su aplicación para determinar la distribución histórica de la mosca de la fruta en la zona media del Estado de San Luis Potosí.

Objetivos particulares

1. Consulta bibliográfica de los conceptos desarrollados en la teoría Geoestadística.
2. Desarrollar y presentar a través del software incluido en el sistema computacional denominado R", la metodología que usa la geoestadística para la interpolación de datos espaciales.
3. Determinar la distribución histórica de la mosca de la fruta en la zona media del Estado de San Luis Potosí utilizando herramientas de la geoestadística

Capítulo 2

DESCRIPCIÓN UNIVARIADA Y BIVARIADA

2.1. Descripción univariada

El manejo de la información es fundamental para mejorar las prácticas realizadas en las distintas áreas del conocimiento humano. Por lo que la información que se posea sobre las diferentes variables estudiadas debe estar organizada de tal manera que sea de fácil lectura y manejo para su análisis. Para ello, el ser humano se vio en la necesidad de crear una ciencia que reduzca la información a valores numéricos para una mejor y más fácil interpretación de los fenómenos, dando a ésta el nombre de **Estadística**. Muchos temas de la estadística tratan acerca de la organización, presentación y resumen de los datos. Puede ser que sólo se necesite un resumen de los datos o bien se necesite realizar una gráfica sobre el comportamiento de la distribución de los mismos, por lo que tener un conjunto con una gran cantidad de datos, la visualización de todos estos de forma conjunta para poder estudiar su distribución no es muy confiable, por lo que frecuentemente es necesario emplear alguna otra estrategia para su análisis.

2.1.1. Tablas de Frecuencia e Histogramas

Una de las más comunes y útil presentacion del conjunto de datos es la tabla de frecuencia y su correspondiente gráfica, el histograma. Una tabla de frecuencia se refiere a que tan frecuentemente los valores observados caen dentro de ciertos intervalos. A estos intervalos se les llama intervalos de clases, clases de frecuencia o simplemente clases. El

Cuadro 2.1: Tabla de frecuencias de 444 datos correspondientes a la mosca de la fruta, obtenidos en la zona media del Estado de San Luis Potosí.

Clases	Número	Porcentaje
$[0, 5)$	243	54.7
$[5, 10)$	62	14.0
$[10, 15)$	56	12.6
$[15, 20)$	23	5.2
$[20, 25)$	17	3.8
$[25, 30)$	8	1.8
$[30, 35)$	15	3.4
$[35, 40)$	9	2.0
$[40, 45)$	1	0.2
$[45, 50)$	4	0.9
$[50, 55)$	2	0.5
$[55, 60)$	0	0.00
$[60, 65)$	2	0.5
$[65, 70)$	0	0.00
$[70, 75)$	2	0.5

cuadro 2.1 muestra una tabla de frecuencia que resume un conjunto de datos acerca de un muestreo en la zona media del Estado de San Luis Potosí realizado sobre la mosca de la fruta y llevado a cabo por personal de Sanidad Vegetal. El muestreo consisitó en colocar 444 trampas distribuidas en terrenos donde existe cultivos frutales (cítricos principalmanete) y se contó durante 42 semanas el número de moscas atrapadas en cada una de dichas trampas. Para el Cuadro 2.1 se utilizaron los valores correspondientes a la semana 4.

La información presentada en el Cuadro 2.1 también puede ser representada gráficamente en un histograma, como en la Figura 2.1, donde se nota que la primera clase tiene la mayor altura, lo que indica que gran parte de los datos, caen en el primer intervalo o clase (0-5]. Es común usar una anchura de clase constante para el histograma, así, la altura de la barra es proporcional al número de valores que caen dentro de la clase, mientras el número de valores que cae dentro de la clase sea mayor, la barra en el histograma, de esa clase será más alta. Si las anchuras de clases varían, en este caso, el área de la barra es la que es proporcional al número de valores dentro de la clase.

Tablas de Frecuencia Acumulativa e Histogramas: Muchos textos de Estadística usan la convención de que los datos se ordenan en forma ascendente para producir tablas de frecuencia acumulativa y descripciones de distribución de frecuencia acumulativa.

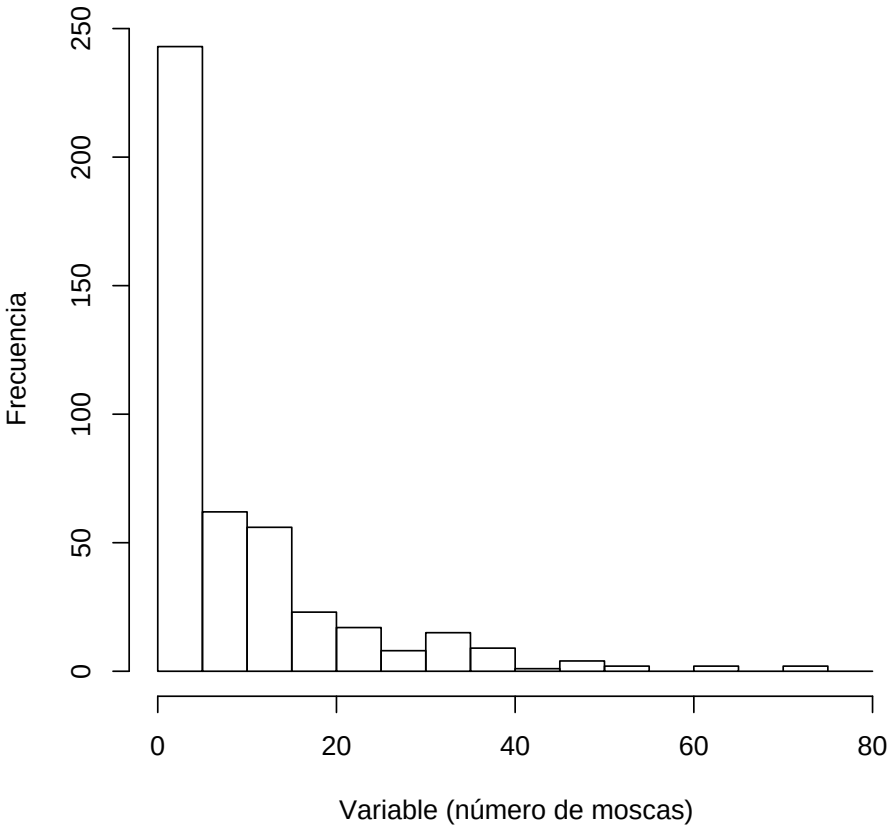


Figura 2.1: Histograma de los 444 valores correspondientes a la mosca de la fruta. Se observa que en las clases [0,5), [5,10) y [10,15) es donde se concentra la mayor parte de los datos, por lo que las barras correspondientes a estas tres clases presentan mayor área.

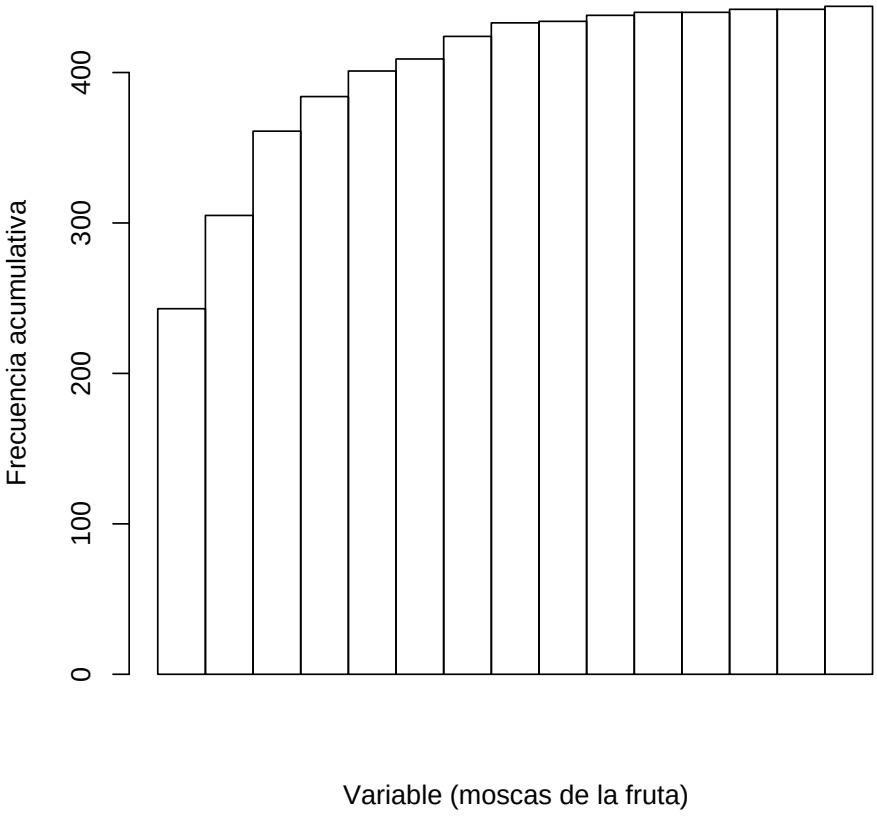


Figura 2.2: Histograma acumulativo de los 444 datos correspondientes a la mosca de la fruta. La mayor cantidad de los datos se concentran en la primera barra, correspondiente a la clase [0,5). Las siguientes clases contienen datos cada vez en menor proporción, por lo que las barras aumentan cada vez en menor proporción.

Cuadro 2.2: Tabla de frecuencia acumulativa de los 444 datos con una anchura de clase de 5 moscas. Mas del 50 % de los datos caen en la primera clase.

Clases	Número	Porcentaje
$[0, 5)$	243	54.7
$[0, 10)$	305	68.7
$[0, 15)$	361	81.3
$[0, 20)$	384	86.1
$[0, 25)$	401	90.3
$[0, 30)$	409	92.1
$[0, 35)$	424	95.5
$[0, 40)$	433	97.5
$[0, 45)$	434	97.7
$[0, 50)$	438	98.6
$[0, 55)$	440	99.1
$[0, 60)$	440	99.1
$[0, 65)$	442	99.5
$[0, 70)$	442	99.5
$[0, 75)$	444	100

En el Cuadro 2.2 se tiene la información del Cuadro 2.1 y presentado en forma acumulativa. En lugar de contar el número de valores dentro de cierta clase, aquí se tiene el número total de valores que se acumulan hasta esa clase. El correspondiente histograma acumulativo, mostrado en la Figura 2.3, es una función no decreciente.

2.1.2. Gráficas de Probabilidad Normal y Lognormal

Algunas de las herramientas de estimación trabajan bajo el supuesto de que la distribución de los valores de datos se aproximan a la distribución Gaussiana o distribución *normal*. La distribución Gaussiana es una de muchas distribuciones que tiene una descripción matemática concisa; tiene propiedades que favorecen su uso en aproximaciones teóricas de estimación. Por lo que es interesante saber qué tan cerca está la distribución de los datos a la distribución Normal. Una gráfica de probabilidad normal es un tipo de gráfica de frecuencia acumulativa que ayuda a decidir si la distribución del conjunto datos se aproxima o no a la distribución normal. En esta gráfica, las observaciones de un conjunto de datos se ordenan de manera no decreciente y luego se grafican contra los valores esperados estandarizados de las observaciones bajo el supuesto de que los datos están distribuidos normalmente. Si el supuesto es verdad, los datos se graficarán en una

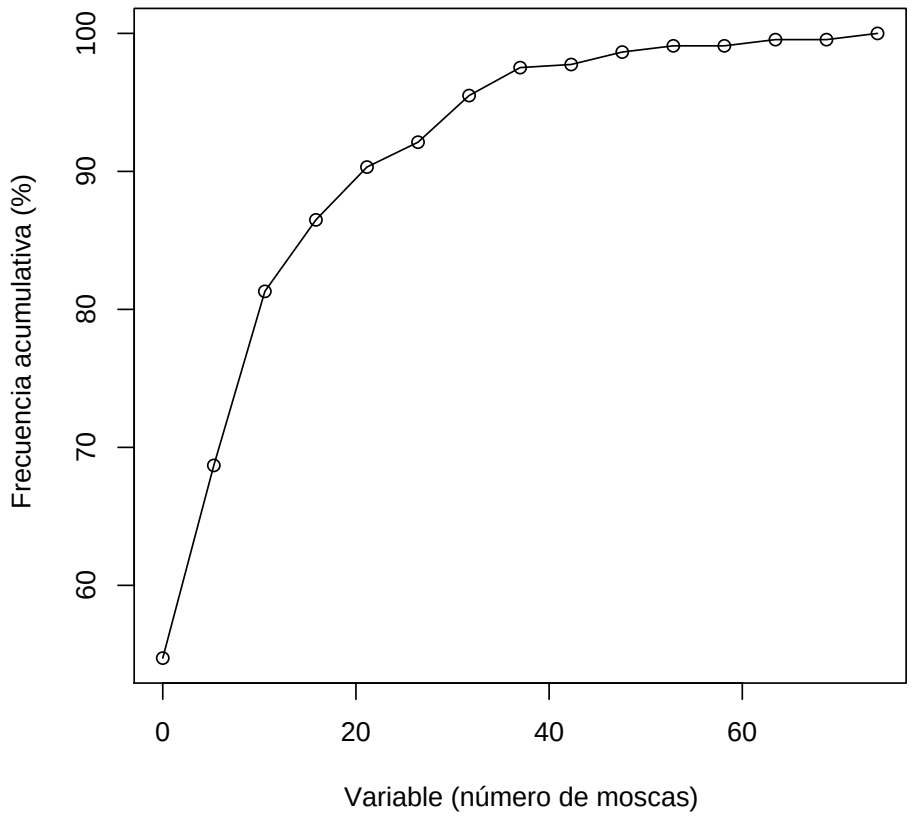


Figura 2.3: Gráfica de probabilidad normal de los 444 datos. Los puntos correspondientes a las frecuencias acumulativas no se encuentran sobre una línea recta, por lo que se puede considerar que el conjunto de datos no se aproxima a una distribución normal.

línea recta.

La Figura 2.3 muestra una gráfica de probabilidad normal de los 444 valores correspondientes a la semana 4 del muestreo a partir de las frecuencias acumulativas dadas en el Cuadro 2.2. Se observa que los puntos correspondientes a las frecuencias acumulativas no se encuentran sobre una línea recta, por lo que se puede considerar que el conjunto de datos no se aproxima a una distribución normal. Pero esto no quiere decir que el conjunto de datos estudiado no sea útil para nuestros propósitos, pues aunque las herramientas de estimación se han desarrollado de manera exhaustiva para el caso en el que la distribución de los datos es la distribución Gaussiana, no necesariamente tiene que cumplirse esta propiedad.

Existen conjuntos de datos de variables que no tienen distribución cercana a la normal, ya que es común tener muchos valores pequeños absolutos y pocos valores muy grandes. Aunque la distribución normal es frecuentemente inapropiada para representar distribuciones asimétricas, una distribución cercanamente relacionada a ella, la distribución lognormal, puede en algunas ocasiones ser una buena alternativa.

Usando una escala logarítmica en el eje X de una gráfica de probabilidad normal, se puede checar la lognormalidad. Igual que en la gráfica de probabilidad normal, en este caso las frecuencias acumulativas se hallarán sobre los puntos correspondientes a una línea recta si los datos están distribuidos en forma lognormal.

2.1.3. Estadísticas de Resumen

Las estadísticas de resumen caen dentro de tres categorías: medidas de localización, de dispersión y de forma.

Las estadísticas del primer grupo son: la media, la mediana y la moda. Ellas pueden dar una idea de donde se encuentra el centro de la distribución. En el segundo grupo se encuentra la varianza, la desviación estándar y el rango intercuartil y son usados para describir la variabilidad de los datos. La forma de la distribución es descrita por los coeficientes de asimetría, de curtosis y de variación. El coeficiente de asimetría proporciona información sobre la longitud de la cola para ciertos tipos de distribución, y permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). El coeficiente de curtosis analiza el grado de concentración que presentan los valores alrededor de la zona central de la distribución y el coeficiente de variación es una medida que indica, porcentualmente, qué tan dispersos o separados están los datos, unos con respecto a otros.

Así que estas estadísticas proporcionan un valioso resumen de la información que no es fácilmente mostrada en el histograma.

Medidas de Localización Muestrales

Media: La media muestral, $\bar{x}$, es el promedio aritmético de los datos.

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.1}$$

El número de datos es n y $x_1, ..., x_n$ son los valores de los datos.

La media representa un valor promedio con el mismo peso de cada una de las observaciones y por consiguiente cada uno de los datos influye de igual forma en el resultado de

ésta, ocasionando que en muchas situaciones cuando se tienen algunos datos que se alejan considerablemente del resto, el valor promedio encontrado no refleja la verdadera realidad del caso. La media muestral de los 444 datos es 8.51 moscas. Este valor se relaciona con el hecho de que en el histograma de la Figura 2.1 gran cantidad de los datos caen en las primeras 3 clases, por lo que la media aritmética se encuentra muy cercana a estas tres clases.

Mediana: De lo expuesto anteriormente se observa la necesidad de introducir otro tipo de medida de localización en la que valores atípicos con respecto al resto no tengan una influencia tan marcada respecto a la medida de localización. A dicha medida, debido a su naturaleza, se le conoce con el nombre de *Mediana*. La mediana, M, es el punto medio de los valores observados cuando se han ordenado de forma no decreciente en cuanto a su magnitud. La mitad de los valores están debajo de la mediana y la otra mitad arriba de la mediana. Una vez que el conjunto de datos está ordenado de tal manera que $x_1 \leq x_2 \leq ... \leq x_n$, la mediana se calcula utilizando la siguiente fórmula:

$$M = \begin{cases} x_{\frac{n+1}{2}} & \text{si } n \text{ es impar} \\ \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{si } n \text{ es par} \end{cases} \tag{2.2}$$

La mediana de los 444 datos es 3. Ambas, la media y la mediana son medidas de la localización del centro de la distribución. La media es muy sensible a valores extremos, no así la mediana. Esto se refleja en el hecho de que la mediana es menor que la media. Aquí se observa claramente como los valores extremos, por ejemplo el 74 influye de manera significativa en la media.

Moda: La moda de un conjunto de datos es el valor que se presenta con mayor frecuencia. La clase con la barra más alta en el histograma da una idea rápida de donde se localiza la moda. Del histograma en la Figura 2.2, se observa que la clase (0-5] tiene los valores más altos. Dentro de esta clase, el valor 0 ocurre más veces que cualquier otro. Por lo que la moda de los datos es 0.

Aunque existen casos en que o bien la moda no existe o bien no es la única. Es decir, si en una serie de datos, la frecuencia de cada uno de estos es la misma, entonces la moda no existe, en tal caso, al conjunto de datos se le llama **amodal** o sin **moda**. Cuando el conjunto de datos tiene más de una moda se dice que es **multimodal**: **bimodal** si son dos modas, **trimodal** si son tres, etc.

Mínimo: El valor mas pequeño en el conjunto de datos es el mínimo. Para los 444 valores, el valor mínimo es 0 moscas.

Máximo: El valor mas grande en el conjunto de datos es el máximo. El máximo para los 444 datos de la mosca de la fruta es 74.

Cuartil inferior y superior: De la misma manera que la mediana parte a los datos a la mitad, los cuartiles parten a los datos en cuatro partes, cada una de ellas contiene el mismo número de datos. Si los valores de los datos son ordenados de manera no decreciente, entonces un cuarto de los datos caen debajo del cuartil inferior o primer cuartil, Q_1, y un cuarto de los datos caen arriba del cuartil superior o tercer cuartil, Q_3.

Un resumen de los 444 valores, incluyendo los cuartiles inferior y superior es:

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   3.000   8.514  12.000  74.000
```

Igual que la mediana, los cuartiles pueden ser fácilmente leídos de la gráfica de frecuencias acumuladas. El valor en el eje X, el cual corresponde al 25 % en el eje Y, es el cuartil inferior y el valor que corresponde al 75 % es el cuartil superior.

Deciles, Percentiles y Cuantiles: La idea de fraccionar a los datos en dos partes con la mediana o en cuartos con los cuartiles puede ser extendido a alguna otra fracción.

Los deciles dividen los datos en décimos. Un décimo de los datos cae debajo del primer decil, dos décimos de los datos caen debajo del segundo decil. El quinto decil corresponde a la mediana.

En muchas aplicaciones al tener un conjunto de datos se requiere conocer los valores de la variable que están por debajo de un porcentaje. Por ejemplo, al realizar un examen a un grupo de 30 personas se quisiera saber la calificación donde el 40 % de los alumnos está por debajo de esta. Por lo que, en este caso, se hace uso de los percentiles. Dado un conjunto de datos, se llama **percentil C** a la cantidad C_p, que representa el número para el cual el C % de los datos son menores que éste. Por ejemplo C_{40} es el valor de la variable que deja por debajo el 40 % de los datos. De esta manera hay relación entre percentil, cuartil y la mediana. El percentil 25 es lo mismo que el primer cuartil, el percentil 50 es el mismo que la mediana y el percentil 75 es el mismo que el tercer cuartil.

Para calcular el C_p de un conjunto de datos $x_1, ..., x_n$, se hace de la forma siguiente:

1. Se ordenan los datos de forma no decreciente, $\tilde{x}_1 \leq \tilde{x}_2 \leq ... \leq \tilde{x}_n$

2. Se determina el C % de los datos, calculando $\tilde{c} = \frac{n}{100}C$

3. Si la cantidad anterior es entera, entonces $C_p = \frac{\tilde{x}_{\tilde{c}}+\tilde{x}_{\tilde{c}+1}}{2}$

4. Si la cantidad anterior no es entera, entonces $C_p = \tilde{x}_{[\tilde{c}]+1}$

En donde $[\tilde{c}]$ representa a la parte entera de $\tilde{c}$. Por ejemplo, si $\tilde{c} = 24{,}7$, $[\tilde{c}] = 24$, si $\tilde{c} = 24{,}2$, $[\tilde{c}] = 24$.

Los cuantiles son una generalización de la idea de fraccionar a los datos. Por ejemplo, si se quiere hablar sobre el valor debajo del cual cae un vigésimo de los datos, se habla de $q_{0,05}$, en lugar de insertar un nuevo nombre para los vigésimos. Así como es cierto que los deciles y percentiles son equivalentes a la mediana y los cuartiles, así también los cuantiles pueden ser escritos como uno de esos estadísticos. Por ejemplo, $q_{0,5}$ es la mediana y $q_{0,75}$ es el cuartil superior.

Medidas de dispersión

Para un análisis más completo de los datos, el estudio de sus medidas de localización no es suficiente, ya que en diferentes conjuntos de datos puede dar medidas de localización iguales y por lo tanto no se tendría el conocimiento de la verdadera forma de su distribución. Por lo que es necesario realizar un estudio de la distribución de los datos con respecto a su valor central, es decir, se necesita un valor que indique una medida para comparar las dispersiones de datos entre dos ó más conjuntos diferentes.

Una medida de dispersión indica qué tan cercanos o separados están los valores con respecto a la media u otra medida de tendencia central (Mediana, Moda, etc.). Esto es, una medida de dispersión indica qué tan confiable es la medida de tendencia central, por ejemplo, el promedio de los datos.

Varianza: La varianza muestral, s^2, está dado por:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2.3}$$

Esto es el promedio del cuadrado de la diferencia de los valores observados de su media. Como involucra las diferencias al cuadrado, la varianza es sensible a valores altos. Por ejemplo, la varianza de los 444 datos del número de moscas es 142.3

Desviación estándar: La desviación estándar, s, es simplemente la raiz cuadrada de la varianza y frecuentemente es usada en lugar de la varianza ya que sus unidades son las

mismas que las unidades de la variable estudiada. Así para los 444 datos que se han estado analizando, la desviación estándar es 11.92

Rango intercuartil: Otra medida de uso para medir la dispersión de los valores observados es el rango intercuartil. El rango intercuartil o RIC, es la diferencia entre el cuartil superior y el inferior y está dado por

$$RIC = Q_3 - Q_1 \tag{2.4}$$

Al contrario de la varianza y la desviación estándar, el rango intercuartil no usa la media como el centro de la distribución, y frecuentemente es preferido si los valores altos tienen mucha influencia sobre la media. El rango intercuartil de los 444 datos, es 11.

Medidas de forma

Coeficiente de asimetría: Un rasgo del histograma es que las estadísticas previas no dan a conocer la simetría ó asimetría de la distribución. La estadística más comúnmente usada para resumir la simetría es llamada *el coeficiente de asimetría*. La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa cómo están distribuidos los datos respecto al eje de simetría. Se dice que la asimetría es positiva cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es simétrica cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como asimetría negativa cuando la mayor cantidad de datos se aglomeran en los valores menores que la media. El coeficiente de asimetría se define como

$$\text{coeficiente de asimetría} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{s^3} \tag{2.5}$$

Donde $\bar{x}$ es el promedio aritmético de los datos como se mencionó anteriormente y s es la desviación estándar. Aquí, el numerador es el promedio de la diferencia cúbica entre los datos y su media, y el denominador es el cubo de la desviación estándar.

Si el coeficiente de asimetría es igual a cero, la distribución es simétrica, es decir, existe la misma cantidad de valores a los dos lados de la media. Este valor es difícil de conseguir por lo que se tiende a tomar los valores que son cercanos ya sean positivos o negativos ($\pm$0,5).

Si es mayor que cero, la curva es asimétricamente positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media.

Si es menor que cero, la curva es asimétricamente negativa por lo que los valores se tienden a reunir más en la parte derecha de la media.

El coeficiente de asimetría es más sensible que la media y la varianza a valores extremos Un solo valor grande puede influenciar altamente este coeficiente, ya que la diferencia entre cada dato y la media está elevado al cubo.

De lo visto anteriormente, se puede afirmar que un histograma es sesgado positivamente si tiene una cola larga de valores altos a la derecha, haciendo la mediana menor que la media. Si hay una cola larga de valores pequeños a la izquierda, la mediana es más grande que la media, el histograma es sesgado negativamente. Si la asimetría es cercana a cero, el histograma es aproximadamente simétrico y la mediana esta cerca de la media. Para los 444 datos que se ha descrito en este capítulo, el coeficiente de asimetría es 2.31, indicando una distribución fuertemente asimétrica, en este caso, es asimétrica a la derecha, ya que se observa en el histograma de la Figura 2.1 que los datos tienden a concentrarse más en la parte izquierda que en la derecha de la media aritmética.

Coeficiente de curtosis: Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, se puede identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) ó una baja concentración (Platicúrtica).

Para calcular el coeficiente de Curtosis se utiliza la ecuación:

$$\text{coeficiente de curtosis} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{s^4} - 3 \tag{2.6}$$

Si el coeficiente de curtosis es cercano a cero, la distribución es Mesocúrtica, si es mayor que cero es Leptocúrtica y si es menor que cero, es Platicúrtica. El coeficiente de curtosis de los 444 datos correspondientes a la mosca de la fruta es 6.52, por lo que se considera que su distribución es Leptocúrtica

Coeficiente de variación: El coeficiente de variación, CV, es una estadística que se usa frecuentemente como una alternativa de asimetría para describir la forma de la distribución. Es usado principalmente para las distribuciones donde los valores son todos positivos y la asimetría es también positiva; aunque esto puede ser calculado para otros tipos de distribuciones, su utilidad como un índice de forma llega a ser cuestionable. El coeficiente de variación está definido como el cociente de la desviación estándar y la media:

$$CV = \frac{s}{\bar{x}} \tag{2.7}$$

Un CV cercano a cero indica que los datos están muy juntos (o son muy similares), mientras que un CV muy grande (cercano a 100 % o a 1, o mayor a estos valores) indica que los datos están muy dispersos (o son muy diversos).

Si la estimación es la meta final de un estudio, el coeficiente de variación puede proveer alguna advertencia de problemas. Un coeficiente de variación mayor que 1 indica la presencia de algunos valores muestrales muy grandes que pueden tener un impacto significativo en las estimaciones finales.

El coeficiente de variación para los 444 datos es 1.4, el cual refleja el hecho de que el histograma de la Figura 2.1 tiene una cola de valores grandes que se repiten pocas veces y se alejan demasiado de la media aritmética.

2.2. Descripción Bivariada

Al análisis realizado a datos con las herramientas univariadas discutidas anteriormente, se le conoce como análisis descriptivo univariado, ya que pueden ser usadas para describir la distribución de los datos de una variable. Si el estudio comprende dos variables simultáneamente, se conoce como análisis descriptivo bivariado, si comprende más de dos variables, se conoce como análisis descriptivo multivariado. Para cada uno de estos análisis, existen técnicas específicas que ayudan a visualizar mejor los datos y hacer un análisis más a detalle. Algunas de las características más importantes de conjuntos de datos obtenidos en fenómenos que se investigan en ciencias de la tierra son la relación y la independencia entre variables. Por lo que a continuación se estudiarán las formas de describir la relación entre dos variables.

Como se mencionó anteriormente, el muestreo de la mosca de la fruta en el estado de San Luis Potosí se hizo de manera consecutiva durante 42 semanas. Además de los 444 datos ya analizados y que corresponden a la semana 4, se analizará otros 444 datos en las mismas localizaciones correspondientes a la semana 8.

2.2.1. Comparando Dos Distribuciones

Suponer que se está interesado en comparar dos distribuciones. Una presentación de sus histogramas junto con algunas estadísticas de resumen revelarán su relación o total diferencia. Desafortunadamente si las distribuciones son muy similares este método de comparación no ayuda a descubrir las pequeñas diferencias que exista entre ellas.

Al hacer un análisis bivariado el primer paso será (al igual que en el caso de la descripción univariada), representar los datos en una tabla de frecuencias. Ahora, a cada caso le

Cuadro 2.3: Estadísticas de resumen de los datos de la semana 4 y semana 8

Estadistica	semana 4	semana 8
n	444	444
$\bar{x}$	8.51	8.44
s	11.92	12.85
CV	1.4	1.52
min	0	0
Q_1	1	0
M	3	3
Q_3	12	10
max	74	75

corresponde no un valor, sino dos (uno para cada una de las variables).

Los pares de valores así formados constituyen la distribución bidimensional. La tabla de frecuencias consiste ahora en una tabla de doble entrada en la que se recogen tanto las frecuencias de cada una de las variables por separado como los pares de puntuaciones que cada caso obtiene en ambas variables (frecuencia conjunta).

Las puntuaciones pueden aparecer sin agrupar o agrupadas en intervalos, no teniendo por qué ser el número de intervalos de las dos variables iguales entre sí, así como tampoco de la misma amplitud.

Los histogramas de los valores de la semana 4 y semana 8 son mostrados en la Figura 2.4. y sus estadísticas son presentadas en el Cuadro 2.3. Hay mucha similitud entre las distribuciones de las dos variables. La distribución de los datos de ambas semanas son sesgadas positivamente, tienen una media muy parecida al igual que la mediana y la desviación estándar. Además de lo anterior, las estadísticas de resumen del Cuadro 2.3 permiten comparar, entre otras cosas, las medianas y los cuartiles de las dos distribuciones, las cuales no difieren mucho.

Para una buena comparación visual de dos distribuciones se puede usar la llamada gráfica cuantil-cuantil ó **gráfica q-q**. Esta gráfica es comúnmente usada cuando hay alguna razón para esperar que las distribuciones sean similares. Una gráfica q-q es una gráfica en la cual los cuantiles de las dos distribuciones son graficadas una contra la otra. La información contenida en el Cuadro 2.4 está presentada como una gráfica q-q en la figura 2.5. Los cuantiles de semana 8 sirven como la coordenada en X mientras que los de la semana 4 sirven como la coordenada en Y. Si las dos distribuciones que están siendo comparadas tienen el mismo número de datos, el cálculo de los cuantiles de cada distribución no es un

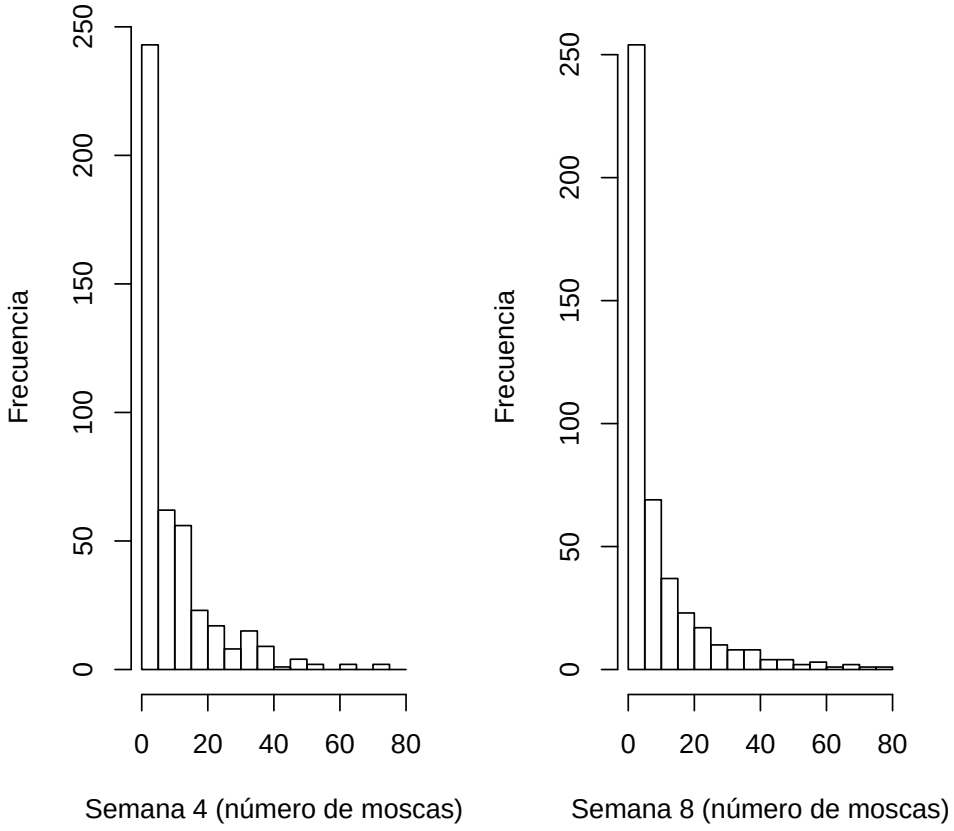


Figura 2.4: Histogramas de los 444 valores correspondientes a los datos de la semana 4 y la semana 8, respectivamente

Cuadro 2.4: Comparación de los cuantiles de los datos del número de moscas correspondientes a las semanas 4 y 8.

Frecuencia acumulativa	Quantil semana 4	Quantil semana 8
0.05	0.00	0.00
0.01	0.00	0.00
0.15	0.00	0.00
0.20	0.00	0.00
0.25	1.00	0.00
0.30	1.00	1.00
0.35	2.00	1.00
0.40	2.00	2.00
0.45	3.00	3.00
0.50	3.00	3.00
0.55	5.00	4.00
0.60	6.00	5.00
0.65	7.00	6.00
0.70	10.00	8.00
0.75	12.00	10.25
0.80	14.00	14.00
0.85	18.00	18.55
0.90	23.70	24.00
0.95	32.85	37.00
1	74.00	75.00

paso necesario para hacer la gráfica q-q. De hecho, se puede ordenar el conjunto de datos de cada distribución en orden ascendente y entonces graficar los correspondientes pares de valores.

Una gráfica q-q de dos distribuciones idénticas graficará una línea recta $y = x$. Para distribuciones que son muy similares, las pequeñas separaciones de la gráfica q-q de la línea recta $y = x$ revelarán donde ellas son diferentes. Para el caso de las distribuciones de los valores de la semana cuatro y ocho, son similares dentro del área seleccionada, por lo que su gráfica q-q está cerca de la línea recta.

Si una gráfica q-q de dos distribuciones es una línea recta pero diferente a la de $y = x$, entonces las dos distribuciones tienen la misma forma pero su localización y dispersión pueden diferir.

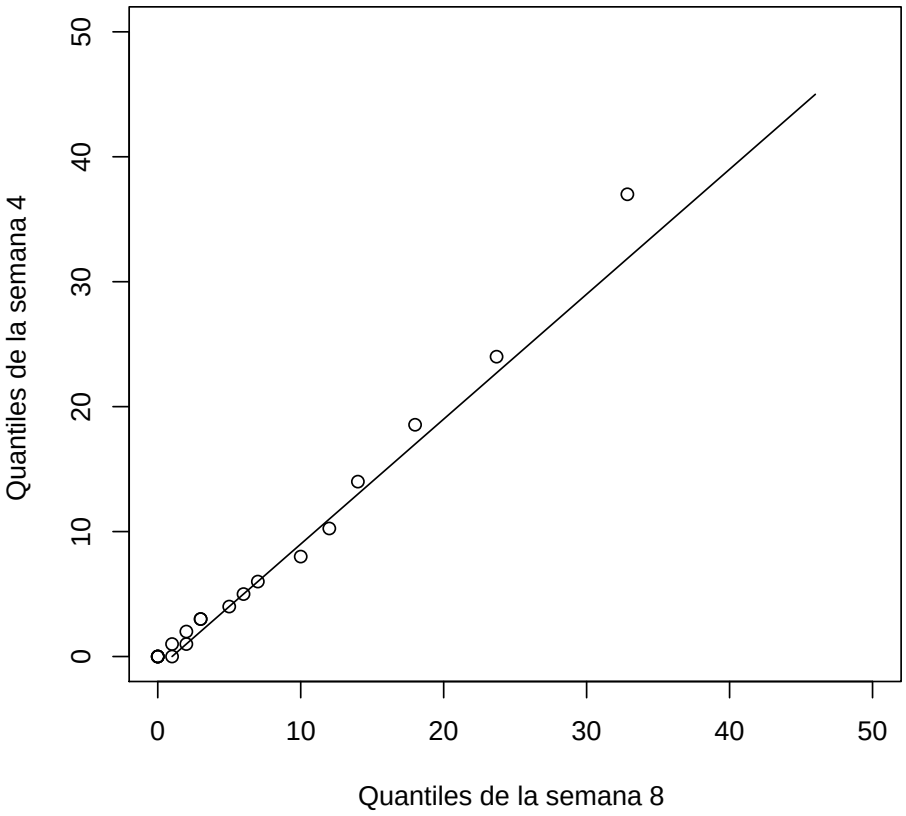


Figura 2.5: Gráfica q-q de la distribución de los 444 valores correspondientes a la semana 4 contra los 444 valores correspondientes a la semana 8.

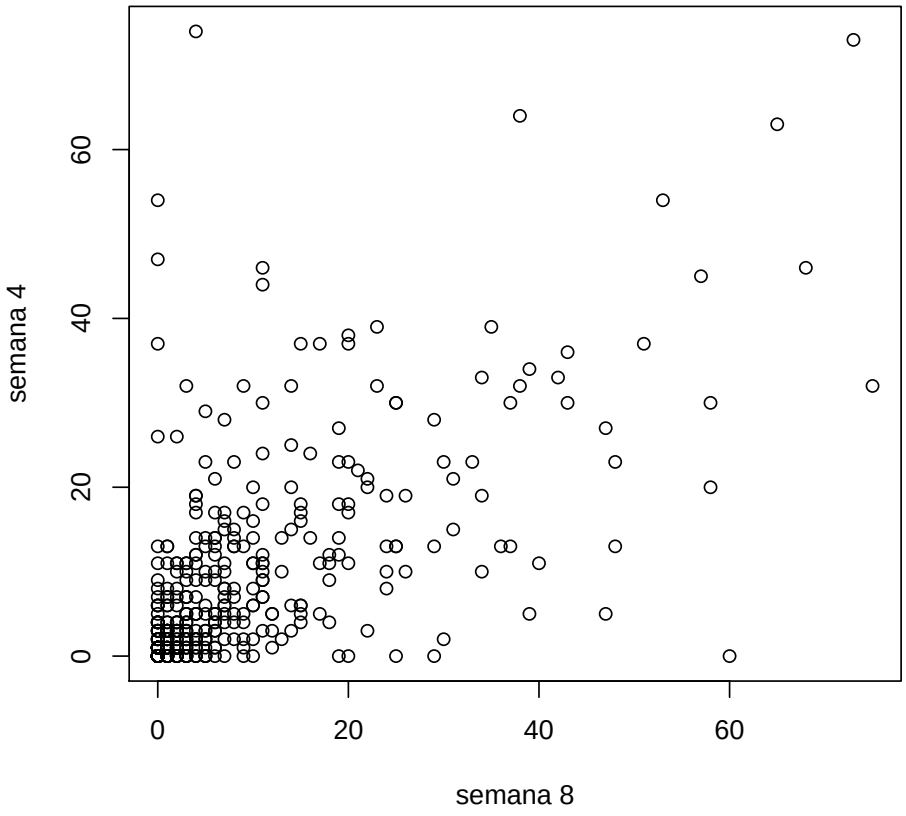


Figura 2.6: Diagrama de dispersión de los 444 datos correspondientes a la semana cuatro contra los datos que corresponden a la semana ocho

2.2.2. Diagrama de dispersión

Un despliegue más común de datos bivariados es el **diagrama de dispersión** o nube de puntos. Para construirlo, representamos cada par observado como un punto en el plano cartesiano. Este será una gráfica de los datos en la cual la coordenada X corresponde a los valores de una variable y la coordenada Y a los valores de la otra variable.

Al observar esta nube de puntos podemos apreciar si los puntos se agrupan cerca de alguna recta o no. Si es así, decimos que existe una correlación lineal y la recta se denomina recta de regresión.

Se dirá que hay correlación lineal fuerte si la nube de puntos se aproxima mucho a la recta y débil (o menos fuerte) si la nube de puntos se aleja de la recta. En el caso del

ejemplo, podemos observar que, a simple vista, la correlación lineal parece ser débil. No obstante, la apreciación visual para verificar la existencia de correlación no es suficiente.

La recta de regresión es ascendente cuando la correlación es positiva o directa (al aumentar el valor de una variable, el valor de la otra también aumenta) o descendente cuando la correlación es negativa o inversa (al aumentar el valor de una variable, el valor de la otra disminuye).

Además de proveer una buena percepción de como están relacionadas las dos variables el **diagrama de dispersión** también ayuda a observar la presencia de datos aberrantes. Al comenzar el estudio de un conjunto de datos espacialmente continuo es necesario observar y corroborar los datos; el éxito del método de estimación depende de que los datos sean confiables. Valores un poco erráticos pueden tener un gran impacto en la estimación.

2.2.3. Correlación

En un sentido más amplio, hay tres modelos que se pueden observar en un **diagrama de dispersión**: las variables son correlacionados positivamente, correlacionados negativamente o no correlacionados. Dos variables son positivamente correlacionadas si al aumentar el valor de una variable, el valor de la otra también aumenta, y similarmente con los valores más pequeños de cada variable. Dos variables son negativamente correlacionadas si al aumentar el valor de una variable, el valor de la otra disminuye.

La posibilidad final es que las dos variables no estén relacionadas. Un incremento en una variable no tiene aparentemente efecto en la otra. En este caso, se dice que las variables son no correlacionadas.

Coeficiente de correlación. El coeficiente de correlación, $\hat{\rho}$, es la estadística más comúnmente usada para resumir la relación entre dos variables y está dado por:

$$\rho = \frac{\frac{1}{n}\sum_{i=n}^{n}(x_i - \bar{x})(y_i - \bar{x})}{s_x s_y} \tag{2.8}$$

El número de datos es n; $x_1, ..., x_n$ son los valores de los datos para la primer variable, $\bar{x}$ es su media, y s_x es su desviación estándar; $y_1, ..., y_n$ son los valores de la segunda variable, $\bar{y}$ es su media, y s_y es su desviación estándar.

El numerador en la Ecuación 3.1 es llamado la covarianza de x e y, y se denota por C_{xy} estos es,

$$C_{XY} = \frac{1}{n}\sum_{i=n}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{2.9}$$

es la covarianza entre los datos x e y.

Dividiendo la covarianza por las desviaciones estándar de las dos variables, garantiza que el coeficiente de correlación siempre estará entre -1 y +1, además de proveer un índice que es independiente de la magnitud de los valores de los datos.

La covarianza correspondientes a los datos de mosca de la fruta para las semanas 4 y 8 es 95.8015, la desviación estándar de la semana 4 es 11.92 y de la semana 8 es 12.85. Por lo que el coeficiente de correlación para las semanas 4 y 8 es 0.63.

El coeficiente de correlación mide qué tan cerca de una línea recta caen los pares de valores observados. Si $\hat{\rho} = 1$, la correlación será fuerte y el **diagrama de dispersión** será una línea recta con una pendiente positiva; si $\hat{\rho} = -1$ la correlación también será fuerte, pero el **diagrama de dispersión** será una línea recta con pendiente negativa. Para $\mid \hat{\rho} \mid < 1$ el **diagrama de dispersión** aparecerá como una nube de puntos que se vuelve mas gordo y mas difuso conforme $\mid \hat{\rho} \mid$ decrece de 1 a 0. Si $C_{XY} = 0$, y por tanto, para $\hat{\rho} = 0$ la correlación es nula. La relación lineal es nula. El coeficiente de correlación para las semanas 4 y 8 es de 0.63, reflejando el hecho de que en el diagrama de dispersión se observe una nube de puntos e indicando que la correlación entre los dos conjuntos de datos es pobre.

Se debe notar que $\hat{\rho}$ provee una medida de la relación *lineal* entre dos variables. Si la relación entre dos variables no es lineal, el coeficiente de correlación puede ser una estadística de resumen muy pobre. El valor de $\hat{\rho}$ es frecuentemente un buen indicador de qué tan exitoso puede ser el tratar de predecir el valor de una variable a partir de la otra con una ecuación lineal. Si $\mid \hat{\rho} \mid$ es grande, entonces para un valor dado de una variable, la otra variable es restringida a un rango pequeño de posibles valores. De otra manera, si $\mid \hat{\rho} \mid$ es pequeño, entonces conociendo el valor de una variable, éste no nos ayuda para predecir el valor de la otra, al conciderar una relación lineal.

2.2.4. Regresión Lineal

Como se vio anteriormente, una fuerte relación entre dos variables puede ayudarnos a predecir una variable si la otra es conocida. La forma más simple para este tipo de predicción es la regresión lineal, en la cual se asume que la dependencia de una variable de la otra puede ser descrita por la ecuación de una línea recta:

$$y = \alpha x + \beta \tag{2.10}$$

La pendiente, α, y la constante, β, se obtiene a partir de:

$$\alpha = \hat{\rho}\frac{s_y}{s_x} \quad \beta = \bar{y} - \alpha\bar{x} \tag{2.11}$$

La pendiente, α, es igual a el coeficiente de correlación multiplicado por el cociente de las desviaciones estándar, con s_y siendo la desviación estándar de la variable que estamos tratando de predecir y s_x la desviación estándar de la variable conocida. Una vez que la pendiente es conocida, la constante, β, puede ser calculada usando las medias de las dos variables, $\bar{x}$ y $\bar{y}$.

Si se usa los 444 pares de datos de las semanas 4 y 8 para calcular una ecuación de regresión lineal para predecir la semana 4 a partir de la semana 8, se tiene:

$$\alpha = 0{,}63\frac{11{,}92}{12{,}85} = 0{,}58 \quad \beta = 8{,}51 - 0{,}58(8{,}44) = 3{,}62 \tag{2.12}$$

Donde la ecuación lineal para predecir los valores de la semana 4 a partir de los valores conocidos de la semana 8 es:

$$semana4 = 0{,}58 * semana8 + 3{,}62 \tag{2.13}$$

En la Figura 2.7 esta línea es superpuesta en el diagrama de dispersión. Esta regresión lineal no se ve muy bien por el hecho de tener una nube de puntos muy difuso en el diagrama de dispersión y por asumir que la dependencia entre la semana 4 y la semana es lineal.

La ecuación 2.12 da una predicción para los datos de la semana 4 dado los datos de la semana 8. Se puede también estar interesado en predecir los datos de la semana 8 si son conocidos los datos de la semana 4.

En la ecuación 2.10, y es la variable desconocida y x es conocida, así que para calcular la ecuación de regresión lineal que predice la semana ocho de la semana 4 es:

$$\alpha = 0{,}63\frac{12{,}85}{11{,}92} = 0{,}68 \quad \beta = 8{,}44 - 0{,}68(8{,}51) = 2{,}65 \tag{2.14}$$

La ecuación lineal para predecir los valores de la semana 8 a partir de los valores conocidos de la semana 4 es:

$$semana8 = 0{,}68 * semana4 + 2{,}65 \tag{2.15}$$

La línea de regresión de la ecuación (2.14) es también mostrada en la Figura 2.7. En

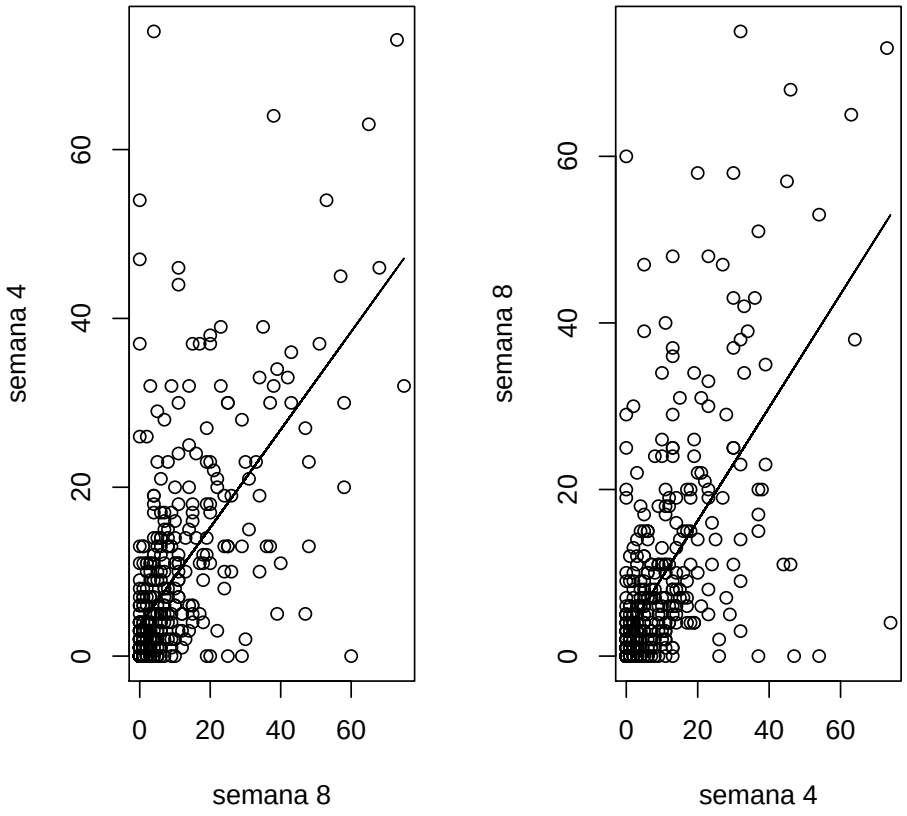


Figura 2.7: Lineas de Regresión lineal superpuestas a los diagramas de dispersión. La regresión lineal de la semana cuatro dada la semana ocho y de la semana ocho dada la semana cuatro, respectivamente.

esta Figura se tiene graficada la semana ocho en el eje y y la semana cuatro en el eje x para enfatizar el hecho de que ahora los valores de la semana 8 son desconocidos. Aunque se observa en la Figura 2.7 que las dos regresiones lineales son las mismas, esto no es cierto, y esto se puede verificar en las ecuaciones 2.12 y 2.14.

Capítulo 3

DEFINICIONES BÁSICAS DE GEOESTADÍSTICA

3.1. Definición de geostadística

La geoestadística es una rama de la estadística que trata fenómenos espaciales (Journel & Huijbregts, 1978, citado por Giraldo 2002). Su interés primordial es la estimación, predicción y simulación de dichos fenómenos (Myers, 1987, citado por Giraldo). Esta herramienta ofrece una manera de describir la continuidad espacial, que es un rasgo distintivo esencial de muchos fenómenos naturales, y proporciona adaptaciones de las técnicas clásicas de regresión para tomar ventajas de esta continuidad (Isaaks & Srivastava, 1989). Petitgas (1996), la define como una aplicación de la teoría de probabilidades a la estimación estadística de variables espaciales (Giraldo, 2002). Se puede entonces sugerir la idea de interpretar este fenómeno en términos de Función Aleatoria (FA), es decir, a cada punto $\mathbf{x}$ del espacio se le asocia una Variable Aleatoria (VA) $Z(\mathbf{x})$, para dos puntos diferentes $\mathbf{x}$ e $\mathbf{y}$, se tendrán dos VAs $Z(\mathbf{x})$ y $Z(\mathbf{y})$ diferentes pero no independientes, y es precisamente su grado de correlación el encargado de reflejar la continuidad del fenómeno en estudio, de modo que el éxito de esta técnica es la determinación de la función de correlación espacial de los datos (Zhang, 1992, citado en monografias.com).

Cuando el objetivo es hacer predicción, la geoestadística opera básicamente en dos etapas. La primera es el análisis estructural, en la cual se describe la correlación entre puntos en el espacio. En la segunda fase se hace predicción en sitios de la región no muestreados por medio de la técnica del kriging. Este es un proceso que consiste en una combinación lineal de pesos asociados a cada localización donde fue muestreado un valor $Z(\mathbf{x_i})$ $(i = 1, 2, 3, ..., n)$ del fenómeno estudiado. Los pesos asignados a los valores muestrales son

apropiadamente determinados por la estructura espacial de correlación establecida en la primera etapa y por la configuración de muestreo (Petitgas, 1996, citado por Giraldo).

La aplicación de la geoestadística para la estimación de reservas en minas, es probablemente su uso mejor conocido. Sin embargo, las técnicas de estimación pueden ser usadas dondequiera que se hace una medición continua en una muestra, en un espacio o tiempo particular, donde el valor muestral es afectado por su posición y su relación con sus vecinos. Los fundamentos básicos de geoestadística se presentan a continuación.

3.2. Variable regionalizada

Una variable medida en el espacio de forma que presente una estructura de correlación, se dice que es una variable regionalizada (Giraldo, 2006). De manera más formal se puede definir como un proceso estocástico con dominio contenido en un espacio euclidiano d-dimensional R^d , es decir, como el proceso estocástico $\{Z(\mathbf{x}) : \mathbf{x}\epsilon\mathbf{D} \subset \mathbf{R^d}\}$. Si d=2, $Z(\mathbf{x})$ puede asociarse a una variable medida en un punto $\mathbf{x}$ del plano (Díaz-Francés, 1993). En términos prácticos $Z(\mathbf{x})$ puede verse como una medición de una variable aleatoria (p.ej. concentración de un contaminante) en un punto $\mathbf{x}$ de una región de estudio (Giraldo, 2006).

En su libro, Giraldo menciona que un proceso estocástico es una colección de variables aleatorias indexadas; esto es, para cada $\mathbf{x}$ en el conjunto de índices D, $Z(\mathbf{x})$ es una variable aleatoria. En el caso de que las mediciones sean hechas en una superficie, entonces $Z(\mathbf{x})$ puede interpretarse como la variable aleatoria asociada a ese punto del plano ($\mathbf{x}$ representa las coordenadas, planas o geográficas, y Z la variable en cada una de ellas). Estas variables aleatorias pueden representar la magnitud de una variable ambiental medida en un conjunto de coordenadas de la región de estudio.

3.2.1. Momentos de una Variable Regionalizada.

Sea $\{Z(\mathbf{x}) : \mathbf{x}\epsilon D \subset R^d\}$ el proceso estocástico que define la variable regionalizada. Para cualesquier $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\ \epsilon D$ contenido en el espacio euclidiano R^d , el vector aleatorio $\mathbf{Z} = [Z(\mathbf{x_1}), Z(\mathbf{x_2}), ..., Z(\mathbf{x_n})]^T$ está definido por su función de distribución conjunta $F[z_1, z_2, ..., z_n] = P[Z(\mathbf{x_1}) \leq z_1, Z(\mathbf{x_2}) \leq z_2, ..., Z(\mathbf{x_n}) \leq z_n]$.

Conocidas las densidades marginales univariadas y bivariadas se pueden establecer los siguientes valores esperados:

$E(Z(\mathbf{x_i})) = m(\mathbf{x}_i)$

$V(Z(\mathbf{x_i})) = E\{[Z(\mathbf{x_i}) - m(\mathbf{x_i})]^2\} = \sigma_i^2$
$C(Z(\mathbf{x_i}), Z(\mathbf{x_j})) = E\{[Z(\mathbf{x_i}) - m(\mathbf{x_i})][Z(\mathbf{x_j}) - m(\mathbf{x_j})]\}$
$\gamma(Z(\mathbf{x_i}), Z(\mathbf{x_j})) = \frac{1}{2}V[Z(\mathbf{x_i}) - Z(\mathbf{x_j})] = \frac{1}{2}E[\{Z(\mathbf{x_i}) - Z(\mathbf{x_j})\}^2]$, siempre y cuando $m(x_i) = m(x_j)$

A este último se le denomina semivariograma.

3.3. Covarianza espacial

Obtener la covarianza para los posibles resultados de lanzar un dado, es fácil porque los resultados son independientes. Pero los valores de una variable regionalizada tienden a estar relacionados. En general, valores de una variable regionalizada en dos lugares cercanos uno del otro son similares, mientras que estos al estar en lugares ampliamente separados son menos parecidos. Se está familiarizado en usar la covarianza para determinar la relación entre 2 variables para observaciones pareadas. Así, para las observaciones $z_{i,1}$, $z_{i,2}$ para $i = 1, 2, ...n$, de las variables, $\mathbf{z_1}$, y $\mathbf{z_2}$, respectivamente,

$$\hat{C}(\mathbf{z_1}, \mathbf{z_2}) = \frac{1}{n}\sum_{i=1}^{n}\{z_{i,1} - \bar{z}_1\}\{z_{i,2} - \bar{z}_2\} \tag{3.1}$$

donde $\bar{z}_1$ y $\bar{z}_2$ son las medias de $\mathbf{z_1}$ y $\mathbf{z_2}$, respectivamente. Si las unidades $i = 1, 2, ...n$ fueran observadas aleatoriamente, entonces $\hat{C}(\mathbf{z_1}, \mathbf{z_2})$ estima la covarianza poblacional sin sesgo.

Esta definición se puede extender para relacionar 2 variables aleatorias. El concepto y su expresión matemática fueron desarrolladas originalmente para análisis de series de tiempo durante los años 1920 y 1930, y han sido muy usados para pronosticar. Tienen su analogía en el espacio, Yaglom (1987) los presenta en este contexto como predicción espacial.

En el nuevo conjunto espacial $\mathbf{z_1}$ y $\mathbf{z_2}$ se convierten en $Z(\mathbf{x_1})$ y $Z(\mathbf{x_2})$, es decir, son los conjuntos de valores de la misma propiedad, Z, en dos lugares $\mathbf{x_1}$ y $\mathbf{x_2}$ y la letra mayúscula, Z representa que son variables aleatorias.

Su covarianza es

$$C(\mathbf{x_1}, \mathbf{x_2}) = E[\{Z(\mathbf{x_1}) - \mu(x_1)\}\{Z(\mathbf{x_2}) - \mu(x_2)\}] \tag{3.2}$$

Donde $\mu(x_1)$ y $\mu(x_2)$ son las medias de Z en $\mathbf{x_1}$ y $\mathbf{x_2}$, respectivamente. La ecuación (3.2) es análoga a la ecuación (3.1). Desafortunadamente, su solución no está disponible, ya que solamente se tiene una realización de Z en cada punto. A diferencia de la ecuación

(3.1) donde se tiene n observaciones para cada una de las variables y por tanto se puede conocer su media, en la ecuación (3.2) no se conoce la media. Sólo suponiéndose que los valores en diferentes lugares son distintas observaciones del atributo, podría superarse el inconveniente encontrado, ya que la media sería constante con independencia de los puntos considerados. Esto es lo que se denomina estacionariedad (Moral 2002).

3.3.1. Estacionaridad

Estacionaridad significa que la distribución del proceso aleatorio tiene ciertos atributos que son los mismos en todos lados. Comenzando por el primer momento, se asume que la media $\mu = E[Z(\mathbf{x})]$ sobre la cual oscilan las realizaciones individuales, es la misma para todo $\mathbf{x}$. Esto nos permite reemplazar $\mu(x_1)$ y $\mu(x_2)$ por el valor singular μ, el cual se puede estimar para muestras repetidas.

La siguiente consideración es el segundo momento: si $\mathbf{x_1}$ y $\mathbf{x_2}$ coinciden, la ecuación (3.2) define la varianza, ya que si μ es la misma, entonces:

$$C(\mathbf{x}_1, \mathbf{x}_2) = C(\mathbf{x}, \mathbf{x}) = E[\{Z(\mathbf{x}) - \mu\}\{Z(\mathbf{x}) - \mu\}] = E[\{Z(\mathbf{x}) - \mu\}^2] = \sigma^2 \tag{3.3}$$

la cual se asume que es finita y al igual que la media, es la misma en todas partes. Cuando $\mathbf{x_1}$ y $\mathbf{x_2}$ no coinciden, su covarianza depende de su separación y no de su posición absoluta. Así, para algún par de puntos, $\mathbf{x_i}$ y $\mathbf{x_j}$ separados por el vector $\mathbf{h} = \mathbf{x_i} - \mathbf{x_j}$, se tiene

$$C(\mathbf{x}_i, \mathbf{x}_j) = E[\{Z(\mathbf{x}_i) - \mu\}\{Z(\mathbf{x}_j) - \mu\}] \tag{3.4}$$

la cual es constante para alguna distancia de separación $\mathbf{h}$. Esta consistencia de la media, varianza y covarianza, que depende sólo de la separación y no de la posición absoluta, es decir, la consistencia del primer y segundo momento del conjunto o proceso, constituye la estacionaridad de segundo orden ó estacionaridad débil. Notar que los momentos son del proceso aleatorio del cual se tiene una única realización y que nunca se podrá conocer sus valores exactamente.

De la misma manera en la que cada v.a. tiene su función de distribución, cada par de v.a.'s $Z(\mathbf{x_i})$ y $Z(\mathbf{x_j})$ tienen su función de distribución conjunta, la cual está dada por,

$F(z_i, z_j) = Prob[Z(\mathbf{x_i}) \leq z_i, Z(\mathbf{x_j}) \leq z_j]$, para cualquier par de valores reales z_i y z_j.

Además de tener asociada una función de densidad de probabilidades conjunta, que en el caso contínuo, es la derivada de F en los puntos (z_i, z_j) en donde F es derivable. Si $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ son puntos en R^2 (en el plano), entonces se espera que la función de

distribución conjunta F de $[Z(\mathbf{x_1}), Z(\mathbf{x_2})]$ sea la misma que la función de distribución conjunta de $[Z(\mathbf{x_3}), Z(\mathbf{x_4})]$, ..., y de $[Z(\mathbf{x_{n-1}}), Z(\mathbf{x_n})]$.

3.4. Función covarianza

Se puede escribir la ecuación (3.4) como:

$$\begin{aligned} cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] &= E[\{Z(\mathbf{x})-\mu\}\{Z(\mathbf{x}+\mathbf{h})-\mu\}] \\ &= E[\{Z(\mathbf{x})\}\{Z(\mathbf{x}+\mathbf{h})\} - \mu(Z(\mathbf{x})) - \mu(Z(\mathbf{x}+\mathbf{h})) + \mu^2] \\ &= E[\{Z(\mathbf{x})\}\{Z(\mathbf{x}+\mathbf{h})\}] - \mu E[(Z(\mathbf{x}))] - \mu E[(Z(\mathbf{x}+\mathbf{h}))] + \mu^2 \quad (3.5) \\ &= E[\{Z(\mathbf{x})\}\{Z(\mathbf{x}+\mathbf{h})\}] - \mu^2 - \mu^2 + \mu^2 \\ &= E[\{Z(\mathbf{x})\}\{Z(\mathbf{x}+\mathbf{h})\}] - \mu^2 \end{aligned}$$

Por lo tanto

$$cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] = E[\{Z(\mathbf{x})\}\{Z(\mathbf{x}+\mathbf{h})\}] - \mu^2$$

Esto es, la covarianza es una función sólo de la distancia **h**. El prefijo "auto" en la función de autocovarianza sirve para hacer referencia a que se tiene la covarianza de Z consigo misma, es decir, se tiene la misma característica medida en diferentes sitios o lugares. Lo anterior, describe la dependencia entre valores de $Z(\mathbf{x})$ con un cambio de distancia, refiriéndose a ella, simplemente como la **función covarianza**. Frecuentemente se usa la notación $C(\mathbf{h})$ para denotar a la $cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})]$. Es importante notar que es necesario que $Z(\mathbf{x})$ tenga segundo momento finito para que la covarianza exista y sea finita.

La autocovarianza depende de la escala en la que Z es medido, por lo que, frecuentemente es más conveniente utilizar a la función de autocorrelación, ya que ésta es una medida adimensional. La función de autocorrelación la denotamos por " $\rho(\mathbf{h})$ " y se define como,

$$\rho(\mathbf{h}) = \frac{cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})]}{\sigma_{\mathbf{x}+\mathbf{h}}\sigma_{\mathbf{x}}} = \frac{C(\mathbf{h})}{\sigma^2_{\mathbf{x}}} = \frac{C(\mathbf{h})}{C(\mathbf{0})} \quad (3.6)$$

Donde $C(\mathbf{0})$ es la covarianza a la longitud **0**.

3.5. Variación intrínseca y el semivariograma

Se puede representar un proceso aleatorio estacionario por el modelo

$$Z(\mathbf{x}) = \mu + \epsilon(\mathbf{x}) \tag{3.7}$$

Es decir, que el valor de Z en el lugar $\mathbf{x}$ es la media del proceso mas un componente aleatorio proveniente de una distribución con media cero y función de covarianza

$$\begin{aligned} C(\mathbf{h}) &= cov[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] \\ &= cov[\mu + \epsilon(\mathbf{x}), \mu + \epsilon(\mathbf{x}+\mathbf{h})] \\ &= cov[\varepsilon(\mathbf{x}), \varepsilon(\mathbf{x}+\mathbf{h})] \\ &= E[\{\varepsilon(\mathbf{x})\}\{\varepsilon(\mathbf{x}+\mathbf{h})\}] - \mu_\epsilon^2, \quad \mu_\epsilon = 0 \\ &= E[\{\varepsilon(\mathbf{x})\}\{\varepsilon(\mathbf{x}+\mathbf{h})\}] \end{aligned} \tag{3.8}$$

En la práctica, en algunas ocasiones no se cumplen las hipótesis de estacionaridad (Moral 2002). Por ejemplo, cuando hay una cierta tendencia no es posible asumir que la media es constante. El problema es que la media parezca cambiar a través de una región y que la varianza incremente sin límite así como incrementa el área de interés (la varianza no es finita). Otras veces, aunque la media sea constante, puede que la covarianza no exista. En este sentido la covarianza no puede ser definida. No se podría insertar un valor de μ en la ecuación (3.6).

Matheron (1965) reconoció el problema que esto traía y la solución que le dió fue una mayor contribución a la geoestadística práctica. Él consideró que, si en general, la media no es constante, sí se puede considerar como constante para $|\mathbf{h}|$ cercano a cero, en cuyo caso la diferencia esperada sería cero, es decir:

$$E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] = 0 \tag{3.9}$$

Él reemplazó la covarianza por la varianza de diferencias como medida de relación espacial, que al igual que la covarianza, depende de la distancia y no de la posición absoluta. Esto lleva a:

$$\begin{aligned} Var[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] &= E\{[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2\} - \{E[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]\}^2 \\ &= E\{[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2\} - 0 \\ &= E\{[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]^2\} \\ &= 2\gamma(\mathbf{h}) \end{aligned} \tag{3.10}$$

$\gamma(\mathbf{h})$ se define en la siguiente sección.

Las ecuaciones (3.10) y (3.11) constituyen las hipótesis intrínsecas de Matheron, las cuales son:

1. Z($\mathbf{x}$) tiene esperanza finita y constante para todo punto en el dominio. Lo que implica que la esperanza de los incrementos es cero.

2. Para cualquier vector $\mathbf{h}$, la varianza del incremento está definida y es una función que únicamente depende de la distancia $|h|$.

Una función aleatoria estacionaria de 2^{o} orden es siempre intrínseca; lo recíproco no es necesariamente cierto (Moral 2002). Si se restringe la validez de las condiciones anteriores a un entorno determinado y para distancias limitadas, se dice que la función aleatoria es cuasiestacionaria o cuasiintrínseca.

Este paso permitió a los practicantes no imponer la estacionaridad de segundo orden donde las suposiciones no se cumplían o eran dudosa. De esta manera se abrió un campo muy extenso de aplicaciones. $\gamma(\mathbf{h})$ es conocido como la semivarianza a la distancia de separación $\mathbf{h}$. El prefijo semi se refiere al hecho de que es la mitad de una varianza. Para este caso, es la mitad de la varianza de una diferencia. Como una función de $\mathbf{h}$ esto es el semivariograma, aún cuando usualmente se usa el variograma.

3.6. Equivalencia del semivariograma y la función covarianza

La existencia de la covarianza implica que la varianza existe, es finita y no depende de $\mathbf{h}$, es decir $V(Z(\mathbf{x_i})) = C(0) = \sigma^2$. Así mismo la estacionaridad de segundo orden implica la siguiente relación entre la función de covarianza y de la semivarianza.

$$
\begin{aligned}
\gamma(\mathbf{h}) &= \gamma[Z(\mathbf{x}+\mathbf{h}), Z(\mathbf{x})] \\
&= \frac{1}{2}E\{[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]^2\} \\
&= \frac{1}{2}E\{[Z(\mathbf{x}+\mathbf{h}) - \mu - Z(\mathbf{x}) + \mu]^2\} \\
&= \frac{1}{2}E\{[(Z(\mathbf{x}+\mathbf{h}) - \mu) - (Z(\mathbf{x}) - \mu)]^2\} \\
&= \frac{1}{2}E\{[Z(\mathbf{x}+\mathbf{h}) - \mu]^2 - 2[Z(\mathbf{x}+\mathbf{h}) - \mu][Z(\mathbf{x}) - \mu] + [Z(\mathbf{x}) - \mu]^2\} \\
&= \frac{1}{2}E\{[Z(\mathbf{x}+\mathbf{h}) - \mu]^2\} + \frac{1}{2}E\{[Z(\mathbf{x}) - \mu]^2\} - E\{[Z(\mathbf{x}+\mathbf{h}) - \mu][Z(\mathbf{x}) - \mu]\} \\
&= \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2 - C(\mathbf{h}) \\
&= \sigma^2 - C(\mathbf{h}) \\
&= C(0) - C(\mathbf{h})
\end{aligned}
\tag{3.11}
$$

El concepto de estacionaridad es muy útil en la modelación de series temporales (Box & Jenkins, 1976). En este contexto es fácil la identificación, puesto que sólo hay una dirección de variación (el tiempo). En el campo espacial existen múltiples direcciones y por lo tanto se debe asumir que en todas, el fenómeno es estacionario. Cuando la esperanza de la variable no es la misma en todas las direcciones o cuando la covarianza o correlación dependan del sentido en que se calculan, no habrá estacionariedad. Si la correlación entre los datos no depende de la dirección en la que ésta se calcule se dice que el fenómeno es **isotrópico**, en caso contrario se hablará de **anisotropía**.

En casos prácticos resulta compleja la identificación de la estacionaridad. La isotropía es estudiada a través del cálculo de funciones de autocovarianza o de semivarianza muestrales en varias direcciones. Si éstas tienen formas considerablemente distintas puede no ser válido el supuesto de isotropía

3.7. Características de las funciones de correlación espacial

A continuación se considera las características más importantes de las funciones de autocorrelación, covarianza y el variograma (Webster, 2001).

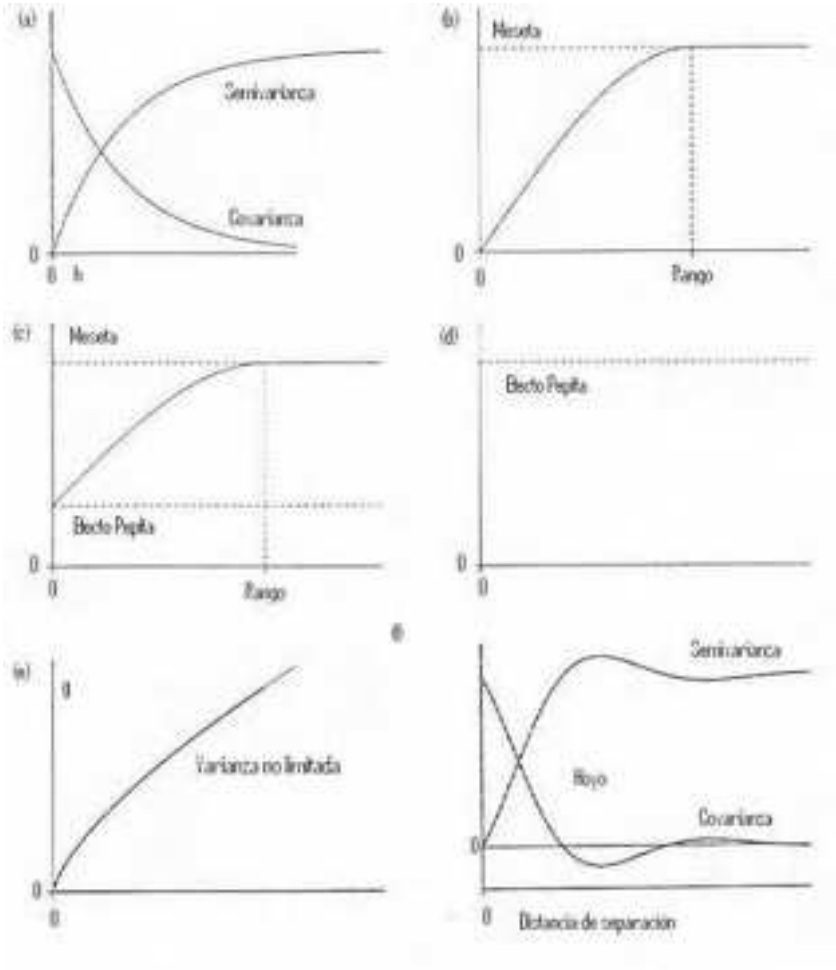


Figura 3.1: Funciones teóricas para correlación espacial: (a) variograma típico y función de covarianza equivalente; (b) variograma limitado mostrando la meseta y el rango; (c) variograma no limitado con efecto pepita; (d) variograma efecto pepita puro; (e) variograma no limitado; (f) variograma y función covarianza ilustrando el efecto hoyo.

Autocorrelación: Al igual que el coeficiente de correlación ordinario, la función de autocorrelación varía entre -1 y 1. De la ecuación (3.7) su valor a la longitud 0 es 1.

Simetría: De la suposición de estacionaridad:

$$\begin{aligned} C(\mathbf{h}) &= E[\{Z(\mathbf{x}) - \mu\}\{Z(\mathbf{x}+\mathbf{h}) - \mu\}] \\ &= E[\{Z(\mathbf{x}+\mathbf{h}) - \mu\}\{Z(\mathbf{x}) - \mu\}] \\ &= C(-\mathbf{h}) \end{aligned} \tag{3.12}$$

Es decir, la autocovarianza es simétrica en el espacio. Lo mismo es cierto para el variograma $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$ para todo $\mathbf{h}$. Esto se cumple para las tres funciones. Lo anterior significa basta considerar solamente las longitudes positivas de las funciones.

Positiva semidefinida: La matriz de covarianza para algún número de puntos es positiva semidefinida. Esto es, su determinante:

$$\begin{vmatrix} C(\mathbf{x_1},\mathbf{x_1}) & C(\mathbf{x_1},\mathbf{x_2}) & \dots & C(\mathbf{x_1},\mathbf{x_n}) \\ C(\mathbf{x_2},\mathbf{x_1}) & C(\mathbf{x_2},\mathbf{x_2}) & \dots & C(\mathbf{x_2},\mathbf{x_n}) \\ \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{x_n},\mathbf{x_1}) & C(\mathbf{x_n},\mathbf{x_2}) & \dots & C(\mathbf{x_n},\mathbf{x_n}) \end{vmatrix}$$

y todos sus menores principales son positivos o cero. Esto es necesario porque la varianza de alguna suma lineal de las variables aleatorias.

$$Y(\mathbf{x}) = \lambda_1 Z(\mathbf{x}_1) + \lambda_2 Z(\mathbf{x}_2) + ... + \lambda_n Z(\mathbf{x}_n) \tag{3.13}$$

debe ser positivo o cero; una varianza no puede ser negativa. La covarianza y autocorrelación son funciones positivas semidefinidas. De la misma manera, el variograma debe ser negativo semidefinido.

Continuidad: Muchas variables medioambientales son variables continuas; los procesos estocásticos que se utilizan para representarlos son continuos, así también lo son las funciones de autocovarianza y variograma de una longitud continua. Crucialmente, $C(\mathbf{h})$ y $\gamma(\mathbf{h})$ son continuas en $\mathbf{h} = \mathbf{0}$ (continuas en el origen), y si esto es así, deben ser continuos en cualquier lado. Así $C(\mathbf{h})$ declina de algún valor positivo $C(\mathbf{0}) = \sigma^{\mathbf{2}}$ a 0 (valores más pequeños a distancias de separación mayores), ver Figura 3.1(a). Dada la simetría, el va-

riograma incrementa de 0 en $\mathbf{h} = \mathbf{0}$ (y alcanza a la varianza); es decir, debe pasar a través del origen si el proceso es continuo, ver Figura 3.1 (a)-(b).

Si esto no se cumple, entonces se debe tener una secuencia continua de posiciones en el espacio de valores los cuales no están relacionados. Esto se manifiesta de manera evidente en el variograma muestral; los valores calculados parecen alcanzar algún valor positivo en la ordenada cuando la distancia $\mathbf{h}$ se aproxima a $\mathbf{0}$, ver Figura 3.1(c). Esta discontinuidad es conocida como efecto pepita.

Monótona creciente: Los variogramas de la Figura 3.1(b)-(c) son funciones monótonas crecientes, es decir, la varianza incrementa con incrementos en la distancia de separación. Los valores de $\gamma(\mathbf{h})$ para $|\mathbf{h}|$ cortos, muestran que los $Z(\mathbf{x})$ son similares y cuando $|\mathbf{h}|$ incrementa, $Z(\mathbf{x})$ y $Z(\mathbf{x}+\mathbf{h})$ disminuyen, en promedio, su similitud. Visto desde el punto de vista de correlación, $\rho(\mathbf{h})$ se incrementa cuando la distancia de separación se acorta y se dice que el proceso está autocorrelacionado ó es espacialmente dependiente.

Efecto Pepita: El semivariograma por definición es nulo en el origen, pero como se mencionó antes, las funciones obtenidas pueden presentar discontinuidad en el origen, a esta discontinuidad se le llama efecto pepita y se representa por C_0.

Meseta y rango: Los semivariogramas de procesos estacionarios de segundo orden alcanzan cotas superiores a partir de las que $\gamma(\mathbf{h})$ permanece constante aún con el aumento de h, como en la Figura 3.1(b)-(c). Esta cota, el máximo, es conocido como meseta. Si la meseta no es finita, el semivariograma define un fenómeno natural que cumple sólo con la hipótesis intrínseca. La meseta se define por C_1 o por $(C_0 + C_1)$ cuando existe efecto pepita.

La meseta puede obtenerse trazando una línea paralela a la abscisa que se ajuste a los puntos de mayor valor del semivariograma, su valor se lee en la intersección de esta línea con la ordenada.

Un semivariograma puede alcanzar su meseta a una distancia de separación finita, en tal caso, se dice que el variograma tiene un rango, también conocido como el rango de correlación. El rango es la distancia a la cual la autocorrelación se convierte en 0, Figura 3.1(c). Esta distancia marca el límite de la dependencia espacial. Algunos variogramas alcanzan su meseta asintóticamente, por lo que no tienen estrictamente un rango. En tales casos y para propósitos prácticos su rango efectivo es usualmente tomado como la distancia de separación a la cual alcanzan 0.95 por ciento de su meseta.

Variograma no limitado: Si, como en la Figura 3.1(e), el semivariograma incrementa de manera indefinida con incrementos en la distancia de separación entonces el proceso no es estacionario de segundo orden. El proceso puede ser intrínseco pero la covarianza no existe.

Efecto hoyo: En algunos casos el semivariograma decrece de su máximo a un mínimo local luego crece nuevamente, ver 3.3(f). Este máximo es equivalente a un mínimo en la función covarianza, la cual parece como un hoyo. Esta forma se repite regularmente en el proceso. Un variograma que continúa fluctuando en una forma de onda con incrementos en la distancia de separación significa mayor regularidad.

Anisotropía: La variación espacial no es necesariamente la misma en todas las direcciones. Si el proceso es anisotrópico, entonces lo es el variograma, como lo es la función de covarianza, si existe. La anisotropía puede tomar varias formas. La pendiente puede variar. Si el semivariograma tiene una meseta entonces la variación en la pendiente permitirá determinar la variación en el rango. Si la variación con dirección es tal que una simple transformación de las coordenadas espaciales lo remueve, entonces se tiene una anisotropía geométrica. La anisotropía geométrica está presente cuando los semivariogramas en diferentes direcciones tienen la misma meseta pero distintos alcance.

Cuando los semivariogramas tienen diferentes mesetas y alcances en diferentes direcciones, se dice que existe anisotropía zonal, puede ser corregido separando el semivariograma en sus componentes isotrópicos horizontal y anisotrópico vertical.

Tendencia: Indica que los valores medidos aumentan o disminuyen dramáticamente en la zona estudiada con el aumento de la distancia. Esto puede ser resuelto aplicando polinomios a la ecuación del semivariograma.

Capítulo 4

FASES DE UN ESTUDIO GEOESTADISTICO

En todo trabajo geoestadístico se distinguen tres etapas:

1. Análisis exploratorio de los datos: En esta fase se estudian los datos muestrales sin tener en cuenta su distribución geográfica. Es una etapa de aplicación de la estadística. Se comprueba la consistencia de los datos, eliminándose aquellos que sean erróneos, y se identifican las distribuciones de las cuales provienen.

2. Análisis estructural: Se estudia la continuidad espacial de la variable. En esta etapa se calcula el variograma experimental, o cualquier otra función que nos explique la variabilidad espacial, se ajusta a los datos un variograma teórico y se analiza e interpreta dicho ajuste al modelo paramétrico seleccionado.

3. Predicciones: Estimaciones de la variable estudiada en los puntos no muestrales, considerando la estructura de correlación espacial seleccionada e integrando la información obtenida de forma directa en los puntos muestrales, así como la obtenida indirectamente en forma de tendencias conocidas u observadas. También se pueden realizar simulaciones, teniendo en cuenta los patrones de continuidad espacial elegidos.

La fase del análisis estructural es la más crítica. La obtención de modelos geoestadísticos realistas conlleva un estudio riguroso del semivariograma o de cualquier función análoga que caracterice la variación espacial del atributo. En diversos trabajos, suelen usarse diferentes algoritmos geoestadísticos sin analizar previamente las posibles estructuras espaciales, tomándose, por defecto, los métodos que existen en los programas utilizados. La estética de los resultados suele esconder ese gran error.

4.1. El Semivariograma

Se ha visto que el semivariograma es una gráfica que describe la diferencia esperada entre pares de muestras separados por una distancia **h**, o lo que es lo mismo, la varianza de los incrementos de la variable regionalizada en las localizaciones separadas una distancia **h**. Ahora se discutirá cómo calcular la función de semivariograma experimental.

Cuando se definió la estacionaridad débil en el capítulo anterior se mencionó que se asumía que la varianza de los incrementos de la variable regionalizada era finita. A esta función denotada por $2\gamma(\mathbf{h})$ se le denomina variograma. Utilizando la definición teórica de la varianza en términos del valor esperado de una variable aleatoria, se tiene:

$$\begin{aligned} 2\gamma(\mathbf{h}) &= var[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})] \\ &= E[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))^2] - (E[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))])^2 \\ &= E[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))^2] - 0 \\ &= E[(Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h}))^2] \end{aligned} \tag{4.1}$$

La mitad del variograma, $\gamma(\mathbf{h})$, se conoce como la función de semivarianza y caracteriza las propiedades de dependencia espacial del proceso, midiendo el grado de correlación existente entre los valores de la variable en cada punto y la distancia entre aquellos. Dada una realización del fenómeno, la función de semivarianza es estimada, por el método de momentos, a través del semivariograma experimental, se calcula mediante (Wackernagel, 1995):

$$\bar{\gamma}(\mathbf{h}) = \frac{1}{2m(h)} \sum_{i=1}^{m(h)} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + h)]^2 \tag{4.2}$$

Donde $m(\mathbf{h})$ es el número de pares de datos puntuales separados por el vector particular de distancia **h**, **h** es el incremento, $Z(\mathbf{x}_i)$ son los valores experimentales y $\mathbf{x}_i$ son las localizaciones donde son medidos los valores $Z(\mathbf{x}_i)$. El semivariograma responde a la pregunta: ¿Qué tan parecidos son los puntos en el espacio a medida que estos se encuentran más alejados?.

La forma más fácil de desplegar los valores del semivariograma es graficándolos. Poniendo en el eje de las abscisas las distancias de separación **h** entre las muestras y en la ordenada el valor del semivariograma para cada distancia de separación. Por definición, **h** comienza en cero por lo que $\bar{\gamma}(\mathbf{h})$ también comienza en cero, ya que es un promedio de valores al cuadrado (Clarck, 2001). Si se considera el caso cuando h es igual a cero (Clarck, 2001) y

se toma dos muestras exactamente en la misma posición y medimos sus valores, la diferencia entre las dos debe ser cero, así que $\gamma(\mathbf{h})$ y $\bar{\gamma}(\mathbf{h})$ deben pasar siempre por el origen. Al suponer que las dos muestras se mueven un poco separándose una cierta distancia, se espera alguna diferencia entre los dos valores, por lo que el semivariograma tendrá algún valor positivo pequeño. Cuando las muestras se van separando a distancias cada vez mayores, la diferencia va creciendo. En el caso ideal cuando las distancias son muy grandes los valores muestrales se vuelven independientes y el valor del semivariograma será mas o menos constante, alcanzando su meseta como se mencionó anteriormente.

A continuación se presenta un ejemplo ilustrativo del cálculo de la función de semivarianza experimental: Suponga que se tienen mediciones de una variable hipotética cuyos valores están comprendidos entre 20 y 50 unidades que dependen de la característica que se esta midiendo y su configuración en una región de estudio es como se presenta en el esquema de la Figura 4.1. Como se indica en la representación, la distancia entre cada par de puntos contiguos es de 100 unidades y se calculará bajo esta situación el semivariograma experimental. Por simplicidad se calcularán sólo los semivariogramas en sentido (izquierda-derecha) e (inferior-superior), debido a que para obtener un semivariograma experimental en el que se tenga en cuenta, además de la distancia, la orientación, se requiere calcular la distancia euclidiana entre los pares de puntos dirigidos hacia la orientación elegida.

En primer lugar en sentido izquierda-derecha se encuentran todas las parejas de puntos que están a una distancia de 100 unidades y se calcula el semivariograma como:

$$\begin{aligned}\bar{\gamma}(100) =& \frac{1}{2*54}[(45-41)^2+(41-43)^2+(43-24)^2+\ldots+(44-33)^2+(33-30)^2\\ &+(27-23)^2+(23-38)^2+(38-33)^2+\ldots+(30-32)^2+(32-41)^2\\ &+(31-46)^2+(46-43)^2+(43-21)^2+\ldots+(37-40)^2+(40-37)^2\\ &+(20-25)^2+(25-32)^2+(32-36)^2+\ldots+(40-33)^2+(33-41)^2\\ &+(40-46)^2+(46-20)^2+(20-34)^2+\ldots+(20-50)^2+(50-42)^2\\ &+(33-47)^2+(47-49)^2+(49-33)^2+\ldots+(41-28)^2+(28-31)^2]\\ &=69{,}11\end{aligned}$$

Análogamente para la distancia de 200 unidades

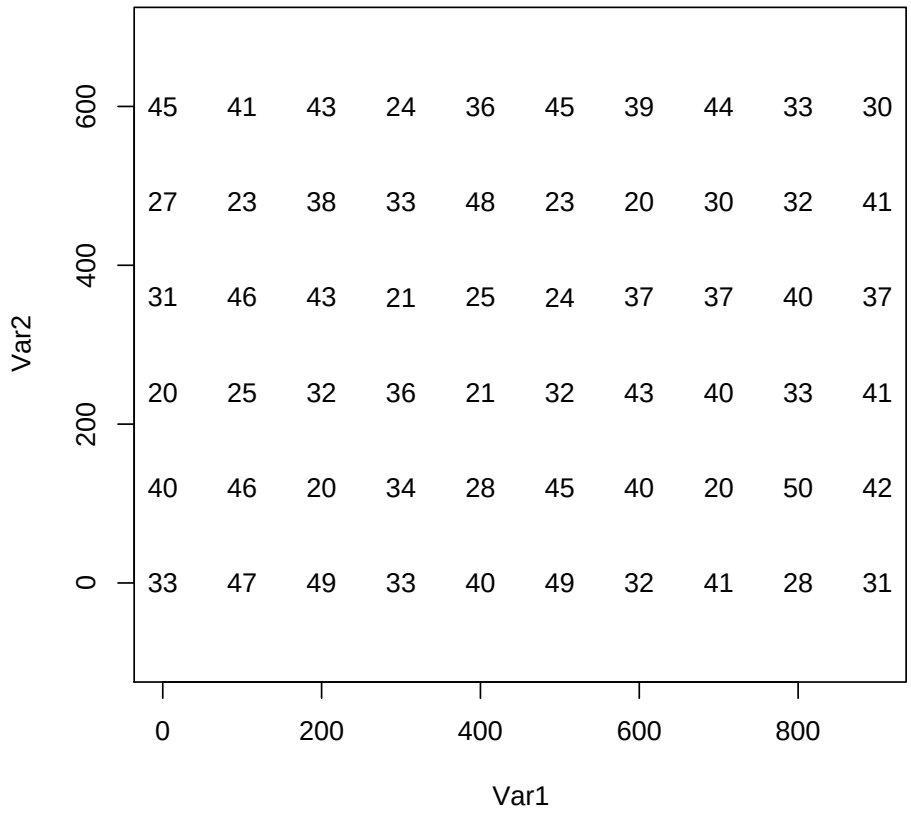


Figura 4.1: 60 datos generados aleatoriamente entre 20 y 50 para ejemplificar el cálculo del semivariograma experimental

Cuadro 4.1: Valores de la función de semivarianza experimental en dos direcciones para el conjunto de datos hipotéticos.

Distancia	Semivarianza Sentido Izquierda-Derecha	Semivarianza Sentido Inferior-Superior
100	69.11	88.89
200	84.72	74.76
300	80.20	94.68
400	80.9	90.17

$$\begin{aligned}\bar{\gamma}(200) =& \frac{1}{2*48}[(45-43)^2+(41-24)^2+(43-36)^2+\ldots+(39-33)^2+(44-30)^2\\ &+(27-38)^2+(23-33)^2+(38-48)^2+\ldots+(20-32)^2+(30-41)^2\\ &+(31-43)^2+(46-21)^2+(43-25)^2+\ldots+(37-40)^2+(37-37)^2\\ &+(20-32)^2+(25-36)^2+(32-21)^2+\ldots+(43-33)^2+(40-41)^2\\ &+(40-20)^2+(46-34)^2+(20-28)^2+\ldots+(40-50)^2+(20-42)^2\\ &+(33-49)^2+(47-33)^2+(49-40)^2+\ldots+(32-28)^2+(41-31)^2]\\ &=84{,}73\end{aligned}$$

Similarmente se procede para otras distancias y para el sentido inferior-superior. Los valores calculados del semivariograma se muestran en la Tabla 4.1 y la gráfica del semivariograma se presenta en la Figura 4.2.

4.1.1. Modelos teóricos de semivarianza

Cada semivarianza calculada para una separación particular es solamente un estimador de una semivarianza media para esa separación y como tal está sujeta a error. Este error, que crece grandemente por las fluctuaciones muestrales, da al variograma experimental una apariencia mas o menos inestable. El verdadero variograma que representa la variación regional es continua y es el variograma que realmente se quiere conocer.

Por lo que una vez que se han definido los puntos del variograma experimental, será necesario ajustar un modelo teórico a dichos puntos. Esto se debe a la imposibilidad de trabajar con el semivariograma experimental, que carece de una función matemática precisa (o al menos difícil de caracterizar), y a la necesidad de extender los valores del semivariograma más allá de la distancia máxima definida. También, por el contrario, será

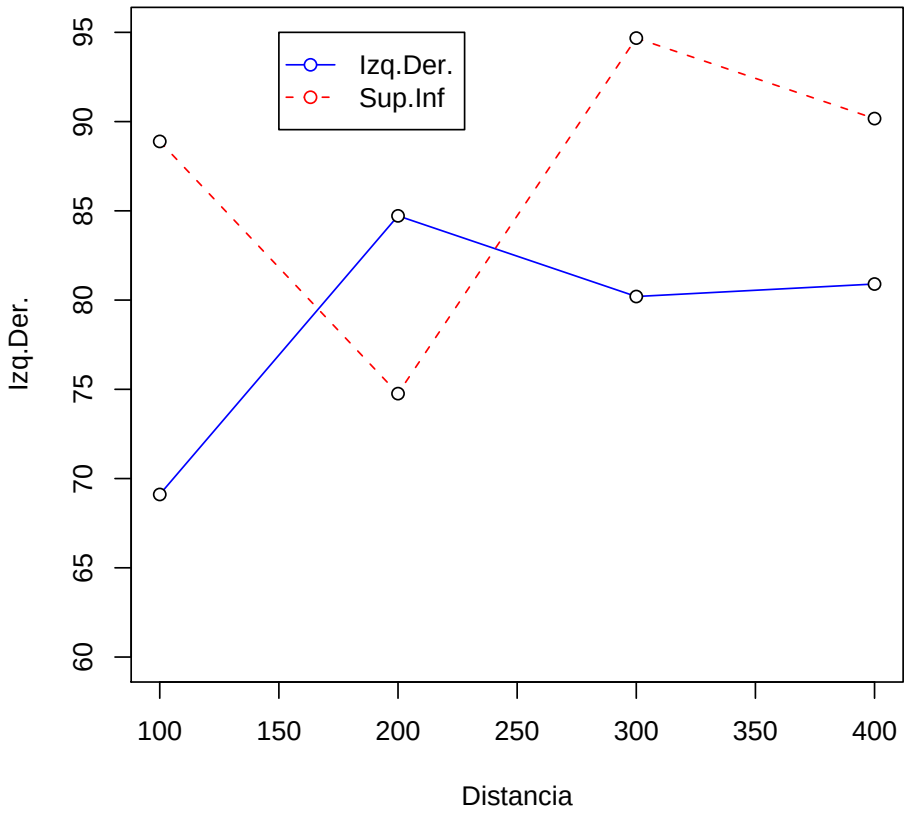


Figura 4.2: Función de semivarianza experimental en dos direcciones para un conjunto de datos hipotéticos.

necesario la extensión del variograma a distancias próximas a cero y así también para poder cuantificar el grado y escala de variación espacial. Este modelo teórico será utilizado posteriormente para interpolar en el espacio el valor de la variable en sitios no muestreados, ya que para ello se requiere semivariogramas a separaciones para las cuales no se tienen comparaciones directamente y dicho modelo ayudará para su cálculo.

Existen numerosos modelos que se utilizan en geostadística, siendo los más comúnmente usados el modelo esférico, el modelo exponencial, el modelo gaussiano y el modelo lineal.

Modelo Esférico: El modelo esférico para el semivariograma, está dado matemáticamente como:

$$\gamma(\mathbf{h}) = \begin{cases} C\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^2}\right) & \text{si} \quad h \leq a \\ \\ C & \text{si} \quad h > a \end{cases}$$

Donde C es el valor de la meseta, a el rango y h la distancia. Este modelo tiene un crecimiento rápido cerca al origen, pero los incrementos marginales van decreciendo para distancias grandes, hasta que para distancias superiores al rango los incrementos son nulos.

Modelo Exponencial:

$$\gamma(\mathbf{h}) = C\left[1 - exp\left(\frac{-h}{a}\right)\right] \tag{4.3}$$

Alcanza la meseta asintóticamente. Se considera que el rango, a, es la distancia para la cual el valor del variograma es del 95 % de la meseta. Como el esférico, muestra un crecimiento lineal próximo al origen; sin embargo, crece de forma más rápida y luego se estabiliza más gradualmente.

Modelo Gaussiano:

$$\gamma(\mathbf{h}) = C\left[1 - exp\left(\frac{-h^2}{a^2}\right)\right] \tag{4.4}$$

Al igual que el modelo exponencial, el modelo tiende a alcanzar la meseta asintóticamente, y el rango se define como la distancia a la cual el variograma alcanza el 95 % de la meseta. El principal distintivo de este modelo es su forma parabólica cerca al origen.

Modelo lineal y lineal generalizado: Hay otros modelos que no tienen meseta. Su uso puede ser delicado debido a que en algunos casos indican la presencia de no estacionariedad en alguna dirección (Clarck, 2001).

Su fórmula matemática es la siguiente:

$$\gamma(\mathbf{h}) = C_0 + bh^{\alpha} \quad 0 < \alpha < 2 \tag{4.5}$$

Obviamente, cuando el parámetro α es igual a uno el modelo es lineal. Para $0 < \alpha < 2$ el modelo es lineal generalizado

4.2. Predicción espacial

Muchas propiedades de los fenómenos podrían ser medidos en un número infinito de lugares, pero en la práctica son medidos en pocos, principalmente por razones de economía. Si se está interesado en conocer sus valores en otra parte, se deben estimar de los datos que se pueden obtener. Lo mismo se cumple si se quiere estimar sobre áreas mas grandes para las cuales no ha sido posible medir u observar las propiedades directamente. Por lo que se considera el problema general de estimar valores en lugares no muestreados. La estimación es la tarea para la cual la geoestadística fue inicialmente desarrollada, y generalmente es llamada **kriging**.

Kriging provee una solución al problema de estimación basado en un modelo contínuo de variación espacial estocástica. Esto hace el mejor uso del conocimiento existente tomando en cuenta la manera en que una propiedad varía en el espacio a través del modelo del variograma. En su formulación original el kriging estima el valor en un lugar simplemente como una suma lineal o promedio ponderado de los datos en su vecindad. Los pesos son asignados a los datos muestrales dentro de la vecindad del punto o bloque a ser estimado de tal manera que minimice la varianza de la estimación (o varianza del kriging) y las estimaciones son insesgadas.

Antes de la descripción de los principios de la estimación geoestadística, conviene revisar una serie de consideraciones acerca de la estimación en sí. Existen diversos métodos de estimación, cuyo uso dependerá del tipo de problema que se trate de resolver y del conocimiento que se tenga del mismo. Previamente a la elección de un método particular, se debe estar en condiciones de determinar las siguientes cuestiones:

1. La estimación a realizar, ¿será local o global?

2. ¿Se desea una estimación puntual o para extensiones mayores, en bloques?

Mediante los métodos de estimación geoestadística, conocidos como krigeado o krigeaje (kriging en la literatura inglesa, en honor de Danie Krige, quien formuló por primera

vez esta metodología en 1951), se puede responder a las dos cuestiones planteadas, ya que contempla todas esas posibilidades. Sin embargo, en ocasiones, ciertos métodos de estimación tradicionales generan unos resultados muy semejantes a los del krigeado (Isaaks y Srivastava, 1989), sobre todo cuando los datos son abundantes.

Las principales características que hacen del krigeado un método de estimación muy superior a los tradicionales (inverso ponderado de la distancia, la triangulación, el poligonal, etc.), son:

1. Mediante el krigeado se puede obtener estimaciones mayores o menores que la de los datos muestrales. Con otros métodos los valores estimados se limitan al intervalo definido por los datos muestrales.

2. Mientras que los métodos tradicionales utilizan el concepto euclidiano de la distancia para el cálculo de los pesos que se aplicarán a cada dato muestral, el krigeado considera tanto la distancia como la geometría de la localización de las muestras.

3. Mediante el krigeado se minimiza la varianza del error esperado (diferencia entre el valor real y el estimado). Como el valor real en un punto no muestral es desconocido, el krigeado emplea un modelo conceptual con una función aleatoria asociada a los valores reales.

Además, los métodos geoestadísticos muestran una gran flexibilidad para la interpolación, pudiéndose estimar valores puntuales o en bloques, así como métodos para incorporar información secundaria que esté relacionada con la variable principal. Todos estos métodos dan lugar a unas superficies muy suaves, además de una estimación de la varianza en todos los puntos, lo cual no puede realizarse con otros métodos de interpolación.

La idea fundamental del krigeado es consecuencia de los conceptos relacionados con la dependencia espacial, tratados en el apartado anterior: los lugares que disten menos entre sí tendrán unos valores de los atributos más semejantes que los correspondientes a los puntos o bloques que estén más separados. En la naturaleza, esto suele cumplirse y, además, las variables naturales generalmente se distribuyen de una forma continua.

4.2.1. Teoría de kriging ordinario

El krigeado ordinario se caracteriza por considerar que se producen fluctuaciones locales de la media, limitando el dominio de estacionaridad de la misma a un ámbito local: $m(\mathbf{x})$ =constante, pero desconocida.

Suponer que se hacen n mediciones de la región de estudio de la variable de interés Z en los puntos $\mathbf{x}_i$, $i = 1, 2, ..., n$. Así, se tienen realizaciones de las variables $Z(\mathbf{x}_1)$, $Z(\mathbf{x}_2)$, $\ldots$,$Z(\mathbf{x}_n)$. Se desea predecir en el punto donde no hubo medición, sea este punto $Z(\mathbf{x}_0)$. En esta circunstancia, el método kriging ordinario propone que el valor de la variable puede predecirse como una combinación lineal de las n variables aleatorias, así:

$$\begin{aligned} Z^*(\mathbf{x}_0) &= \lambda_1 Z(\mathbf{x}_1) + \lambda_2 Z(\mathbf{x}_2) + \lambda_3 Z(\mathbf{x}_3) + \ldots + \lambda_n Z(\mathbf{x}_n) \\ &= \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) \end{aligned} \tag{4.6}$$

En donde los λ_i representan los pesos o ponderaciones de los valores originales. Dichos pesos se calculan en función de la distancia entre los puntos muestreados y el punto donde se va a hacer la correspondiente predicción. La suma de los pesos debe ser igual a uno para que la esperanza del predictor sea igual a la esperanza de la variable. Esto último se conoce como el requisito de insesgamiento.

Estadísticamente la propiedad de insesgamiento se expresa a través de:

$E[Z^*(\mathbf{x}_0)] = E[Z(\mathbf{x}_0)]$

Para demostrar que la suma de las ponderaciones debe ser igual a 1, se asume que el proceso es estacionario de media μ (desconocida) y haciendo uso de las propiedades del valor esperado:

$E[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)] = \mu$

$\sum_{i=1}^{n} \lambda_i E[Z(\mathbf{x}_i)] = \mu$

$\sum_{i=1}^{n} \lambda_i \mu = \mu$

$\Rightarrow \sum_{i=1}^{n} \lambda_i = 1$

Se dice que $Z^*(\mathbf{x}_0)$ es el mejor predictor lineal, en este caso, porque los pesos se obtienen de tal manera que minimicen la varianza del error de predicción, es decir que minimicen la expresión:

$V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)]$

Esta última es la característica distintiva de los métodos kriging, ya que existen otros métodos de interpolación como el de distancias inversas o el poligonal, que no garantizan varianza mínima de predicción (Samper y Carrera, 1990).

La estimación de los pesos se obtiene minimizando:

$V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)]$ sujeto a $\sum_{i=1}^{n} \lambda_i = 1$

Se tiene que

$$V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = V[Z^*(\mathbf{x}_0)] - 2COV[Z^*(\mathbf{x}_0), Z(\mathbf{x}_0)] + V[Z(\mathbf{x}_0)]$$

Desagregando las componentes de la ecuación anterior se obtiene lo siguiente:

$$V[Z^*(\mathbf{x}_0)] = V[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j COV[Z(\mathbf{x}_i), Z(\mathbf{x}_j)]$$

En adelante se usará la siguiente notación: $COV[Z(\mathbf{x}_i), Z(\mathbf{x}_j)]] = C_{ij}$ y $V[Z(\mathbf{x}_0)] = \sigma^2$

De lo anterior:

$$\begin{aligned} COV[Z^*(\mathbf{x}_0), Z(\mathbf{x}_0)] &= COV[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i), Z(\mathbf{x}_0)] \\ &= \sum_{i=1}^{n} \lambda_i COV[Z(\mathbf{x}_i), Z(\mathbf{x}_0)] \\ &= \sum_{i=1}^{n} \lambda_i C_{i0} \end{aligned}$$

Entonces reemplazando, se tiene que:

$$V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2 \sum_{i=1}^{n} \lambda_i C_{i0} + \sigma^2 \tag{4.7}$$

Luego se debe minimizar la función anterior sujeta a la restricción $\sum_{i=1}^{n} \lambda_i = 1$. Este problema de minimización con restricciones se resuelve mediante el método de multiplicadores

de Lagrange.

$\sigma_k^2 = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^n \lambda_i C_{i0} + \sigma^2 + 2\mu[\sum_{i=1}^n \lambda_i - 1]$

Siguiendo el procedimiento acostumbrado para obtener valores extremos de una función, se deriva e iguala a cero, en este caso con respecto a λ_i y μ

$$\frac{d(\sigma_k^2)}{d\lambda_1} = 2\lambda_1 C_{11} + 2\sum_{j=2}^{n} \lambda_j C_{1j} - 2C_{10} + 2\mu = 2\sum_{j=1}^{n} \lambda_j C_{1j} - 2C_{10} + 2\mu = 0 \tag{4.8}$$

$\Rightarrow \sum_{j=1}^n \lambda_j C_{1j} + \mu = C_{10}$

De manera análoga se determinan las derivadas con respecto a $\lambda_2, \ldots, \lambda_n$:

$$\frac{d(\sigma_k^2)}{d\lambda_2} = 2\lambda_2 C_{22} + 2\sum_{j=1, j\neq 2}^{n} \lambda_j C_{2j} - 2C_{20} + 2\mu = 2\sum_{j=1}^{n} \lambda_j C_{2j} - 2C_{20} + 2\mu = 0 \tag{4.9}$$

$\Rightarrow \sum_{j=1}^n \lambda_j C_{2j} + \mu = C_{20}$

$\vdots$

$$\frac{d(\sigma_k^2)}{d\lambda_n} = 2\lambda_n C_{nn} + 2\sum_{j=1}^{n-1} \lambda_j C_{nj} - 2C_{n0} - 2\mu = 2\sum_{j=1}^{n} \lambda_j C_{nj} - 2C_{n0} + 2\mu = 0 \tag{4.10}$$

$\Rightarrow \sum_{j=1}^n \lambda_j C_{nj} + \mu = C_{n0}$

Por último derivamos con respecto a μ

$$\frac{d(\sigma_k^2)}{d\mu} = 2\sum_{i=1}^{n} \lambda_i - 2 = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i = 1 \tag{4.11}$$

De (4.8), (4.9), (4.10), (4.11) resulta un sistema de (n + 1) ecuaciones con (n + 1) incógnitas, que matricialmente puede ser escrito como:

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

$$\mathbf{C}_a * \lambda = \mathbf{C_0}$$

Donde $\mathbf{C}_a$ es la matriz de covarianzas aumentada, $\mathbf{C_0} = (C_{10}, C_{20}, \dots, C_{n0}, 1)'$ y $\lambda = (\lambda_\mathbf{1}, \lambda_\mathbf{2}, \dots, \lambda_\mathbf{n}, \mu)'$

Los pesos que minimizan el error de predicción se determinan mediante la función de covarianza a través de

$$\lambda = \mathbf{C}_a^{-1} * \mathbf{C}_0$$

Encontrando los pesos se calcula la predicción en el punto $\mathbf{x}_0$. De forma análoga se procede para cada punto donde se quiere hacer predicción.

Varianza de Predicción del Kriging Ordinario

Multiplicando (4.8), (4.9) y (4.10) por λ_i se obtiene:

$\lambda_i\{\sum_{j=1}^n \lambda_j C_{ij} + \mu\} = \lambda_i C_{i0}$ para todo $i = 1, 2, \dots, n$

Sumando las n ecuaciones

$\sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j C_{ij} + \sum_{j=1}^n \lambda_i \mu = \sum_{i=1}^n \lambda_i C_{i0}$ para todo $i = 1, 2, \dots, n$

$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_{ij} = \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu$

Sustituyendo la expresión anterior en (4.7)

$$\sigma_k^2 = \sigma^2 + \sum_{i=1}^n \lambda_i C_{i0} - \sum_{i=1}^n \lambda_i \mu - 2\sum_{i=1}^n \lambda_i C_{i0}$$

$$\sigma_k^2 = \sigma^2 - \sum_{i=1}^{n} \lambda_i C_{i0} - \mu \tag{4.12}$$

Estimación de Ponderaciones por medio de la Función de Semivarianza

Los pesos λ_i pueden ser estimados a través de la función de semivarianza, para lo cual se debe recordar la relación entre las funciones de covariograma y de semivarianza establecida en la ecuación (3.12). Antes de esto es conveniente tener en cuenta la siguiente notación:

$\sigma^2 = V[Z(\mathbf{x})]$, $\gamma_{ij} = \gamma(\mathbf{h})$, donde $\mathbf{h}$ es la distancia entre los puntos i y j y análogamente $C_{ij} = C(\mathbf{h})$

La relación entre las dos funciones en cuestión es la siguiente:

$\gamma(h) = \sigma^2 - C(\mathbf{h})$

Por lo que:

$$\gamma_{ij} = \sigma^2 - C_{ij} \Rightarrow C_{ij} = \sigma^2 - \gamma_{ij} \tag{4.13}$$

Reeemplazando esta relación en (4.8),(4.9) y (4.10) se determinan los pesos óptimos λ_i en términos de la función de semivarianza:

$$\begin{aligned}
\frac{d(\sigma_k^2)}{d\lambda_1} &= \sum_{j=1}^{n} \lambda_j C_{1j} + \mu - C_{10} = \sum_{j=1}^{n} \lambda_j (\sigma^2 - \gamma_{1j}) + \mu - (\sigma^2 - \gamma_{10}) \\
&= \sigma^2 \sum_{j=1}^{n} \lambda_j - \sum_{j=1}^{n} \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
&= \sigma^2 - \sum_{j=1}^{n} \lambda_j \gamma_{1j} + \mu - \sigma^2 + \gamma_{10} \\
&= -\sum_{j=1}^{n} \lambda_j \gamma_{1j} + \mu + \gamma_{10}
\end{aligned}$$

De manera que, al igualar a cero la derivada, se obtiene que,

$$\sum_{j=1}^{n} \lambda_j \gamma_{1j} - \mu = \gamma_{10}$$

Similarmente,

$\frac{d(\sigma_k^2)}{d\lambda_2} = -\sum_{j=1}^{n} \lambda_j \gamma_{2j} + \mu + \gamma_{20}$

Al igualar a cero la derivada, se obtiene que
$\sum_{j=1}^{n} \lambda_j \gamma_{2j} - \mu = \gamma_{20}$
$\vdots$
$\frac{d(\sigma_k^2)}{d\lambda_n} = -\sum_{j=1}^{n} \lambda_j \gamma_{nj} + \mu + \gamma_{n0}$

Al igualar a cero la derivada, se obtiene que
$\sum_{j=1}^{n} \lambda_j \gamma_{nj} - \mu = \gamma_{n0}$

El sistema de ecuaciones se completa con (4.11). De acuerdo con lo anterior los pesos se obtienen en términos del semivariograma a través del sistema de ecuaciones:

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma 10 \\ \gamma_2 0 \\ \vdots \\ \gamma_n 0 \\ 1 \end{pmatrix}$$

Para establecer la expresión de la correspondiente varianza del error de predicción en términos de la función de semivarianza se reemplaza (4.13) en (4.12), de donde:

$$\begin{aligned}\sigma_k^2 &= \sigma^2 - \sum_{i=1}^{n} \lambda_i C_{i0} - \mu \\ &= \sigma^2 - \sum_{i=1}^{n} \lambda_i (\sigma^2 - \gamma_{i0}) - \mu \\ &= \sigma^2 - \sigma^2 \sum_{i=1}^{n} \lambda_i + \sum_{i=1}^{n} \lambda_i \gamma_{i0} - \mu \\ &= \sum_{i=1}^{n} \lambda_i \gamma_{i0} - \mu\end{aligned}$$

Validación del kriging

Existen diferentes métodos para evaluar la bondad de ajuste del modelo de semivariograma elegido con respecto a los datos muestrales y por ende de las predicciones hechas con kriging. El más empleado es el de validación cruzada, que consiste en excluir la observación de uno de los n puntos muestrales y con los $n-1$ valores restantes y el modelo de semivariograma escogido, predecir vía kriging el valor de la variable en estudio en la ubicación del punto que se excluyó. Se piensa que si el modelo de semivarianza elegido describe adecuadamente la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el valor predicho debe ser pequeña. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de n "errores de predicción" . Los parámetros del modelo a validar (pepita, meseta y rango) se van modificando en un procedimiento de prueba y error hasta obtener los estadísticos de validación cruzada adecuados. Estos estadísticos son los siguientes:

Media de los errores de estimación (MEE)

$$MEE = \frac{1}{n} \sum_{i=1}^{n} [Z^*(\mathbf{x}_i) - Z(\mathbf{x}_i)] \tag{4.14}$$

donde $Z^*(\mathbf{x}_i)$ es el valor estimado de la variable de interés en el punto $\mathbf{x}_i$ y $Z(\mathbf{x}_i)$ es el valor medido de la variable de interés en el punto $\mathbf{x}_i$ y n es el número de puntos muestrales utilizado en la interpolación. La MEE no debe ser significativamente distinta de 0 (prueba de t), en cuyo caso, indicaría que el modelo de semivariograma permite el cálculo de estimaciones no sesgadas.

Error cuadrado medio (ECM)

$$ECM = \frac{1}{n}\sum_{i=1}^{n}[(Z^*(\mathbf{x}_i) - Z(\mathbf{x}_i))^2] \tag{4.15}$$

Un modelo de semivariograma se considera adecuado si, como regla práctica, el ECM es menor que la varianza de los valores muestrales.

Una forma descriptiva de hacer la validación cruzada es mediante un gráfico de dispersión de los valores observados contra los valores predichos. En la medida en que la nube de puntos se ajuste más a una línea recta que pasa por el origen, mejor será el modelo de semivariograma utilizado para realizar el kriging.

Representación de las predicciones

Una vez que se ha hecho la predicción en un conjunto de puntos diferentes de los muestrales vía kriging, se debe elaborar un mapa que dé una representación global del comportamiento de la variable de interés en la zona estudiada. Los más empleados son los mapas de contornos, los mapas de residuos y los gráficos tridimensionales. En el caso de los mapas de contornos, en primer lugar se divide el área de estudio en un enmallado y se hace la predicción en cada uno de los nodos de este mismo. Posteriormente se unen los valores predichos con igual valor, generando así las líneas de contorno (isolíneas de distribución). Este gráfico permite identificar la magnitud de la variable en toda el área de estudio. Es conveniente acompañar el mapa de interpolaciones de la variable con los correspondientes mapas de isolíneas de los errores y de las varianzas de predicción (posiblemente estimados a través de métodos matemáticos), con el propósito de identificar zonas de mayor incertidumbre respecto a las predicciones.

Ilustración

Suponga que se tiene una configuración de datos como la que se presenta en la Figura 4.3; se ha etiquetado el punto que se va a estimar como la localización 0, y las localizaciones de las muestras de 1 a 7. Las coordenadas de estos ocho puntos están en la Tabla 4.2, junto con los valores muestrales disponibles.

Para calcular los pesos del kriging ordinario, primero se debe decidir qué modelo de continuidad espacial se quiere que tenga el modelo de función aleatoria. Para conservar este ejemplo relativamente simple, se calcularán todas las covarianzas de la siguiente función.

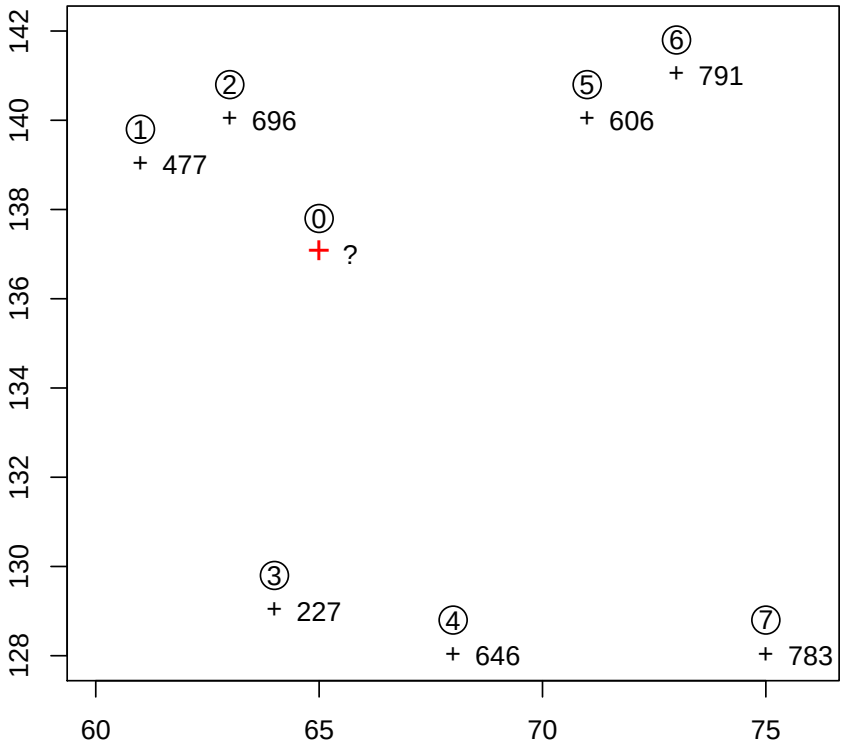


Figura 4.3: Un ejemplo de una configuración de datos para ilustrar el estimador kriging. El valor muestral es dado a la derecha del signo mas

Cuadro 4.2: Coordenadas y valores muestrales para los datos mostrados en la Figura 4.4

Número de muestra	X	Y	V	Distancia de 65E,137N
1	61	139	477	4.5
2	63	140	696	3.6
3	64	129	227	8.1
4	68	128	646	9.5
5	71	140	606	46.7
6	73	141	791	8.9
7	75	128	783	13.5

Cuadro 4.3: Tabla de distancias euclidianas, de la Figura 4.3, entre todos los pares posibles de localizaciones de los 7 datos

Localización	0	1	2	3	4	5	6	7
0	0.00	4.47	3.61	8.06	9.49	6.71	8.94	13.45
1	4.47	0.00	2.24	10.44	13.04	10.05	12.17	17.80
2	3.61	2.24	0.00	11.05	13.00	8.00	10.05	16.97
3	8.06	10.04	11.05	0.00	4.12	13.04	15.00	11.05
4	9.49	13.04	13.00	4.12	0.00	12.37	13.93	7.00
5	6.71	10.05	8.00	13.04	12.37	0.00	2.24	12.65
6	8.94	12.17	10.05	15.00	13.93	2.24	0.00	13.15
7	13.45	17.80	16.97	11.05	7.00	12.65	13.15	0.00

$$\tilde{C}(\mathbf{h}) = \begin{cases} C_0 + C_1 & \text{si } |\text{h}|=0 \\ \\ C_1 exp(\frac{-3|h|}{a}) & \text{si } |\text{h}|>0 \end{cases} \tag{4.16}$$

Usando la Ecuación 4.13, esta función de covarianza corresponde al siguiente variograma:

$$\tilde{\gamma}(\mathbf{h}) = \begin{cases} 0 & \text{si } |\text{h}|=0 \\ \\ C_0 + C_1(1 - exp(\frac{-3|h|}{a})) & \text{si } |\text{h}|>0 \end{cases} \tag{4.17}$$

Los Geoestadísticos normalmente definen la continuidad espacial en su modelo de función aleatoria a través del variograma y resuelven el sistema de kriging ordinario usando la covarianza. En este ejemplo se usará la función de covarianza.

Para usar la covarianza de la ecuación 4.16, se ignorará la posibilidad de anisotropía; la covarianza entre valores de datos de cualesquiera dos localizaciones dependerán sólo de la distancia entre ellos y no de la dirección. Para desarrollar los pasos del kriging ordinario, se usarán los siguientes parámetros para la función de la ecuación 4.16:

$C_0 = 0$, $a = 10$, $C_1 = 10$ (pepita cero, meseta 10 y rango 10)

Ahora la función de covarianza tiene la siguiente expresión:

$$\tilde{C}(\mathbf{h}) = 10exp(\frac{-3|h|}{10}) = 10e^{-0,3|h|} \tag{4.18}$$

Habiendo escogido una función de covarianza de la cual se puede calcular todas las covarianza requeridas para el modelo de función aleatoria, se construirán ahora las matrices $\mathbf{C}_a$ y $\mathbf{C_0}$. Usando la tabla 4.3, la cual provee las distancias entre cada par de localizaciones y Ecuación 4.16, la matriz $\mathbf{C}_a$ es:

$$\mathbf{C}_a = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & C_{17} & 1 \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} & C_{27} & 1 \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} & C_{37} & 1 \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} & C_{47} & 1 \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} & C_{57} & 1 \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} & C_{67} & 1 \\ C_{61} & C_{72} & C_{73} & C_{74} & C_{65} & C_{76} & C_{77} & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 10{,}00 & 5{,}11 & 0{,}44 & 0{,}20 & 0{,}49 & 0{,}26 & 0{,}05 & 1{,}00 \\ 5{,}11 & 10{,}00 & 0{,}36 & 0{,}20 & 0{,}91 & 0{,}49 & 0{,}06 & 1{,}00 \\ 0{,}44 & 0{,}36 & 10{,}00 & 2{,}90 & 0{,}20 & 0{,}11 & 0{,}36 & 1{,}00 \\ 0{,}20 & 0{,}20 & 2{,}90 & 10{,}00 & 0{,}24 & 0{,}15 & 1{,}22 & 1{,}00 \\ 0{,}49 & 0{,}91 & 0{,}20 & 0{,}24 & 10{,}00 & 5{,}11 & 0{,}22 & 1{,}00 \\ 0{,}26 & 0{,}49 & 0{,}11 & 0{,}15 & 5{,}11 & 10{,}00 & 0{,}19 & 1{,}00 \\ 0{,}05 & 0{,}06 & 0{,}36 & 1{,}22 & 0{,}22 & 0{,}19 & 10{,}00 & 1{,}00 \\ 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 & 0{,}00 \end{pmatrix}$$

El vector $\mathbf{C_0}$ es

$$\mathbf{C_0} = \begin{pmatrix} C_{10} \\ C_{20} \\ C_{30} \\ C_{40} \\ C_{50} \\ C_{60} \\ C_{70} \\ 1 \end{pmatrix} = \begin{pmatrix} 2{,}61 \\ 3{,}39 \\ 0{,}89 \\ 0{,}58 \\ 1{,}34 \\ 0{,}68 \\ 0{,}18 \\ 1{,}00 \end{pmatrix}$$

La inversa de $\mathbf{C}_a$ es

$$\mathbf{C}_a^{-1} = \begin{pmatrix} 0{,}127 & -0{,}077 & -0{,}013 & -0{,}009 & -0{,}008 & -0{,}009 & -0{,}012 & 0{,}136 \\ -0{,}077 & 0{,}129 & -0{,}010 & -0{,}008 & -0{,}015 & -0{,}008 & -0{,}011 & 0{,}121 \\ -0{,}013 & -0{,}010 & 0{,}098 & -0{,}042 & -0{,}010 & -0{,}010 & -0{,}014 & 0{,}156 \\ -0{,}009 & -0{,}008 & -0{,}042 & 0{,}102 & -0{,}009 & -0{,}009 & -0{,}024 & 0{,}139 \\ -0{,}008 & -0{,}015 & -0{,}010 & -0{,}009 & 0{,}130 & -0{,}077 & -0{,}012 & 0{,}118 \\ -0{,}009 & -0{,}008 & -0{,}010 & -0{,}009 & -0{,}077 & 0{,}126 & -0{,}013 & 0{,}141 \\ -0{,}012 & -0{,}011 & -0{,}014 & -0{,}024 & -0{,}012 & -0{,}013 & 0{,}085 & 0{,}188 \\ 0{,}136 & 0{,}121 & 0{,}156 & 0{,}139 & 0{,}118 & 0{,}141 & 0{,}188 & -2{,}180 \end{pmatrix}$$

El conjunto de pesos que proveerá estimadores insesgados con mínima varianza es calculado multiplicando $\mathbf{C}_a^{-1} * \mathbf{C}_0$.

$$W = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ \mu \end{pmatrix} = \mathbf{C}_a^{-1} * \mathbf{C}_0 = \begin{pmatrix} 0{,}173 \\ 0{,}318 \\ 0{,}129 \\ 0{,}086 \\ 0{,}151 \\ 0{,}057 \\ 0{,}086 \\ 0{,}907 \end{pmatrix}$$

La Figura 4.4 muestra los valores muestrales con sus correspondientes pesos. El estimador resultante es:

$$\begin{aligned} Z_0^* &= \sum_{i=1}^{n} \lambda_i Z_i \\ &= (0{,}173)(477) + (0{,}318)(696) + (0{,}129)(227) + (0{,}086)(646) \\ &+ (0{,}151)(606) + (0{,}057)(791) + (0{,}086)(783) \\ &= 592{,}7 \end{aligned} \tag{4.19}$$

La varianza de estimación minimizada es

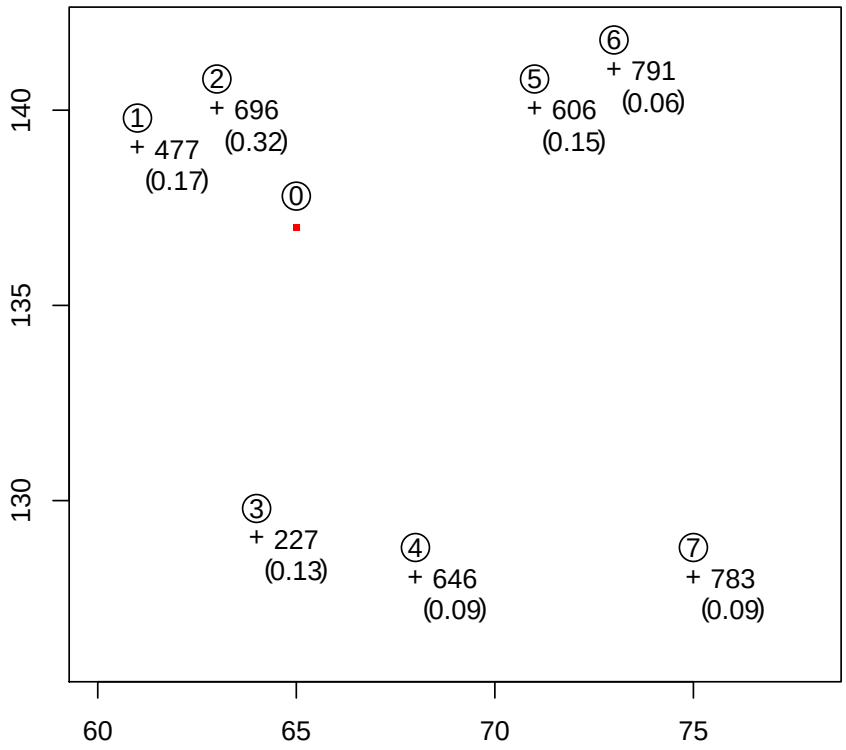


Figura 4.4: Los pesos del kriging ordinario para las siete muestras usando el modelo de covarianza exponencial isotrópico dado en la Ecuación 4.16. El valor muestral está a la derecha del signo "mas" mientras los pesos del Kriging son mostrados en paréntesis

$$\begin{aligned}\sigma_k^2 &= \sigma^2 - \sum_{i=1}^{n} \lambda_i C_{i0} - \mu \\ &= 10 - (0{,}173)(2{,}61) - (0{,}318)(3{,}39) - (0{,}129)(0{,}89) - (0{,}086)(0{,}58) \\ &- (0{,}151)(1{,}34) - (0{,}057)(0{,}68) - (0{,}086)(0{,}18) - 0{,}907 \\ &= 7{,}14\end{aligned} \tag{4.20}$$

4.2.2. Kriking Simple

Suponga que hay una variable regionalizada estacionaria con media (m) y covarianza conocidas. De manera análoga a como se define en modelos lineales (por ejemplo en diseño de experimentos) el modelo establecido en este caso es igual a la media más un error aleatorio con media cero. La diferencia es que en este caso los errores no son independientes.

Sea $Z(\mathbf{x})$ la variable de interés medida en el sitio $\mathbf{x}$.

$E[Z(\mathbf{x})] = \mu$
$Z(\mathbf{x}) = \mu + \epsilon(\mathbf{x})$, con $E[\epsilon(\mathbf{x})] = 0$

El predictor de la variable de interés en un sitio $\mathbf{x}_0$ donde no se tiene información se define como:

$$Z^*(\mathbf{x}_0) = \mu + \epsilon^*(\mathbf{x}_0)$$

Aquí $\epsilon^*(\mathbf{x}_0)$ corresponde a la predicción del error aleatorio en el sitio $\mathbf{x}_0$. Despejando de la ecuación anterior $\epsilon^*(\mathbf{x}_0) = Z^*(\mathbf{x}_0) - \mu$.

El predictor del error aleatorio se define por:

$$\epsilon^*(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i) = \sum_{i=1}^{n} \lambda_i (Z(\mathbf{x}_i) - \mu)$$

de donde el predictor de la variable de estudio es:

$$Z^*(\mathbf{x}_0) = \mu + [\textstyle\sum_{i=1}^{n} \lambda_i (Z(\mathbf{x}_i) - \mu)] = \mu + \sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i)$$

El predictor es insesgado si:

$E[Z^*(\mathbf{x}_0)] = E[Z(\mathbf{x}_0)] = \mu$. Luego el predictor será insesgado cuando $E[\epsilon^*(\mathbf{x}_0)] = 0$

$E[\epsilon^*(\mathbf{x}_0)] = \sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i) = \sum_{i=1}^{n} \lambda_i(0) = 0$. Por consiguiente, a diferencia del kriging ordinario, en este caso no existen restricciones para las ponderaciones tendientes al cumplimiento de la condición de insesgamiento.

La estimación de los pesos del método kriging ordinario se obtiene de tal forma que se minimice $V[\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)]$.

$$
\begin{aligned}
V[\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)] &= E[\{\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)\}^2] \\
&= E[\{(\sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i)) - \epsilon(\mathbf{x}_0)\}^2] \\
&= E[(\sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i))(\sum_{j=1}^{n} \lambda_j \epsilon(\mathbf{x}_j))] - 2E[(\sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i))(\epsilon(\mathbf{x}_0))] + E[(\epsilon(\mathbf{x}_0))^2] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j E[(\epsilon(\mathbf{x}_i))(\epsilon(\mathbf{x}_j))] - 2 \sum_{i=1}^{n} \lambda_i E[(\epsilon(\mathbf{x}_i))(\epsilon(\mathbf{x}_0))] + E[(\epsilon(\mathbf{x}_0))^2]
\end{aligned}
$$

Usando:

1. $E[\epsilon(\mathbf{x}_0)] = 0$
2. $E[(\epsilon(\mathbf{x}_i))(\epsilon(\mathbf{x}_j))] = COV[\epsilon(\mathbf{x}_i), (\epsilon(\mathbf{x}_j))] = C_{ij}$
3. $E[(\epsilon(\mathbf{x}_o))^2] = \sigma^2$

$$
V[\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2 \sum_{i=n}^{n} \lambda_i C_{i0} + \sigma^2 \tag{4.21}
$$

derivando con respecto a λ_1 se tiene:

$$\frac{dV[\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)]}{d\lambda_1} = \frac{d}{d\lambda_1}(\lambda_1^2 C_{11} + 2\lambda_1 \sum_{j=2}^{n} \lambda_j C_{1j} + \sum_{i=2}^{n}\sum_{j=2}^{n} \lambda_i \lambda_j C_{ij} - 2\lambda_1 C_{10} - 2\sum_{j=2}^{n} \lambda_i C_{i0} + \sigma^2)$$

$$= 2\lambda_1 C_{11} + 2\sum_{j=2}^{n} \lambda_j C_{1j} - 2C_{10}$$

$$= 2\sum_{i=1}^{n} \lambda_i C_{1i} - 2C_{10}$$

Igualando a cero

$$\sum_{i=1}^{n} \lambda_i C_{1i} = C_{10}$$

En general para cualquier $i = 1, 2, \ldots, n$, se obtiene:

$$\sum_{j=1}^{n} \lambda_j C_{ij} = C_{i0} \tag{4.22}$$

Con las n ecuaciones resultantes se construye el siguiente sistema de ecuaciones:

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

De manera que si la matriz de covarianzas es invertible entonces el vector de lamndas es:

$\lambda = \mathbf{C}_a^{-1} * \mathbf{C}_0$.

Varianza de Predicción Kriging Simple

Se tiene de (4.21) que:

$$V[\epsilon^*(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)] = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sigma^2$$

$$\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} \lambda_j C_{ij} - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sigma^2$$

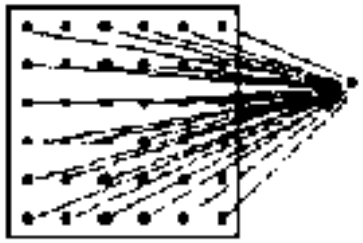


Figura 4.5: Enmallado regular de puntos dentro del bloque. La covarianza de un punto al bloque corresponde a la covarianza promedio entre el punto muestreado i y todos los puntos dentro del bloque

reemplazando (4.22) en esta última expresión se tiene que:

$\sigma_k^2 = \sum_{i=1}^n \lambda_i C_{i0} - 2\sum_{i=1}^n \lambda_i C_{i0} + \sigma^2$
$\sigma_k^2 = \sigma^2 - \sum_{i=1}^n \lambda_i C_{i0}$

4.2.3. Kriging en Bloques

En los dos métodos kriging hasta ahora descritos el objetivo ha estado centrado en la predicción puntual. A menudo, sin embargo, se requiere estimar un bloque, o más precisamente, estimar el valor promedio de la variable dentro de un área local.

El valor promedio dentro del bloque es estimado por :

$$\bar{Z}(A) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i) \tag{4.23}$$

Del sistema de ecuaciones para el kriging ordinario se tiene:

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

$$\mathbf{C}_a * \lambda = \mathbf{C_0}$$

Consecuentemente el vector del lado derecho de la igualdad en el sistema de arriba debe modificarse para incluir las covarianzas respecto al bloque. La covarianza de un

punto al bloque corresponde a la covarianza promedio entre el punto muestreado i y todos los puntos dentro del bloque (en la práctica un enmallado regular de puntos dentro del bloque es usado como se muestra en la Figura 4.6). El sistema de ecuaciones del kriging en bloques está dado por:

$$\begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{pmatrix} = \begin{pmatrix} \bar{C}_{1A} \\ \bar{C}_{2A} \\ \vdots \\ \bar{C}_{nA} \\ 1 \end{pmatrix}$$

$$\mathbf{C}_a * \lambda = \mathbf{C_0}$$

donde el vector de covarianzas al lado derecho de la igualdad en el sistema anterior contiene las covarianzas entre las variables $Z(\mathbf{x}_1)$, $Z(\mathbf{x}_2)$, ..., $Z(\mathbf{x}_n)$ y el bloque A donde se quiere hacer la estimación.

$\bar{C_{iA}} = \frac{1}{|A|} \sum_{j/j \epsilon A} C_{iA}$

La varianza del error de predicción del kriging en bloques está dada por:
$\sigma^2 = \bar{C}_{AA} - (\sum_{i=1}^{n} \lambda_i \bar{C}_{iA} + \mu)$, con $\bar{C}_{AA} = \frac{1}{|A|^2} \sum_{i/i \epsilon A} \sum_{j/j \epsilon A} C_{ij}$ igual a la covarianza entre pares de puntos dentro del bloque.

4.2.4. Kriging Universal

En los supuestos hechos hasta ahora respecto a los métodos kriging se ha asumido que la variable regionalizada es estacionaria (al menos se cumple con la hipótesis intrínseca). En muchos casos, la variable no satisface estas condiciones y se caracteriza por exhibir una tendencia. Por ejemplo en hidrología los niveles piezométricos de una acuífero pueden mostrar una pendiente global en la dirección del flujo (Samper y Carrera, 1990). Para tratar este tipo de variables es frecuente descomponer la variable $Z(\mathbf{x})$ como la suma de la tendencia, tratada como una función determinística, más una componente estocástica estacionaria de media cero. Asumir que:

$Z(\mathbf{x}) = m(\mathbf{x}) + \epsilon(\mathbf{x})$

con $E[\epsilon(\mathbf{x})] = 0$, $V[\epsilon(\mathbf{x})] = \sigma^2$ y por consiguiente $E[Z(\mathbf{x})] = m(\mathbf{x})$

La tendencia puede expresarse mediante:

$m(\mathbf{x}) = \sum_{l=1}^{p} a_l f_l(\mathbf{x})$

donde las funciones $f_l(\mathbf{x})$ son conocidas y p es el número de términos empleados para ajustar $m(\mathbf{x})$.

El predictor kriging universal se define como:

$Z^*(\mathbf{x}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)$

este será insesgado si:

$E[Z^*(\mathbf{x}_0)] = m(\mathbf{x}_0)$

$E[\sum_{i=1}^{n} \lambda_i Z(\mathbf{x}_i)] = m(\mathbf{x}_0)$

$(\sum_{i=1}^{n} \lambda_i m(\mathbf{x}_i)) = m(\mathbf{x}_0)$

$\sum_{i=1}^{n} \lambda_i (\sum_{l=1}^{p} a_l f_l(\mathbf{x}_i)) = \sum_{l=1}^{p} a_l f_l(\mathbf{x}_0)$

$\sum_{l=1}^{p} a_l (\sum_{i=1}^{n} \lambda_i f_l(\mathbf{x}_i)) = \sum_{l=1}^{p} a_l f_l(\mathbf{x}_0) \rightarrow \sum_{i=1}^{n} \lambda_i f_l(\mathbf{x}_i) = \sum_{l=1}^{p} f_l(\mathbf{x}_0)$

La obtención de los pesos en el kriging universal, análogo a los otros métodos kriging, se hace de tal forma que la varianza del error de predicción sea mínima.

$$
\begin{aligned}
V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] &= E[(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))^2] \\
&= E[((\sum_{i=1}^{n} \lambda_i (m(\mathbf{x}_i) - \epsilon(\mathbf{x}_i))) - (m(\mathbf{x}_0) - \epsilon(\mathbf{x}_0)))^2] \\
&= E\{[(\sum_{i=1}^{n} \lambda_i m(\mathbf{x}_i) - m(\mathbf{x}_0)) + (\sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i) - \epsilon(\mathbf{x}_0))]^2\} \\
&= E[(\sum_{i=1}^{n} \lambda_i \epsilon(\mathbf{x}_i) - \epsilon(\mathbf{x}_0))^2] \\
&= \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} \lambda_j E[(\epsilon(\mathbf{x}_i)\epsilon(\mathbf{x}_j))] - 2\sum_{i=1}^{n} \lambda_i E[\epsilon(\mathbf{x}_i)\epsilon(\mathbf{x}_0)] + E[(\epsilon(\mathbf{x}_0))^2]
\end{aligned}
$$

Usando

$C_{ij} = COV[\epsilon(\mathbf{x}_i), \epsilon(\mathbf{x}_j)]$
$\sigma^2 = E[(\epsilon(\mathbf{x}_i))^2]$

se tiene

$V[Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sigma^2$

Luego incluyendo la restricción dada por la condición de insesgamiento, se debe minimizar:

$\sigma_{ku}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j C_{ij} - 2\sum_{i=1}^{n} \lambda_i C_{i0} + \sigma^2 + \sum_{l=1}^{p} \mu_l [\sum_{i=1}^{n} \lambda_i f_l(\mathbf{x}_i) - f_l(\mathbf{x}_0)]$

o en términos de la función de semivarianza

$\sigma_{ku}^2 = -\sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j \gamma_{ij} + 2\sum_{i=1}^{n} \lambda_i \gamma_{i0} + \sum_{l=1}^{p} \mu_l [\sum_{i=1}^{n} \lambda_i f_l(\mathbf{x}_i) - f_l(\mathbf{x}_0)]$

derivando la expresión anterior respecto a λ_1, λ_2, ..., λ_n, μ_1, μ_2, ..., μ_p e igualando a cero las correspondientes derivadas se obtienen las siguientes ecuaciones:

$\sum_{j=1}^{n} \lambda_j \gamma_{ij} + \sum_{l=1}^{p} \mu_l f_l(\mathbf{x}_i) = \gamma_{i0} \; i = 1, 2, \ldots, n$
$\sum_{j=1}^{n} \lambda_j f_l(\mathbf{x}_j) = f_l(\mathbf{x}_0) \; j = 1, 2, \ldots, p$

en términos matriciales

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & f_{11} & \dots & f_{p1} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & f_{12} & \dots & f_{p2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & f_{1n} & \dots & f_{pn} \\ f_{11} & f_{12} & \dots & f_{1n} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ f_{p1} & f_{p2} & \dots & f_{pn} & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ f_{10} \\ \vdots \\ f_{p0} \end{pmatrix}$$

donde $f_{lj} = f_l(\mathbf{x}_j)$ es la l-ésima función en el punto j-ésimo.

La varianza de predicción del kriging universal está dada por (Samper y Carrera, 1990):

$\sigma^2_{ku} = \sum_{i=1}^{n} \lambda_i \gamma_{i0} + \sum_{l=1}^{p} \mu_l f_l(\mathbf{x}_0)$.

Nótese que si $p = 1$ y $f_l(\mathbf{x}) = 1$, el sistema de ecuaciones del kriging universal y la varianza de predicción coinciden con las del kriging ordinario. En este orden de ideas puede decirse que el kriging ordinario es un caso particular del kriging universal.

Capítulo 5

EL PAQUETE geoR PARA EL ANÁLISIS DE DATOS GEOESTADÍSTICOS

5.1. Introducción

El paquete geoR provee funciones para el análisis de datos geoestadísticos usando el software R. Este documento ilustra algunas de las capacidades del paquete. El objetivo es familiarizar al lector con los comandos de geoR para analizar datos y mostrar algunos de los resultados gráficos que pueden ser producidos. Los comandos usados aquí son solo ilustrativos, proveen ejemplos básicos del uso del paquete. No se realizará un análisis definitivo del conjunto de datos usado a lo largo de los ejemplos ni se cubrirán todos los detalles de las capacidades del paquete. En lo que sigue:

1. Los comandos de R son mostrados en *una fuente de escritura como este*
2. el resultado correspondiente, si lo hay, es mostrado en `una fuente de escritura como este`

Típicamente, los argumentos son usados por default por las funciones llamadas pero el usuario puede inspeccionar otros argumentos de las funciones usando las funciones `args` y `help`. Por ejemplo, para ver todos los argumentos para la función `variog` escribir `args(variog)` y/o `help(variog)`.

5.2. Comenzando una sesión y cargando los datos.

5.2.1. Instalando los paquetes sp y geoR

El paquete **geoR** trabaja conjuntamente con el paquete **sp**. Ambos pueden ser descargados de internet, al igual que el programa **R**. Más específicamente **R** puede ser descargado de la página http://www.r-project.org

sp puede ser descargado de la página http://www.r-project.org

geoR puede ser descargado de la página http://www.r-project.org

Para lograr que el paquete **geoR** funcione correctamente se debe bajar la versión 2.7.2 de **R** y una vez instalado se debe instalar el paquete **sp**. Los archivos del paquete **sp** al igual que el de **geoR** vienen en una carpeta comprimida. En la Figura 5.1 se muestra los pasos para intalar el paquete **sp** al programa. Sólo se indica la instalación del paquete sp pero es lo mismo para instalar el paquete **geoR**.

5.2.2. Cargando los paquetes sp y geoR

En la Figura 5.2 se muestra el procedimiento para cargar el paquete **sp**.

Al hacer lo mismo para cargar el paquete **geoR** se estará en condiciones de escribir los comandos que abajo se teclean y así poder hacer el análisis geoestadístico a un conjunto de datos.

Otra manera de cargar cualquier paquete es el siguiente: después de comenzar una sesión en **R**, se escribe el comando `library` o `require` y entre paréntesis se escribe el paquete que se desea cargar. Así, una vez instalado el paquete **geoR** éste se carga de la siguiente manera:

```
> library(geoR)
```

Si el paquete es cargado correctamente un mensaje será desplegado. Estos dos paquetes se deben cargar siempre que se inicie una sesión en R y se desee analizar datos espaciales, al contrario de su instalación, que es una sola vez.

5.2.3. Trabajando con datos

Típicamente, los datos son guardados como un objeto (o lista) de la clase `geodata`

Un objeto de este tipo de clase contiene dos elementos obligatorios: las coordenadas de la locación de los datos como primer elemento (`$coords`) y los valores de los datos como

Figura 5.1: Instalando el paquete sp. Hacer click en Instalar paquete(s) a partir de archivos Zip locales. Abrir la carpeta contenedora del archivo Zip. Seleccionar el archivo Zip y hacer clic nuevamente en Abrir para que el paquete quede instalado automáticamente.

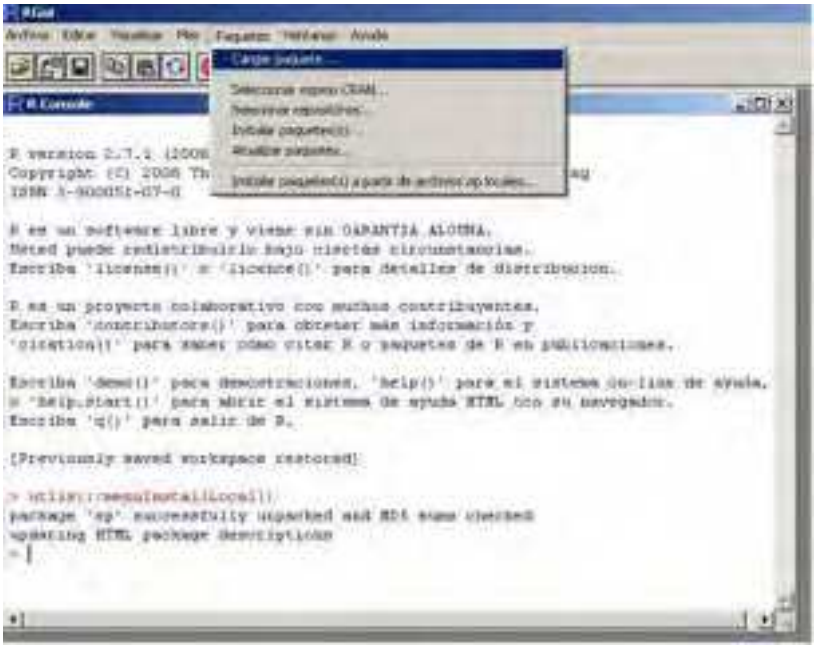


Figura 5.2: Cargando el paquete **sp**. Hacer clic en Cargar paquete y aparecerá una lista de todos los paquetes disponibles, por lo que se debe se debe seleccionar sp y dar clic en ok para que el paquete quede cargado al programa y listo para usarse.

segundo elemento (`$data`) la cual puede ser un vector o una matriz. Es una matriz si se mide más de una propiedad en el mismo punto.

Los objetos de la clase `geodata` pueden tener otros elementos tales como covariables y coordenadas de los límites del área de estudio, no importando que forma tenga la figura que forma los límites.

Hay pocos conjuntos de datos incluidos en la distribución del paquete. Para el ejemplo incluido en este documento se usará el conjunto de datos simulados `s100`, el cual contiene las coordenadas de las localizaciones de los datos y los datos en si. Para cargar un conjunto de datos se utiliza la función `data` de la siguiente manera

```
> data(s100)
```

donde al conjunto de datos se le tiene asignado el nombre de s100 y al agregar la extensión ($coords) devuelve el elemento referente a las coordenadas. El comando:

```
> s100$data
```

proporciona el vector referente a los datos. La lista de todos los conjuntos de datos incluidos en el paquete es dado por `data(package="geoR")`

5.3. Herramientas Exploratorias

Un resumen rápido del objeto `geodata` puede ser obtenido usando un método de resumen, el cual devolverá información de las coordenadas y valores de los datos, tales como el mínimo y máximo de las coordenada, de los datos y de las distancia entre los datos.

```
> summary(s100)

Number of data points: 100

Coordinates summary
        Coord.X    Coord.Y
min 0.005638006 0.01091027
max 0.983920544 0.99124979

Distance summary
        min         max
0.007640962 1.278175109
```

```
Data summary
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-1.1680  0.2730  1.1050  0.9307  1.6100  2.8680

Other elements in the geodata object
[1] "cov.model" "nugget"    "cov.pars"  "kappa"     "lambda"
```

Los elementos de la lista `$covariate`, `$borders` y/o `units.m` serán resumidos también si están presentes en el objeto `geodata`.

5.3.1. Graficando las localizaciones de los datos y valores

La función `plot.geodata` muestra un despliegue de las localizaciones de los datos y la gráfica de los datos contra las coordenadas. Para un objeto de la clase `geodata` el comando `plot(s100)` produce los resultados mostrados en la Figura 5.3 y son los mismos que produce la función `plot.geodata (s100)`.

La función `points.geodata (s100)` muestra la localización de cada uno de los datos, esto lo hace graficando los puntos de sus coordenadas. Hay opciones para especificar tamaños de los puntos, modelos y colores, los cuales pueden juntarse para ser proporcionales a los valores de los datos, es decir, se puede indicar que un dato es mucho mayor que otros asignando el tono de un color mas alto o bien haciendo que ese punto en la gráfica sea relativamente mayor que los demás. Algunos ejemplos de resultados gráficos son ilustrados por los comandos de abajo y corresponden a las gráficas mostradas en la Figura 5.4

```
> data(s100)
> par(mfrow = c(2, 2))
> points(s100, xlab = "Coord X", ylab = "Coord Y")
> points(s100, xlab = "Coord X", ylab = "Coord Y", pt.divide = "rank.prop")
> points(s100, xlab = "Coord X", ylab = "Coord Y", cex.max = 1.7,
+     col = gray(seq(1, 0.1, l = 100)), pt.divide = "equal")
> points(s100, pt.divide = "quintile", xlab = "Coord X",
+     ylab = "Coord Y")
```

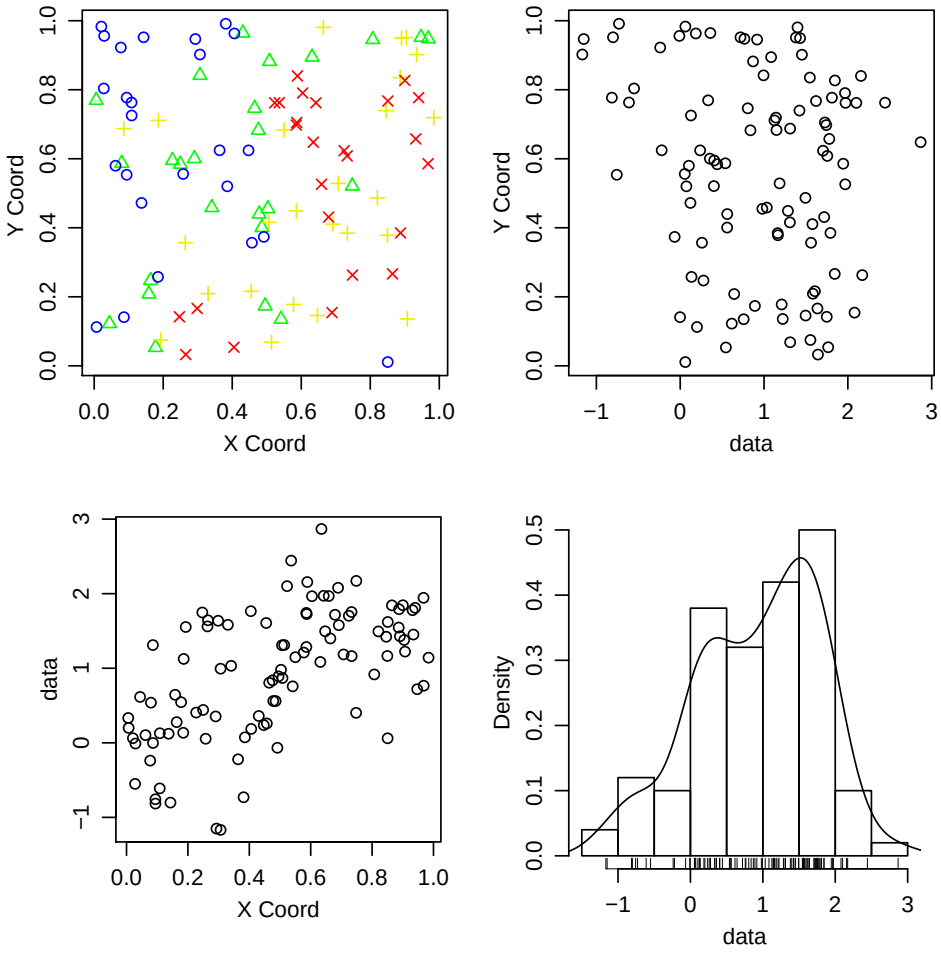


Figura 5.3: Gráfica producida por `plot.geodata(s100)` o por `plot(s100)`

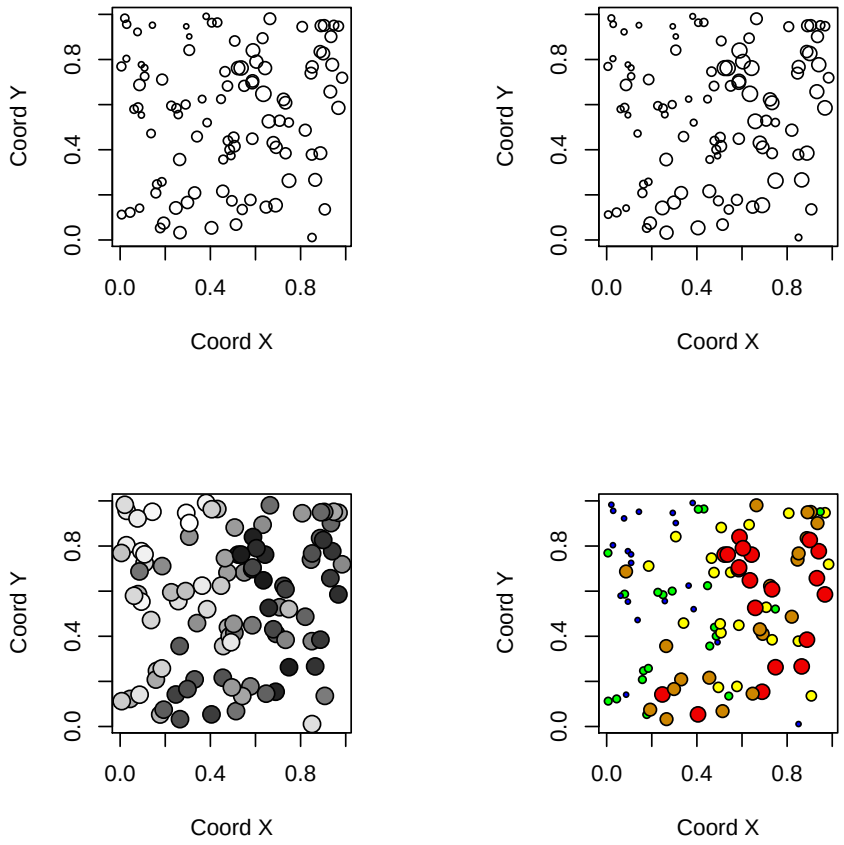


Figura 5.4: Gráfica producida por `points.geodata`. Los puntos corresponden a las localizaciones de los datos. Mientras más grande sea el círculo, mayor es el valor del dato. Para los circulos con colores, mientras más oscuro sea el color significa que el valor del dato es mayor en comparación con los demás.

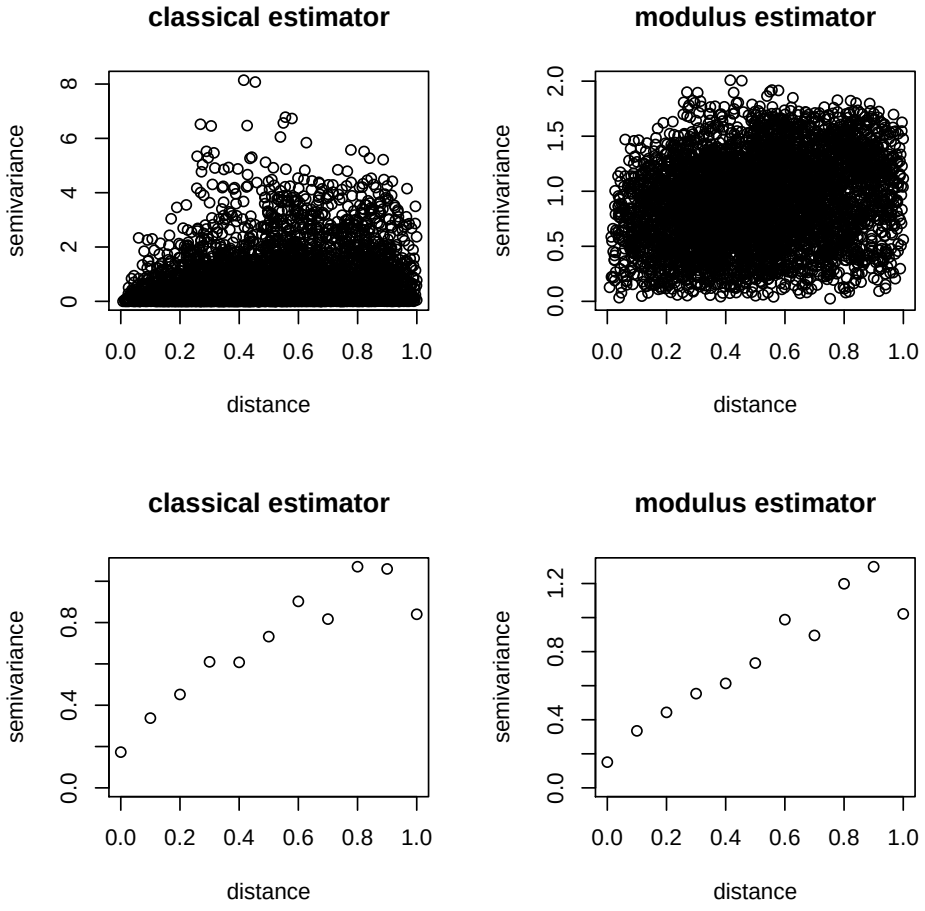


Figura 5.5: Graficando los resultados de `variog`

5.3.2. Semivariograma empírico

El Semivariograma empírico es calculado usando la función `variog`. La Figura 5.5 muestra el semivariograma empírico de los datos graficados anteriormente

```
> data(s100)
bin1<-variog(s100,uvec=seq(0,1,l=11))
```

La función `variog` devuelve varios resultados. Los primeros tres son los mas importantes y contienen las distancias, la semivarianza estimada y el numero de pares de datos involucrados para el cálculo de la semivarianza , respectivamente. Para el ejemplo, se asignó el nombre bin1 para el cálculo del semivariograma clásico utilizando la función `variog`, el argumento **uvec** indica las distancias para las cuales será calculado el semivariograma, los cuales son: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 y 1.0 y al mismo tiempo proporciona

la distancia máxima, que es 1. Por lo que la función `names` o bien `attributes` devuelve los resultados obtenidos con esta función.

```
> data(s100)
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> names(bin1)
```

```
 [1] "u"                "v"                "n"
 [4] "sd"               "bins.lim"         "ind.bin"
 [7] "var.mark"         "beta.ols"         "output.type"
[10] "max.dist"         "estimator.type"   "n.data"
[13] "lambda"           "trend"            "pairs.min"
[16] "nugget.tolerance" "direction"        "tolerance"
[19] "uvec"             "call"
```

Donde:

`bin1$u` contiene las distancia de separación para las que se calculó el semivariograma

`bin1$v` contiene los valores del semivariogramas para cada una de las distancias

`bin1$n` contiene el número total de pares para cada una de las distancias.

Los variogramas direccionales pueden ser calculados también por la función `variog` usando los argumentos `direction` y `tolerance`. Por ejemplo, para calcular un variograma para la dirección 60 grados con un ángulo de tolerancia (22.5 grados) el comando sería.

```
> data(s100)
> vario60 <- variog(s100, max.dist = 1, direction = pi/3)
```

Para un cálculo rápido en cuatro direcciones se usa la función `variog4` la cual calcula por default los variogramas para los ángulos de dirección 0, 45, 90 y 135 grados. Las direcciones anteriores vienen por default en el argumento `direction` de `variog4` y al igual que en `vario60` y `variog`, esta función también incluye al argumento `tolerance`.

```
> data(s100)
> vario.4 <- variog4(s100, max.dist = 1)
```

La figura 5.6 muestra los variogramas direccionales obtenidos con las funciones `vario60` y `vario.4`, respectivamente y los comandos son:

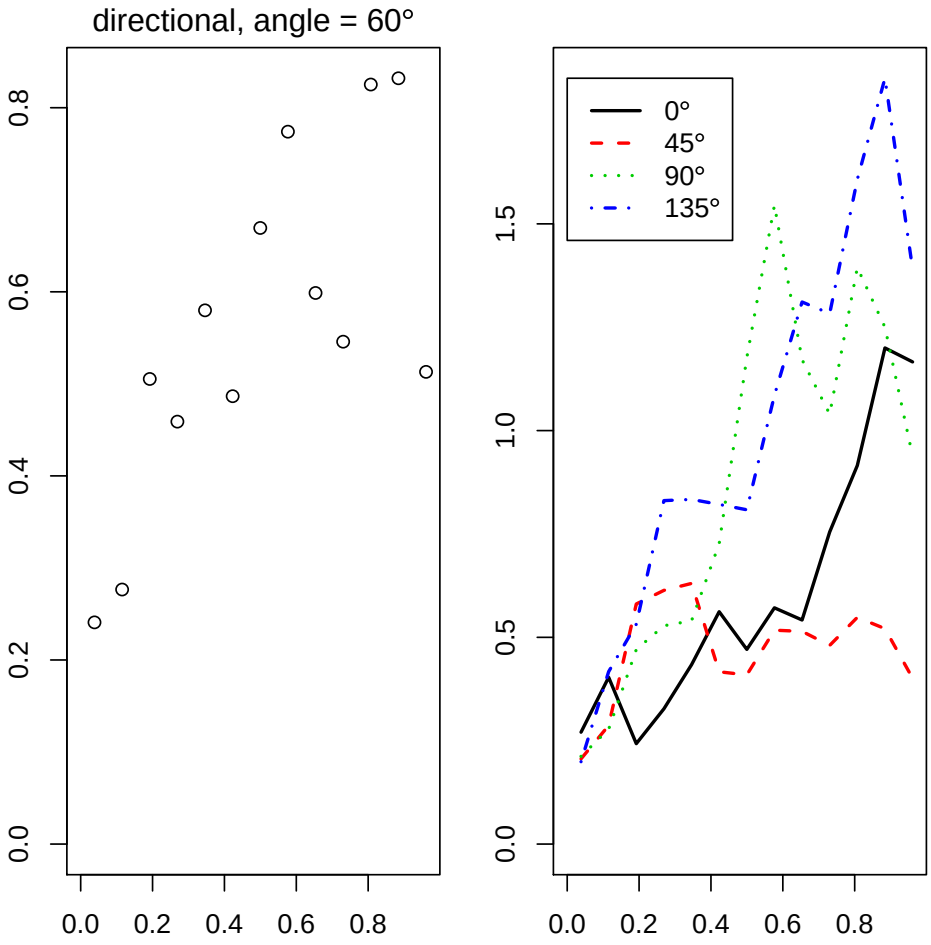


Figura 5.6: Variogramas direccionales

```
> data(s100)
> vario60 <- variog(s100, max.dist = 1, direction = pi/3)
> par(mfrow = c(1, 2), mar = c(3, 3, 1.5, 0.5))
> plot(vario60)
> title(main = expression(paste("directional, angle = ",
+     60 * degree)))
> vario.4 <- variog4(s100, max.dist = 1)
> plot(vario.4, lwd = 2)
```

5.4. Estimación de los Parámetros

Los variogramas teóricos y empíricos pueden ser comparados gráfica y visualmente. Por ejemplo en la figura 5.7 se muestra el modelo del variograma teórico usado para simular los datos s100 y dos variogramas estimados. Los cuales son creados a partir del código mostrado abajo. La función `lines.variomodel` adiciona una línea con un modelo de variograma especificado por sus argumentos.

```
> data(s100)
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> plot(bin1)
> lines.variomodel(cov.model = "exp", cov.pars = c(1,
+     0.3), nugget = 0, max.dist = 1, lwd = 3)
> smooth <- variog(s100, option = "smooth", max.dist = 1,
+     n.points = 100, kernel = "normal", band = 0.2)
> lines(smooth, type = "l", lty = 2)
> legend(0.4, 0.3, c("empirical", "exponential model",
+     "smoothed"), lty = c(1, 1, 2), lwd = c(1, 3, 1))
```

En la práctica usualmente no se conocen los parámetros verdaderos, por lo que se tienen que estimar por algún método. Los parámetros del modelo pueden estimarse usando el paquete **geoR**.

1. A ojo: tratando diferenes modelos sobre el variogramas empíricos (usando la función `lines.variomodel`),

2. Por ajuste de mínimos cuadrados de los variogramas empíricos:con opciones de mínimos cuadrados ordinarios y ponderados (usando la función `variofit`),

3. Método basado en la verosimilitud: con opciones para máxima verosimilitud y máxima verosimilitud con restricciones (usando la función `likfit`),

4. Método Bayesiano: usando la función `krige.bayes`

El ajuste a ojo, consiste en dibujar curvas de funciones de variogramas teóricos sobre un variograma empírico, cambiando el modelo del variograma y/o sus parámetros y, por último, escogiendo uno de ellos. Los siguientes comandos muestran como adicionar una línea con un modelo de variograma teórico a una gráfica de variograma empírico. Tres modelos diferentes de variogramas son usados y graficados junto con el variograma empírico en la Figura 5.8.

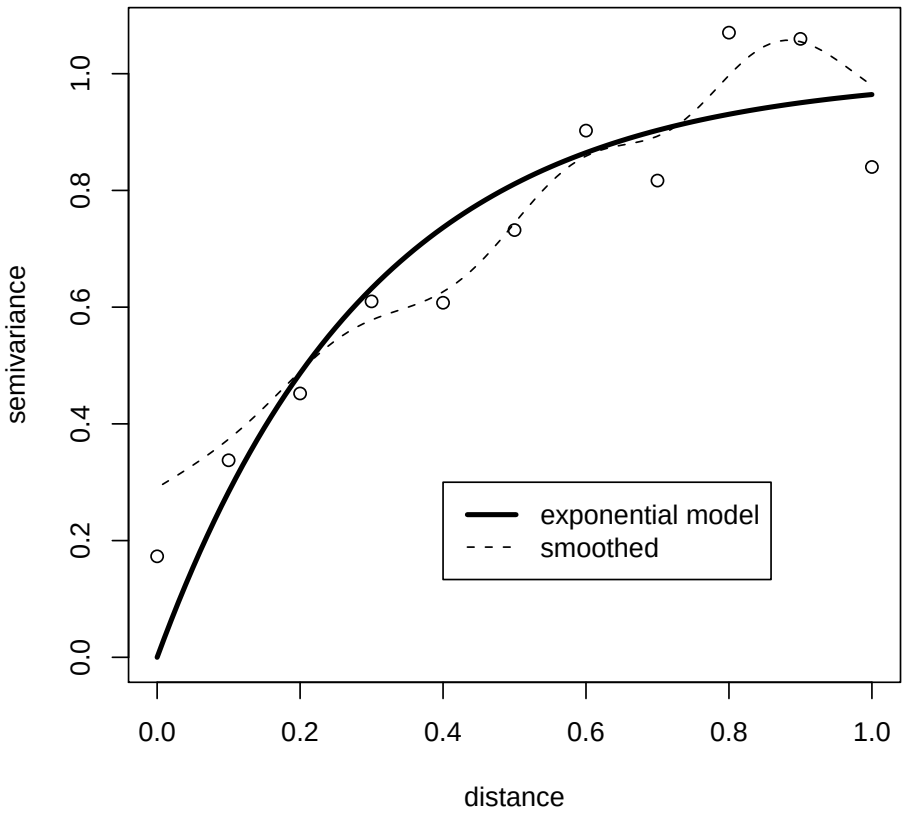


Figura 5.7: Curvas de variogramas teóricos ajustados al variograma empírico. Se observa que el modelo aplanado se ajusta mejor que el modelo exponencial

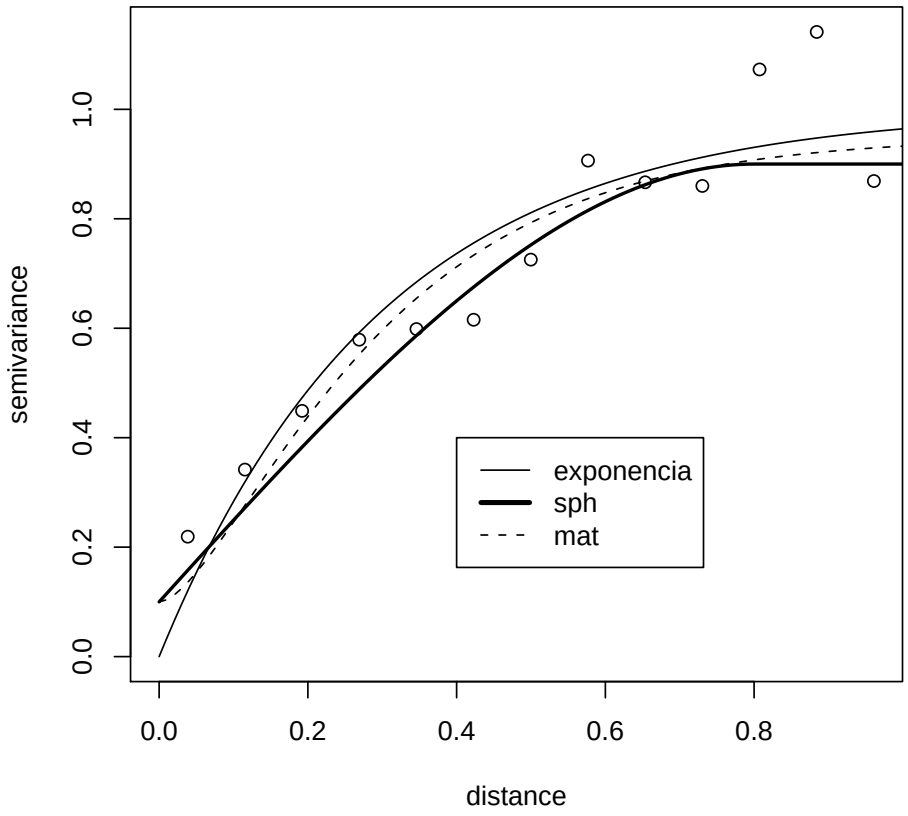


Figura 5.8: Tres curvas de variogramas teóricos adicionadas al variograma empírico. Los parámetros meseta y rango son dados con el argumento `cov.pars`, la pepita con el argument `nug`.

```
> data(s100)
> plot(variog(s100, max.dist = 1))
> lines.variomodel(cov.model = "exp", cov.pars = c(1,
+     0.3), nug = 0, max.dist = 1)
> lines.variomodel(cov.model = "mat", cov.pars = c(0.85,
+     0.2), nug = 0.1, kappa = 1, max.dist = 1, lty = 2)
> lines.variomodel(cov.model = "sph", cov.pars = c(0.8,
+     0.8), nug = 0.1, max.dist = 1, lwd = 2)
```

`variofit` estima los parámetros ajustando un modelo paramétrico a un variograma empírico usando mínimos cuadrados ordinarios o ponderados y `likfit` estima los parámetros

por máxima verosimilitud o por máxima verosimilitud restringida.

Cuando se usan las funciones de estimación paramétrica `variofit` y `likfit` el parámetro de efecto pepita puede ser estimado o bien poner un valor fijo. Lo mismo aplica para suavidad, anisotropía y parámetros de transformación. Las opciones para tomar en cuenta la tendencia también están incluidas. La tendencia puede ser incluida como funciones polinomiales de las coordenas y/o funciones lineales de las covarianzas dadas.

Un ejemplo utilizando la función `likfit` está dado a continuación. Aquí `ini.cov.pars` son los valores iniciales para los parámetros de covarianza: meseta parcial y rango.

Los métodos para `print()` y `summary()` han sido escrito para resumir los objetos resultantes. Aquí el argumento `ini.cov.pars` indica los valores de inicio de los parámetros de covarianza: meseta y rango, respectivamente.

```
> data(s100)
> ml <- likfit(s100, ini.cov.pars = c(1, 0.5))
> ml

likfit: estimated model parameters:
    beta    tausq  sigmasq      phi
"0.7766" "0.0000" "0.7517" "0.1827"
Practical Range with cor=0.05 for asymptotic range: 0.547355

likfit: maximised log-likelihood = -83.57

> summary(ml)

Summary of the parameter estimation
-----------------------------------
Estimation method: maximum likelihood

Parameters of the mean component (trend):
  beta
0.7766

Parameters of the spatial component:
   correlation function: exponential
      (estimated) variance parameter sigmasq (partial sill) =  0.7517
      (estimated) cor. fct. parameter phi (range parameter)  =  0.1827
```

```
   anisotropy parameters:
      (fixed) anisotropy angle = 0  ( 0 degrees )
      (fixed) anisotropy ratio = 1

Parameter of the error component:
      (estimated) nugget =  0

Transformation parameter:
      (fixed) Box-Cox parameter = 1 (no transformation)

Practical Range with cor=0.05 for asymptotic range: 0.547355

Maximised Likelihood:
   log.L n.params      AIC      BIC
"-83.57"      "4"  "175.1"  "185.6"

non spatial model:
   log.L n.params      AIC      BIC
"-125.8"      "2"  "255.6"  "260.8"

Call:
likfit(geodata = s100, ini.cov.pars = c(1, 0.5))
```

Los comandos de abajo muestran como ajustar modelos usando diferentes métodos con opciones para fijar o estimar el parámetro de efecto pepita pero no ilustra rasgos tales como estimación de tendecia, anisotropía, suavidad y parámetro de transformación Box Cox. `fix.nugget = T` indica que el efecto pepita está fijo.

Ajustando modelo con pepita fijado a cero.

```
> data(s100)
> ml <- likfit(s100, ini.cov.pars = c(1, 0.5), fix.nugget = T)
> reml <- likfit(s100, ini.cov.pars = c(1, 0.5), fix.nugget = T,
+     method = "RML")
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> ols <- variofit(bin1, ini.cov.pars = c(1, 0.5), fix.nugget = T,
+     weights = "equal")
> wls <- variofit(bin1, ini.cov.pars = c(1, 0.5), fix.nugget = T)
```

Ajustando modelo con valor fijo para el efecto pepita.

```
> data(s100)
> ml.fn <- likfit(s100, ini = c(1, 0.5), fix.nugget = T,
+     nugget = 0.15)
> reml.fn <- likfit(s100, ini = c(1, 0.5), fix.nugget = T,
+     nugget = 0.15, method = "RML")
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> ols.fn <- variofit(bin1, ini = c(1, 0.5), fix.nugget = T,
+     nugget = 0.15, weights = "equal")
> wls.fn <- variofit(bin1, ini = c(1, 0.5), fix.nugget = T,
+     nugget = 0.15)
```

Ajustando modelo pepita estimado

```
> ml.n <- likfit(s100, ini = c(1, 0.5), nug = 0.5)
> reml.n <- likfit(s100, ini = c(1, 0.5), nug = 0.5,
+     method = "RML")
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> ols.n <- variofit(bin1, ini = c(1, 0.5), nugget = 0.5,
+     weights = "equal")
> wls.n <- variofit(bin1, ini = c(1, 0.5), nugget = 0.5)
```

Ahora, los comandos para graficar modelos ajustados contra variograma empírico como se muestra en la Figura 5.7 y 5.8 son mostrados en la figura 5.9 y el código está escrito abajo:

```
> par(mfrow = c(1, 3))
> plot(bin1, main = expression(paste("fixed ", tau^2 ==
+     0)))
> lines(ml, max.dist = 1)
> lines(reml, lwd = 2, max.dist = 1)
> lines(ols, lty = 2, max.dist = 1)
> lines(wls, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS", "WLS"),
+     lty = c(1, 1, 2, 2), lwd = c(1, 2, 1, 2), cex = 0.7)
> plot(bin1, main = expression(paste("fixed ", tau^2 ==
```

```
+     0.15)))
> lines(ml.fn, max.dist = 1)
> lines(reml.fn, lwd = 2, max.dist = 1)
> lines(ols.fn, lty = 2, max.dist = 1)
> lines(wls.fn, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS", "WLS"),
+     lty = c(1, 1, 2, 2), lwd = c(1, 2, 1, 2), cex = 0.7)
> plot(bin1, main = expression(paste("estimated ", tau^2)))
> lines(ml.n, max.dist = 1)
> lines(reml.n, lwd = 2, max.dist = 1)
> lines(ols.n, lty = 2, max.dist = 1)
> lines(wls.n, lty = 2, lwd = 2, max.dist = 1)
> legend(0.5, 0.3, legend = c("ML", "REML", "OLS", "WLS"),
+     lty = c(1, 1, 2, 2), lwd = c(1, 2, 1, 2), cex = 0.7)
```

5.5. Validación cruzada

La función `xvalid` realiza validación cruzada usando la estrategia de eliminar un dato o usando un conjunto diferente de localizaciones proveídas por el usuario através del argumento `location.xvalid`. Para la primera estrategia, los datos puntuales son eliminados uno por uno y predichos por kriging usando los datos que no fueron eliminados. Los comandos de abajo ilustran la validación cruzada para los modelos ajustados por máxima verosimilitud y mínimos cuadrados ponderados. En las siguientes dos llamadas los parámetros del modelo permanecen igual para la predicción en cada localización.

```
> data(s100)
> xv.ml <- xvalid(s100, model = ml)
> xv.wls <- xvalid(s100, model = wls)
```

Los resultados gráficos se muestran para los resultados de la validación cruzada donde se usa la estrategia de eliminar un dato combinada con los estimadores de mínimos cuadrados ponderados. Los residuales de la validación cruzada se obtienen sustrayendo los datos observados menos el valor predicho. Los residuales estandarizados son obtenidos dividiendolos por la raíz cuadrada de la varianza de la predicción (varianza del kriging). Por default se producen las 10 gráficas mostradas en la Figura 5.10 pero el usuario puede restringir esta

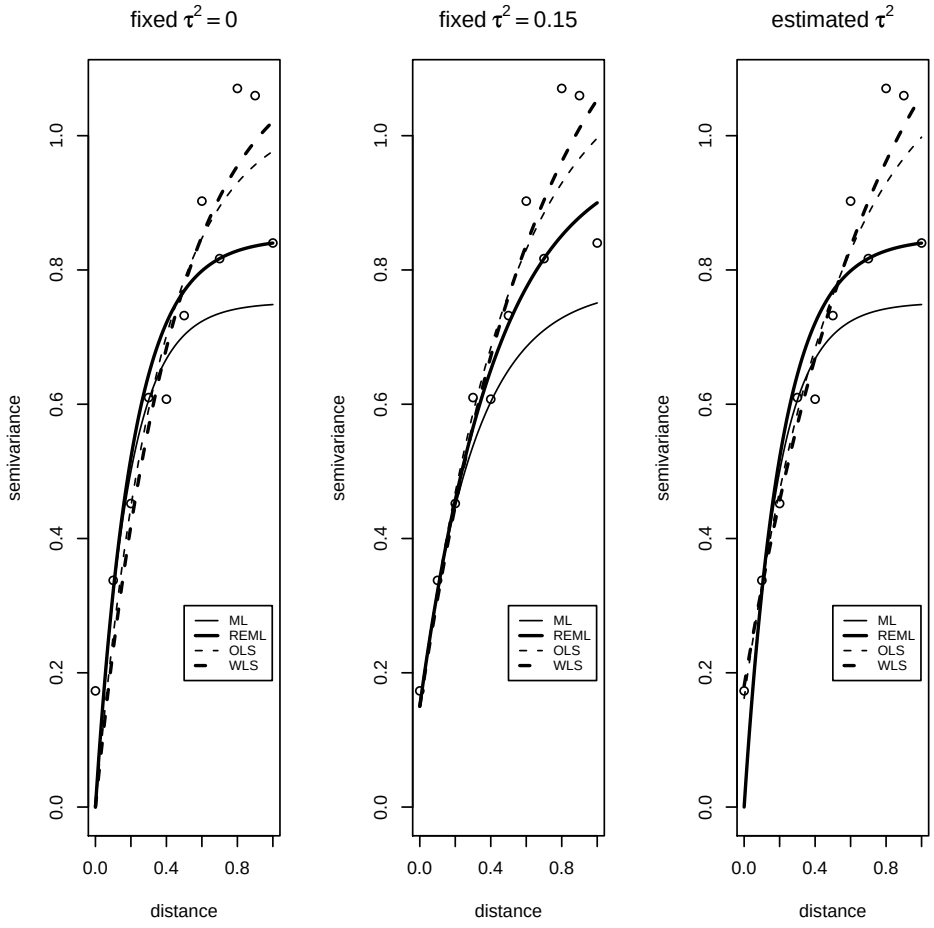


Figura 5.9: Variograma empírico y modelos ajustados por diferentes métodos

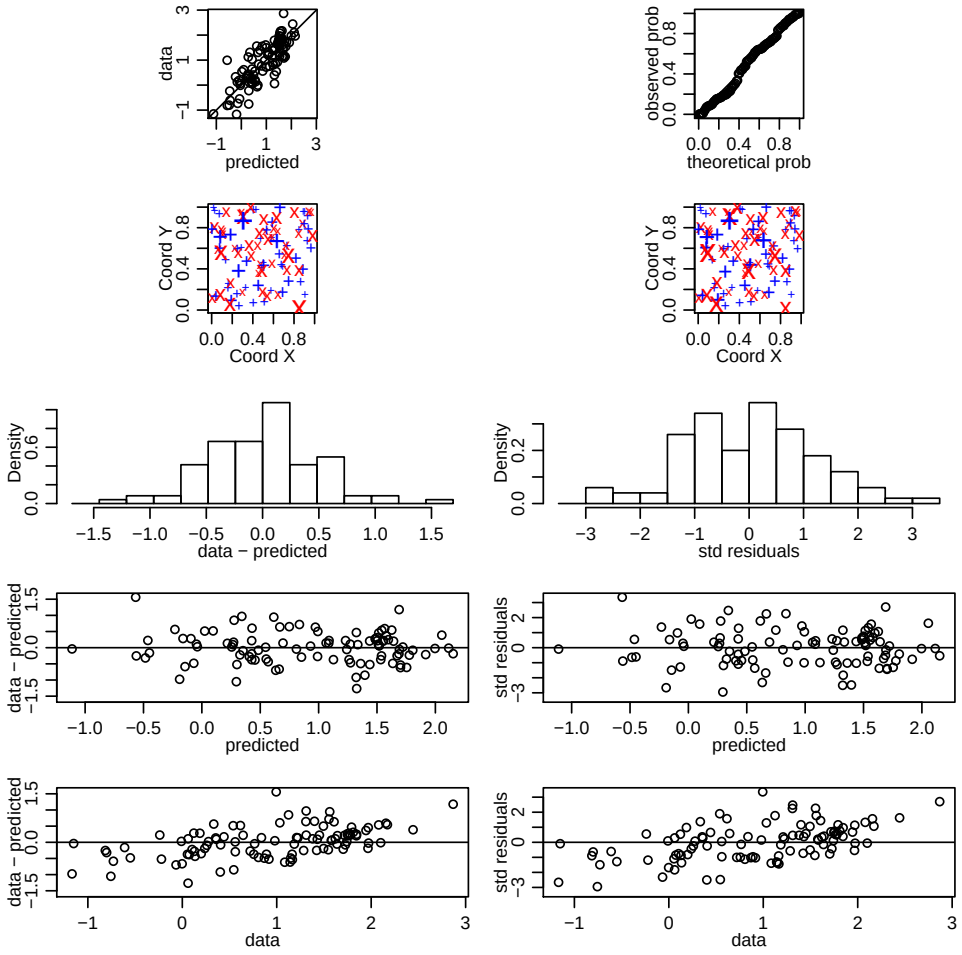


Figura 5.10: Gráficas resultantes con la validación cruzada.

opción usando los argumentos de dicha función. La gráfica de dispersión de los datos eliminados contra su correspondiente valor predicho mostrada en la parte superior izquierda indica que los puntos siguen una línea recta.

```
> par(mfcol = c(5, 2), mar = c(3, 3, 1, 0.5), mgp = c(1.5,
+     0.7, 0))
> plot(xv.wls)
```

Una variación de este método es ilustrado por las siguientes dos llamadas, donde los parámetros del modelo son re-estimados cada vez que un punto es removido del conjunto de datos. Aunque correr estos comandos puede consumir algo de tiempo.

```
> xvR.ml <- xvalid(s100, model = ml, reest = TRUE)
```

```
> xvR.wls <- xvalid(s100, model = wls, reest = TRUE,
+     variog.obj = bin1)
```

5.6. Interpolación espacial

La interpolación geoestadistica convencional (kriging) puede ser realizada con opciones para:

1. kriging simple
2. kriging ordinario
3. kriging con tendencia (universal)

Hay opciones adicionales para la transformación Box-Cox (y transformación de regreso de los resultados) y modelos anisotrópicos. Las simulaciones pueden ser dibujadas de los resultados de las distribuciones predichas, si se piden.

Como primer ejemplo considerar la predicción en cuatro localizaciones con etiquetas 1, 2, 3, 4 e indicado en la figura 5.11. Para agregar estas etiquetas se teclean los siguientes comandos:

```
> data(s100)
> plot(s100$coords, xlim = c(0, 1.2), ylim = c(0, 1.2),
+     xlab = "Coord X", ylab = "Coord Y")
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> text(loci, as.character(1:4), col = "red")
```

El comando para relizar kriging ordinario usando los parámetros estimados por mínimos cuadrados ponderados con efecto pepita fijado aparece abajo. El argumento `locations` indica los puntos en los que se harán las predicciones, que en este caso incluye 4 puntos colocados en forma de matriz para poder ser leídas por la función. El argumento `krige` es una lista que define los parámetros del modelo y el tipo de kriging a utilizar, que para este ejemplo y generalmente, es dado por la función `krige.control`, donde `obj.model` es una lista con los parámetros del modelo, que en este caso, son los resultados obtenidos por la función `variofit`.

```
> data(s100)
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
```

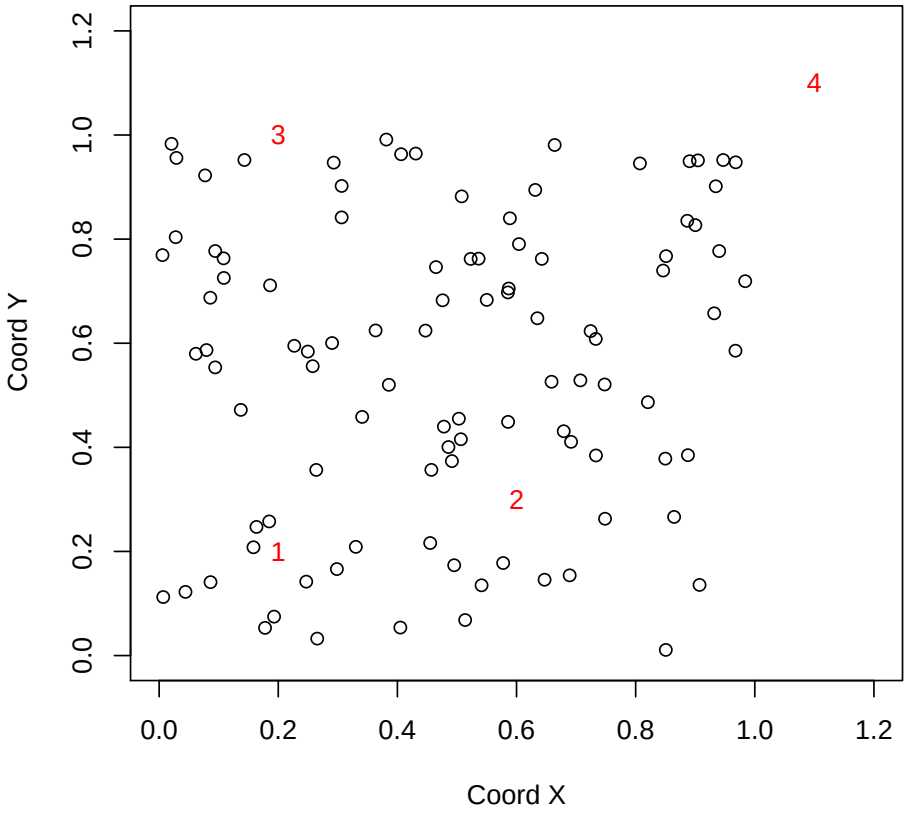


Figura 5.11: Localizaciones de datos y puntos para ser predichos

```
> wls <- variofit(bin1, ini = c(1, 0.5), fix.nugget = T)
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> kc4 <- krige.conv(s100, locations = loci,
+krige = krige.control(obj.model = wls))
```

El resultado es una lista que incluye los valores predichos (`kc4$predict`) y las varianzas del kriging (`kc4$krige.var`). Considerar ahora un segundo ejemplo. El objetivo es realizar predicción en una rejilla cubriendo el área y mostrar los resultados. De nuevo, se usa el kriging ordinario. Los comandos de abajo definen una rejilla de localizaciones y realizan la predicción en estas localizaciones. La función `expand.grid` crea la rejilla, creando un cuadrado donde cada lado de distancia igual a 1 tiene 51 divisiones, por lo que los puntos a predecir son las intersecciones de estas divisiones. En la Figura 5.12 se muestra el área que cubre esta rejilla.

```
> data(s100)
> plot(s100$coords, xlim = c(0, 1.2), ylim = c(0, 1.2),
+     xlab = "Coord X", ylab = "Coord Y")
> polygon(x = c(0, 1, 1, 0), y = c(0, 0, 1, 1), lty = 2)
> ml <- likfit(s100, ini = c(1, 0.5), fix.nugget = T)
> pred.grid <- expand.grid(seq(0, 1, l = 51), seq(0,
+     1, l = 51))
> kc <- krige.conv(s100, locations = pred.grid,
+krige = krige.control(obj.m = ml))
```

Un método para la función `image` puede usarse para desplegar los valores predichos como se muestra en la Figura 5.13, así como otros resultados de la predicción devueltos por `krige.conv`. Correr estos comandos puede ser tardado.

```
> data(s100)
> ml <- likfit(s100, ini = c(1, 0.5), fix.nugget = T)
> pred.grid <- expand.grid(seq(0, 1, l = 51), seq(0,
+     1, l = 51))
> kc <- krige.conv(s100, loc = pred.grid,
+krige = krige.control(obj.m = ml))
> image(kc, locations = pred.grid, col = gray(seq(1,
+     0.1, l = 30)), xlab = "Coord X", ylab = "Coord Y")
```

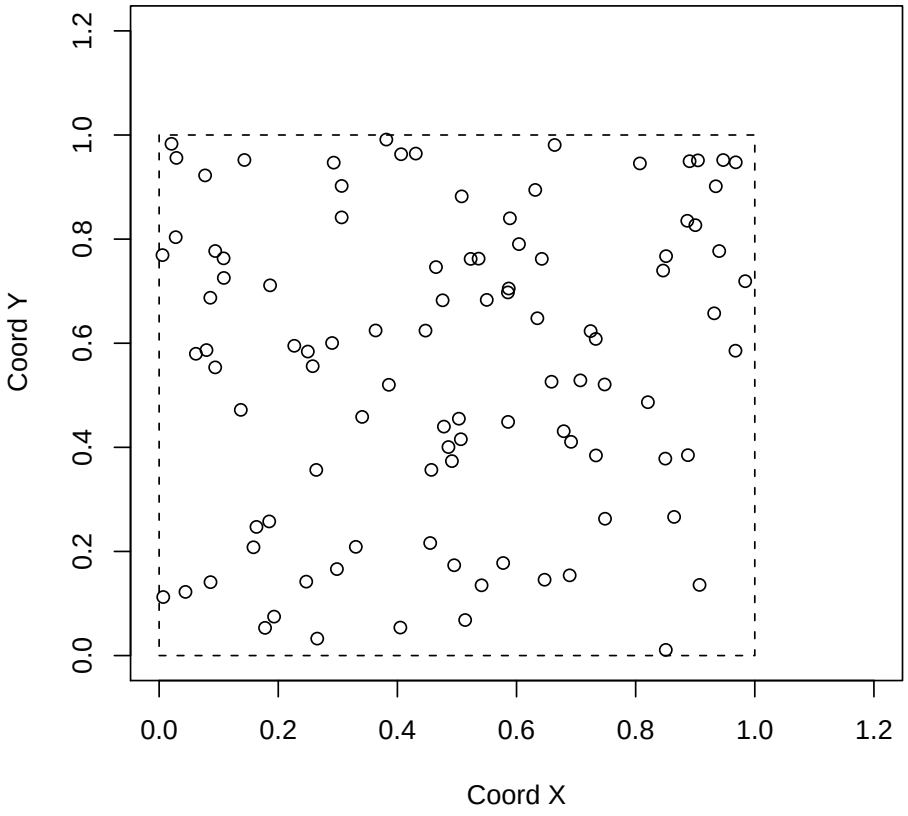


Figura 5.12: Localizaciones de los datos y rejilla donde se harán las predicciones

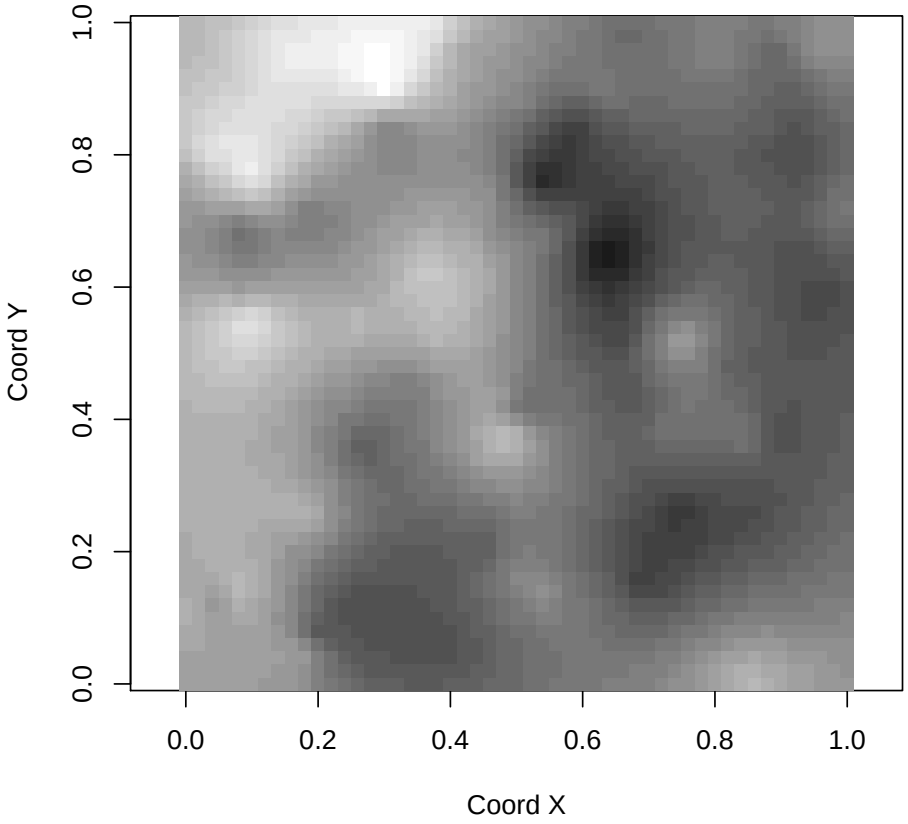


Figura 5.13: Mapa de estimaciones del kriging.

5.7. Análisis Bayesiano

El análisis Bayesiano para modelos Gausianos es implementado por la función `krige.bayes`. Esto puede ser realizado para diferentes grados de incertidumbre, es decir, los parámetros del modelo pueden ser tratados como fijos o aleatorios.

Como un ejemplo, considerar un modelo sin pepita incluyendo incertidumbre en los parámetros media, meseta y rango. La predicción en las cuatro localizaciones indicadas arriba es realizada escribiendo un comando como:

```
> data(s100)
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> bsp4 <- krige.bayes(s100, loc = loci,
+          prior = prior.control(phi.discrete = seq(0,
+     5, l = 101), phi.prior = "rec"),
+          output = output.control(n.post = 5000))
```

Los histogramas mostrando la distribución posterior para los parámetros del modelo pueden ser graficados tecleando el comando de abajo y mostrados en la Figura 5.14

```
> data(s100)
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> bsp4 <- krige.bayes(s100, loc = loci,
+     prior = prior.control(phi.discrete = seq(0,
+     5, l = 101), phi.prior = "rec"),
+          output = output.control(n.post = 5000))
> par(mfrow = c(1, 3), mar = c(3, 3, 1, 0.5), mgp = c(2,
+     1, 0))
> hist(bsp4$posterior$sample$beta, main = "", xlab = expression(beta),
+     prob = T)
> hist(bsp4$posterior$sample$sigmasq, main = "",
+          xlab = expression(sigma^2), prob = T)
> hist(bsp4$posterior$sample$phi, main = "", xlab = expression(phi),
+     prob = T)
```

Usando resúmenes de estas distribuciones posteriores (medias, medianas o modas) se puede checar los variogramas estimados Bayesianos contra el variograma empírico, los

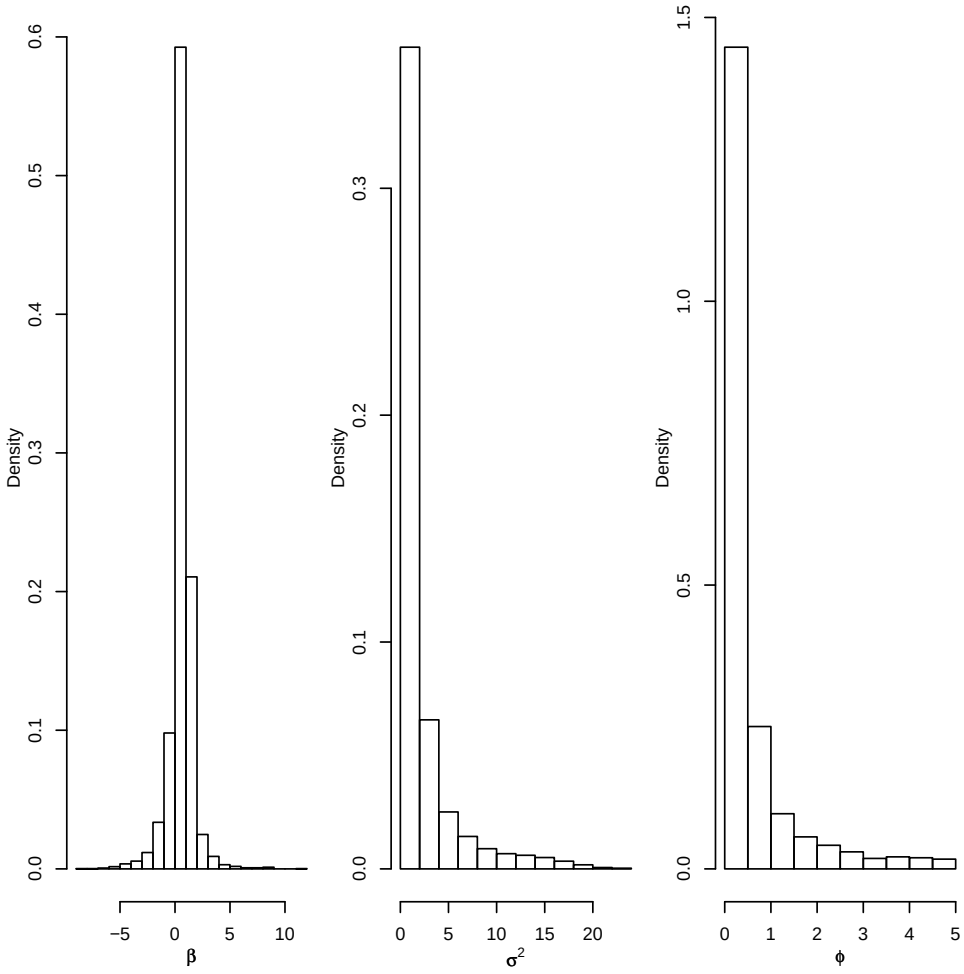


Figura 5.14: Histogramas de muestras de la distribución posterior

cuales son mostrados en la Figura 5.15 y el código está escrito abajo. Notar que también es posible comparar estas estimaciones con otros variogramas ajustados calculados en la sección 3.

```
> data(s100)
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> plot(bin1, ylim = c(0, 1.5))
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> bsp4 <- krige.bayes(s100, loc = loci,
+     prior = prior.control(phi.discrete = seq(0,5, l = 101),
+     phi.prior = "rec"),output = output.control(n.post = 5000))
> lines(bsp4, max.dist = 1.2, summ = mean)
> lines(bsp4, max.dist = 1.2, summ = median, lty = 2)
> lines(bsp4, max.dist = 1.2, summ = "mode", post = "par",
+     lwd = 2, lty = 2)
> legend(0.25, 0.4, legend = c("variogram posterior mean",
+     "variogram posterior median", "parameters posterior mode"),
+     lty = c(1, 2, 2), lwd = c(1, 1, 2), cex = 0.8)
```

En la figura 5.16 se muestran las distribuciones predichas en las cuatro localizaciones seleccionadas. Las líneas cortadas muestran las distribuciones Gaussianas con media y varianza dados por los resultados del kriging ordinario obtenidos en la sección 4. Las líneas completas corresponden a la predicción Bayesiana. La gráfica muestra resultados de densidad de estimación usando muestras de las distribuciones predichas.

```
> data(s100)
> bin1 <- variog(s100, uvec = seq(0, 1, l = 11))
> wls <- variofit(bin1, ini = c(1, 0.5), fix.nugget = T)
> loci <- matrix(c(0.2, 0.6, 0.2, 1.1, 0.2, 0.3, 1, 1.1),
+     ncol = 2)
> kc4 <- krige.conv(s100, locations = loci,
+    krige = krige.control(obj.m = wls))
> par(mfrow = c(2, 2))
> for (i in 1:4) {
+     kpx <- seq(kc4$pred[i] - 3 * sqrt(kc4$krige.var[i]),
+         kc4$pred[i] + 3 * sqrt(kc4$krige.var[i]), l = 100)
```

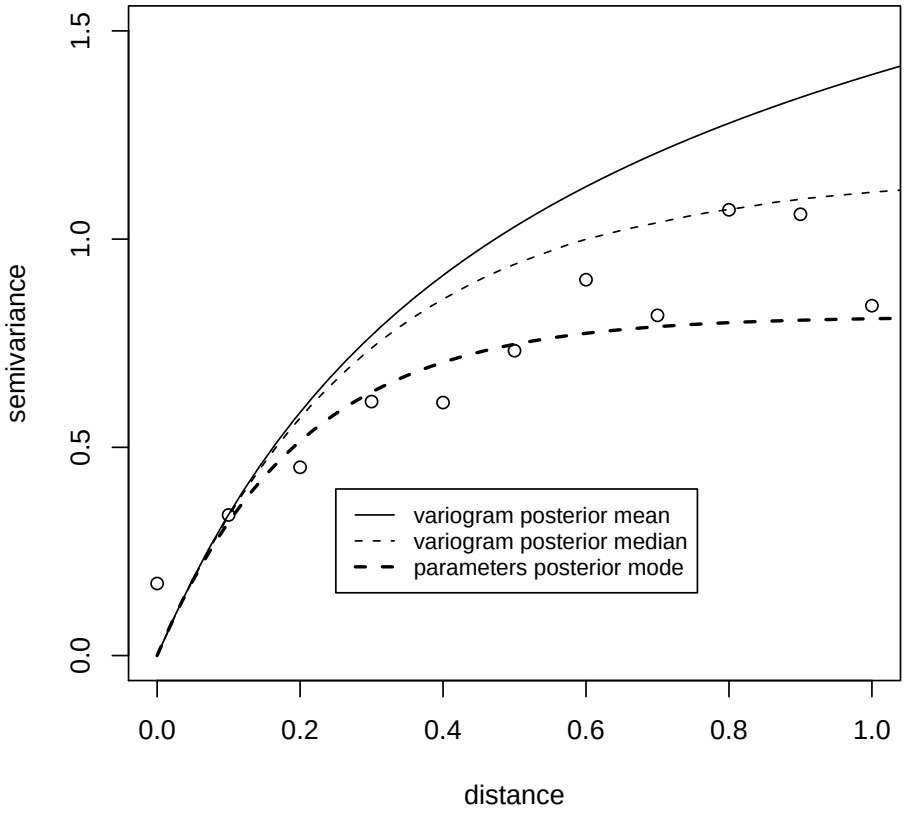


Figura 5.15: Modelos de variogramas basados en las distribuciones posteriores

```
+     kpy <- dnorm(kpx, mean = kc4$pred[i], sd = sqrt(kc4$krige.var[i]))
+     bsp4 <- krige.bayes(s100, loc = loci,
+     prior = prior.control(phi.discrete = seq(0,5, l = 101),
+     phi.prior = "rec"), output = output.control(n.post = 5000))
+     bp <- density(bsp4$predic$simul[i, ])
+     rx <- range(c(kpx, bp$x))
+     ry <- range(c(kpy, bp$y))
+     plot(cbind(rx, ry), type = "n", xlab = paste("Location",
+         i), ylab = "density", xlim = c(-4, 4), ylim = c(0,
+         1.1))
+     lines(kpx, kpy, lty = 2)
+     lines(bp)
+ }
```

Considerar ahora, bajo la misma suposición del modelo, obtener simulaciones de las distribuciones predichas en una rejilla de puntos que cubren el área. Los comandos para definir la rejilla y realizar la predicción Bayesiana son:

```
> data(s100)
> pred.grid <- expand.grid(seq(0, 1, l = 31), seq(0,
+     1, l = 31))
> bsp <- krige.bayes(s100, loc = pred.grid,
+     prior = prior.control(phi.discrete = seq(0, 5, l = 51)),
+     output = output.control(n.predictive = 2))
```

La Figura 7.4 muestra los mapas con los resúmenes y simulaciones de la distribución predicha y pueden ser graficadas como sigue:

```
> data(s100)
> pred.grid <- expand.grid(seq(0, 1, l = 31), seq(0,
+     1, l = 31))
> bsp <- krige.bayes(s100, loc = pred.grid,
+     prior = prior.control(phi.discrete = seq(0,5, l = 51)),
+     output = output.control(n.predictive = 2))
> par(mfrow = c(2, 2), mar = c(3, 3, 1, 0.5), mgp = c(1.5,
+     0.7, 0))
> image(bsp, loc = pred.grid, main = "predicted", col = gray(seq(1,
```

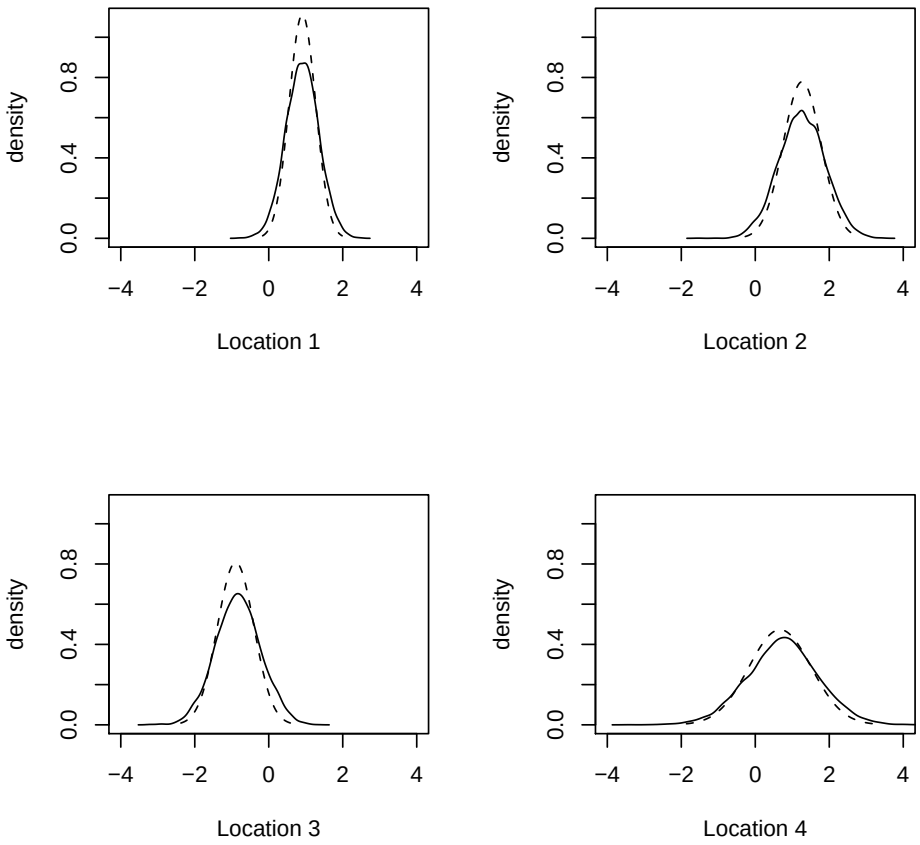


Figura 5.16: Distribuciones Predichas en los cuatro localizaciones seleccionadas

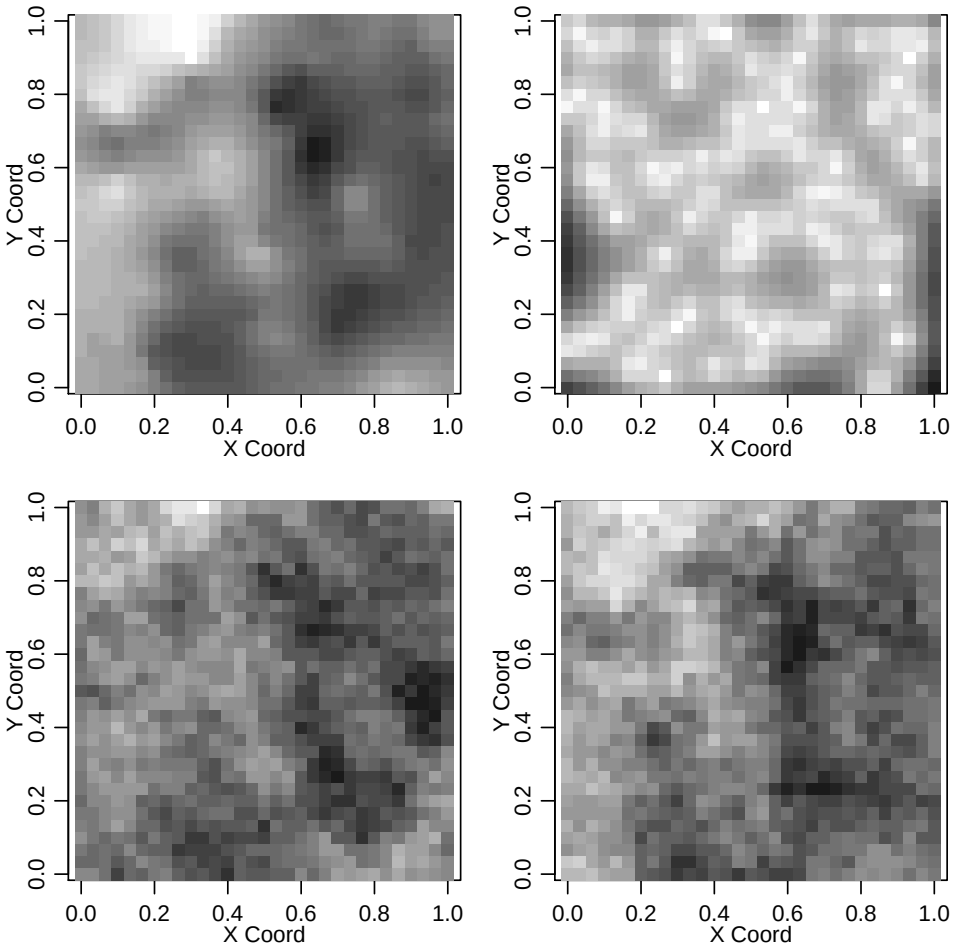


Figura 5.17: Mapas obtenidos de la distribución predicha

```
+     0.1, l = 30)))
> image(bsp, val = "variance", loc = pred.grid, main = "prediction variance",
+     col = gray(seq(1, 0.1, l = 30)))
> image(bsp, val = "simulation", number.col = 1, loc = pred.grid,
+     main = "a simulation from\nthe predictive distribution",
+     col = gray(seq(1, 0.1, l = 30)))
> image(bsp, val = "simulation", number.col = 2, loc = pred.grid,
+     main = "another simulation from \n the predictive distribution",
+     col = gray(seq(1, 0.1, l = 30)))
```

Capítulo 6

APLICACIÓN DE LA GEOESTADÍSTICA: ESTUDIO DE LA MOSCA DE LA FRUTA EN EL ESTADO DE S.L.P.

6.1. Introducción

En el presente capítulo se realiza el análisis Estadístico y Geoestadístico del comportamiento que presenta la distribución de la mosca de la fruta en cada una de las semanas muestreadas.

Como se vio anteriormente, a partir de una muestra de alguna propiedad medida en un área determinada se puede estimar la medición en cualquier punto de la misma. Por lo que el objetivo que se pretende es la estimación $Z(\mathbf{x_0})$ de la función aleatoria número de individuos, $Z(\mathbf{x})$, en una localización cualquiera $\mathbf{x}_0$ no muestreada. Por consiguiente, el objetivo último de este trabajo es generar, a partir de una muestra del número de capturas de mosca de la fruta en una determinada zona geográfica, un mapa de incidencia de la zona en estudio. Como el muestreo se hizo durante 42 semana consecutivas, una propuesta es repetir el procedimiento para cada semana y con esto determinar su distribución histórica.

El número de moscas a lo largo de un determinado dominio es una función aleatoria con argumento espacial o, en otros términos, una variable regionalizada (Matheron, 1965) por estar distribuida y autocorrelacionada en el espacio, siendo la muestra una realización de la misma. El proceso de estimación del número de moscas no sólo necesitará de la información experimental proporcionada por la muestra, sino también de la información

estructural relativa al tipo de dependencia espacial suministrada por el covariograma, semivariograma o correlograma.

Para llevar a cabo el análisis geoestadístico se dispuso del programa R, y más específicamente de la librería `geoR`. La elección del mismo se debe a que con él se pueden efectuar todas las fases de un estudio geoestadístico, proporcionando gráficas de gran calidad con diferentes opciones para mostrar la información disponible y además por su fácil acceso y uso ya que tanto el programa como sus librerías se pueden descargar de internet gratuitamente. La base de datos fue proporcionada por parte de las oficinas de Sanidad Vegetal ubicadas en el Distrito Federal y consta de las coordenadas de las 444 trampas y del número de capturas para cada una de las 42 semanas de muestreo. Dispuesta esta información en un bloc de notas, en forma tabulada, se procedió a la realización del estudio geoestadístico.

6.2. Área de Estudio

El estudio se ha localizado en la zona media del Estado de San Luis Potosí, abarcando los municipios de Ciudad Fernández, Pinihuan, Rioverde y San Ciro donde existen plantaciones de frutales. La parcela dominante corresponde principalmente a cultivo de Naranja y en menor cantidad de Pomelo, Guayaba, Carambolo, Chabacano, Chicozapote, Ciruela, Durazno y Granada.

Se localizaron las parcelas donde existen cultivos de frutales y se distribuyeron un total de 444 trampas para la captura de las moscas y se estableció su localización espacial a través de sus coordenadas espaciales (x,y) medidas en metros.

Se efectuaron conteos semanales de las moscas capturadas en cada trampa durante 42 semanas.

La distribución de las trampas no es regular, como tampoco lo es su distancia de separación, por lo que no se puede decir con exactitud la superficie del área de estudio. En la Figura 6.1 se muestra la distribución de las trampas.

6.3. Descripción Univariada

El estudio Estadístico está comprendido por los análisis Univariado y Bivariado. En el APÉNDICE A se presenta cada uno de los valores obtenidos en el análisis Univariado, para cada una de las semanas muestreadas.

Cada conjunto de datos tienen el 0 como mínimo y el máximo varía de semana a

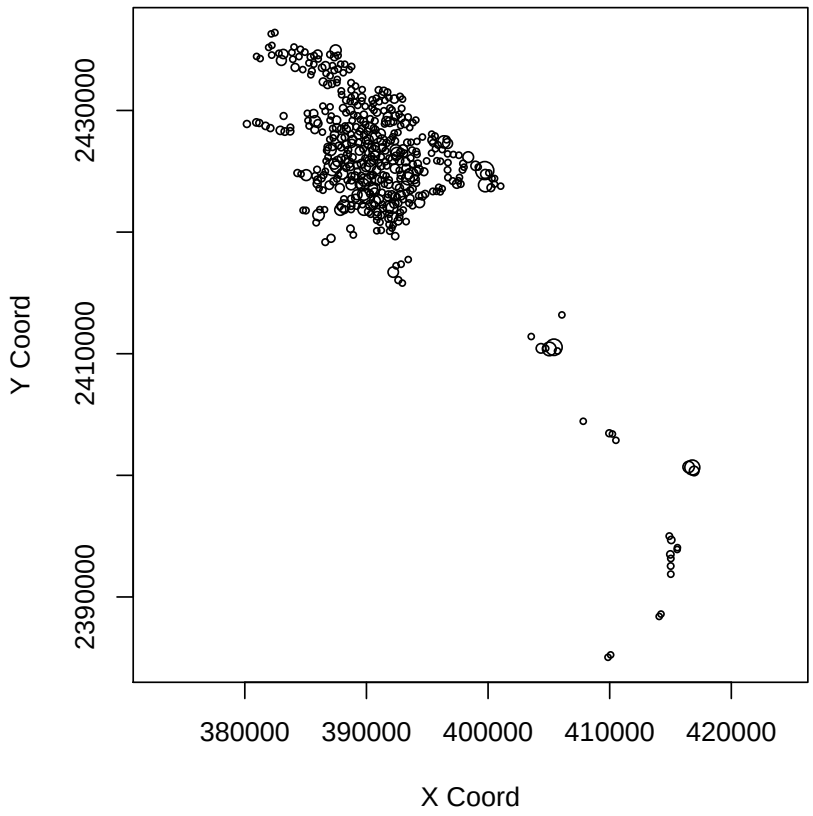


Figura 6.1: Gráfica de distribución de las trampas. Los puntos corresponden al valor en cada una de las trampas correspondientes al muestreo de la semana 4.

semana: las primeras 19 semanas tienen como máximo un número menor a 100 capturas, 6 semanas entre la 20 y la 37 presentan mayor incidencia y son la 20, 23, 32, 33, 34, 35, 36 y 37, de entre ellas, las semanas 20 y 23 presentan datos atípicos, por ejemplo existe trampas donde tiene 100 capturas mas que la semana anterior y la que le sigue tiene menos de 5. Las últimas 7 semanas tienen 100 capturas en promedio como máximo.

Los Coeficiente de Asimetría y de Variación son relativamente grandes para las semanas 1, 15, 16, esto por el hecho de que los máximos para estas semanas son muy elevados con respecto a los valores restantes. Los datos que presentan mayor media y varianza corresponden a las semanas 20, 23, 33, 34 y 35, ya que en estas semanas es donde se reporta mayor densidad de población, y de ellos, quien presenta un coeficiente de varianza mayor, corresponden a la semana 33.

6.3.1. Histogramas

A continuación se da la interpretación de los histogramas presentados en el APÉNDICE B para cada conjunto de datos. Se construyó un histograma para cada una de las semanas para ver la dispersión que presentan los datos. Para comparar los distintos histogramas se utilizó una anchura de clase de 5 unidades

En todos los casos se detectaron distribuciones sesgadas a la derecha, ya que la mayoría de los datos se reúnen más en la parte izquierda del valor de la media, en el intervalo (0,5], esto es visto también por el coeficiente de asimetría, que es positivo para cada conjunto de datos. Los coeficientes de curtosis para cada una de las semanas es mayor que cero, por lo que el grado de concentración que presentan los valores en la región central de la distribución es alta. Como alternativa para ver la asimetría de la forma de la distribución se muestra también el Coeficiente de Variación, el cual es mayor que uno para cada conjunto de datos, indicando que los datos están muy dispersos. Incluso, para los conjuntos de datos para las semanas 1, 14, 15, 16, 17, 18, 19, 21, 24, 25, 26, 27, 32, 33 y 35 este coeficiente es mayor que dos.

6.4. Descripción Bivariada

En este tipo de análisis, se estudia el comportamiento que tienen entre cada par de conjuntos de datos para ver cual es la relación entre cada una de las semanas muestreadas.

La covarianza mide el grado de variación de una variable con respecto a la otra y el coeficiente de correlación mide el grado de la dependencia lineal entre dos variables. Se observa (VER APÉNDICE C) de acuerdo al coeficiente de correlación que, en general,

existe dependencia lineal entre las semanas muestreadas consecutivamente, es decir, existe dependencia lineal entre los conjuntos de datos correspondientes a la semana 4 y 5, al igual que entre la semana 5 y 6, etc., excepto para las semanas muestreadas de la 13 a la 22, que su coeficiente de correlación entre semanas muestreadas consecutivamente es menor a 0.6, esto sucede igual para las semanas 26 a 29 y para las primeras 4 semanas. A medida que se separan las fechas de muestreo (semanas) la dependencia lineal entre semanas disminuye considerablemente.

6.5. Análisis Espacial

En la fase del análisis estructural de los datos, se eligió la función variograma para la caracterización de la continuidad espacial. Se construyeron los semivariogramas experimentales para cada una de las semanas y posteriormente, con la ayuda del progrma "R" se ajustaron los modelos teóricos con la función de máxima verosimilitud incluida en dicho programa, para ello, se hizo uso del criterio de información de Akaike para obtener los parámetros de los modelos. Se comenzó buscando la distancia máxima y el incremento de la distancia para considerarlos en el cálculo

Se observa en la Figura 6.1, que la distancia entre los puntos más alejados es igual a 58,260.82 metros. Una magnitud igual a la mitad de esta medida nos permitirá abarcar todos los puntos si nos situásemos en el centro de la misma. Por lo que fue conveniente considerar una distancia máxima de 30 000 m, que es aproximadamente la mitad de la distancia entre los dos puntos más alejados. El siguiente parámetro que se fijó es el incremento de la distancia. No se sabe exactamente la distancia de separación de las observaciones, pero se puede tener una noción al acercarse más a la gráfica de posicionamiento de las observaciones. La gráfica de la Figura 6.2 muestra este acercamiento. Al hacer un cálculo de las distancias, se concluye que la mayoría está una separación de alrededor de 600 m. Por lo que se consideró este número como la distancia promedio entre trampas adyacentes. Para el programa "R", para el cálculo del variorama no se tomó en cuenta una tolerancia dimensional y tampoco se consideró la existencia de anisotropía. Así, en este caso, el primer punto del variograma se obtuvo mediante el emparejamiento de cada dato muestral con aquellos que disten menos de 600 m; el segundo punto se obtiene emparejando los datos muestrales con aquellos que disten entre 600 m y 1200 m, y así hasta completar la distancia máxima de 30 000 m.

Finalmente se representan gráficamente los valores de $\gamma(\mathbf{h})$ en función de $\mathbf{h}$. En el

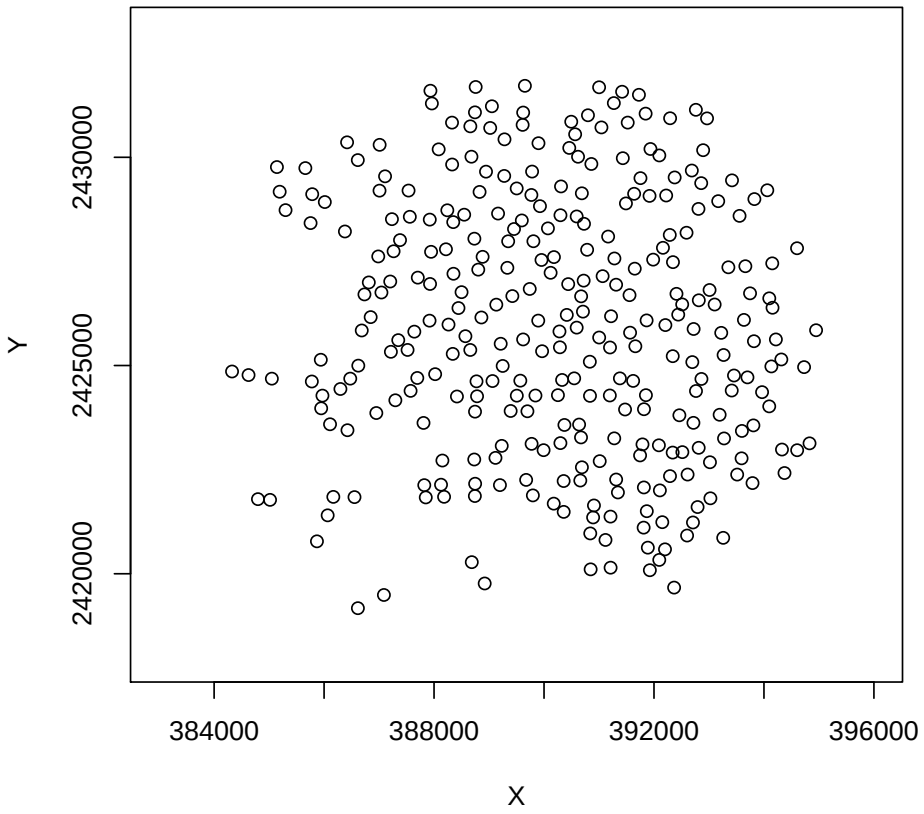


Figura 6.2: Gráfica de distribución de las trampas reducidas a una zona donde presenta la mayor concentración. Se observa que, en general, presentan un patrón de distancia de separación.

APÉNDICE E se muestran estos semivariogramas experimentales. Al mismo tiempo se trazan los modelos ajustados: un modelo Esférico, ajustado por mínimos cuadrados ponderados y dos modelos ajustados por máxima verosimilitud, el modelo Esférico y el Exponencial. A simple vista se observa que para los semivariogramas de las semanas 9, 40, 41 y 42 no existe buen ajuste por mínimos cuadrados y por máxima verosimilitud el ajuste no es muy bueno para las semanas 15, 21, 33, 39, 40, 41 y 42. Sin embargo según se puede observar en el Cuadro 6.1, en general existe un mejor ajuste al utilizar el modelo Esférico ajustado por el método de máxima verosimilitud. R utiliza la función `variofit` para el ajuste por mínimos cuadrados ponderados y `likfit` por máxima verosimilitud. En el APÉNDICE F se muestra los valores de los parámetros encontrados por mínimos cuadrados ponderados por el método de máxima verosimilitud de los modelos mencionados anteriormente. Se observa que los efectos pepita y meseta fueron muy variados entre semanas. Se puede ver en las tres tablas que en las semanas 33, 34, 23, 35 y 20 se presentan los mayores efectos pepita y meseta, que de igual manera es en estas semanas donde se presenta mayor media y varianza y es donde se reporta mayor densidad de población. El valor de la meseta afecta a la variación de las estimaciones, una meseta más alta indica mayor variación en las estimaciones. Para el caso del rango, un valor grande significa un comportamiento más continuo, las estimaciones dan como resultado mapas bastante lisos para la variable de interés.

6.5.1. Validación de los modelos ajustados

La validación de los diferentes modelos ajustados a los semivariogramas experimentales de cada una de las semanas muestreadas se realizó con el procedimientos denominado validación cruzada. La estrategia consiste en eliminar los datos uno por uno y predecirlos por kriging usando los datos que no fueron eliminados junto con el modelo de semivariograma a validar. Dado que si el modelo de semivarianza elegido describe bien la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el valor predicho debe ser pequeña. La Tabla 6.1 muestra el error cuadrado medio de los modelos obtenidos por mínimos cuadrados ponderados y por máxima verosimilitud para cada uno de los 42 conjuntos de datos. De acuerdo a esta tabla los modelos encontrados por máxima verosimilitud, presentan mejor ajuste que los que fueron encontrados por mínimos cuadrados ponderados, ya que tienen menor Error Cuadrado Medio. El ajuste por mínimos cuadrados ponderados para las semanas 18 y 24 presentan menor Error Cuadrado Medio en comparación con los encontrados por máxima verosimilitud.

Cuadro 6.1: Comparaciones de los modelos ajustados por Mínimos Cuadrados Ponderados y por Máxima Verosimilitud. En general, los modelos ajustados por Máxima Verosimilitu presentan menor Error cuadrático Medio (ECM).

SEMANA	ESFÉRICO (MCP)	ESFÉRICO (MV)	EXPONENCIAL (MV)
1	17.88	17.13	17.25
2	14.16	13.91	13.90
3	5.39	5.3	5.30
4	125.97	117.38	119.90
5	207.18	189.69	186.77
6	164.12	146.33	148.98
7	100.11	78.18	79.22
8	131.39	127.48	127.05
9	146.26	105.6	107.77
10	160.66	159.99	160.12
11	154.92	141.57	143.85
12	224.39	222.53	218.84
13	170.99	151.64	150.07
14	51.73	49.65	49.85
15	35.48	34.39	33.47
16	22.71	20.97	20.87
17	6.85	5.17	5.16
18	2.23	2.26	2.25
19	16.73	16.46	16.46
20	1884.03	1626.3	1641.78
21	5.48	4.54	4.55
22	17.57	15.07	15.29
23	686.45	600.59	603.74
24	78.03	78.19	78.04
25	75.05	66	66.57
26	55.02	53.21	53.76
27	5.35	5.3	5.26
28	100.72	81.86	82.00

SEMANA	ESFÉRICO (MCP)	ESFERICO (MV)	EXPONENCIAL (MV)
29	72.15	71.85	74.09
30	107.88	98.14	98.37
31	83.35	78.82	79.84
32	164.04	134.47	128.39
33	511.08	495.79	500.96
34	550.8	489.89	489.45
35	1023	803.38	800.93
36	261.07	223.72	229.48
37	248.25	225.42	222.81
38	254.1	250.97	253.71
39	262.57	257.95	260.77
40	230.06	214.86	217.22
41	317.78	274.93	276.43
42	188.28	197.09	200.60

6.5.2. Interpolación

Una vez determinados los modelos se procedió a realizar las estimaciones para la densidad de población de insectos en una parte del área de estudio y así poder tener un mejor conocimiento de la dispersión de dichos insectos en esa área. La Figura 6.3 muestra un rectángulo que limita esta área, el cual tiene 11500 m en el lado menor y 13000 m en el lado mayor cubriendo un área total de 149,500,000 metros cuadrados, aproximadamente 14 950 hectáreas. Se definió esta área para poder visualizar mejor el comportamiento de la variable en estudio ya que el área total de estudio también abarca lugares donde no existe plantación de frutales y para poder hacer las estimaciones se necesita de un área donde la variable sea contínua.

Para determinar el modelo de variograma que mejor describe el comportamiento de la variable, se escogió el variograma de menor error cuadrado medio. En la Tabla 6.2 se presenta el modelo de semivariograma ajustado para cada semana una vez evaluados los diferentes modelos con los resultados de la validación cruzada. Estos modelos fueron empleados para obtener las estimaciones de la densidad de las moscas en puntos no muestreados mediante la técnica del kriging ordinario, que permitió elaborar los mapas de incidencia (VER APÉNDICE G). Se utilizó este método ya que asume que las medias locales no están necesariamente relacionadas (lo más cercanamente) a la media poblacional.

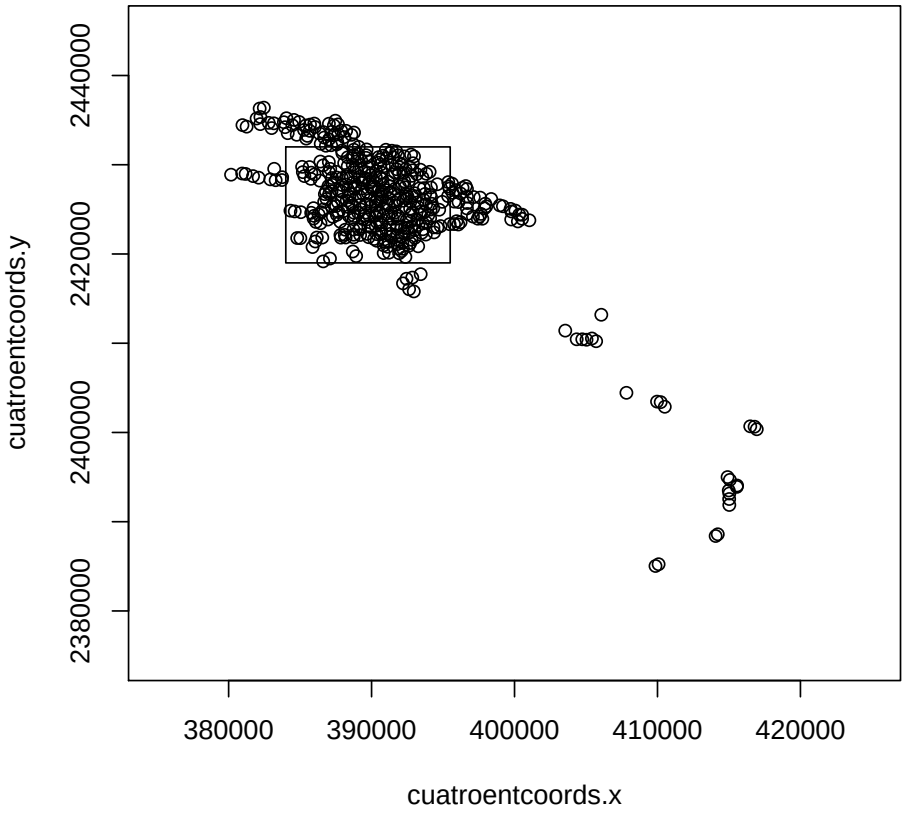


Figura 6.3: Gráfica de las limitaciones donde se hicieron las estimaciones.

Cuadro 6.2: Parámetros (efecto pepita, meseta y rango) de los modelos ajustados a los semivariogramas del número de moscas de la fruta para cada conjunto de datos.

SEMANA	MODELO	pepita	meseta	rango
1	ESFERICO	13.71	4.15	1333.16
2	EXPONENCIAL	13.17	1.31	2499.99
3	EXPONENCIAL	3.43	2.19	384.66
4	ESFERICO	82.38	57.6	1656.94
5	EXPONENCIAL	77.89	159.28	466.62
6	ESFERICO	101.82	79.56	1750.62
7	ESFERICO	58.71	36.36	1753.63
8	EXPONENCIAL	91.75	91.74	1699.99
9	ESFERICO	7.53	130.42	717
10	ESFERICO	135.7	43.49	2499.89
11	ESFERICO	50.85	122.11	862.5
12	EXPONENCIAL	6.54	244.91	280.61
13	EXPONENCIAL	41.66	139.02	381.11
14	ESFERICO	32.43	20.7	1068.58
15	EXPONENCIAL	0	38.04	353.09
16	EXPONENCIAL	13.28	33.45	26604.49
17	EXPONENCIAL	3.56	16.29	35396.72
18	ESFERICO	1.22	0.45	4054.68
19	ESFERICO	15.74	1.46	5152.89
20	ESFERICO	1042.39	1234.78	1780
21	ESFERICO	3.42	1.45	1731.81
22	ESFERICO	1.13	17.99	746.67
23	ESFERICO	372.71	473.78	1780
24	EXPONENCIAL	62.23	27.65	875.05
25	ESFERICO	44.44	43.77	1837.7
26	ESFERICO	39.11	26.42	2030.99
27	EXPONENCIAL	2.50	17.59	4299.99
28	ESFERICO	60.28	30.82	1511.47
29	ESFERICO	48.79	49.96	3283.99

SEMANA	MODELO	pepita	meseta	rango
30	ESFERICO	86.44	31.47	11493.11
31	ESFERICO	56.11	47.74	3664.48
32	EXPONENCIAL	0	203.44	561.24
33	ESFERICO	432.31	204.84	3480.5
34	EXPONENCIAL	144.99	637.31	703.72
35	EXPONENCIAL	165	998.78	405.41
36	ESFERICO	133.63	318.87	3407.7
37	EXPONENCIAL	71.01	218.25	557.46
38	ESFERICO	211.82	111.25	3705
39	ESFERICO	224.7	94.06	3507.72
40	ESFERICO	170.29	94.96	2499.9
41	ESFERICO	238.05	92.76	2999.9
42	ESFERICO	86.69	187.26	396.25

6.6. Interpretación de Resultados

Como se puede observar en la primera Figura del APÉNDICE G, la concentración de mosca de la fruta en la primera semana es baja y aumenta drásticamente en la segunda semana. Las partes de color claro indican la zona donde la densidad de individuos es grande en comparación con las zonas donde el color es más tenue.

Si se comparan las representaciones gráficas de las 42 semanas, se puede analizar la evolución temporal de la población de insectos en la región.

Fuera de observar que las mayores densidades se encuentran dentro del área del estudio y en algunas semanas avanzan hacia las orillas, los mapas obtenidos no muestran un patrón común en la distribución de los insectos. De la semana 2 a la semana 10 permanece constante, de la semana 11 a la semana 15 la densidad permanece baja y constante. De la semana 16 a la semana 19, la densidad vuelve a aumentar y disminuye rápidamente a partir de la semana 20 hasta ser casi cero en las semana 28, 32 y 35. Para las últimas semanas no se podría decir con exactitud el comportamiento de la variable, ya que como se vió en la sección anterior, el ajuste de un buen semivariograma teórico no fué posible debido a la dispersión tan irregular de los puntos del semivariograma experimental.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1. Se hace una presentación de los conceptos básicos necesarios para poder realizar un análisis espacial de datos, esperando que sea útil a investigadores y personas en general que quieran aplicar la Geoestadística.

2. El programa "R" fué de gran ayuda para hacer el análisis de los datos de la mosca de la fruta, presenta de forma fácil y gratuita las funciones necesarias para hacer el análisis geoestadístico a detalle. Esto una vez leído y repasado el manual existente en inglés o bien el que se presenta junto con el trabajo de tesis del pasante.

3. La aplicación de la Geoestadística para el análisis de los datos espaciales que se hizo en el presente trabajo y más precisamente la obtención de los mapas de distribución de la mosca de la fruta para los diferentes cultivos de la zona media de San Luis Potosí es de gran trascendencia, ya que ello permite conocer los lugares de mayor concentración de dicha plaga, predecir su cambio de posición en el tiempo y así conocer el área de mayor o menor ataque, esto con el fin de optimizar el recurso destinado para su control.

RECOMENDACIONES

1. Es recomendable que se extendiera el uso de la Geoestadística en los sistemas agrícolas como una herramienta para determinar la distribución de las diferentes plagas que atacan los cultivo para un mejor control de las mismas.

2. Sería deseable la revision contínua de las trampas durante todo el año, incluso durante varios años, ya que con ello se puede hacer un análisis más detallado sobre la densidad de mosca de la fruta en dicha área. Con ello podría hacerse un análisis de series de tiempo para poder observar si existe cierta estacionaridad y ver si hay meses o periodos en el año en los que la plaga llega a tener una densidad constante, mayor o menor en comparación con los otros meses.

3. Al igual que la Estadística clásica tiene su propia forma de recolectar información para un análisis detallado, la Geoestadística también tiene un método apropiado para recolectar información o medir la propiedad de la variable que se está estudiando. Desafortunadamente, el diseño de muestreo utilizado para la recolección de los datos con los que se trabajó no fue el adecuado, por lo que se recomienda utilizar un método de recolección adecuado si se desea continuar haciendo análisis de este tipo en la misma área. El método apropiado en Geoestadística es dividir toda el área de estudio en forma de cuadrícula, así, la distancia entre cada punto de norte a sur deberá ser la misma al igual que la de este-oeste, pero la distancia entre ambas direcciones puede ser diferente. Estas distancias y de ello, el total de puntos muestreados dependerán de la cantidad de recurso con que se disponga y del punto de vista del especialista según el tema que trate.

4. Extender el uso del paquete computacional "R" por su fácil manejo y acceso gratuito.

Apéndice A

ESTADÍSTICOS DE RESUMEN

Cuadro A.1: Estadísticas de Resumen para cada una de las 42 semanas muestreadas.

semana	media	moda	mediana	minimo	maximo	Q1	Q3
1	1.74	0	0	0	71	0	2
2	2.32	0	1	0	29	0	3
3	1.40	0	0	0	20	0	2
4	8.51	0	3	0	74	1	12
5	10.91	0	5	0	98	1	14.25
6	9.53	0	4	0	94	1	13
7	7.10	0	3	0	82	0	10
8	8.44	0	3	0	75	0	10.25
9	7.87	0	3	0	79	0	10
10	8.78	0	3	0	95	0	12
11	7.77	0	3	0	92	0	9
12	10.45	0	4	0	98	1	13
13	7.80	0	3	0	97	0	10
14	3.29	0	1	0	73	0	4
15	1.60	0	0	0	108	0	1
16	0.61	0	0	0	86	0	0
17	0.51	0	0	0	29	0	0
18	0.55	0	0	0	17	0	0
19	1.65	0	0	0	46	0	2
20	28.37	0	10	0	396	3	30

semana	media	moda	mediana	minimo	maximo	Q1	Q3
21	0.86	0	0	0	22	0	1
22	1.73	0	0	0	37	0	1
23	17.35	0	6	0	218	1	19
24	3.82	0	0	0	90	0	3
25	4.61	0	1	0	84	0	4
26	3.38	0	0	0	92	0	3
27	0.53	0	0	0	34	0	0
28	5.01	0	1	0	120	0	6
29	5.46	0	2	0	82	0	7
30	5.44	0	2	0	115	0	7
31	5.07	0	1	0	85	0	6
32	7.53	0	3	0	158	0	8
33	11.68	0	3.5	0	290	0	12
34	16.91	0	5	0	213	0	19
35	17.85	0	7	0	546	1	23
36	12.83	0	5	0	122	1	17
37	12.50	0	6	0	128	1	16
38	13.50	0	5	0	95	1	18.25
39	13.51	0	6	0	128	1	19
40	10.17	0	3	0	117	0	13
41	13.13	0	6	0	119	1	17
42	11.88	0	6	0	101	1	17

Cuadro A.2: Estadísticas de Resumen para cada una de las 42 semanas muestreadas.

semana	VARIANZA	DESV.EST	RIC	CA	CC	CV
1	17.90	4.23	2	10.73	164.37	2.43
2	14.42	3.80	3	2.94	12.11	1.64
3	5.67	2.38	2	2.90	12.42	1.70
4	142.30	11.93	11	2.32	6.60	1.40
5	249.38	15.79	13.25	2.61	8.06	1.45
6	187.27	13.68	12	2.50	7.83	1.44
7	99.89	9.99	10	2.57	9.90	1.41
8	165.61	12.87	10.25	2.47	6.75	1.52
9	145.94	12.08	10	2.45	7.44	1.53
10	179.93	13.41	12	2.59	8.63	1.53
11	182.22	13.50	9	3.15	12.08	1.74
12	252.50	15.89	12	2.72	8.54	1.52
13	186.12	13.64	10	3.41	14.12	1.75
14	53.17	7.29	4	5.16	36.76	2.22
15	35.23	5.94	1	13.50	234.28	3.71
16	18.13	4.26	0	18.47	367.69	7.03
17	5.37	2.32	0	8.87	92.72	4.51
18	2.53	1.59	0	5.77	43.12	2.92
19	17.24	4.15	2	5.82	44.99	2.52
20	2413.53	49.13	27	3.62	17.22	1.73
21	4.80	2.19	1	5.36	38.49	2.55
22	20.20	4.49	1	4.41	23.46	2.61
23	889.14	29.82	18	3.43	15.01	1.72
24	92.57	9.62	3	4.77	29.78	2.52
25	91.57	9.57	4	4.13	23.80	2.07
26	66.60	8.16	3	5.18	39.00	2.42
27	8.62	2.94	0	7.49	62.98	5.55
28	92.12	9.60	6	5.46	50.42	1.92
29	86.87	9.32	7	3.41	16.55	1.71
30	105.29	10.26	7	4.87	36.96	1.89
31	94.22	9.71	6	3.92	20.62	1.92

semana	VARIANZA	DESV.EST	RIC	CA	CC	CV
32	231.29	15.21	8	5.37	39.17	2.02
33	702.02	26.50	12	6.44	55.35	2.27
34	908.39	30.14	19	3.36	14.00	1.78
35	1319.47	36.32	22	8.30	105.29	2.03
36	346.96	18.63	16	2.59	8.27	1.45
37	306.55	17.51	15	2.66	9.98	1.40
38	347.85	18.65	17.25	2.01	4.03	1.38
39	350.18	18.71	18	2.41	7.80	1.39
40	278.20	16.68	13	2.88	10.63	1.64
41	343.05	18.52	16	2.65	9.05	1.41
42	254.30	15.95	16	2.39	7.16	1.34

Apéndice B

HISTOGRAMAS

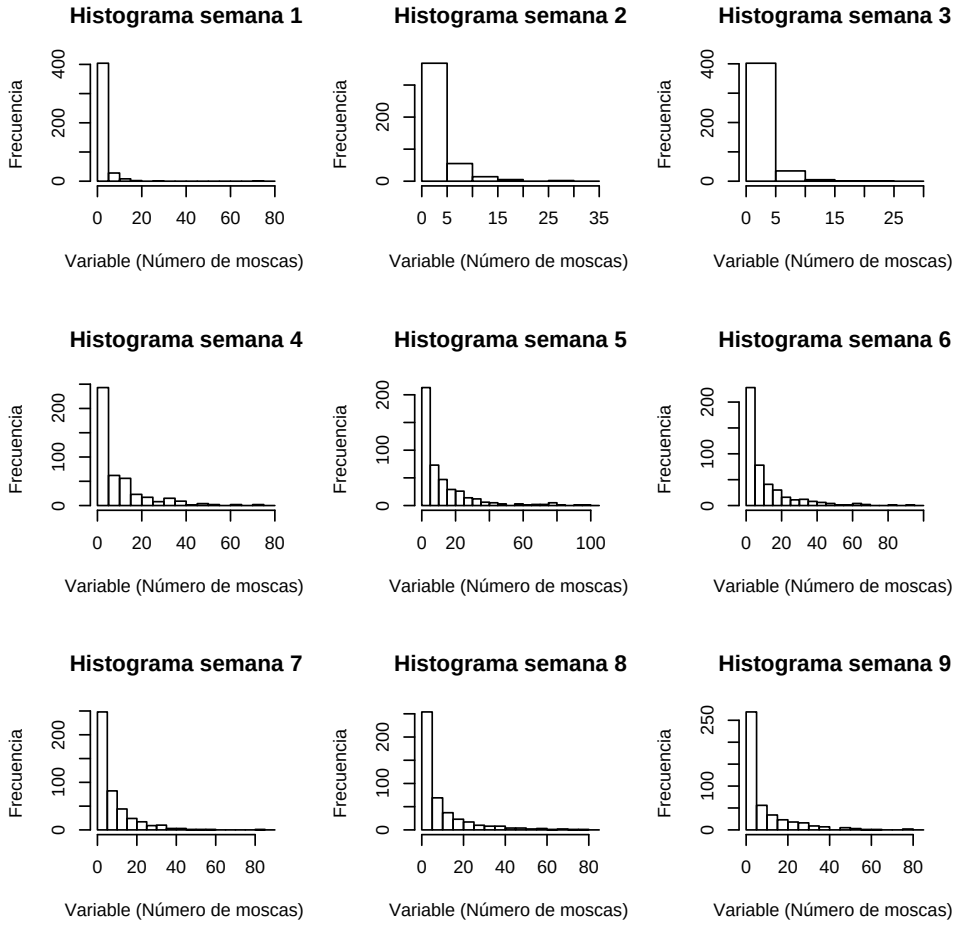


Figura B.1: Histogramas correspondientes a los conjuntos de datos para las primeras 9 semanas muestreadas

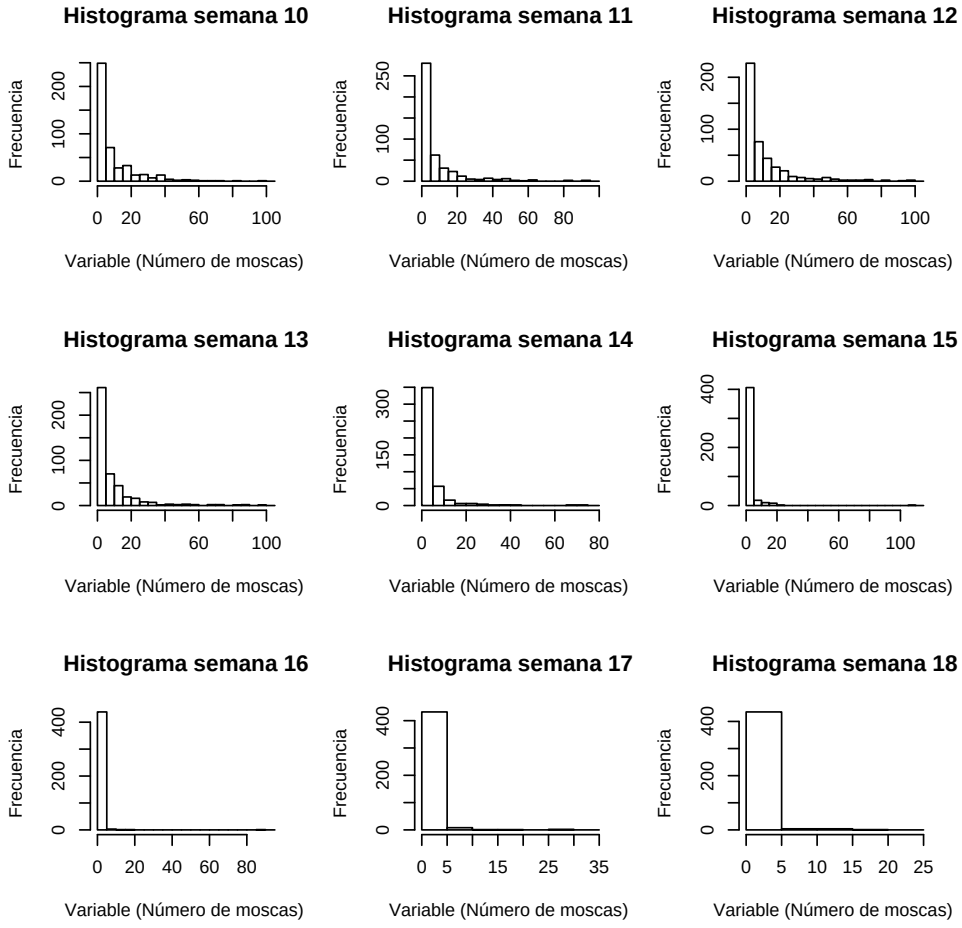


Figura B.2: Histogramas correspondientes a los conjuntos de datos para las semanas de la 10 a la 18

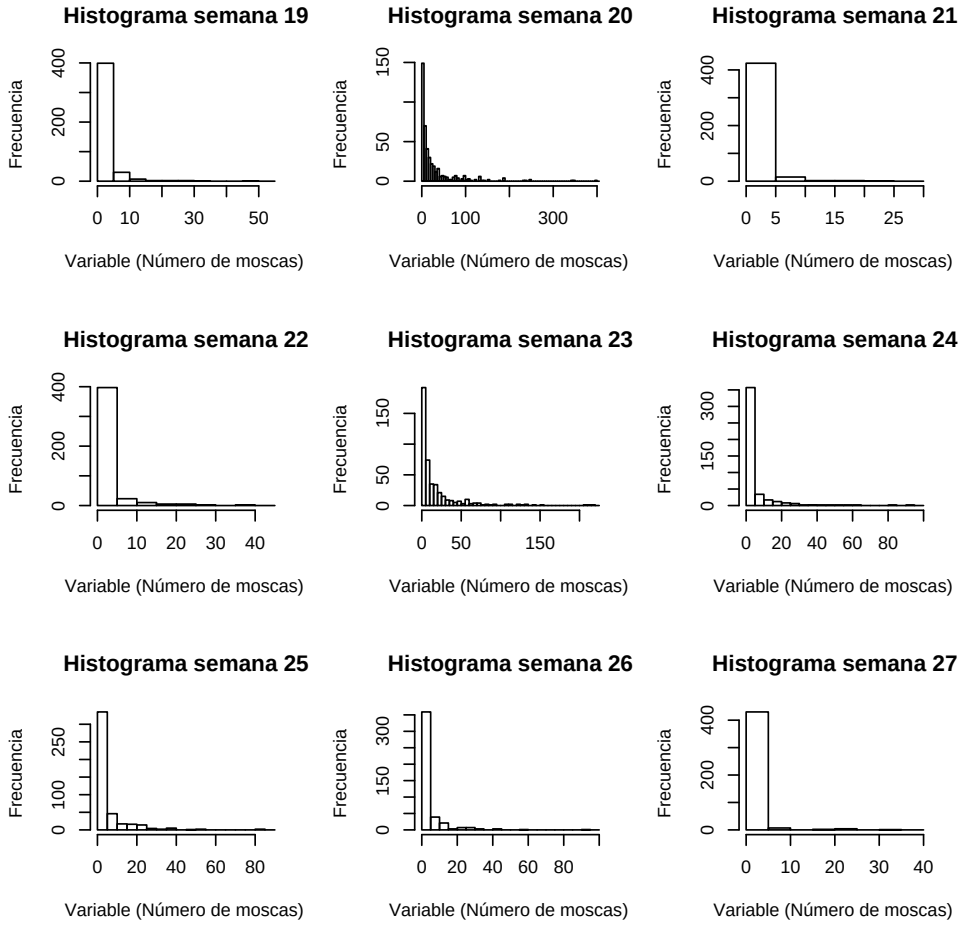


Figura B.3: Histogramas correspondientes a los conjuntos de datos para las semanas de la 19 a la 27

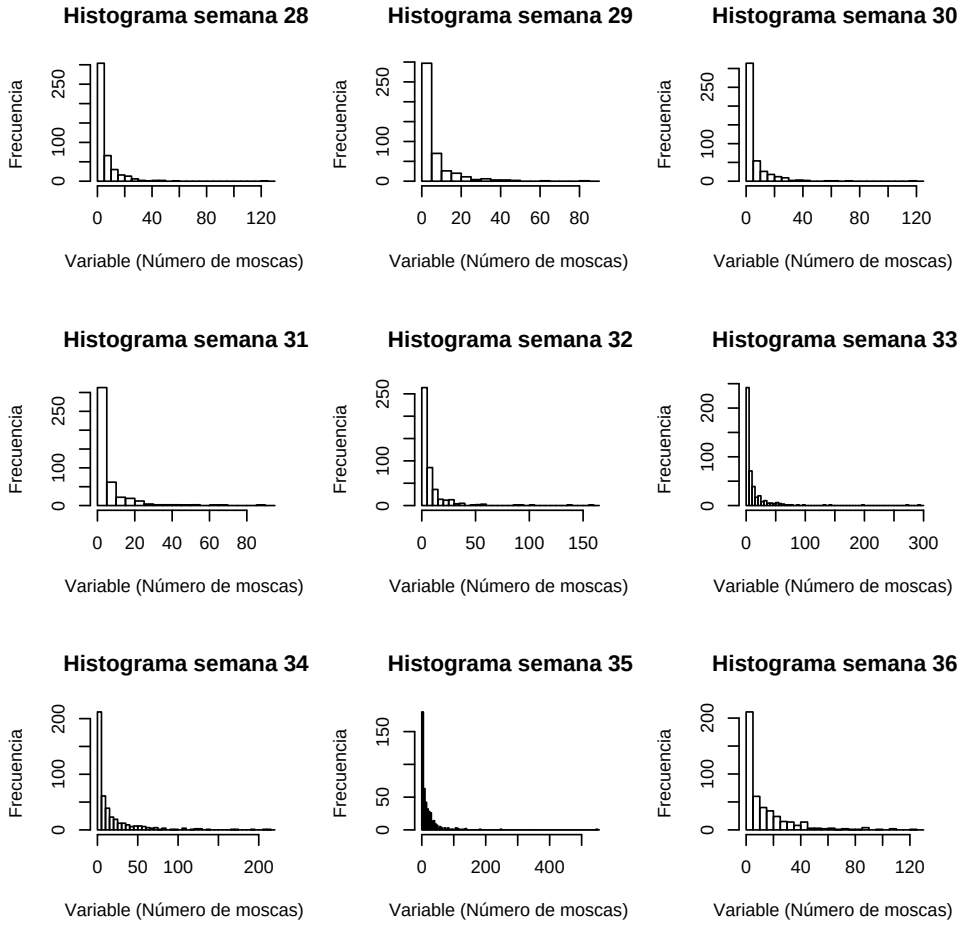


Figura B.4: Histogramas correspondientes a los conjuntos de datos para las semanas de la 28 a la 36

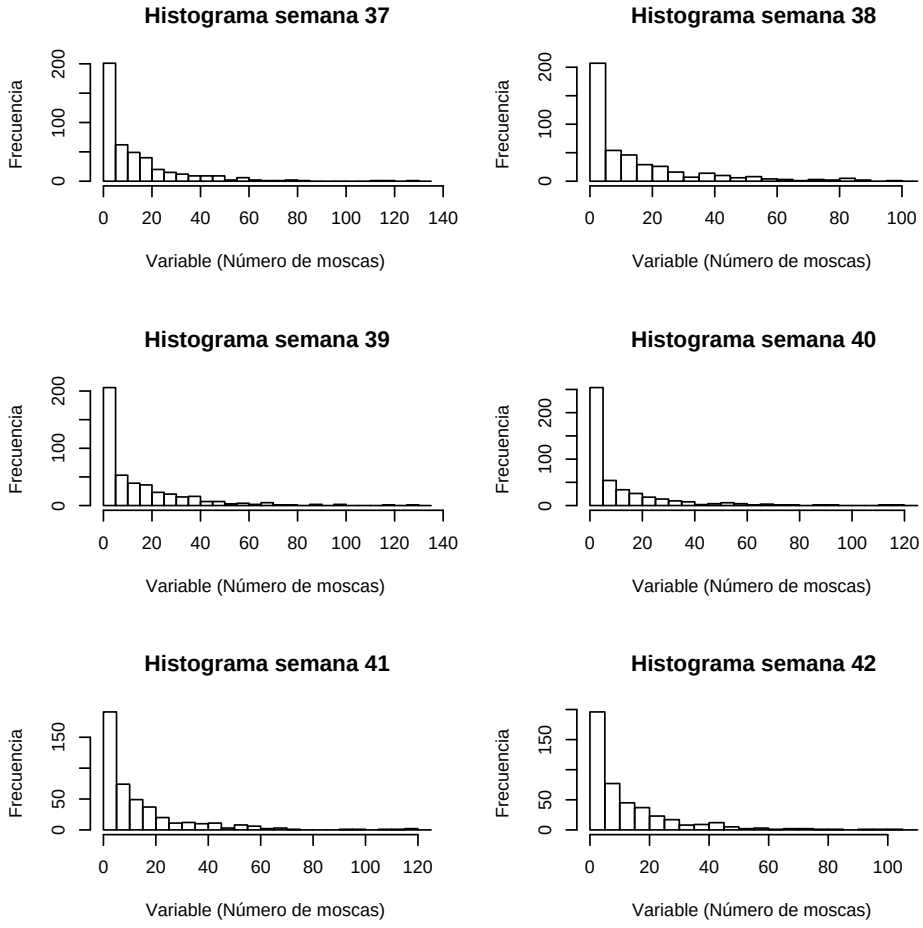


Figura B.5: Histogramas correspondientes a los conjuntos de datos para las semanas de la 37 a la 42

Apéndice C

CORRELACIONES

Cuadro C.1: Correlaciones correspondientes a las semanas de la 1 a la 11

semana	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11
S01	1	0.48	0.22	0.25	0.13	0.13	0.12	0.16	0.06	0.20	0.17
S02	0	1	0.41	0.49	0.40	0.33	0.28	0.29	0.29	0.33	0.29
S03	0	0	1	0.56	0.47	0.44	0.43	0.33	0.39	0.18	0.15
S04	0	0	0	1	0.61	0.65	0.45	0.63	0.45	0.39	0.31
S05	0	0	0	0	1	0.59	0.65	0.55	0.59	0.47	0.38
S06	0	0	0	0	0	1	0.59	0.67	0.53	0.43	0.38
S07	0	0	0	0	0	0	1	0.56	0.59	0.48	0.41
S08	0	0	0	0	0	0	0	1	0.58	0.51	0.47
S09	0	0	0	0	0	0	0	0	1	0.58	0.58
S10	0	0	0	0	0	0	0	0	0	1	0.73
S11	0	0	0	0	0	0	0	0	0	0	1

Cuadro C.2: Correlaciones correspondientes a las semanas de la 12 a la 22

semana	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
S01	0.20	0.08	0.07	0.03	0.04	0.05	0.19	0.11	0.06	0.07	0.10
S02	0.30	0.21	0.12	0.14	0.10	0.13	0.18	0.30	0.17	0.09	0.17
S03	0.17	0.17	0.09	0.08	-0.02	-0.01	0.07	0.19	0.17	0.07	0.14
S04	0.32	0.24	0.18	0.09	0.13	0.18	0.29	0.34	0.30	0.20	0.27

semana	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
S05	0.32	0.33	0.22	0.10	-0.01	-0.01	0.06	0.28	0.28	0.07	0.21
S06	0.30	0.26	0.18	0.09	0.09	0.10	0.26	0.34	0.34	0.18	0.23
S07	0.33	0.34	0.28	0.16	-0.02	-0.05	0.02	0.19	0.31	0.05	0.21
S08	0.50	0.30	0.36	0.11	0.27	0.16	0.36	0.28	0.28	0.29	0.26
S09	0.40	0.42	0.36	0.21	0.05	0.02	0.17	0.26	0.29	0.20	0.19
S10	0.63	0.47	0.35	0.13	0.04	0.07	0.25	0.33	0.43	0.24	0.42
S11	0.69	0.60	0.41	0.16	0.18	0.16	0.30	0.34	0.51	0.24	0.45
S12	1	0.65	0.56	0.17	0.18	0.09	0.28	0.22	0.43	0.22	0.48
S13	0	1	0.49	0.26	0.05	0.00	0.16	0.22	0.52	0.23	0.45
S14	0	0	1	0.18	0.09	0.06	0.16	0.12	0.28	0.15	0.22
S15	0	0	0	1	0.04	0.11	0.12	0.10	0.13	0.13	0.08
S16	0	0	0	0	1	0.24	0.38	0.09	0.09	0.24	0.24
S17	0	0	0	0	0	1	0.40	0.16	0.04	0.24	0.11
S18	0	0	0	0	0	0	1	0.49	0.24	0.52	0.29
S19	0	0	0	0	0	0	0	1	0.38	0.38	0.26
S20	0	0	0	0	0	0	0	0	1	0.25	0.57
S21	0	0	0	0	0	0	0	0	0	1	0.36
S22	0	0	0	0	0	0	0	0	0	0	1

Cuadro C.3: Correlaciones correspondientes a las semanas de la 23 a la 32

semana	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32
S01	0.05	0.06	0.05	0.04	0.00	0.02	0.08	0.11	0.10	0.01
S02	0.16	0.13	0.15	0.13	0.01	0.10	0.23	0.27	0.25	0.18
S03	0.16	0.15	0.15	0.13	0.01	0.08	0.16	0.08	0.16	0.19
S04	0.27	0.27	0.25	0.19	-0.04	0.18	0.18	0.18	0.21	0.24
S05	0.26	0.27	0.23	0.17	0.01	0.16	0.15	0.13	0.12	0.15
S06	0.32	0.34	0.25	0.18	0.01	0.24	0.19	0.11	0.16	0.23
S07	0.29	0.30	0.25	0.22	-0.02	0.18	0.14	0.09	0.08	0.09
S08	0.25	0.27	0.19	0.16	0.00	0.18	0.17	0.10	0.14	0.20
S09	0.27	0.29	0.22	0.16	0.03	0.18	0.16	0.12	0.08	0.06
S10	0.40	0.32	0.37	0.33	0.04	0.26	0.24	0.18	0.18	0.17
S11	0.48	0.43	0.45	0.30	0.12	0.32	0.22	0.12	0.10	0.05
S12	0.39	0.34	0.36	0.31	0.07	0.24	0.24	0.18	0.20	0.19

semana	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32
S13	0.50	0.42	0.46	0.38	0.11	0.30	0.25	0.15	0.12	0.16
S14	0.28	0.19	0.25	0.17	0.21	0.21	0.13	0.08	0.07	0.07
S15	0.12	0.12	0.11	0.09	0.01	0.06	0.08	0.04	-0.02	0.00
S16	0.05	0.06	0.06	0.02	0.01	0.03	0.05	0.02	0.01	0.04
S17	0.02	0.02	0.01	-0.01	-0.02	0.05	0.21	0.20	0.25	0.15
S18	0.19	0.15	0.20	0.17	-0.03	0.12	0.09	0.07	0.01	0.01
S19	0.34	0.34	0.30	0.24	-0.03	0.23	0.12	0.04	0.02	0.00
S20	0.99	0.83	0.91	0.74	0.16	0.65	0.48	0.34	0.31	0.31
S21	0.23	0.09	0.11	0.23	0.10	0.28	0.19	0.17	0.04	0.13
S22	0.51	0.35	0.45	0.41	0.07	0.41	0.24	0.16	0.16	0.25
S23	1	0.77	0.88	0.78	0.23	0.73	0.51	0.36	0.33	0.30
S24	0	1	0.67	0.45	-0.03	0.34	0.25	0.16	0.17	0.21
S25	0	0	1	0.67	0.08	0.48	0.45	0.33	0.33	0.28
S26	0	0	0	1	0.12	0.42	0.43	0.32	0.24	0.22
S27	0	0	0	0	1	0.26	0.13	0.18	0.10	0.09
S28	0	0	0	0	0	1	0.49	0.31	0.28	0.24
S29	0	0	0	0	0	0	1	0.68	0.63	0.44
S30	0	0	0	0	0	0	0	1	0.73	0.45
S31	0	0	0	0	0	0	0	0	1	0.64
S32	0	0	0	0	0	0	0	0	0	1.00

Cuadro C.4: Correlaciones correspondientes a las semanas de la 32 a la 44

semana	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42
S01	0.00	0.03	0.03	0.11	0.15	0.13	0.09	0.14	0.09	0.12
S02	0.17	0.18	0.20	0.23	0.18	0.19	0.13	0.20	0.24	0.21
S03	0.22	0.23	0.24	0.21	0.21	0.17	0.17	0.21	0.33	0.28
S04	0.29	0.32	0.30	0.29	0.33	0.29	0.28	0.28	0.38	0.34
S05	0.21	0.26	0.26	0.29	0.30	0.25	0.28	0.26	0.40	0.31
S06	0.33	0.32	0.33	0.33	0.32	0.36	0.35	0.28	0.44	0.35
S07	0.18	0.24	0.19	0.27	0.26	0.31	0.34	0.26	0.38	0.33
S08	0.30	0.32	0.30	0.23	0.26	0.27	0.27	0.23	0.35	0.32
S09	0.15	0.19	0.12	0.18	0.23	0.24	0.34	0.21	0.38	0.25
S10	0.23	0.31	0.21	0.26	0.30	0.34	0.35	0.19	0.27	0.24

semana	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42
S11	0.05	0.14	0.13	0.17	0.26	0.24	0.31	0.17	0.18	0.12
S12	0.19	0.24	0.21	0.22	0.25	0.24	0.28	0.17	0.22	0.16
S13	0.12	0.24	0.22	0.26	0.29	0.21	0.28	0.23	0.22	0.19
S14	0.07	0.08	0.06	0.08	0.11	0.11	0.14	0.08	0.17	0.11
S15	0.00	0.02	0.03	0.04	0.09	0.01	0.06	0.02	0.06	0.06
S16	0.00	0.04	0.01	0.02	0.05	0.01	0.00	-0.01	-0.01	-0.02
S17	0.02	0.04	0.05	-0.01	0.11	0.04	0.01	-0.02	-0.02	-0.01
S18	0.03	0.03	0.01	0.03	0.07	0.03	0.05	0.03	-0.03	0.05
S19	0.06	0.11	0.11	0.12	0.12	0.14	0.18	0.18	0.15	0.16
S20	0.22	0.44	0.35	0.45	0.40	0.42	0.38	0.33	0.32	0.30
S21	0.07	0.06	0.09	0.16	0.21	0.14	0.12	0.21	0.05	0.13
S22	0.19	0.33	0.32	0.36	0.34	0.28	0.30	0.22	0.19	0.23
S23	0.21	0.42	0.32	0.44	0.37	0.43	0.37	0.33	0.30	0.30
S24	0.14	0.34	0.28	0.36	0.36	0.29	0.32	0.29	0.34	0.23
S25	0.24	0.43	0.31	0.39	0.33	0.37	0.31	0.25	0.27	0.26
S26	0.17	0.31	0.23	0.35	0.25	0.36	0.31	0.26	0.17	0.23
S27	-0.02	-0.07	0.00	0.06	-0.01	0.06	0.02	0.06	-0.02	0.07
S28	0.13	0.27	0.22	0.28	0.25	0.35	0.24	0.24	0.19	0.22
S29	0.26	0.39	0.29	0.36	0.29	0.34	0.20	0.23	0.23	0.23
S30	0.25	0.29	0.22	0.33	0.22	0.35	0.16	0.16	0.10	0.16
S31	0.44	0.47	0.47	0.43	0.33	0.37	0.17	0.16	0.20	0.20
S32	0.54	0.69	0.70	0.57	0.51	0.41	0.22	0.29	0.26	0.29
S33	1	0.67	0.54	0.49	0.42	0.49	0.42	0.35	0.48	0.46
S34	0	1	0.68	0.67	0.62	0.60	0.48	0.38	0.47	0.45
S35	0	0	1	0.67	0.57	0.49	0.40	0.37	0.47	0.44
S36	0	0	0	1	0.72	0.68	0.56	0.50	0.48	0.48
S37	0	0	0	0	1	0.66	0.61	0.60	0.55	0.47
S38	0	0	0	0	0	1	0.75	0.55	0.50	0.49
S39	0	0	0	0	0	0	1	0.62	0.60	0.55
S40	0	0	0	0	0	0	0	1	0.61	0.61
S41	0	0	0	0	0	0	0	0	1	0.67
S42	0	0	0	0	0	0	0	0	0	1

Apéndice D

COVARIANZAS

Cuadro D.1: Covarianzas correspondientes a las semanas de la 1 a la 11

SEMANA	SO1	SO2	SO3	SO4	SO5	SO6	SO7	SO8	SO9	S10	S11
SO1	17.9	7.6	2.2	12.7	8.4	7.6	5.0	8.5	3.3	11.5	9.9
SO2	0	14.4	3.7	22.1	24.2	17.3	10.7	14.0	13.2	17.0	14.7
SO3	0	0	5.7	15.9	17.6	14.3	10.3	10.2	11.3	5.7	4.8
SO4	0	0	0	142.3	115.2	106.8	53.6	96.0	64.2	61.8	50.0
SO5	0	0	0	0	249.4	127.9	101.9	111.5	112.9	100.2	80.3
SO6	0	0	0	0	0	187.3	80.1	117.3	87.3	78.4	70.1
SO7	0	0	0	0	0	0	99.9	71.6	71.5	64.4	55.4
SO8	0	0	0	0	0	0	0	165.6	89.9	87.4	81.5
SO9	0	0	0	0	0	0	0	0	145.9	93.7	94.5
S10	0	0	0	0	0	0	0	0	0	179.9	132.3
S11	0	0	0	0	0	0	0	0	0	0	182.2

Cuadro D.2: Covarianzas correspondientes a las semanas de la 12 a la 22

SEMANA	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
SO1	13.3	4.8	2.0	0.7	0.7	0.5	1.3	1.9	13.4	0.7	1.9
SO2	18.2	11.0	3.3	3.2	1.6	1.1	1.1	4.7	32.5	0.8	3.0
SO3	6.3	5.6	1.6	1.1	-0.2	-0.1	0.3	1.9	20.0	0.3	1.5
SO4	61.2	39.5	15.4	6.7	6.7	4.8	5.6	16.6	175.4	5.2	14.5

SEMANA	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21	S22
SO5	81.2	71.0	25.8	9.5	-0.5	-0.2	1.4	18.6	215.3	2.3	14.9
SO6	64.6	47.8	18.4	7.3	5.2	3.0	5.6	19.2	227.1	5.4	14.3
SO7	52.3	47.0	20.0	9.2	-0.7	-1.2	0.2	7.8	150.5	1.1	9.4
SO8	103.1	52.0	33.3	8.6	14.9	4.6	7.4	14.8	174.4	8.2	15.0
SO9	76.6	69.2	31.9	15.0	2.8	0.7	3.2	13.0	170.3	5.2	10.5
SO10	133.6	85.3	34.7	10.4	2.5	2.3	5.4	18.5	282.2	6.9	25.2
SO11	147.4	110.3	40.6	12.8	10.1	4.9	6.4	18.8	341.4	7.1	27.5
SO12	252.5	140.7	64.8	16.3	12.2	3.3	7.0	14.5	334.9	7.7	34.6
SO13	0	186.1	48.6	20.9	2.7	0.0	3.5	12.3	351.3	7.0	27.3
SO14	0	0	53.2	8.0	2.7	1.0	1.9	3.5	101.2	2.4	7.3
SO15	0	0	0	35.2	0.9	1.5	1.2	2.5	37.4	1.6	2.0
SO16	0	0	0	0	18.1	2.4	2.6	1.5	18.8	2.2	4.5
SO17	0	0	0	0	0	5.4	1.5	1.5	4.4	1.2	1.2
SO18	0	0	0	0	0	0	2.5	3.2	18.4	1.8	2.1
SO19	0	0	0	0	0	0	0	17.2	76.7	3.5	4.9
SO20	0	0	0	0	0	0	0	0	2413.5	27.3	125.9
SO21	0	0	0	0	0	0	0	0	0	4.8	3.6
SO22	0	0	0	0	0	0	0	0	0	0	20.2

Cuadro D.3: Covarianzas correspondientes a las semanas de la 23 a la 32

SEMANA	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32
SO1	6.4	2.3	2.2	1.3	-0.1	0.7	3.0	5.0	4.0	0.9
SO2	18.3	4.8	5.6	4.2	0.1	3.6	8.2	10.4	9.4	10.2
SO3	11.4	3.4	3.3	2.6	0.1	1.9	3.6	1.9	3.6	6.7
SO4	96.6	30.7	28.5	18.6	-1.4	20.1	20.3	21.6	24.7	42.9
SO5	122.7	40.3	35.1	21.8	0.6	24.9	22.3	20.6	18.3	35.7
SO6	129.9	44.5	33.0	20.6	0.4	31.5	24.3	16.0	21.3	47.0
SO7	87.1	28.9	24.0	17.6	-0.6	17.2	13.0	9.4	7.5	13.0
SO8	95.0	33.1	23.0	16.7	0.1	22.0	20.9	13.5	18.0	39.4
SO9	96.0	33.2	25.3	15.3	0.9	21.2	18.1	14.9	9.5	11.1
SO10	160.1	41.8	48.1	35.7	1.5	32.9	29.8	24.5	23.2	35.6
SO11	192.9	55.8	58.2	33.5	4.7	40.8	27.4	16.2	12.9	11.0
SO12	186.5	51.5	54.5	39.8	3.5	37.1	36.2	28.6	30.2	46.7

SEMANA	S23	S24	S25	S26	S27	S28	S29	S30	S31	S32
SO13	201.5	55.0	60.5	42.6	4.5	38.9	31.7	20.3	15.3	33.9
SO14	60.4	13.4	17.7	9.9	4.4	15.0	9.0	5.7	4.8	7.9
SO15	20.8	6.9	6.1	4.2	0.2	3.4	4.7	2.5	-1.1	0.2
SO16	7.0	2.5	2.6	0.7	0.1	1.2	1.8	0.7	0.4	2.6
SO17	1.4	0.4	0.2	-0.1	-0.1	1.1	4.6	4.6	5.6	5.3
SO18	9.2	2.3	3.0	2.2	-0.1	1.9	1.3	1.2	0.2	0.4
SO19	42.6	13.8	11.9	8.3	-0.3	9.0	4.8	1.8	0.7	0.2
SO20	1443.2	390.8	426.3	296.4	23.3	306.3	218.3	170.8	148.2	230.9
SO21	14.8	1.9	2.2	4.0	0.7	5.9	3.9	3.8	0.9	4.3
SO22	67.9	15.0	19.3	15.1	1.0	17.6	9.9	7.4	6.8	17.0
SO23	889.1	219.7	251.6	189.9	20.3	207.6	142.0	111.4	94.5	138.0
SO24	0	92.6	61.6	35.1	-0.9	31.3	22.1	15.8	15.4	30.8
SO25	0	0	91.6	52.3	2.3	43.9	40.2	32.4	30.5	40.8
SO26	0	0	0	66.6	2.9	33.0	32.6	27.2	19.3	27.1
SO27	0	0	0	0	8.6	7.3	3.4	5.5	2.9	4.0
SO28	0	0	0	0	0	92.1	43.6	30.5	26.3	35.3
SO29	0	0	0	0	0	0	86.9	65.4	56.8	61.7
SO30	0	0	0	0	0	0	0	105.3	72.3	70.1
SO31	0	0	0	0	0	0	0	0	94.2	95.0
SO32	0	0	0	0	0	0	0	0	0	231.3

Cuadro D.4: Covarianzas correspondientes a las semanas de la 33 a la 42

SEMANA	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42
SO1	-0.5	3.7	5.0	8.9	10.8	10.2	7.4	10.2	6.7	7.8
SO2	16.8	20.1	27.9	16.5	11.8	13.1	9.1	12.5	16.6	12.5
SO3	14.2	16.4	20.7	9.4	8.7	7.8	7.7	8.2	14.7	10.7
SO4	90.4	113.6	129.0	64.0	68.1	65.3	62.0	56.4	83.8	63.8
SO5	88.2	125.8	151.5	85.0	84.1	73.9	82.6	69.0	117.2	77.9
SO6	120.2	131.6	165.5	84.9	77.4	91.5	90.7	64.7	111.9	76.4
SO7	48.1	73.7	69.0	50.9	44.9	58.1	63.4	43.1	69.8	52.3
SO8	101.2	124.9	142.4	55.5	58.9	63.7	64.8	49.7	84.5	66.6
SO9	49.2	69.4	53.7	41.2	48.6	53.3	77.8	42.5	84.0	48.0
SO10	81.8	124.8	104.5	65.4	69.8	84.7	87.3	43.6	67.7	51.5

SEMANA	S33	S34	S35	S36	S37	S38	S39	S40	S41	S42
SO11	18.1	58.8	64.6	43.6	61.1	61.4	77.8	37.8	44.0	26.6
SO12	78.1	113.6	122.1	65.8	69.9	72.4	84.2	44.4	63.7	41.3
SO13	44.3	100.7	109.7	65.1	69.5	53.9	70.9	52.6	54.8	42.3
SO14	14.4	16.9	16.5	10.6	14.0	14.9	18.6	9.7	22.9	12.9
SO15	-0.4	4.4	6.1	3.9	9.6	0.8	7.1	1.7	6.5	5.8
SO16	0.4	4.9	1.4	1.7	3.4	0.6	0.3	-0.4	-1.2	-1.3
SO17	1.4	2.6	4.0	-0.4	4.6	1.6	0.3	-0.8	-0.7	-0.5
SO18	1.3	1.6	0.5	0.8	1.9	1.0	1.6	0.8	-0.9	1.3
SO19	6.6	13.7	15.9	9.7	9.1	10.6	14.1	12.7	11.6	10.3
SO20	290.8	646.8	619.6	413.1	344.4	384.1	348.6	273.5	292.3	237.5
SO21	3.8	3.9	7.4	6.4	8.2	5.7	5.0	7.6	2.0	4.6
SO22	22.2	44.8	52.8	29.9	27.1	23.6	25.0	16.2	16.2	16.8
SO23	167.0	374.4	352.0	241.8	192.6	237.7	205.2	163.3	166.0	141.4
SO24	35.8	98.5	98.8	64.8	60.7	51.7	57.8	47.0	60.3	35.6
SO25	62.0	125.1	108.8	70.1	55.7	65.4	55.7	39.4	47.8	39.1
SO26	36.3	77.2	67.0	53.9	35.1	54.9	46.8	35.6	25.1	29.7
SO27	-1.3	-5.9	-0.5	3.0	-0.7	3.5	0.9	3.2	-1.3	3.3
SO28	34.2	79.4	77.9	50.0	41.8	62.2	44.0	38.1	34.0	33.6
SO29	65.3	108.2	97.4	62.7	47.7	58.4	34.0	35.9	40.5	34.3
SO30	69.2	89.5	82.1	63.1	39.3	66.1	30.8	26.8	19.3	26.2
SO31	112.6	138.4	167.5	78.0	56.0	67.1	30.6	25.2	36.2	31.6
SO32	218.5	314.1	384.7	160.6	136.4	117.5	63.7	72.4	74.0	70.6
SO33	702.0	537.1	520.1	239.9	195.8	241.8	208.7	152.8	235.8	192.3
SO34	0	908.4	749.8	374.1	329.6	336.1	270.6	192.5	264.0	217.8
SO35	0	0	1319.5	455.7	365.2	333.8	269.5	226.2	316.2	252.3
SO36	0	0	0	347.0	233.8	236.3	195.4	155.5	165.8	141.4
SO37	0	0	0	0	306.6	215.0	198.8	175.1	177.6	129.8
SO38	0	0	0	0	0	347.9	260.9	169.8	174.1	144.7
SO39	0	0	0	0	0	0	350.2	192.1	208.4	163.8
SO40	0	0	0	0	0	0	0	278.2	188.6	161.5
SO41	0	0	0	0	0	0	0	0	343.1	196.8
SO42	0	0	0	0	0	0	0	0	0	254.3

Apéndice E

DIFERENTES MODELOS AJUSTADOS A LOS SEMIVARIOGRAMAS EXPERIMENTALES

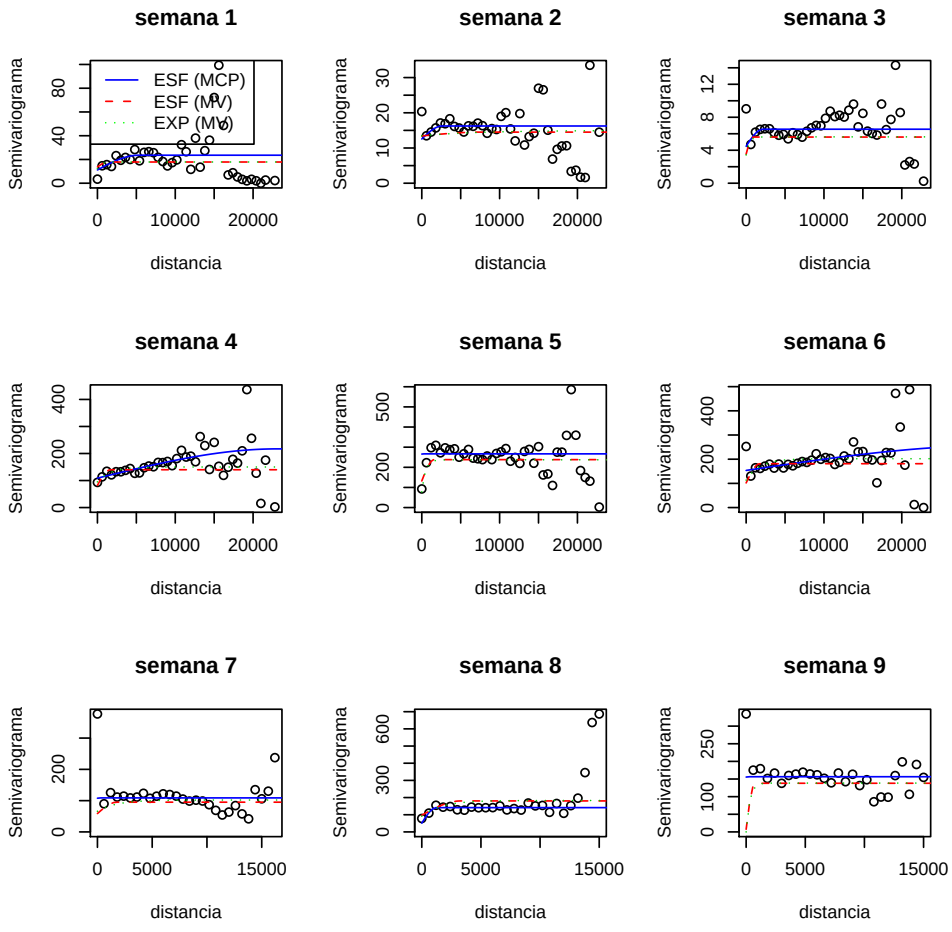


Figura E.1: Gráfica de los variogramas direccionales y modelos ajustados por mínimos cuadrados ponderados (MCP) y por máxima verosimilitud (MV)

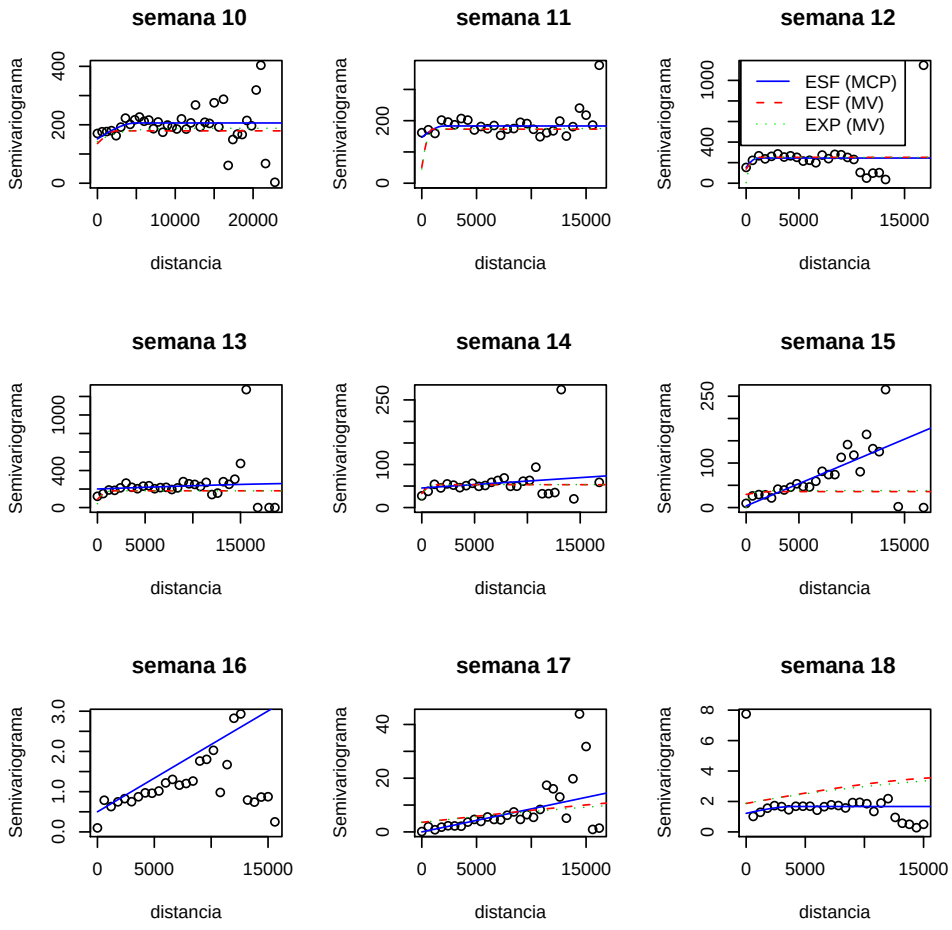


Figura E.2: Gráfica de los semivariogramas direccionales y modelos ajustados por mínimos cuadrados ponderados (MCP) y por máxima verosimilitud (MV)

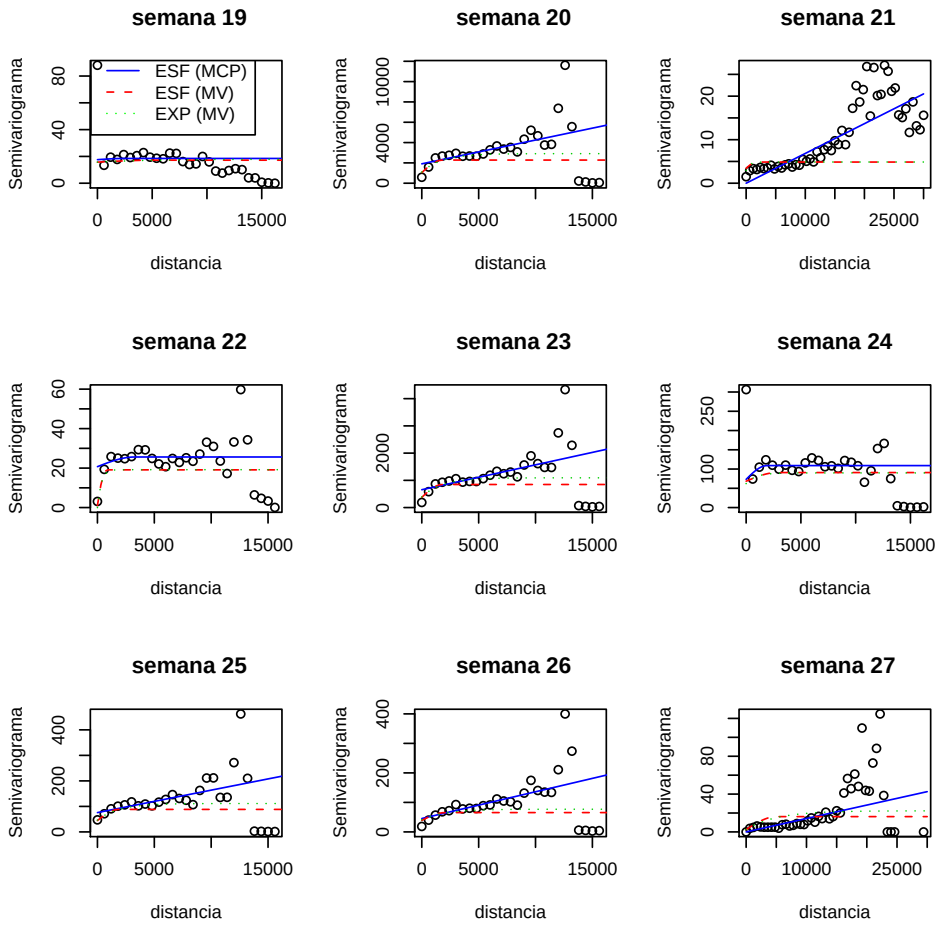


Figura E.3: Gráfica de los variogramas direccionales y modelos ajustados por mínimos cuadrados ponderados (MCP) y por máxima verosimilitud (MV).

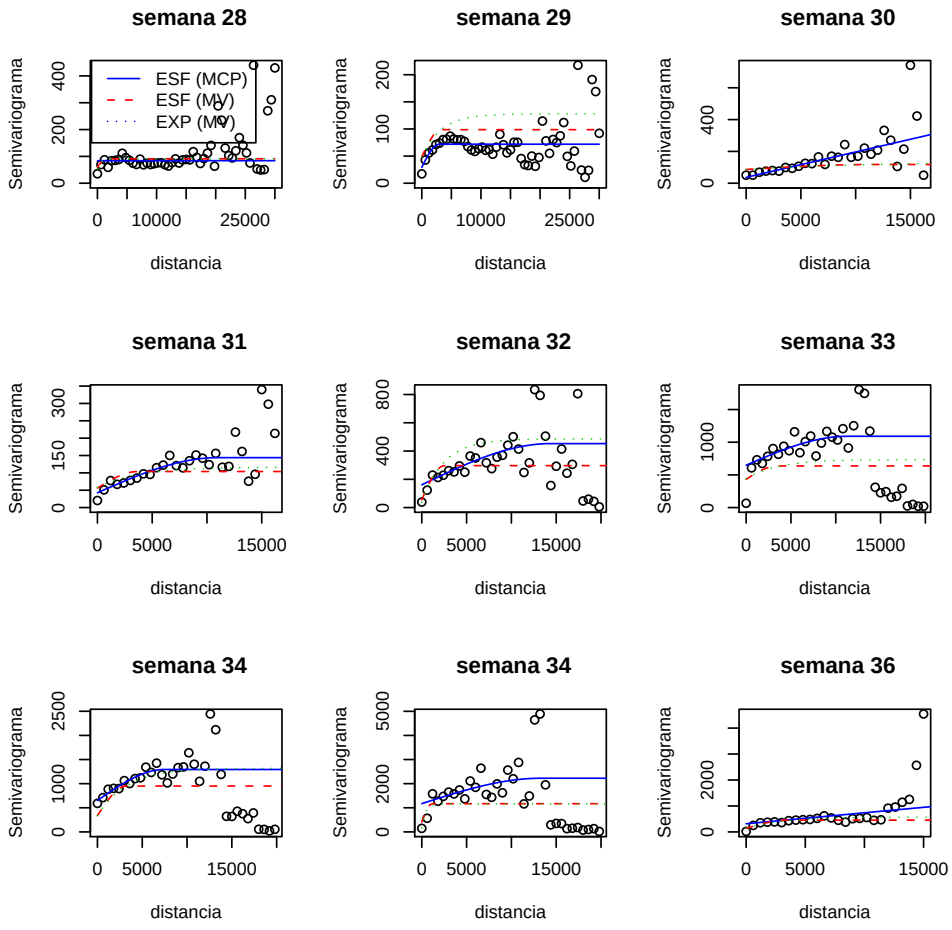


Figura E.4: Gráfica de los semivariogramas direccionales y modelos ajustados por mínimos cuadrados ponderados (MCP) y por máxima verosimilitud (MV)

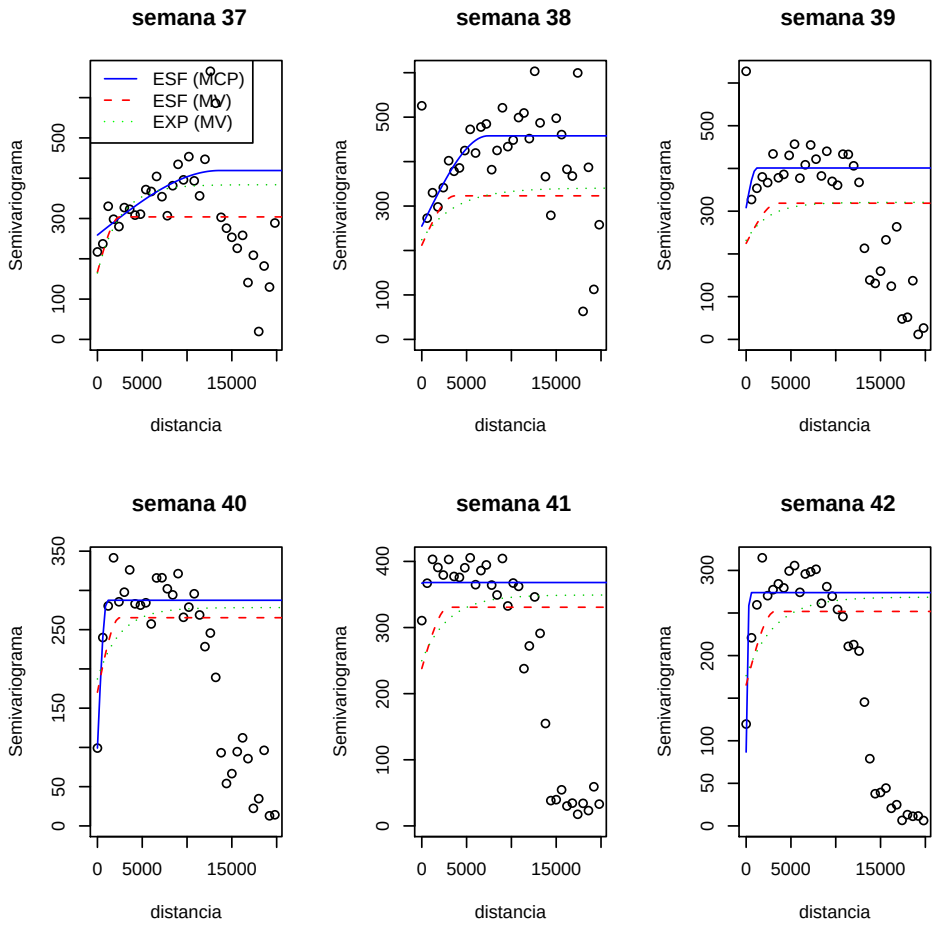


Figura E.5: Gráfica de los semivariogramas direccionales y modelos ajustados por mínimos cuadrados ponderados (MCP) y por máxima verosimilitud (MV)

Apéndice F

DIFERENTES PARÁMETROS OBTENIDOS

Cuadro F.1: Parámetros obtenidos mediante mínimos cuadrados ponderados para ajustar un modelo Esférico a cada conjunto de datos

SEMANA	PEPITA	MESETA	RANGO
1	11.17	12.41	5261.39
2	12.56	3.69	2603.54
3	4.48	2.07	2140.69
4	108.93	108.17	23066.73
5	266.77	0.00	4038.94
6	153.55	103.16	32281.74
7	108.72	0.00	2161.89
8	51.26	90.17	1224.66
9	156.50	0.00	4091.71
10	151.52	54.65	4974.74
11	146.87	36.22	2015.98
12	154.59	90.27	1071.91
13	200.62	64.64	26268.14
14	45.42	1214.87	1133328.68
15	3.73	21504.83	3227843.40
16	0.50	25.07	224813.97
17	0.00	1294.40	2263822.94

SEMANA	PEPITA	MESETA	RANGO
18	1.22	0.45	4054.68
19	17.57	0.91	2445.28
20	1907.92	230689.60	1478042.80
21	0.00	713.66	1565298.17
22	20.71	4.91	3242.92
23	655.28	79809.59	1309541.40
24	73.26	35.70	1881.42
25	75.69	13111.46	2234711.09
26	45.41	8098.23	1339443.62
27	0.00	3011.48	3182959.87
28	84.02	0.00	8485.79
29	29.24	42.66	2919.04
30	36.87	36766.95	3454061.22
31	42.67	101.18	10949.87
32	160.90	291.71	14342.54
33	648.30	442.48	11232.20
34	651.24	642.15	7472.59
35	1173.13	1052.37	13259.02
36	315.52	154595.40	5504295.60
37	258.90	160.19	13627.81
38	254.78	203.17	7400.52
39	308.17	92.91	1190.85
40	99.17	188.29	1079.72
41	368.09	0.00	2720.23
42	86.69	187.26	396.25

Cuadro F.2: Parámetros obtenidos por máxima verosimilitud para ajustar un modelo Esférico a cada conjunto de datos

SEMANA	PEPITA	MESETA	RANGO
1	13.71	4.15	1333.16
2	13.42	1.09	7478.00
3	3.53	2.07	866.00
4	82.38	57.60	1656.94

SEMANA	PEPITA	MESETA	RANGO
5	132.83	105.88	1584.77
6	101.82	79.56	1750.62
7	58.71	36.36	1753.63
8	98.00	83.07	3290.78
9	7.53	130.42	717.00
10	135.70	43.49	2499.89
11	50.85	122.11	862.50
12	137.28	114.61	1202.50
13	88.84	92.34	1236.97
14	32.43	20.70	1068.58
15	29.49	6.40	2498.97
16	13.24	17.36	21177.00
17	3.56	11.08	36206.00
18	1.87	1.77	19030.80
19	15.74	1.46	5152.89
20	1042.39	1234.78	1780.00
21	3.42	1.45	1731.81
22	1.13	17.99	746.67
23	372.71	473.78	1780.00
24	68.45	22.46	2824.30
25	44.44	43.77	1837.70
26	39.11	26.42	2030.99
27	2.47	13.70	4999.90
28	60.28	30.82	1511.47
29	48.79	49.96	3283.99
30	86.44	31.47	11493.11
31	56.11	47.74	3664.48
32	54.88	241.74	2499.00
33	432.31	204.84	3480.50
34	340.85	609.93	3241.90
35	408.07	762.66	1151.00
36	133.63	318.87	3407.70
37	166.87	137.25	2499.90
38	211.82	111.25	3705.00

SEMANA	PEPITA	MESETA	RANGO
39	224.70	94.06	3507.72
40	170.29	94.96	2499.90
41	238.05	92.76	2999.90
42	165.46	86.45	3228.00

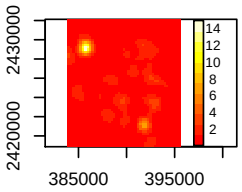
Cuadro F.3: Parámetros obtenidos por máxima verosimilitud para ajustar un modelo Exponencial a cada conjunto de datos

semana	PEPITA	MESETA	RANGO
1	12.34	5.53	417.41
2	13.47	1.51	7477.97
3	3.43	2.19	384.64
4	95.12	54.99	1656.94
5	77.89	159.28	466.58
6	113.65	88.26	1750.62
7	66.16	35.81	1753.62
8	91.57	91.50	1678.93
9	0.00	138.45	300.70
10	142.17	45.99	2499.89
11	45.79	129.01	381.02
12	6.55	244.91	280.62
13	41.66	139.02	381.11
14	19.70	33.55	306.40
15	0.00	38.04	353.09
16	13.27	27.38	21177.00
17	3.56	16.63	36206.00
18	1.88	2.69	19030.80
19	15.74	2.20	5152.88
20	1102.71	1807.35	1780.00
21	2.94	1.93	565.74
22	0.00	19.19	315.69
23	391.25	701.49	1780.00
24	62.23	27.65	875.05
25	46.47	64.84	1837.70

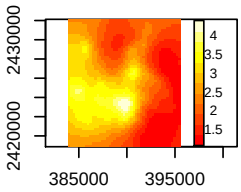
SEMANA	PEPITA	MESETA	RANGO
26	41.45	35.50	2030.99
27	2.55	19.75	4999.90
28	54.23	36.94	563.79
29	48.44	79.46	3283.99
30	86.27	48.71	11493.11
31	58.22	58.37	3664.48
32	38.46	447.74	2499.00
33	442.74	291.67	3480.50
34	334.73	984.30	3241.90
35	165.00	998.78	405.41
36	146.67	432.06	3407.70
37	164.94	218.96	2499.90
38	224.34	116.29	3705.00
39	231.81	88.87	2499.90
40	187.20	90.95	2499.89
41	249.79	99.46	2999.89
42	176.46	92.24	3227.99

Apéndice G

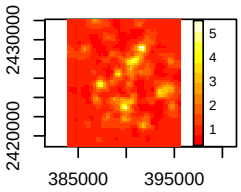
PREDICCIONES REALIZADAS



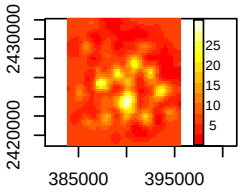
Predicción semana 1



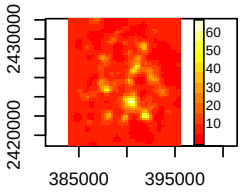
Predicción semana 2



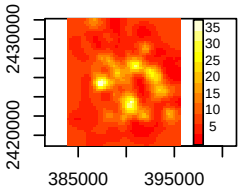
Predicción semana 3



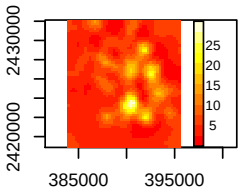
Predicción semana 4



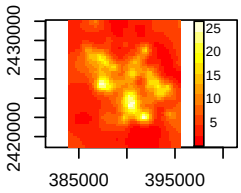
Predicción semana 5



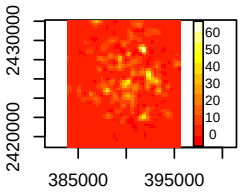
Predicción semana 6



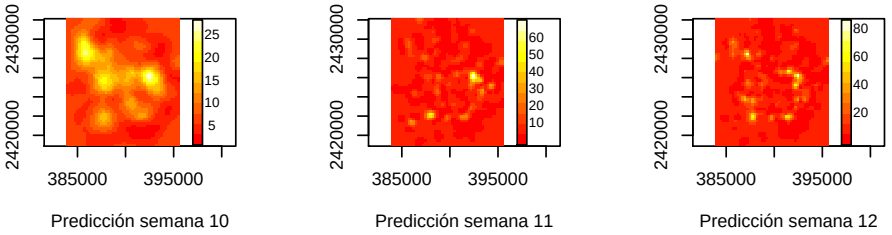
Predicción semana 7



Predicción semana 8



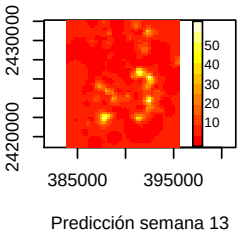
Predicción semana 9



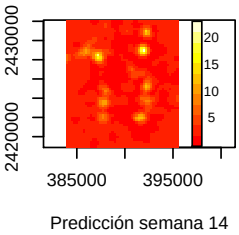
Predicción semana 10

Predicción semana 11

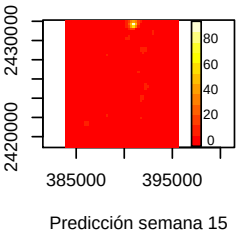
Predicción semana 12



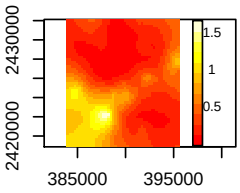
Predicción semana 13



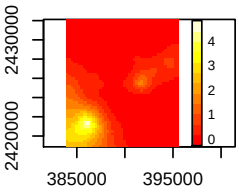
Predicción semana 14



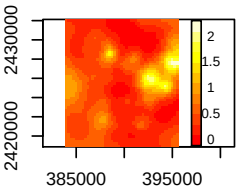
Predicción semana 15



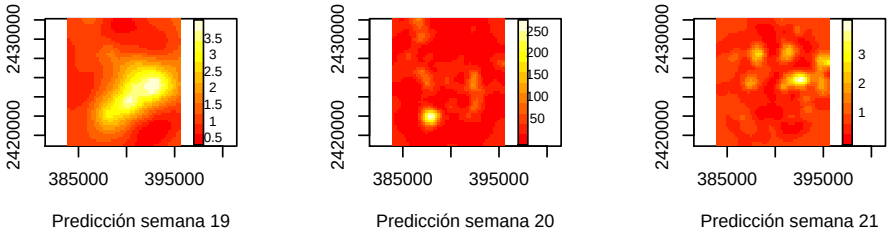
Predicción semana 16



Predicción semana 17



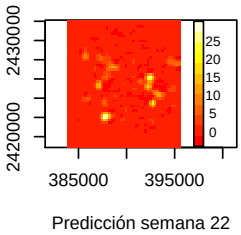
Predicción semana 18



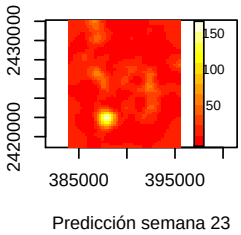
Predicción semana 19

Predicción semana 20

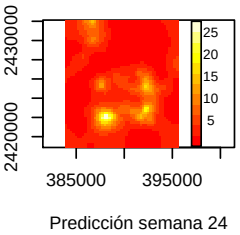
Predicción semana 21



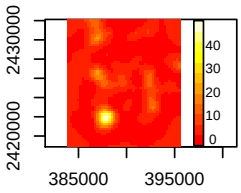
Predicción semana 22



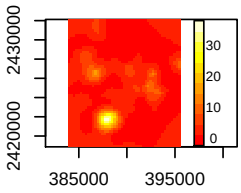
Predicción semana 23



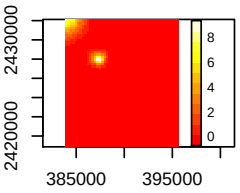
Predicción semana 24



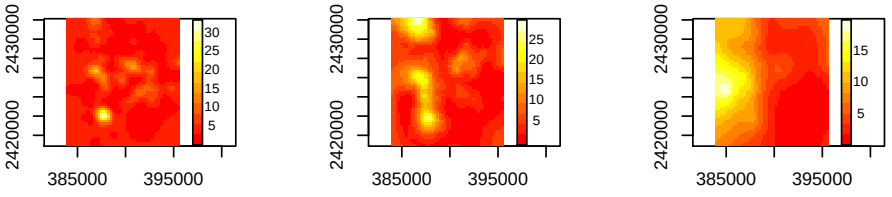
Predicción semana 25



Predicción semana 26



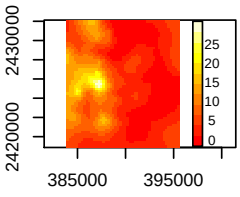
Predicción semana 27



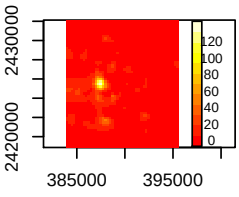
Predicción semana 28

Predicción semana 29

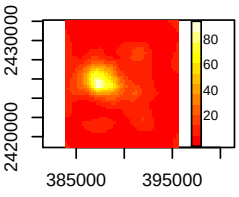
Predicción semana 30



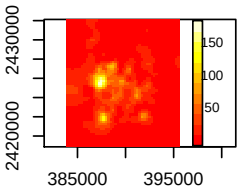
Predicción semana 31



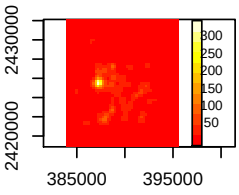
Predicción semana 32



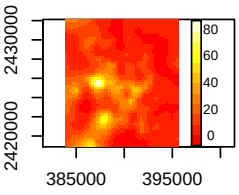
Predicción semana 33



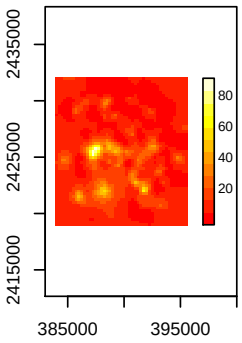
Predicción semana 34



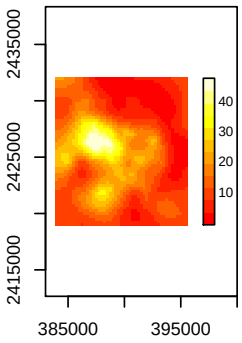
Predicción semana 35



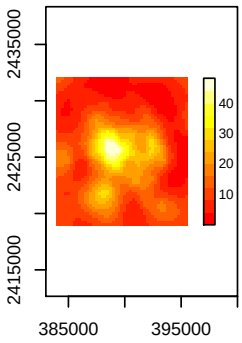
Predicción semana 36



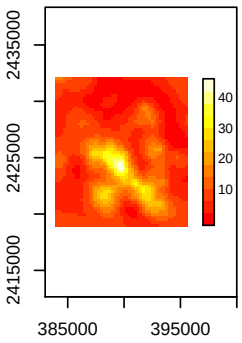
Predicción semana 37



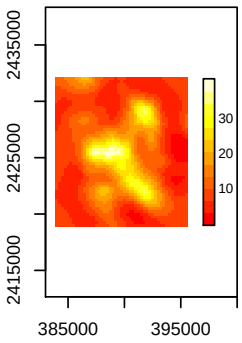
Predicción semana 38



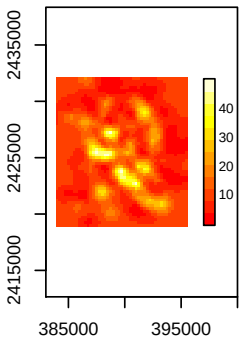
Predicción semana 39



Predicción semana 40



Predicción semana 41



Predicción semana 42

Bibliografía

[1] Box, G. E. P. & Jenkins, G.M. . (1976). *Time Series Analysis Forecasting and Control*. Holden -Day, San Francisco.

[2] Castillo, A., Espinoza, J.C., Valle, J. & Infante, F. (2006) *Dispersion del parasitoide Africano Phymastichus LaSalle (Hymenoptera: Eulophidae) en un nuevo agroecosistema*. Folia Entomologia Mexicana, anño/vol. 45, número 003, Sociedad Mexicana de Entomología, A.C. Xalapa, Méx.

[3] Cañada T., M. R. 2004. *Estudio aplicación de la geoestadística al estudio de la variabilidad espacial del ozono en los veranos de la comunidad de Madrid*.

[4] Chauvet, P. 1994. *Aide-Memoire de Géostatistique Miniere*. École des Mines de Paris, 210p.

[5] Clarck, I. 2001. *Practical Geostatistics*. Geostokos Limited, Alloa Business Centre, Whins Road, Alloa, Central Scotland. Geostokos Limited, Alloa Business Centre, Whins Road, Alloa, Central Scotland.

[6] Díaz F., E. (1993). *Introducción a Conceptos Básicos de Geoestadística*. Memorias Seminario Estadística y Medio Ambiente. Centro de Investigación en Matemáticas, CIMAT. Guanajuato, México.

[7] Gallardo, A. (2006). *Geoestadística*. Departamento de Ecología y Biología Animal. Facultad de Biología, Campus de Lagoas-Marcosende, Universidad de Vigo, 36310 Vigo.

[8] Giraldo, R. 2002. *Introducción a la Geoestadística: Teoría y Aplicación*. Departamento de Estadística, Universidad Nacional de Colombia, 97 p.

[9] http://www.r-project.org

[10] Isaaks, E.H., & Srivastava, R.M. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press. New York, 561 pp

[11] Journel, A.G. & Huijbregts, Ch.J. 1978. *Mining Geostatistics.* Academic, London, 600 pp

[12] Matheron, G. 1965. *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aleatories aux sciences de la nature.* . Masson. París.

[13] Maximiano, C. A. 2007. *Tesis Teoría Geoestadística Aplicada al Análisis de la Variabilidad Espacial Arqueológica Intra-Site*. Universidad Autónoma de Barcelona.

[14] Miranda-Salas, Marcelo and Condal, Alfonso R *Importancia del análisis estadístico exploratorio en el proceso de interpolación espacial: caso de estudio Reserva Forestal Valdivia*. Bosque (Valdivia), ago. 2003.

[15] monografias.com. *Elementos de Geoestadística*.

[16] Moral G., F.J. y Marquez S., J.R. (2002). *Ejemplo de representación gráfica de una variable regionalizada*. Universidad de Extremadura. España.

[17] Moral G., F.J. 2003. *Estudio Representación gráfica de la distribución espacial de una plaga en una plantación mediante el uso de técnicas geoestadísticas*. Universidad de Extremadura, Badajoz.

[18] Petitgas, P. 1996. *Geostatistics and Their Applications to Fisheries Survey Data 5:* 114-142. In: B. A. Megrey and E. Mosknes, (E). Computers and Fisheries Research. Chapman-Hall, Londres.

[19] Samper, C J. & Carrera, R J., 1990. *Geoestadística. Aplicaciones a la hidrología subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Barcelona, 484 pp.

[20] Wackernagel, H. (1995). *Multivariate Geostatistics. An Introduction with Applications*. Springer-Verlag, Berlin.

[21] Webster, R. & Oliver, M.A. 2001. *Geostatistics for Environmental Scientists*. Ed. John Wiley and Sons Ltd. Chichester, 271 pp.

[22] Yaglom, A.M. (1987). *Correlation Theory of Stationary and Related Random Functions: Basic Results*. Springer, New York. Vol. 1.