

# Cómo funciona Kriging

**Fuente:** ArcGIS Resource Center

(<http://help.arcgis.com/es/arcgisdesktop/10.0/help/index.html#/009z00000076000000>)

Kriging es un procedimiento geoestadístico avanzado que genera una superficie estimada a partir de un conjunto de puntos dispersados con valores z. A diferencia de otros métodos de interpolación en el conjunto de herramientas de Interpolación, utilizar la herramienta Kriging en forma efectiva implica una investigación interactiva del comportamiento espacial del fenómeno representado por los valores z antes de seleccionar el mejor método de estimación para generar la superficie de salida.

## ¿Qué es kriging?

Las herramientas de interpolación IDW (Distancia inversa ponderada) y Spline son consideradas métodos de interpolación determinísticos porque están basados directamente en los valores medidos circundantes o en fórmulas matemáticas especificadas que determinan la suavidad de la superficie resultante. Hay una segunda familia de métodos de interpolación que consta de métodos geoestadísticos, como kriging, que está basado en modelos estadísticos que incluyen la autocorrelación, es decir, las relaciones estadísticas entre los puntos medidos. Gracias a esto, las técnicas de estadística geográfica no solo tienen la capacidad de producir una superficie de predicción sino que también proporcionan alguna medida de certeza o precisión de las predicciones.

Kriging presupone que la distancia o la dirección entre los puntos de muestra reflejan una correlación espacial que puede utilizarse para explicar la variación en la superficie. La herramienta Kriging ajusta una función matemática a una cantidad especificada de puntos o a todos los puntos dentro de un radio específico para determinar el valor de salida para cada ubicación. Kriging es un proceso que tiene varios pasos, entre los que se incluyen, el análisis estadístico exploratorio de los datos, el modelado de variogramas, la creación de la superficie y (opcionalmente) la exploración de la superficie de varianza. Este método es más adecuado cuando se sabe que hay una influencia direccional o de la distancia correlacionada espacialmente en los datos. Se utiliza a menudo en la ciencia del suelo y la geología.

## La fórmula de kriging

El método kriging es similar al de IDW en que pondera los valores medidos circundantes para calcular una predicción de una ubicación sin mediciones. La fórmula general para ambos interpoladores se forma como una suma ponderada de los datos:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

- donde:  
 $Z(s_i)$  = el valor medido en la ubicación  $i$   
 $\lambda_i$  = una ponderación desconocida para el valor medido en la ubicación  $i$   
 $s_0$  = la ubicación de la predicción  
 $N$  = la cantidad de valores medidos

En IDW, la ponderación,  $\lambda_i$ , depende exclusivamente de la distancia a la ubicación de la predicción. Sin embargo, con el método kriging, las ponderaciones están basadas no solo en la distancia entre los puntos medidos y la ubicación de la predicción, sino también en la disposición espacial general de los puntos medidos. Para utilizar la disposición espacial en las ponderaciones, la correlación espacial debe estar cuantificada. Por lo tanto, en un kriging ordinario, la ponderación,  $\lambda_i$ , depende de un modelo ajustado a los puntos medidos, la distancia a la ubicación de la predicción y las relaciones espaciales entre los valores medidos alrededor de la ubicación de la predicción. En las siguientes secciones se describe cómo se utiliza la fórmula general de kriging para crear un mapa de la superficie de predicción y un mapa de la precisión de las predicciones.

## Crear un mapa de la superficie de predicción con el método kriging

Para llevar a cabo una predicción con el método de interpolación de kriging, es necesario realizar dos tareas:

- Descubrir las reglas de dependencia.
- Realizar las predicciones.

A fin de completar estas dos tareas, kriging atraviesa un proceso de dos pasos:

1. Crea los variogramas y las funciones de covarianza para calcular los valores de dependencia estadística (denominada autocorrelación espacial) que dependen del modelo de autocorrelación (ajustar un modelo).
2. Prevé los valores desconocidos (hacer una predicción).

Se dice que en este método los datos se utilizan dos veces, debido a estas dos tareas bien distintivas: la primera vez, para calcular la autocorrelación espacial de los datos, y la segunda, para hacer las predicciones.

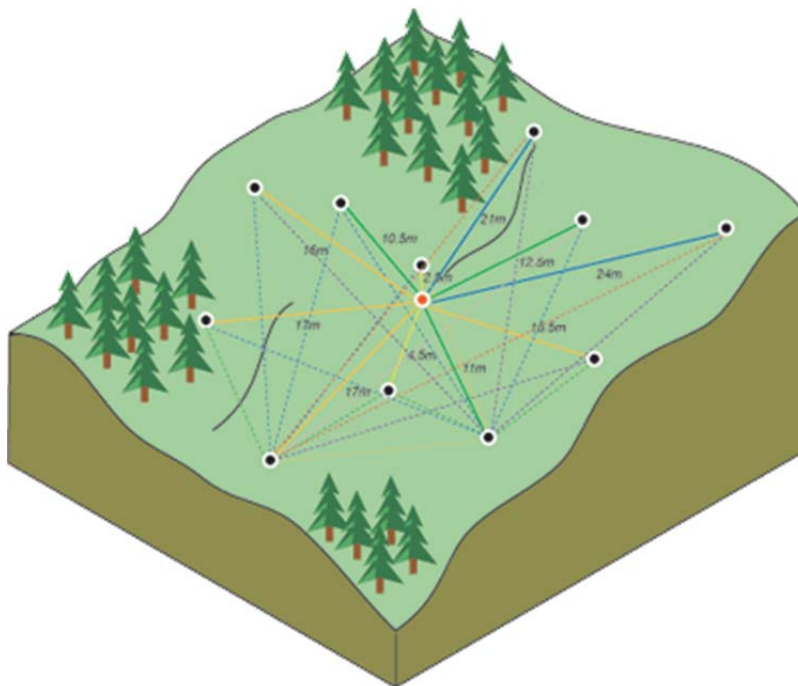
## Variografía

El ajuste de un modelo, o modelado espacial, también se conoce como análisis estructural o variografía. En el modelado espacial de la estructura de los puntos medidos, se comienza con un gráfico del semivariograma empírico, calculado con la siguiente ecuación para todos los pares de ubicaciones separados por la distancia  $h$ :

$$\text{Semivariogram}(\text{distance}_h) = 0.5 * \text{average}((\text{value}_i - \text{value}_j)^2)$$

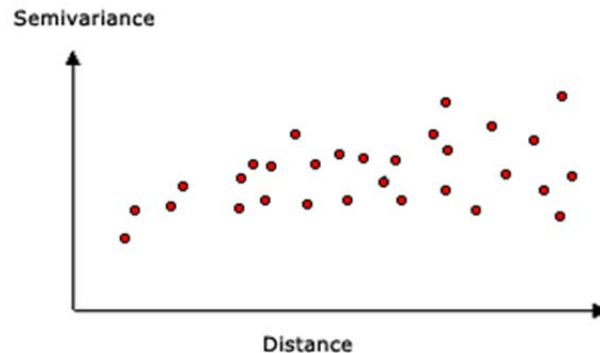
La fórmula implica calcular la diferencia cuadrada entre los valores de las ubicaciones asociadas.

En la imagen a continuación se muestra la asociación de un punto (en color rojo) con todas las demás ubicaciones medidas. Este proceso continúa con cada punto medido.



Cálculo de la diferencia cuadrada entre las ubicaciones asociadas

A menudo, cada par de ubicaciones tiene una distancia única y suele haber varios pares de puntos. La diagramación de todos los pares rápidamente se vuelve imposible de administrar. En lugar de diagramar cada par, los pares se agrupan en bins de intervalo. Por ejemplo, calcule la semivarianza promedio de todos los pares de puntos que están a más de 40 metros de distancia pero a menos de 50 metros. El semivariograma empírico es un gráfico de los valores de semivariograma promediados en el eje Y, y la distancia (o intervalo) en el eje X (consulte el diagrama a continuación).



Ejemplo de gráfico de semivariograma empírico

La autocorrelación espacial cuantifica un principio básico de geografía: es más probable que las cosas que están más cerca sean más parecidas que las que están más alejadas. Entonces, los pares de ubicaciones que están más cerca (extremo izquierdo del eje X de la nube de semivariograma) deberían tener valores más similares (parte inferior en el eje Y de la nube de semivariograma). A medida que los pares de ubicaciones estén más separados entre sí (hacia la derecha en el eje X de la nube de semivariograma), deberían ser más distintos y tener una diferencia cuadrada más grande (hacia arriba en el eje Y de la nube de semivariograma).

## Ajustar un modelo al semivariograma empírico

El siguiente paso es ajustar un modelo a los puntos que forman el semivariograma empírico. El moldeado del semivariograma es un paso clave entre la descripción espacial y la predicción espacial. La aplicación principal de kriging es la predicción de los valores de atributo en las ubicaciones que no fueron muestreadas. El semivariograma empírico proporciona información sobre la autocorrelación espacial de los datasets. Sin embargo, no suministra información para todas las direcciones y distancias posibles. Por esta razón, y para asegurar que las predicciones de kriging tengan varianzas de kriging positivas, es necesario ajustar un modelo (es decir, una función o curva continua) al semivariograma empírico. En resumen, esto es similar al análisis de regresión, en el que se ajusta una línea o curva continua a los puntos de datos.

Para ajustar un modelo al semivariograma empírico, seleccione una función que sirva como modelo, por ejemplo, un tipo esférico que se eleve y nivele las distancias más grandes que sobrepasan un determinado rango (vea el ejemplo del modelo esférico más abajo). Existen desviaciones de los puntos en el semivariograma empírico con respecto al modelo; algunos están por encima de la curva del modelo y algunos están por debajo. Sin embargo, si suma la distancia de cada punto por encima de la línea y la distancia de cada punto por debajo, los dos valores deberían ser similares. Existen varios modelos de semivariograma para elegir.

## Modelos de semivariograma

La herramienta Kriging proporciona las siguientes funciones para elegir el modelado del semivariograma empírico:

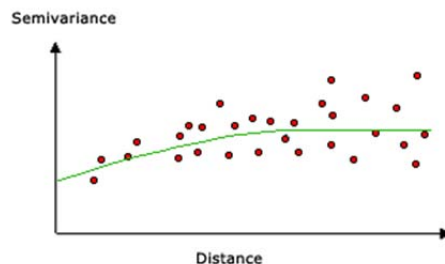
- Circular
- Esférica
- Exponencial
- Gaussiana
- Lineal

El modelo seleccionado influye en la predicción de los valores desconocidos, en particular cuando la forma de la curva cercana al origen difiere significativamente. Cuanto más pronunciada sea la curva cercana al origen, más influirán los vecinos más cercanos en la predicción. Como resultado, la superficie de salida será menos suave. Cada modelo está diseñado para ajustarse a diferentes tipos de fenómenos de forma más precisa.

En los siguientes diagramas se muestran dos modelos comunes y se identifican las diferencias de las funciones:

### Un ejemplo del modelo esférico

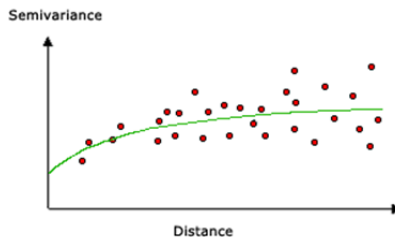
En este modelo se muestra una disminución progresiva de la autocorrelación espacial (así como un aumento en la semivarianza) hasta cierta distancia, después de la cual la autocorrelación es cero. El modelo esférico es uno de los que más se utilizan.



Ejemplo de modelo esférico

## Un ejemplo del modelo exponencial

Este modelo se aplica cuando la autocorrelación espacial disminuye exponencialmente cuando aumenta la distancia. En este caso, la autocorrelación desaparece por completo solo a una distancia infinita. El modelo exponencial también es un modelo comúnmente utilizado. La elección de qué modelo se va a utilizar está basada en la autocorrelación espacial de los datos y en el conocimiento previo del fenómeno.



Ejemplo de modelo exponencial

## Comprender un semivariograma: rango, meseta y nugget

Como se indicó previamente, el semivariograma muestra la autocorrelación espacial de los puntos de muestra medidos. Tal como lo expresa un principio básico de la geografía (las cosas más cercanas son más parecidas), los puntos medidos que están cerca por lo general tendrán una diferencia cuadrada menor que la de aquellos que están más distanciados. Una vez diagramados todos los pares de ubicaciones después de haber sido colocados en un bin, se ajusta un modelo para estas ubicaciones. El rango, la meseta y el nugget se utilizan, generalmente, para describir estos modelos.

### Rango y meseta

Al observar el modelo de un semivariograma, notará que a una determinada distancia, el modelo se nivela. La distancia a la que el modelo comienza a aplanarse se denomina rango. Las ubicaciones de muestra separadas por distancias más cortas que el rango están autocorrelacionadas espacialmente, mientras que las ubicaciones que están más alejadas que el rango, no lo están.

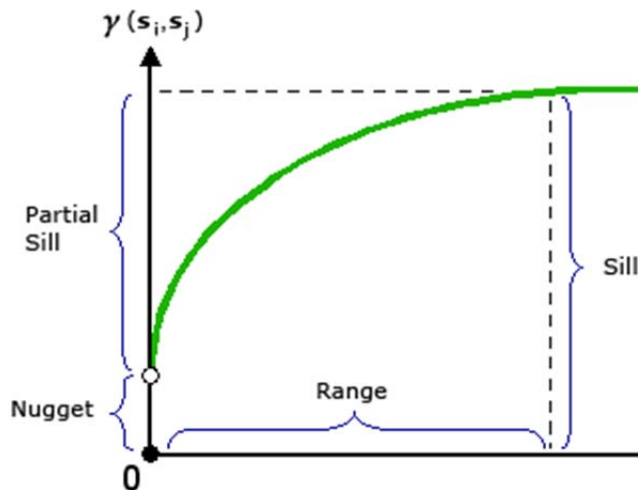


Ilustración de componentes de rango, meseta y nugget

El valor en el cual el modelo de semivariograma alcanza el rango (el valor en el eje Y) se denomina meseta. Una meseta parcial es la meseta menos el nugget. El nugget se describe en la siguiente sección.

## Nugget

En teoría, a una distancia de separación cero (por ej. intervalo = 0), el valor del semivariograma es 0. No obstante, a una distancia de separación infinitamente inferior, el semivariograma a menudo muestra un efecto nugget, que es un valor mayor que 0. Si el modelo de semivariograma intercepta el eje Y en 2, entonces el nugget es 2.

El efecto nugget puede atribuirse a errores de medición o a fuentes espaciales de variación a distancias que son menores que el intervalo de muestreo (o a ambas cosas). Los errores de medición ocurren debido al error inherente a los dispositivos de medición. Los fenómenos naturales pueden variar espacialmente en un rango de escalas. La variación a microescalas más pequeñas que las distancias de muestreo aparecerán como parte del efecto nugget. Antes de recopilar datos, es importante lograr comprender las escalas de variación espacial en las que está interesado.

## Realizar una predicción

Cuando haya descubierto la dependencia o autocorrelación en sus datos (consulte la sección Variografía más arriba) y haya finalizado con el primer uso de los datos (usar la información espacial de los datos para calcular las distancias y modelar la autocorrelación espacial) puede realizar una predicción utilizando el modelo ajustado. Después de esto, se aparta el semivariograma empírico.

Ahora puede utilizar los datos para realizar predicciones. Al igual que la interpolación de IDW, kriging forma ponderaciones a partir de los valores medidos circundantes para prever ubicaciones sin mediciones. Asimismo, los valores medidos que estén más cerca de las ubicaciones sin mediciones tienen la mayor influencia. Sin embargo, las ponderaciones de kriging para los puntos medidos circundantes son más sofisticadas que las del método IDW. Este último utiliza un algoritmo simple basado en la distancia, mientras que las ponderaciones de kriging provienen de un semivariograma que se desarrolló observando la naturaleza espacial de los datos. Para crear una superficie continua del fenómeno, se realizan predicciones para cada ubicación, o centro de celda, en el área de estudio basadas en el semivariograma y la disposición espacial de los valores medidos que son cercanos.

## Los métodos kriging

Existen dos métodos kriging: ordinario y universal.

El kriging ordinario es el más general y más utilizado de los métodos kriging y es el predeterminado. Presupone que el valor medio constante es desconocido. Esa es una presuposición razonable a menos que haya una razón científica para rechazarla.

El kriging universal presupone que hay una tendencia de invalidación en los datos, por ejemplo, un viento prevaleciente, y puede modelarse a través de la función determinística polinómica. Esta función polinómica se resta de los puntos medidos originalmente y la autocorrelación se modela a partir de los errores aleatorios. Una vez que el modelo se ajusta a los errores aleatorios y antes de realizar una predicción, se vuelve a sumar la función polinómica a las predicciones para obtener resultados significativos. El kriging universal solo se debe utilizar si se conoce una tendencia en los datos y se puede dar una justificación científica para describirla.



## Gráficos de semivariograma

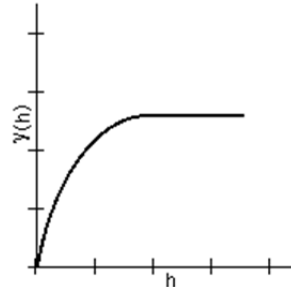
Kriging es un procedimiento complejo que requiere de un mayor conocimiento sobre las estadísticas espaciales que lo que se puede abarcar en este tema. Antes de utilizar este método, debe comprender profundamente sus principios básicos y evaluar la adecuación de sus datos para realizar un modelo con esta técnica. Si no comprende bien este procedimiento, se recomienda que repase algunas de las referencias enumeradas al pie de este tema.

Kriging se basa en una teoría de variables regionalizadas que presupone que la variación espacial en el fenómeno representado por los valores  $z$  es estadísticamente homogénea a lo largo de toda la superficie (por ej., se puede observar el mismo patrón de variación en todas las ubicaciones sobre la superficie). Esta hipótesis de homogeneidad espacial es fundamental para la teoría de variables regionalizadas.

## Modelos matemáticos

A continuación se presentan las formas generales y las ecuaciones de los modelos matemáticos utilizados para describir la semivarianza.

### SPHERICAL



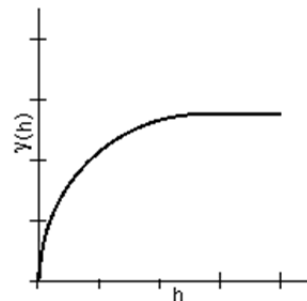
$$\gamma(h) = c_0 + c \left( \frac{3h}{2\alpha} - \frac{1}{2} \left( \frac{h}{\alpha} \right)^3 \right) \quad 0 < h \leq \alpha$$

$$\gamma(h) = c_0 + c \quad h > \alpha$$

$$\gamma(0) = 0$$

Ilustración de modelo de semivarianza esférica

### CIRCULAR



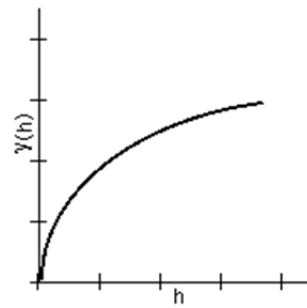
$$\gamma(h) = c_0 + c \left( 1 - \frac{2}{\pi} \cos^{-1} \left( \frac{h}{\alpha} \right) + \sqrt{1 - \frac{h^2}{\alpha^2}} \right) \quad 0 < h \leq \alpha$$

$$\gamma(h) = c_0 + c \quad h > \alpha$$

$$\gamma(0) = 0$$

Ilustración de modelo de semivarianza circular

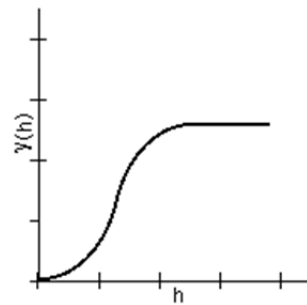
### EXPONENTIAL



$$\gamma(h) = c_0 + c \left( 1 - \exp\left(-\frac{h}{\tau}\right) \right) \quad h > 0$$
$$\gamma(0) = 0$$

Ilustración de modelo de semivarianza exponencial

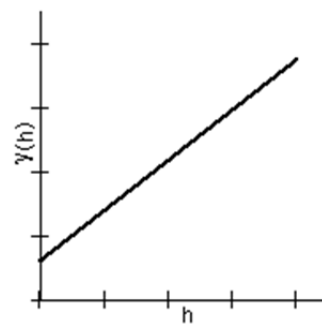
### GAUSSIAN



$$\gamma(h) = c_0 + c \left( 1 - \exp\left(-\frac{h^2}{\tau^2}\right) \right) \quad h > 0$$
$$\gamma(0) = 0$$

Ilustración de modelo de semivarianza gaussiana

### LINEAR



$$\gamma(h) = c_0 + c \left( \frac{h}{\alpha} \right) \quad 0 < h \leq \alpha$$

$$\gamma(h) = c_0 + c \quad h > \alpha$$

$$\gamma(0) = 0$$

Ilustración de modelo de semivarianza lineal

## Referencias

Burrough, P. A. Principles of Geographical Information Systems for Land Resources Assessment. New York: Oxford University Press. 1986.

Heine, G. W. "A Controlled Study of Some Two-Dimensional Interpolation Methods". COGS Computer Contributions 3 (n.º 2): 60–72. 1986.

McBratney, A. B. y R. Webster. "Choosing Functions for Semi-variograms of Soil Properties and Fitting Them to Sampling Estimates". Journal of Soil Science 37: 617–639. 1986.

Oliver, M. A. "Kriging: A Method of Interpolation for Geographical Information Systems". International Journal of Geographic Information Systems 4: 313–332. 1990

Press, W. H., S. A. Teukolsky, W. T. Vetterling y B. P. Flannery. Numerical Recipes in C: The Art of Scientific Computing. New York: Cambridge University Press. . 1988.

Royle, A. G., F. L. Clausen y P. Frederiksen. "Practical Universal Kriging and Automatic Contouring". Geoprocessing 1: 377–394. 1981.