

Importancia, usos y limitaciones de S+Spatial Stats en S-plus

Sergio Daniel Carrero Arevalo

April 18, 2022

1 Introducción

S-PLUS, es una implementación comercial que se hizo en el lenguaje de programación ES donde se disponen capacidades de programación orientada a objetos y algoritmos analíticos avanzados. De hecho diferentes investigadores remarcen al programa como de los más importantes para análisis de datos, modelación estadística y extracción de datos o “data mining”. Ahora bien, SPATIALSTATS es un módulo para datos, análisis y la generación de gráficos involucrando su comportamiento espacial.

2 Como funciona S+SPATIAL STATS de S-plus

Para entender lo anterior primero debemos saber que los datos espaciales consisten en mediciones u observaciones tomadas en ubicaciones específicas o dentro de regiones específicas, siendo las ubicaciones puntos o áreas referenciadas y los datos las observaciones específicas como a que se usa la construcción como vivienda o como comercial en una determinada área o región, teniendo un patrón regular o irregular, además los datos pueden ser continuos si por ejemplo medimos un nematodo ubicado en un punto particular, siendo este una muestra central o un dato discreto si evaluamos las ubicaciones de un ave a través de sus nidos “obtenemos un dato de conteo”. La variabilidad espacial se debe modelar en función de la distancia que hay entre los distintos sitios de muestreo donde los datos más similares son los que deben estar más cerca.

Es importante recordar que los datos Lattice pueden ser de manera regular (uniforme o coherentes dentro de una cuadrícula) o irregular (casos de covid dentro de una ciudad), como se vio en Bogotá al inicio de la pandemia donde en marzo la zona de la mayoría de los casos era en la región de usaquén, chapinero alto y los Rosales o como meses después se evidenciaba ya en la región sur en la zona de Kenedy, Bosa y Soacha. Para poder entender esto es importante identificar los patrones de puntos cuando justamente son las ubicaciones la variable que nos interesa donde debemos entender si hay aleatoriedad, agrupamiento o regularidad. Después de haber entendido un dato se deben hacer modelados y predicciones para ello el analista debe evaluar si existe variación a gran escala y variación a pequeña escala para evidenciar posibles errores en la medición, o si hay patrones de dependencia. Cuando se quiere confirmar dependencia espacial los modelos de dependencia dependen del área a evaluar y pueden verse afectados por valores atípicos en una región en específico, además de tener en cuenta las condiciones de contorno (efectos de borde). Así mismo se debe asumir una estacionariedad para poder hacer las inferencias ya que son invarianza de ubicación de cada uno de los datos implicando que las relaciones dentro de cada subregión o subconjunto de puntos siguen siendo las mismas sin importar donde residen los puntos en el espacio.

2.1 De estas se pueden encontrar 3 tipos de estacionariedad

1. Estacionariedad estricta: Equivalencia por traslación y rotación
2. Estacionariedad débil: Media constante y covarianza independiente de ubicación
3. Estacionariedad de incrementos: La varianza del incremento es decir de la diferencia de primer orden entre 2 puntos debe ser independiente de la ubicación.

2.2 Conclusión

En cuanto a conceptos adicionales debemos tener en cuenta la isotropía (hace referencia a que el proceso espacial ha evolucionado homogéneamente independientemente de la dirección) y el anisotropico (a la correlación o covarianza resultante difiere de la dirección) y finalmente la escala (nos dan a entender que procesos o patrones se han dado a mayor área son el resultado de procesos particulares en diferentes escalas). Para esto Spatial STATS tiene una elevada cantidad de funciones para entender el dato como variogramas estándar, robustos o empiricos, Kriging ordinario o universal si el dato es geoestadístico o si es un dato lattice encuentra vecinos o grupos de vecinos, correlaciones como Moran y Geary, regresión espacial. Y finalmente, para patrones de puntos puede trazar mapas de puntos sin distorsión espacial, además de estimar la intensidad y estacionalidad por diferentes métodos, sin embargo esta diferente flexibilidad se gana por un sacrificio en la capacidad de memoria de la computadora, además de que hay ciertos tipos de datos en los que se queda corto como al no tener ningún paquete de procesamiento de imágenes.