



Taller Evaluado

Estadística Espacial

Problema 1 (Geoestadístico) Considere la base de datos **geoR::camg**; *contenido de magnesio medido en muestras de suelo tomadas de la capa 0-20 cm en 178 ubicaciones dentro de una determinada zona de estudio dividida en tres subzonas*. Con el objetivo de crear un mapa de la zona de interés por la variable “mg020” (contenido de magnesio), responda las siguientes preguntas:

1. Dar un formato apropiado a la base de datos para trabajar en **R** a través de la librería *geoR*.
2. Realice un análisis exploratorio a través de gráficas apropiadas y responda las siguientes preguntas:
 - (a) ¿Existe una posible presencia de valores atípicos?
 - (b) ¿Se debe remover alguna tendencia antes de analizar la dependencia espacial?
 - (c) ¿Se justifica alguna transformación (como Box-Cox) en los datos?
3. Haga un análisis detallado de dependencia espacial usando el variograma como herramienta y responda las siguientes preguntas:
 - (a) ¿Se justifica el uso de un modelo geoestadístico?
 - (b) Proponga tres modelos de variogramas identificando valores iniciales para el nugget, el sill y el rango. Además, indique claramente el tipo de media asumida.
4. Ajuste los modelos de variograma teóricos propuestos en el ítem anterior, mediante máxima verosimilitud restringida. Para cada modelo de covarianza estime el modelo con media constante y media lineal. Utilice el criterio de Akaike (AIC) para seleccionar el modelo más apropiada. Reporte los AIC en una tabla donde también incluya el AIC del modelo no espacial.
5. Realice un kriging apropiado (con la función de media y función de covarianza escogida en los puntos anteriores) para este conjunto de datos e interprételo. Grafique el kriging y la varianza del kriging estimado.

Problema 2 (Datos de Área) Considere el conjunto de datos sobre suicidios en 32 municipios de Londres (excluyendo City London) en el periodo de 1089-1993 para hombres y mujeres combinados. Las variables que se registran son:

- Número de suicidios observados en el periodo estudiado (y),
- Número de casos de suicidio esperados (E),
- índice de privación social (x_1),
- índice de fragmentación social (x_2), que refleja la falta de conexiones sociales y de sentido de comunidad.

Además, de las variables del sector geográfico, con el objetivo de explicar y , a través de $SMR = y/E$.

1. Cargue los archivos “LDNSuicides.shp” (polígono espacial) y “LondonSuicides.RData” (datos). Luego unir en el polígono espacial el data.frame que contiene las variables.
2. Aplique los test de Moran y Geary para la variable SMR . Utilice la matriz \mathbf{W} estandarizada por fila. ¿Se justifica un modelo espacial?
3. Considere los siguientes modelos para $SMR = y/E$:

$$(M1): \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \eta, \quad \eta \sim \mathcal{N}(0, \sigma^2);$$

$$(M2): \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \eta, \quad \eta \sim SAR(\rho);$$

$$(M3): \quad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \eta, \quad \eta \sim CAR(\rho)$$

Ajuste cada uno de estos modelos utilizando las funciones apropiadas de la librería “spatialreg”. En el caso del modelo SAR usar el errorsarlm.

- (a) En base al AIC ¿Cuál de los 3 modelos escogería usted?.
- (b) Para el modelo seleccionado, según los p-valores ¿son todas las covariables significativas?
- (a) En caso de ser necesario, reduzca (en el sentido de eliminar alguna covariable) el modelo anterior.
4. Para el modelo final, graficar el SMR estimado y el real. Para ello guarde en el polígono espacial el ajuste y el residuo del modelo final.
5. Aplique los test de Moran y geary a los residuos del modelo final. ¿Existe dependencia espacial? Grafique los residuos de este modelo y los residuos del modelo lineal con las mismas covariables.