

AEDD: Trabajo Estadística Espacial

Grupo XX

Curso 2023/2024

Este trabajo debe entregarse en formato pdf, incluyendo el código R utilizado, las correspondientes salidas y los comentarios (o interpretaciones de los resultados) pertinentes (para ello se recomienda emplear RMarkdown, a partir de un fichero *.Rmd* o un fichero *.R* mediante spin, que también debe entregarse).

Se empleará el conjunto de datos `temp_andalu` almacenado en el archivo `temp_andalu_X.RData`, donde *X* es el número de grupo, que contiene temperaturas medias (`temp`) mensuales (correspondientes al mes `attr(temp_andalu, "fecha")`, seleccionado al azar de los últimos dos años) en las estaciones de la Red de Información Agroclimática de Andalucía. Las posiciones de las estaciones (`lon`, `lat`) están en coordenadas proyectadas UTM en kilómetros. El objetivo de este trabajo es realizar un análisis geoestadístico de estos datos, utilizando las herramientas explicadas durante las clases.

```
load("temp_andalu_1.RData") # Renombrar según grupo
str(temp_andalu)
```

```
## 'data.frame':   101 obs. of  3 variables:
## $ lon : num  373 501 405 553 191 ...
## $ lat : num  4207 4067 4159 4077 4117 ...
## $ temp: num  20 21.4 21.3 22.5 20.9 ...
## - attr(*, "fecha")= chr "10/2022"
```

La fecha límite de entrega de este trabajo es el **29 de diciembre de 2023** (para poder disponer de la corrección antes del periodo de exámenes). En la evaluación, se tendrá en cuenta la correcta resolución de los ejercicios, la claridad y corrección en la redacción del trabajo, la interpretación y discusión de los resultados y el correcto funcionamiento de los scripts entregados. En caso de que se considere oportuno, se podrá solicitar a los alumnos la presentación oral del trabajo.

Ejercicios

1. Lectura de datos y análisis exploratorio
 - a. Realizar un breve análisis descriptivo del conjunto de datos y generar un histograma de la respuesta.
 - b. Convertir el `data.frame` a un objeto `sf`, definiendo el CRS mediante la cadena de texto `"+proj=utm +zone=30 +ellps=WGS84 +units=km"`, y representar la distribución espacial de la respuesta. ¿Se observa algún patrón (más o menos claro)?
 - c. Completar el análisis descriptivo de la variabilidad de gran escala. ¿Aparentemente hay tendencia espacial? En caso afirmativo, sugerir un modelo lineal.
 - d. Volver a representar la distribución espacial de la respuesta (o solo las posiciones de observación si surgen dificultades) junto con los límites administrativos de Andalucía (que se pueden obtener empleando el paquete `mapSpain`).
2. Modelado de la dependencia espacial

- a. Analiza la variabilidad de pequeña escala empleando el estimador clásico del semivariograma, considerando 20 saltos hasta un salto máximo de 290.
 - b. Ajustar a las estimaciones piloto un modelo (isotrópico) de semivariograma exponencial y uno esférico, mediante WLS. ¿Con cuál se obtiene un mejor ajuste? Interpretar las estimaciones obtenidas de los parámetros del semivariograma.
 - c. Emplear medidas de validación cruzada, considerando 10 grupos, para seleccionar el modelo final (establecer la semilla igual al número de grupo multiplicado por 10; no es necesario generar gráficos).
3. Predicción espacial
- a. Crear una rejilla de predicción de dimensiones 150x75 que cubra los límites de Andalucía e intersecarla con dicha región (o una rejilla que cubra las posiciones de observación si surgen dificultades).
 - b. Empleando el modelo obtenido en el ejercicio anterior, calcular las predicciones y varianzas kriging en la rejilla de predicción, y representarlas.