



Universidad Nacional de Colombia

FACULTAD DE CIENCIAS

TALLER 2 - TEST DE MANTEL

*Estadística Espacial*

Integrantes:

John Anderson Guarín López  
German Camilo Vasquez Herrera

31 de marzo de 2024

# **Análisis de la aleatoriedad espacial de los datos de temperatura máxima diaria de Bogotá en el año 2023**

## **Resumen**

El objetivo de este estudio es determinar si la medición de la temperatura máxima diaria en Bogotá, durante el año 2023, presenta aleatoriedad espacial (es decir, que la ubicación de las estaciones de medición no influye en la medición de la temperatura máxima diaria en Bogotá) o si la ubicación espacial de cada estación radica en las mediciones realizadas en el año 2023. Para ello, se tiene en cuenta la información reportada por el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) y, en base a los datos obtenidos, poder utilizar el test de Mantel, como herramienta estadística, para determinar si se presenta aleatoriedad espacial o correlación espacial, entre las mediciones del año 2023.

## **Palabras Clave**

Temperatura, Estaciones de Medición, Aleatoriedad Espacial, Correlación Espacial, Test de Mantel

## **Abstract**

The objective of this study is to determine whether the measurement of the daily maximum temperature in Bogotá, during 2023, presents spatial randomness (i.e., that the location of the measurement stations doesn't influence the measurement of the daily maximum temperature in Bogotá) or if the spatial location of each station affects the measurements taken in 2023. To do this, the information reported by IDEAM (Institute of Hydrology, Meteorology, and Environmental Studies) is taken into account, and based on the data obtained, the Mantel test is used as a statistical tool to determine whether spatial randomness or spatial correlation is present among the measurements of the 2023.

## **Keywords**

Temperature, Measurement Stations, Spatial Randomness, Spatial Correlation, Mantel Test.

## **Introducción**

El análisis de la distribución espacial de variables climáticas, como la temperatura, es fundamental para comprender los patrones climáticos regionales y su variabilidad. Esto permite saber si en un entorno específico, la distribución de dichas variables está basada de acuerdo al área donde se realiza la medición, como por ejemplo, las localidades en Bogotá en donde se encuentran cada una de las estaciones para la

medición de dichas variables. En este informe, se presenta un análisis de la aleatoriedad espacial de los datos de temperatura recopilados en Bogotá D.C., con el fin de determinar si la medición de la temperatura máxima de la ciudad radica en la ubicación de las estaciones de medición o, en caso contrario, que dicha medición no depende de su ubicación espacial.

Se utilizarán herramientas estadísticas y de análisis espacial para explorar patrones climáticos en cada una de las estaciones de medición seleccionadas y evaluar la posible influencia de factores geográficos y ambientales en la temperatura máxima diaria. Este estudio tiene como objetivo evaluar si la medición de la temperatura máxima diaria de Bogotá depende de la ubicación espacial de cada una de las estaciones de medición o si se presenta aleatoriedad espacial; es decir, la medición de la temperatura máxima no depende de su ubicación.

Para llevar a cabo este estudio, el texto se divide de la siguiente manera: primero, se menciona un marco teórico, donde se describen los aspectos más relevantes por los cuales se deben realizar estudios referentes a la temperatura máxima diaria. Segundo, se propone la metodología con la que se llevará a cabo el análisis de aleatoriedad, mencionando las herramientas a utilizar para llevar a cabo el análisis correspondiente. Tercero, se realizará el análisis respectivo, dado una breve descripción de las herramientas estadísticas utilizadas y la base de datos respectiva. Cuarto, se mencionarán los resultados correspondientes de acuerdo al análisis realizado. Finalmente, se indicarán las conclusiones respectivas de este trabajo, haciendo énfasis en los resultados obtenidos y dejando una reflexión para estudios futuros sobre la aleatoriedad espacial en el estudio de la temperatura máxima diaria.

## Marco Teórico:

La temperatura máxima diaria es un parámetro climático fundamental que describe la temperatura más alta alcanzada en un día específico en una ubicación geográfica determinada. Este valor es de suma importancia en una variedad de contextos, desde la planificación urbana y la agricultura hasta la salud pública y la gestión de desastres naturales. En este marco teórico, exploraremos algunos aspectos relevantes de la temperatura máxima diaria y su importancia en diferentes áreas de estudio.

1. **Influencia del Clima en la Temperatura Máxima Diaria:** La temperatura máxima diaria está estrechamente relacionada con el clima local y regional. Factores como la radiación solar, la humedad atmosférica, la altitud y la cobertura del suelo afectan la cantidad de calor absorbido y retenido por la superficie terrestre, lo que a su vez influye en la temperatura máxima diaria.
2. **Variabilidad Temporal y Espacial:** La temperatura máxima diaria exhibe una variabilidad tanto temporal como espacial. A lo largo del año, esta variabilidad puede manifestarse en forma de estaciones distintas, con temperaturas máximas más altas en verano y más bajas en invierno. Además, las diferencias

geográficas, como la latitud y la proximidad al mar, también pueden influir en la variación de la temperatura máxima diaria.

3. **Impacto en la Salud Humana:** Las temperaturas máximas extremas pueden tener serias implicaciones para la salud humana. Las olas de calor, caracterizadas por temperaturas máximas diarias inusualmente altas durante períodos prolongados, pueden aumentar el riesgo de enfermedades relacionadas con el calor, como golpes de calor y deshidratación. Los grupos vulnerables, como los niños pequeños, los ancianos y las personas con enfermedades crónicas, son particularmente susceptibles a los efectos adversos de las temperaturas extremas.
4. **Impacto en la Agricultura y la Ecología:** La temperatura máxima diaria también desempeña un papel crucial en la agricultura y la ecología. Las plantas tienen rangos de temperatura óptimos para el crecimiento y la reproducción, y las temperaturas máximas extremas pueden causar estrés térmico y daño a los cultivos. Además, los ecosistemas terrestres y acuáticos están influenciados por la temperatura, ya que afecta la distribución y el comportamiento de las especies.
5. **Importancia en la Planificación Urbana y la Gestión de Recursos:** En entornos urbanos, la temperatura máxima diaria puede estar influenciada por fenómenos como el efecto isla de calor urbano, donde las áreas urbanas experimentan temperaturas más altas que las áreas circundantes debido a la concentración de edificios y actividades humanas. La comprensión de la temperatura máxima diaria es crucial para la planificación urbana sostenible y la gestión de recursos, incluida la planificación de infraestructuras y la implementación de medidas de adaptación al cambio climático.

## Metodología

Para llevar a cabo el análisis de la aleatoriedad espacial de los datos de temperatura de Bogotá en el año 2023, se utilizó la base de datos de la página del IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales). Los datos utilizados en este análisis consisten en mediciones de temperatura máxima diaria, registradas en una muestra de 7 estaciones en Bogotá, donde se mide dicho parámetro. Cabe aclarar que se extrajeron los valores de la temperatura máxima en el periodo 01/01/2023 - 31/12/2023. Las estaciones seleccionadas fueron las siguientes:

- C.UNIV.AGROP-UDCA: 365 observaciones.
- COL MIGUEL A. CARO: 353 observaciones.<sup>1</sup>
- COLEGIO SAN CAYETAN: 365 observaciones.
- INEM KENNEDY: 365 observaciones.

---

<sup>1</sup>Datos obtenidos hasta el 19/12/2023

- JARDIN BOTANICO - JOSE CELESTINO MUTIS: 365 observaciones.
- NUEVA GENERACION: 365 observaciones.
- UNIVERSIDAD NACIONAL: 365 observaciones.

Luego, se extrajeron las coordenadas geográficas (longitud y latitud) junto con los valores reportados de temperatura máxima diaria de cada estación, para consolidarlo a través de un archivo CSV.

Ahora bien, antes de realizar el análisis, se llevó a cabo un preprocesamiento de los datos para asegurar su calidad. Se eliminaron las columnas no relevantes del conjunto de datos, dejando únicamente las columnas de longitud, latitud y valor de temperatura. Además, se verificó la presencia de datos faltantes en el conjunto de datos, encontrando ningún valor faltante.

Finalmente, todo el análisis se realizó mediante el uso del software R, como herramienta para verificar la aleatoriedad espacial de los datos de la temperatura máxima diaria en Bogotá para el año 2023.

## Análisis de la aleatoriedad espacial

Para evaluar la aleatoriedad espacial de los datos de temperatura, se utilizó el test de Mantel. Este test, de acuerdo al libro de Schabenberger, & Gotway, 2005, nos indica que se busca resolver la siguiente prueba de hipótesis:

$$\begin{cases} H_0 : \text{Hay Aleatoriedad Espacial} \\ H_1 : \text{Hay Correlación Espacial} \end{cases}$$

Este test compara las matrices de distancias geográficas y de diferencias en los valores de temperatura para determinar si existe una correlación espacial significativa entre ellas. Si el  $p$ -valor de los datos es mayor que el estadístico de prueba podemos rechazar la hipótesis nula, en caso contrario, se concluye que no hay evidencia suficiente para rechazar que haya aleatoriedad espacial.

De forma teórica, para hallar el estadístico de prueba del test de Mantel, se obtiene mediante la siguiente fórmula:

$$M = \sum_{i=1}^n \sum_{j=1}^n W_{ij} U_{ij}$$

Donde  $W_{ij} = ||s_i - s_j||$  y  $U_{ij} = (Z(s_i) - Z(s_j))^2$ .

Para entender de una manera mas amena la ecuación anterior,  $W_{ij}$  representa la distancia entre el punto  $s_i$  y el punto  $s_j$ , los cuales pueden estar asociados por la latitud y la longitud de la observación. Es por ello, que esos puntos se expresan mediante coordenadas, dado que solo se busca calcular la distancia entre dos puntos en el espacio.

En el caso de  $U_{ij}$ , este se refiere a la diferencia entre los valores observados en los puntos  $s_i$  y  $s_j$ , elevada al cuadrado. Como en este caso el estudio se basa en la temperatura máxima diaria,  $U_{ij}$  representa la diferencia entre la temperatura máxima diaria de  $s_i$ , con la temperatura máxima diaria de  $s_j$  al cuadrado.

El test de mantel se basa en obtener las distancias de cada una de las coordenadas especificadas, mediante la siguiente forma:

$$w_{ij} = \begin{cases} 1 & \text{Si los sitios } i \text{ y } j \text{ están conectados} \\ 0 & \text{Si los sitios } i \text{ y } j \text{ no están conectados} \end{cases}$$

Ahora bien, en R hay una función que calcula los resultados del test de Mantel, agregando las coordenadas y el valor de cada una de ellas. Por lo que a continuación se explicará de manera detallada el código referente para calcular el test de Mantel en el caso de la medición de la temperatura máxima diaria en Bogotá.

---

```
# Librerias Utilizadas -----

if (!require(ade4)){install.packages("ade4");library(ade4)}
if (!require(readr)){install.packages("readr");library(readr)}
if (!require(readxl)){install.packages("readxl");library(readxl)}
if (!require(dplyr)){install.packages("dplyr");library(dplyr)}

# Fijar Directorio y Carga de Base de Datos -----

setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
getwd()
temp<-read.csv("temperatura.csv", sep=",", header=T)
temp<-temp[,c(3,4,18)]

est <- read_excel("CNE_IDEAM.xls", col_names = TRUE)
est <- est[, 9:12]
est[, 1:2] <- lapply(est[, 1:2], function(x) as.numeric(as.character(x)))
est_1 <- est %>% filter(est[,3] == "Bogota")

# Verificacion de Datos Faltantes -----

missings <- function(x) return(sum(is.na(x)))
apply(temp,2,missings)
apply(est_1,2,missings)
```

```

# Mapa de Estaciones en Bogota -----

plot(x = est_1$LONGITUD, y = est_1$LATITUD, xlab = "Longitud", ylab =
      "Latitud",
      type = "l", col = 4, lwd = 2, main = "Estaciones en Bogota")
points(temp[,2], temp[,1], col=2, pch=16,
        main="Estaciones de Monitoreo de Temperatura en Bogota",
        xlab="Longitud", ylab="Latitud")

# Test de Mantel -----

coord_dists <- dist(cbind(temp$Longitud, temp$Latitud))
valor_dists <- dist(temp$Valor)
tictoc::tic()
res<-mantel.rtest(coord_dists, valor_dists, nrepet = 1000)
tictoc::toc()
res

plot(res, xlab="M", main="Distribucion de Montecarlo Test de Mantel")

```

---

Primero, se utilizaron las librerías *ade4* y *readr*, para cargar las funciones correspondientes, para el análisis respectivo. Luego, se importa la base de datos obtenida, fijando el directorio de trabajo correspondiente. Entonces, al momento de cargar la base de datos, se comprueba si la base correspondiente tiene datos faltantes, los cuales no se obtuvieron, como se comentó anteriormente.

**Nota:** Hay que aclarar que, aunque una de las estaciones no arrojó todas las observaciones en el año, esto no significó inconveniente alguno para el cálculo del test de Mantel, por lo que podemos utilizar la totalidad de los datos obtenidos por esta estación.

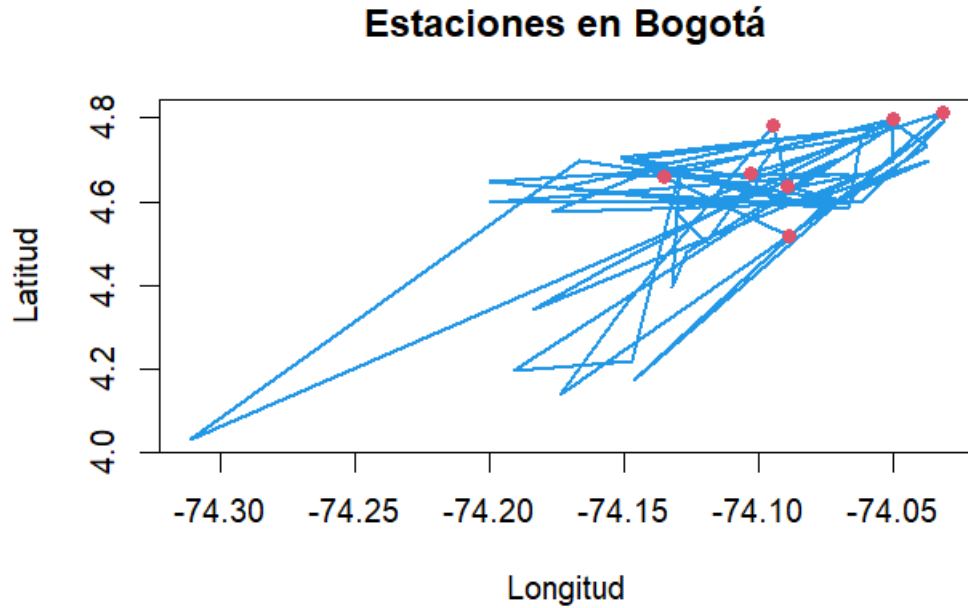
Una vez realizado lo anterior, se establecen las coordenadas, indicando las columnas de  $x$  = Latitud y  $y$  = longitud. Asimismo, se deja en una nueva variable el valor de la temperatura máxima diaria en Bogotá. Finalmente, se utiliza la función *mantel.rtest* para que arroje los resultados que se describirán mas adelante.<sup>2</sup>

A continuación se muestra un mapa de todas las estaciones de medición en Bogotá, donde las líneas azules representan cada una de las estaciones, mientras que los puntos rojos, representan las estaciones escogidas para la medición de la temperatura máxima.

---

<sup>2</sup>Se deja un número de 1000 simulaciones, como las posibles permutaciones, para efectos prácticos de este trabajo

Gráfico 1: Distribución de las estaciones en Bogotá



Fuente: Elaboración propia, a partir de los datos suministrados por la base de datos del IDEAM.

Tal y como se puede observar, las estaciones de medición parecen concentrarse entre coordenadas de latitud (4.4, 4.8); mientras que las coordenadas de longitud, se concentran en  $(-74.15, -74.05)$ . Por lo tanto, veamos si a partir de esta distribución se logrará obtener correlación espacial.

## Resultados

Una vez realizado el test de Mantel, se obtuvieron los siguientes resultados:

- Estadístico de prueba (M): 0,4349915
- $p$ -valor: 0,000999

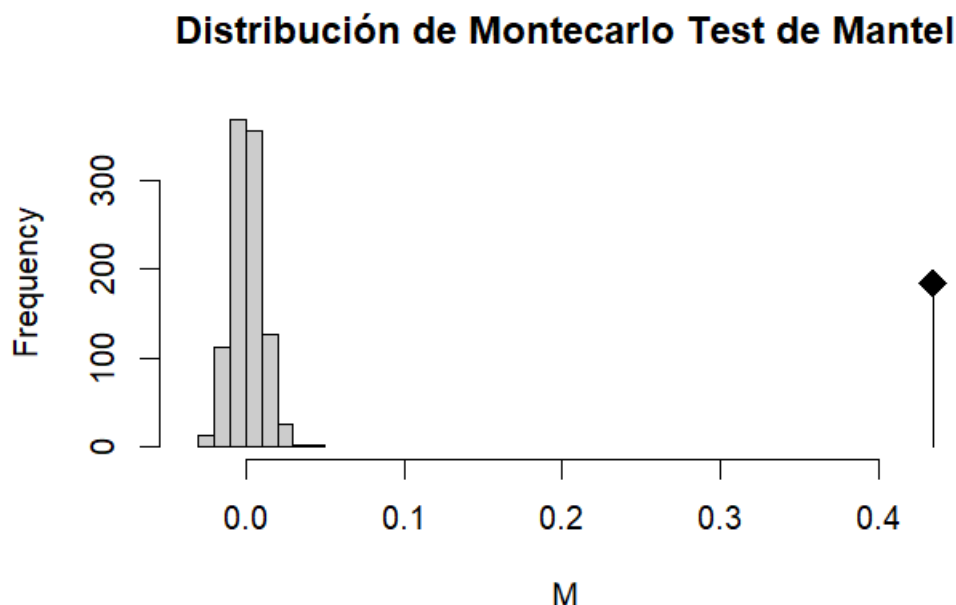
De acuerdo a lo anterior, el  $p$ -valor obtenido es significativamente pequeño ( $p < 0.05$ ), lo que indica que se rechaza la hipótesis nula de existencia de aleatoriedad espacial. Por lo tanto, se concluye que existe evidencia de una estructura espacial no aleatoria en los datos de temperatura máxima diaria en Bogotá, en las estaciones analizadas.

El gráfico de distribución de Montecarlo muestra la distribución de los valores estadísticos generados mediante simulación bajo la hipótesis nula. Se observa que el valor observado del estadístico de prueba (M) se ubica fuera de esta distribución, reforzando el rechazo de la hipótesis nula y afirmando la correlación espacial entre



los datos.

Gráfico 2: Distribución de Montecarlo del Test de Mantel



Fuente: Elaboración propia, a partir de los datos suministrados por la base de datos del IDEAM.

Finalmente, en el gráfico se puede observar que la distribución de los datos bajo la hipótesis nula, se encuentra con valores cercanos a 0, lo cual se aleja bastante del valor observado bajo el Test de Mantel.

## Conclusiones

Los resultados de este análisis sugieren que la distribución de la temperatura máxima diaria en Bogotá estudiada no es aleatoria y presenta una estructura espacial significativa. Este hallazgo puede tener implicaciones importantes para la comprensión de los procesos climáticos locales y regionales, así como para la planificación y gestión de recursos ambientales.

Asimismo, el hecho de presentarse correlación espacial, indica que, de manera local, la detección de la temperatura máxima diaria en la ciudad de Bogotá D.C. radica en la localidad en la que la estación donde se toma la medida, dado que al encontrar indicios de correlación espacial, las localidades adyacentes tienden a obtener medidas similares, comparado con localidades mas alejadas. Por ejemplo, si tomamos la localidad de Suba, con la localidad de Rafael Uribe Uribe, probablemente no encontremos medidas similares de la temperatura máxima diaria, pero si comparamos ahora Suba con Engativá, al ser localidades adyacentes, se obtendrían medidas sim-

ilares que justifican la correlación espacial.

El análisis realizado proporciona evidencia sólida de la existencia de patrones espaciales en los datos de temperatura, lo que subraya la importancia de considerar la dimensión espacial en el estudio del clima y sus impactos. Asimismo, para estudios futuros, sería ideal analizar estaciones de todas las localidades, con el fin de corroborar si la ubicación de las localidades vecinas contribuyen a medidas similares de temperatura máxima de la ciudad, así como a confirmación del resultado de este estudio.

En resumen, la temperatura máxima diaria es un parámetro climático fundamental que influye en una amplia gama de aspectos de la vida humana y los ecosistemas naturales. Su estudio y comprensión son esenciales para abordar los desafíos asociados con el cambio climático, la salud pública, la seguridad alimentaria y la sostenibilidad ambiental. Mediante el uso de herramientas como el análisis espacial y estadístico, como el test de Mantel, podemos profundizar nuestra comprensión de la variabilidad y los patrones de la temperatura máxima diaria, lo que a su vez nos permite tomar decisiones informadas y desarrollar estrategias efectivas de mitigación y adaptación.

## Referencias Bibliográficas

Schabenberger, O., & Gotway, C. (2005). *Statistical methods for spatial data analysis*. CHAPMAN & HALL/CRC.