

Fundamentos de Estadística Espacial

Sesión 1: Introducción y Conceptos Básicos

FRED TORRES CRUZ

Universidad Nacional del Altiplano

9 de septiembre de 2025

1 Introducción a la Estadística Espacial

Contenido de la Presentación

- 1 Introducción a la Estadística Espacial
- 2 Dependencia y Autocorrelación Espacial

Contenido de la Presentación

- 1 Introducción a la Estadística Espacial
- 2 Dependencia y Autocorrelación Espacial
- 3 Tipos de Datos Espaciales

Contenido de la Presentación

- 1 Introducción a la Estadística Espacial
- 2 Dependencia y Autocorrelación Espacial
- 3 Tipos de Datos Espaciales
- 4 R para Análisis Espacial

Propósito de la Sesión

Objetivo Principal

Introducir los conceptos esenciales de la estadística espacial, reconocer los tipos de datos espaciales y configurar el entorno de trabajo en R para elaborar los primeros mapas y visualizaciones.

¿Por qué es importante?

- La información geográfica está presente en la mayoría de datos
- Los métodos estadísticos tradicionales no consideran la dimensión espacial
- La ubicación y proximidad afectan los fenómenos observados
- Necesidad de herramientas especializadas para el análisis espacial

Al finalizar la sesión, las y los participantes serán capaces de:

- RA1** Explicar las diferencias entre estadística tradicional y estadística espacial, y la relevancia de la *Primera Ley de la Geografía* de Tobler
- RA2** Identificar y clasificar tipos de datos espaciales: puntuales, areales y geoestadísticos
- RA3** Configurar un entorno de trabajo en R para análisis espacial e importar datos geográficos
- RA4** Elaborar mapas básicos y temáticos, e interpretar visualmente patrones espaciales

Características del Programa

- **Duración total:** 102 horas académicas
- **Modalidad:** Presencial con componente práctico
- **Sesión actual:** 6 horas (2 bloques de 3 horas)
- **Nivel:** Intermedio–Avanzado

Herramientas de Software

- **R:** Análisis estadístico y programación
- **RStudio:** Entorno de desarrollo integrado
- **QGIS:** Sistema de información geográfica

¿Qué es la Estadística Espacial?

Definition

La **estadística espacial** es el conjunto de métodos estadísticos que incorporan explícitamente la información sobre la ubicación geográfica de las observaciones.

Características Clave

- Considera la posición geográfica
- Analiza patrones espaciales
- Evalúa dependencia espacial
- Incorpora la estructura espacial

Aplicaciones

- **Epidemiología:** Brotes de enfermedades
- **Economía:** Desarrollo regional
- **Ecología:** Distribución de especies
- **Criminalística:** Puntos calientes

Primera Ley de la Geografía de Tobler

Primera Ley de Tobler (1970)

“Everything is related to everything else, but near things are more related than distant things”

“Todo está relacionado con todo, pero las cosas cercanas están más relacionadas que las distantes”

Implicaciones Fundamentales

- **Dependencia espacial:** Los valores en ubicaciones cercanas tienden a ser similares
- **Autocorrelación espacial:** Correlación entre valores y su ubicación
- **Violación de independencia:** Las observaciones no son independientes entre sí
- **Necesidad de métodos especiales:** Los métodos tradicionales pueden ser inadecuados

Estadística Tradicional vs Espacial

Aspecto	Estadística Tradicional	Estadística Espacial
Independencia	Asume observaciones independientes	Considera dependencia espacial
Ubicación	Ignora la posición geográfica	Incorpora coordenadas espaciales
Estructura	Datos tabulares simples	Geometrías y topología
Variabilidad	Homogeneidad en varianza	Heterogeneidad espacial
Escala	Una escala de análisis	Múltiples escalas espaciales

Tipos de Dependencia Espacial

1. Autocorrelación Positiva

- Valores similares se agrupan espacialmente
- **Ejemplo:** Zonas de alta renta tienden a estar juntas
- Patrón más común en la naturaleza

2. Autocorrelación Negativa

- Valores disímiles se ubican cerca
- **Ejemplo:** Patrón de tablero de ajedrez
- Menos frecuente en datos reales

3. Independencia Espacial

- No hay patrón espacial aparente
- **Ejemplo:** Distribución aleatoria
- Supuesto de métodos tradicionales

Medidas Comunes

- Índice de Moran (I): $-1 \leq I \leq 1$
- Índice de Geary (C): $0 \leq C \leq 2$
- Getis-Ord (G): Detección de clusters

Índice de Moran

Fórmula del Índice de Moran

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

Donde:

- n = número de observaciones
- w_{ij} = peso espacial entre i y j
- x_i, x_j = valores observados
- \bar{x} = media de los valores

Interpretación:

- $I > 0$: Autocorrelación positiva
- $I = 0$: Independencia espacial
- $I < 0$: Autocorrelación negativa
- Valor esperado: $E[I] = -1/(n - 1)$

Tipos de Datos Espaciales

Clasificación Fundamental

En estadística espacial, clasificamos los datos según su naturaleza geométrica y proceso generador:

① Datos Puntuales (Point Patterns)

- Eventos localizados en coordenadas específicas

② Datos Areales (Lattice Data)

- Valores agregados en unidades espaciales discretas

③ Datos Geoestadísticos (Continuous)

- Fenómenos continuos medidos en puntos específicos

④ Datos Espaciotemporales

- Incorporan dimensión temporal explícita

Importante

Cada tipo requiere métodos de análisis específicos y diferentes supuestos estadísticos.

Datos Puntuales (Point Patterns)

Características

- Eventos localizados en coordenadas (x, y)
- Puede incluir marcas (atributos)
- Distribución en área de estudio
- Intensidad variable

Preguntas de Investigación

- ¿Es el patrón aleatorio?
- ¿Hay agregación o dispersión?
- ¿Dónde están los clusters?
- ¿Qué factores influyen en la distribución?

Ejemplos

- **Epidemiología:** Casos de enfermedad
- **Criminalística:** Ubicación de delitos
- **Ecología:** Nidos de aves
- **Geología:** Epicentros sísmicos
- **Comercio:** Ubicación de tiendas

Métodos de Análisis

- Función K de Ripley
- Función G del vecino más cercano
- Estimación de densidad kernel
- Análisis de cuadrantes

Datos Areales (Lattice Data)

Características

- Valores agregados por unidades espaciales
- Geometría poligonal definida
- Vecindad entre unidades
- Fronteras fijas

Tipos de Unidades

- **Administrativas:** Países, estados, municipios
- **Regulares:** Grillas, hexágonos
- **Funcionales:** Cuencas, áreas de influencia

Ejemplos

- **Demografía:** Población por municipio
- **Economía:** PIB por región
- **Salud:** Tasas de mortalidad
- **Educación:** Resultados de pruebas
- **Elecciones:** Resultados electorales

Métodos de Análisis

- Autocorrelación espacial (Moran, Geary)
- Modelos autorregresivos espaciales
- Detección de clusters (LISA)
- Cartografía temática

Datos Geoestadísticos (Continuous)

Características

- Fenómeno continuo en el espacio
- Mediciones en ubicaciones específicas
- Correlación espacial continua
- Variabilidad espacial estructurada

Conceptos Clave

- **Semivariograma:** Medida de correlación espacial
- **Rango:** Distancia de dependencia
- **Sill:** Varianza máxima
- **Nugget:** Variabilidad a distancia cero

Ejemplos

- **Clima:** Temperatura, precipitación
- **Suelos:** pH, nutrientes
- **Contaminación:** Concentraciones
- **Topografía:** Elevación
- **Geofísica:** Campos magnéticos

Métodos de Análisis

- Análisis variográfico
- Kriging (interpolación óptima)
- Simulación estocástica
- Modelos de superficie

Semivariograma

Definición

El semivariograma mide la variabilidad espacial como función de la distancia:

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(s) - Z(s + h)] \quad (2)$$

Estimador Empírico

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [Z(s_i) - Z(s_j)]^2 \quad (3)$$

donde $N(h)$ son los pares de puntos separados por distancia h .

Parámetros del Modelo

- **Nugget** (c_0): Discontinuidad en el origen
- **Sill** ($c_0 + c$): Valor asintótico
- **Rango** (a): Distancia de correlación

R para Análisis Espacial

Paquetes Fundamentales

- **sf**: Simple Features (moderno)
- **sp**: Clases espaciales (clásico)
- **terra**: Datos raster (nuevo)
- **raster**: Datos raster (clásico)

Visualización

- **ggplot2**: Grammar of graphics
- **tmap**: Mapas temáticos
- **leaflet**: Mapas interactivos
- **mapview**: Vista rápida

Análisis Estadístico

- **spdep**: Dependencia espacial
- **spatstat**: Patrones de puntos
- **gstat**: Geoestadística
- **spatialreg**: Regresión espacial

Utilidades

- **dplyr**: Manipulación de datos
- **readr**: Lectura de archivos
- **magrittr**: Operadores pipe

Configuración del Entorno

Instalación de Paquetes Básicos

```
1 # Paquetes esenciales
2 install.packages(c("sf", "terra", "sp", "ggplot2", "tmap",
3                   "readr", "dplyr", "spdep"))
4
5 # Verificar instalación de sf y dependencias
6 sf::sf_extSoftVersion()
7
8 # Cargar paquetes
9 library(sf)
10 library(terra)
11 library(ggplot2)
12 library(tmap)
13 library(dplyr)
```

Dependencias Importantes

Objeto sf - Simple Features

Características del Objeto sf

- Integra geometría y atributos en un `data.frame`
- Compatible con el paradigma `tidyverse`
- Implementa estándar ISO 19125 (Simple Features)
- Reemplaza progresivamente a los objetos `sp`

Componentes

- **Geometría:** Columna especial `geometry`
- **Atributos:** Columnas de datos
- **CRS:** Sistema de referencia coordinado
- **Bounding box:** Extensión espacial

Tipos de Geometría

- **POINT:** Puntos
- **LINESTRING:** Líneas
- **POLYGON:** Polígonos
- **MULTI*:** Múltiples geometrías

Creación de Objetos sf

```
1 library(sf); library(dplyr); library(ggplot2)
2
3 # 1) Crear puntos desde una tabla
4 df_pts <- data.frame(
5   id = 1:5,
6   lon = c(-77.05, -77.02, -77.10, -77.03, -77.08),
7   lat = c(-12.05, -12.07, -12.10, -12.04, -12.09),
8   categoria = c("A", "B", "A", "C", "B")
9 )
10
11 pts <- st_as_sf(df_pts, coords = c("lon", "lat"), crs = 4326)
12
13 # 2) Inspección
14 print(pts)
15 st_geometry_type(pts)
16 st_crs(pts)
```

Sistemas de Referencia Coordenados (CRS)

Tipos de CRS

- **Geográfico:** Latitud/Longitud
 - Unidades en grados
 - Ejemplo: WGS84 (EPSG:4326)
- **Proyectado:** Coordenadas planas
 - Unidades en metros
 - Ejemplo: UTM (EPSG:32718)

Operaciones Comunes

- `st_crs()`: Consultar CRS
- `st_transform()`: Transformar CRS
- `st_set_crs()`: Asignar CRS

Códigos EPSG Comunes

- 4326: WGS84 (mundial)
- 32718: UTM Zona 18S (Perú)
- 4326: Geographic WGS84

Importante

Siempre verificar y transformar CRS según el análisis requerido. Las operaciones geométricas requieren sistemas proyectados.

Lectura de Datos Espaciales

```
1 # Leer diferentes formatos
2 # Shapefile
3 distritos <- st_read("datos/distritos.shp")
4
5 # GeoPackage (recomendado)
6 distritos <- st_read("datos/distritos.gpkg", layer = "distritos")
7
8 # GeoJSON
9 puntos <- st_read("datos/puntos.geojson")
10
11 # Desde URL
12 nc <- st_read(system.file("shape/nc.shp", package="sf"))
13
14 # Verificar datos
15 dim(nc)
16 names(nc)
17 st_geometry_type(nc)
18 st_bbox(nc)
```