



Clase 06: Geoestadística y Análisis Espacial

Fundamentos estadísticos para el análisis de datos geoespaciales

Profesor: Francisco Parra O.

4 de septiembre de 2025

USACH - Ingeniería Civil en Informática

Agenda

Introducción: ¿Por qué Geoestadística?

Exploratory Spatial Data Analysis (ESDA)

Interpolación Espacial

Análisis de Patrones de Puntos

Regresión Espacial

Casos de Aplicación

Herramientas y Recursos

Introducción: ¿Por qué Geoestadística?

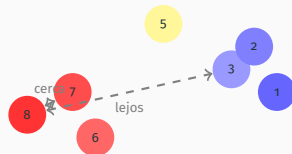
La Primera Ley de la Geografía

Ley de Tobler (1970)

"Todo está relacionado con todo lo demás, pero las cosas cercanas están más relacionadas"

Implicaciones:

- Datos NO independientes
- Proximidad importa
- Métodos clásicos fallan
- Herramientas especiales







Valores similares se agrupan

¿Qué es la Geoestadística?

Definición:


- Rama de la estadística aplicada
- Analiza fenómenos espacialmente correlacionados
- Desarrollada inicialmente para minería (Kriging, 1951)
- Aplicable a cualquier dato georreferenciado

Objetivos principales:

-  Detectar patrones espaciales
-  Interpolar valores desconocidos
-  Identificar clusters y outliers
-  Modelar relaciones espaciales

Aplicaciones en el mundo real:

- **Salud:** Mapeo de enfermedades
- **Ambiente:** Contaminación del aire
- **Inmobiliario:** Valoración de propiedades
- **Crimen:** Hot spots delictuales
- **Agricultura:** Rendimiento de cultivos
- **Minería:** Estimación de reservas

 Sin geoestadística, ignoramos la estructura espacial de los datos

Exploratory Spatial Data Analysis (ESDA)

¿Qué es ESDA?

Conjunto de técnicas para describir y visualizar distribuciones espaciales, identificar localizaciones atípicas, descubrir patrones de asociación espacial y sugerir regímenes espaciales

1. Visualización

- Mapas temáticos
- Cartogramas
- Mapas de calor
- 3D surfaces

2. Análisis Global

- Autocorrelación global
- Tendencias espaciales
- Anisotropía
- Heterogeneidad

3. Análisis Local

- LISA clusters
- Hot/Cold spots
- Outliers espaciales
- Regímenes locales

💡 ESDA es el primer paso: entender los datos antes de modelar

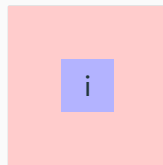
Matrices de Pesos Espaciales

¿Qué son?

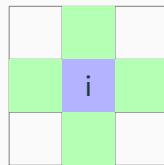
Matrices W que definen las relaciones de vecindad entre observaciones espaciales. Elemento w_{ij} = peso de la relación entre i y j

Tipos principales:

- **Contigüidad:** Queen, Rook
- **Distancia:** K-vecinos, umbral
- **Kernel:** Gaussiano, triangular
- **Custom:** Redes, flujos



Queen: 8 vecinos



Rook: 4 vecinos

Normalización: Por filas para que $\sum_j w_{ij} = 1$

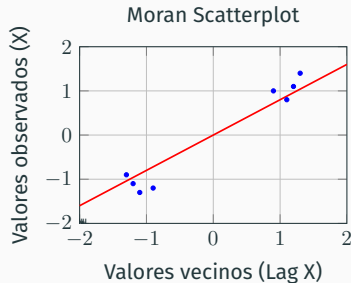
Índice de Moran (I)

$$I = \frac{n}{\sum_{ij} w_{ij}} \frac{\sum_{ij} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

Interpretación:

- $I > 0$: Clustering
- $I = 0$: Aleatorio
- $I < 0$: Dispersión

Rango: $[-1, 1]$ aproximadamente



⚠ Siempre probar significancia estadística con permutaciones

Autocorrelación Espacial Local (LISA)

LISA

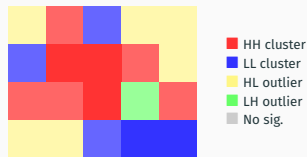
Descomponen el índice global en contribuciones locales para identificar clusters

Moran Local:

$$I_i = \frac{(x_i - \bar{x})}{s^2} \sum_j w_{ij} (x_j - \bar{x})$$

Tipos:

- **HH:** Hot spot
- **LL:** Cold spot
- **HL:** Outlier alto
- **LH:** Outlier bajo



Mapa de clusters LISA

Interpolación Espacial

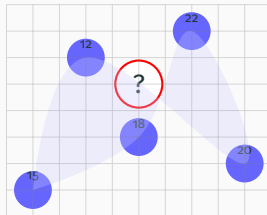
El Problema de la Interpolación

Objetivo

Estimar valores en ubicaciones no muestreadas

Aplicaciones:

- Puntos → Superficie
- Llenar gaps
- Cambiar resolución
- Mapas de predicción



Interpolación



Base: Ley de Tobler

IDW

$$\hat{z} = \frac{\sum w_i z_i}{\sum w_i}$$

$$w_i = 1/d_i^p$$

Pros:

- Simple
- Rápido

Contras:

- Sin incertidumbre
- "Bull's eyes"

Splines

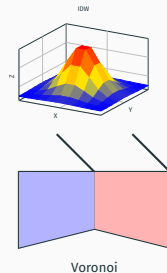
- Suave
- Min. curvatura
- Exacto

Voronoi

- Polígonos
- Constante
- Discontinuo

Trend

- Polinomio
- Tendencia
- Inexacto



Semivariogramas: La Base del Kriging

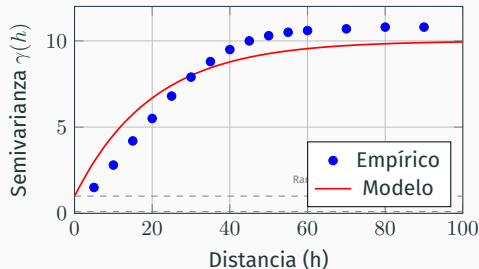
Semivariograma

Variabilidad vs distancia

$$\gamma(h) = \frac{1}{2N(h)} \sum [z_i - z_{i+h}]^2$$

Componentes:

- **Nugget:** Var. en $h=0$
- **Sill:** Var. total
- **Range:** Alcance



Modelos teóricos: Esférico, Exponencial, Gaussiano, Matérn

Kriging: El Mejor Estimador Lineal Insesgado (BLUE)

Tipos de Kriging:

- **Simple:** Media conocida y constante
- **Ordinario:** Media desconocida constante
- **Universal:** Con tendencia (drift)
- **Co-Kriging:** Múltiples variables
- **Indicator:** Variables categóricas

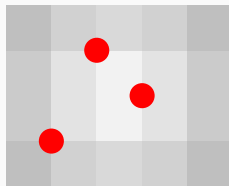
Ecuación de Kriging Ordinario:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$$

con $\sum_{i=1}^n \lambda_i = 1$ (insesgado)

Ventajas:

- ✓ BLUE
- ✓ Con varianza
- ✓ Anisotropía
- ✓ Óptimo



Mapa de incertidumbre:

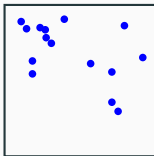
Varianza de Kriging

Análisis de Patrones de Puntos

Patrones Espaciales de Puntos

Point Pattern Analysis

Análisis de distribución: aleatorio, agrupado o regular



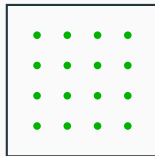
Aleatorio (CSR)
Complete Spatial
Randomness

Poisson



Agrupado
Atracción entre puntos

Clustering



Regular
Repulsión entre puntos

Uniforme

⚠ Patrón varía con escala

Función K de Ripley

Función K

Puntos esperados dentro de radio r

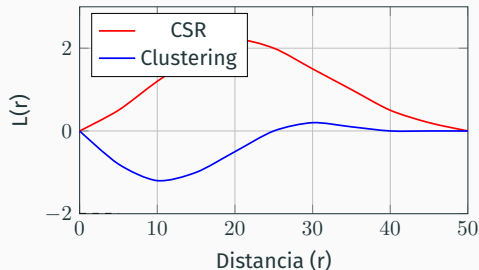
$$K(r) = \lambda^{-1} E[N(r)]$$

Interpretación:

- $K(r) = \pi r^2$: CSR
- $K(r) > \pi r^2$: Clustering
- $K(r) < \pi r^2$: Regularidad

Función L (normalizada):

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r$$



💡 Usar simulaciones Monte Carlo para bandas de confianza

Hot Spot Analysis: Getis-Ord G_i^*

Estadístico G_i^*

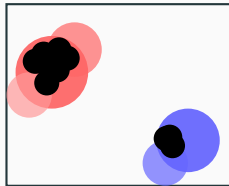
Detecta hot spots y cold spots significativos

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - (\sum_j w_{ij})^2}{n-1}}}$$

Z-scores:

- $> 2,58$: Hot 99 %
- $> 1,96$: Hot 95 %
- $< -1,96$: Cold 95 %
- $< -2,58$: Cold 99 %

Aplicación: Análisis de criminalidad



Hot/Cold Spots de Crimen

💡 Políticas focalizadas

Regresión Espacial

El Problema con OLS en Datos Espaciales

OLS: $y = X\beta + \varepsilon$

Violaciones:

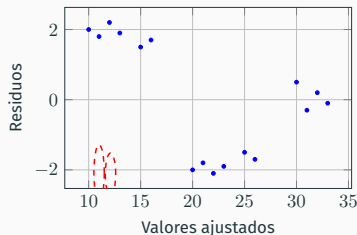
- ⊗ Independencia
- ⊗ Homoscedasticidad
- ⊗ Estacionariedad

Consecuencias:

- Sesgo
- Inferencia errónea
- Mal ajuste
- SE subestimados

Diagnósticos espaciales:

Patrón espacial en residuos



Moran's I detecta autocorrelación

Modelos de Regresión Espacial

SAR:

$$y = \rho W y + X\beta + \varepsilon$$

- ρ : Autocorrelación
- $W y$: Lag espacial
- Spillovers

SEM:

$$y = X\beta + u$$

$$u = \lambda W u + \varepsilon$$

- λ : Error AC
- Var. omitidas

SDM:

$$y = \rho W y + X\beta + W X\theta + \varepsilon$$

- Lags de X y y
- Flexible
- Efectos dir/indir

Selección:

1. Moran's I
2. Tests LM
3. AIC/BIC
4. Especificación

Geographically Weighted Regression (GWR)

GWR

Coeficientes locales que varían espacialmente

Modelo GWR:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i$$

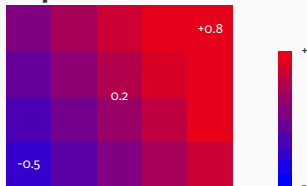
Ventajas:

- Captura heterogeneidad espacial
- Coeficientes locales interpretables
- Identifica regímenes espaciales
- Flexible y visual

Consideraciones:

Ancho de banda crítico

Mapa de coeficientes locales:



β_k para educación

💡 Los coeficientes varían suavemente en el espacio

Casos de Aplicación

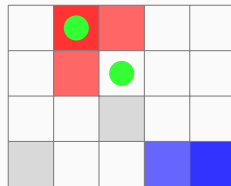
Caso 1: Análisis de Precios Inmobiliarios

Problema:

- Valoración de propiedades en Santiago
- Efectos de vecindario
- Amenidades y des-amenidades

Metodología:

1. ESDA: Moran's I, LISA clusters
2. Kriging de precios/m²
3. GWR con variables:
 - Distancia a metro
 - Áreas verdes
 - Seguridad
 - Educación



Resultados esperados: Clusters de precio

Insights:

- Premium del 15 % cerca del metro
- Efectos spillover entre comunas
- Heterogeneidad en valoración de amenidades

Caso 2: Epidemiología Espacial

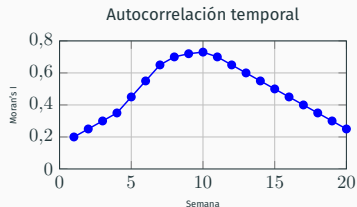
Análisis COVID-19 en Chile:

Datos:

- Casos por comuna
- Movilidad (Google Mobility)
- Densidad poblacional
- Nivel socioeconómico

Análisis realizado:

- Hot spots temporales (G_i^*)
- Difusión espacial (Moran's I)
- Modelo SIR espacial
- Predicción con Kriging espacio-temporal



Hallazgos:

- Clustering inicial en comunas ABC1
- Difusión jerárquica y por contagio
- Persistencia en zonas vulnerables

Herramientas y Recursos

Herramientas para Geoestadística

Python:

- **PySAL:** Suite completa
- **GeoPandas:** DataFrames espaciales
- **scikit-gstat:** Variogramas
- **PyKriging:** Kriging
- **pointpats:** Patrones de puntos
- **mgwr:** GWR

R:

- **sp/sf:** Objetos espaciales
- **gstat:** Geoestadística
- **spdep:** Dependencia espacial
- **spatstat:** Point patterns
- **GWmodel:** GWR
- **tmap:** Visualización

Software especializado:

- **GeoDa:** ESDA visual
- **SAGA GIS:** Geoestadística
- **ArcGIS:** Spatial Analyst
- **QGIS:** Processing toolbox
- **GS+:** Variogramas

💡 Python + PySAL es la combinación más versátil para investigación

Mejores Prácticas

✓ Do's:

- Siempre hacer ESDA primero
- Verificar estacionariedad
- Validar modelos con datos independientes
- Considerar múltiples escalas
- Reportar incertidumbre
- Usar múltiples métodos
- Documentar decisiones

✗ Don'ts:

- Ignorar autocorrelación espacial
- Usar OLS sin diagnósticos
- Interpolar sin validación
- Asumir isotropía siempre
- Extrapolar más allá del rango
- P-hacking espacial
- Olvidar el MAUP

⚠ MAUP: Modifiable Areal Unit Problem - los resultados pueden cambiar con la agregación

Conclusiones

Resumen de la Clase

Conceptos clave aprendidos:

- ✓ Primera Ley de Tobler
- ✓ Autocorrelación espacial (Moran, LISA)
- ✓ Matrices de pesos espaciales
- ✓ Interpolación (IDW, Kriging)
- ✓ Semivariogramas
- ✓ Análisis de patrones de puntos
- ✓ Hot spots (Getis-Ord)
- ✓ Regresión espacial (SAR, SEM, GWR)

Habilidades desarrolladas:

- Detectar dependencia espacial
- Crear superficies interpoladas
- Identificar clusters y outliers
- Modelar con conciencia espacial
- Generar mapas de incertidumbre
- Validar modelos espaciales

Mensaje clave

Datos espaciales = métodos especiales

Libros recomendados:


- .Applied Spatial Data Analysis with R Bivand et al.
- "Geographic Data Science with Python Rey et al.
- "Geostatistics for Natural Resources Isaaks & Srivastava

Práctica con notebooks:

- ESDA de datos chilenos
- Kriging de contaminación en Santiago
- Hot spots de criminalidad
- GWR de precios inmobiliarios

Proyecto sugerido:

- Análisis geoestadístico completo de un fenómeno local
- Comparar métodos de interpolación
- Implementar modelo de regresión espacial
- Desarrollar dashboard interactivo con resultados

 Próxima clase: Machine Learning Geoespacial

Entregables esperados

Análisis ESDA completo

- Autocorrelación global y local
- Clusters y outliers espaciales

Comparación de métodos

- Mínimo 3 métodos de interpolación
- Validación cruzada y métricas

Visualizaciones interactivas

- Mapas con Folium/Plotly
- Dashboard integrado

Reporte final

- Metodología e interpretación
- Recomendaciones aplicadas

¿Preguntas?

✉ francisco.parra.o@usach.cl

🐙 github.com/franciscoparrao

Material y notebooks disponibles en el repositorio del curso