Examen Integrador – Módulo I

# 1. Análisis Exploratorio Espacial de Datos (ESDA)

Para este análisis se utilizó la base de datos Chi-SDOH, correspondiente a las zonas censales de Chicago, que contiene variables sociales, económicas y de salud. Se trabajó con dos variables numéricas:

- vcrimert: Tasa de crimen violento por zona censal.  
- EP\_DISABL: Porcentaje de personas con discapacidad en cada zona.

Se utilizó el software GeoDa para calcular el Índice de Moran Bivariado con dos matrices espaciales: rook y queen.

## 1.1 Resultados del Índice de Moran Bivariado

Con la matriz de pesos 'queen', el valor del índice de Moran fue de -0.030. Este valor indica que no existe autocorrelación espacial significativa entre la tasa de crimen violento y el porcentaje de personas con discapacidad. El valor está muy cerca de cero, lo que refleja una distribución aleatoria.

## Conclusión del Moran I:

No se encontró autocorrelación espacial significativa entre las dos variables analizadas. Esto indica que las zonas con alto crimen violento no tienden a estar rodeadas por zonas con altos porcentajes de discapacidad, y viceversa.

# 2. Regresión Espacial

Se aplicaron tres modelos para evaluar la relación entre las variables: modelo OLS, modelo SAR (Spatial Lag), y modelo SEM (Spatial Error).

## 2.1 Regresión OLS

La regresión lineal ordinaria (OLS) arrojó un coeficiente de pendiente de -0.001 con un p-valor de 0.168, indicando que no hay una relación significativa entre las variables. El R² fue de 0.002, por lo que el modelo explica muy poca variabilidad.

## 2.2 Modelo SAR (Spatial Lag)

En el modelo SAR se obtuvo un coeficiente espacial (rho) de -3.217 con p-valor de 0.000. Aunque el parámetro espacial fue significativo, el modelo presentó problemas graves: R² igual a 0 y valores de AIC y log-likelihood no disponibles. El coeficiente de la variable independiente tampoco fue significativo (p = 0.249).

## 2.3 Modelo SEM (Spatial Error)

En el modelo SEM, el coeficiente espacial (lambda) fue de 1.238 con p-valor de 0.000, lo que indica una fuerte dependencia espacial en los errores. La variable independiente no fue significativa (p = 0.159), pero el modelo es estadísticamente estable y presentó un R² de 0.987. El test de Breusch-Pagan no detectó heterocedasticidad (p = 0.085).

## Conclusión:

El modelo SEM fue el más adecuado, ya que mostró mejores resultados estadísticos, mayor estabilidad y un coeficiente espacial significativo. Aunque no se halló una relación directa entre crimen y discapacidad, sí se identificó una fuerte dependencia espacial.

# 3. Machine Learning

## 3.1 Aprendizaje Supervisado

Un ejemplo real es el uso de modelos de predicción de fraude en tarjetas de crédito. Estos modelos se entrenan con datos etiquetados (fraude o no fraude) y aprenden a predecir nuevas transacciones sospechosas con base en patrones históricos.

## 3.2 Aprendizaje No Supervisado

Un caso común es la segmentación de clientes en marketing usando clustering. Los algoritmos como K-means agrupan clientes según similitudes sin usar etiquetas previas, permitiendo descubrir patrones ocultos en los datos.

## 3.3 Aprendizaje por Refuerzo

Se usa en robótica o videojuegos, donde un agente aprende a tomar decisiones con base en recompensas y penalizaciones. Un ejemplo real es AlphaGo, que aprendió a jugar y ganar a humanos jugando miles de partidas simuladas.