

# **Primera Entrega, Proyecto Final**

## **Modelación y Simulación I**

Julián David Cárdenas Bernal

### **1. Requisitos de Datos**

Se hará uso de la Plataforma Nacional de Datos Abiertos de Colombia, específicamente de la información suministrada por los conjuntos de datos "Sectores críticos por exceso de velocidad" y "Sectores críticos de siniestralidad vial", del primero se proporciona un método de frecuencia de registros de velocidad mayor a 80 km/h en un tramo vial en diferentes municipios y regiones ("Datos Abiertos Colombia", 2021); y del segundo, compartiendo la misma información descriptiva del primero, se indica numéricamente una calificación según análisis espacial GiZScore, una probabilidad de ocurrencia GiPValue, y la cantidad de fallecidos en el sector.

Dado que se incluyen más valores cuantitativos explícitos al adjuntar ambos conjuntos de datos, además de la información categórica, se tiene la expectativa de que los datos sean suficientes para construir un modelo que permita predecir y analizar el número de fallecidos en cuanto a factores de siniestralidad vial de diferentes sectores.

### **2. Enfoque Matemático**

Se utilizará un modelo de regresión para analizar cómo las variables independientes, algunas ya mencionadas, afectan la dependiente, el número de fallecidos. En primer lugar se considera la técnica de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) si es que esta última variable sigue una distribución lineal con respecto a las variables independientes.

El modelo OLS es adecuado como punto de partida para otros análisis de regresión espacial en caso de ser necesarios ("ArcGIS Pro", s.f.). De este modo, se asume que la relación entre las variables independientes y dependientes

es aproximadamente lineal, que los errores de las predicciones son independientes, con su varianza siendo constante a lo largo de todas las observaciones y, en el caso del modelo por OLS, que estos tengan una distribución normal, para así considerar sus inferencias estadísticas.

### **3. Plan de Implementación**

Para el procesamiento y limpieza de los datos, se manejarán los valores faltantes, ya sea por imputación o eliminación, se hará la transformación de variables categóricas, posiblemente como valores indicadores para su inclusión en los modelos de regresión, y podrían normalizarse las coordenadas.

Para la construcción del modelo, se implementarán en Python las librerías de pandas y statsmodels en el caso de OLS. Si se detectan patrones espaciales en los datos, se emplearía un modelo de regresión espacial para corregir la dependencia de este tipo y se incluiría la librería geopandas.

Para la validación del modelo OLS, se recurrirá a las diferentes medidas resultantes de su implementación con la librería statsmodels, tales como el R cuadrado, la prueba F, y los test de Durbin-Watson y de Jarque-Bera. Si se requiere un modelo espacial, se validará con el índice de Moran.

### **4. Resultados Esperados**

El modelo debería predecir el número de fallecidos en función de las características de la vía y de siniestralidad, determinando si ciertos factores tienen una relación significativa. Las visualizaciones se expresarían con un mapa de calor de siniestralidad, con gráficos de dispersión e histogramas de los errores.

Los resultados podrían ayudar a priorizar las áreas con mayor riesgo de siniestralidad, orientando las inversiones en infraestructura vial y políticas públicas según los factores significativos.

## Lista de Referencias

- ArcGIS Pro*. (s.f.). <https://pro.arcgis.com/es/pro-app/latest/tool-reference/spatial-statistics/regression-analysis-basics.htm>
- Datos Abiertos Colombia*. (2021, 21 de abril). [https://www.datos.gov.co/Transporte/SECTORES-CRITICOS-POR-EXCESO-DE-VELOCIDAD/24ny-2dhf/about\\_data](https://www.datos.gov.co/Transporte/SECTORES-CRITICOS-POR-EXCESO-DE-VELOCIDAD/24ny-2dhf/about_data)