

Regresión Avanzada

Trabajo Final

Contexto:

En el siguiente trabajo analizaremos datos relativos a la seguridad vial urbana, que presentamos en `SegVial.txt`. El objetivo del trabajo es encontrar características urbanas para determinar la vulnerabilidad de los usuarios viales. De este modo se busca evaluar qué políticas públicas son más eficaces para alcanzar nivel de fatalidad y accidentes de tránsito violentos cero.

Contamos con información de diferentes 24 ciudades de 5 países distintos de Europa de los cuales tenemos en cuenta los accidentes de tránsito, los datos corresponden al año 2018. La métrica de seguridad vial utilizamos el índice MDS (muerte o daño severo). Contamos para esto con información correspondiente a la infraestructura de las carreteras, las características de los vehículos, datos de tráfico y controles establecidos, factores ambientales y la distribución de la *participación modal*.¹

Buscaremos modelos para identificar las características descriptas teniendo en cuenta las características de las dos partes involucradas en el accidente.

Variables independientes son:

- **DenPob**: densidad poblacional (población/ km^2)
- **ArCiclista**: relación entre el área total para ciclistas y el área total para conducir.
- **ArBajaVel**: relación entre el área total limitada para baja velocidad y el área total para conducir.
- **PMPeatones**: participación modal para caminar(%).
- **PMCiclistas**: participación modal para cyclists (%).
- **PMTPublico**: participación modal para transporte público (%).
- **PMVMotor**: participación modal para vehículos con motor (%).

¹porcentaje de viajeros que utilizan un determinado tipo de transporte o el número de viajes que realizan utilizando ese transporte. La participación modal es una componente importante en el desarrollo del transporte sustentable en una ciudad o región.

- **Temp:** temperatura media anual (C).
- **Precipitacion:** precipitación media anual (mm).
- **PBI:** producto bruto interno medio per cápita (EUR).

Por otra parte, como variables de respuesta tenemos el índice MDS que provocan choques del vehículo A al B ($A \rightarrow B$), lo notamos AB.

- **PeatAuto:** Peaton \rightarrow Auto o Taxi.
- **CicAuto:** Ciclista \rightarrow Auto o Taxi.
- **V2RMSM** Vehículo con dos ruedas a motor \rightarrow si mismos.
- **V2RMAuto** Vehículo con dos ruedas a motor \rightarrow Auto o Taxi.
- **AutoSM** Auto o Taxi \rightarrow si mismos.
- **AutoAuto** Auto o Taxi \rightarrow Auto o Taxi.

1. Hacer un análisis exploratorio de los datos.
2. Estandarizar las variables de respuesta de acuerdo a la a población (cada millón de habitantes). Es decir, para cada ciudad i y variable de respuesta y , considerar $y_i^{estan} = \frac{y_i}{Poblacion_i} 10^6$, donde población es una variable de la base de datos.
3. Estandarizar las variables regresoras, de forma tal que tengan media cero y varianza uno.
4. En cada caso ajustar un modelo de regresión parsimonioso.
5. Comparar los modelos ajustados y sacar conclusiones en términos del problema planteado.