

Tarea 5

Profesor: Leo Ferres **Valor:** 4 puntos de homework **Librerías clave:** pandas, geopandas, cmdstanpy, arviz, matplotlib, seaborn

Análisis Socioeconómico y Epidemiológico con Modelado Bayesiano

Objetivo:

Explorar la relación entre indicadores socioeconómicos y datos epidemiológicos (idealmente de dengue) a nivel sub-nacional usando herramientas de modelado bayesiano con Stan y visualización de resultados con ArviZ.

Parte 1: Recolección de Datos

1. **País asignado:** Usa el mismo país de las tareas anteriores.
2. **Datos requeridos:**
 - **Divisiones administrativas nivel 2** (por ejemplo, provincias, departamentos, etc.). Usa el shapefile limpio de tareas anteriores.
 - **Datos socioeconómicos:** Ejemplos: ingresos, pobreza, nivel educativo, acceso a servicios, etc. (puedes usar fuentes como World Bank, UNData, DHS, etc.).
 - **Datos epidemiológicos:** Casos de enfermedades, idealmente *dengue*, para el año más reciente y a la escala geográfica más desagregada posible.

Parte 2: Limpieza y Unificación de Datos

1. Limpia y armoniza las bases para que todos los datos estén disponibles por la misma unidad geográfica.
2. Crea un `DataFrame` que contenga:
 - Una fila por unidad administrativa.
 - Columnas con variables socioeconómicas.
 - Una columna con número de casos (o tasa) de la enfermedad.

Parte 3: Modelo Bayesiano con Stan

1. **Aprende e implementa:** Estudia los siguientes materiales para familiarizarte con inferencia bayesiana y Stan:

- Charla en R Rosario (con ejemplos en RStan)
- Stan tutorials en YouTube
- Manual de usuario de Stan
- cmdstanpy

2. **Modelo sugerido:** Haz una regresión bayesiana (e.g. Poisson o binomial negativa si hay sobredispersión) donde:

$$\text{Casos}_i \sim \text{Distribución apropiada}(\mu_i) \log(\mu_i) = \beta_0 + \beta_1 \cdot X_{1,i} + \dots + \beta_k \cdot X_{k,i}$$

donde X_j son variables socioeconómicas.

3. Usa `cmdstanpy` para correr el modelo y `arviz` para analizar los resultados:
 - Trazos de cadenas (`traceplots`)
 - Distribuciones posteriores
 - Resumen de parámetros

Parte 4: Análisis de Resultados

1. ¿Qué variables están más asociadas a la presencia de dengue u otra enfermedad?
2. ¿Qué zonas muestran más riesgo?
3. ¿Cómo es la incertidumbre en los parámetros?

Entrega

1. Un notebook `.ipynb` o script `.py` con:
 - Código para recolección, limpieza, análisis y modelado.
2. Un archivo `README.md` con:
 - Descripción de fuentes de datos
 - Supuestos del modelo
 - Principales resultados e interpretación
 - Gráficos generados con `arviz` y `matplotlib`

Criterios de Evaluación (4 puntos)

Criterio	Puntos
Datos bien integrados y representados	1.0
Implementación correcta del modelo en Stan	1.5
Análisis con ArviZ y visualización	1.0
Interpretación clara y crítica de resultados	0.5