**Maestría en Ciencia y Análisis de Datos- Universidad Mayor de San Andrés**

**Modelos lineales y modelos lineales generalizados**

NOMBRE:

INFORMACION DE CONTACTO:

Correo electrónico:

Número de teléfono celular (opcional):

**PRACTICA 3 y PRACTICA 4**

**Cada pregunta tiene un valor de 15 puntos,**

**Límite de entrega: Lunes 5 a medianoche.**

**Entrega en la plataforma Moodle. Trabajar con un nivel de significancia de 5%**

**Copiar capturas de pantalla de los resultados R para cada respuesta, cuando corresponda**

**Práctica 3:** Utilizando los datos de crimenes.csv (link GitHub debajo), construya un modelo para el numero de crímenes violentos (variable dependiente), considerando como variables explicativas el tipo de institución educativa (tipo) y la región:

1. Estime modelos de Poisson y quasi-Poisson con offset (2 puntos).
2. Estime modelos de Poisson y quasi-Poisson sin offset (2 puntos).
3. Estime un modelo binomial negativo con offset (2 puntos).
4. Estime modelos de Poisson y binomiales negativos con inflación de ceros con offset (2 puntos).
5. Compare los modelos y decida cuál de los modelos es el más apropiado (2 puntos). Explique.
6. Estime el modelo final que escogió con errores estándar robustos basados en estimadores sandwich, e interprete la significancia estadística y el significado de los estimadores (5 puntos).

<https://raw.githubusercontent.com/rogon666/UMSA/main/MLMLG/datos/crimenes.csv>

**Práctica 4:** Utilizando los datos de empleados\_balanceado.csv (link GitHub debajo), construya un modelo logit (logístico) para la probabilidad de renunciar (variable dependiente), considerando como variables explicativas : la edad, la distancia al trabajo, el ingreso mensual, el número de empresas en las que trabajo previamente(num\_empresas\_trabajadas), si trabaja horas extras o no (horas\_extras), el porcentaje de incremento salarial que recibió (incremento\_salarial\_porcentaje), los años que lleva trabajando en la empresa (anios\_en\_la\_empresa), los años que lleva trabajando en el puesto (anios\_en\_el\_puesto), los años desde el ultimo ascenso (anios\_desde\_ultimo\_ascenso) y los años que lleva trabajando con el actual supervisor (anios\_con\_actual\_supervisor).

1. Estime el modelo logit saturado con todas las variables explicativas (2 puntos).
2. Estime el modelo logit saturado con todas las variables explicativas utilizando estimadores sandwich para los errores estándar (2 puntos).
3. Reduzca el modelo logit robusto (basado en estimadores sándwich) excluyendo las variables no significativas (2 puntos).
4. Estime un modelo Gompit (utilizando una función de enlace cloglog) con las variables que incluyo en el modelo reducido logit (2 puntos).
5. Compare los modelos y decida cuál de los modelos es el más apropiado (2 puntos). Explique.
6. Interprete la significancia estadística y los estimadores del modelo final reducido que escogió, empleando estimadores sandwich para los errores estándar, ya sea Gompit o Logit (5 puntos).

<https://raw.githubusercontent.com/rogon666/UMSA/main/MLMLG/datos/empleados_balanceado.csv>