

1. Estimación bajo un modelo Gamma

Breve explicación: Los Gamma-ray bursts (GRBs) son uno de los fenómenos más exóticos de los que estudia la astronomía moderna. Fueron descubiertos por casualidad durante los 60's por los satélites americanos y rusos y aparecieron como explosiones con emisiones de rayos gamma con una duración de 0.1-100 segundos en lugares aleatorios del cielo. A partir de mediados de los 90's se empezó a ver a los GRBs como un efecto secundario del nacimiento de un agujero negro. Los GRBs de larga duración se asocian a la emisión de un chorro de material energético a velocidades cercanas a la de la luz (relativistas) de una estrella supergigante antes de su colapso final y explosión de supernova. Los GRBs de corta duración se asocian a dos estrellas neutrónicas que se espiralan y emergen como un agujero negro, otra vez eyectando un chorro de material energético a velocidades relativistas. Como el chorro se expande y se enfría, según cálculos astrofísicos, un resplandor de radiación en longitudes de onda larga, tales como rayos-X, bandas visibles y de radio, deberían ser detectados y deberían decaer en tiempos de escala horas-días-semanas.

El conjunto de datos GRB contiene las 63 mediciones del brillo en la escala espectral 0.4-4.5 keV a tiempos que van entre 2 minutos y 5 días después de la explosión. Las columnas registran las siguientes variables: 1) el tiempo de observación en segundos y 2) X-ray flux (en unidades de 10^{-11} ergcm²s, 2-10 keV).

- a) Realizar un scatterplot de la variable TIME vs. FLUX y otro tomando logaritmo en ambas variables. ¿Cuál le parece la escala más adecuada para analizar los datos?
- b) Blustin et al. (2006) proponen analizar los datos usando un modelo lineal a trozos teniendo en cuenta un punto de corte δ de manera de permitir que los puntos con valores de $\log(TIME)$ mayores a δ puedan tener otra pendiente. A partir de los gráficos, ¿parece razonable esta propuesta?
- c) En este ítem proponemos realizar tres ajustes usando la familia de distribución Gamma con link log considerando los siguientes modelos para μ_i , la esperanza de $FLUX_i$:

$$\mathcal{M}_1 : \log(\mu_i) = \beta_0 + \beta_1 \log(TIME)$$

$$\mathcal{M}_2 : \log(\mu_i) = \beta_0 + \beta_1 \log(TIME) + \beta_2(\log(TIME) - 7)(\log(TIME) > 7)$$

$$\mathcal{M}_3 : \log(\mu_i) = \beta_0 + \beta_1 \log(TIME) + \beta_2(\log(TIME) - 8.5)(\log(TIME) > 8.5)$$

de manera que en los modelos \mathcal{M}_2 y \mathcal{M}_3 los puntos con $\log(TIME)$ mayores a $\delta = 7$ y 8.5 , respectivamente, puedan tener otra pendiente.

Ajustar los tres modelos, evaluar la significación de los coeficientes en cada \mathcal{M}_j , $j = 1, 2, 3$ y elegir el modelo que le parece más adecuado para los datos. Justificar.

d) Mediante la función `resid` computar los residuos de Pearson y deviance para cada modelo. Graficar los residuos vs. los valores ajustados en cada \mathcal{M}_j y realizar boxplots de los residuos computados. Comentar.

2. Para el Ejercicio 2 de la Guía 4 computar los residuos de Pearson y deviance. Graficar los residuos vs. los valores ajustados y realizar boxplots de los residuos computados. Interpretar.

3. **Bondad de ajuste caso Binomial.** Cuando evaluamos la bondad del ajuste global de un modelo debemos tener ciertos cuidados.

Consideremos el conjunto de datos dados en el archivo `birthwt`. Los datos corresponden a un estudio de factores de riesgo de bajo peso en recién nacidos. Los datos fueron recogidos en el Baystate Medical Center, Springfield, Massachusetts, en 1986. El siguiente cuadro describe las variables consideradas.

Variable	Nombre
Código de identificación	ID
Bajo peso al nacer (0= peso al nacer \neq 2500gr) (1= peso al nacer $<$ 2500gr)	LOW
Edad de la madre en años	AGE
Peso en libras en la última menstruación	LWT
Raza(1=Blanca, 2=Negra, 3=Otros)	RACE
Fuma durante el embarazo (1= Si, 0=No)	SMOKE
Historia de trabajo prematuro (0=ninguna, 1=uno, 2=dos, etc.)	PTL
Historia de hipertensión (1=si, 0=no)	HT
Presencia de irritación uterina (1=si, 0=no)	UI
Nro. visitas al médico durante el primer trimestre (0=ninguna, 1=uno, 2=dos, etc.)	FTV
Peso al nacer en gramos	BWT

Se desea explicar el bajo peso en niños (LOW) a partir de un conjunto de variables explicativas.

- a)* Defina las variables factoriales en los casos que sean necesarios.
- b)* A partir de LWT cree una variable dicotómica LWD que tome valores 1 si el peso de la madre antes de la última menstruación es menor a 110, 0 en caso contrario.
- c)* Realice un ajuste de la variable LOW usando en modelo las variables AGE, LWD, RACE, SMOKE, HT, UI, donde la variable LWD es la definida en el ítem anterior. Analice la significación de cada una de las variables?
- d)* Observe los valores de los estadísticos de bondad de ajuste y su significación. ¿Cuáles son sus conclusiones?
- e)* Mediante el test de Hosmer-Lemeshow (función `hoslem.test` de la librería `ResourceSelection`) evaluar la significación de la bondad del ajuste y comparar con las conclusiones obtenidas en el ítem anterior.