

Modelos Lineales con R

Práctica 8: GLM.

Profesor: Andrés García Medina
andres.garcia.medina@uabc.edu.mx

Fecha de entrega: miercoles 29 de mayo, 2024 (11:59 pm).

Instrucciones: Subir un documento pdf a classroom con las respuestas de cada uno de los ejercicios solicitados. Adjunta el código fuente en formato `.r` o `.ipynb`. Justificar detalladamente cada una de sus respuestas.

Ejercicio 1: Toxicidad de la cipermetrina en orugas del tabaco

Los siguientes datos corresponden a un estudio de toxicidad de la cipermetrina en orugas del tabaco¹. Se contó con 240 orugas, 120 machos y 120 hembras. Se dividieron los 120 machos en 6 grupos de 20 orugas cada uno, similarmente las orugas hembra fueron separadas en 6 grupos. A los diferentes grupos se les aplicaron 6 dosis crecientes de cipermetrina y se registraron las diferentes tasas de mortandad en cada grupo después de 3 días de exposición a la cipermetrina. Se observaron los siguientes datos:

dosis	1	2	4	8	16	32
$\log_2(dosis)$	0	1	2	3	4	5
machos	1	4	9	13	18	20
hembras	0	2	6	10	12	16

Suponemos que $y_{ij} \sim \text{Binomial}(n, p_i)$, de modo que

$$\begin{aligned} f(y_{ij}) &= \binom{n}{y_{ij}} p_{ij}^{y_{ij}} (1 - p_{ij})^{n - y_{ij}} \\ &= \exp \left\{ \frac{1}{1/n} \left(\frac{y_{ij}}{n} \log \frac{p_{ij}}{1 - p_{ij}} + \log(1 - p_{ij}) \right) + \log \left(\binom{n}{y_{ij}} \right) \right\} \end{aligned}$$

¹En Venables y Ripley (2002), estos datos se analizan mediante una regresión logística con logaritmo de dosis en base 2 como variable explicativa.

Así, para la variable aleatoria y_{ij}/n , i.e. proporción observada de muertes, tenemos una distribución de la familia exponencial con un predictor lineal, y asociamos al modelo una μ_{ij} vía la liga canónica:

$$\text{logit}(p_{ij}) = \eta_{ij} = x_{ij}^T \beta = \beta_0 + \beta_1 \text{género}_i + \beta_2 \text{ldosis}_j + \beta_3 \text{género}_i * \text{ldosis}_j$$

Si codificamos a los machos como 0 y a las hembras como 1, entonces, bajo la parametrización anterior, la matriz X es de la forma

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 1 & 0 & 3 & 0 \\ 1 & 0 & 4 & 0 \\ 1 & 0 & 5 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 4 & 4 \\ 1 & 1 & 5 & 5 \end{bmatrix}$$

- (a) Determine explícitamente los siguiente parámetros del modelo de regresión Binomial:
 - θ
 - ϕ
 - $a(\phi)$
 - $b(\theta)$
 - $c(y, \phi)$
 - $g_c(\mu)$: liga canónica
 - $\mathbb{E}(Y)$
 - $Var(Y)$
- (b) Ajuste un modelo lineal generalizado (regresión Binomial) a través de la función `glm` de **R**. Se espera que obtenga los siguientes resultados: $\beta_0 = -2.819$, $\beta_1 = -0.175$, $\beta_2 = 1.259$, $\beta_3 = -0.353$.
- (c) Considere el Método de scoring de Fisher (Algoritmo Newton-Raphson Modificado: variante IRLWS) para determinar las coeficientes a través de la implementación de la siguiente función recursiva:

$$\beta^{k+1} = (X^T U X)^{-1} X^T U y^* \quad (1)$$

donde

$$\eta = X\beta^k \quad (2)$$

$$X^T = [x_1, \dots, x_n]_{p \times n} \quad (3)$$

$$D = \text{diag}(1/g'(\mu_i)) \quad (4)$$

$$W = \text{diag}(1/\text{Var}(y_i)) \quad (5)$$

$$U = DWD \quad (6)$$

$$y^* = \eta + D^{-1}(y - \mu) \quad (7)$$

$$(8)$$

Considere un vector inicial $\beta = [-1, 0, 0, 0]^T$, una tolerancia mínima (norma mínima de delta) de 1×10^{-4} y un máximo de 100 iteraciones. Se recomienda usar la librería `lm` para estimar el vector β_k en cada interacción, en esta caso se debe pasar el vector de pesos U a la función.

- (d) Grafique la curva estimada diferenciando el comportamiento esperado para la hembras (H) y machos (M). Se espera que obtenga una figura como la mostrada en1

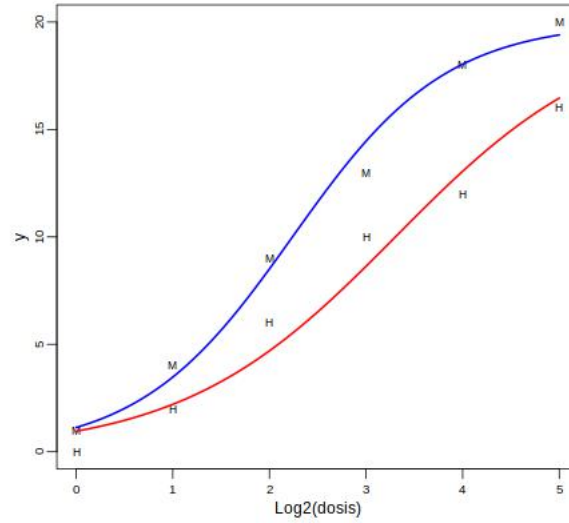


Figure 1:

Ejercicio 2

Considere el estudio transversal de pacientes con una forma de cáncer de piel llamada melanoma maligno (Roberts et al. 1981). Para una muestra de $n = 400$ pacientes, se registró el sitio del tumor y su tipo histológico. La cuestión de interés es si existe alguna asociación entre el tipo de tumor y su localización. Intente responder esta pregunta con ayuda de los modelos lineales generalizados. Los datos se muestran en la figura2

Tumor type	Site			Total
	Head & neck	Trunk	Extrem -ities	
Hutchinson's melanotic freckle	22	2	10	34
Superficial spreading melanoma	16	54	115	185
Nodular	19	33	73	125
Indeterminate	11	17	28	56
Total	68	106	226	400

Figure 2: