# Tema 13. Regresión Logística. Ejercicios

*Abdelmalik Moujahid, Iñaki Inza y Pedro Larrañaga*
*Departamento de Ciencias de la Computación e Inteligencia Artificial*
*Universidad del País Vasco–Euskal Herriko Unibertsitatea*

1. Supongamos que estamos interesados en estudiar la relación entre las variables predictoras $X_1$ ="Estatus social", con posibles valores 0 (baja) y 1 (alta), $X_2$ ="Fumador", con posibles valores 0 (no) y 1 (sí) y $X_3$ ="Presión sanguínea sistólica" (variable continua), y la variable clase $C$ ="Mortalidad por enfermedad cardiovascular".

   Después de analizar datos de 200 individuos durante un período de años fijado, se han obtenido los siguientes dos modelos de regresión logística:

Modelo 1

| Variable | Coeficiente |
|---|---|
| $X_1$ | $-0{,}5200$ |
| $X_2$ | $-0{,}5200$ |
| $X_3$ | $-0{,}5600$ |
| $X_1 \times X_2$ | $0{,}1750$ |
| $X_1 \times X_3$ | $-0{,}0330$ |
| Constante | $-1{,}1800$ |

Modelo 2

| Variable | Coeficiente |
|---|---|
| $X_1$ | $-0{,}5000$ |
| $X_2$ | $-0{,}4200$ |
| $X_3$ | $0{,}0100$ |
| Constante | $-1{,}1900$ |

   a) Establecer las fórmulas de regresión logística correspondientes a cada uno de los modelos anteriores.

   b) Expresar cada uno de los modelos de regresión logística anterior por medio de su manera logit.

   c) Usando el modelo 1 de regresión logística, calcular la probabilidad de mortalidad por enfermedad cardiovascular para un individuo de clase social alta, fumador y con una presión sanguínea sistólica de 150.

   d) Usando el modelo 2 de regresión logística, calcular la probabilidad de mortalidad por enfermedad cardiovascular para cada uno de los siguientes individuos:

Persona 1 clase social alta, fumador y con presión sanguínea de 150

Persona 2 clase social baja, fumador y con presión sanguínea de 150

   e) Teniendo en cuenta los resultados del apartado anterior, calcular el *risk ratio* (RR) comparando la persona 1 y la persona 2. Interpretar el valor de dicho RR.

2. Supongamos que tenemos 8 variables predictoras, denotadas por $X_1$, $X_2$, $X_3$, $X_4$, $X_5$, $X_6$, $X_7$ y $X_8$, y una variable predictora $C$. A partir de una muestra de tamaño 609 se han obtenido dos modelos que denotamos por Modelo $A$ y Modelo $B$, y cuyas estimaciones de los parámetros, así como sus correspondientes desviaciones estándard aparecen reflejadas en Tabla 1 y Tabla 2.

| Variable | Coeficiente | Desviación estándar | Valor Chi-cuadrado | Significatividad |
|---|---|---|---|---|
| Constante | $-6{,}7727$ | 1,1401 | 35,29 | 0,0000 |
| $X_1$ | 0,5976 | 0,3520 | 2,88 | 0,0896 |
| $X_2$ | 0,0322 | 0,0152 | 4,51 | 0,0337 |
| $X_3$ | 0,0087 | 0,0033 | 7,17 | 0,0074 |
| $X_4$ | 0,3695 | 0,2936 | 1,58 | 0,2083 |
| $X_5$ | 0,8347 | 0,3052 | 7,48 | 0,0062 |
| $X_6$ | 0,4393 | 0,2908 | 2,28 | 0,1309 |

Tabla 1: Información de parámetros correspondiente al Modelo $A$

| Variable | Coeficiente | Desviación estándar | Valor Chi-cuadrado | Significatividad |
|---|---|---|---|---|
| Constante | $-4{,}0474$ | 1,2549 | 10,40 | 0,0013 |
| $X_1$ | $-12{,}6809$ | 3,1042 | 16,69 | 0,0000 |
| $X_2$ | 0,0349 | 0,0161 | 4,69 | 0,0303 |
| $X_3$ | $-0{,}0055$ | 0,0042 | 1,70 | 0,1917 |
| $X_4$ | 0,3665 | 0,3278 | 1,25 | 0,2635 |
| $X_5$ | 0,7735 | 0,3272 | 5,59 | 0,0181 |
| $X_6$ | 1,0468 | 0,3316 | 9,96 | 0,0016 |
| $X_7$ | $-2{,}3299$ | 0,7422 | 9,85 | 0,0017 |
| $X_8$ | 0,0691 | 0,0143 | 23,18 | 0,0000 |

Tabla 2: Información de parámetros correspondiente al Modelo $B$

Además se conoce que $-2ln\widehat{L}_A = 400{,}41$, mientras que $-2ln\widehat{L}_B = 347{,}28$.

Teniendo en cuenta toda la información anterior, se pide:

a) Utilizar el test de la razón de verosimilitud para testar el modelo $B$ frente al modelo $A$.

b) En el modelo $A$, utilizar el test de Wald para tratar de eliminar alguna variable del modelo.

c) Igual que en el apartado b), para el modelo $B$.