

# Análisis de datos

Ejercicios de Regresión logística

2º de Biología

Departamento de Matemáticas

Universidad Autónoma de Madrid

2011/12

# Ejercicio 1

El 4 de julio de 1999 una tormenta con vientos que excedían las 90 millas por hora azotó el nordeste de Minnesota, en EE.UU., causando graves daños en los bosques de un parque natural de la zona. Los científicos analizaron los efectos de la tormenta determinando para más de 3600 árboles del parque su diámetro en cm (variable  $D$ ), y una medida de la severidad local de la tormenta (variable  $S$ ). Los árboles que habían muerto fueron codificados con  $Y = 1$ , y los árboles supervivientes con  $Y = 0$ . Los datos se encuentran en el fichero `tormenta.sav`. Utilizaremos un modelo de regresión logística para expresar la probabilidad de no supervivencia de los árboles mediante las variables  $D$  y  $S$ . Utilizando SPSS, se pide:

- a. Hallar las estimaciones puntuales de los parámetros del modelo. La probabilidad estimada de que un árbol sobreviva a la tormenta, ¿crece o decrece con el diámetro?
- b. Obtener un intervalo de confianza de nivel 95 % para el coeficiente del diámetro,  $\beta_1$ .
- c. Obtener un intervalo de confianza de nivel 95 % para  $\exp(\beta_1)$ .
- d. La variable  $D$ , ¿tiene una influencia significativa en el modelo? ¿Y la variable  $S$ ? Dar respuestas al nivel de significación del 5 %.
- e. Estimar la probabilidad de que no sobreviva un árbol cuyo diámetro es de 30 cm situado en una zona en que la fuerza de la tormenta viene dada por  $S = 0,8$ .

## Los resultados que se obtienen con SPSS son:

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup>						
D	,097	,005	346,022	1	,000	1,102
S	4,424	,189	545,122	1	,000	83,412
Constante	-3,543	,127	774,463	1	,000	,029

a. Variable(s) introducida(s) en el paso 1: D, S.

**a.** Los parámetros del modelo son:

$$\beta_0 = -3,543, \beta_D = 0,097, \beta_S = 4,424.$$

Dado que  $\beta_D > 0$ , la probabilidad de que un árbol **no** sobreviva a la tormenta crece con el diámetro (y por tanto, la probabilidad de que sobreviva decrece con el diámetro).

**b.**

$$IC_{95\%}(\beta_D) = (0,097 \pm 1,96 \cdot 0,005) = (0,0872, 0,1068)$$

**c.**

$$IC_{95\%}(e^{\beta_D}) = (e^{0,0872}, e^{0,1068}) = (1,091, 1,113)$$

**d.** Al nivel de significación  $\alpha = 0,05$  ambos coeficientes son significativamente distintos de cero ya que para ambos el  $p$ -valor (Sig.) del estadístico de contraste es menor que 0,05.

**e.** La probabilidad de **no** sobrevivir se estima en

$$\hat{p} = \frac{1}{1 + e^{-(-3,543 + 0,097 \cdot 30 + 4,424 \cdot 0,8)}} = 0,95$$

por tanto, la probabilidad de sobrevivir sería 0,05.

## Ejercicio 2.

Con los mismos datos del ejercicio anterior, ajustar un modelo de regresión logística simple que incluya sólo el diámetro como variable regresora para explicar la probabilidad de no supervivencia de los árboles, y responder a las siguientes cuestiones relacionadas con este modelo:

- a.** A nivel  $\alpha = 0,01$ , es posible afirmar que el diámetro influye en la probabilidad de que un árbol sobreviva?
  
- b.** Escribe la regla para clasificar un árbol como superviviente o no superviviente en función de su diámetro.

Los resultados que se obtienen con SPSS son:

Variables en la ecuación

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup> D	,098	,005	405,223	1	,000	1,102
Constante	-1,702	,083	422,608	1	,000	,182

a. Variable(s) introducida(s) en el paso 1: D.

**a.** A la vista de estos resultados, el  $p$ -valor (Sig.) del estadístico de contraste para el coeficiente de  $D$  es menor que 0,01, por tanto se puede afirmar que es significativamente distinto de cero es decir, se puede afirmar que *el diámetro influye en la probabilidad de que un árbol sobreviva.*

**b.** Dado que  $-1,702 + 0,098D = 0$  da  $D = 17,37$ , la regla será:

- ▶ no sobrevive si:  $D > 17,37$
- ▶ sobrevive si:  $D < 17,37$