

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



Master en Ciencia de Datos

ENTREGA DE EJERCICIOS

MAE: METODOS AVANZADOS DE ESTADISTICA

Autor: Guillermo Hoyo Bravo

Tutor: José Berrendero

Diciembre 2021

Índice general

1. Tercera Entrega	1
1.1. Ejercicio 6	1

Capítulo 1

Tercera Entrega

1.1. Ejercicio 6

En un experimento descrito en Prentice (1976) se expuso una muestra de escarabajos a cierto pesticida. Tras cinco horas de exposición a distintos niveles de concentración del pesticida algunos de los escarabajos murieron y otros sobrevivieron. Los resultados para cada dosis aparecen en la tabla siguiente:

Dosis (log CS2 mg/l)	N. insectos	N. muertos
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	53
1.8610	62	61
1.8839	60	60

Formula un modelo de regresión logística para analizar estos datos y estima la probabilidad de que muera un escarabajo expuesto durante cinco horas a una dosis de concentración 1.8

La función glm ajusta la regresión logística (generalized linear model), también será necesario especificar que la variable respuesta sea una binomial. La variable independiente X será la concentración al pesticida expuesto. La variable respuesta Y será 1 si el escarabajo muere y 0 si sobrevive.

Este es el código y los resultados obtenidos de él:

```
# Definimos nuestros datos
dosis <- c(1.6907, 1.7242, 1.7552, 1.7842,
          1.8113, 1.8369, 1.8610, 1.8839)
supervivencia <- data.frame(
  total = c(59, 50, 62, 56, 63, 59, 62, 60),
  muertos = c(6, 13, 18, 28, 52, 53, 61, 60)
)
supervivencia["vivos"] <- supervivencia$total - supervivencia$muertos
y <- as.matrix(supervivencia[c("muertos", "vivos")])

# Definimos el modelo
model <- glm(y~dosis, family=binomial)
summary(model)
```

```
Call:
glm(formula = y ~ dosis, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6858  -0.4106   0.8411   1.3552   1.6363

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -59.301     5.166  -11.48  <2e-16 ***
dosis         33.490     2.903   11.54  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 270.712  on 7  degrees of freedom
Residual deviance:  12.944  on 6  degrees of freedom
AIC: 43.085

Number of Fisher Scoring iterations: 4
```

El modelo logístico es:

$$P(Y = 1|x) = \frac{1}{1 + e^{-59,301 + 33,490x}}$$

Prediciendo que un escarabajo muera con $X = 1.8$ generamos el siguiente código:

```
prob <- predict(model, data.frame(dosis=1.8), type='response')
prob
```

Con resultado; 1: 0.727425830185318

Por tanto, la probabilidad de que un escarabajo muera despues de haber estado expuesto durante 5 horas a una concentración de pesticida de 1.8[logCS₂mg/l]esde0727.