

# Bondad de ajuste

## Contents

1	Introduccion	1
2	R-cuadrado en regresión de Poisson	1
3	Contraste para un grupo de coeficientes	3
4	Contraste de bondad de ajuste	3

## 1 Introduccion

Se estima el siguiente modelo de regresión logística:

```
d = read.csv("datos/Aircraft_Damage.csv")
d$bomber = factor(d$bomber, labels = c("A4", "A6"))
m1 = glm(damage ~ bomber + load + experience, data = d, family = poisson)
summary(m1)

##
## Call:
## glm(formula = damage ~ bomber + load + experience, family = poisson,
##      data = d)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.406023  0.877489 -0.463  0.6436
## bomberA6     0.568772  0.504372  1.128  0.2595
## load        0.165425  0.067541  2.449  0.0143 *
## experience -0.013522  0.008281 -1.633  0.1025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 53.883 on 29 degrees of freedom
## Residual deviance: 25.953 on 26 degrees of freedom
## AIC: 87.649
##
## Number of Fisher Scoring iterations: 5
```

El objetivo es analizar como de bueno es el modelo de regresión logística que se ha estimado.

## 2 R-cuadrado en regresión de Poisson

Se puede definir un  $R^2$  de manera similar a como se hizo en regresión lineal. La manera habitual es:

$$R^2 = 1 - \frac{D_1}{D_0}$$

donde D es la desviación del modelo (deviance en inglés). Se calcula como el doble de la verosimilitud del modelo calculada en los parámetros estimados (en valor absoluto):

$$D = |2\log L(\hat{\beta})|$$

$$\log L(\hat{\beta}) = \sum_{i=1}^n (-\lambda_i + y_i \log(\lambda_i) - \log(y_i!))$$

$$\lambda_i = \exp(x_i^T \beta)$$

Se definen dos desviaciones:

- D1: la desviación del modelo analizado.
- D0: la desviación del modelo en el que solo se estima  $\beta_0$ .

```
source("funciones/poisson_funciones.R")
(D1 = abs(2*poisson_logL(coef(m1),d$damage,model.matrix(m1))) )

## [1] 79.64922
m0 = glm(damage ~ 1, data = d, family = poisson)
summary(m0)

##
## Call:
## glm(formula = damage ~ 1, family = poisson, data = d)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.4274    0.1474   2.899  0.00374 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 53.883  on 29  degrees of freedom
## Residual deviance: 53.883  on 29  degrees of freedom
## AIC: 109.58
##
## Number of Fisher Scoring iterations: 5
(D0 = abs(2*poisson_logL(coef(m0),d$damage,model.matrix(m0))) )

## [1] 107.5791
(R2 = 1 - D1/D0)

## [1] 0.2596219
```

Si  $R^2 \approx 1$  el modelo se ajusta muy bien a los datos, y  $R^2 \approx 0$  implica un mal ajuste. Es decir,  $R^2 \approx 0$  significa que la verosimilitud de ambos modelos es muy parecida, luego  $\beta_1 \approx \beta_2 \approx \beta_k \approx 0$ .

### 3 Contraste para un grupo de coeficientes

Supongamos que tenemos dos modelos:

$$\lambda_i = \exp(x_i^T \beta)$$

$$\lambda_{A,i} = \exp(x_{A,i}^T \beta_A)$$

donde  $\beta_A$  es un subconjunto de  $\beta$ . Supongamos que  $\dim(\beta_A) = m$  y  $\dim(\beta) = k$ , con  $m < k$ . Si  $\beta_B$  representa los parámetros que están en  $\beta$  pero no están en  $\beta_A$ , se puede resolver el siguiente contraste:

$$H_0 : \beta_B = 0, \quad H_1 : \beta_B \neq 0$$

En el caso de que la hipótesis nula sea cierta, se tiene que:

$$G = D_A - D_1 \sim \chi_{k-m}^2$$

donde  $D_1$  es la desviación del modelo con parámetros  $\beta$  y  $D_A$  es la desviación del modelo con parámetros  $\beta_A$ .

- Si  $G < \chi_{\alpha}^2$  no se rechaza la hipótesis nula.
- Si  $G \geq \chi_{\alpha}^2$  se rechaza la hipótesis nula.

Es decir, valores grandes del estadístico significa que la verosimilitud de ambos modelos es muy diferente, luego  $\beta_B \neq 0$ . Para valores pequeños de G, ambos modelos son muy parecidos, luego los regresores  $\beta_B$  no aportan nada al modelo, es decir,  $\beta_B = 0$ .

Por ejemplo, queremos resolver el contraste con hipótesis nula  $\beta_2 = \beta_3 = 0$ :

```
mA = glm(damage ~ bomber, data = d, family = poisson)
(DA = abs(2*poisson_logL(coef(mA), d$damage, model.matrix(mA))))
```

```
## [1] 91.97952
```

```
(G = DA - D1)
```

```
## [1] 12.3303
```

```
# valor crítico del contraste
```

```
k = length(coef(m1))
```

```
m = length(coef(mA))
```

```
qchisq(0.95, df = k-m)
```

```
## [1] 5.991465
```

Luego se rechaza la hipótesis nula.

### 4 Contraste de bondad de ajuste

Utilizando el contraste anterior entre el modelo con todos los regresores y el modelo con solo  $\beta_0$  se puede analizar la bondad del modelo. Es decir, se puede contrastar:

- H0: el modelo estimado NO es adecuado ( $\beta_1 = \beta_2 = \dots = \beta_k = 0$ )
- H1: el modelo estimado es adecuado.

El estadístico del contraste es

$$G = D_0 - D_1 \sim \chi^2_{k-1}$$

En este caso:

```
(G = D0 - D1)
```

```
## [1] 27.9299
k = length(coef(m1))
(pvalor = 1-pchisq(G, k-1))

## [1] 3.757192e-06
```

Luego el modelo es muy adecuado.