

Análisis de los residuos en el modelo logit

Contents

1	Residuos del modelo logit	1
2	Ejemplo	1

1 Residuos del modelo logit

Los residuos del modelo se definen como siempre:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\pi}_i, \quad i = 1, \dots, n$$

Se definen también los residuos de Pearson:

$$r_i = \frac{e_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$$

El gráfico de los residuos de Pearson frente a los valores previstos, $\hat{\pi}_i$, mostrará los datos siguiendo dos curvas:

- los puntos con $y = 0$ estarán en $r_i = \frac{\hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$.
- los puntos con $y = 1$ estarán en $r_i = \frac{1 - \hat{\pi}_i}{\sqrt{\hat{\pi}_i(1 - \hat{\pi}_i)}}$.

El análisis de estos gráficos no es tan directo como en el caso de regresión lineal, ya que los datos seguirán esas curvas independientemente de si el modelo es adecuado o no.

2 Ejemplo

Se estima el siguiente modelo de regresión logística:

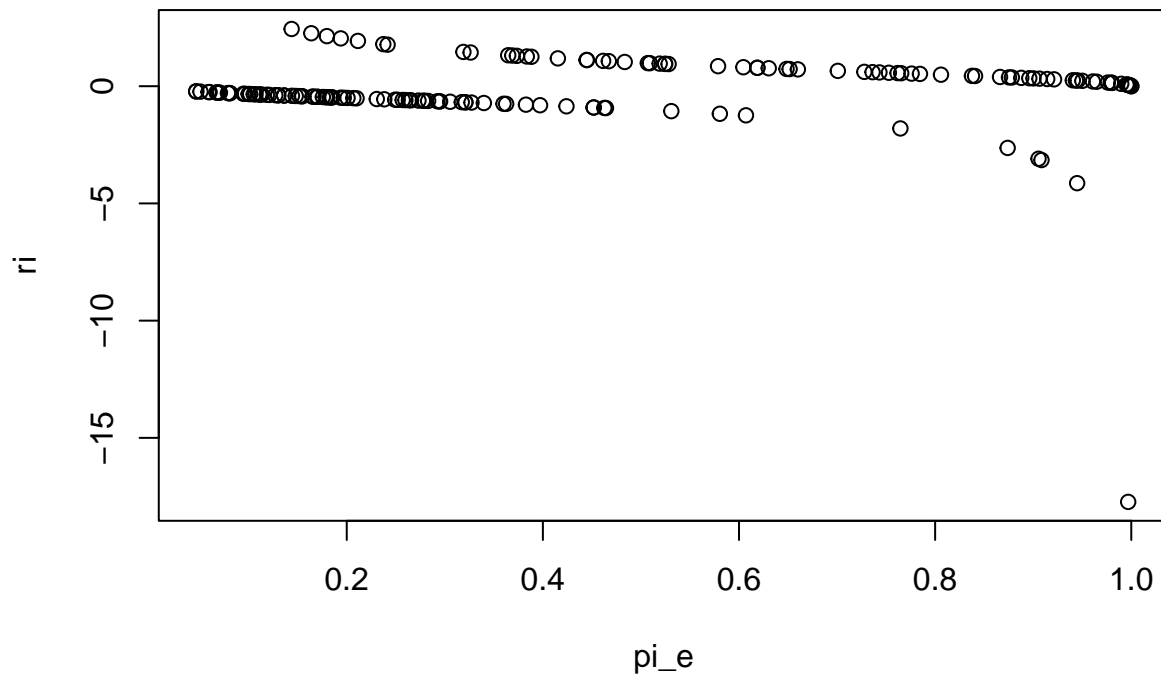
```
d = read.csv("datos/MichelinNY.csv")
m = glm(InMichelin ~ Food + Decor + Service + Price, data = d, family = binomial)
summary(m)
```

```
##
## Call:
## glm(formula = InMichelin ~ Food + Decor + Service + Price, family = binomial,
##      data = d)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3923  -0.6723  -0.3810   0.7169   1.9694
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -11.19745     2.30896  -4.850 1.24e-06 ***
```

```
## Food          0.40485    0.13146    3.080  0.00207 **
## Decor          0.09997    0.08919    1.121  0.26235
## Service       -0.19242    0.12357   -1.557  0.11942
## Price          0.09172    0.03175    2.889  0.00387 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 225.79  on 163  degrees of freedom
## Residual deviance: 148.40  on 159  degrees of freedom
## AIC: 158.4
##
## Number of Fisher Scoring iterations: 6
```

Los residuos son:

```
pi_e = predict(m, type = "response")
ei = d$InMichelin - pi_e
ri = ei/sqrt(pi_e*(1-pi_e))
plot(pi_e,ri)
```



Se observa que hay un dato con comportamiento diferente al resto.

```
pos = which.max(ri)
d[pos,]
```

```
##      InMichelin Restaurant.Name Food Decor Service Price
## 135           1      Paradou    19    17      18     38
```

Se trata del restaurante Paradou, que a pesar de tener puntuaciones muy bajas está incluido en la Guía Michelin.