

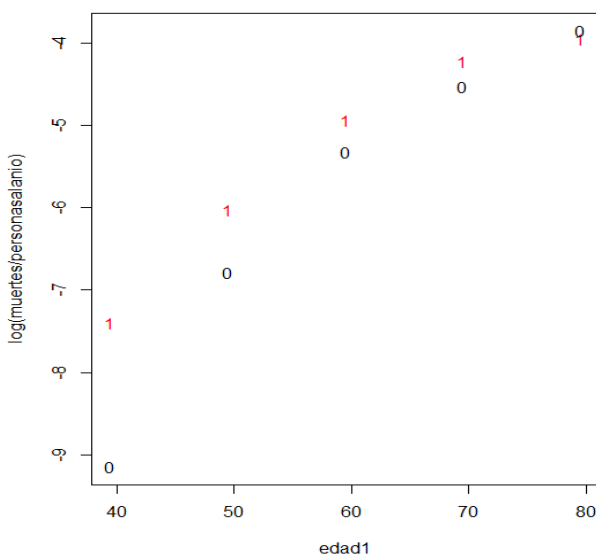
EJEMPLO REGRESION POISSON PARA MODELAR TASAS

```
> coronarias
  edad edad1 edad2 muertes personasanio fumadores
1 35-44 39.5    1     32      52407         1
2 45-54 49.5    2    104      43248         1
3 55-64 59.5    3    206      28612         1
4 65-74 69.5    4    186      12663         1
5 75-84 79.5    5    102       5317         1
6 35-44 39.5    1     2      18790         0
7 45-54 49.5    2    12      10673         0
8 55-64 59.5    3    28       5710         0
9 65-74 69.5    4    28       2585         0
10 75-84 79.5    5    31       1462         0
```

Los datos corresponden a las muertes por enfermedad coronaria de doctores británicos seguidos de 1951 a 1961.
¿La tasa de muerte es mayor en los fumadores? Si sí, ¿qué tanto?

La relación del $\log(\text{muertes}/\text{personasanio})$ y edad no es exactamente lineal.
En este caso en cada grupo hay diferente número de personas y además en cada grupo puede haber hasta 10 años de diferencia en edad, por lo que se debe tomar en cuenta las diferencias en exposición al riesgo, es decir la suma de años-persona.
La variable edad es categórica.
Las variables edad1 y edad2 son numéricas, y son prácticamente lo mismo, pues una es transformación lineal de la otra.

Log(tasas de mortalidad) para fumadores y no fumadore



MODELOS

Con las variables explicativas edad (categórica) y fumadores:

```
modelo1<-glm(muertes~edad+fumadores,offset=log(personasanio),family=poisson,coronarias)
```

Con la variable explicativas edad1 (numérica) y fumadores:

```
modelo2<-glm(muertes~edad1+fumadores,offset=log(personasanio),family=poisson,coronarias)
```

Con la variable explicativas edad2 (numérica), fumadores y la interacción:

```
modelo3<-glm(muertes~edad2 +fumadores+edad2*fumadores,offset=log(personasanio),family=poisson,coronarias)
```

Con la variable explicativas edad1 (numérica), fumadores y la interacción:

```
modelo33<-glm(muertes~edad1+fumadores+edad1*fumadores,offset=log(personasanio),family=poisson,coronarias)
```

Se crea la variable edad22 para permitir que el modelo se curve.

```
edad22<-edad2*edad2
```

Con la variable explicativas edad2 (numérica), edad22, fumadores y la interacción:

```
modelo4<-glm(muertes~edad2+edad22+fumadores+edad2*fumadores,offset=log(personasanio),family=poisson,coronarias)
```

¿Qué pasa si quitamos la interacción?

```
modelo5<-glm(muertes~ edad22 + edad2 +fumadores,offset=log(personasanio),family=poisson, coronarias)
```

Tabla para los 6 modelos ajustados con los valores de deviance, gl y pvalue

```

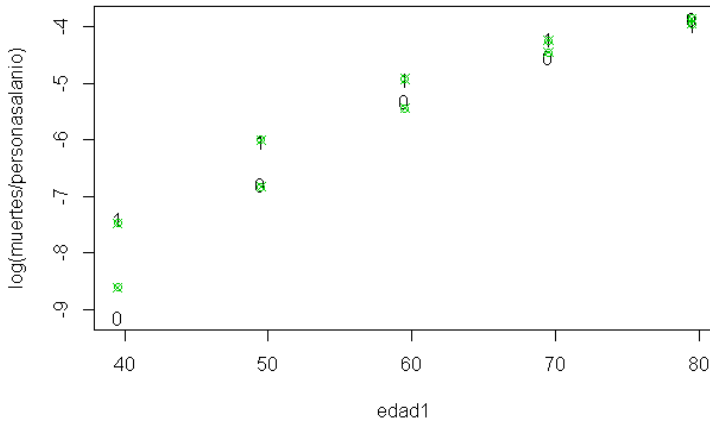
      [,1]      [,2]          [,3] [,4]
[1,] "modelo"  "deviance"    "g1" "p-value"
[2,] "modelo1" "12.132366396288" "4"  "0.0163936252689151"
[3,] "modelo2" "69.1820803831422" "7"  "2.16138218434025e-12"
[4,] "modelo3" "59.8934929458888" "6"  "4.7309267614537e-11"
[5,] "modelo33" "59.8934929458888" "6"  "4.7309267614537e-11"
[6,] "modelo4" "1.63537013146436" "5"  "0.896939299974585"
[7,] "modelo5" "12.1755453094149" "6"  "0.0581652458635171"

```

El mejor modelo es el modelo4. Cabe mencionar que el modelo5 y el modelo1, tienen casi la misma deviance , pero hay diferencias en el pvalue.

USANDO MODELO 4

tasas de mortalidad para fumadores y no fumadores



```
> summary(modelo4)
```

Call:

```
glm(formula = muertes ~ edad2 + edad22 + fumadores + edad2 *
     fumadores, family = poisson, data = coronarias, offset = log(personasalario))
```

Deviance Residuals:

```

      1          2          3          4          5          6          7          8          9         10
 0.43820 -0.27329 -0.15265  0.23393 -0.05700 -0.83049  0.13404  0.64107 -0.41058 -0.01275

```

Coefficients:

```

      Estimate Std. Error z value Pr(>|z|)
(Intercept) -10.79176    0.45008 -23.978 < 2e-16 ***
edad2        2.37648    0.20795  11.428 < 2e-16 ***
edad22       -0.19768    0.02737  -7.223 5.08e-13 ***
fumadores     1.44097    0.37220   3.872 0.000108 ***
edad2:fumadores -0.30755    0.09704  -3.169 0.001528 **
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

```

Null deviance: 935.0673 on 9 degrees of freedom
Residual deviance: 1.6354 on 5 degrees of freedom
AIC: 66.703

```

Number of Fisher Scoring iterations: 4

Por los signos de los parámetros estimados, vemos que la tasa de mortalidad crece con la edad; que es más grande en los fumadores y en el caso de la interacción edad2:fumadores, al ser negativa, ayuda a modelar el hecho de que las diferencias entre fumadores y no fumadores van siendo menos con la edad.

```

2.056824e-05  1.076692e+01  8.206353e-01  4.224800e+00  7.352475e-01

```

También se puede hacer usando el modelo binomial, y queda algo muy semejante, pues para n grandes y probabilidades chicas ambas distribuciones se parecen

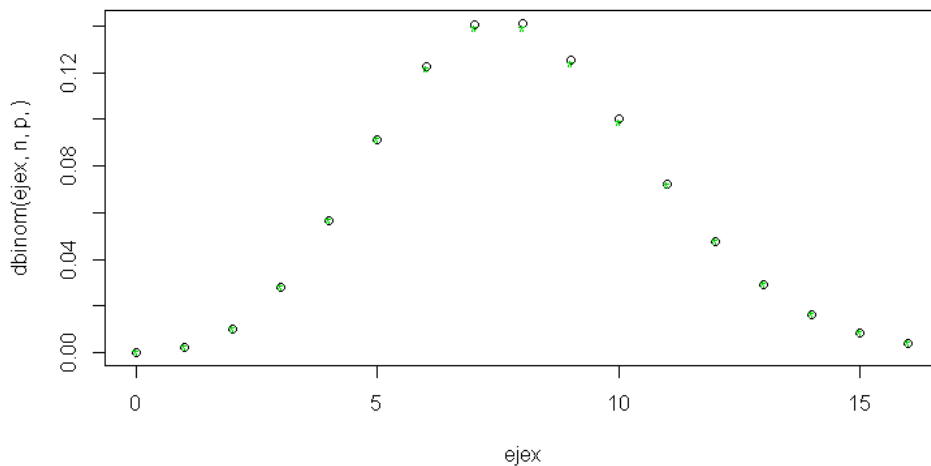
```
props<-muertes/personasano  
modelito2<-glm(props~edad2*fumadores+edad2, family=binomial(link =  
"logit"),weights=personasano,coronarias)  
modelito2  
> modelito2
```

Call: glm(formula = props ~ edad2 * fumadores + edad2, family = binomial(link = "logit"), data = coronarias, weights = personasano)

Coefficients:
(Intercept) edad2 fumadores edad2 edad2:fumadores
-10.7989 2.3789 1.4457 -0.1971 -0.3085

Degrees of Freedom: 9 Total (i.e. Null); 5 Residual
Null Deviance: 939.7
Residual Deviance: 1.641 AIC: 66.63

bastante cercanas n=400 p=0.02



no muy cercanas n=400 p=0.5

