

GLM MODELOS GAUSSIANOS

Los modelos Gaussianos son equivalente a las regresiones.

Lo bueno de todo es que las restricciones son más relajadas y puedes añadir variables cualitativas a las entradas.

Vamos a practicar precisamente este hecho y vamos a añadir 1 variable cualitativa a una regresión lineal múltiple.

¡A por ello!

Tenemos el siguiente modelo:

```
call:
lm(formula = consumo ~ peso + Caballos, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0168690 -0.0049601  0.0007663  0.0040027  0.0142186

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.305e-03  4.094e-03   1.540  0.13435
peso         1.149e-02  1.620e-03   7.089 8.45e-08 ***
Caballos      7.479e-05  2.312e-05   3.235  0.00303 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00664 on 29 degrees of freedom
Multiple R-squared:  0.8471,    Adjusted R-squared:  0.8365
F-statistic: 80.33 on 2 and 29 DF,  p-value: 1.494e-12
```

Escrito en forma de modelo matemático:

$$\text{Consumo} = 6.305e-03 + 1.149e-02 \cdot \text{peso} + 7.479e-05 \cdot \text{Caballos}$$

Queremos ver qué variable cualitativa puede tener influencia en el consumo:

Vamos a probar dos de ellas:

- am = automático, manual
- vs = tipo de motor

MODELO 1 - CALCULA UN MODELO LINEAL CON AM

Calcula el siguiente modelo:

$$\text{Consumo} = b + a1*\text{peso} + a2*\text{Caballos} + a3*\text{AM}$$

Copia el resultado del modelo y los residuos:

```
Call:
glm(formula = consumo ~ wt + hp + am, family = gaussian, data = df)

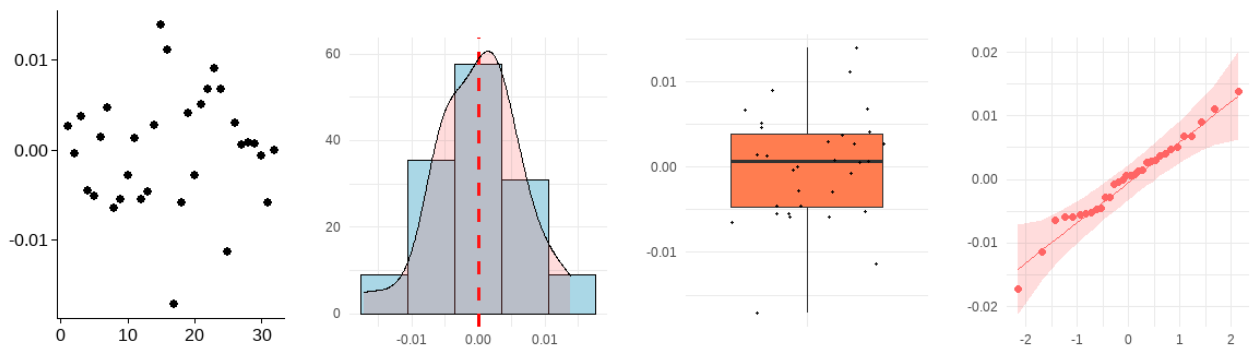
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.164e-03  7.033e-03   0.734  0.46892
wt          1.184e-02  2.408e-03   4.916 3.49e-05 ***
hp          7.277e-05  2.556e-05   2.847  0.00817 **
am          7.377e-04  3.663e-03   0.201  0.84184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4.560272e-05)

Null deviance: 0.0083625  on 31  degrees of freedom
Residual deviance: 0.0012769  on 28  degrees of freedom
AIC: -223.32

Number of Fisher Scoring iterations: 2
```

Comprobacion residuos



Los residuos siguen una distribución normal:

```
Shapiro-Wilk normality test

data:  residuos
W = 0.97918, p-value = 0.7753
```



¿Según los parámetros que ves en la tabla y los p-valores que puedes afirmar?

Se puede afirmar que las variables peso y caballos influyen de forma estadísticamente significativa en el consumo, ya que los valores p son inferiores al nivel de significación ($3,49\text{e-}05$ y $0,00817$, respectivamente).

¿Influye el AM en el consumo?

La variable AM no influye en el consumo ya que el valor p es superior al nivel de significación ($p = 0,84184$).

MODELO 2 - CALCULA UN MODELO LINEAL CON VS

Calcula el siguiente modelo:

$$\text{Consumo} = b + a_1 \cdot \text{peso} + a_2 \cdot \text{Caballos} + a_3 \cdot \text{VS}$$

Copia el resultado del modelo y los residuos:

```
Call:
glm(formula = consumo ~ wt + hp + vs, family = gaussian, data = df)

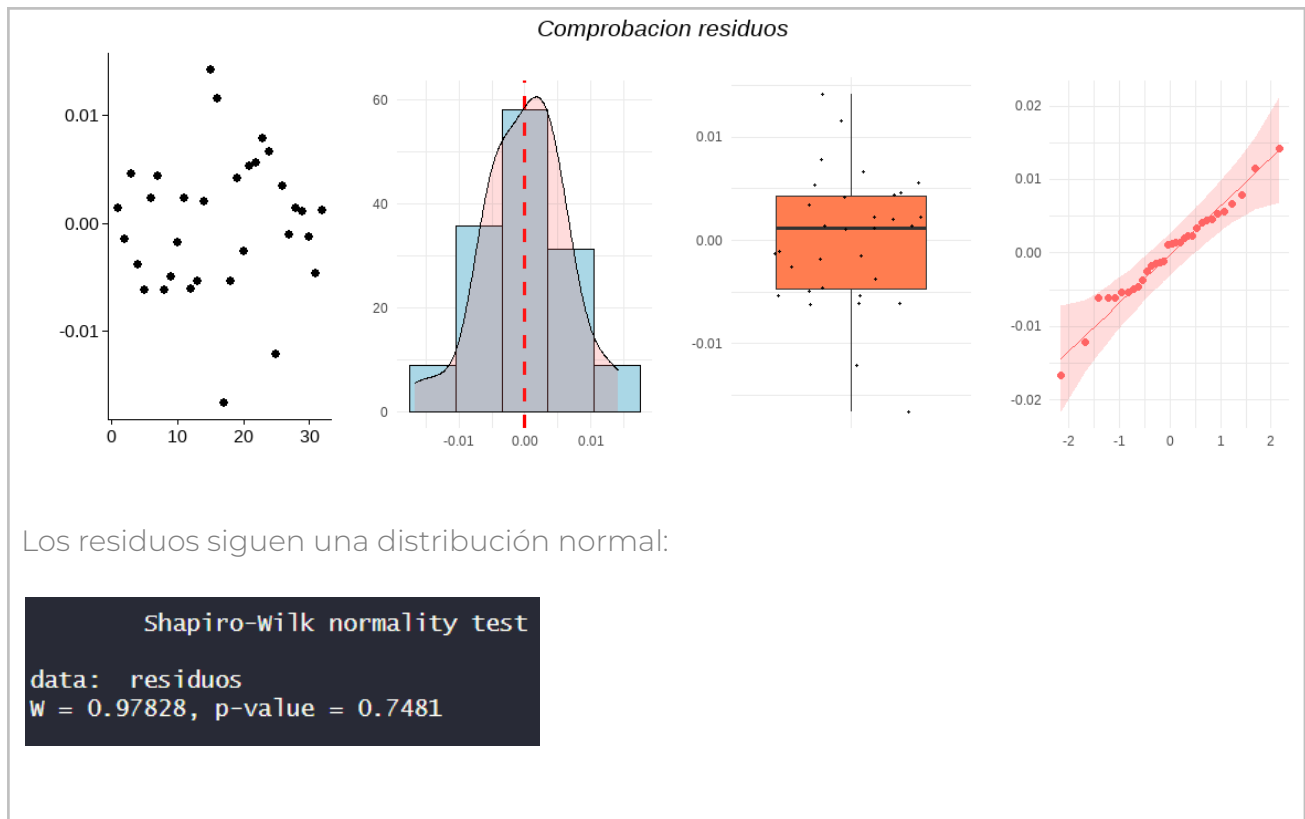
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.554e-03  6.270e-03   1.524  0.1388
wt           1.131e-02  1.654e-03   6.840 1.97e-07 ***
hp           6.361e-05  2.843e-05   2.238  0.0334 *
vs          -2.409e-03  3.497e-03  -0.689  0.4967
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 4.490804e-05)

    Null deviance: 0.0083625  on 31  degrees of freedom
Residual deviance: 0.0012574  on 28  degrees of freedom
AIC: -223.81

Number of Fisher Scoring iterations: 2
```

GLM MODELOS GAUSSIANOS



¿Según los parámetros que ves en la tabla y los p-valores que puedes afirmar?

Se puede afirmar que las variables peso y caballos influyen de forma estadísticamente significativa en el consumo, ya que los valores p son inferiores al nivel de significación ($1,97e-07$ y $0,0334$, respectivamente).

¿Influye el VS en el consumo?

La variable VS no influye en el consumo ya que el valor p es superior al nivel de significación ($p = 0,4967$).

Estamos flexibilizando la capacidad de los modelos de regresión lineal con la ayuda de los modelos lineales generalizados.

¡Vamos a por más!

Adelante