

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:

GLM MODELOS GAUSSIANOS

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

ECUACIÓN DEL MODELO: $\text{mpg} = 36.087 - 2.084 \cdot \text{am}[=0] - 0.037 \cdot \text{hp} - 2.879 \cdot \text{wt}$

Estimaciones de parámetro

Parámetro	B	Desv. Error	95% de intervalo de confianza de Wald		Contraste de hipótesis		
			Inferior	Superior	Chi-cuadrado de Wald	gl	Sig.
(Intersección)	36,087	1,6242	32,903	39,270	493,645	1	,000
[am=0]	-2,084	1,2875	-4,607	,440	2,619	1	,106
[am=1]	0 ^a	.	.	.	.	.	.
hp	-,037	,0090	-,055	-,020	17,399	1	,000
wt	-2,879	,8465	-4,538	-1,219	11,563	1	,001
(Escala)	5,634 ^b	1,4085	3,452	9,197			

Variable dependiente: mpg

Modelo: (Intersección), am, hp, wt

a. Definido en cero porque este parámetro es redundante.

b. Estimación de máxima verosimilitud.

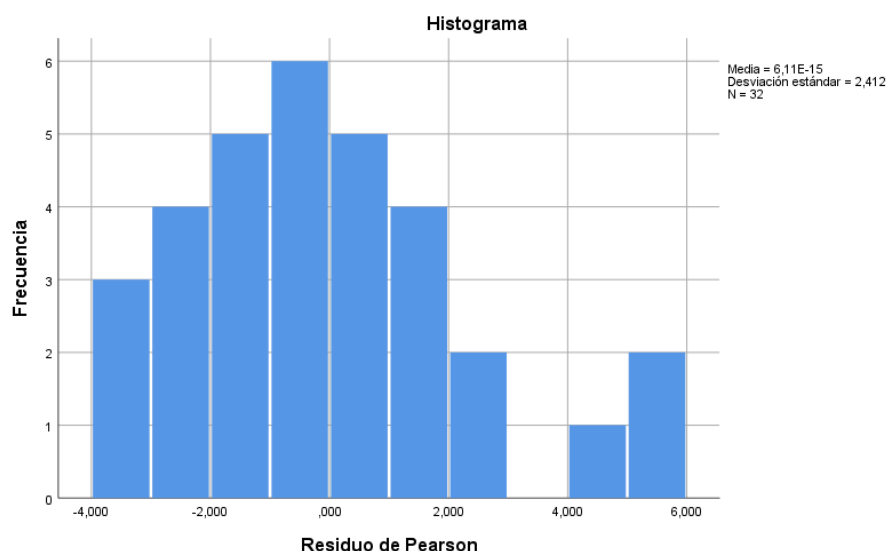
RESIDUOS

Pruebas de normalidad

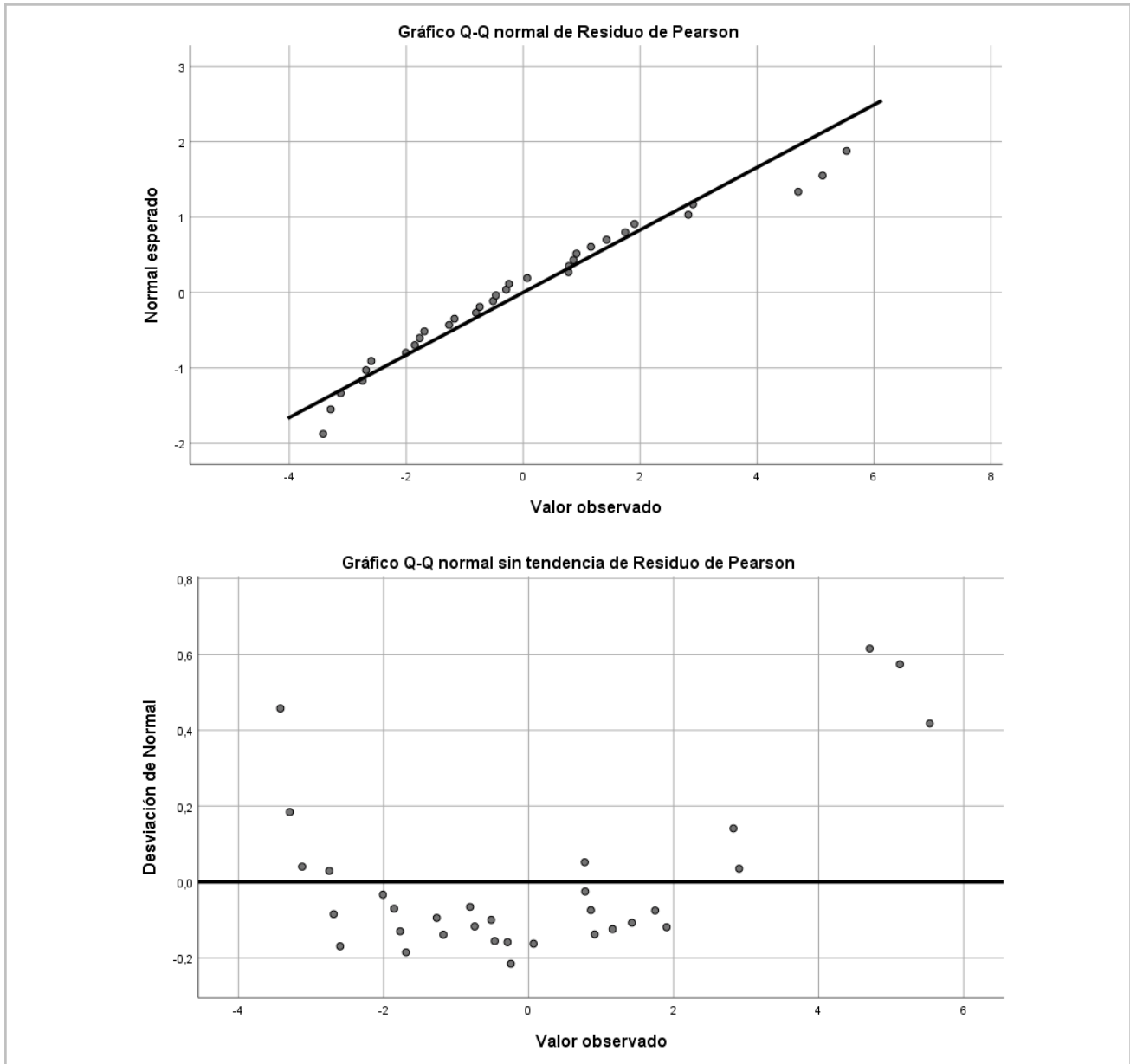
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Residuo de Pearson	,103	32	,200*	,945	32	,106

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors



GLM MODELOS GAUSSIANOS



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

El peso (wt) y los caballos (hp) tienen una relación inversa con la variable de salida (mpg) y su contribución al modelo es estadísticamente significativa. Sin embargo, la variable tipo de transmisión (am) no aporta al modelo, pues no es estadísticamente significativa.

Los residuos tienen media = 0, pero no siguen claramente una distribución normal, como se puede ver en las pruebas de normalidad, el histograma y el Q-Q plot.

¿Influye el AM en el consumo?

No. Su contribución no es estadísticamente significativa.

MODELO 2 - CALCULA UN MODELO LINEAL CON VS

Calcula el siguiente modelo:

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

ECUACIÓN DEL MODELO: $\text{mpg} = 36.750 - 0.025*hp - 3.780*wt - 1.368*vs[=0]$

Estimaciones de parámetro

Parámetro	B	Desv. Error	95% de intervalo de confianza de Wald		Contraste de hipótesis		
			Inferior	Superior	Chi-cuadrado de Wald	gl	Sig.
(Intersección)	36,750	1,5587	33,695	39,805	555,880	1	,000
hp	-,025	,0103	-,046	-,005	6,107	1	,013
wt	-3,780	,5985	-4,953	-2,607	39,886	1	,000
[vs=0]	-1,368	1,2656	-3,848	1,113	1,168	1	,280
[vs=1]	0 ^a	.	.	.	.	.	.
(Escala)	5,881 ^b	1,4702	3,603	9,599			

Variable dependiente: mpg

Modelo: (Intersección), hp, wt, vs

a. Definido en cero porque este parámetro es redundante.

b. Estimación de máxima verosimilitud.

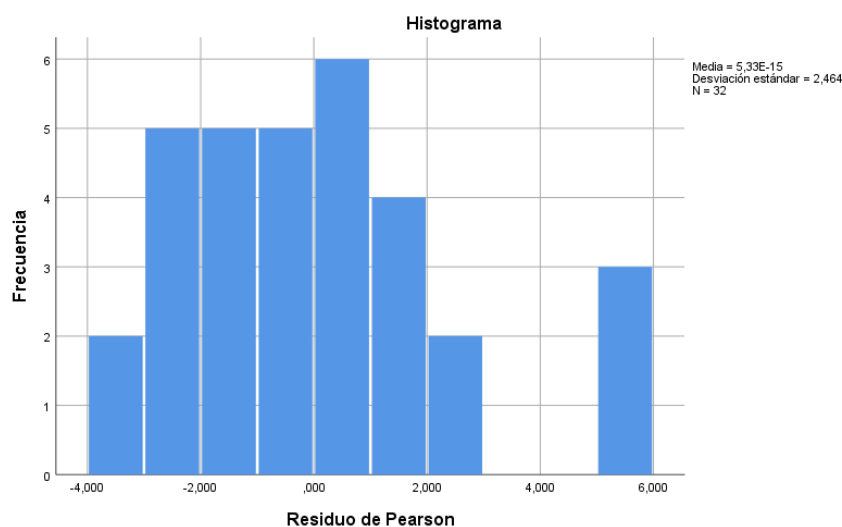
RESIDUOS

Pruebas de normalidad

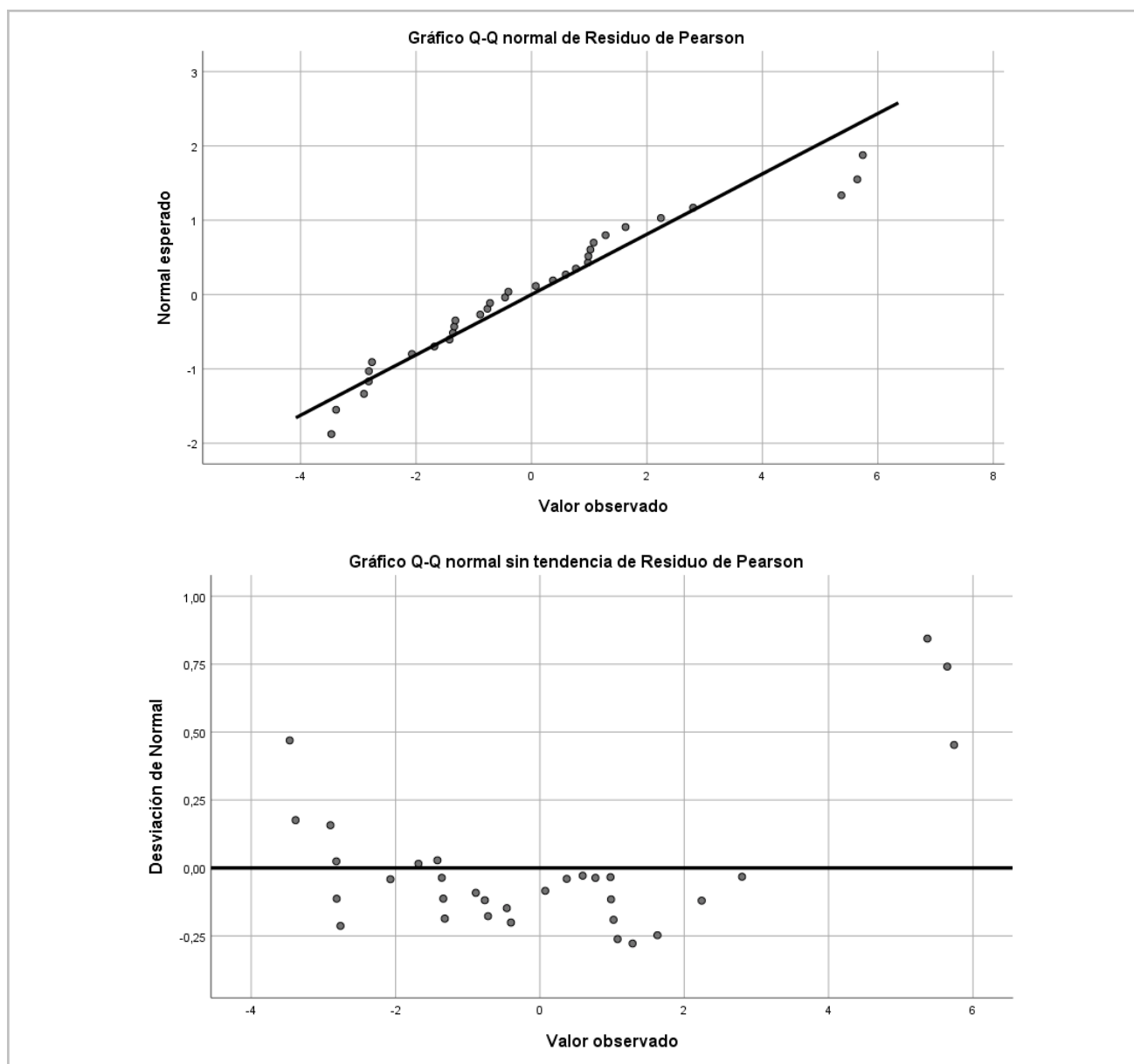
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Residuo de Pearson	,114	32	,200*	,922	32	,024

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors



GLM MODELOS GAUSSIANOS



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

El peso y los caballos tienen una relación inversa con la variable de salida (mpg) y su contribución al modelo es estadísticamente significativa. Sin embargo, la variable tipo de motor no tiene aportación al modelo, pues no es estadísticamente significativa.

Los residuos tienen media = 0, pero no siguen una distribución normal, como se puede ver en las pruebas de normalidad, el histograma y el Q-Q plot.

¿Influye el VS en el consumo?

No. Su contribución no es estadísticamente significativa.



GLM MODELOS GAUSSIANOS

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

¡Vamos a por más!

Adelante