

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA



Ingeniería en Software y Tecnologías Emergentes
Estadística Avanzada

Practica 6 y 7. Regresión Lineal simple, R cuadrada y
normalidad de los residuos

ALUMNO: Fernando Haro Calvo

MATRICULA: 372106

GRUPO: 932

PROFESOR: Juan Iván Nieto Hipólito

27 de septiembre del 2023

Práctica 6 y 7 Regresión Lineal

▼ Clase

Estadística Avanzada

Instrucciones

Dados datos sobre el peso y las emisiones de CO2 de una lista de 36 carros:

- Encontrar el mejor modelo (aprox. $R^2 = |1|$).
- Reportar el código, valor de R^2 obtenido.
- Graficar el histograma de los errores (media y distribución estándar del error).
- Concluir si los errores se aproximan a una distribución normal.

Código

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score

"""
Definir datos
"""
x = [99, 95, 95, 90, 105, 105, 90, 92, 98, 99, 99, 101, 99, 94, 97, 97, 99, 104, 104, 105, 94, 99, 99, 99, 99, 102, 104, 114, 109, 114, 115]
y = [790, 1160, 929, 865, 1140, 929, 1109, 1365, 1112, 1150, 980, 990, 1112, 1252, 1326, 1330, 1365, 1280, 1119, 1328, 1584, 1428, 1365, 14]

plt.figure(1)
plt.scatter(x, y)

"""
Definir modelo
"""
model = np.polyfit(x, y, 9) # Calcular polinomio de grado 9 (mejor valor de R^2)

"""
Calcular el valor de R^2
"""
predict = np.poly1d(model)
r2 = r2_score(y, predict(x)) # Calcular el valor de R^2
print("El valor de R-squared es:", r2)

"""
Graficar modelo
"""
x_axis = range(90, 120)
y_axis = predict(x_axis)
plt.plot(x_axis, y_axis, c="g")

#####
#  ERRORES  #
#####

"""
Calculo de los errores y-ypred
"""
errores = []
for n in y:
    errores = y - predict(x)

# print("Errores: ", errores)

media_error = np.mean(errores)
print("Media de los errores =", media_error)

ds_error = np.std(errores)
print("Desviación estándar de los errores =", ds_error)
```

```

"""
Graficar histograma de los errores
"""
plt.figure(2)
plt.title("HISTOGRAMA DE LOS ERRORES")
plt.xlabel("Errores")
plt.ylabel("Frecuencia")
plt.hist(erroses, bins=9)
plt.show()

```

Salidas

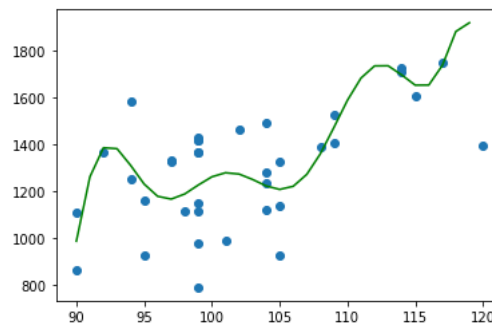
Mejor valor de R^2 :

```

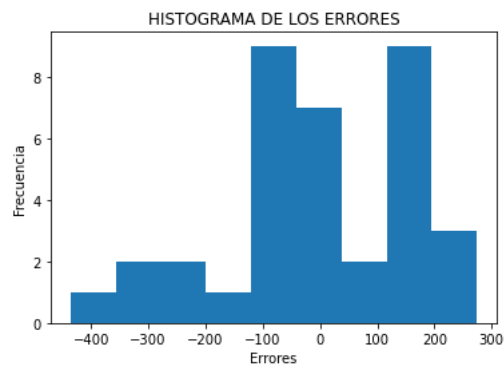
El valor de R-squared es: 0.5143998605096718
Media de los errores = -0.004180908203125
Desviación estándar de los errores = 166.36417649020925

```

Grafica del modelo:



Histograma de errores:



Podemos concluir que los errores no siguen una distribución normal, esto gracias a que tiene 2 grandes ojivas resultando en 2 modas diferentes, una estando en la parte derecha de la gráfica, descartando totalmente la posibilidad de que la distribución sea normal.