

Regresión lineal simple - Guía para ejecutar la actividad

1. Crear un diagrama de dispersión de la variable “temp2” y las variables independientes “edad”, “dilataci” y “expulsiv”.

Cargar la base de datos “Epidural.xlsx” en Spyder:

```
import pandas as pd

data = pd.read_excel("Epidural.xlsx")
```

Para crear un diagrama de dispersión, sigue los siguientes pasos:

Primero, hay que importar las librerías necesarias.

```
import matplotlib.pyplot as plt
import numpy as np
```

En segundo lugar, selecciona las variables para el diagrama de dispersión.

```
x = data['edad']
y = data['TEMP2']
```

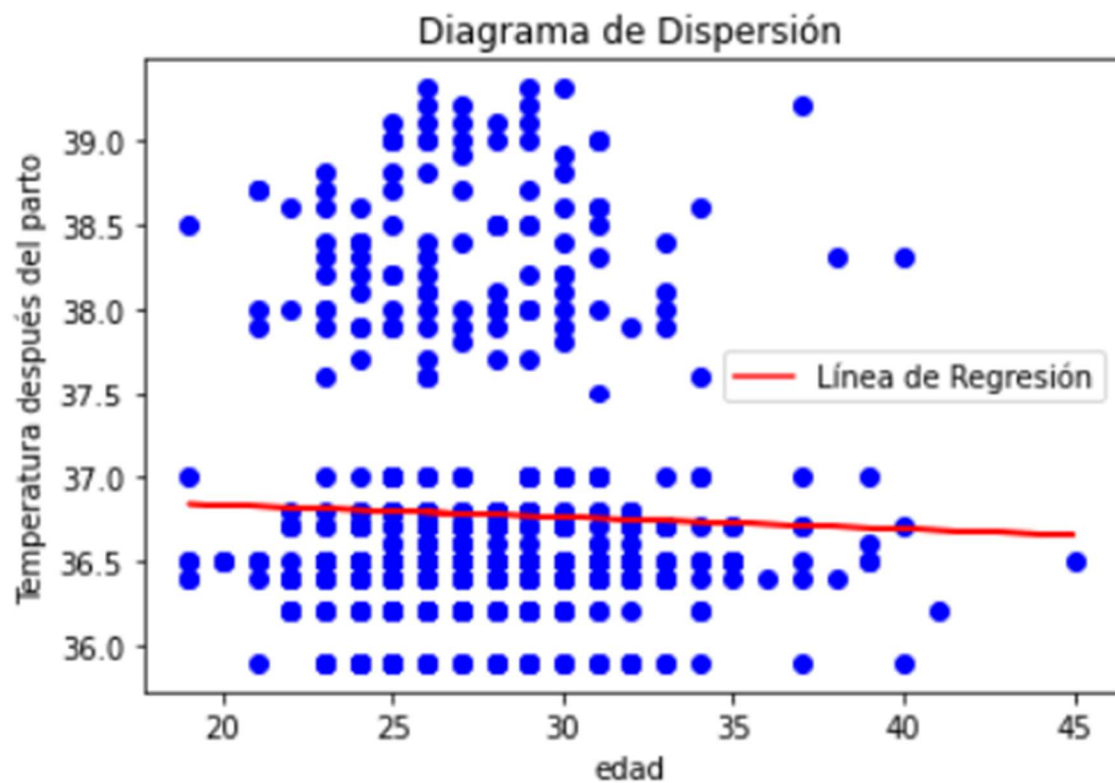
En tercer lugar, calcula la pendiente y la intersección de la línea de regresión. Puedes usar la función polyfit() de numpy para calcularlas.

```
import numpy as np
m, b = np.polyfit(x, y, 1)
```

En cuarto lugar, crea el gráfico de dispersión con la función scatter() y agrega la línea de regresión con la función plot() de matplotlib:

```
plt.scatter(x, y, color='b')
plt.plot(x, m*x + b, color='r', label='Línea de Regresión')
plt.xlabel('edad')
plt.ylabel('Temperatura después del parto')
plt.title('Diagrama de Dispersión')
plt.legend()
plt.show()
```

En la consola superior derecha, si clicas en la pestaña “Plots”, verás este gráfico:



Como podemos observar, la línea de regresión es ligeramente descendente, lo que sugiere una relación negativa entre la variable dependiente (“temp2”) y la independiente (“edad”). No obstante, el ángulo de inclinación de la línea, que indica la fuerza de la relación, apenas está empujada, lo que apunta a que la relación entre ambas variables es muy débil.

En quinto lugar, genera el mismo gráfico con las variables independientes “DILATACI” y “EXPULSIV” e interpreta la fuerza de su relación con la variable dependiente “TEMP2”. Ten en cuenta que los nombres de las variables deben escribirse exactamente igual que en la base de datos que has cargado. Estas, en concreto, están recogidas en mayúsculas.

```

x = data['DILATACI']
y = data['TEMP2']

plt.scatter(x, y, color='b')
plt.plot(x, m*x + b, color='r', label='Línea de Regresión')
plt.xlabel('DILATACI')
plt.ylabel('TEMP2')
plt.title('Diagrama de Dispersión')
plt.legend()
plt.show()

x = data['EXPULSIV']
y = data['TEMP2']

plt.scatter(x, y, color='b')
plt.plot(x, m*x + b, color='r', label='Línea de Regresión')
plt.xlabel('EXPULSIV')
plt.ylabel('TEMP2')
plt.title('Diagrama de Dispersión')
plt.legend()
plt.show()

```

2. Calcular los coeficientes de correlación entre las variables. Se utilizará el coeficiente r de Pearson, que varía entre -1 y 1. Cuanto mayor sea su valor absoluto, mayor será la intensidad de la relación. Determina cuál es la relación más intensa.

Los valores de r cercanos a 1 indican una relación positiva fuerte, mientras que los valores cercanos a -1 indican una relación negativa fuerte. Si r es igual a 0, no hay relación lineal entre las dos variables.

Primero, tienes que calcular la matriz de correlación:

```

variables = ['TEMP2', 'edad', 'DILATACI', 'EXPULSIV']

data = data[variables]

corr_matrix = data.corr()

```

Después, solicita la visualización de la tabla de correlaciones, que aparecerá en la consola inferior derecha. Pega la tabla en tu documento de la actividad.

```

print(corr_matrix)

```

¿Observas una relación significativa entre la variable “TEMP2” y las demás variables independientes?

3. Construir un modelo lineal que cuantifique la relación entre “temp2” y la variable que presente la relación más intensa. Para ello, escribe la ecuación de la recta de regresión.

Inserta el siguiente código en Spyder y ejecútalo:

```
# Seleccionar las variables que se utilizarán para la regresión
x = data['DILATACI']
y = data['TEMP2']

# Calcular los coeficientes de la recta de regresión lineal simple
corr_coef = np.corrcoef(x, y)[0][1] # Coeficiente de correlación
beta1 = corr_coef * (y.std() / x.std()) # Pendiente
beta0 = y.mean() - beta1 * x.mean() # Intercepto

# Imprimir los valores de beta0 y beta1
print('Intercepto: ', beta0)
print('Pendiente: ', beta1)
```

A lo largo del ejercicio, hemos comprobado que no hay una relación significativa entre la variable dependiente “temp2” y la edad, el tiempo de dilatación y la duración del expulsivo. No obstante, para este ejercicio, es conveniente que calcules la ecuación de la recta de regresión con la variable “DILATACI” para practicar.

$y = \beta_0 + \beta_1 x$

- y es la variable dependiente
- x es la variable independiente
- β_0 es el intercepto (el valor de “y” cuando “x” es igual a 0).
- β_1 es la pendiente

Sustituye β_0 y β_1 por los valores que aparecen en la consola inferior derecha.

Envía la actividad a tu tutor en un documento de pdf. En dicho documento, deberás incluir todos los elementos que estimes oportunos: texto explicativo, imágenes, volcados de pantalla, código de Python, etc., para demostrar que has resuelto las cuestiones solicitadas.