

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA



Ingeniería en Software y Tecnologías Emergentes
Estadística Avanzada

Práctica 8 y 9. Regresión Múltiple

ALUMNO: Fernando Haro Calvo

GRUPO: 932

PROFESOR: Juan Iván Nieto Hipólito

04 de octubre del 2023

Práctica 8 y 7 Regresión Múltiple

▼ Clase

Estadística Avanzada

Instrucciones

1. Realiza una búsqueda de un modelo de regresión múltiple que describa el comportamiento de algún sistema.
2. Define la variable independiente y las variables independientes del modelo.
3. Realiza una predicción en base al modelo adquirido.

Datos

Hoja de datos:

[student-mat.csv](#)

Fuente:

<https://vitalflux.com/linear-regression-datasets-csv-excel/>

Procedimiento

Elegí una base de datos con información de 395 estudiantes de Estados Unidos para predecir su rendimiento escolar en base a factores relacionados con su vida personal.

```
# Importamos librerías necesarias
import numpy as np
import pandas as pd
from sklearn import linear_model

datos = pd.read_csv("student-mat.csv")

# Variables independientes
x = datos[['travelttime', # El tiempo de viaje desde casa a la escuela (1-4)
```

```

        'studytime', # El tiempo de estudio semanal del estudiante (1-4)
        'famrel', # La calidad de las relaciones familiares (1-5)
        'freetime', # Tiempo libre fuera de la escuela (1-5)
        'health' # Estado de salud del estudiante (1-5)
    ]]

# Variable dependiente (PROMEDIO DE LOS 3 PERIODOS)
# G1, G2 Y G3 siendo calificaciones parciales (0-20)
y = (datos[('G1')] + datos[('G2')] + datos[('G3')]) / 60 * 100

regr = linear_model.LinearRegression() # Obtiene el modelo de regresion lineal

regr.fit(x, y) # Realiza el calculo de regresion lineal con los parámetros x, y

r2 = regr.score(x, y)
print("Coefficient of determination (r2):", r2)
print("Slope (y):", regr.coef_)
print("Intercept:", regr.intercept_)

```

Resultados:

```

Coefficient of determination (r2): 0.03726500820120815
Slope (y): [-3.04418317  2.63771762  0.38409374  0.40254126 -0.98199685]
Intercept: 53.109043747113546

```

Ya que obtuvimos el modelo de los datos, vamos a predecir el promedio de un estudiante imaginario, con los siguientes datos:

- **1** Tiempo de viaje de casa a la escuela
- **4** Tiempo de estudio semanal
- **5** Calidad de relación familiar
- **5** Tiempo libre fuera de la escuela
- **5** Estado de salud

```

predictedProfit = regr.predict([[1, 4, 5, 5, 5]])
print("\nPromedio General:", round(np.mean(y), 2), "de 100")
print("Promedio predicho del estudiante:", round(predictedProfit[0], 2), "de 100\n")

```

Resultados:

```
Promedio General: 53.4 de 100  
Promedio predicho del estudiante: 59.64 de 100
```