TAREA 4

MÉTODOS DE ANÁLISIS DE DATOS

HAROLD RICARDO RESTREPO SEPÚLVEDA

El objetivo es reproducir en Python el modelo de regresión para el ejercicio visto en clase, para el cual los datos son:

|  |  |  |
| --- | --- | --- |
| quantity sold | price | advertising |
| 8500 | 2 | 2800 |
| 4700 | 5 | 200 |
| 5800 | 3 | 400 |
| 7400 | 2 | 500 |
| 6200 | 5 | 3200 |
| 7300 | 3 | 1800 |
| 5600 | 4 | 900 |

En Excel se tuvo la siguiente salida:



En Python se va a realizar la regresión lineal de dos maneras:

1. Usando Scikit-learn.
2. Usando Statsmodels con el cual podemos también tener los test estadísticos.
3. Usando Scikit-learn

Al ejecutar el script “Modelo\_usando\_Scikit-learn.py” obtenemos el siguiente resultado.

Intercepto: 8536.213882431086

Coeficientes

price -835.722351

advertising 0.592228

Raiz del error medio cuadrático: 234.7340232659235

R2: 0.9806814305770988

En este resultado se puede ver que los valores del intercepto, los coeficientes y el R-cuadrado son los mismos que se tuvieron en el Excel. Sin embargo, no tenemos los valores de los test estadísticos.

1. Usando Statsmodels.

En la siguiente imagen se tiene el resultado del modelo después de ejecutar el script “Modelo\_usando\_Statsmodels.py”. Se puede ver que los valores de los coeficientes son los mismos obtenidos en el modelo realizado en Excel, De igual forma ocurre con los valores de los test estadísticos.

