

Julian Peña Reyes

Análisis de datos

TALLER 4

Elaborar en Python la regresión múltiple para el ejercicio propuesto en clase.

VAR RESPUESTA		
cantidad vendida	price	advertiseing
8500	\$2.00	2800
4700	\$5.00	200
5800	\$3.00	400
7400	\$2.00	500
6200	\$5.00	3200
7300	\$3.00	1800
5600	\$4.00	900

Solución

Creamos nuestro código para agregar los datos en 3 columnas y hacer la regresión multivariada.

```
import pandas as pd
import numpy as np
import sklearn.linear_model as LinearRegression
import matplotlib.pyplot as plt
from sklearn import datasets, linear_model
import statsmodels.api as sm
import statsmodels.stats.diagnostic as smd

Q = np.array([8500,4700,5800,7400,6200,7300,5600])
P = np.array([2,5,3,2,5,3,4])
A = np.array([2800,200,400,500,3200,1800,900])

X_multiple = pd.DataFrame({"P":P,"A": A})

print(X_multiple.describe())

y_multiple = Q
```

```

from sklearn.model_selection import train_test_split
#Separo los datos de "train" en entrenamiento y prueba para probar
los algoritmos
X_train, X_test, y_train, y_test = train_test_split(X_multiple,
y_multiple, test_size=0.2)

#Defino el algoritmo a utilizar
lr_multiple = linear_model.LinearRegression()

#Entreno el modelo
lr_multiple.fit(X_train, y_train)

#Realizo una predicción
Y_pred_multiple = lr_multiple.predict(X_test)

print('DATOS DEL MODELO REGRESIÓN LINEAL MULTIPLE')
print()
print('Valor de las pendientes o coeficientes "a":')
print(lr_multiple.coef_)
print('Valor de la intersección o coeficiente "b":')
print(lr_multiple.intercept_)

print('Precisión del modelo:')
print(lr_multiple.score(X_train, y_train))

X_train = sm.add_constant(X_train, prepend=True)
modelo = sm.OLS(endog=y_train, exog=X_train,)
modelo = modelo.fit()
print(modelo.summary())

```

Resultados

	P	A
count	7.000000	7.000000
mean	3.428571	1400.000000
std	1.272418	1215.181742
min	2.000000	200.000000
25%	2.500000	450.000000
50%	3.000000	900.000000
75%	4.500000	2300.000000
max	5.000000	3200.000000

DATOS DEL MODELO REGRESIÓN LINEAL MULTIPLE

OLS Regression Results

Dep. Variable:	QuantitySold	R-squared:	0.962			
Model:	OLS	Adj. R-squared:	0.943			
Method:	Least Squares	F-statistic:	50.27			
Date:	Sat, 05 Dec 2020	Prob (F-statistic):	0.00146			
Time:	07:12:49	Log-Likelihood:	-48.142			
No. Observations:	7	AIC:	102.3			
Df Residuals:	4	BIC:	102.1			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	8536.2139	386.912	22.062	0.000	7461.975	9610.453
Price	-835.7224	99.653	-8.386	0.001	-1112.404	-559.041
Advertising	0.5922	0.104	5.676	0.005	0.303	0.882
=====						
Omnibus:	nan	Durbin-Watson:	3.358			
Prob(Omnibus):	nan	Jarque-Bera (JB):	0.744			
Skew:	-0.771	Prob(JB):	0.689			
Kurtosis:	2.585	Cond. No.	6.07e+03			
=====						

Acá tenemos la regresión y sus datos como en el Excel.