actividad-5-1

September 27, 2024

Jacob Valdenegro A01640992

Roberto Rillo A01642022

Verifiar los resultados utilizando Python.

[19]:
```
import numpy as np

X = np.array([[1, 2],
              [1, 3],
              [1, 5],
              [1, 7],
              [1, 9]])

# Calcular la matriz HAT
H = X @ np.linalg.inv(X.T @ X) @ X.T

# Valores reales de y
y = np.array([5, 8, 7, 10, 12])

# Valores predichos: ŷ = H * y
y_pred = H @ y

# Coeficientes de regresión:   = (X^T X)^(-1) X^T y
beta = np.linalg.inv(X.T @ X) @ X.T @ y

# Imprimir resultados
print("Matriz HAT (H):\n", H)
print("\nValores predichos (ŷ):\n", y_pred)
print("\nCoeficientes de la regresión ( ):\n", beta)
```

```
Matriz HAT (H):
 [[ 0.51219512  0.41463415  0.2195122   0.02439024 -0.17073171]
 [ 0.41463415  0.34756098  0.21341463  0.07926829 -0.05487805]
 [ 0.2195122   0.21341463  0.20121951  0.18902439  0.17682927]
 [ 0.02439024  0.07926829  0.18902439  0.29878049  0.40853659]
 [-0.17073171 -0.05487805  0.17682927  0.40853659  0.6402439 ]]

Valores predichos (ŷ):
```

```
 [ 5.6097561   6.48170732  8.22560976  9.9695122  11.71341463]

Coeficientes de la regresión ( ):
 [3.86585366 0.87195122]
```

Utilizando los Datos “Cirugía de Hígado” obtener la matriz HAT, los valores predichos, yˆ, y los coeficientes de la ecuación de regresión utilizando el método de matrices.

[ ]:
```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
import numpy as np

# Cargar el archivo Excel
archivo = 'Covarianza, Correlación y HAT.xlsx'
datos = pd.read_excel(archivo)

# Verifica que los datos se cargaron correctamente
print(datos.head())  # Muestra las primeras filas
```

```
   Factor Coagulación  Índice pronóstico  Función de enzima  \
0                 6.7                 62                 81
1                 5.1                 59                 66
2                 7.4                 57                 83
3                 6.5                 73                 41
4                 7.8                 65                115

   Función de hígado  Edad  Género  Alcohol (moderado)  Alcohol (severo)  \
0               2.59    50       0                   1                 0
1               1.70    39       0                   0                 0
2               2.16    55       0                   0                 0
3               2.01    48       0                   0                 0
4               4.30    45       0                   0                 1

   Sobrevivencia (días)
0                   695
1                   403
2                   710
3                   349
4                  2343
```

[ ]:
```
# Variables predictoras: todas menos 'Sobrevivencia (días)'
X = datos.drop(columns=['Sobrevivencia (días)'])

# Variable objetivo
y = datos['Sobrevivencia (días)']

# Escalar las variables predictoras
scaler = MinMaxScaler()
```

```
X_scaled = scaler.fit_transform(X)

# Convertir a matriz NumPy
X_scaled = np.array(X_scaled)
y = np.array(y)
```

[ ]:
```
X_scaled = np.hstack([np.ones((X_scaled.shape[0], 1)), X_scaled])

# Calcular la matriz HAT
XTX_inv = np.linalg.inv(X_scaled.T @ X_scaled)
H = X_scaled @ XTX_inv @ X_scaled.T

# Calcular los coeficientes
beta = XTX_inv @ X_scaled.T @ y

# Calcular los valores predichos
y_pred = X_scaled @ beta

# Valores de apalancamiento (diagonal de la matriz HAT)
apalancamiento = np.diag(H)
```

[ ]:
```
# Imprimir resultados
print("\nCoeficientes de la regresión ( ):")
print(beta)

print("\nValores predichos (ŷ):")
print(y_pred)

print("\nValores de apalancamiento:")
print(apalancamiento)

print("\nMatriz HAT:")
print(H)
```

```
Coeficientes de la regresión ( ):
[-575.86374747  453.26085291  738.23244281  854.14304366  429.43164879
   25.63892773   13.09258075  -41.26764482  195.70703222]

Valores predichos (ŷ):
[ 706.25623722  430.82292124  732.22983195  425.03957611 1454.587552
  317.98512422  561.43434506  620.57029786  814.99238167  761.68391557
  989.9519364   228.02181205 1394.13461057 1050.29172024  937.95668893
  732.48434715  418.09976441   58.32293366  857.99665125  939.03472439
  555.53153245  366.02513011  485.52130235  817.77817446  861.31877768
  669.35808718  592.75876392 1619.88673751  399.48541635  596.03964243
  257.21831085  163.93595245  594.67195134 1143.59947369  512.12738287
```

```
  702.86128551  629.09497899  563.49249113  556.62971999  607.66774824
  423.46168884  569.50576666 1363.25897377  625.40911054  688.74643782
  590.69719298  948.046226   1333.78069235  712.00261732 1065.38071841
  462.70846831  649.64082214  690.30288166  958.05440017  330.98090005
  811.81615789  635.24982009  367.64186795  629.61451147  469.62155034
  154.73023565  369.81392347  403.26975861  902.08295831  600.59893803
  772.29123143  207.7278467   396.43186123 1229.83955625  633.00436345
  424.81983957  475.31740512  993.2862035   197.38922035 1290.08457201
  605.07146825 1006.51638101  817.39664738  616.69071501  333.13872555
  621.92856032  678.01871248  455.39524007  842.81200029  215.61353579
 1044.1331181   629.57478633  526.79553852  902.07905149  722.37107654
  212.15682923  745.77909508  242.52437636  565.21119037  222.12633537
  839.18478601  535.39633923  868.71602558  607.10394711  277.08092447
 1086.53421927  698.14563064 1093.148018    566.51505469  568.33491022
  584.36793863  385.90989128  459.71998145]

Valores de apalancamiento:
[0.03714196 0.06106075 0.09511866 0.07234516 0.11942645 0.08065189
 0.07642418 0.07843142 0.05461582 0.07009281 0.07159644 0.04962392
 0.13226894 0.05700003 0.07881398 0.07780303 0.11185145 0.10082189
 0.05958639 0.08650172 0.06043351 0.11175274 0.12660129 0.03285494
 0.06822654 0.05336978 0.06117105 0.21015924 0.0637799  0.04627201
 0.07765548 0.1382091  0.07017845 0.07616134 0.06096139 0.05026388
 0.09800896 0.2127648  0.03475134 0.08648634 0.05249874 0.14410901
 0.18602117 0.03727933 0.12612543 0.05678346 0.06061771 0.12127331
 0.05897746 0.13267214 0.05612883 0.12188549 0.05172967 0.11945507
 0.08982727 0.06521368 0.15044357 0.08925879 0.09766654 0.1493418
 0.09388399 0.10980073 0.05614709 0.0760336  0.04568778 0.05055969
 0.11489641 0.05145828 0.10340551 0.06096153 0.05972407 0.06740464
 0.09018131 0.10956971 0.11351692 0.09159203 0.05941746 0.05413813
 0.08807509 0.12946688 0.03965555 0.09675421 0.08221149 0.05637406
 0.09542024 0.10265085 0.10503717 0.08818435 0.0813272  0.03353834
 0.04151066 0.08399356 0.05901144 0.08170849 0.10319491 0.05440942
 0.06707476 0.09243314 0.05273941 0.05580399 0.07672453 0.08439488
 0.12997958 0.03064927 0.07615033 0.10349403 0.06246334 0.05864247]

Matriz HAT:
[[ 0.03714196  0.01278513  0.03153001 … -0.0107616   0.03002929
   0.00244981]
 [ 0.01278513  0.06106075  0.04827415 …  0.02342793  0.0143413
   0.03831749]
 [ 0.03153001  0.04827415  0.09511866 …  0.02121118  0.02055136
   0.03079561]
 …
 [-0.0107616   0.02342793  0.02121118 …  0.10349403  0.00417997
   0.03037262]
 [ 0.03002929  0.0143413   0.02055136 …  0.00417997  0.06246334
  -0.00802438]
```

```
[ 0.00244981  0.03831749  0.03079561 …  0.03037262 -0.00802438
  0.05864247]]
```

La suma de los valores de apalancamiento (o leverages) es igual al número de coeficientes de la regresión en un modelo de regresión lineal, estos valores tambien reflejan qué tan influyentes son en la estimación de los coeficientes de regresión.

La ecuación de regresión seria:

$\hat{y}$ = -575.86 * X1 + 453.26 * X2 + 738.23 * X3 + 854.14 * X4 + 429.43 * X5 + 25.64 * X6 + 13.09 * X7 - 41.27 * X8 + 195.71