Matriz HAT

Ejercicio 1:

```python
import numpy as np

X1 = np.array([2, 3, 5, 7, 9])
y = np.array([5, 8, 7, 10, 12])
X = np.column_stack((np.ones(X1.shape[0]), X1))

# Xt * X
XTX = X.T @ X

#Inversa
INV = np.linalg.inv(X.T @ X)

# Matriz HAT
HAT = X @ np.linalg.inv(X.T @ X) @ X.T

# Coeficientes β
beta = np.linalg.inv(X.T @ X) @ X.T @ y

# valores predichos ŷ
y_hat = X @ beta

print("Coeficientes (β):", beta)
print("XT * X:\n", XTX)
print("Inversa:\n", INV)
print("Matriz HAT (H):\n", HAT)
print("Valores predichos (ŷ):", y_hat)
```

```
Coeficientes (β): [3.86585366 0.87195122]
XT * X:
 [[  5.  26.]
 [ 26. 168.]]
Inversa:
 [[ 1.02439024 -0.15853659]
 [-0.15853659  0.0304878 ]]
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]
```

Las predicciones del modelo parecen ser adecuadas, a pesar de que en ciertos casos parece subestimar o sobreestimar valores, en general se encuentra bastante cerca de los valores observados. Como los coeficientes eran altos, y predice correctamente o cercano a, podemos decir que el modelo es bueno con este set de datos.

Ejercicio 2:

```python
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler

df = pd.read_csv('mHat.csv')
print(df.head())

X = df.drop(columns=['Sobrevivencia (días)'])
y = df['Sobrevivencia (días)']

scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

X_scaled = np.c_[np.ones(X_scaled.shape[0]), X_scaled]

X_matrix = np.array(X_scaled)
y_matrix = np.array(y).reshape(-1, 1)

beta = np.linalg.inv(X_matrix.T @ X_matrix) @ X_matrix.T @ y_matrix

# (ŷ = X * β)
y_pred = X_matrix @ beta

# Matriz HAT (H = X (X'X)^-1 X')
HAT_matrix = X_matrix @ np.linalg.inv(X_matrix.T @ X_matrix) @ X_matrix.T

apalancamiento  = np.diag(HAT_matrix)

print("Coeficientes de la ecuación de regresión (β):\n", beta)
print("\nValores predichos (ŷ):\n", y_pred)
print("\nMatriz HAT:\n", HAT_matrix)
print("\nValores de apalancamiento:\n", apalancamiento )
```

Resultado:

	Factor Coagulación	Índice pronóstico	Función de enzima	Función de hígado	Edad	Género	Alcohol (moderado)	Alcohol (severo)	Sobrevivencia (días)
0	6.7	62	81	2.59	50	0	1	0	695
1	5.1	59	66	1.70	39	0	0	0	403
2	7.4	57	83	2.16	55	0	0	0	710

```
3               6.5             73              41              2.01    48      0               0
        0               349
4               7.8             65              115             4.30    45      0               0
        1               2343
Coeficientes de la ecuación de regresión (β):
 [[-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]]

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]]

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]
```

El modelo parece predecir bien la sobrevivencia en días. Hay valores de apalancamiento que nos indican una gran influencia de ciertas variables sobre nuestro modelo, como el “alcohol severo”, la “función de enzima” y “factor de coagulación”, mientras que el género, edad y alcohol moderado parecen no tener un impacto tan significativo. Muchas de estas variables pueden estar conectadas, por lo que se requiere de un análisis más profundo, de la misma manera el alcohol severo presentó un signo positivo inesperado, por lo que tendríamos que checar otras interacciones y colinealidad.

Ecuación (reemplazando con los coeficientes):

Sobrevivencia (días) =−575.86 + 453.26×Factor de Coagulación + 738.23×Índice de pronóstico + 854.14×Función de enzima + 429.43×Función de hígado + 25.64×Edad + 13.09×Género − 41.27×Alcohol (moderado) + 195.71×Alcohol (severo)

$X_1 = 2, 3, 5, 7, 9$ $\quad Y = 5, 8, 7, 10, 12$

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{pmatrix} \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 5 & 26 \\ 26 & 168 \end{pmatrix}$$

$$(X^T X)^{-1} = \frac{1}{5(168) - 26(26)} \begin{pmatrix} 168 & -26 \\ -26 & 5 \end{pmatrix}$$

$$\downarrow$$

$$840 - 676 = 164$$

$$(X^T X)^{-1} = \frac{1}{164} \begin{pmatrix} 168 & -26 \\ -26 & 5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1.024 & -0.159 \\ -0.15 & 0.030 \end{pmatrix}$$

$$X(X^T X)^{-1} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{pmatrix} \begin{pmatrix} 1.024 & -0.159 \\ -0.15 & 0.030 \end{pmatrix}$$

$$X(X^TX)^{-1} = \begin{pmatrix} 0.707 & -0.097 \\ 0.548 & -0.062 \\ 0.232 & -0.006 \\ -0.08 & 0.054 \\ -0.402 & 0.115 \end{pmatrix}$$

$$H = X^TX(X^TX)^{-1} = \begin{pmatrix} 0.707 & -0.097 \\ 0.548 & -0.062 \\ 0.232 & -0.006 \\ -0.08 & 0.054 \\ -0.402 & 0.115 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{pmatrix}$$

$$H = \begin{pmatrix} 0.512 & 0.414 & 0.219 & 0.024 & -0.171 \\ 0.414 & 0.347 & 0.213 & 0.079 & -0.054 \\ 0.219 & 0.213 & 0.201 & 0.189 & 0.177 \\ 0.024 & 0.079 & 0.189 & 0.248 & 0.408 \\ -0.171 & -0.054 & 0.177 & 0.408 & 0.640 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{pmatrix} \begin{pmatrix} 5 \\ 8 \\ 7 \\ 10 \\ 12 \end{pmatrix} = \begin{pmatrix} 42 \\ 247 \end{pmatrix}$$

$$\hat{B} = \begin{pmatrix} 1.024 & -0.159 \\ -0.15 & 0.030 \end{pmatrix} \begin{pmatrix} 42 \\ 247 \end{pmatrix} = \begin{matrix} 3.8711 \\ 0.862 \end{matrix}$$

$$\hat{y} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \\ 1 & 7 \\ 1 & 9 \end{pmatrix} \begin{pmatrix} 3.8711 \\ 0.862 \end{pmatrix} = \begin{matrix} 5.598 \\ 6.46 \\ 8.184 \\ 9.908 \\ 11.632 \end{matrix}$$