

# ACTIVIDAD 3. Regresión Lineal Múltiple

Instituto Tecnológico y de Estudios Superiores de Mty

Grace Aviance Silva Aróstegui A01285158

Concentración TC3006C:

Fecha, 03 de septiembre del 2024.

Inteligencia Artificial Avanzada para Ciencia de Datos

Módulo: Estadística Prof: Ramiro Zermeño Díaz

Campus Guadalajara, Zapopan.

## Dataset: Cirugía de Hígado y Supervivencia

Realizar las transformaciones adecuadas a las variables predictoras.

```
import scipy.stats as stats
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler, StandardScaler

# 0.0s

df = pd.read_csv("DF-Higado.csv")
df.columns

# 0.0s

Index(['Factor Coagulación', 'Índice pronóstico', 'Función de enzima',
       'Función de hígado', 'Edad', 'Género', 'Alcohol (moderado)',
       'Alcohol (severo)', 'Supervivencia (días)'],
      dtype='object')

# Escalamos las variables del dataset

# Variables predictoras
# No escalamos Género, Alcohol Moderado, Alcohol Severo por ser categóricos
# Género: 0 = Mujer, 1 = Hombre
# Alcohol (Moderado): 0 = No, 1 = Sí
# Alcohol (severo): 0 = No, 1 = Sí
x = df[['Factor Coagulación', 'Índice pronóstico', 'Función de enzima',
       'Función de hígado', 'Edad']]
# Variable de respuesta
y = df['Supervivencia (días)']
scaler = MinMaxScaler()
x_scaled = scaler.fit_transform(x)
x_scaled

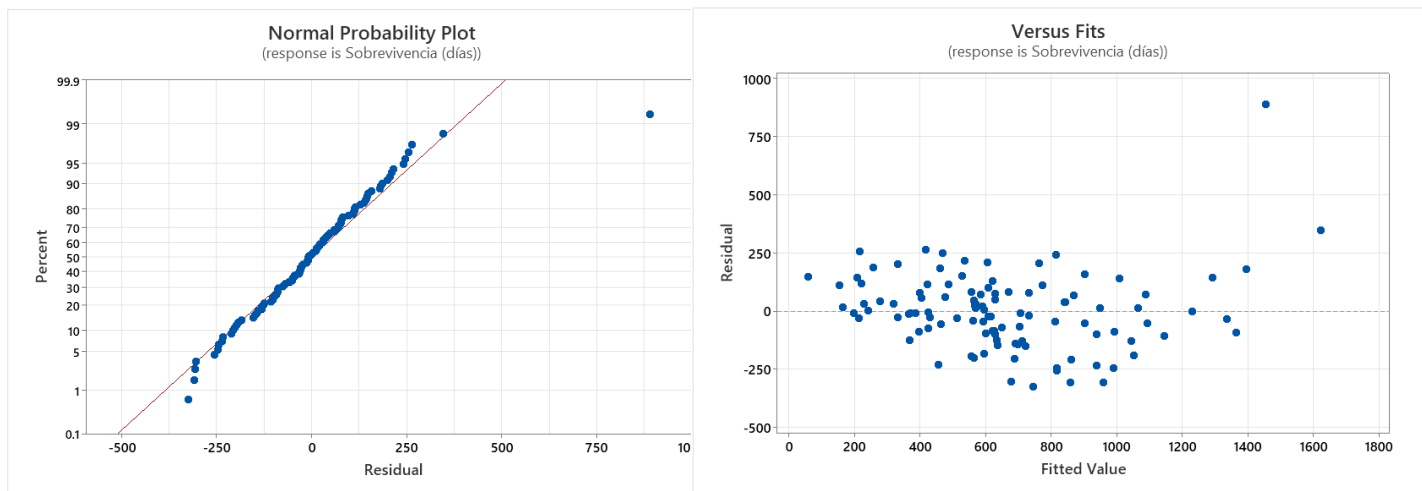
# 0.0s
```

```
... array([[0.47674419, 0.59340659, 0.60416667, 0.32685512, 0.5
],
[0.29069767, 0.56043956, 0.44791667, 0.16961131, 0.225
],
[0.55813953, 0.53846154, 0.625
, 0.25088339, 0.625
],
[0.45348837, 0.71428571, 0.1875
, 0.22438163, 0.45
],
[0.60465116, 0.62637363, 0.95833333, 0.62897527, 0.375
],
[0.37209302, 0.32967033, 0.51041667, 0.12014134, 0.875
],
[0.36046512, 0.41758242, 0.41666667, 0.20671378, 0.475
],
[0.12790698, 0.65934066, 0.60416667, 0.32332155, 0.975
],
[0.39534884, 0.64835165, 0.72916667, 0.31095406, 0.7
],
[0.12790698, 0.74725275, 0.73958333, 0.29328622, 0.45
],
[0.43023256, 0.83516484, 0.625
, 0.59893993, 0.175
],
[0.47674419, 0.47252747, 0.20833333, 0.19787986, 0.675
],
[0.37209302, 0.96703297, 0.94791667, 0.56713781, 0.825
],
[0.37209302, 0.82417582, 0.67708333, 0.56713781, 0.55
],
[0.59302326, 0.59340659, 0.45833333, 0.46996466, 0.7
],
[0.55813953, 0.72527473, 0.46875
, 0.29328622, 0.85
],
[0.39534884, 0.84615385, 0.05208333, 0.39575972, 0.15
],
[0.12790698, 0.47252747, 0.1875
, 0.14310954, 0.225
],
[0.54651163, 0.65934066, 0.53125
, 0.49823322, 0.725
],
[0.34883721, 0.53846154, 0.66666667, 0.40282686, 0.825
],
[0.30232558, 0.48351648, 0.55208333, 0.37279152, 0.225
],
[0.09302326, 0.82417582, 0.3125
, 0.06713781, 0.925
],
[0.47674419, 0.1978022
, 0.46875
, 0.24028269, 0.
],
[0.37209302, 0.64835165, 0.65625
, 0.46996466, 0.475
],
[0.43023256, 0.56043956, 0.80208333, 0.39045936, 0.15
],
...
[0.37209302, 0.68131868, 0.42708333, 0.31448763, 0.475
],
[0.3255814
, 0.61538462, 0.60416667, 0.10954064, 0.8
],
[0.5
, 0.9010989
, 0.10416667, 0.36042403, 0.45
],
[0.61627907, 0.40659341, 0.33333333, 0.30388693, 0.325
],
[0.22093023, 0.65934066, 0.38541667, 0.23498233, 0.725
]])
```

Realizar el modelo de regresión con las variables significativas.

Coefficients					
Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-575.9	97.5	-5.91	0.000	
Factor Coagulación	453	126	3.61	0.000	1.58
Índice pronóstico	738	103	7.15	0.000	1.32
Función de enzima	854.1	97.0	8.80	0.000	1.63
Función de hígado	429	145	2.97	0.004	2.40
Edad	25.6	57.9	0.44	0.659	1.02
Género					
1	13.1	34.1	0.38	0.702	1.07
Alcohol (moderado)					
1	-41.3	38.5	-1.07	0.287	1.36
Alcohol (severo)					
1	195.7	50.1	3.91	0.000	1.44

$y = -575,9 + 453 * F.Cuagulación + 738 * I.Pronostico + 854,1 * F.Enzima + 429 * F.Higado + 25,6 * Edad$



Interpretar la tabla ANOVA, R2, R2 ajustada, p-values y FIV.

### Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-575.9	97.5	-5.91	0.000	
Factor Coagulación	453	126	3.61	0.000	1.58
Índice pronóstico	738	103	7.15	0.000	1.32
Función de enzima	854.1	97.0	8.80	0.000	1.63
Función de hígado	429	145	2.97	0.004	2.40
Edad	25.6	57.9	0.44	0.659	1.02
Género					
1	13.1	34.1	0.38	0.702	1.07
Alcohol (moderado)					
1	-41.3	38.5	-1.07	0.287	1.36
Alcohol (severo)					
1	195.7	50.1	3.91	0.000	1.44

### Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
171.427	77.53%	75.71%	72.05%

Podemos observar los respectivos p-values de cada una de las variables así como FIV. El  $R^2 = 77,53\%$  y  $R^2_{ajustada} = 75.71\%$