Samuel García Berenfeld
A01642317

M1.4 Regresión Lineal Múltiple

Realizar las transformaciones adecuadas a las variables predictoras.

Con los datos proporcionados, lo que hice fue simplemente ponerlos todos (menos la variable a predecir) en una escala de 0 a 1, con el fin de normalizar los datos de tal manera de que todos estén en la misma escala. En mi opinión, esto le da demasiado peso a las variables binarias, ya que 1 es el mayor valor posible y 0 el menor. Pero se me hace la única manera lógica de que todos estén en una escala.

Realizar el modelo de regresión con las variables significativas.

Al hacer la regresión con estos datos obtenemos el modelo

Regression Equation

Sobrevivencia (días) = -575.9 + 453 Factor Coagulación + 738 Índice pronóstico
+ 854.1 Función de enzima + 429 Función de hígado + 25.6 Edad
+ 195.7 Alcohol (severo) - 41.3 Alcohol (Moderado) + 13.1 Género

Podemos observar también nuestra tabla de coeficientes.

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
Alcohol (severo)	195.7	50.1	3.91	0.000	1.44
Alcohol (Moderado)	-41.3	38.5	-1.07	0.287	1.36
Género	13.1	34.1	0.38	0.702	1.07

Probar si se deben agregar interacciones o términos polinomiales.

De lo observado, no tenemos suficiente información para determinar si las interacciones o los términos polinomiales ayudarían al modelo, ya que nuestras variables significativas tienen valores p aceptables. Las variables que parecen no ser importantes para el modelo (género, edad y alcohol moderado) intuitivamente no deberían afectar la sobrevivencia del individuo y las demás parecen ayudar al modelo lo suficiente por si solas.

Samuel García Berenfeld
A01642317

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

Nuestro análisis de la varianza otorga lo siguiente:

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	8	10037033	1254629	42.69	0.000
Factor Coagulación	1	382582	382582	13.02	0.000
Índice pronóstico	1	1502570	1502570	51.13	0.000
Función de enzima	1	2276466	2276466	77.46	0.000
Función de hígado	1	259327	259327	8.82	0.004
Edad	1	5769	5769	0.20	0.659
Alcohol (severo)	1	448520	448520	15.26	0.000
Alcohol (Moderado)	1	33713	33713	1.15	0.287
Género	1	4327	4327	0.15	0.702
Error	99	2909332	29387		
Total	107	12946365			

Aquí podemos observar nuestros valores p. Todos los valores por arriba de 0.05 podrían considerarse como no significativos al explicar nuestro modelo, en este caso se trata de edad, alcohol (moderado) y género. Los demás valores pueden considerarse como significativos para nuestro modelo.

Por otro lado, podemos observar también nuestra tabla de coeficientes:

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
Alcohol (severo)	195.7	50.1	3.91	0.000	1.44
Alcohol (Moderado)	-41.3	38.5	-1.07	0.287	1.36
Género	13.1	34.1	0.38	0.702	1.07

Aquí podemos observar nuestro VIF para cada variable del modelo. En general, estamos buscando valores cercanos a 1, lo que nos indica que las variables no tienen correlación entre sí, por lo que todas son significativas individualmente. Para nuestro modelo, el valor VIF se ve bien para todas las variables, excepto quizás función de hígado. Esto tiene sentido porque es probable que en este modelo, el factor de alcohol de los pacientes se relacione al funcionamiento de este órgano.

Finalmente, nuestro resumen del modelo arroja:

Model Summary

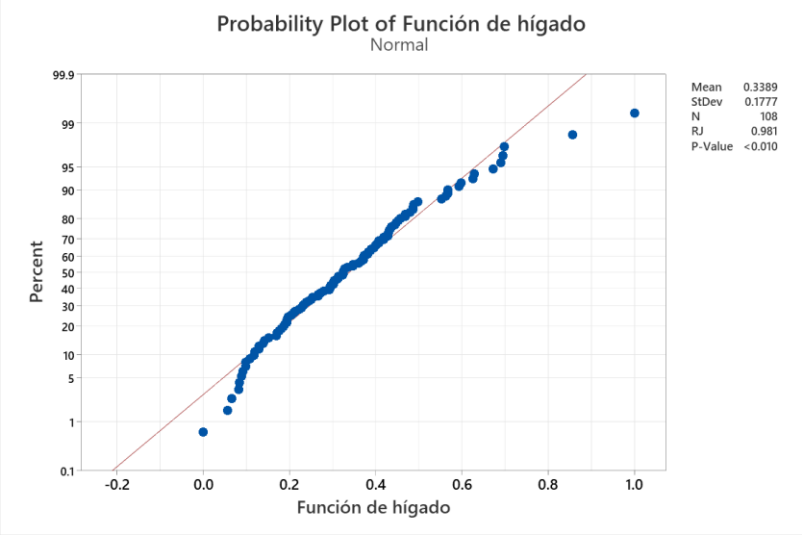
S	R-sq	R-sq(adj)	R-sq(pred)
171.427	77.53%	75.71%	72.05%

Aquí podemos encontrar la r cuadrada y la r cuadrada ajustada. Estos valores nos demuestran si

Verificar el cumplimiento de los supuestos.

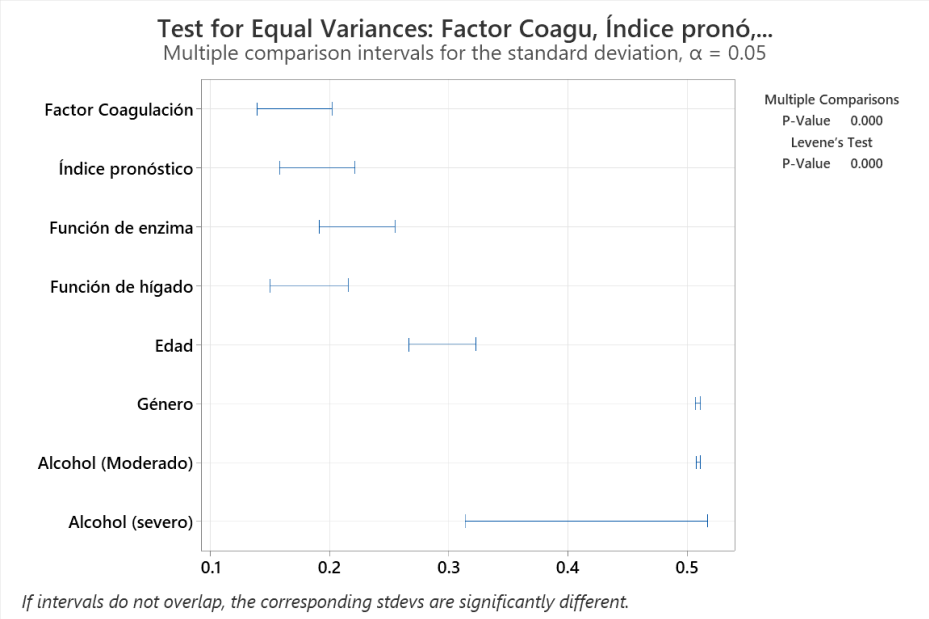
Normalidad – Prueba Shapiro-Wilks:

Para todos los datos se hizo este análisis de normalidad y todas las variables arrojaron el valor de normal. Esto significa que, al ser comparado con una distribución normal, los datos parecen tener una distribución similar. Aquí está un ejemplo obtenido, pero los demás se incluyen en el documento de minitab adjunto.



Samuel García Berenfeld
A01642317

Homocedasticidad – Prueba de Bartlett:



95% Bonferroni Confidence Intervals for Standard Deviations

Sample	N	StDev	CI
Factor Coagulación	108	0.165718	(0.129414, 0.217717)
Índice pronóstico	108	0.184682	(0.151013, 0.231726)
Función de enzima	108	0.217973	(0.185167, 0.263255)
Función de hígado	108	0.177669	(0.140763, 0.230075)
Edad	108	0.289358	(0.259036, 0.331626)
Género	108	0.501986	(0.495806, 0.521446)
Alcohol (Moderado)	108	0.502245	(0.497020, 0.520708)
Alcohol (severo)	108	0.397618	(0.318782, 0.508831)

Individual confidence level = 99.375%

Tests

Method	Test Statistic	P-Value
Multiple comparisons	—	0.000
Levene	26.79	0.000

Independencia – Prueba Durbin-Watson:

Samuel García Berenfeld
A01642317

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.77028

Un valor cercano a 2 me indica que los valores en mi modelo son independientes satisfactoriamente, esto puede ser corroborado con las observaciones anteriores al hablar del VIF.