

EL MODELO LOGÍSTICO Y LA CURVA ROC

Ahora vamos a practicar el modelo logístico y cómo aplicarlo.

En estas situaciones la variable respuesta es una variable cualitativa (normalmente de dos grupos)

Se trata de proporcionar la probabilidad de pertenecer a un grupo u otro con la ayuda de un modelo lineal logístico.

¡Empezamos!

EL OBJETIVO DEL MODELO

Se quiere ver la influencia de distintas variables como la presión sistólica, el peso, la edad con la gravedad del paciente.

En este caso las variables son las siguientes:

- IAH30: pacientes con valores altos de apnea. Gravedad del paciente.
 - 0 = No grave
 - 1 = grave
- TAS_m: presión sistólica
- Sexo: masculino o femenino
- Edad: edad de los pacientes

¿Qué variable es la de salida?

IAH30, que clasifica a los pacientes en graves o no graves.

¿Y las de entrada?

TAS_m, sexo y edad.

¿Según este modelo qué objetivo perseguimos? ¿Qué podremos ver?

El objetivo es determinar si la variable IAH30 está influida por la TAS, sexo y edad, y estimar la probabilidad de que un paciente sea grave en función de estas variables entrada.

MODELO 1 – LA IAH30 CON LA TAS_M

Calcula el modelo logístico con la ayuda de IAH30 en función de la TAS_M.

EL MODELO LOGÍSTICO Y LA CURVA ROC

$$\text{Logist(IAH30)} = b + a * \text{TAS_M}$$

Resumen del modelo

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	196,423 ^a	,128	,171

a. La estimación ha terminado en el número de iteración 4 porque las estimaciones de parámetro han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	3,174	7	,868

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
Paso 1 ^a	TAS_m	,047	,011	18,348	1	,000	1,048	1,026	1,071
	Constante	-6,162	1,478	17,386	1	,000	,002		

a. Variables especificadas en el paso 1: TAS_m.

ECUACIÓN DEL MODELO:
 $\text{Logit(IAH30)} = 0.047 * \text{TAS} - 6.16$

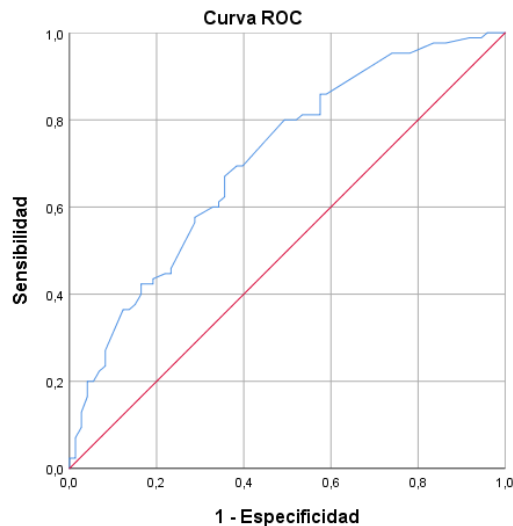
Con los resultados de la tabla, ¿qué puedes afirmar de este modelo?

En este modelo logístico binomial, observamos que la probabilidad de ser un paciente grave aumenta de forma significativa cuando mayor es la tensión arterial sistólica. Esto se explica porque el coeficiente de la TAS_m es positivo (0.074) y es estadísticamente significativo ($p < 0.001$).

CALCULA LA CURVA ROC DEL MODELO

Calcula la curva ROC del modelo anterior y dibújala

EL MODELO LOGÍSTICO Y LA CURVA ROC



Los segmentos de diagonal se generan mediante empates.

¿Qué valor tiene la AUC? (área debajo de la curva ROC)

Área bajo la curva

Variables de resultado de prueba: Probabilidad pronosticada

Área	Desv. Error ^a	Significación asintótica ^b	95% de intervalo de confianza asintótico	
			Límite inferior	Límite superior
,708	,041	,000	,628	,789

Las variables de resultado de prueba: Probabilidad pronosticada tienen, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

a. Bajo el supuesto no paramétrico

b. Hipótesis nula: área verdadera = 0,5

CALCULA Y COMPARA MODELOS

Calcula dos modelos más:

$$\text{Modelo 1 : } \text{Logis(IAH30)} = b + a1 \cdot \text{TAS_M}$$

$$\text{Modelo 2 : } \text{Logis(IAH30)} = b + a1 \cdot \text{TAS_M} + a2 \cdot \text{edad}$$

$$\text{Modelo 3 : } \text{Logis(IAH30)} = b + a1 \cdot \text{TAS_M} + a2 \cdot \text{edad} + a3 \cdot \text{sexo}$$

Copia los valores de estos tres modelos aquí:

EL MODELO LOGÍSTICO Y LA CURVA ROC

MODELO 1

Variables en la ecuación

			Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
		B						Inferior	Superior
Paso 1ª	TAS_m	,047	,011	18,348	1	,000	1,048	1,026	1,071
	Constante	-6,162	1,478	17,386	1	,000	,002		

a. Variables especificadas en el paso 1: TAS_m.

ECUACIÓN DEL MODELO 1:
 $\text{Logit(IAH30)} = 0.047 * \text{TAS} - 6.16$

MODELO 2

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Paso 1ª	TAS_m	,042	,011	14,079	1	,000	1,043	1,020	1,065
	edad	,038	,018	4,434	1	,035	1,038	1,003	1,075
	Constante	-7,100	1,574	20,360	1	,000	,001		

a. Variables especificadas en el paso 1: TAS_m, edad.

ECUACIÓN DEL MODELO 2:
 $\text{Logit(IAH30)} = 0.042 * \text{TAS} + 0.038 * \text{EDAD} - 7.1$

MODELO 3

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP(B)	
								Inferior	Superior
Paso 1 ^a	TAS_m	,040	,011	12,815	1	,000	1,041	1,018	1,064
	edad	,043	,018	5,504	1	,019	1,044	1,007	1,082
	sexo(1)	-,660	,411	2,583	1	,108	,517	,231	1,156
	Constante	-6,664	1,599	17,361	1	,000	,001		

a. Variables especificadas en el paso 1: TAS_m, edad, sexo.

ECUACIÓN DEL MODELO 3:
 $\text{Logit(IAH30)} = 0.040 * \text{TAS} + 0.043 * \text{EDAD} - 0.66 * \text{sexo}[=1] - 6.664$

¿Cuál de ellos es el mejor modelo según el AIC o BIC?

MODELO 1:

Bondad de ajuste^a

EL MODELO LOGÍSTICO Y LA CURVA ROC

	Valor	gl	Valor/gl
Desviación	61,134	53	1,153
Desviación escalada	61,134	53	
Chi-cuadrado de Pearson	50,182	53	,947
Chi-cuadrado de Pearson escalado	50,182	53	
Logaritmo de la verosimilitud ^b	-48,808		
Criterio de información Akaike (AIC)	101,615		
AIC corregido para muestras finitas (AICC)	101,693		
Criterio de información bayesiana (BIC)	107,741		
AIC coherente (CAIC)	109,741		

Variable dependiente: IAH30

Modelo: (Intersección), TAS_m

a. Los criterios de información están en un formato de cuanto más pequeño mejor.

b. La función de logaritmo de la verosimilitud completa se visualiza y utiliza en el cálculo de los criterios de información.

MODELO 2:

Bondad de ajuste^a

	Valor	gl	Valor/gl
Desviación	175,197	145	1,208
Desviación escalada	175,197	145	
Chi-cuadrado de Pearson	143,147	145	,987
Chi-cuadrado de Pearson escalado	143,147	145	
Logaritmo de la verosimilitud ^b	-91,758		
Criterio de información Akaike (AIC)	189,515		
AIC corregido para muestras finitas (AICC)	189,671		
Criterio de información bayesiana (BIC)	198,703		
AIC coherente (CAIC)	201,703		

Variable dependiente: IAH30

Modelo: (Intersección), TAS_m, edad

a. Los criterios de información están en un formato de cuanto más pequeño mejor.

b. La función de logaritmo de la verosimilitud completa se visualiza y utiliza en el cálculo de los criterios de información.

MODELO 3:

Bondad de ajuste^a

	Valor	gl	Valor/gl
Desviación	178,087	147	1,211
Desviación escalada	178,087	147	
Chi-cuadrado de Pearson	149,825	147	1,019
Chi-cuadrado de Pearson escalado	149,825	147	
Logaritmo de la verosimilitud ^b	-91,816		

EL MODELO LOGÍSTICO Y LA CURVA ROC

Criterio de información Akaike (AIC)	191,632		
AIC corregido para muestras finitas (AICC)	191,894		
Criterio de información bayesiana (BIC)	203,883		
AIC coherente (CAIC)	207,883		

Variable dependiente: IAH30

Modelo: (Intersección), TAS_m, edad, sexo

a. Los criterios de información están en un formato de cuanto más pequeño mejor.

b. La función de logaritmo de la verosimilitud completa se visualiza y utiliza en el cálculo de los criterios de información.

El mejor modelo es el 1 porque tiene menor BIC.

Interpreta los resultados del mejor modelo

- Modelo 1. Es el más sencillo de los 3, donde la TAS_m contribuye de forma estadísticamente significativa al modelo. El valor del BIC es 107.741.
- El modelo 2 incluye la edad, que contribuye de forma estadísticamente significativa al modelo ($p = 0.035$). Sin embargo, el valor del BIC es 198.703.
- El modelo 3 incluye el sexo, pero su contribución al modelo no es estadísticamente significativa ($p = 0.108$). El valor del BIC es 203.883.

Teniendo en cuenta todo ello, el mejor modelo sería el 1, pues es el más sencillo y tiene menor BIC.

Has aplicado el modelo logístico y has entendido con un ejemplo la curva ROC.

Ya eres un científico mucho más avanzado en estadística que la mayoría ☺