

Aplicación del Modelo de Regresión Logística

Autor: Lic. Verónica Cuenca Ramallo

Palabra clave: Aplicación del modelo de regresión logística

METODOLOGIA

Para obtener el modelo se considera las siguientes variables:

VARIABLE DEPENDIENTE

¿Si sabe leer y escribir?

0 SI

1 NO

VARIABLE INDEPENDIENTE

¿Cuál es el último nivel y curso más alto de instrucción que aprobó?

Los datos obtenidos para la aplicación del modelo fueron obtenidos de la base de datos del municipio de Tacacoma en la ENCUESTA A HOGARES proporcionado en el INTITUTO IETA (Instituto de Estadística Teórico y aplicado) de la carrera de ESTADISTICA.

La aplicación del modelo se realizó en la propia Institución, teniendo los resultados obtenidos.

EL MODELO

Para realizar la aplicación del modelo de REGRESIÓN LOGISTICA caracterizado de la siguiente forma:

Si tenemos una variable que describe una respuesta en forma dicotómica (Éxito o Fracaso) y se quiere estudiar el efecto con otras variables (Independientes) tienen sobre ella, el modelo de REGRESIÓN LOGISTICA BINARIA puede ser de gran utilidad para lo siguiente:

- Estimar la probabilidad de que se presente el evento de interés, dado los valores de las variables independientes.

- Evaluar la influencia que cada variable independiente tiene sobre la respuesta en forma (ODD RATIO) Una OR mayor que uno indica aumento en la probabilidad del evento y una OR menor que uno, implica una disminución.

El modelo de REGRESIÓN LOGISTICA puede escribirse como:

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

Donde p es la probabilidad de que ocurra el evento del interés, dado el valor de las variables independientes, podemos calcular directamente la estimación de la probabilidad de que ocurra el evento de interés de la siguiente forma:

$$P_{15Ni} = \frac{e^{b_0+b_1P_{15Ni}}}{1 + e^{b_0+b_1P_{15Ni}}}$$

P15: ¿Sabe leer y escribir?

P16: ¿Cuál es el último nivel y curso más alto de instrucción que aprobó?

En este caso también se puede establecer la significancia de los coeficientes usando diferentes estadísticos.

RESULTADOS OBTENIDOS

Para realizar el análisis se consideraron los dos criterios, mencionados anteriormente.

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	576,309	1	,000
	Bloque	576,309	1	,000
	Modelo	576,309	1	,000

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	27,451 ^a	,432	,966

a. La estimación ha finalizado en el número de iteración 11 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Prueba de Hosmer y Lemeshow

Paso	Chi cuadrado	gl	Sig.
1	,512	2	,774

La interpretación de los cuadros:

$H_0: x \simeq \log\left(\frac{p}{1-p}\right)$ si se ajusta al modelo

$H_1: x \neq \log\left(\frac{p}{1-p}\right)$ no se ajusta al modelo

El estadístico HOSMER- LEMESHOW

$$\chi_{HL}^2 = \sum_{K=1}^g \frac{(O_{1K} - E_{1K})^2}{E_{1K}(1 - \varepsilon_K)}$$

El p es dado por :

$Pr\chi^2 \geq \chi_{HL}^2$ donde χ^2 es la CHICUADRADA con $(g - 2)$

Grados de libertad.

Por lo tanto se acepta la hipótesis y el modelo es adecuado.

Tabla de contingencias para la prueba de Hosmer y Lemeshow

		15. ¿Sabe leer y escribir? = Si		15. ¿Sabe leer y escribir? = No		Total
		Observado	Esperado	Observado	Esperado	
Paso 1	1	43	43,000	0	,000	43
	2	462	461,989	0	,011	462
	3	400	399,512	0	,488	400
	4	24	24,499	89	88,501	113

Tabla de clasificación^a

Observado		Pronosticado		Porcentaje correcto
		15. ¿Sabe leer y escribir? Si	No	
Paso 1	15. ¿Sabe leer y escribir? Si	927	2	99,8
	No	3	86	96,6
Porcentaje global				99,5

a. El valor de corte es ,500

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	P16N	-3,943	,615	41,166	1	,000	,019
	Constante	48,499	7,820	38,459	1	,000	1,155E21

a. Variable(s) introducida(s) en el paso 1: P16N.

Matriz de correlaciones

		Constant	P16N
Paso 1	Constant	1,000	-,997
	P16N	-,997	1,000

Entonces:

$$b_0 = 48,499$$

$$b_1 = - 3,943$$

donde.

Y= Sabe leer y escribir

X= Grado de enseñanza

Teniendo los resultados se concluyó que:

$$P_{15Ni} = \frac{e^{48,499 - 3,943 P_{16N}}}{1 + e^{48,499 - 3,943 P_{16N}}}$$

Conclusión

MIENTRAS MENOS GRADO DE INSTRUCCIÓN TENGA MENOR SERA LA PROBABILIDAD QUE SEPA LEER Y ESCRIBIR.

“QUIEN REPITE LAS PALABRAS SIN ENTENDER, ES COMO EL LORO QUE NO SABE LO QUE DICE”

“¡Venturoso aquel a quien el cielo dio un pedazo de pan, sin que le quede obligación de agradecerse a otro que al mismo cielo!”

José Martí