

Análisis de regresión logística

Hasta ahora nos hemos centrado en regresiones cuya variable dependiente (Y) es de escala, sin embargo ¿qué sucede si tenemos una variable dependiente dicotómica? En la práctica de psicología esta situación es muy común (desarrollar o no un trastorno, pasar o repetir de año, colaborar o no colaborar, etc), en estos casos se opta por trabajar con una regresión logística.

La regresión logística no nos permite predecir valores (como si lo hacen los modelos de regresión vistos anteriormente), sino que nos permite evaluar la probabilidad de que se dé un valor específico de Y. Este modelo emplea el método de “Máxima verosimilitud” para la estimación y el contraste de hipótesis. Recibe el nombre de “logística” porque se basa en el logaritmo neperiano $\ln \frac{p}{1-p}$ llamado “Logit”, éste mide de modo continuo como cambian las posibilidades de que se dé el suceso en cuestión.

La interpretación es similar a los casos de regresión antes vistos:

El ajuste del modelo es interpretado a través del R^2 de Nagelkerke, la interpretación de este “psedo R^2 ” es similar a la regresión lineal y no lineal, pero dadas las características de las variable dicotómica, ya no podemos hablar de un porcentaje de la varianza explicada por la variables independientes. La interpretación más aceptada es que hay una verosimilitud de X% del modelo con las VI planteadas que en un modelo en el que no hay ninguna VI.

Para evaluar el sentido del aporte de cada una de las VI del modelo se utilizan los β . En este tipo de modelo es importante tener en cuenta cómo ha sido codificada la variable dependiente (qué significa 0 y qué significa 1) porque, en el caso de un $\beta > 0$ se implica que existe una asociación positiva entre la VI y la VD (a mayor X más posibilidad de 1 en Y), mientras que un $\beta < 0$ implica una relación inversa entre la VI y la VD (a mayor X, menos posibilidad de 1 en Y).

Para evaluar la importancia negativa de cada VI empleamos $\text{Exp}(\beta)$. A más alejado se encuentre ese número de 1, mayor será su importancia. En caso la VI tenga un $\beta < 0$ $\text{Exp}(\beta) < 1$, lo inverso sucederá en el caso de un $\beta > 0$. El $\text{Exp}(\beta)$ nos señala cuál será el incremento en la probabilidad de que ocurra Y por cada punto de incremento en la variable X.

Imaginemos que un grupo de psicólogos está investigando algunos elementos asociados a la conducta pro-social. Para realizar este estudio han optado por un diseño experimental, se fabricó un situación en la que una persona se tropieza y cae en la vereda y se evaluó si este recibía ayuda o no (No ayuda = 0; Ayuda = 1). Se incorporaron en el modelo 3 variables independientes: Atractivo de la persona a recibir la ayuda (X_1), Distancia de la persona que se cae (X_2) y Cantidad de personas alrededor (X_3). Se hipotetizaba que las variables X_1 estaría asociado directamente a la conducta de ayudar, mientras que X_2 y X_3 , tendrían un efecto inverso; a mayor distancia del suceso y mientras más gente haya alrededor, menor será la responsabilidad percibida por el sujeto para ayudar. Tras realizar una serie de ensayos los investigadores obtuvieron un modelo con un R^2 de 0.725 (Existe una verosimilitud de 72.5% en este modelo, en comparación con un modelo sin ninguna variable independiente) y se obtuvieron los siguientes datos en lo concerniente a β y $\text{Exp}(\beta)$:

	β	$\text{Exp}(\beta)$
Atractivo de la persona a recibir la ayuda (X_1),	2.482	11.965
Distancia de la persona que se cae (X_2)	-0.040	0.961
Cantidad de personas alrededor (X_3)	-0.102	0.903

Los resultados son los esperados, X_2 y X_3 , son variables con un $\beta < 0$, mientras que X_1 presenta un $\beta > 0$. Ésta última variable es también la más importante en el modelo. El incremento de una unidad en X_1 aumenta las posibilidades de recibir ayuda 11.965 veces, mientras que cada incremento de X_2 y X_3 disminuye las posibilidades de ayudar en una proporción de 0.961 y 0.903 respectivamente.

Ejercicio de Regresión logística (Regresión logística.sav)

Tomando como punto de partida el creciente aumento en la tasa de suicidio adolescente, un grupo de investigadores japoneses decide investigar las variables psicológicas de riesgo y protección ante el suicidio en adolescentes de 3ro a 5to de secundaria de colegios de varios distritos de la Tokio. Los investigadores tenían particular interés en estudiar la probable conexión directa entre la depresión (medida con el Inventario de Depresión de Beck) y el haber tenido al menos algún intento de suicidio.

Utilice la variable “Depresión” y determine si es que la hipótesis de los investigadores es estadísticamente probable. Al ser una variable dicotómica, los adolescentes sólo podían responder si habían intentado suicidarse o no. Para aquellos casos en los que la variable dependiente es dicotómica debes utilizar una Regresión con un modelo logístico. Estos son los pasos que debes seguir:

1. Para corroborar el funcionamiento global del modelo se utiliza la prueba ómnibus. Revisa la significación y evalúa la calidad del modelo.

```
LOGISTIC REGRESSION VARIABLES Suicidio
/METHOD=ENTER Depresión
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Pruebas omnibus sobre los coeficientes del modelo

		Chi cuadrado	gl	Sig.
Paso 1	Paso	8,858	1	,003
	Bloque	8,858	1	,003
	Modelo	8,858	1	,003

La significación es menor a .05 ($p = .003$). El modelo es, por tanto, significativo.

b) La regresión logística tiene una forma de medir ajuste similar a la de la regresión lineal. Sin embargo, en el caso de la regresión logística los estadísticos que miden ajuste son llamados Pseudo R^2 y no se interpretan exactamente de la misma manera. Para este curso utilizaremos el R cuadrado de Nagelkerke. Utilizando los cuadros de “Ajuste del modelo” responde ¿cuál es el ajuste del Modelo?

Resumen del modelo

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	220,287 ^a	,048	,067

a. La estimación ha finalizado en el número de iteración 4 porque las estimaciones de los parámetros han cambiado en menos de ,001.

Asimismo, podemos concluir que hay una verosimilitud de 6,7% en este modelo en comparación con un modelo en el que no se postula ninguna VI como asociada a la tentativa de suicidio. El modelo es prácticamente irrelevante.

c) Recuerda que una de las funciones de la Regresión es predecir. Extrae la Tabla de clasificación (o Matriz de confusión) y determina qué tan adecuado es el pronóstico resultado del modelo ¿Qué se pronostica mejor? ¿La presencia o ausencia de tentativa de suicidio?

Tabla de clasificación^a

	Observado		Pronosticado		
			Ha intentado suicidarse		Porcentaje correcto
			No	Si	
Paso 1	Ha intentado suicidarse	No	108	12	90,0
		Si	55	5	8,3
	Porcentaje global				62,8

a. El valor de corte es ,500

Se concluye que el modelo clasifica bien el 90% de los casos en los que no hubo intento de suicidio y sólo el 8,3% de casos donde sí hubo. El modelo no es del todo adecuado, al no predecir adecuadamente a los que sí han tenido al menos un intento de suicidio.

d) Para evaluar el “sentido” de cada una de las variables del modelo revisa el Beta y su significación en el cuadro de variables en la ecuación ¿Es significativa la variable identificada? ¿La relación es directa o inversa?

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ^a	Depresión	-,040	,014	8,504	1	,004	,961
	Constante	1,279	,686	3,475	1	,062	3,593

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

Depresión tiene B1 = -,040. Es decir, a mayor depresión, menor probabilidad de haber intentado suicidarse.

e) Finalmente, para evaluar la importancia relativa de la(s) variable(s) significativas utilizarás Exp(B). Este indicador se interpreta de la siguiente manera: Por cada punto adicional en la variable “Depresión” la nueva probabilidad de haber intentado suicidarse (Recuerda que la variable interés es dicotómica, donde 1=No y 2=Si) es igual a Exp(B) veces más que la anterior ¿Cuál es el puntaje Exp (B) en nuestro caso?

En nuestro caso, por cada punto adicional en la variable depresión, la nueva probabilidad de haber intentado suicidarse es de .961.