*E*conomía, XLIV, 48 (julio-diciembre, 2019), pp. 9-35
ISSN 1315-2467, Depósito legal pp: 198702me336

Funciones de enlace alternativas en modelos de respuesta binomial

Alternative link functions in binomial response models

Malinda Coa Ravelo* y Ernesto Ponsot Balaguer**

Códigos JEL: C01, C02, C13, C25
Recibido: 10/02/2019, Revisado: 06/03/2019, Aceptado: 10/04/2019

Resumen
En esta investigación se consideran funciones de enlace alternativas a la función logit en el ajuste de modelos de respuesta binomial. El modelo logit es el más utilizado en el análisis de datos categóricos con respuesta dicotómica; sin embargo, algunas funciones de enlace alternativas pueden ser más apropiadas en aplicaciones concretas. Junto al enlace logit, se exploran en este trabajo los enlaces probit, cauchit, cloglog, loglog, clog y log. Partiendo del modelo binomial saturado, para cada enlace mencionado, se desarrollan las ecuaciones necesarias para el cómputo de los estimadores y de sus varianzas, y se discute su comportamiento en algunas situaciones, sugiriendo al investigador sobre el modelo apropiado a seleccionar de acuerdo a los objetivos que persiga.
Palabras Claves: modelos lineales generalizados, función de enlace, modelos de respuesta binomial, estimación.

Abstract
In this research, alternative link functions to the logit function are considered in the adjustment of binomial response models. The logit model is the most used within the analysis of categorical data with dichotomous response; however, some alternative link functions may be more appropriate in specific applications. Along with the logit link, the probit, cauchit, cloglog, loglog, clog and log links are explored in this research. Starting from the saturated binomial model, for each mentioned link, the necessary equations for the calculation of the estimators and their variances are developed, and their behavior is discussed in some interesting situations, suggesting the researcher concerning the most suitable model to select according to the objectives that are pursued.
Key Words: generalized linear models, link function, binomial response models, estimation.

*MSc en Estadística por la Universidad de Los Andes. Escuela de Estadística. Universidad de Los Andes. Mérida, Venezuela. Correo electrónico: malinda@ula.ve
**PhD en Estadística. Escuela de Ciencias Matemáticas y Computacionales. Universidad de Investigación de Tecnología Experimental Yachay. Imbabura, Ecuador. Correo electrónico: eponsot@yachaytech.edu.ec

1. Introducción

Los modelos lineales generalizados constituyen una ampliación de los modelos lineales que permiten considerar distribuciones no normales de los errores (binomial, Poisson, gamma, entre otras) y varianzas no constantes. Existen variables dependientes, que por su naturaleza (discreta por ejemplo) no pueden ser consideradas normales, o bien violan los supuestos del modelo lineal. Los modelos lineales generalizados ofrecen una alternativa para tratarlos (Nelder y Wedderburn, 1972; McCulloch y Searle, 2000; Dobson, 2002; Agresti, 2015). En este contexto, los modelos lineales generalizados apelan a una función de enlace, la cual se encarga de linealizar dicha relación mediante la transformación de la variable respuesta.

Otra de las utilidades de la función de enlace es la de conseguir que las predicciones del modelo queden acotadas. Es por ello que un aspecto clave para la construcción de un modelo satisfactorio es la escogencia de una función de enlace apropiada, pues esto garantizará que, independientemente de la entrada, el modelo producirá predicciones en el rango adecuado (Li, 2014). Si la forma funcional está mal especificada, entonces las estimaciones de los coeficientes y las inferencias basadas en ellos, pueden ser engañosas.

En los modelos lineales generalizados hay funciones llamadas enlaces canónicos que se aplican por defecto a cada una de las distribuciones (McCullagh y Nelder, 1989), tal como el enlace logit para el modelo binomial; sin embargo, existen otras funciones disponibles que también pueden relacionar las predicciones. A pesar de que las funciones de enlace canónico garantizan máxima información y una interpretación simple de los parámetros de regresión, ellas no siempre garantizan el mejor ajuste a un conjunto de datos dado (Czado y Santner, 1992). Por lo general, la escogencia de la función de enlace es arbitraria, pero si está mal especificada puede permitir un sesgo sustancial en los parámetros de regresión y las estimaciones de las medias de la respuesta. El logit es la función

de enlace canónico para datos de respuesta binomial, pero probit también es popular (McCullagh y Nelder, 1989; Hosmer y Lemeshow, 2000; Collett 2002), entre otras alternativas que también son utilizadas.

El investigador tiende a adoptar el enlace logit para respuestas de datos binomiales, por las ventajas que ofrece al ser la función de enlace canónico y su facilidad para la interpretación de los resultados en la forma de logaritmos de posibilidad (*log odds*) y razones de posibilidad (*odds ratios*). Pero, como se ha dicho, la función de enlace logit no siempre garantiza un buen ajuste para todos los datos de respuesta binomial. Al respecto, Czado y Munk (2000) afirman que, en algunas aplicaciones, el ajuste completo del modelo puede ser mejorado utilizando funciones de enlace no canónicas.

Con base en los argumentos antes señalados y frente un determinado conjunto de datos, surge entonces la pregunta: ¿Cuál función de enlace escoger? Partiendo del modelo binomial saturado, en este trabajo se muestra la derivación matemática-analítica que da una respuesta a esta pregunta. Asi, se obtiene un conjunto de ecuaciones que permiten estimar y evaluar, mediante cálculos y gráficos sencillos, los diversos modelos que surgen al aplicar diferentes funciones de enlace en el ajuste de datos binomiales.

Además de la introducción, el artículo se organiza de la siguiente manera: la sección 2 presenta algunas definiciones y consideraciones sobre los modelos lineales generalizados; en la sección 3, se plantea y describe el modelo binomial saturado, mientras que en la sección 4 se presentan las funciones de enlace más utilizadas. A continuación, en la sección 5, se presenta el desarrollo analítico de los estimadores y varianzas para el modelo binomial saturado general, así como para cada modelo obtenido cuando se utiliza una función de enlace particular; estos resultados constituyen el punto central de este trabajo de investigación. En la sección 6, se presenta un ejemplo de aplicación de lo obtenido en la sección 5, para luego finalizar presentando las conclusiones en la sección 7.

2. Modelo lineal generalizado (MLG)

La unicidad de muchos métodos estadísticos fue demostrada utilizando la idea del Modelo Lineal Generalizado (MLG). Este modelo se define en términos de un conjunto de variables aleatorias independientes $Y_1, Y_2, ..., Y_n$, cada una con distribución perteneciente a la familia exponencial y las siguientes propiedades (Dobson, 2002):
(Dobson, 2002):

- La distribución de cada Y_i tiene forma canónica y depende de un parámetro simple θ_i, llamado parámetro canónico; así,
$$f(y_i;\theta_i) = \exp[y_i\theta_i - b(\theta_i) + c(y_i)].$$
- Todos los Y_i se distribuyen de la misma forma (es decir, todos son normales o todos binomiales, y así).

Entonces, la función de densidad de probabilidad conjunta de $Y_1, Y_2, ..., Y_n$ es

$$f(y_1,\cdots,y_n;\theta_1,\cdots,\theta_n) = \prod_{i=1}^{n} \exp\left[y_i\theta_i - b(\theta_i) + c(y_i)\right]$$
$$= \exp\left[\sum_{i=1}^{n} y_i\,\theta_i - \sum_{i=1}^{n} b\,(\theta_i) + \sum_{i=1}^{n} c\,(y_i)\right].$$

Para el modelo especificado es de interés un conjunto de parámetros $\beta_1, ..., \beta_m$ (donde $m \leq n$), más que los parámetros θ_i. Suponiendo que $E[Y_i] = \mu_i$, donde μ_i es alguna función de θ_i, para el MLG hay una transformación de μ_i tal que

$$\eta_i = g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$$

En esta ecuación:

- g es una función monótona diferenciable, llamada función de enlace.
- x_i es un vector $m \times 1$ de variables explicativas (covariables para niveles del factor),

$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{im} \end{bmatrix} \quad \text{de modo que} \quad \mathbf{x}_i' = [x_{i1} \cdots x_{im}].$$

- El vector x_i' es la *i*-ésima fila de la matriz de diseño **X**.
- $\boldsymbol{\beta}$ es el vector m × 1 de parámetros $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}$.

Así, un MLG consta de tres componentes:

1. Las variables respuestas $Y_1, Y_2, \ldots, Y_n$, de las cuales se asume que comparten la misma distribución proveniente de la familia exponencial. Este componente suele llamarse componente aleatorio.
2. Un conjunto de parámetros $\boldsymbol{\beta}$, y variables explicativas
$$\mathbf{X} = \begin{bmatrix} x_1' \\ \vdots \\ x_n' \end{bmatrix} = \begin{bmatrix} x_{11} \cdots x_{1m} \\ \vdots \quad \ddots \quad \vdots \\ x_{n1} \cdots x_{nm} \end{bmatrix}.$$
Siendo $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, con $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)'$, llamado el predictor lineal o componente sistemático.
3. Una función de enlace *g*, monótona y diferenciable, tal que
$$\eta_i = g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$$
donde $\mu_i = E[Y_i]$. La función de enlace que usa el parámetro canónico como $g(\mu)$, es llamada enlace canónico.

2.1. Estimación máximo verosímil de un MLG

Considere las variables aleatorias independientes $Y_1, Y_2, \ldots, Y_n$ que satisfacen las condiciones de un MLG. Se desea estimar los parámetros $\boldsymbol{\beta}$, tales que $E[Y_i] = \mu_i$ y $g(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta}$. Para cada Y_i, la función log-verosímil es

$$l_i = \log f(y_i; \theta_i) = y_i\theta_i - b(\theta_i) + c(y_i)$$

y la función log-verosímil para todos los Y_i es

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n} l_i = \sum y_i\theta_i - \sum b(\theta_i) + \sum c(y_i).$$

A fin de obtener el estimador máximo verosímil para el parámetro β_j se necesita diferenciar l_i. Puede demostrarse que las ecuaciones normales para un MLG son

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{\mathrm{V}[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = 0, \qquad j = 1,2,\ldots,m,$$

donde $\eta_i = \mathbf{x}_i' \boldsymbol{\beta} = \sum_{j=1}^{m} \beta_j \, x_{ij} = g(\mu_i)$ para la función de enlace g.

Estas ecuaciones tienen la forma $\mathbf{X'DV}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$, donde $\mathbf{V}$ denota la matriz diagonal de varianzas de las observaciones, y $\mathbf{D}$ a la matriz diagonal con elementos $\partial\mu_i/\partial\eta_i$. Aunque $\boldsymbol{\beta}$ no aparece en estas ecuaciones, está implícitamente a través de $\boldsymbol{\mu}$, puesto que $\mu_i = g^{-1}(\sum_{j=1}^{m} \beta_j \, x_{ij})$. Diferentes funciones de enlace producen diferentes conjuntos de ecuaciones. Las ecuaciones de verosimilitud son funciones no lineales de $\boldsymbol{\beta}$ que deben ser resueltas a través de métodos iterativos.

2.2. Distribución asintótica de un MLG

Bajo condiciones de regularidad, para un n grande el estimador máximo verosímil $\hat{\boldsymbol{\beta}}$ de β de un MLG es eficiente y tiene una distribución aproximadamente normal (Agresti, 2015). La matriz de covarianzas de esta distribución es la inversa de la matriz de información de Fisher J, la cual tiene como elementos $\mathrm{E}[-\partial^2 \mathrm{L}(\beta)/\partial\beta_\mathrm{h}\,\partial\beta_\mathrm{j}]$ Es posible comprobar que

$$\mathrm{E}\left[-\frac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_h \partial \beta_j}\right] = \sum_{i=1}^{n} \frac{x_{ih} x_{ij}}{\mathrm{V}[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2.$$

Sea W una matriz diagonal con elementos

$$w_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{V[Y_i]}.$$

Entonces, generalizando a la matriz completa W y con la matriz de diseño X, se tiene que

$$J = \mathbf{X'WX}$$

La forma de W, y por lo tanto de J, dependen de la función de enlace g, dado que $\partial\eta_i/\partial\mu_i = g'(\mu_i)$, entonces la distribución asintótica de $\hat{\boldsymbol{\beta}}$ para un MLG $\boldsymbol{\eta} = \mathrm{g}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$ es

$$\widehat{\boldsymbol{\beta}} \sim \text{AN}[\boldsymbol{\beta}, (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}] \quad [1]$$

donde AN significa Asintóticamente Normal.

3. Modelo de respuesta binomial

Muchas de las variables de interés en el área económica son de naturaleza cualitativa: una persona posee o no posee casa, tiene seguro contra invalidez o no lo tiene, participa en la fuerza laboral o no participa, son algunos entre otros muchos ejemplos. En los modelos donde la variable de interés (Y) es cualitativa, el objetivo es encontrar la probabilidad de que un acontecimiento suceda, tal como poseer una casa, tener un seguro o participar en la fuerza laboral.

Los modelos binomiales son los adecuados para modelar estadísticamente las situaciones antes señaladas. Es por ello que los mismos, en especial los modelos logit y probit, han proliferado en la literatura económica, desde la década de 1980 en adelante, erigiéndose como marco para el estudio de las decisiones de los agentes económicos (Enchautegui, 2000). Este auge, particularmente en microeconomía, surge por la mayor disponibilidad de datos microeconómicos en las últimas décadas aunado a su facilidad de análisis, dados los avances tecnológicos desde entonces.

El modelo binomial más simple es aquel donde se considera solo una variable respuesta (Y) y una variable explicativa (A). En el cuadro 1, considere el factor A con *k* niveles. Sea y_i el número de éxitos observado en el i-ésimo nivel del factor *A* y n_i el número total de observaciones para dicho nivel. Bajo esta disposición de datos agrupados según niveles de la variable explicativa categórica (factor), $n_{\cdot} = \sum n_i$.

Cuadro 1. Número de éxitos y total de respuesta Y versus los niveles del factor A

	Y	
A	Nro. de éxitos	Total
1	y_1	n_1
2	y_2	n_2
⋮	⋮	⋮
$k-1$	y_{k-1}	n_{k-1}
k	y_k	n_k
Total	$y_.$	$n_.$

Fuente: Elaboración propia.

Las respuestas correspondientes a los distintos niveles de A, se asumen independientes entre sí y provenientes de una población binomial en el número de éxitos (Y = 1). Esto es,

$$Y_i \overset{Ind}{\sim} \text{Bin}(y_i; n_i, p_i), \ \forall\ i = 1, \cdots, k-1, k$$

donde Y_i es la variable aleatoria que representa el número de éxitos en la i-ésima población y p_i, considerada constante, es la probabilidad de éxito asociada. Suponer una distribución binomial en el número de éxitos de la variable respuesta Y_i en cada nivel del factor explicativo, implica que $V[Y_i] = n_i p_i (1 - p_i)$ y $E[Y_i] = n_i p_i$.

Es posible utilizar el enfoque de los MLG sobre los datos del cuadro 1 cuando se supone que las variables respuestas $Y_1, \ldots, Y_k$ son independientes y siguen una distribución binomial, pues la misma pertenece a la familia exponencial de distribuciones. En cuanto a la función de enlace, es posible utilizar cualquier función que sea monótona y diferenciable; sin embargo, la escogencia de la misma da lugar a los modelos binomiales más importantes presentes en la literatura (modelo logit, modelo probit, modelo cloglog, por mencionar algunos, y entre los cuales destaca el primero).

3.1. Modelo de respuesta binomial saturado

Entre los posibles modelos que pueden ajustarse para el conjunto de datos del cuadro 1 está el modelo saturado, el cual postula un número de parámetros igual al número de datos con que se cuenta,

reproduciendo exactamente la respuesta observada. Este modelo no da cuenta de la variabilidad de los datos, y es de poca utilidad para su análisis estadístico; sin embargo, a menudo sirve para comparar con otros ajustes del modelo y, dada la simplicidad de sus cálculos, es muy útil a fin de obtener una primera aproximación del comportamiento de los datos. Otro aspecto a considerar en el ajuste de modelos para la tabla de clasificación mencionada es la parametrización de referencia (Ponsot, 2011), en donde el investigador selecciona aquellos niveles de los factores que estarán explícitamente representados en el modelo, proponiendo uno cualquiera de ellos como aquel de referencia.

Así pues, en su acepción más simple, para los datos del cuadro 1 se puede postular un modelo saturado ($m = k$) y la parametrización de referencia, con lo cual la matriz de diseño $\mathbf{X}$ resulta cuadrada (de orden $k \times k$) e invertible. Siendo k el nivel de referencia, dicha parametrización conduce al siguiente modelo:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{k-2} \\ \eta_{k-1} \\ \eta_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{k-2} \\ \beta_{k-1} \\ \beta_k \end{bmatrix}. \qquad [2]$$

Matricialmente, se tiene que $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, lo cual implica que $\boldsymbol{\beta} = \mathbf{X}^{-1}\boldsymbol{\eta}$. Al ser saturado, el modelo planteado en [2] no cuenta con los grados de libertad suficientes para el cálculo de los estadísticos del *deviance* ni de Pearson (Ponsot, 2011). Sin embargo, aún se pueden estimar sus parámetros ($\boldsymbol{\beta}$) y determinar su significación estadística.

4. Funciones de enlace para modelos de respuesta binomial

De entre las funciones de enlace más utilizadas para modelar datos binomiales están (McCullagh y Nelder, 1989; Piegorsch, 1992; Collett, 2002; Hardin y Hilbe, 2007; Hilbe, 2009; Koenker y Yoon, 2009; Tutz, 2011; Hosmer y Cols., 2013):

- Logit o función logística: $g(p_i) = \log[p_i/(1-p_i)]$.
  La función de enlace logit es la inversa de la función de distribución logística, $\mathrm{F}(x) = 1/(1+\exp(-x)) = \exp(x)/(1+\exp(x))$. Logit es la función por defecto (enlace canónico), y la más utilizada en datos binomiales. Un MLG que la utilice como enlace, es llamado modelo de regresión logística o modelo logit.
- Probit o función normal inversa: $g(p_i) = \Phi^{-1}(p_i)$.
  La función de enlace probit utiliza la inversa de la función de distribución normal estándar, $\mathrm{F}(x) = \Phi(x) = (2\pi)^{-1} \int_{-\infty}^{x} \exp(\mathrm{z}^2/2)\mathrm{dz}$. El análisis probit es el método preferido para entender las relaciones dosis-respuesta, siendo ampliamente utilizado en problemas de toxicología y farmacocinética. También tiene uso generalizado en la econometría. Un MLG que utiliza el enlace probit, es llamado modelo de regresión probit o modelo probit.
- Cloglog o función loglog complementario: $g(p_i) = \log[-\log(1-p_i)]$.
  El enlace cloglog es la inversa de la función valor extremo mínimo acumulada o Gompertz (también llamada la distribución Gompertz), $\mathrm{F}(x) = 1 - \exp(-\exp(x))$. Se usa ampliamente para modelar datos de supervivencia, con los que es posible estimar cocientes de riesgo o riesgos relativos. Los MLG que utilizan esta función de enlace, se conocen como modelos de regresión loglog complementario o solo modelos loglog complementario.

Las funciones de enlace logit, probit y cloglog han sido extensivamente utilizadas en una amplia variedad de aplicaciones, en campos tan diversos como la medicina, ingeniería, economía y psicología, entre otros. Otras funciones de enlace encontradas en la literatura, pero con aplicaciones muy particulares son:

- Loglog o función loglog: $g(p_i) = -\log\{-\log(p_i)\}$.
  El enlace loglog es la contraparte de la función loglog complementario, y su función de distribución es $\mathrm{F}(x) = \exp\{-\exp[-(x)]\}$. Se necesita el signo menos adicional en la exponenciación interna para que los coeficientes de los dos modelos loglog tengan los mismos signos. Este enlace es rara vez utilizado porque su

comportamiento es inapropiado para $p < 0.5$ (por lo general, la región de interés). Un MLG que utilice esta función de enlace es llamado modelo de regresión loglog o modelo loglog.

- Clog o función log complementario: $g(p_i) = -\log(1 - p_i)$. Con función de distribución $F(x) = 1 - \exp(-x)$, este modelo puede producir probabilidades menores que cero. Sin embargo, cuando esto ocurre, dichos valores pueden ser truncados en cero obteniéndose una función continua y no decreciente. El modelo ha sido aplicado a una amplia variedad de situaciones experimentales como, por ejemplo, en investigaciones epidemiológicas donde las razones de riesgo son de interés. También, puede resultar en casos de truncamiento de datos, que es cuando los datos politómicos o continuos son truncados en una respuesta dicotómica. El MLG que apela a esta función de enlace, es denominado modelo de regresión clog o modelo clog, simplemente.
- Log o función logarítmica: $g(p_i) = \log(p_i)$. La expresión para la función de distribución de este modelo es $F(x) = \exp(x)$. El enlace log es el enlace canónico en modelos de conteos de Poisson, pero cuando se utiliza como función de enlace en modelos de respuesta binomial se denomina modelo log-binomial, y ha sido propuesto como un enfoque útil para calcular el riesgo relativo. La función de enlace log produce coeficientes que son más fáciles de interpretar que los *odds ratios* que se producen por el modelo de regresión logística. El problema del modelo radica en que puede predecir valores fuera del rango 0 a 1; sin embargo, cuando esto ocurre, los valores son truncados permitiendo la convergencia y obtención de cocientes de riesgo.
- Cauchit o función Cauchy-Lorentz: $g(p_i) = \tan[\pi(p_i - 0.5)]$. El enlace cauchit utiliza la inversa de la función de distribución Cauchy (también conocida como distribución Cauchy-Lorentz), $\mathrm{F}(x) = \tan^{-1}(x)/\pi + 0.5$. Si un MLG utiliza esta función de enlace, es llamado modelo de regresión cauchit o solo modelo cauchit.

Sobre las funciones de enlace

En la figura 1, se representan gráficamente las funciones de enlace más utilizadas en el ajuste de modelos de respuesta binomial. Algunas acotaciones sobre ellas son:

- Logit y probit son las funciones de enlace más utilizadas. Ambas son curvas sigmoidales, simétricas alrededor de $p = 0.5$, y producen resultados similares en muchos análisis a menos que hayan muchas probabilidades pequeñas o grandes. Se requeriría una cantidad de datos muy grande para mostrar que una es mejor que la otra.
- La función logística es esencialmente lineal entre $p = 0.2$ y $p = 0.8$, pero fuera de este rango se vuelve marcadamente no lineal. Las funciones logit y probit están bastante relacionadas linealmente en el intervalo $0.1 \leq p \leq 0.9$.
- La función cauchit también es simétrica en $p = 0.5$ y mantiene relación lineal con logit y probit entre $p = 0.2$ y $p = 0.8$, pero, en comparación con las mismas, tiende al infinito mucho más rápido fuera de este rango.
- Las funciones cloglog y loglog también son sigmoidales pero, a diferencia de logit y probit, no son simétricas con respecto a $p = 0.5$. Es por esto que el uso de estos enlaces está limitado a aquellas situaciones donde es apropiado tratar las probabilidades de éxito p de una manera asimétrica.
- Los enlaces logit y cloglog son similares para valores pequeños de p. Sin embargo, cuando p se aproxima a 1, cloglog tiende mucho más lento al infinito que logit. Entre los enlaces logit y loglog ocurre lo contrario: son similares para valores grandes de p, pero cuando p tiende a 0, loglog tiende más rápido al infinito que logit.
- Por otra parte, clog y log son funciones de enlace muy particulares ya que aplican solo para valores positivos o negativos de la predictora lineal, respectivamente. También, mientras la primera es similar a logit y loglog para valores grandes de p, la segunda lo es a logit y cloglog para valores pequeños de este parámetro.

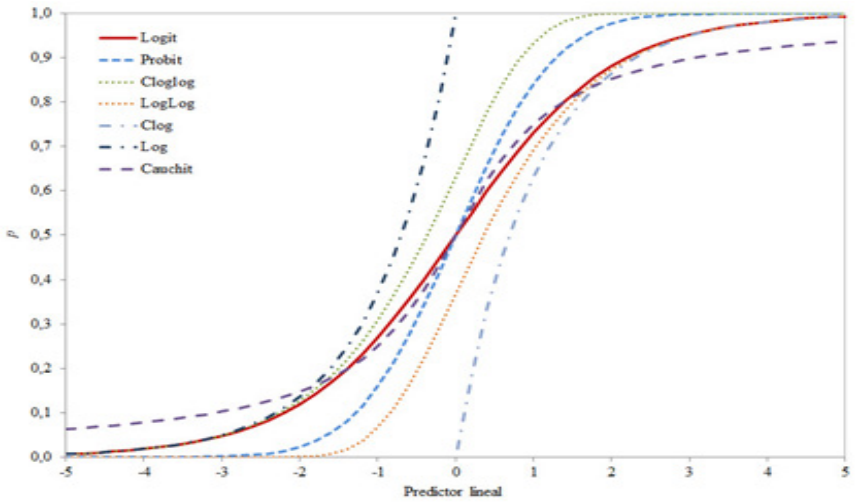


Figura 1. Funciones de enlace.
Fuente: Elaboración propia.

Dadas las características señaladas, los enlaces logit, probit y cauchit son conocidas como funciones de enlace simétricas, mientras que cloglog, loglog, clog y log como funciones de enlace asimétricas. Entre las ventajas de logit con respecto a los otros modelos destacan:

1. Es preferido dada la amplia variedad de estadísticos de ajuste asociados. Además, proporciona interpretación directa, es decir, sus parámetros son fácilmente interpretables en términos de *log-odds* (para β) y *odds* (para $\exp(\beta)$).
2. Si todas las variables son categóricas, la hipótesis $H_0: \boldsymbol{\beta} = \mathbf{0}$ equivale a la declaración de que la variable respuesta y las variables explicativas son independientes.
3. Es el enlace canónico para la distribución binomial, por lo tanto, es conveniente matemáticamente.
4. A pesar de ser muy similares, desde el punto de vista computacional, es más conveniente que probit. Esta última requiere de integración numérica, lo que implica mayores recursos computacionales.

5. Son particularmente apropiados para el análisis de datos que han sido recolectados retrospectivamente, tal como en los casos de estudio-control.

5. Estimadores y varianzas en el modelo binomial saturado

Suponga que se particionan las matrices del modelo [2] de la siguiente manera:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_{k-2} \\ \eta_{k-1} \\ -- \\ \eta_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}_{k-1} \\ \eta_k \end{bmatrix} \quad ; \quad \begin{bmatrix} \beta_1 \\ -- \\ \beta_2 \\ \vdots \\ \beta_{k-2} \\ \beta_{k-1} \\ \beta_k \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \boldsymbol{\beta}_k \end{bmatrix} \quad ; \quad \mathbf{X} = \begin{bmatrix} \mathbf{j} & \mathrm{I} \\ 1 & \mathbf{0}' \end{bmatrix}.$$

Así, dicho modelo puede expresarse como

$$\begin{bmatrix} \boldsymbol{\eta}_{k-1} \\ \eta_k \end{bmatrix} = \begin{bmatrix} \mathbf{j} & \mathrm{I} \\ 1 & \mathbf{0}' \end{bmatrix} \begin{bmatrix} \beta_1 \\ \boldsymbol{\beta}_k \end{bmatrix}$$

y dado que $\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{0}' & 1 \\ \mathrm{I} & -\mathbf{j} \end{bmatrix}$ entonces

$$\begin{bmatrix} \beta_1 \\ \boldsymbol{\beta}_k \end{bmatrix} = \begin{bmatrix} \mathbf{0}' & 1 \\ \mathrm{I} & -\mathbf{j} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_{k-1} \\ \eta_k \end{bmatrix} = \begin{bmatrix} \eta_k \\ \boldsymbol{\eta}_{k-1} - \mathbf{j}\eta_k \end{bmatrix}.$$

Como $\beta_1 = \eta_k$, $\boldsymbol{\beta}_k = \boldsymbol{\eta}_{k-1} - \mathbf{j}\eta_k$, y $\eta_i = g(p_i) = g(\mu_i)$ si $\mu_i = n_i p_i$, en términos generales las estimaciones de los parámetros estarán dadas por

$$\hat{\beta}_j = \begin{cases} g(\hat{p}_k) & \text{si } j = 1 \\ g(\hat{p}_{j-1}) - g(\hat{p}_k) & \text{si } j = 2, \cdots, k-2, k-1, k. \end{cases} \qquad [3]$$

En tanto, la varianza de los parámetros estimados según [1] es

$$\mathrm{V}[\widehat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} = \mathbf{X}^{-1}\mathbf{W}^{-1}(\mathbf{X}')^{-1} = \mathbf{X}^{-1}\mathbf{W}^{-1}(\mathbf{X}^{-1})'$$

donde $\mathbf{W} = \mathrm{diag}\{w_1, w_2, \ldots, w_{k-2}, w_{k-1}, \mathrm{w}_k\}$, con $w_i = \frac{(\partial \mu_i / \partial \eta_i)^2}{\mathrm{V}[Y_i]}$, para $i = 1, 2, \ldots, k-2, k-1, k$. Particionando $\mathbf{W}$ hasta el término w_{k-1}, esto es haciendo $\mathbf{W} = \mathrm{diag}\{\mathbf{w}_{k-1}, w_k\} = \begin{bmatrix} \mathbf{w}_{k-1} & \mathbf{0} \\ \mathbf{0}' & w_k \end{bmatrix}$, se puede determinar $\mathrm{V}[\widehat{\boldsymbol{\beta}}]$ mediante

$$\begin{aligned}\mathrm{V}[\widehat{\boldsymbol{\beta}}] &= \begin{bmatrix}\mathbf{0}' & 1\\ \mathrm{I} & -\mathbf{j}\end{bmatrix}\begin{bmatrix}(\mathbf{w}_{k-1})^{-1} & \mathbf{0}\\ \mathbf{0}' & (w_k)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{0}' & 1\\ \mathrm{I} & -\mathbf{j}\end{bmatrix}'\\ &= \begin{bmatrix}\mathbf{0}' & (w_k)^{-1}\\ (\mathbf{w}_{k-1})^{-1} & -\mathbf{j}(w_k)^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathrm{I}\\ 1 & -\mathbf{j}'\end{bmatrix}\\ &= \begin{bmatrix}(w_k)^{-1} & -\mathbf{j}'(w_k)^{-1}\\ -\mathbf{j}(w_k)^{-1} & (\mathbf{w}_{k-1})^{-1}+\mathbf{J}(w_k)^{-1}\end{bmatrix}.\end{aligned}$$

Como $\mathrm{V}[\hat{\beta}_1] = (w_k)^{-1}$ y $\mathrm{V}[\widehat{\boldsymbol{\beta}}_{k-1}] = \mathrm{diag}\{(\mathbf{w}_{k-1})^{-1} + \mathbf{J}(w_k)^{-1}\}$ la varianza de los estimadores estará dada en términos generales por

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases}(w_k)^{-1} & \text{si } j = 1\\ (w_{j-1})^{-1} + (w_k)^{-1} & \text{si } j = 2,\cdots,k-2,k-1,k.\end{cases} \qquad [4]$$

5.1. Estimadores y varianzas, según la función de enlace

En el cuadro 1, considere $\hat{p}_j = y_j/n_j$ y $\mathrm{V}[Y_j] = n_j\hat{p}_j(1-\hat{p}_j)$, para $j = 1, \ldots, k$. Si n_j es relativamente grande, $\hat{p}_j$ será una estimación razonablemente buena de p_j. A continuación, se desarrollan las estimaciones de los parámetros $\widehat{\boldsymbol{\beta}}$ y sus varianzas $\mathrm{V}[\widehat{\boldsymbol{\beta}}]$ del modelo saturado, y bajo la parametrización por referencia, para las funciones de enlace antes mencionadas, en los modelos de respuesta binomial.

Enlace logit

Sea $\eta_i = g(p_i) = \mathrm{logit}(p_i) = \log[p_i/(1-p_i)] = \log[\mu_i/(n_i-\mu_i)] = g(\mu_i)$, si $\mu_i = n_i p_i$. De [3] se tiene que las estimaciones de los β_j están dadas por

$$\hat{\beta}_j = \begin{cases}\mathrm{logit}(\hat{p}_k) & \text{si } j = 1\\ \mathrm{logit}(\hat{p}_{j-1}) - \mathrm{logit}(\hat{p}_k) & \text{si } j = 2,\cdots,k.\end{cases} \qquad [5]$$

En cuanto a las varianzas de los estimadores, $\partial\eta_i/\partial\mu_i = 1/[n_i p_i(1-p_i)]$, por lo que $\partial\mu_i/\partial\eta_i = n_i p_i(1-p_i) = \mathrm{V}[Y_i]$. Consecuentemente, $w_i = \mathrm{V}[Y_i]$, y por [4] se obtiene que

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases}1/\mathrm{V}[Y_k]) & \text{si } j = 1\\ 1/\mathrm{V}[Y_{j-1}] + 1/\mathrm{V}[Y_k] & \text{si } j = 2,\cdots,k.\end{cases} \qquad [6]$$

Enlace probit

Sea $\eta_i = g(p_i) = \text{probit}(p_i) = \Phi^{-1}(p_i) = \Phi^{-1}(\mu_i/n_i) = g(\mu_i)$ si $\mu_i = n_i p_i$. De [3]:

$$\hat{\beta}_j = \begin{cases} \text{probit}(\hat{p}_k) & \text{si } j = 1 \\ \text{probit}(\hat{p}_{j-1}) - \text{probit}(\hat{p}_k) & \text{si } j = 2, \cdots, k. \end{cases} \quad [7]$$

Para las varianzas de los estimadores, se tiene que $\mu_i = n_i \Phi(\eta_i)$, por lo que $\partial\mu_i/\partial\eta_i = n_i \phi(\eta_i) = n_i \phi[\Phi^{-1}(p_i)]$, y $w_i = \{n_i \phi[\Phi^{-1}(p_i)]\}^2/\text{V}[Y_i]$. Por [4] se tiene entonces que

$$\text{V}[\hat{\beta}_j] = \begin{cases} \text{V}[Y_k]/\{n_k \phi[\Phi^{-1}(\hat{p}_k)]\}^2 & \text{si } j = 1 \\ \text{V}[Y_{j-1}]/\{n_{j-1} \phi[\Phi^{-1}(\hat{p}_{j-1})]\}^2 + & \\ \text{V}[Y_k]/\{n_k \phi[\Phi^{-1}(\hat{p}_k)]\}^2 & \text{si } j = 2, \cdots, k. \end{cases} \quad [8]$$

Enlace cauchit

Sea $\eta_i = g(p_i) = \text{cauchit}(p_i) = \tan[\pi(p_i - 0.5)] = \tan[\pi(\mu_i/n_i - 0.5)] = g(\mu_i)$ si $\mu_i = n_i p_i$. De [3]:

$$\hat{\beta}_j = \begin{cases} \text{cauchit}(\hat{p}_k) & \text{si } j = 1 \\ \text{cauchit}(\hat{p}_{j-1}) - \text{cauchit}(\hat{p}_k) & \text{si } j = 2, \cdots, k. \end{cases} \quad [9]$$

En cuanto a las varianzas de los estimadores, $\mu_i = n_i \tan^{-1}(\eta_i)/\pi + 0.5$, por lo tanto $\partial\mu_i/\partial\eta_i = n_i/[\pi(1 + \eta_i^2)] = n_i/\{\pi\{1 + \tan^2[\pi(p_i - 0.5)]\}\}$ o también $\partial\mu_i/\partial\eta_i = n_i/\{\pi \sec^2[\pi(p_i - 0.5)]\}$. De aquí que $w_i = n_i^2/\{\text{V}[Y_i]\pi^2 \sec^4[\pi(p_i - 0.5)]\}$, y de [4] se obtiene que

$$\text{V}[\hat{\beta}_j] = \begin{cases} \{\text{V}[Y_k]\pi^2 \sec^4[\pi(\hat{p}_k - 0.5)]\}/n_k^2 & \text{si } j = 1 \\ \{\text{V}[Y_{j-1}]\pi^2 \sec^4[\pi(\hat{p}_{j-1} - 0.5)]\}/n_{j-1}^2 + & \\ \{\text{V}[Y_k]\pi^2 \sec^4[\pi(\hat{p}_k - 0.5)]\}/n_k^2 & \text{si } j = 2, \cdots, k. \end{cases} \quad [10]$$

Enlace cloglog

Sea $\eta_i = g(p_i) = \text{cloglog}(p_i) = \log[-\log(1 - p_i)] = \log[-\log(1 - \mu_i/n_i)]$ si $\mu_i = n_i p_i$. De [3]:

$$\hat{\beta}_j = \begin{cases} \text{cloglog}(\hat{p}_k) & \text{si } j = 1 \\ \text{cloglog}(\hat{p}_{j-1}) - \text{cloglog}(\hat{p}_k) & \text{si } j = 2, \cdots, k. \end{cases} \quad [11]$$

Para las varianzas de los estimadores, $\partial\eta_i/\partial\mu_i = -1/[n_i(1-p_i)\log(1-p_i)]$, por lo que $\partial\mu_i/\partial\eta_i = -n_i(1-p_i)\log(1-p_i)$. De aquí que $w_i = [n_i(1-p_i)\log(1-p_i)]^2/\mathrm{V}[Y_i]$, y por [4] entonces

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases} \mathrm{V}[Y_k]/[n_k(1-\hat{p}_k)\log(1-\hat{p}_k)]^2 & \text{si } j=1 \\ \mathrm{V}[Y_{j-1}]/[n_{j-1}(1-\hat{p}_{j-1})\log(1-\hat{p}_{j-1})]^2 + & \\ \mathrm{V}[Y_k]/[n_k(1-\hat{p}_k)\log(1-\hat{p}_k)]^2 & \text{si } j=2,\cdots,k. \end{cases} \quad [12]$$

Enlace loglog

Sea $\eta_i = g(p_i) = \mathrm{loglog}(p_i) = -\log[-\log(p_i)] = -\log[-\log(\mu_i/n_i)]$ si $\mu_i = n_i p_i$. De [3]:

$$\hat{\beta}_j = \begin{cases} \mathrm{loglog}(\hat{p}_k) & \text{si } j=1 \\ \mathrm{loglog}(\hat{p}_{j-1}) - \mathrm{loglog}(\hat{p}_k) & \text{si } j=2,\cdots,k. \end{cases} \quad [13]$$

Para las varianzas de los estimadores, se tiene que $\partial\eta_i/\partial\mu_i = -1/[n_i p_i \log(p_i)]$, por lo que $\partial\mu_i/\partial\eta_i = -n_i p_i \log(p_i)$. De aquí que $w_i = [n_i p_i \log(p_i)]^2/\mathrm{V}[Y_i]$, y por [4] entonces

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases} \mathrm{V}[Y_k]/[n_k\hat{p}_k\log(\hat{p}_k)]^2 & \text{si } j=1 \\ \mathrm{V}[Y_{j-1}]/[n_{j-1}\hat{p}_{j-1}\log(\hat{p}_{j-1})]^2 + & \\ \mathrm{V}[Y_k]/[n_k\hat{p}_k\log(\hat{p}_k)]^2 & \text{si } j=2,\cdots,k. \end{cases} \quad [14]$$

Enlace clog

Sea $\eta_i = g(p_i) = \mathrm{clog}(p_i) = -\log(1-p_i) = -\log(1-\mu_i/n_i) = g(\mu_i)$, si $\mu_i = n_i p_i$. De [3]:

$$\hat{\beta}_j = \begin{cases} \mathrm{clog}(\hat{p}_k) & \text{si } j=1 \\ \mathrm{clog}(\hat{p}_{j-1}) - \mathrm{clog}(\hat{p}_k) & \text{si } j=2,\cdots,k. \end{cases} \quad [15]$$

Para las varianzas de los estimadores, $\partial\eta_i/\partial\mu_i = 1/[n_i(1-p_i)]$ y $\partial\mu_i/\partial\eta_i = n_i(1-p_i)$, por lo que $w_i = [n_i(1-p_i)]^2/\mathrm{V}[Y_i]$, y por [4] se tiene que

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases} \mathrm{V}[Y_k]/[n_k(1-\hat{p}_k)]^2 & \\ \mathrm{V}[Y_{j-1}]/[n_{j-1}(1-\hat{p}_{j-1})]^2 + \mathrm{V}[Y_k]/[n_k(1-\hat{p}_k)]^2 & \end{cases} \quad [16]$$

Enlace log

Sea $\eta_i = g(p_i) = \log(p_i) = \log(\mu_i/n_i) = g(\mu_i)$, De [3]:

$$\hat{\beta}_j = \begin{cases} \log(\hat{p}_k) & \text{si } j = 1 \\ \log(\hat{p}_{j-1}) - \log(\hat{p}_k) & \text{si } j = 2, \cdots, k. \end{cases} \quad [17]$$

En cuanto a las varianzas de los estimadores, $\partial\eta_i/\partial\mu_i = 1/\mu_i = 1/[n_i p_i]$ y $\partial\mu_i/\partial\eta_i = n_i p_i$, por lo que $w_i = (n_i p_i)^2/\mathrm{V}[Y_i]$, y por [4] se tiene que

$$\mathrm{V}[\hat{\beta}_j] = \begin{cases} \mathrm{V}[Y_k]/(n_k\hat{p}_k)^2 & \text{si } j = 1 \\ \mathrm{V}[Y_{j-1}]/(n_{j-1}\hat{p}_{j-1})^2 + \mathrm{V}[Y_k]/(n_k\hat{p}_k)^2 & \text{si } j = 2, \cdots, k. \end{cases} \quad [18]$$

6. Ejemplo

En el cuadro 2, se presenta una situación en que el interés se centra en estudiar la relación entre una variable respuesta Y y un factor explicativo A con tres niveles. Se muestran allí las frecuencias observadas para cada nivel.

Cuadro 2. Ejemplo $Y(0,1)$ versus $A(1,2,3)$

	Y		
A	0	1	Total
1	33	20	350
2	256	94	350
k	178	172	350
Total	764	286	1050

Fuente: Elaboración propia

De [2], el modelo saturado empleando la parametrización con el tercer nivel del factor como referencia es como sigue:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} = \begin{bmatrix} g(p_1) \\ g(p_2) \\ g(p_3) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}.$$

Sean $\hat{p}_1 = 0.0571$, $\hat{p}_2 = 0.2686$ y $\hat{p}_3 = 0.4914$. El cuadro 3 contiene las estimaciones de los parámetros de la predictora lineal y sus varianzas, según la función de enlace empleada, las cuales fueron

calculadas mediante las ecuaciones desarrolladas en la sección 5. El cuadro 3 también contiene las pruebas χ^2 de Wald para $H_0 : \beta_i = 0$, $i = 1,2,3$, con el fin de comprobar la significación estadística de los parámetros estimados.

Cuadro 3. Estimación de β_i y prueba de Wald ($H_0 : \beta_i = 0$), según función de enlace

Enlace	$\hat{\beta}_i$	$V[\hat{\beta}_j]$	χ^2	p-valor	Conclusión
logit	−0.034	0.011	0.1	0.7484	No rechazar
	−2.769	0.065	118.9	< 0.0000	Rechazar
	−0.968	0.026	36.0	< 0.0000	Rechazar
probit	−0.022	0.004	0.1	0.7484	No rechazar
	−1.558	0.016	149.8	< 0.0000	Rechazar
	−0.596	0.001	36.8	< 0.0000	Rechazar
cauchit	−0.027	0.007	0.1	0.7485	No rechazar
	−5.484	1.500	20.0	< 0.0000	Rechazar
	−0.863	0.025	30.0	< 0.0000	Rechazar
cloglog	−0.391	0.006	25.4	< 0.0000	Rechazar
	−2.442	0.056	106.4	< 0.0000	Rechazar
	−0.771	0.017	35.5	< 0.0000	Rechazar
loglog	0.342	0.006	20.0	< 0.0000	Rechazar
	−1.393	0.012	167.2	< 0.0000	Rechazar
	−0.615	0.010	36.6	< 0.0000	Rechazar
clog	0.676	0.003	165.6	< 0.0000	Rechazar
	−0.617	0.003	129.9	< 0.0000	Rechazar
	−0.363	0.004	34.7	< 0.0000	Rechazar
log	−0.710	0.003	170.7	< 0.0000	Rechazar
	−2.152	0.050	92.4	< 0.0000	Rechazar
	−0.604	0.011	34.1	< 0.0000	Rechazar

Fuente: Cálculos propios

Los resultados del cuadro 3 permiten explorar los posibles modelos binomiales y contestar, inicialmente, a la pregunta: *¿cuál función de enlace se ajusta mejor?* Una inspección al cuadro confirma que las funciones de enlace asimétricas cloglog, loglog, clog y log ajustan bien el conjunto de datos, pues los tres parámetros estimados mediante ellas resultan significativos según la prueba de Wald. Lo contrario ocurre para los modelos con funciones de enlace simétricas como logit, probit y cauchit, donde el parámetro $\hat{\beta}_1$ resulta no significativo.

A continuación, cabe considerar la pregunta: *¿cuál función de enlace escoger entre aquellas con buen ajuste?* La respuesta dependerá de los objetivos de la investigación que se lleva a cabo:

- Si lo que se desea es precisión en las estimaciones, el mejor modelo es aquel cuyos estadísticos χ^2 sean mayores pues, en este sentido, aumenta la potencia de la prueba.
- Por el contrario, si lo que se desea es una interpretación simple, entre los modelos con buen ajuste, se escogerá aquel que ofrezca esta cualidad.

Para el ejemplo, si el objetivo es la precisión en las estimaciones, se recomendaría el modelo clog, pues es el que tiene los valores más altos de los estadísticos χ^2 entre los modelos con buen ajuste. Obsérvese que este modelo también es el que presenta varianzas menores, lo cual implica una mayor precisión en las estimaciones. Por el contrario, si el objetivo es la interpretación, el modelo log sería el escogido, pues es el que mejor detenta esta característica dada su simplicidad. La situación ideal se presenta cuando un modelo ofrece precisión en las estimaciones y, además, facilidad en su interpretación. Sin embargo, en el ejemplo, ninguno ofrece ambas cualidades.

Otro aspecto relevante de las ecuaciones obtenidas en la sección 5, es la posibilidad de representar gráficamente las varianzas de los estimadores a fin de evaluar su desempeño en distintos escenarios.

Así, en la figura 2, se ilustra el comportamiento de la $V[\hat{\beta}_j]$ según las ecuaciones [6], [8] y [10], correspondientes a las funciones de enlace simétricas logit, probit y cauchit. Fijando el tamaño de muestra, n_j, en cada nivel del factor ($n_j = 350$ para $j = 1, \ldots, k$, con $k = 3$, como en el ejemplo), en las figuras 2(a) y 2(b), se despliega la $V[\hat{\beta}_1]$ cuando la función de enlace es logit. Específicamente, cuando $j = 1$, la $V[\hat{\beta}_j]$ es simétrica en $\hat{p}_k = 0.5$, manteniéndose pequeña y casi constante en el intervalo $0.2 < \hat{p}_k < 0.8$ y aumentando a medida que $\hat{p}_k \to 0$ o $\hat{p}_k \to 1$ tal como se evidencia en la figura 2(a). Por otra parte, cuando $j = 2, \ldots, k$, la $V[\hat{\beta}_j]$ tiene un comportamiento como el que se muestra en la figura 2(b), donde se observa que las mismas se mantienen pequeñas, siempre que $0.2 < (\hat{p}_{j-1}, \hat{p}_k) < 0.8$. Fuera de este intervalo, la $V[\hat{\beta}_1]$ tiende al aumento, manteniendo un comportamiento similar al de $V[\hat{\beta}_1]$ pero replicado en dos dimensiones (para $\hat{p}_{j-1}$ y $\hat{p}_k$).

Las figuras 2(c) y 2(d) ejemplifican la $V[\hat{\beta}_j]$ cuando en el modelado se apela a una función de enlace probit, donde se observa un comportamiento similar en estas varianzas como cuando se utiliza una función de enlace logit: simetría en $\hat{p}_j = 0.5$, con poca variación entre $0.2 < \hat{p}_j < 0.8$ y creciente cuando $\hat{p}_j$ tiende a 0 o a 1, para todo $j = 1, \ldots, k$. Al caso, solo es posible argumentar dicha similitud en el comportamiento de las varianzas, más no es posible compararlas directamente pues las funciones de enlace que las originan provienen de distribuciones con medias y varianzas distintas (Tutz, 2011). Por último, las figuras 2(e) y 2(f) presentan la $V[\hat{\beta}_j]$ cuando se utiliza una función de enlace cauchit. Se observa que, estructuralmente, son similares a las obtenidas con logit y probit, pero crecen muy rápido a medida que p_j se acerca a 0 o 1, o a ambos.

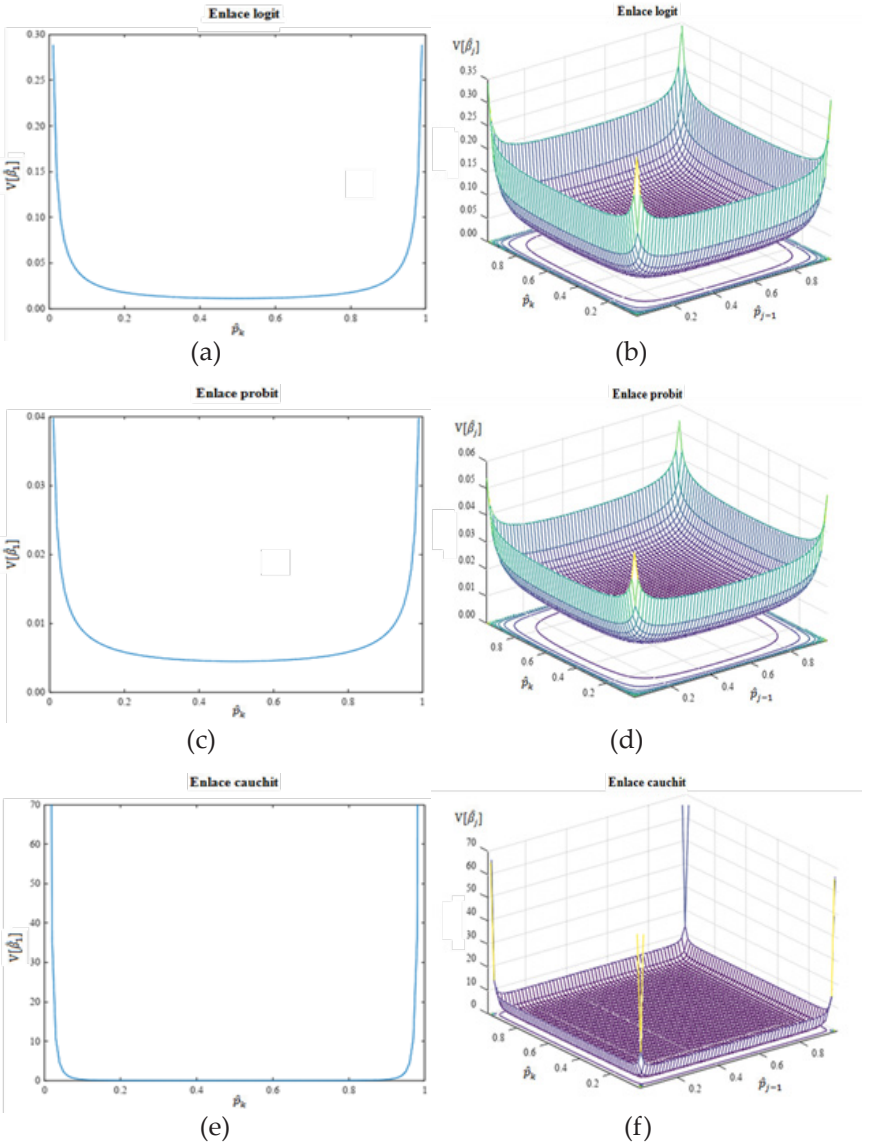
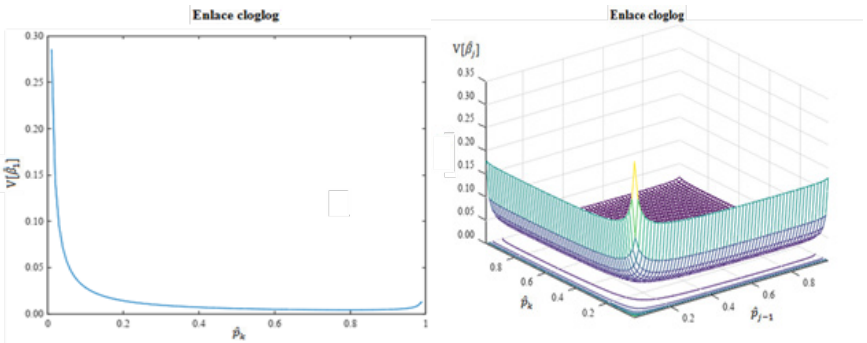


Figura 2. $V[\hat{\beta}_j]$ según funciones de enlace simétricas logit, probit y cauchit (con $nj = 350$, para $j = 1, \ldots, k$).
Fuente: Elaboración propia.

De igual manera, en la figura 3 se despliega la $\mathrm{V}[\hat{\beta}_j]$ cuando se modelan los datos mediante las funciones cloglog (figuras 3(a) y 3(b), ecuación [12]) y su contraparte loglog (figuras 3(c) y 3(d), ecuación [14]). Manteniendo fijo $n_j = 350$, como en los gráficos anteriores, acá se evidencian las características asimétricas propias de las distribuciones de procedencia de estas funciones de enlace. Para el caso, las varianzas de los estimadores cuando se modela con cloglog son muy similares a las obtenidas mediante el enlace logit, para valores de $p_j < 0.5$; en tanto, para valores de $p_j > 0.5$, la misma tiende a cero, para luego crecer lentamente cuando $p_j \rightarrow 1$. En contraste, ocurre lo contrario cuando se modela con el enlace loglog: las varianzas de los estimadores se hacen similares a las del enlace logit cuando $p_j > 0.5$, tienden a cero para $p_j < 0.5$, y luego presentan un crecimiento lento cuando $p_j \rightarrow 0$. Este comportamiento es de esperar en las varianzas de estas funciones de enlace, pues son complementarias.

Por último, en la figura 4, se despliega la $\mathrm{V}[\hat{\beta}_j]$ cuando se modelan los datos mediante las funciones de enlace clog (figuras 4(a) y 4(b), ecuación [16]) y su contraparte log (figuras 4(c) y 4(d)), ecuación [18]). Nótese el comportamiento similar y complementario a las varianzas presentadas en la figura 3, con la diferencia de que cuando se utiliza el enlace clog y $p_j \rightarrow 0$, o se recurre al enlace log y $p_j \rightarrow 1$, las varianzas tienden a cero y no al aumento, como en los casos de cloglog y loglog.



(a) (b)

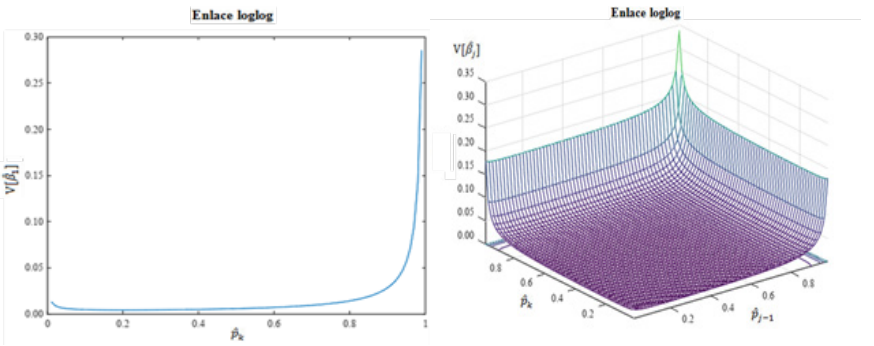
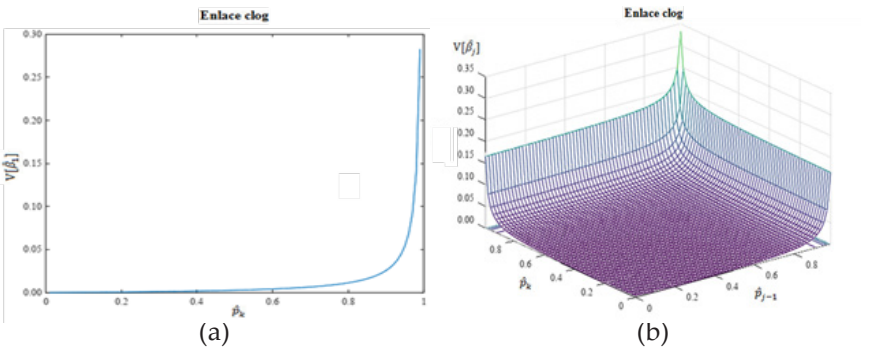


Figura 3. $V[\hat{\beta}_j]$ según funciones de enlace asimétricas cloglog y loglog (con $n_j = 350$, para $j = 1, \ldots, k$).
Fuente: Elaboración propia.



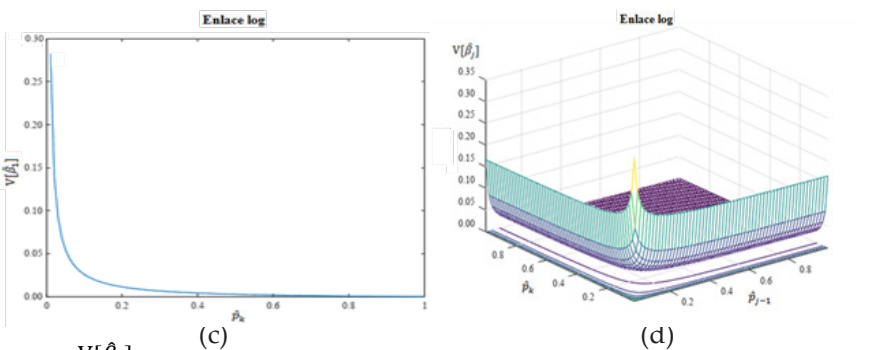


Figura 4. $V[\hat{\beta}_j]$ según funciones de enlace asimétricas clog y log (con $n_j = 350$, para $j = 1, \ldots, k$).
Fuente: Elaboración propia.

7. Conclusiones

En el contexto de los modelos lineales generalizados se utiliza principalmente el enlace logit, aprovechando que se trata del enlace canónico para la función de masas binomial. Sin embargo, existen funciones de enlace alternativas que podrían mejorar sustancialmente las inferencias y predicciones, pero rara vez son consideradas y evaluadas. Por ejemplo, en aplicaciones con conjuntos de datos grandes y que se distribuyen a lo largo del gradiente completo de las funciones de enlace ajustadas, sobre todo en los extremos, una función de enlace asimétrica ajustaría mejor que la tradicional logit.

En la escogencia de la función de enlace, dos aspectos son relevantes: la bondad del ajuste y la facilidad de interpretación. Las ecuaciones desarrolladas en este estudio permiten al investigador indagar y evaluar, a través de cálculos y gráficos simples, sobre cuál función de enlace resulta apropiada cuando se considera el ajuste de un modelo binomial en una tabla de contingencia con respuesta dicotómica y un factor explicativo con k niveles. Ante diversos escenarios planteados, empleando las ecuaciones derivadas, es posible responder a las preguntas: ¿Qué función de enlace se ajusta mejor y tiene una interpretación más simple? Las respuestas a estas interrogantes sirven de guía para el investigador, ante la disyuntiva de seleccionar el mejor modelo posible que se ajuste a su situación particular de datos, atendiendo a los objetivos que se propone.

En el modelado, el uso de paquetes estadísticos muchas veces hace transparente el ajuste de modelos para el investigador, lo que limita la compresión del proceso subyacente. En esta investigación se pudo manejar explícitamente, y a un nivel básico, todos los componentes del modelo binomial, a través de un enfoque que permite entender completamente los principios que gobiernan los modelos lineales generalizados, en general, y en particular, aquellos que rigen al modelo binomial.

La deducción de las ecuaciones obtenidas, así como su análisis principalmente gráfico, hacen sencilla la tarea de decidir sobre los modelos, en el caso de que se postule un modelo binomial saturado. Esta situación es principalmente ilustrativa y de naturaleza introductoria, pero sirve de base a futuros estudios y generalizaciones que los autores se proponen realizar cuando se postulen modelos no-saturados, o bien con un número cualquiera de factores explicativos, así como el tratamiento de las observaciones anómalas en cada uno de estos contextos.

8. Referencias

Agresti, Alan (2015). *Foundations of linear and generalized linear models.* New Jersey: John Wiley & Sons, 480 pp. DOI: 10.1111/biom.12759.

Collett, David (2002). *Modelling binary data.* Second Edition. EEUU: Chapman & Hall, 408 pp. DOI: 10.1201/b16654.

Czado, Claudia y Munk, Axel (2000). "Noncanonical links in generalized linear models-when is the effort justified?". *Journal of statistical planning and inference*, 87, 2 (June, 2000), pp. 317-345. DOI: 10.1016/s0378-3758(99)00195-0.

Czado, Claudia y Santner, Thomas (1992). "The effect of link misspecification on binary regression inference". *Journal of statistical planning and inference*, 33, 2 (November, 1992), pp. 213-231. DOI: 10.1016/0378-3758(92)90069-5.

Dobson, Annette (2002). *An introduction to generalized linear models*. Second Edition. Florida: Chapman & Hall, 225 pp. DOI: 10.1201/9781420057683.

Enchautegui, María (2000)."Módulo de estudio sobre modelos Probit y Logit." Puerto Rico: Universidad de Puerto Rico, 772 pp.

Hardin, James y Hilbe, Joseph (2007). *Generalized linear models and extensions*. Fourth Edition. Texas: Stata Press, 598 pp.

Hilbe, Joseph (2009). *Logistic regression models*. First Edition. New York: Chapman & Hall, 656 pp. DOI: 10.1201/9781420075779.

Hosmer, David y Lemeshow, Stanley (2000). *Applied logistic regression*. Second Edition. Canada: John Wiley & Sons, 383 pp. DOI: 10.1002/0471722146.

Koenker, Roger y Yoon, Jungmo (2009). "Parametric links for binary choice models: A fisherian-bayesian colloquy". *Journal of Econometrics*, 152, 2 (October, 2009), pp. 120-130. DOI: 10.1016/j.jeconom.2009.01.009.

Li, Jingwei (2014). *Choosing the proper link function for binary data*. Tesis doctoral no publicada. Austin: The University of Texas, 37 pp.

McCullagh, Peter y Nelder, John (1989). *Generalized linear models*. Second Edition. London: Chapman & Hall, 532 pp. DOI: 10.1007/978-1-4899-3242-6.

McCulloch, Charles y Searle, Shayle (2000). *Generalized linear, and mixed models*. First Edition. New York: John Wiley & Sons Inc, 335 pp. DOI: 10.1002/0471722073.

Nelder, John y Wedderburn, Robert (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society*, 135, 3, pp. 370-384. DOI: 10.2307/2344614.

Piegorsch, Walter (1992). "Complementary log regression for generalized linear models". *The American Statistician*, 46, 2 (May, 1992), pp. 94-99. DOI: 10.1080/00031305.1992.10475858.

Ponsot, Ernesto (2011). *Estudio de la agrupación de niveles del factor en el modelo logit binomial*. Tesis doctoral. Mérida, Venezuela: Instituto de Estadística Aplicada y Computación de la Universidad de Los Andes, 2011, 179 pp.

Tutz, Gerhard (2011). *Regression for categorical data*. UK: Cambridge University Press, 561 pp. DOI: 10.1017/cbo9780511842061