Revista Colombiana de Estadística
Volumen 28 N° 1. pp. 77 a 96. Junio 2005

Una comparación entre la inferencia basada en las estadísticas de Wald y razón de verosimilitud en los modelos logit y probit vía Monte Carlo

Artur José Lemonte[*]
Luis Hernando Vanegas[**]

Resumen

Presentamos un estudio que evalúa y compara el desempeño de puebas de hipótesis e intervalos de confianza basados en la estadística de Wald con los basados en la estadística de razón de verosimilitud para los modelos logit y probit. Esta comparación se hace a través de las tasas de cobertura de los intervalos de confianza, tasas superior e inferior de los intervalos de confianza, y la potencia de la prueba de significancia. Se emplearon métodos de simulación de Monte Carlo. También se compararon las estadísticas de Wald y de razón de verosimilitud en los modelos logit y probit en presencia de errores en la especificación del modelo.

Palabras Claves: *Modelo logit, modelo probit, error de especificación, simulación de Monte Carlo, test de razón de verosimilitud, test de Wald.*

Abstract

In this paper we present a study which evaluates and compares the performance of hypothesis testing and confidence intervals based on Wald's statistic with those based on the likelihood ratio statistic for probit and logit models. We compare the rate of coverage of the confidence intervals, the maximum and minimum confidence interval coverage rate, and the power of the significance tests employing Monte Carlo simulation methods. The Wald and likelihood ratio statistics are also compared for the logit and probit models in the presence of errors in model specification.

Keywords: *Logit model, Probit model, Specification errors, Monte Carlo methods, Likelihood ratio test, Wald's test.*

[*]Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brasil. E-mail: artur@cox.de.ufpe.br
[**]Departamento de Estatística, Universidade Federal de Pernambuco, Cidade Universitária, Recife/PE, 50740-540, Brasil. E-mail: hernando@cox.de.ufpe.br

1. Introducción

Los métodos de regresión han sido la herramienta principal para describir la relación existente entre una variable respuesta y una o más variables explicativas. En particular, cuando la variable respuesta es dicotómica, los modelos de regresión logit y probit vienen siendo los métodos más aplicados en muchos campos del conocimiento, como por ejemplo Medicina o Biología, cuando el interés primario del análisis de datos está en evaluar y cuantificar la influencia de una o más variables sobre un evento de interés. Este análisis es aplicado usando intervalos de confianza y pruebas de significancia para los parámetros del modelo. A través de la estimación puntual o por intervalo de los parámetros, es posible calcular medidas de asociación como el riesgo relativo y el riesgo relativo indirecto. Además, la significancia y la parsimonia del modelo son evaluadas a través de pruebas de significancia. Por lo tanto, es necesario contar con una metodología de inferencia confiable que permita obtener conclusiones válidas de los datos.

Este artículo tiene como objetivo evaluar y comparar el desempeño de la inferencia basada en la estadística de Wald con la basada en la estadística de razón de verosimilitud, siendo estas dos estrategias, las más usadas para la construcción de intervalos de confianza y pruebas de hipótesis en modelos de respuesta dicotómica. Es nuestro interés comparar características tales como: tasas de cobertura de los intervalos de confianza, tasas superior e inferior de los intervalos de confianza y potencia de la prueba de significancia.

En la sección 2, exponemos algunos conceptos en modelos para respuesta binaria; en la sección 3 presentamos los resultados obtenidos a través de simulaciones de Monte Carlo sobre el desempeño de las pruebas de significancia basadas en las estadísticas de Wald y de razón de verosimilitud en modelos logit y probit; en la sección 4 se resumen algunas conclusiones.

2. Modelo para respuesta binaria

Supongamos que para cada individuo o unidad experimental k, tenemos el vector $(y_k, x_{k1}, x_{k2}, \dots, x_{kp})$ donde y_k puede asumir sólo uno de dos valores posibles, denotados por conveniencia 1 (éxito) y 0 (fracaso), y sea $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})$ un conjunto de variables observadas para explicar o predecir el valor de y_k. Denotamos la probabilidad de éxito condicionada por la información en el vector $\mathbf{x}_k$ como:

$$\pi(\mathbf{x}_k) = P\left[Y_k = 1 | x_{k1}, \dots, x_{kp}\right] = P\left[Y_k = 1 | \mathbf{x}_k\right] \tag{1}$$

Suponer que la dependencia de la probabilidad de éxito sobre el vector $\mathbf{x}_k$ ocurre a partir de una combinación lineal implica que el valor de π puede asumir valores mayores que uno o menores que cero, generando inconsistencias con las leyes de la probabilidad. Una forma simple de solucionar este problema es usando una función g, llamada función de enlace, que proyecte el intervalo $(0, 1)$ en toda la recta. Esta solución conduce al siguiente modelo:

$$g\left[\pi(\mathbf{x}_k)\right] = \sum_{j=0}^{p} \beta_j x_{kj}, \tag{2}$$

donde $(\beta_0, \beta_1, \ldots, \beta_p)$ es un conjunto de parámetros desconocidos y $x_{k0} = 1$.

Se consideran diferentes funciones de enlace, dependiendo del fenómeno en estudio y de la relación entre la probabilidad de éxito y las variables explicativas. Tres funciones usadas comúnmente en la práctica son:

1. Logit:
$$g[\pi(\mathbf{x}_k)] = \log\left(\frac{\pi(\mathbf{x}_k)}{1-\pi(\mathbf{x}_k)}\right)$$
2. Probit:
$$g[\pi(\mathbf{x}_k)] = \Phi^{-1}(\mathbf{x}_k)$$
3. log-log:
$$g[\pi(\mathbf{x}_k)] = -\log\left\{-\log[\pi(\mathbf{x}_k)]\right\}$$

donde $\Phi(\cdot)$ es la función de distribución acumulada normal estándar. En este artículo se consideran sólamente 1 y 2.

La relación entre la probabilidad de éxito y el conjunto de variables explicativas puede expresarse con el siguiente modelo:

$$y = \mu + \varepsilon, \tag{3}$$

donde $\mu = \mathrm{E}\left(Y|\mathbf{x}\right) = \pi\left(\mathbf{x}\right)$ es la componente sistemática y ε la componente aleatoria con la distribución siguiente:

$$\varepsilon = \begin{cases} 1-\pi(\mathbf{x}) & \text{con probabilidad } \pi(\mathbf{x}) \\ -\pi(\mathbf{x}) & \text{con probabilidad } 1-\pi(\mathbf{x}) \end{cases}$$

y, además, $\mathrm{E}(\varepsilon) = 0$ y $\mathrm{Var}(\varepsilon) = \pi(\mathbf{x})\left(1-\pi(\mathbf{x})\right)$. La relación entre la probabilidad de éxito y el conjunto de variables explicativas se determina especificando la forma de la función de enlace g, dada en (2).

2.1. Estimación Puntual de los Parámetros

Los parámetros del vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ se estiman usando datos muestrales obtenidos con un diseño experimental controlado o por medio de registros históricos existentes. Estos datos consisten de n ocurrencias independientes del fenómeno en estudio, pudiendo ser expresados en la forma:

$$\mathbf{y} = (y_1, \ldots, y_n)^\top \quad \text{y} \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{bmatrix}$$

donde se asume que las variables en $\mathbf{X}$ pueden explicar parte de la variabilidad de la respuesta Y.

El método de estimación más usado es el de máxima verosimilitud que consiste proponer como estimaciones de los parámetros del modelo, los valores que maximizam la función de verosimilitud:

$$L(\mathbf{y},\mathbf{X}) = \prod_{k=1}^{n} \pi(\mathbf{x}_k)^{y_k} \left(1-\mathbf{x}_k\right)^{1-y_k} \tag{4}$$

o, equivalentemente, aquellos que maximizan la función de log-verosimilitud:

$$l(\mathbf{y},\mathbf{X}) = \sum_{k=1}^{n} \left\{ y_k \log[\pi(\mathbf{x}_k)] + (1-y_k)\log[1-\pi(\mathbf{x}_k)] \right\}$$

A partir de la ecuación anterior los parámetros serán estimados substituyendo $\pi(\mathbf{x}_k)$ por:

$$g^{-1}\left(\sum_{j=0}^{p} \beta_j x_{kj}\right)$$

maximizando la función resultante con respecto a $\boldsymbol{\beta}$. En el caso del modelo logit, la función a ser maximizada es:

$$l(\boldsymbol{\beta};\mathbf{y},\mathbf{X}) = \sum_{k=1}^{n}\sum_{j=0}^{p} y_k \beta_j x_{kj} - \sum_{k=1}^{n} \log\left(1+\exp\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)\right) \tag{5}$$

cuyo gradiente está dado por:

$$\frac{\partial l\left(\boldsymbol{\beta};\mathbf{y},\mathbf{X}\right)}{\partial \beta_r} = \sum_{k=1}^{n}\left(y_k - \frac{\exp\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)}{1+\exp\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)}\right) x_{kr}$$

Analogamente, para el modelo probit, la función a ser maximizada es:

$$\begin{aligned} l(\boldsymbol{\beta};\mathbf{y},\mathbf{X}) \quad = \quad & \sum_{k=1}^{n}\left\{ y_k \log\left[\Phi\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)\right]\right. \\ & \left. + (1-y_k)\log\left[1-\Phi\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)\right]\right\} \end{aligned} \tag{6}$$

cuyo gradiente está dado por:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial \beta_r} \quad = \quad & \sum_{k=1}^{n}\left[f\left(\sum_{j=0}^{p}\beta_j x_{kj}\right) x_{kr}\left(\frac{y_k}{\Phi\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)}\right.\right. \\ & \left.\left. - \frac{(1-y_k)}{1-\Phi\left(\sum_{j=0}^{p}\beta_j x_{kj}\right)}\right)\right] \end{aligned}$$

Los estimadores de máxima verosimilitud para los modelos logit y probit no tienen fórmulas analíticas para su cálculo, pues estos dependen de un sistema de ecuaciones no lineales, por tanto las estimaciones se obtienen maximizando las funciones (5) y (6) a través de métodos de optimización no lineal como, por ejemplo, BFGS, Newton-Raphson, etc. En este artículo se utilizó el método cuasi-Newton BFGS para la maximización de las funciones (5) y (7) considerando gradiente analítico. Con estas dos funciones, acontecen rara vez problemas con la convergencia de los métodos de optimización, pues usualmente esto ocurre cuando alguna combinación de variables, discrimina perfectamente los éxitos de los fracasos en la respuesta (McCullagh & Nelder 1989).

Asintóticamente los estimadores de máxima verosimilitud de los parámetros de los modelos logit y probit son insesgados y tienen varianza igual a la inversa de la matriz de información de Fisher (Hosmer & Lemeshow 1989), es decir:

$$\operatorname{Var}(\widehat{\boldsymbol{\beta}}) = -\left[\mathrm{E}\left(\frac{\partial^2 l(\boldsymbol{\beta};\mathbf{y},\mathbf{X})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top}\right)\right]^{-1} = (\mathbf{X}^\top G\mathbf{X})^{-1}$$

En el modelo logit, G es dada por:

$$G = \begin{bmatrix} \pi(\mathbf{x}_1)(1-\pi(\mathbf{x}_1)) & 0 & \cdots & 0 \\ 0 & \pi(\mathbf{x}_2)(1-\pi(\mathbf{x}_2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi(\mathbf{x}_n)(1-\pi(\mathbf{x}_n)) \end{bmatrix}$$

y en el modelo probit por:

$$G = \begin{bmatrix} \frac{f\left(\sum_{i=0}^p \beta_i x_{1i}\right)}{\pi(\mathbf{x}_1)(1-\pi(\mathbf{x}_1))} & 0 & \cdots & 0 \\ 0 & \frac{f\left(\sum_{i=0}^p \beta_i x_{2i}\right)}{\pi(\mathbf{x}_2)(1-\pi(\mathbf{x}_2))} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{f\left(\sum_{i=0}^p \beta_i x_{ni}\right)}{\pi(\mathbf{x}_n)(1-\pi(\mathbf{x}_n))} \end{bmatrix}$$

La función $f(\cdot)$ representa la densidad de la distribución normal estándar. Así, la varianza estimada de los estimadores se obtiene sustituyendo las cantidades poblacionales por las cantidades muestrales, es decir, sustituyendo $\boldsymbol{\beta}$ y $\pi(\mathbf{x}_k)$ por sus valores estimados.

2.2. Estimación por intervalo de los Parámetros

En el caso de la estimación por intervalo de los parámetros, se dispone de dos estratégias diferentes, basadas en las estadísticas de máxima verosimilitud (LR) y de Wald, cuya distribución asintótica es Ji-cuadrado. La estadística LR para el j-ésimo elemento del vector $\boldsymbol{\beta}$, evaluada en el punto γ está definida por:

$$\mathrm{LR}_j(\gamma) = -2\left[l^*(\boldsymbol{\beta}^\star;\mathbf{y},\mathbf{X}) - l^{**}(\boldsymbol{\beta};\mathbf{y},\mathbf{X})\right], \quad j = 0,1,\ldots,p \tag{7}$$

donde $l^{**}(\boldsymbol{\beta}; \mathbf{y}, \mathbf{X})$ es el máximo de la función log-verosimilitud (definida para el caso logit y probit en las ecuaciones (5) y (6) respectivamente) en relación a $\boldsymbol{\beta}$ y $l^{*}(\boldsymbol{\beta}^{\star}; \mathbf{y}, \mathbf{X})$ es el máximo de la función log-verosimilitud en relación a $\boldsymbol{\beta}^{\star} = (\beta_0, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_p)$ siendo $\beta_j = \gamma$. Así, el intervalo de confianza para β_j, considerando un nivel de confianza $(1-\alpha)$, donde α representa el nivel de significancia, utilizando la estadística LR es:

$$\mathrm{IC}[\beta_j, 1-\alpha] = \{\beta \mid \mathrm{LR}_j(\beta) < q_{1-\alpha}\}, \tag{8}$$

donde $q_{1-\alpha}$ es el percentil $(1-\alpha)$ de la distribución Ji-cuadrado con un grado de libertad. Análogamente, la estadística de Wald para el j-ésimo elemento del vector $\boldsymbol{\beta}$ evaluada γ está definida por:

$$\mathrm{W}_j(\gamma) = \frac{(\widehat{\beta}_j - \gamma)^2}{\widehat{\mathrm{Var}}(\widehat{\beta}_j)}, \tag{9}$$

donde $\widehat{\beta}_j$ y $\widehat{\mathrm{Var}}(\widehat{\beta}_j)$ son los valores estimados de β_j y de la varianza del estimador de β_j respectivamente. Así, el intervalo de confianza para β_j utilizando la estadística de Wald está dado por:

$$\mathrm{IC}[\beta_j, 1-\alpha] = \{\beta \mid \mathrm{W}_j(\beta) < q_{1-\alpha}\} \tag{10}$$

Normalmente en aplicaciones prácticas, el intervalo de confianza para β_j basado en la estadística de Wald es más usado que el basado en LR, pues este último no tiene una fórmula analítica para su cálculo, siendo más difícil de calcular y más "costoso" computacionalmente.

Es posible probar hipótesis sobre β_j utilizando su intervalo de confianza, verificando si $\mathrm{IC}[\beta_j, 1-\alpha]$ contiene el valor de β_{j0} (valor atribuido para β_j en $\mathcal{H}_0$). Entonces, una prueba de hipótesis para β_j con un nivel de significancia α puede describirse de la forma siguiente:

$$\begin{aligned} \mathcal{H}_0 &: \beta_j = \beta_{j0} \\ \mathcal{H}_1 &: \beta_j \neq \beta_{j0} \end{aligned} \tag{11}$$

con la regla de decisión: rechazar $\mathcal{H}_0$ si $\beta_{j0} \notin \mathrm{IC}[\beta_j, 1-\alpha]$ o, equivalentemente, rechazar $\mathcal{H}_0$ si $\mathrm{W}_j(\beta_{j0}) > q_{1-\alpha}$ en el caso de la estadística de Wald y rechazar $\mathcal{H}_0$ si $\mathrm{LR}_j(\beta_{j0}) > q_{1-\alpha}$ para LR. Particularmente, para evaluar la significancia de β_j en el modelo, las estadísticas de prueba para Wald y LR son representadas por $\mathrm{W}_j(0)$ y $\mathrm{LR}_j(0)$ respectivamente. Dada la importancia de la prueba (11), evaluamos sus características:

- P(error tipo 1), denotado por α: probabilidad de concluir que β_j es significativo cuando realmente no lo es;

- P(error tipo 2), denotado por δ: probabilidad de concluir que β_j no es significativo cuando realmente lo es;

- Potencia de la prueba, dada por $1-\delta$: probabilidad de concluir que β_j es significativo cuando realmente lo es.

3. Experimentos de Monte Carlo

Los resultados presentados en esta sección corresponden al modelo de la ecuación (3), con $\pi(x) = g^{-1}(\beta_0 + \beta_1 x)$, es decir, apenas una variable explicativa. Este modelo fue considerado en diferentes escenarios de simulación, definidos por las características siguientes:

(i) probabilidad de éxito en la población, $\mathcal{P} = P(Y = 1)$;

(ii) fuerza y dirección de la asociación entre la respuesta y la variable explicativa, representada por β_1;

(iii) tamaño de la muestra, n.

Estas cantidades son especificadas en las tablas. En las Tablas 1 y 2, presentamos la potencia de la prueba de significancia para β_1 dada por:

$$1 - \widehat{\delta} = 1 - \frac{\#\{0 \in \text{IC}[\beta_1, 1-\alpha]\}}{r}$$

donde r es el número de réplicas de Monte Carlo ($r = 5000$). En las Tablas 3 y 4 tenemos la potencia de la pruebas de significancia para β_1 cuando las pruebas de Wald y de razón de verosimilitud se aplican con error en la especificación del modelo. En las Tablas 5 a 22 presentamos la tasa de cobertura del intervalo de confianza para β_1, dada por:

$$\text{TC}[\beta_1, 1-\alpha] = \frac{\#\{\beta_1 \in \text{IC}[\beta_1, 1-\alpha]\}}{r}$$

tambien presentamos la tasa superior, representada por TSup, que es la frecuencia con que el límite superior de los r intervalos de confianza no excede el verdadero valor de β_1, la tasa inferior, representada por TInf, que es la frecuencia con que el límite inferior de los r intervalos de confianza exceden el verdadero valor de β_1.

3.1. Resultados

En esta sección presentamos los resultados obtenidos. Todo el proceso de simulación fue realizado utilizando el lenguaje de programación `Ox` (Doornik 2001).

En las Tablas 1 y 2 el comportamiento de la inferencia en los modelos logit y probit es similar. La potencia aumenta con el tamaño de la muestra y con la fuerza de la asociación entre la respuesta y la variable explicativa, medida a través del valor de β_1. Además, el comportamiento de la potencia cuando $\mathcal{P} = 0.3$ es bastante similar a $\mathcal{P} = 0.7$, lo que indica que la potencia es simétrica con respecto a $\mathcal{P} = 0.5$, caso donde la inferencia fue más eficiente, pues presentó en todos los casos mayor potencia. Este hecho puede explicarse de la siguiente forma: cuando $\mathcal{P}$ está próximo a cero, es necesario considerar una muestra "grande" de observaciones para obtener un número de éxitos (individuos con la característica de interés) suficiente para poder estudiar adecuadamente la asociación entre la respuesta y la variable explicativa; análogamente cuando $\mathcal{P}$ está próximo de uno, es necesario considerar una muestra "grande" de observaciones para obtener un número de fracasos (indivíduos sin la característica de interés) suficiente para poder estudiar adecuadamente la asociación entre las variables respuesta y explicativa. De tal manera, para un tamaño de la muestra fijo, la eficacia de la inferencia estadística será mayor en el caso de $\mathcal{P} = 0.5$ con relación a los otros valores de $\mathcal{P}$. Se nota también que la potencia de las pruebas de Wald y LR fueron similares en todos los casos, evidenciando que para los modelos logit y probit, estas pruebas presentan la misma eficiencia.

En las Tablas 5 a 13, considerando $n = 100$, que los intervalos de confianza para β_1 en el modelo logit presentan tasas de cobertura muy similares esperadas; es decir, presentan valores próximos de 90%, 95% y 99% respectivamente. En el caso de los valores de TInf y TSup, considerando la estadística de Wald, se observa que los valores de TInf disminuyen a medida que el valor de β_1 aumenta, confirmando los resultados obtenidos por Hauck & Donner (1977) y Jennings (1986), segun los cuales, la distribución de la estadística de prueba el test de Wald converge a la distribución Ji-cuadrado con un grado de liberdad (esto es, la distribución de la estadística en $\mathcal{H}_0$) cuando $\beta_1 \longrightarrow \infty$.

Para la estadística LR, los valores de TInf y TSup presentaron comportamiento más uniforme, manteniendo en casi todos los casos la simetría del intervalo de confianza, es decir, el valor de TInf está próximo del valor de TSup. En las Tablas 14 a 22, los intervalos de confianza para β_1 en el modelo probit presentan tasas de cobertura muy similares a las esperadas, es decir, presentan valores próximos de 90%, 95% y 99% respectivamente. En relación a los valores de TInf y TSup, es posible observar que el valor de TInf en la mayoría de los casos es mayor que los valores de TSup, indicando que la distribución de $\widehat{\beta}_1$ no es simétrica en relación a β_1.

En la Tabla 3 consideramos la presencia de errores en la especificación, es decir, generamos ocurrencias del modelo logit y estimamos un modelo probit. En la práctica, conocer el comportamiento de la inferencia en esta situación es muy importante, pues la selección del modelo se debe en muchos casos a razones subjetivas del investigador. Por lo tanto, es importante evaluar el desempeño de la inferencia cuando la función de enlace usada es inadecuada. Se observa que la potencia aumenta con el tamaño de la muestra y con el valor de β_1. Además, puede observarse que la prueba de Wald fue más eficiente en todos los casos; este test no se afecta por el error en la especificación del modelo, pues no depende fuertemente de la función de enlace utilizada para el modelo aplicado. La prueba LR es fuertemente afectada por el error en la especificación del modelo, presentando tasas muy bajas comparadas con las de la Tabla 3 donde el modelo generado y el estimado fueron el mismo.

En la Tabla 4 también consideramos la presencia de error en la especificación del modelo; en este caso, generamos ocurrencias del modelo probit y estimamos un modelo logit. El desempeño de las prueba de Wald y de razón de verosimilitud son similares a los presentados en la Tabla 3, es decir, la potencia crece con el tamaño de la muestra y con el valor de β_1, siendo más eficiente la prueba de Wald en todos los casos. El desempeño de la prueba LR fue también bastante afectado por el error de especificación, presentando tasas muy bajas comparadas con las de la Tabla 2 donde el modelo generado y el estimado fueron el mismo.

Tabla 1: Poder de la prueba de Wald y LR en el modelo logit

			LR			Wald		
$\mathcal{P}$	n	β_1	10%	5%	1%	10%	5%	1%
0,3	50	0,2	0,153	0,085	0,022	0,152	0,083	0,019
		0,7	0,578	0,441	0,210	0,576	0,436	0,191
		1,2	0,912	0,849	0,627	0,911	0,845	0,596
	80	0,2	0,181	0,109	0,029	0,183	0,110	0,029
		0,7	0,800	0,690	0,435	0,802	0,693	0,438
		1,2	0,992	0,981	0,919	0,992	0,981	0,921
	100	0,2	0,239	0,152	0,049	0,238	0,149	0,046
		0,7	0,919	0,856	0,663	0,916	0,852	0,656
		1,2	0,999	0,998	0,985	0,999	0,998	0,985
0,5	50	0,2	0,152	0,084	0,021	0,152	0,084	0,021
		0,7	0,639	0,509	0,260	0,639	0,510	0,260
		1,2	0,954	0,905	0,724	0,954	0,905	0,727
	80	0,2	0,188	0,111	0,035	0,188	0,111	0,035
		0,7	0,848	0,748	0,488	0,848	0,748	0,488
		1,2	0,996	0,990	0,951	0,996	0,990	0,951
	100	0,2	0,266	0,170	0,059	0,267	0,170	0,058
		0,7	0,956	0,917	0,769	0,956	0,917	0,769
		1,2	0,999	0,999	0,997	0,999	0,999	0,997
0,7	50	0,2	0,140	0,075	0,017	0,142	0,077	0,020
		0,7	0,566	0,437	0,208	0,568	0,441	0,224
		1,2	0,901	0,830	0,609	0,901	0,832	0,631
	80	0,2	0,183	0,104	0,032	0,181	0,102	0,032
		0,7	0,767	0,656	0,399	0,765	0,652	0,397
		1,2	0,987	0,967	0,875	0,986	0,966	0,873
	100	0,2	0,249	0,161	0,054	0,249	0,161	0,054
		0,7	0,928	0,874	0,706	0,931	0,877	0,715
		1,2	0,999	0,998	0,991	0,999	0,998	0,992

Tabla 2: Poder de la prueba de Wald y LR en el modelo probit

			LR			Wald		
$\mathcal{P}$	n	β_1	10%	5%	1%	10%	5%	1%
0,3	50	0,2	0,230	0,140	0,047	0,231	0,142	0,044
		0,4	0,554	0,414	0,193	0,554	0,413	0,191
		0,7	0,923	0,863	0,660	0,923	0,864	0,659
	80	0,2	0,324	0,212	0,077	0,325	0,213	0,077
		0,4	0,773	0,661	0,401	0,775	0,663	0,402
		0,7	0,995	0,985	0,928	0,995	0,985	0,930
	100	0,2	0,447	0,327	0,133	0,447	0,325	0,132
		0,4	0,906	0,840	0,630	0,904	0,841	0,625
		0,7	0,999	0,998	0,989	0,999	0,998	0,990
0,5	50	0,2	0,232	0,146	0,041	0,233	0,148	0,042
		0,4	0,585	0,454	0,219	0,584	0,455	0,218
		0,7	0,941	0,887	0,684	0,942	0,887	0,687
	80	0,2	0,339	0,224	0,079	0,339	0,223	0,080
		0,4	0,793	0,681	0,413	0,793	0,683	0,419
		0,7	0,994	0,987	0,938	0,994	0,987	0,938
	100	0,2	0,491	0,361	0,171	0,491	0,363	0,169
		0,4	0,934	0,883	0,706	0,935	0,883	0,706
		0,7	1,000	0,999	0,995	1,000	0,999	0,996
0,7	50	0,2	0,217	0,131	0,040	0,217	0,133	0,042
		0,4	0,548	0,417	0,192	0,549	0,419	0,197
		0,7	0,914	0,845	0,641	0,915	0,847	0,646
	80	0,2	0,315	0,207	0,074	0,314	0,206	0,072
		0,4	0,744	0,629	0,373	0,744	0,626	0,367
		0,7	0,990	0,979	0,906	0,990	0,981	0,904
	100	0,2	0,461	0,341	0,149	0,464	0,342	0,153
		0,4	0,916	0,856	0,672	0,915	0,858	0,675
		0,7	0,916	0,856	0,672	0,915	0,858	0,675

Tabla 3: Poder de la prueba de Wald y LR generando logit y estimando probit

			LR			Wald		
$\mathcal{P}$	n	β_1	10%	5%	1%	10%	5%	1%
0,3		0,2	0,122	0,067	0,017	0,153	0,087	0,020
	50	0,7	0,185	0,116	0,031	0,281	0,182	0,060
		1,2	0,304	0,198	0,079	0,579	0,437	0,209
		0,2	0,137	0,079	0,022	0,181	0,109	0,029
	80	0,7	0,237	0,146	0,050	0,414	0,289	0,111
		1,2	0,438	0,322	0,144	0,801	0,690	0,434
		0,2	0,161	0,091	0,024	0,238	0,150	0,047
	100	0,7	0,310	0,214	0,075	0,554	0,436	0,203
		1,2	0,438	0,322	0,144	0,915	0,690	0,434
0,5		0,2	0,111	0,060	0,012	0,153	0,085	0,020
	50	0,7	0,170	0,096	0,026	0,311	0,208	0,069
		1,2	0,299	0,202	0,075	0,636	0,511	0,256
		0,2	0,140	0,075	0,015	0,187	0,112	0,035
	80	0,7	0,238	0,144	0,040	0,459	0,326	0,133
		1,2	0,432	0,317	0,133	0,846	0,750	0,490
		0,2	0,159	0,095	0,025	0,266	0,170	0,060
	100	0,7	0,308	0,206	0,081	0,639	0,512	0,276
		1,2	0,552	0,434	0,250	0,957	0,916	0,769
0,7		0,2	0,115	0,066	0,014	0,141	0,075	0,018
	50	0,7	0,168	0,100	0,030	0,274	0,177	0,061
		1,2	0,284	0,193	0,073	0,566	0,437	0,210
		0,2	0,138	0,074	0,018	0,181	0,104	0,031
	80	0,7	0,231	0,149	0,047	0,396	0,277	0,107
		1,2	0,425	0,317	0,142	0,766	0,653	0,391
		0,2	0,166	0,096	0,026	0,246	0,160	0,053
	100	0,7	0,314	0,213	0,079	0,583	0,456	0,239
		1,2	0,563	0,453	0,244	0,930	0,875	0,708

Tabla 4: Poder de la prueba de Wald y LR generando probit y estimando logit

			LR			Wald		
$\mathcal{P}$	n	β_1	10%	5%	1%	10%	5%	1%
0,3		0,2	0,119	0,067	0,013	0,230	0,138	0,040
	50	0,4	0,163	0,096	0,026	0,554	0,407	0,176
		0,7	0,281	0,175	0,056	0,924	0,860	0,638
		0,2	0,134	0,074	0,017	0,326	0,215	0,080
	80	0,4	0,234	0,139	0,040	0,776	0,663	0,404
		0,7	0,431	0,301	0,111	0,995	0,985	0,930
		0,2	0,160	0,093	0,025	0,442	0,322	0,130
	100	0,4	0,308	0,207	0,074	0,905	0,835	0,625
		0,7	0,552	0,417	0,185	0,999	0,998	0,989
0,5		0,2	0,115	0,057	0,012	0,237	0,145	0,042
	50	0,4	0,175	0,097	0,025	0,589	0,454	0,220
		0,7	0,297	0,185	0,062	0,943	0,886	0,687
		0,2	0,126	0,074	0,016	0,339	0,224	0,080
	80	0,4	0,231	0,140	0,038	0,794	0,681	0,414
		0,7	0,417	0,296	0,107	0,994	0,987	0,938
		0,2	0,171	0,099	0,027	0,492	0,361	0,168
	100	0,4	0,315	0,212	0,077	0,934	0,883	0,706
		0,7	0,567	0,432	0,207	1,000	0,999	0,995
0,7		0,2	0,112	0,059	0,013	0,220	0,135	0,045
	50	0,4	0,171	0,098	0,026	0,548	0,417	0,205
		0,7	0,294	0,189	0,058	0,914	0,846	0,657
		0,2	0,128	0,072	0,018	0,314	0,203	0,074
	80	0,4	0,222	0,139	0,044	0,743	0,625	0,373
		0,7	0,407	0,285	0,117	0,990	0,978	0,906
		0,2	0,166	0,096	0,029	0,466	0,344	0,155
	100	0,4	0,322	0,209	0,077	0,917	0,858	0,683
		0,7	0,577	0,443	0,210	0,999	0,999	0,993

Tabla 5: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	0,2	90%	TInf	0,0564	0,0506	0,05
				TC	0,8970	0,8980	0,90
				TSup	0,0470	0,0514	0,05
			95%	TInf	0,0270	0,0252	0,025
				TC	0,9498	0,9490	0,95
				TSup	0,0232	0,0256	0,025
			99%	TInf	0,0056	0,0044	0,005
				TC	0,9890	0,9880	0,99
				TSup	0,0054	0,0080	0,005

Tabla 6: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	0,7	90%	TInf	0,0608	0,0498	0,05
				TC	0,8914	0,8938	0,90
				TSup	0,0478	0,0564	0,05
			95%	TInf	0,0128	0,0226	0,025
				TC	0,9452	0,9442	0,95
				TSup	0,0420	0,0332	0,025
			99%	TInf	0,0070	0,0020	0,005
				TC	0,9876	0,9890	0,99
				TSup	0,0054	0,0090	0,005

Tabla 7: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	1,2	90%	TInf	0,0646	0,0470	0,05
				TC	0,8888	0,8940	0,90
				TSup	0,0466	0,0590	0,05
			95%	TInf	0,0334	0,0166	0,025
				TC	0,9470	0,9498	0,95
				TSup	0,0196	0,0336	0,025
			99%	TInf	0,0054	0,0018	0,005
				TC	0,9902	0,990	0,99
				TSup	0,0044	0,0086	0,005

Tabla 8: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	0,2	90%	TInf	0,0526	0,0484	0,05
				TC	0,8934	0,8936	0,90
				TSup	0,0540	0,0580	0,05
			95%	TInf	0,0246	0,0226	0,025
				TC	0,9504	0,9486	0,95
				TSup	0,0250	0,0288	0,025
			99%	TInf	0,0044	0,0032	0,005
				TC	0,9900	0,9904	0,99
				TSup	0,0056	0,0064	0,005

Tabla 9: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	0,7	90%	TInf	0,0618	0,0524	0,05
				TC	0,8926	0,8926	0,90
				TSup	0,0448	0,0550	0,05
			95%	TInf	0,0322	0,0244	0,025
				TC	0,9454	0,9468	0,95
				TSup	0,0224	0,0288	0,025
			99%	TInf	0,0076	0,0038	0,005
				TC	0,9872	0,9870	0,99
				TSup	0,0052	0,0092	0,005

Tabla 10: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	1,2	90%	TInf	0,0604	0,0476	0,05
				TC	0,8948	0,8936	0,90
				TSup	0,0448	0,0588	0,05
			95%	TInf	0,0302	0,0180	0,025
				TC	0,9464	0,9466	0,95
				TSup	0,0234	0,0354	0,025
			99%	TInf	0,0058	0,0024	0,005
				TC	0,9892	0,9840	0,99
				TSup	0,0050	0,0136	0,005

Tabla 11: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	0,2	90%	TInf	0,0526	0,0526	0,05
				TC	0,8946	0,8946	0,90
				TSup	0,0528	0,0528	0,05
			95%	TInf	0,0284	0,0284	0,025
				TC	0,9430	0,9430	0,95
				TSup	0,0286	0,0286	0,025
			99%	TInf	0,0064	0,0064	0,005
				TC	0,9860	0,9866	0,99
				TSup	0,0076	0,007	0,005

Tabla 12: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	0,7	90%	TInf	0,0554	0,0466	0,05
				TC	0,8980	0,8976	0,90
				TSup	0,0466	0,0552	0,05
			95%	TInf	0,0280	0,0202	0,025
				TC	0,9452	0,9464	0,95
				TSup	0,0268	0,0314	0,025
			99%	TInf	0,0064	0,0034	0,005
				TC	0,9872	0,9876	0,99
				TSup	0,0064	0,0090	0,005

Tabla 13: Tasas de cobertura − modelo logit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	1,2	90%	TInf	0,0622	0,0462	0,05
				TC	0,8950	0,8986	0,90
				TSup	0,0228	0,0552	0,05
			95%	TInf	0,0296	0,0180	0,025
				TC	0,9476	0,9506	0,95
				TSup	0,0228	0,0314	0,025
			99%	TInf	0,0060	0,0022	0,005
				TC	0,9894	0,9870	0,99
				TSup	0,0046	0,0108	0,005

Tabla 14: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	0,2	90%	TInf	0,0564	0,0524	0,05
				TC	0,8984	0,8998	0,90
				TSup	0,0452	0,0478	0,05
			95%	TInf	0,0290	0,0264	0,025
				TC	0,9462	0,9480	0,95
				TSup	0,0248	0,0256	0,025
			99%	TInf	0,0060	0,0056	0,005
				TC	0,9890	0,9886	0,99
				TSup	0,0050	0,0058	0,005

Tabla 15: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	0,4	90%	TInf	0,0600	0,0536	0,050
				TC	0,8948	0,8968	0,90
				TSup	0,0452	0,0496	0,05
			95%	TInf	0,0326	0,0284	0,025
				TC	0,9446	0,9452	0,95
				TSup	0,0228	0,0264	0,025
			99%	TInf	0,0056	0,0032	0,005
				TC	0,9888	0,9904	0,99
				TSup	0,0056	0,0064	0,005

Tabla 16: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,3	100	0,7	90%	TInf	0,0642	0,0520	0,05
				TC	0,8960	0,9014	0,90
				TSup	0,0398	0,0466	0,05
			95%	TInf	0,0304	0,0210	0,025
				TC	0,9488	0,9530	0,95
				TSup	0,0208	0,0260	0,025
			99%	TInf	0,0072	0,0036	0,005
				TC	0,9878	0,9888	0,99
				TSup	0,0050	0,0076	0,005

Tabla 17: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	0,2	90%	TInf	0,0542	0,0508	0,05
				TC	0,8946	0,8958	0,90
				TSup	0,0512	0,0534	0,05
			95%	TInf	0,0272	0,0258	0,025
				TC	0,9478	0,9472	0,95
				TSup	0,0250	0,0270	0,025
			99%	TInf	0,0054	0,0046	0,005
				TC	0,9888	0,9892	0,99
				TSup	0,0058	0,0062	0,005

Tabla 18: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	0,4	90%	TInf	0,0600	0,0542	0,05
				TC	0,8946	0,8962	0,90
				TSup	0,0454	0,0496	0,05
			95%	TInf	0,0330	0,0280	0,025
				TC	0,9444	0,9480	0,95
				TSup	0,0226	0,0240	0,025
			99%	TInf	0,0078	0,0054	0,005
				TC	0,9858	0,9870	0,99
				TSup	0,0064	0,0076	0,005

Tabla 19: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,5	100	0,7	90%	TInf	0,0604	0,0498	0,05
				TC	0,8952	0,8992	0,90
				TSup	0,0444	0,0510	0,05
			95%	TInf	0,0324	0,0218	0,025
				TC	0,9428	0,9492	0,95
				TSup	0,0248	0,0290	0,025
			99%	TInf	0,0070	0,0040	0,005
				TC	0,9886	0,9886	0,99
				TSup	0,0044	0,0074	0,005

Tabla 20: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	0,2	90%	TInf	0,0576	0,0550	0,05
				TC	0,8940	0,8948	0,90
				TSup	0,0484	0,0502	0,05
			95%	TInf	0,0302	0,0296	0,025
				TC	0,9456	0,9456	0,95
				TSup	0,0242	0,0248	0,025
			99%	TInf	0,0054	0,0046	0,005
				TC	0,9902	0,9900	0,99
				TSup	0,0044	0,0054	0,005

Tabla 21: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	0,4	90%	TInf	0,0552	0,0522	0,05
				TC	0,8864	0,8968	0,90
				TSup	0,0584	0,0510	0,05
			95%	TInf	0,0288	0,0256	0,025
				TC	0,9472	0,9460	0,95
				TSup	0,0240	0,0284	0,025
			99%	TInf	0,0066	0,0078	0,005
				TC	0,9872	0,9850	0,99
				TSup	0,0062	0,0092	0,005

Tabla 22: Tasas de cobertura − modelo probit

					Intervalo de Confianza		
$\mathcal{P}$	n	β_1	$1-\alpha$		LR	Wald	Esperado
0,7	100	0,7	90%	TInf	0,0568	0,0512	0,05
				TC	0,8954	0,8960	0,90
				TSup	0,0478	0,0528	0,05
			95%	TInf	0,0282	0,0246	0,025
				TC	0,9472	0,9466	0,95
				TSup	0,0478	0,0288	0,025
			99%	TInf	0,0062	0,0054	0,005
				TC	0,9870	0,9856	0,99
				TSup	0,0068	0,0090	0,005

4. Conclusiones

En este artículo se estudió la inferencia en modelos logit y probit basada en dos estadísticas muy utilizadas en la práctica: la de razón de verosimilitud y la de Wald. El objetivo fue comparar el desempeño de estas estadísticas en estos modelos. Para esto, analizamos algunas características, tales como la potencia de la prueba de significancia y las tasas de cobertura de los intervalos de confianza. Se concluyó que para los modelos logit y probit las dos estadísticas poseen desempeño muy similar, pues presentaron en todos los casos, intervalos de confianza con tasas de cobertura próximas de los valores nominales. Además, el desempeño de la potencia en relación con el tamaño de la muestra y con la fuerza de la asociación fue el esperado: aumenta a medida que el tamaño de la muestra aumenta y el valor de β_1 se aleja de cero. De acuerdo con la literatura sobre inferencia en el modelo logit, no se recomienda la prueba de Wald, pues, para altos grados de asociación, su potencia decrece para el nivel de significancia. Este comportamiento sugiere que esta prueba no debe usarse, ya que puede conducir a conclusiones equivocadas sobre la significancia de los parámetros en un análisis de regresión usando el modelo logit.

En relación al desempeño de las estadísticas en la presencia de errores en la especificación del modelo, se concluye que la prueba de razón de verosimilitud es fuertemente afectada por este tipo de error, pues presentó en todos los casos una potencia menor que la de la prueba basada en la estadística de Wald. Además, esta última tuvo un desempeño similar al de los casos donde no estaba presente el error de especificación del modelo, es decir, esta prueba no se afectó por la presencia de error en la especificación del modelo. Por lo tanto, en la práctica recomendamos utilizar las dos pruebas para la inferencia de los modelos logit y probit, y verificar si las conclusiones obtenidas en relación a la significancia de los parámetros son las mismas. En caso contrario, recomendamos la estadística de Wald, pues en la práctica es más común equivocarse en la especificación del modelo, que estar en la presencia de una asociación exageradamente grande.

Agradecimientos

Artur José Lemonte agradece el apoyo financiero concedido por la CAPES, y Luis Hernando Vanegas a la CNPq. Los autores agradecen a los profesores de la Maestria en Estadística de la Universidad Federal de Pernambuco por los conocimientos transmitidos, y en especial a los profesores Francisco Cribari Neto y Klaus Leite Pinto Vasconcellos. Los autores también agradecen a Renata Nunes de Souza y a Polyane Alves Santos por su ayuda en la realización de las simulaciones.

Bibliografía

Doornik, J. (2001), *Ox: An Object-Oriented Matrix Language*, 4 edn, Timberlake Consultants Press, London.

Hauck, W. W. & Donner, A. (1977), 'Wald's test as applied to hipotesis in logit analysis', *JASA* **72**, 851–853.

Hosmer, D. & Lemeshow, S. (1989), *Applied Logistic Regression*, Wiley & Sons, New York.

Jennings, D. E. (1986), 'Judgind inference adequacy logistic regression', *JASA* **81**, 471–476.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, Chapman & Hall, New York.