Revista de Matemática: Teoría y Aplicaciones
ISSN: 1409-2433
mta.cimpa@ucr.ac.cr
Universidad de Costa Rica
Costa Rica

LLINÁS SOLANO, HUMBERTO; ARTETA CHARRIS, MARTHA; TILANO HERNÁNDEZ, JORGE

EL MODELO DE REGRESIÓN LOGÍSTICA PARA EL CASO EN QUE LA VARIABLE DE RESPUESTA PUEDE ASUMIR UNO DE TRES NIVELES: ESTIMACIONES, PRUEBAS DE HIPÓTESIS Y SELECCIÓN DE MODELOS

Revista de Matemática: Teoría y Aplicaciones, vol. 23, núm. 1, enero, 2016, pp. 173-197
Universidad de Costa Rica
San José, Costa Rica

Revista de Matemática: Teoría y Aplicaciones 2016 **23**(1) : 173–197
cimpa – ucr issn: 1409-2433 (Print), 2215-3373 (Online)

EL MODELO DE REGRESIÓN LOGÍSTICA PARA EL CASO EN QUE LA VARIABLE DE RESPUESTA PUEDE ASUMIR UNO DE TRES NIVELES: ESTIMACIONES, PRUEBAS DE HIPÓTESIS Y SELECCIÓN DE MODELOS

THE REGRESSION LOGISTICS MODEL IN CASE THE RESPONSE VARIABLE ASSUMES ONE OF THREE LEVELS: ESTIMATIONS, PROOF OF HYPOTHESIS AND MODEL SELECTION

Humberto Llinás Solano*, Martha Arteta Charris†
Jorge Tilano Hernández‡

Received: 21 Jan 2014; Revised: 1 Oct 2015;
Accepted: 21 Oct 2015

---

*Departamento de Matemáticas y Estadística, Universidad del Norte, Km 5 Vía Puerto Colombia, Barranquilla, Colombia. E-Mail: hllinas@uninorte.edu.co
†Misma dirección que/*Same address as*: H. Llinás. E-Mail: mmarteta@uninorte.edu.co
‡Misma dirección que/*Same address as*: H. Llinás. E-Mail: jtilano@uninorte.edu.co

Resumen

Este tratado sigue el siguiente esquema: se presenta, primero el vector score y la matriz de información de los modelos logístico y saturado multinomial con tres posibles niveles de respuesta a partir de la primera y segunda derivada de la función de verosimilitud respecto a los parámetros de los modelos; las relaciones entre el vector score y la matriz de información; la estandarización multivariante de las variables de entrada de cada modelo; las respectivas distribuciones asintóticas; las pruebas de comparación y selección de modelos que abarcan para la variable politómica con tres niveles los modelos logístico y saturado, logístico y submodelo, logístico con el modelo nulo, y logístico con el submodelo de una variable explicativa menos.

Palabras clave: modelo logístico; logit multinomial; vector score; matriz de información de Fisher; distribuciones asintóticas; pruebas de hipótesis.

Abstract

This approach follows the following scheme: first, the vector score and the information matrix from the logistics models and saturated multinomials with three possible response levels starting from the first and second derivative of the function of likelihood with respect to the parameters of the models; the relationship between the vector score and the information matrix; the multivariant standardization of the entry variables of each model; the respective asymptotic distributions; proof of comparisons and model selections that include the polytomic variable with three levels, logistic logistical and saturated models, logistical and submodel, logistical with null model, and logistical with the submodel of a less explanatory variable.

Keywords: logistic model; logit multinomial; vector score; Fisher´s information matrix; asymptotic distributions; hypothesis testing.

Mathematics Subject Classification: 62E20.

1 Introducción

En Llinás [6] se examinó el modelo de regresión logística en donde la variable de respuesta es dicotómica. En ese trabajo se demostró la existencia y unicidad de las estimaciones de máxima verosimilitud (ML) de este modelo. Además, basada en la teoría asintótica de estas ML-estimaciones y del vector score, se encontraron aproximaciones para las diferentes desviaciones $-2\log L$, donde L es la función de verosimilitud. Con ayuda de estas aproximaciones, se obtuvieron estadísticos para diferentes pruebas de hipótesis, cada una con distribución chi-cuadrada. La teoría asintótica fue desarrollada para el caso de variables aleatorias independientes, pero no idénticamente distribuidas. En ese trabajo se hizo siempre la distinción entre datos agrupados y no agrupados.

Siguiendo esta metodología, Llinás [7] estudió el modelo logístico multinomial para el caso en el que la variable de interés puede tomar uno de tres valores posibles. Se describieron modelos relacionados como el nulo, completo y saturado. Para cada modelo, se presentó y demostró teoremas de estimación para los correspondientes parámetros.

Teniendo en cuenta los resultados encontrados en [7] y basados de igual manera en la metodología empleada en [6], en este trabajo se demuestra la existencia y unicidad de las ML-estimaciones del modelo multinomial para el caso en el que la variable de interés puede tomar uno de tres valores posibles. También se utiliza las distribuciones asintóticas de estas ML-estimaciones y del vector score para encontrar aproximaciones para las diferentes desviaciones $-2\log L$, las cuales son útiles para obtener estadísticos para diferentes pruebas de hipótesis, cada una con distribucón chi-cuadrada. La teoría asintótica se desarrolla también para el caso de variables aleatorias independientes, pero no identicamente distribuidas, se hace siempre la distinción entre datos agrupados y no agrupados, dando los detalles que no son encontrados en la corriente literatura (por ejemplo, [1], [2] y [4]).

Este artículo está organizado en seis secciones, en las cuales se presenta un análisis teórico detallado sobre los modelos logísticos multinomial (donde la variable de respuesta toma sólo un valor de tres posibles) describiendo los supuestos básicos, propiedades y características que tienen dichos modelos y presentando y demostrando los resultados más importantes que se conocen sobre las estimaciones de sus parámetros, distribuciones asintóticas y pruebas de comparación de modelos, dando los detalles que, así reunidos, no se encuentran comúnmente en la literatura.

2 Modelo logístico multinomial para tres niveles

En esta sección presentaremos los resultados más importantes introducidos en Llinás [7]. Las demostraciones de los teoremas presentados en esta sección, se pueden encontrar en ese trabajo.

2.1 Modelo básico

La variable de interés Y puede asumir tres niveles: 0, 1 ó 2. Para cada $r = 0, 1, 2$, sea $p_r := P(Y = r)$ la probabilidad de que Y tome el valor r. Definamos:

$$U_{ri} = \begin{cases} 1, & \text{si } Y_i = r; \\ 0, & \text{de otra forma,} \end{cases}$$

donde $r = 0, 1, 2$ y $i = 1, \ldots, n$. Observe que $U_{ri} \sim \mathcal{B}(1, p_{ri})$. Fijando $y = (y_1, \cdots, y_n)^T$, obtenemos el logaritmo de la función de verosimilitud en el parámetro $2n$-dimensional $p = (p_{01}, p_{11}, \ldots, p_{0n}, p_{1n})^T$:

$$\mathcal{L}(p) = \sum_{i=1}^{n} [u_{0i} \ln p_{0i} + u_{1i} \ln p_{1i} + (1 - u_{0i} - u_{1i}) \ln(1 - p_{0i} - p_{1i})]. \quad (1)$$

2.2 El modelo completo

El *modelo completo* es caracterizado por el supuesto de que todos p_{ri} (con $r = 0, 1, 2$ y $i = 1, \ldots, n$) son considerados como parámetros.

Teorema 2.1 *En el modelo completo, las ML-estimaciones de p_{ri} son $\widehat{p}_{ri} = U_{ri}$ con valores $\widehat{p}_{ri} = u_{ri}$ para $r = 0, 1, 2$ y $i = 1, \ldots, n$. Además, $\mathcal{L}_c := \mathcal{L}(y) = 0$.*

2.3 El modelo nulo

El *modelo nulo* es caracterizado por el supuesto de que para cada $r = 0, 1, 2$, todos los p_{ri} ($i = 1, \cdots, n$) son considerados iguales; es decir, se tienen dos parámetros p_0 y p_1. En este caso, (1) será:

$$\mathcal{L}(p) \;=\; n[\overline{u}_0 \ln p_0 + \overline{u}_1 \ln p_1 + (1 - \overline{u}_0 - \overline{u}_1) \ln(1 - p_0 - p_1)]. \quad (2)$$

Teorema 2.2 *En el modelo nulo, la ML-estimación de p_r es $\hat{p}_r = \overline{U}_r$ con valor $\hat{p}_r = \overline{u}_r$. Además, $\mathcal{L}_o := \mathcal{L}(\widehat{p}) < 0$ si y sólo si $0 < \overline{u}_0 + \overline{u}_1 < 1$.*

2.4 El modelo saturado y supuestos

El modelo saturado está caracterizado por los supuestos 1 y 2 presentados en la sección 5 de Llinás [7]. Teniendo en cuenta las notaciones n_j, Z_{rj}, z_{rj}, p_{rj} introducidas en ese trabajo, en el modelo saturado el logaritmo de la función de máxima verosimilitud será

$$\mathcal{L}(p) = \sum_{j=1}^{J} [z_{0j} \ln p_{0j} + z_{1j} \ln p_{1j} + (n_j - z_{0j} - z_{1j}) \ln(1 - p_{0j} - p_{1j})] \,. \quad (3)$$

Teorema 2.3 *En el modelo saturado, las ML-estimaciones de* p_{rj} *son* $\tilde{p}_{rj} = \frac{Z_{rj}}{n_j}$*, con valores* $\tilde{p}_{rj} = \frac{z_{rj}}{n_j}$*,para* $j = 1, \cdots, J$ *y* $r = 0, 1, 2$*. Además,*

$$\mathcal{L}(\widetilde{p}) = \sum_{j=1}^{J} n_j [\tilde{p}_{0j} \ln \tilde{p}_{0j} + \tilde{p}_{1j} \ln \tilde{p}_{1j} + (1 - \tilde{p}_{0j} - \tilde{p}_{1j}) \ln(1 - \tilde{p}_{0j} - \tilde{p}_{1j})]. \quad (4)$$

También se cumple: $\mathcal{L}_s := \mathcal{L}(\tilde{p}) < 0$ *para* $0 < \tilde{p}_{rj} < 1$.

2.5 El modelo logístico y supuestos

Se hacen los supuestos 1 y 2 de la sección 2.4, donde adicionalmente se supone que la matriz de diseño

$$C = \begin{pmatrix} 1 & x_{11} & \cdots & x_{K1} \\ 1 & x_{12} & \cdots & x_{K2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1J} & \cdots & x_{KJ} \end{pmatrix}$$

tiene rango completo $Rg(C) = 1 + K \leq J$. Aquí, x_{kj} es como se explicó en el supuesto 1 del modelo saturado (véase Llinás [7]). Para llegar a un modelo logístico se toma como referencia una de las categorías de la variable dependiente Y, sin pérdida de generalidad tome la referencia como el nivel 2 y se hace el supuesto adicional:

3.
$$\begin{aligned} g_{0j}(x_j) &= \ln\left(\frac{p_{0j}}{p_{2j}}\right) = \delta_0 + \beta_{10}x_{1j} + \cdots + \beta_{K0}x_{Kj} && (5) \\ g_{1j}(x_j) &= \ln\left(\frac{p_{1j}}{p_{2j}}\right) = \delta_1 + \beta_{11}x_{1j} + \cdots + \beta_{K1}x_{Kj} && (6) \end{aligned}$$

donde $x_j := (1, x_{1j}, \ldots, x_{Kj})^T$. Sea

$$\alpha = (\beta_0, \beta_1)^T = (\delta_0, \beta_{10}, \cdots, \beta_{K0}, \delta_1, \beta_{11}, \cdots, \beta_{K1})^T$$

el vector de los $2(1+K)$ parámetros en el modelo. Nótese que el supuesto sobre $Rg(C) = 1 + K$, hace identificable al parámetro α. Para una observación x_j en la población j y para cada r, la probabilidad p_{rj} se calcula de la siguiente manera:

$$p_{rj} = \frac{\exp\{g_r(x_j)\}}{\sum_{s=0}^{2} \exp\{g_s(x_j)\}}. \quad (7)$$

Si $A(x_j) := \sum_{s=0}^{2} \exp\{g_s(x_j)\}$, el logaritmo de la función de verosimilitud se puede escribir en función de α como:

$$\mathcal{L}(\alpha) = \sum_{j=1}^{J} \left[Z_{0j} g_0(x_j) + (n_j - Z_{0j} - Z_{2j}) g_1(x_j) - n_j \ln\left(A(x_j)\right)\right]. \tag{8}$$

3 Score e información para el modelo saturado

En esta sección se presentan resultados asintóticos tanto para el vector score como para la matriz de información en el modelo saturado, resaltando el hecho de se tiene el caso de variables independientes pero no idénticamente distribuidas, dando los detalles que, así reunidos no se encuentran en la literatura. Estos resultados son importantes para las pruebas de comparación que se describirán en la sección 6.

Teorema 3.1 *Sean n_j, p_{rj} y p como se explicó en la sección 5 de Llinás [7] y $v_{rj} := p_{rj}(1 - p_{rj})$. Además, sean $S(p)$, $\tilde{\Im}^*$ y Z^* como en el apéndice A.1. Supóngase que existe $\lim_{n_j \to \infty} \frac{n_j}{n v_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$ para todo $j = 1, 2, \ldots J$ y todo $r = 0, 1$. Entonces para el modelo saturado, se cumplen (cuando $n \to \infty$) y para J fijo:*

(a) $\frac{1}{n} S(p) \stackrel{a}{=} \frac{1}{n} S^*(p)$, *donde* $S^*(p) = \tilde{\Im}(\tilde{p} - p)$, *siendo* $\widetilde{p}$ *la Ml-estimación de* p *en el modelo saturado.*

(b) $\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}}{\partial p^2}\right) \stackrel{a}{=} \frac{1}{n}\tilde{\Im}^*$. *Es decir, para todo $j = 1, 2, \ldots, J$ y todo $r, r' = 0, 1$ se tiene que*

$$\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} - \frac{1}{n} E\left(\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}}\right) \xrightarrow{P} 0.$$

(c) $Z^* \xrightarrow{d} \mathcal{N}_{2J}(0, I)$.

(d) $\frac{1}{\sqrt{n}} S(p) \xrightarrow{d} \mathcal{N}_{2J}(0, \tilde{\Xi})$ *siendo* $\tilde{\Xi} = \lim_{n \to \infty} \left(\frac{\tilde{\Im}}{n}\right)$.

Aquí, $\stackrel{a}{=}$ significa equivalencia asintótica; $\xrightarrow{P}$; convergencia en probabilidad; $\xrightarrow{d}$, convergencia en distribución; $\mathcal{N}_{2J}$, la distribución normal $2J$-dimensional e I, la matriz identidad $2J$-dimensional.

Demostración.

(a) Dado que $Z_{rj} = n_j \tilde{p}_{rj}$; $r = 0, 1$; se tiene que $S(p)$ es el vector

$$\left(\frac{n_1(\tilde{p}_{01} - p_{01})}{v_{01}}, \cdots, \frac{n_J(\tilde{p}_{0J} - p_{0J})}{v_{0J}}, \frac{n_1(\tilde{p}_{11} - p_{11})}{v_{11}}, \cdots, \frac{n_J(\tilde{p}_{1J} - p_{1J})}{v_{1J}} \right)^T.$$

Se deduce que $S(p) = [diag(\tilde{\Im})](\tilde{p} - p)$. Sea $S^*(p) := \tilde{\Im}(\tilde{p} - p)$. Entonces,

$$\frac{S(p) - S^*(p)}{n} = \begin{bmatrix} \frac{n_1}{nv_{01}v_{11}} p_{01}p_{11}(\tilde{p}_{11} - p_{11}) \\ \vdots \\ \frac{n_J}{nv_{0J}v_{1J}} p_{0J}p_{1J}(\tilde{p}_{1J} - p_{1J}) \\ \frac{n_1}{nv_{01}v_{11}} p_{01}p_{11}(\tilde{p}_{01} - p_{01}) \\ \vdots \\ \frac{n_J}{nv_{0J}v_{1J}} p_{0J}p_{1J}(\tilde{p}_{0J} - p_{0J}) \end{bmatrix}.$$

Para un j fijo se tiene,

$$\lim_{n_j \to \infty} \frac{n_j p_{0j} p_{1j}}{n v_{0j} v_{1j}} (\tilde{p}_{rj} - p_{rj}) = \frac{1}{\sigma_{0j}^2} \frac{p_{0j}p_{1j}}{v_{1j}} \lim_{n_j \to \infty} \left(\frac{Z_{rj}}{n_j} - p_{rj} \right) \xrightarrow{P} 0.$$

Luego, $\frac{1}{n} S(p) \stackrel{a}{=} \frac{1}{n} S^*(p)$.

(b) Teniendo en cuenta A.1c, tenemos:

$$\frac{1}{n} \left(-\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} \right) - \frac{1}{n} E \left(-\frac{\partial^2 \mathcal{L}}{\partial p_{r'j} \partial p_{rj}} \right) = \\ = \pm \frac{1}{n} \left[\left(\frac{Z_{rj}}{n_j} - p_{rj} \right) \frac{n_j}{nv_{rj}} \frac{(1 - 2p_{rj})}{v_{rj}} \right].$$

Ahora, se sabe que $\left(\frac{Z_{rj}}{n_j} - p_{rj} \right) \xrightarrow{p} 0$ si $n_j \to \infty$ por la ley débil de los grandes números y que $\frac{n_j}{v_{rj}} \frac{(1-2p_{rj})}{v_{rj}} \xrightarrow{n_j \to \infty} \frac{1}{\sigma_{rj}^2} \frac{(1-2p_{rj})}{v_{rj}}$.
Entonces, $\left(\frac{Z_{rj}}{n_j} - p_{rj} \right) \frac{n_j}{nv_{rj}} \frac{(1-2p_{rj})}{v_{rj}} \xrightarrow{p} 0$.

(c) Como las variables Z_{rj} convergen a la distribución normal por el teorema central del límite. Luego, $Z \xrightarrow{d} \mathcal{N}_{2J}(m, V)$ cuando $n_j \to \infty$, donde

$$V = \left(\begin{array}{ccc|ccc} n_1 v_{01} & \cdots & 0 & -n_1 p_{01} p_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & n_J v_{0J} & 0 & \cdots & -n_J p_{0J} p_{1J} \\ \hline -n_1 p_{01} p_{11} & \cdots & 0 & n_1 v_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -n_J p_{0J} p_{1J} & 0 & \cdots & n_J v_{1J} \end{array}\right).$$

Entonces,

$$Z^* = \tilde{\Im}^{-1/2}(V^*)^{-1}(Z - m) \sim \mathcal{N}_{2J}(0, I).$$

Por lo tanto, para J fijo, $Z^* \xrightarrow{d} \mathcal{N}_{2J}(0, I)$ cuando $n \to \infty$.

(d) Considerando el teorema A.1d, se tiene que $(1/\sqrt{n})S(p) = \left(\frac{\tilde{\Im}}{n}\right)^{1/2} Z^*$. Además, por el supuesto $\left(\frac{\tilde{\Im}}{n}\right)^{1/2} \xrightarrow{cs} \tilde{\Xi}^{1/2}$, donde $\xrightarrow{cs}$ significa convergencia casi segura. Por lo tanto, $\left(\frac{\tilde{\Im}}{n}\right)^{1/2} \xrightarrow{P} \tilde{\Xi}^{1/2}$. Luego, por el teorema 2.1.2 de Richard [8], se cumple que $\frac{1}{\sqrt{n}} S(p) \xrightarrow{d} \mathcal{N}_{2J}(0, \tilde{\Xi})$. ∎

4 Score e información para el modelo logístico

Análogo a la sección anterior, en ésta se presentan resultados asintóticos tanto para el vector score como para la matriz de información en el modelo logístico, resaltando también el hecho de se tiene el caso de variables independientes pero no idénticamente distribuidas, dando los detalles que, así reunidos no se encuentran en la literatura. Estos resultados también son importantes para las pruebas de comparación que se describirán en la sección 6.

Teorema 4.1 *Sean n_j y p_{rj} como se explicó en la sección 5 de Llinás [7] y $v_{rj} := p_{rj}(1 - p_{rj})$. Además, sean $S(\alpha)$ y $\Im$ como en el apéndice A.2. Considerando los supuestos de los modelos saturado y logístico se tiene:*

(a) La matriz $\Im$ es de rango completo, $Rg(\Im) = 2(1 + K)$, y definida positiva.

(b) Supóngase que $\lim_{n_j \to \infty} \frac{n_j}{n v_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$ *existe. Entonces existe una matriz* Ξ, *definida positiva y de tamaño* $2(1+K) \times 2(1+K)$ *tal que*

$$\frac{1}{\sqrt{n}} S(\alpha) \xrightarrow{d} N_{2(1+K)}(0, \Xi), \; n \to \infty, \quad J \textit{ fijo.}$$

Aquí, $\xrightarrow{d}$ *significa convergencia en distribución y* $\mathcal{N}_{2J}$*, la distribución normal* $2J$*-dimensional.*

Demostración.

(a) Teniendo en cuenta que $Rg(C) = 1 + K$, que $\Im$ es cuadrada de $2(1+K)$ y que $Rg\left(A^T A\right) = Rg(A)$, tenemos

$$Rg(\Im) = Rg\left[\left(\left(I_2 \otimes C^T\right) V^{\frac{1}{2}}\right)\left(V^{\frac{1}{2}}(I_2 \otimes C)\right)\right] = 2(1+K).$$

Aquí, $\otimes$ es el símbolo para el producto Kronecker de matrices e I_2, la matriz identidad de orden 2. Esto indica que $\Im$ es de rango completo $2(1+K)$. Ahora se debe probar que $\Im$ es definida positiva. Sea $u \neq 0$ cualquier vector columna de $2(1+K)$. Entonces,

$$\begin{aligned} u^T \Im u &= u^T \left[\left(I_2 \otimes C^T\right) V (I_2 \otimes C)\right] u \\ &= \left(V^{\frac{1}{2}}(I_2 \otimes C) u\right)^T \left(V^{\frac{1}{2}}(I_2 \otimes C) u\right) \end{aligned}$$

pero $V^{\frac{1}{2}}(I_2 \otimes C) u$ es un vector columna de $2J$. Por lo tanto, para todo $u \neq 0$, se cumple que $u^T \Im u \geq 0$. Pero, $u^T \Im u = 0$ si y solo si se cumple que $V^{\frac{1}{2}}(I_2 \otimes C) u = 0$. Ahora,

- Para $r = r'$ los componentes $n_j v_{rj} > 0$, $v_{rj} = p_{rj}(1 - p_{rj})$.
- Para $r \neq r'$, los componentes $-n_j p_{rj} p_{r'j} < 0$.

Luego, $u = 0$. Entonces, $\forall u \neq 0$, $u^T \Im u > 0$. Por tanto, $\Im$ es definida positiva.

(b) Sea $\lambda := (\lambda_0, \cdots, \lambda_{2K+1})^T$ cualquier vector de números reales. Se desea probar que

$$\lambda^T \left(\frac{1}{\sqrt{n}} S(\alpha)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda^T S_i(\alpha)$$

tiene distribución asintótica normal unidimensional, donde $S_i(\alpha)$ es el vector score de la observación i. En efecto,

- Primero, tenemos que

$$E\left(\lambda^T S_i\left(\alpha\right)\right)=\lambda^T E\left(S_i\left(\alpha\right)\right)=\lambda^T\left(0\right)=0.$$

- La matriz de información $\Im_i$ (para una observación Y_i en la población j) correspondiente al modelo logístico viene dada por

$$\Im_i=Cov\big((S_i)(\alpha)\big)=E\{S_i(\alpha)\left[S_i(\alpha)\right]^T\}=C^* v_{rj}$$

donde $v_{rj}=V(U_{ri})$ y

$$C^*=\begin{pmatrix} x_{0j}^2 & \dots & x_{0j}x_{kj} & x_{0j}^2 & \dots & x_{0j}x_{kj} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{kj}x_{0j} & \dots & x_{kj}^2 & x_{kj}x_{0j} & \dots & x_{kj}^2 \\ x_{0j}^2 & \dots & x_{0j}x_{kj} & x_{0j}^2 & \dots & x_{0j}x_{kj} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{kj}x_{0j} & \dots & x_{kj}^2 & x_{kj}x_{0j} & \dots & x_{kj}^2 \end{pmatrix}.$$

Ahora, se sabe que $\Im_i$ es definida positiva $\forall i$ en j. Por lo tanto,

$$V\left(\lambda^T S_i\left(\alpha\right)\right)=\lambda^T Cov\left(S_i\left(\alpha\right)\right)\lambda=\lambda^T\Im_i\lambda>0.$$

Es claro que $V=V^*\widetilde{\Im}V^*$ donde V^* es como en el teorema A.1d. Ahora, $\frac{1}{n}\Im=\left[\left(I_2\otimes C^T\right)V^*\right]\frac{\widetilde{\Im}}{n}\left[V^*\left(I_2\otimes C\right)\right]$. Por consiguiente, por el teorema 3.3.1b de Tilano [9], se tiene que $\frac{1}{n}\widetilde{\Im}\underset{n\longrightarrow\infty}{\longrightarrow}\widetilde{\Xi}$. Entonces, $\frac{1}{n}\Im\underset{n\longrightarrow\infty}{\longrightarrow}\left[\left(I_2\otimes C^T\right)V^*\right]\widetilde{\Xi}\left[V^*\left(I_2\otimes C\right)\right]$. Ahora, sea $\Xi:=\left[\left(I_2\otimes C^T\right)V^*\right]\widetilde{\Xi}\left[V^*\left(I_2\otimes C\right)\right]$. Luego, con todo lo anterior, se obtiene que $\frac{1}{n}\Im\underset{n\longrightarrow\infty}{\longrightarrow}\Xi$ y $Rg\left(\Xi\right)=2\left(1+K\right)$.

Ahora se debe probar que $\frac{1}{\sqrt{n}}S\left(\alpha\right)\xrightarrow{d}N_{2(k+1)}\left(0,\Xi\right),\ n\to\infty,\quad J$ fijo. Para el caso no agrupado , esta matriz cumple $\frac{1}{J}\Im\left(\alpha\right)\overset{J\to\infty}{\longrightarrow}\Xi$. Considerando lo anterior y sabiendo que Ξ es definida positiva se tiene

$$\frac{1}{n}\sum_{i=1}^{n}V\left(\lambda^T S_i\left(\alpha\right)\right)=\lambda^T\frac{1}{n}\Im\lambda\underset{n\to\infty}{\longrightarrow}\lambda^T\Xi\lambda>0.$$

Además se cumple la condición de Lindberg, es decir,

$$\forall\varepsilon>0,\ \frac{1}{n}\sum_{i=1}^{n}E_{\lambda^T S_i(\alpha)}\left(\left[\lambda^T S_i\left(\alpha\right)\right]^2\cdot 1_{\left\{\left[\lambda^T S_i(\alpha)\right]^2>\varepsilon^2 n\right\}}\right)\underset{n\to\infty}{\longrightarrow}0.$$

Por el teorema 1.1.3 de Llinás [5] se tiene

$$\lambda^T \left(\frac{1}{\sqrt{n}} S\left(\alpha\right) \right) \xrightarrow{d} N_1 \left(0, \lambda^T \Xi \lambda\right).$$

Por consiguiente, al aplicar el teorema central del límite multivariado (teorema 1.1.4 de Llinás [5]) se concluye que:

$$\frac{1}{\sqrt{n}} S\left(\alpha\right) \xrightarrow{d} N_{2(1+K)} \left(0, \Xi\right), \ n \to \infty, \ \text{para } J \text{ fijo.} \ \blacksquare$$

5 Estimación de parámetros logísticos

5.1 Existencia y cálculos de parámetros logísticos

En esta sección se presenta y demuestra el teorema de existencia de las ML-estimaciones de los parámetros logísticos, explicando brevemente el método iterativo que se utiliza para calcularlas: el método de Newton-Raphson.

Teorema 5.1 *(teorema de existencia) Sea Z como en la sección 5 de Llinás [7], $m := E(Z)$ y $V := Cov(Z)$. Entonces, las ML-estimaciones $\hat{\alpha}$ de α existen, son únicas y son calculadas según la siguiente fórmula de recursión*

$$\begin{aligned} \hat{\alpha}^{(0)} &= 0 \\ \hat{\alpha}^{(t+1)} &= \hat{\alpha}^{(t)} + \left[(I_2 \otimes C^T) \hat{V}^{(t)} (I_2 \otimes C) \right]^{-1} (I_2 \otimes C^T)(Z - \hat{m}^{(t)}). \end{aligned}$$

Además asintóticamente tenemos:

$$\begin{aligned} \sqrt{n}(\hat{\alpha} - \alpha) &\stackrel{a}{=} Cov^{-1} \left(\frac{1}{\sqrt{n}} \cdot \frac{\partial \mathcal{L}}{\partial \alpha} \right) \cdot \left(\frac{1}{\sqrt{n}} \cdot \frac{\partial \mathcal{L}}{\partial \alpha} \right) \\ &= \sqrt{n} \left[(I_2 \otimes C^T) V (I_2 \otimes C) \right]^{-1} (I_2 \otimes C^T)(Z - m). \end{aligned}$$

Aquí, $\stackrel{a}{=}$ significa equivalencia asintótica, $\otimes$ es el símbolo para el producto Kronecker de matrices e I_2, la matriz identidad de orden 2.

Demostración. Se sabe que $\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\Im$ es una matriz de rango completo $2(1+K)$.

Entonces existen únicamente las ML-estimaciones $\hat{\alpha}$ como soluciones de las $2(1+K)$ ecuaciones $S(\alpha) := \frac{\partial \mathcal{L}}{\partial \alpha} = (I_2 \otimes C^T)(Z - m) = 0$ (véase teorema A.1a), por lo que debe cumplirse que $\frac{\partial \mathcal{L}(\hat{\alpha})}{\partial \alpha} = 0$.

Para un $k = 0, \ldots, K$ fijo,

$$\frac{\partial \mathcal{L}}{\partial \beta_{rk}} = \sum_{j=1}^{J} x_{kj}(Z_{rj} - n_j p_{rj}) = \sum_{j=1}^{J} x_{kj} \left(Z_{rj} - n_j \frac{e^{g_r(x_j)}}{1 + e^{g_0(x_j)} + e^{g_1(x_j)}} \right).$$

Usando la aproximación de Taylor, si α_1 es un punto que está entre α y $\widehat{\alpha}$, entonces,

$$\frac{\partial \mathcal{L}(\alpha)}{\partial(\alpha)} = \frac{\partial \mathcal{L}(\hat{\alpha})}{\partial \alpha} + \frac{\partial^2 \mathcal{L}(\alpha_1)}{\partial \alpha^2}(\alpha - \hat{\alpha}).$$

Considerando que $\frac{\partial \mathcal{L}(\hat{\alpha})}{\partial(\alpha)} = 0$, luego, teniendo en cuenta el teorema A.1, esta expresión puede escribirse como:

$$\begin{aligned}
\hat{\alpha} - \alpha &= \left(\frac{-\partial^2 \mathcal{L}(\alpha_1)}{\partial \alpha^2} \right)^{-1} \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} \\
&= \left[(I_2 \otimes C^T) V_1 (I_2 \otimes C) \right]^{-1} (I_2 \otimes C^T)(Z - m)
\end{aligned}$$

siendo $V_1 = V(\alpha_1)$. Como α_1 es un punto del segmento de línea que une a α y $\hat{\alpha}$ entonces,

$$\alpha_1 = t\hat{\alpha} + (1-t)\alpha \qquad t \in [0,1].$$

Bajo el supuesto de que $\widehat{\alpha}$ es fuertemente consistente para α, es decir, $\hat{\alpha} \xrightarrow{c.s} \alpha$, cuando $n \to \infty$. Por componentes se tiene que $\alpha_1 \xrightarrow{c.s} \alpha$. Entonces, es válido que $\alpha_1 \xrightarrow{p} \alpha,\ n \to \infty$ (propiedad de convergencia) por componentes. Ahora,

$$\sqrt{n}(\hat{\alpha} - \alpha) = \left[(I_2 \otimes C^T) \frac{1}{n} V_1 (I_2 \otimes C) \right]^{-1} \cdot \frac{1}{\sqrt{n}} (I_2 \otimes C^T)(Z - m).$$

Como $\alpha_1 \xrightarrow{p} \alpha$, entonces

$$\begin{aligned}
&\underbrace{\left[(I_2 \otimes C^T) \frac{1}{n} V_1 (I_2 \otimes C) \right]^{-1} \cdot \frac{1}{n} (I_2 \otimes C^T)(Z - m)}_{\sqrt{n}(\hat{\alpha}-\alpha)} \\
&\qquad \stackrel{a}{=} \underbrace{\left[(I_2 \otimes C^T) \frac{1}{n} V (I_2 \otimes C) \right]^{-1} \frac{1}{\sqrt{n}} (I_2 \otimes C^T)(Z - m).}_{\sqrt{n}[(I_2 \otimes C^T) V (I_2 \otimes C)]^{-1} \cdot (I_2 \otimes C^T)(Z-m)}
\end{aligned}$$

Es decir,

$$\hat{\alpha}^{(t+1)} = \hat{\alpha}^{(t)} + \left[(I_2 \otimes C^T) V^{(t)} (I_2 \otimes C) \right]^{-1} (I_2 \otimes C^T)(Z - \hat{m}^{(t)}). \ \blacksquare$$

5.2 Distribuciones asintóticamente normales y chi-cuadradas

En esta sección, se presentan y demuestran teoremas asintóticos (normales y chi-cuadradas) que se necesitarán en la sección 6 para realizar pruebas de comparación.

Distribuciones asintóticamente normales

Teorema 5.2 *Supóngase que* $\lim\limits_{n_j\to\infty}\frac{n_j}{nv_{rj}} = \frac{1}{\sigma^2_{rj}} > 0$ *existe. Entonces, existe una matriz* $\tilde{\Xi}$ *definida positiva tal que* $\sqrt{n}(\tilde{p}-p) \xrightarrow{d} \mathcal{N}_{2J}(0, \widetilde{\Xi}^{-1})$, $n\to\infty$, J *fijo, siendo* $\tilde{p}$ *la estimación de* p *en el modelo saturado y*

$$\tilde{\Xi} = \left(\begin{array}{cccc|cccc}
\frac{1}{\sigma^2_{01}} & 0 & \cdots & 0 & -\frac{p_{01}p_{11}}{\sigma^2_{01}v_{11}} & 0 & \cdots & 0 \\
0 & \frac{1}{\sigma^2_{02}} & \cdots & 0 & 0 & -\frac{p_{02}p_{12}}{\sigma^2_{02}v_{12}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\sigma^2_{0J}} & 0 & 0 & \cdots & -\frac{p_{0J}p_{1J}}{\sigma^2_{0J}v_{1J}} \\
\hline
-\frac{p_{01}p_{11}}{\sigma^2_{01}v_{11}} & 0 & \cdots & 0 & \frac{1}{\sigma^2_{11}} & 0 & \cdots & 0 \\
0 & -\frac{p_{02}p_{12}}{\sigma^2_{02}v_{12}} & \cdots & 0 & 0 & \frac{1}{\sigma^2_{12}} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & -\frac{p_{0J}p_{1J}}{\sigma^2_{0J}v_{1J}} & 0 & 0 & \cdots & \frac{1}{\sigma^2_{1J}}
\end{array}\right).$$

Demostración. Se sabe que $\frac{1}{n}S(p) \stackrel{a}{=} \frac{1}{n}\tilde{\Im}(\tilde{p}-p)$. Entonces, $\frac{1}{n}\tilde{\Im}(\tilde{p}-p) \stackrel{a}{=} \frac{1}{n}\tilde{\Im}^{1/2}Z^*$. Por lo tanto, $\sqrt{n}(\tilde{p}-p) \stackrel{a}{=} \left(\frac{\tilde{\Im}}{n}\right)^{-1/2}Z^*$. Ahora por hipótesis, tenemos que $\left(\frac{\tilde{\Im}}{n}\right)^{-1/2}Z^* \xrightarrow{n\to\infty} \tilde{\Xi}^{-1/2}Z^*$. Además, $Z^* \xrightarrow{d} \mathcal{N}_{2J}(0,I)$. Entonces, se tiene que $\tilde{\Xi}^{-1/2}Z^* \xrightarrow{d} \mathcal{N}_{2J}(0,\widetilde{\Xi}^{-1})$, cuando $n\to\infty$, y J es fijo. Aplicando el teorema 2.2.1.c. de Tilano [9] resulta $\sqrt{n}(\tilde{p}-p) \xrightarrow{d} \mathcal{N}_{2J}(0,\tilde{\Xi}^{-1})$. ■

Teorema 5.3 *Para los datos no agrupados (se supone que* $\frac{1}{J}\Im$ *tenga un límite* Ξ *definida positiva cuando* $J\to\infty$*) son válidas las siguientes afirmaciones para* $J\to\infty$*:*

(a) *Si* $X^* := \Im^{-1/2}(I_2\otimes C^T)(U-m)$, *entonces,* $X^* \xrightarrow{d} \mathcal{N}_{2(1+K)}(0,I)$, $J\to\infty$.

(b) $\sqrt{J}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}_{2(1+K)}(0,\Xi^{-1})$.

(c) $\Im^{1/2}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}_{2(1+K)}(0,I)$, $J\longrightarrow\infty$.

(d) $\frac{\hat{\beta}_{rk}-\beta_{rk}}{\sqrt{\hat{V}(\hat{\beta}_{rk})}} \xrightarrow{d} \mathcal{N}_1(0,1);\, k=0,\ldots,K;\; r=0,1;\; \beta_{r0}=\delta_r.$

Demostración.

(a) Es claro que para el caso particular de datos no agrupados en donde $n_j = 1$, para todo j y $J = n$, se tiene que:

$$X^* = \Im^{-1/2}S(\alpha) = \left(\frac{1}{J}\Im\right)^{-1/2} \cdot \frac{S(\alpha)}{\sqrt{J}}.$$

Se sabe que $\left(\frac{1}{n}\Im\right) \xrightarrow{n\to\infty} \Xi$. En este caso, $n = J$. Entonces se cumple que $\left(\frac{1}{J}\Im\right)^{-1/2} \xrightarrow{J\to\infty} \Xi^{-1/2}$.

Se sabe por la observación análoga del teorema 3.8.2 de Llinás [5] que

$$\frac{1}{\sqrt{J}}S(\alpha) \xrightarrow{d} \mathcal{N}_{2(1+K)}(0,\Xi), \text{ cuando } J \to \infty.$$

Por lo tanto, por los teoremas T.1.1.1b y 1.3.1 de Llinás [5], se tiene

$$\left(\frac{1}{J}\Im\right)^{-1/2}\frac{S(\alpha)}{\sqrt{J}} \xrightarrow{d} \mathcal{N}_{2(k+1)}(0,I), J \to \infty.$$

(b) Se sabe por el teorema 5.1 que $\sqrt{n}(\hat{\alpha}-\alpha) \stackrel{a}{=} Cov^{-1}\left(\frac{1}{\sqrt{n}}\cdot\frac{\partial\mathcal{L}}{\partial\alpha}\right)\cdot\left(\frac{1}{\sqrt{n}}\frac{\partial\mathcal{L}}{\partial\alpha}\right)$.

Como $J = n$, entonces, $\sqrt{J}(\hat{\alpha}-\alpha) \stackrel{a}{=} J\,cov^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\frac{1}{\sqrt{J}}S(\alpha)$.

Por los teoremas 3.1b., T.1.1.1b.,c. y T.1.3.1 de Llinás [5], se tiene que $\sqrt{J}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}_{2(k+1)}(0,\Xi^{-1})$, cuando $J \to \infty$.

(c) Se sabe por el inciso (a) que:

$$\Im^{-1/2}(I_2 \otimes C^T)(U-m) \xrightarrow{d} \mathcal{N}_{2(1+k)}(0,I),\; J \to \infty.$$

Por el teorema 5.1 $\sqrt{J}(\hat{\alpha}-\alpha) \stackrel{a}{=} \sqrt{J}(\Im)^{-1}(I_2 \otimes C^T)(Z-m)$. Por lo tanto, $\Im^{1/2}(\hat{\alpha}-\alpha) \stackrel{a}{=} \Im^{-1/2}(I_2 \otimes C^T)(Z-m)]$. Por el teorema 1.1.1c de Llinás [5] se cumple que $\Im^{1/2}(\hat{\alpha}-\alpha) \xrightarrow{d} \mathcal{N}_{2(1+K)}(0,I)$, cuando $J \to \infty$.

(d) Al aplicar el hecho de que el vector aleatorio de las estimaciones $2(1+K)$ de los parámetros del modelo sigue una distribución normal multivariada se verifica que cada marginal de la forma $\frac{\hat{\beta}_{rk}-\beta_{rk}}{\sqrt{\hat{V}(\hat{\beta}_{rk})}}$ estandarizada sigue también una distribución univariada $\mathcal{N}_1(0,1)$. ■

Teorema 5.4 *Supóngase que* $\lim\limits_{n_j \to \infty} \dfrac{n_j}{n v_{rj}} = \dfrac{1}{\sigma_{rj}^2} > 0$ *existe. Entonces asintóticamente se tiene:*

(a) $(\widehat{\alpha} - \alpha)^T Cov^{-1}(\widehat{\alpha})(\widehat{\alpha} - \alpha) \xrightarrow{d} \chi^2_{2(1+K)}$, *cuando* $n \to \infty$.

(b) Para cada subparámetro de dimensión $s < 2(1 + K)$

$$(\widehat{\gamma} - \gamma)^T \widehat{Cov}^{-1}(\widehat{\gamma})(\widehat{\gamma} - \gamma) \xrightarrow{d} \chi^2(s), \quad n \to \infty$$

siendo $\widehat{Cov}(\widehat{\gamma})$ *la matriz de covarianzas estimada de* $\widehat{\gamma}$.

(c) En particular: $\dfrac{\left(\widehat{\beta}_{rk} - \beta_{rk}\right)^2}{\widehat{V}(\widehat{\beta}_{rk})} \xrightarrow{d} \chi^2(1)$*;* $k = 0, 1, \cdots, K$*;* $r = 0, 1$*;* $\beta_{r0} = \delta_r$, $n \to \infty$.

Demostración.

(a) Por el teorema 5.3b. para el caso de datos no agrupados se tiene que la matriz de covarianzas asintótica de $\sqrt{n}\widehat{\alpha}$ es Ξ^{-1}. Por lo tanto, de los teoremas 1.2.1b de Llinás [5] y 5.3b, se cumple que:

$$A := \widehat{Cov}^{-1/2}(\sqrt{n}\widehat{\alpha})[\sqrt{n}(\widehat{\alpha} - \alpha)] \xrightarrow{d} \mathcal{N}_{2(1+K)}(0, I), \quad n \to \infty.$$

De este modo, $A^T A = (\widehat{\alpha} - \alpha)^T \widehat{Cov}^{-1}(\widehat{\alpha})(\widehat{\alpha} - \alpha)$ converge en distribución a $\chi^2_{2(K+1)}$ cuando $n \to \infty$.

(b) Como γ es un subparámetro de α de dimensión s, se sigue que

$$(\widehat{\gamma} - \gamma)^T \widehat{Cov}^{-1}(\widehat{\gamma})(\widehat{\gamma} - \gamma) \xrightarrow{d} \chi^2(s).$$

(c) Del mismo modo, al particularizar 5.4 b, se sigue que:

$$\frac{\left(\widehat{\beta}_{rk} - \beta_{rk}\right)^2}{\widehat{V}(\widehat{\beta}_{rk})} \xrightarrow{d} \chi^2(1), n \to \infty. \ \blacksquare$$

6 Pruebas de comparación de modelos y selección de modelos

Con base en la teoría asintótica para las ML-estimaciones y el vector score, en esta sección, se encuentran aproximaciones para las diferentes deviaciones $-2\mathcal{L}(\hat{\theta})$. A partir de ellas, se obtienen estadísticas para distintas pruebas de comparación de modelos (Logístico vs Saturado, Submodelo vs Logístico y Nulo vs Logístico), con distribución asintótica chi-cuadrada. Estas pruebas de hipótesis sirven como criterios para escoger uno o varios submodelos de un modelo logístico sin perder información estadísticamente significativa.

6.1 Comparación de un modelo logístico con el modelo saturado correspondiente

Teorema 6.1 *Supóngase que* $\lim\limits_{n_j\to\infty} \frac{n_i}{n v_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$ *existe. Entonces, se cumple que* $2\left[\mathcal{L}\left(\tilde{p}\right) - \mathcal{L}\left(p\right)\right] \overset{a}{=} \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \tilde{\Im}^{-1} \left(\frac{\partial \mathcal{L}}{\partial p}\right)$*, donde* $\tilde{p}$ *es la ML-estimación de* p *en el modelo saturado.*

Demostración. Por la aproximación de Taylor, para cualquier par de puntos $\tilde{p}$ y p, existe un p_1, que está entre ellos, tal que:

$$\mathcal{L}\left(p\right) = \mathcal{L}\left(\tilde{p}\right) + \left[\frac{\partial \mathcal{L}\left(\tilde{p}\right)}{\partial p}\right]^T \left(p - \tilde{p}\right) + \frac{1}{2}\left(p - \tilde{p}\right)^T \frac{\partial^2 \mathcal{L}\left(p_1\right)}{\partial p^2}\left(p - \tilde{p}\right),$$

pero $\frac{\partial \mathcal{L}(\tilde{p})}{\partial p} = 0$ porque la ML-estimación $\tilde{p}$ es una solución del sistema de ecuaciones $\frac{\partial \mathcal{L}(p)}{\partial p} = 0$. Por consiguiente se tiene que

$$\begin{aligned} 2\left[\mathcal{L}\left(\tilde{p}\right) - \mathcal{L}\left(p\right)\right] &= \left(\tilde{p} - p\right)^T \left(\frac{-\partial^2 \mathcal{L}\left(p_1\right)}{\partial p^2}\left(\tilde{p} - p\right)\right) \\ &\overset{a}{=} \left[\tilde{\Im}^{-1}\frac{\partial \mathcal{L}}{\partial p}\right]^T \left(\frac{-\partial^2 \mathcal{L}\left(p_1\right)}{\partial p^2}\right)\left[\tilde{\Im}^{-1}\frac{\partial \mathcal{L}}{\partial p}\right], \end{aligned}$$

y por el teorema A.1c de Tilano [9] y la no singularidad de $\tilde{\Im}$,

$$2\left[\mathcal{L}\left(\tilde{p}\right) - \mathcal{L}\left(p\right)\right] \overset{a}{=} \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \tilde{\Im}^{-1}\left(-\frac{\partial^2 \mathcal{L}\left(p_1\right)}{\partial p^2}\right)\tilde{\Im}^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right).$$

Por ley fuerte de los grandes números, $\tilde{p}_{rj} = \frac{Z_{rj}}{n_j} \overset{c.s.}{\to} p_{rj}$, $n_j \to \infty$, para cada $r = 0, 1$ y $j = 1, 2, \cdots, J$. Dado que $\lim\limits_{n_j\to\infty} \frac{n_j}{n v_{rj}} = \frac{1}{\sigma_{rj}^2} > 0$ existe, se tiene que: $\tilde{p} \overset{c.s.}{\to} p,\ n \to \infty$ por componentes. Esto implica que $p_1 \xrightarrow{c.s.} p$ y por tanto, $p_1 \xrightarrow{p} p$, $n \to \infty$ por componentes, por consiguiente, para cada $r = 0, 1$ y $j = 1, 2, \cdots, J$ se tiene $\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_{r1})}{\partial p_{rj}^2}\right) - \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_r)}{\partial p_{rj}^2}\right) \overset{p}{\to} 0;\ n \longrightarrow \infty, J$ fijo. Esto implica que $\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p^2}\right) \overset{a}{=} \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p)}{\partial p^2}\right) \overset{a}{=} \frac{1}{n}\widetilde{\Im}^*$, por el teorema 3.1a de Tilano [9]. Entonces, $2\left[\mathcal{L}(\tilde{p}) - \mathcal{L}(p)\right] \overset{a}{=} \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \widetilde{\Im}^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right)$ ya que $\widetilde{\Im}^{-1}\widetilde{\Im}^*\widetilde{\Im}^{-1} \overset{a}{=} \widetilde{\Im}^{-1}$. ■

Teorema 6.2 *Para los casos de datos no agrupados y agrupados, es válido que:*

$$2\left[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}(\alpha)\right] \overset{a}{=} \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)^T Cov^{-1}\left(\frac{\partial \mathcal{L}}{\partial \alpha}\right) \cdot \frac{\partial \mathcal{L}}{\partial \alpha}$$

donde $\widehat{\alpha}$ es una ML-estimación consistente de α en el modelo logístico.

Demostración. Se sabe por la aproximación de Taylor que para cualquier par de puntos $\widehat{\alpha}$ y α, existe un α_2 entre ellos, tal que

$$\mathcal{L}(\alpha) = \mathcal{L}(\widehat{\alpha}) + \left[\frac{\partial \mathcal{L}(\widehat{\alpha})}{\partial \alpha}\right]^T (\alpha - \widehat{\alpha}) + \frac{1}{2}(\alpha - \widehat{\alpha})^T \cdot \frac{\partial^2 \mathcal{L}(\alpha_2)}{\partial \alpha^2} \cdot (\alpha - \widehat{\alpha}).$$

Pero, como $\frac{\partial \mathcal{L}(\widehat{\alpha})}{\partial \alpha} = 0$ por ser $\widehat{\alpha}$ la ML-estimación que satisface $\frac{\partial \mathcal{L}}{\partial \alpha} = 0$, se obtiene :

$$2\left[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}(\alpha)\right] = (\widehat{\alpha} - \alpha)^T \left(-\frac{\partial^2 \mathcal{L}(\alpha_2)}{\partial \alpha^2}\right) \cdot (\widehat{\alpha} - \alpha).$$

Como α_2 es un punto del segmento de línea que une α y $\widehat{\alpha}$, entonces ,

$$\alpha_2 = t\widehat{\alpha} + (1-t)\,\alpha\ t \in [0, 1].$$

Bajo el supuesto que $\widehat{\alpha}$ es fuertemente consistente para α, es decir, $\widehat{\alpha} \overset{c.s}{\to} \alpha$, cuando $n \to \infty$ por componentes. Se tiene que $\alpha_2 \xrightarrow{c.s} \alpha, n \to \infty$ por componentes. Por tanto, por el teorema A.1c. $\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(\alpha_2)}{\partial \alpha^2}\right) \overset{a}{=} \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(\alpha)}{\partial \alpha^2}\right) = \frac{1}{n}\Im(\alpha)$. Luego, $2\left[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}(\alpha)\right] \overset{a}{=} \sqrt{n}\,(\widehat{\alpha} - \alpha)^T \cdot \frac{\Im(\alpha)}{n}\sqrt{n}\,(\widehat{\alpha} - \alpha)$. Aplicando el teorema 5.1 se tiene que:

$$2\left[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}(\alpha)\right] \overset{a}{=} \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)^T \cdot Cov^{-1}\left(\frac{\partial \mathcal{L}}{\partial \alpha}\right) \cdot \frac{\partial \mathcal{L}}{\partial \alpha}. \ ■$$

Teorema 6.3 *La LR-estadística de prueba (según el método de cocientes de funciones de verosimilitud) para la hipótesis:*
H_0*: El modelo logístico es válido con todas las variables explicativas* $X_1, \cdots, X_K$
H_1 *: El Modelo saturado correspondiente es válido con sus* J *poblaciones.*
Es equivalente a la llamada desviación que tiene el modelo logístico del modelo saturado, $D^*(M) = 2\log\left(\frac{\mathcal{L}(\widetilde{p})}{\mathcal{L}(\widehat{\alpha})}\right) = 2\left[\mathcal{L}\left(\widetilde{p}\right) - \mathcal{L}\left(\widehat{\alpha}\right)\right]$*, con las siguientes característica:* $D^*\left(M\right) \xrightarrow{d} \chi^2\left[2J - 2\left(1+K\right)\right]$*, bajo* H_0*, cuando* $n \to \infty$ *y* J *es fijo.*

Demostración. En el teorema A.3b.c. se demostró que $I_{2J} - P_{2J}$ es una proyección con rango $2J - 2\left(1+K\right)$. Por consiguiente, existe una matriz ortogonal U de $m \times 2J$ tal que $I_{2J} - P_{2J} = U^T U$, con $m = 2J - 2\left(1+K\right)$. Considerando lo anterior y los teoremas A.1b. de Tilano [9] y 1.3.1 de Llinás [5], se tiene que $UZ^* \xrightarrow{d} \mathcal{N}_m\left(0, I\right), n \to \infty$, J fijo. Por consiguiente, por el teorema 1.3.2 de Llinás [5], se tiene que $\left(UZ^*\right)^T\left(UZ^*\right) \xrightarrow{d} \chi^2_m$, cuando $n \to \infty$ y para J fijo. ■

6.2 Comparación de un modelo logístico con algún sub-modelo.

El siguiente teorema muestra la prueba de comparación de un modelo logístico con un submodelo. Si, en el trabajo práctico, se ha llegado a un submodelo del modelo logístico inicial mediante un proceso de eliminación de variables explicativas, entonces, se espera que la prueba dada en este teorema no rechace H_o (p-valor alto). Se espera esto para poder reemplazar el modelo inicial por el submodelo. En caso contrario, la reducción significaría una pérdida de información estadísticamente significativa. Para la prueba se tendrán en cuenta los teoremas A.4 y A.5.

Teorema 6.4 *La LR- estadística de prueba para la situación señalada en el teorema A.4 es equivalente a la estadística*

$$D^*\left(L\right) := 2\log\left(\frac{L\left(\widehat{\alpha}\right)}{L\left(\widehat{\alpha_0}\right)}\right) = 2\left[\mathcal{L}\left(\widehat{\alpha}\right) - \mathcal{L}\left(\widehat{\alpha_0}\right)\right]$$

con las siguientes características:

(a) $D^*\left(L\right) \stackrel{a}{=} \left(X^*\right)^T\left(I_{2(1+K)} - P_{2(1+K)}\right)\left(X^*\right)$*, bajo* H_0.

(b) $D^*\left(L\right) \xrightarrow{d} \chi^2_{2\left(K-\widetilde{K}\right)}$*, bajo* H_0*,* $J \longrightarrow \infty$.

(c) $\widehat{\gamma}^T Cov^{-1}(\widehat{\gamma})\,\widehat{\gamma} \xrightarrow{d} \chi^2_{2\left(K-\widetilde{K}\right)}$*, bajo* H_0*,* $\quad J \longrightarrow \infty$.

Demostración.

(a) Análogamente el teorema 6.2 se tiene que

$$2\left[\mathcal{L}\left(\alpha\right)-\mathcal{L}\left(\widehat{\alpha_0}\right)\right]\overset{a}{=}\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right)^T Cov^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right)\cdot\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right).$$

Por tanto, $D^*\left(L\right)=2\left[\mathcal{L}\left(\widehat{\alpha}\right)-\mathcal{L}\left(\widehat{\alpha_0}\right)\right]$. Bajo H_0 vale que $\mathcal{L}\left(\alpha\right)=\mathcal{L}\left(\alpha_0\right)$, entonces

$$\begin{aligned}
&D^*\left(L\right)=\\
&\quad= 2\left[\mathcal{L}\left(\widehat{\alpha}\right)-\mathcal{L}\left(\alpha\right)-\mathcal{L}\left(\widehat{\alpha_0}\right)+\mathcal{L}\left(\alpha_0\right)\right]\\
&\quad\overset{a}{=} \left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T Cov^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\cdot\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)-\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right)^T Cov^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right)\cdot\left(\frac{\partial\mathcal{L}}{\partial\alpha_0}\right)\\
&\quad\overset{a}{=} \left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T Cov^{-1}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\cdot\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)-\left(T_0\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T\left(TT^T\right)^{-1}T_0\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\\
&\quad= \left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T\left\{\Im^{-1}-\left(T\Im^{-\frac{1}{2}}\right)^T\left(TT^T\right)^{-1}T\Im^{-\frac{1}{2}}\right\}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right).
\end{aligned}$$

Ahora, por el teorema A.5, se tiene que

$$\begin{aligned}
D^*\left(L\right) &= \left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T\left\{\Im^{-1}-\Im^{-\frac{1}{2}}P_{2(1+k)}\Im^{-\frac{1}{2}}\right\}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)\\
&= \underbrace{\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)^T\Im^{-\frac{1}{2}}}_{(X^*)^T}\left(I_{2(1+k)}-P_{2(1+k)}\right)\underbrace{\Im^{-\frac{1}{2}}\left(\frac{\partial\mathcal{L}}{\partial\alpha}\right)}_{X^*}.
\end{aligned}$$

(b) Puesto que $Rg\left(I_{2(1+k)}-P_{2(1+k)}\right)=2(1+\tilde{k})$, la matriz $I-P$ es cuadrada, semi-definida positiva $|I-P|=0\quad(\geq 0)$, entonces existe una matriz ortogonal $U_{m\times 1}$ tal que $I-P=UU^T$, con $m=2(k-\tilde{k})$. De (a) y el teorema 1.3.1 de [2], $UX^*\longrightarrow N_m(0,I)$ bajo $H_0, J\longrightarrow\infty, m=2(k-\tilde{k})$. Luego por el teorema 1.3.2 de [2], $D^*\left(L\right)=(UX^*)^T(UX)\overset{d}{\longrightarrow}\chi^2(m)$, bajo H_0, $J\longrightarrow\infty$.

(c) Se sabe que γ es un sub parámetro de α con dimensión $2(k-\tilde{k})$. Entonces, por el teorema 5.4b. $(\widehat{\gamma}-\gamma)^T Cov^{-1}(\widehat{\gamma})(\widehat{\gamma}-\gamma)\overset{d}{\longrightarrow}\chi^2_{2(k-\tilde{k})}$, $J\longrightarrow\infty$. Ahora, bajo H_0: $\gamma=0$, el término γ desaparece, con lo que se cumple que $\widehat{\gamma}^T Cov^{-1}(\widehat{\gamma})\widehat{\gamma}\overset{d}{\longrightarrow}\chi^2_{2(k-\tilde{k})}$. ■

Comparación de un modelo logístico con el modelo nulo

Teorema 6.5 *Para la hipótesis*
H_0*: El modelo logístico vale (sólo con el intercepto).*
H_1 *: El modelo logístico vale con* $X_1, \cdots, X_K$.
Se puede tomar alternativamente, una de las estadísticas de prueba siguientes:

(a) $2\log\left(\frac{\mathcal{L}(\widehat{\alpha})}{\mathcal{L}(\widehat{\alpha_0})}\right) = 2\left[\mathcal{L}(\widehat{\alpha}) - \mathcal{L}\left(\widehat{\delta_0}\right)\right] \xrightarrow{d} \chi^2(2K)$ *bajo* H_0, $J \longrightarrow \infty$, *donde* $\widehat{\delta}_0 = logit\left(\overline{Y}\right)$ *es la estimación de* δ *en el modelo nulo.*

(b) $\widehat{\beta}^T Cov^{-1}\left(\widehat{\beta}\right)\widehat{\beta} \xrightarrow{d} \chi^2(2K)$, *bajo* H_0, $J \longrightarrow \infty$.

Demostración.

(a) Es un caso particular del teorema 6.4 con $\widetilde{K} = 0$.

(b) β es un subparámetro de α con dimensión $2K$ por el teorema 5.4b. se tiene

$$\left(\widehat{\beta} - \beta\right)^T Cov^{-1}\left(\widehat{\beta}\right)\left(\widehat{\beta} - \beta\right) \xrightarrow{d} \chi^2(2K),\ J \longrightarrow \infty.$$

Bajo $H_0 : \beta = 0$, esta expresión queda reducida y queda demostrado. ■

Comparación de un modelo logístico con un submodelo que tiene una variable explicativa menos

Teorema 6.6 *Para la hipótesis*

H_0: *El submodelo vale con* $X_1, \cdots, X_K$ *sin un* X_k.
H_1 : *El Modelo logístico vale con* $X_1, \cdots, X_K$.
Se puede tomar, alternativamente, una de las dos estadísticas de prueba siguientes:

(a) $2\log\left(\frac{L(\hat{\alpha})}{L(\hat{\alpha}_0)}\right) = 2\left[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\hat{\alpha}_0)\right] \xrightarrow{d} \chi^2(1)$ *bajo* H_0, $J \to \infty$ *donde* $\hat{\alpha}_0$ *es la estimación bajo* H_0

(b) $\frac{\left(\hat{\beta}_k\right)^2}{\hat{V}\left(\hat{\beta}_k\right)} = \left(\frac{\hat{\beta}_k}{SE\left(\hat{\beta}_k\right)}\right)^2 \xrightarrow{d} \chi^2(1)$, *bajo* H_0, $J \to \infty$.

Demostración.

(a) Es un caso particular del teorema 6.4b con $\tilde{K} = K - 1$. Es decir $k - \hat{k} = 1$.

(b) Del teorema 5.4c. se tiene que para cada $k = 0, 1, \cdots, K$

$$\frac{\left(\hat{\beta}_k - \beta_k\right)^2}{\hat{V}\left(\hat{\beta}_k\right)} \xrightarrow{d} \chi^2\left(1\right), \quad J \to \infty.$$

Al considerar $H_0 : \ \beta_k = 0$ se reduce a $\frac{\left(\hat{\beta}_k\right)^2}{\hat{V}\left(\hat{\beta}_k\right)} \xrightarrow{d} \chi^2\left(1\right), \quad J \to \infty$ con lo que queda demostrado. ■

7 Conclusiones

En primera instancia, el desarrollo de este trabajo permitió dar nuevos fundamentos teóricos a los modelos logísticos en el caso particular de tres niveles y la metodología para generalizar los resultados cuando la variable respuesta toma un único valor de R posibles.

En general, se obtuvo un conjunto de pruebas de hipótesis y un análisis detallado de las pruebas de comparación y selección de modelos. Para esto, se volvió a definir el vector score y la matriz de información para los modelos logísticos y saturados heredando propiedades similares a los modelos con variable de respuesta dicotómica.

En el caso del modelo saturado existió la necesidad de recurrir a estimadores asintóticos y se tuvo que expresar los resultados, ya conocidos, en unos términos diferentes.

Referencias

[1] Agresti, A. (1990) *Categorical Data Analysis*. John Wiley and Sons, New York.

[2] Hosmer, D.; Lemeshow, S. (2000) *Applied Logistic Regresssion*, Segunda edición. John Wiley & Sons, New York.

[3] Jiménez, J. (2004) *Notas de clase Algebra lineal II (con aplicaciones en estadística)*. Unibiblos, Universidad Nacional de Colombia.

[4] Kleinbaum, D.; Klein, M. (2002) *Logistic Regression: A self Learning Text*, Segunda edición. Springer, Nueva York.

[5] Llinás, H. (1998) *Modelos Logísticos: Estimaciones, Pruebas de Hipótesis y Selección de Modelos*. Tesis de Maestría, Universidad del Valle, Cali, Colombia.

[6] Llinás H. (2006) "Precisiones en la teoría de modelos logísticos", *Revista Colombiana de Estadística* **29**(2): 293–265.

[7] Llinás, H.; Carreño, C. (2012) "El modelo logístico multinomial para el caso en que la variable de respuesta puede asumir uno de tres niveles y modelos relacionados", *Revista Colombiana de Estadística* **35**(1): 131–138.

[8] Richard, A.; Dean, W. (2002) *Applied Multivariate Statistical Analysis*, Fifth Edition. John Wiley & Sons, New York.

[9] Tilano, J. (2012) *Modelos Saturados: Estimaciones, Pruebas de Hipótesis y Selección de Modelos*. Tesis de Maestría. Universidad del Norte, Barranquilla, Colombia.

A Apéndice

El siguiente teorema presenta propiedades matriciales tanto para el vector score como para la matriz de información en el modelo saturado.

Teorema A.1 *Teniendo en cuenta las notaciones n_j, Z_{rj}, z_{rj}, p_{rj}, Z, p (introducidas en la sección 5 de Llinás [7]), $v_{rj} := p_{rj}(1-p_{rj})$ y $m := E(Z)$, en el modelo saturado se cumple que:*

(a) El vector (aleatorio) score de la muestra es

$$S(p) := \frac{\partial \mathcal{L}}{\partial p} = \begin{pmatrix} \frac{Z_{01}-n_1p_{01}}{v_{01}} \\ \vdots \\ \frac{Z_{0J}-n_Jp_{0J}}{v_{0J}} \\ \frac{Z_{11}-n_1p_{11}}{v_{11}} \\ \vdots \\ \frac{Z_{1J}-n_Jp_{1J}}{v_{1J}} \end{pmatrix}_{2J\times 1} .$$

Importante aclarar que en $\mathcal{L} = \mathcal{L}(p)$ se consideran las variables aleatorias U_i en lugar de los valores u_i. Además, $E(S(p)) = 0$.

(b) *La matriz de información de la muestra es* $\Im(p) := Cov(S(p))$, *de tamaño* $2J \times 2J$ *y está dada por*

$$\left[\begin{array}{ccc|ccc}
n_1/v_{01} & \cdots & 0 & -\frac{n_1p_{01}p_{11}}{v_{01}v_{11}} & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & n_J/v_{0J} & 0 & \cdots & -\frac{n_Jp_{0J}p_{1J}}{v_{0J}v_{1J}} \\
\hline
-\frac{n_1p_{01}p_{11}}{v_{01}v_{11}} & \cdots & 0 & n_1/v_{11} & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & -\frac{n_Jp_{0J}p_{1J}}{v_{0J}v_{1J}} & 0 & \cdots & n_J/v_{1J}
\end{array}\right]$$

la cual es semi-definida positiva. Para mayor simplicidad, sea $\tilde{\Im} := \Im(p)$.

(c) $E(-\frac{\partial^2\mathcal{L}}{\partial p^2}) = -\tilde{\Im}^*$, *donde*

$$\tilde{\Im}^* := \left[\begin{array}{cccc|cccc}
\frac{n_1}{v_{01}} & 0 & \cdots & 0 & -\frac{n_1}{v_{11}} & 0 & \cdots & 0 \\
0 & \frac{n_2}{v_{02}} & \cdots & 0 & 0 & -\frac{n_2}{v_{12}} & \cdots & 0 \\
\vdots & 0 & \ddots & \vdots & \vdots & 0 & \ddots & \vdots \\
0 & 0 & \cdots & \frac{n_J}{v_{0J}} & 0 & 0 & \cdots & -\frac{n_J}{v_{1J}} \\
\hline
-\frac{n_1}{v_{01}} & 0 & \cdots & 0 & \frac{n_1}{v_{11}} & 0 & \cdots & 0 \\
0 & -\frac{n_2}{v_{02}} & \cdots & 0 & 0 & \frac{n_2}{v_{12}} & \cdots & 0 \\
\vdots & 0 & \ddots & \vdots & \vdots & 0 & \ddots & \vdots \\
0 & 0 & \cdots & -\frac{n_J}{v_{0J}} & 0 & 0 & \cdots & \frac{n_J}{v_{1J}}
\end{array}\right].$$

(d) *Sea* $Z^* := \tilde{\Im}^{-1/2}(V^*)^{-1}(Z - m)$, *siendo*

$$V^* := diag\{v_{01}, \cdots, v_{0J}, v_{11}, \cdots, v_{1J}\}.$$

Entonces, $S(p) = (\tilde{\Im}^{1/2})Z^*$ *ó, que es equivalente,* $Z^* = (\tilde{\Im})^{-1/2}S(p)$.

Demostración. Sólo debemos aplicar resultados básicos del cálculo y del álgebra lineal y, obviamente, de la estadística (véase por ejemplo, Jiménez [3] y Richard [8]). ∎

Análogo al teorema anterior, el teorema siguiente presenta propiedades matriciales tanto para el vector score como para la matriz de información en el modelo logístico.

Teorema A.2 *Teniendo en cuenta las notaciones Z, C, m y α, introducidas anteriormente, en un modelo logístico se cumple:*

(a) El vector aleatorio score de la muestra es $S(\alpha) := \frac{\partial \mathcal{L}}{\partial \alpha} = (I_2 \otimes C^T)(Z - m)$ un vector columna de tamaño $2(1+K)$. Además, $E(S(\alpha)) = 0$. Aquí, $\otimes$ es el símbolo para el producto Kronecker de matrices.

(b) La matriz de información de la muestra viene dada por la expresión $\Im(\alpha) := Cov(S(\alpha)) = (I_2 \otimes C^T) V (I_2 \otimes C)$, matriz de tamaño $2(1+K) \times 2(1+K)$. Se denota $\Im := \Im(\alpha)$. Aquí, I_2 es la matriz identidad de orden 2.

(c) $\dfrac{\partial \mathcal{L}}{\partial \alpha} = \left(I_2 \otimes C^T\right) V^* \tilde{\Im}^{1/2} Z^*$, vector columna de tamaño $2\,(1+k)$. Aquí, $\tilde{\Im}$, V^* y Z^* son como en el teorema A.1.

(d) $E\left(\frac{\partial^2 \mathcal{L}}{\partial \alpha^2}\right) = \frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\Im$.

Demostración. Sólo se debe aplicar resultados básicos del cálculo y del álgebra lineal y, obviamente, de la estadística (véase por ejemplo, Jiménez [3] y Richard [8]). ■

Teorema A.3 *Son válidas las siguientes proposiciones:*

(a) La matriz $P_{2J} := \widetilde{T}^T \cdot \left(\widetilde{T} \cdot \widetilde{T}^T\right)^{-1} \cdot \widetilde{T}$ de tamaño $2J \times 2J$ es una proyección (en el sentido que es simétrica e independiente), siendo $\widetilde{T} := \left(I_2 \otimes C^T\right) V^* \tilde{\Im}^{1/2}$ una matriz de $2\,(1+K) \times 2J$.

(b) La matriz $I_{2J} - P_{2J}$ es una proyección, donde I_{2J} es la matriz idéntica de tamaño $2J \times 2J$.

(c) $Rg\,(I_{2J} - P_{2J}) = 2J - 2\,(1+K)$.

Demostración. Sólo se debe aplicar resultados básicos del cálculo y del álgebra lineal. ■

Teorema A.4 *Para la situación de hacer la hipótesis:*
H_0: Un sub-modelo logístico vale con $X_1, \cdots, X_{\widetilde{K}}$.
H_1 : El Modelo logístico vale con $X_1, \cdots, X_K$, $\widetilde{K} < K$.
Son válidas las siguientes afirmaciones:

(a) $T_0 := \dfrac{\partial \alpha}{\partial \alpha_0}$ *es una matriz de* $2(1+\widetilde{K}) \times 2\,(1+K)$ *de la forma:*

$$\frac{\partial \alpha}{\partial \alpha_0} = \left.\left(\underbrace{\begin{matrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{matrix}}_{2\left(1+\widetilde{K}\right)} \middle| \underbrace{\begin{matrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{matrix}}_{2\left(K-\widetilde{K}\right)} \right)\right\} 2\,(1+K).$$

Por tanto, $\frac{\partial \mathcal{L}}{\partial \alpha_0} = T_0 \cdot \frac{\partial \mathcal{L}}{\partial \alpha}$.

(b) $Cov\left(\frac{\partial \mathcal{L}}{\partial \alpha_0}\right) = TT^T$ *es una matriz cuadrada de tamaño* $2\left(1+\widetilde{K}\right)$*, siendo* $T := T_0 \Im^{\frac{1}{2}}$ *una matriz de* $2\left(1+\widetilde{K}\right) \times 2\,(1+K)$, α *el vector de los* $2\,(1+K)$ *parámetros en el modelo logístico y* α_0 *el vector* α *restringido bajo* $H_0 : \beta_{r,1+\widetilde{k}} = \beta_{r,2+\widetilde{k}} = \cdots = \beta_{r,K} = 0$*,* $r = 0, 1$.

Demostración. Sólo debemos considerar la regla de la cadena y el teorema 5.1. ■

Teorema A.5 *Para la situación señalada en el teorema A.4, son válidas las siguientes proposiciones:*

(a) *La matriz* $P_{2(1+K)} := T^T \left(TT^T\right)^{-1} T$ *de tamaño* $2\,(1+K) \times 2\,(1+K)$, *es una proyección con rango* $2\left(1+\widetilde{K}\right)$, *siendo* $T = T_0 \Im^{\frac{1}{2}}$.

(b) *La matriz* $I_{2(1+K)} - P_{2(1+K)}$ *es una proyección, donde* $I_{2(1+K)}$ *es la matriz idéntica de tamaño* $2\,(1+K) \times 2\,(1+K)$.

(c) $Rg\left(I_{2(1+k)} - P_{2(1+K)}\right) = 2\left(K - \widetilde{K}\right)$.

Demostración. Sólo se debe aplicar resultados básicos del cálculo y del álgebra lineal. ■