Revista Colombiana de Estadística
ISSN: 0120-1751
revcoles_fcbog@unal.edu.co
Universidad Nacional de Colombia
Colombia

Llinás, Humberto Jesús
Precisiones en la teoría de los modelos logísticos
Revista Colombiana de Estadística, vol. 29, núm. 2, diciembre, 2006, pp. 239-265
Universidad Nacional de Colombia
Bogotá, Colombia

Revista Colombiana de Estadística
Volumen 29 N° 2. pp. 239 a 265. Diciembre 2006

Precisiones en la teoría de los modelos logísticos

Accuracies in the Theory of the Logistic Models

Humberto Jesús Llinás[a]

Universidad del Norte, Barranquilla, Colombia

Resumen

Se estudian los modelos logísticos, como una clase de modelos lineales generalizados (MLG). Se demuestra un teorema sobre la existencia y unicidad de las estimaciones de máxima verosimilitud (abreviadas por ML) de los parámetros logísticos y el método para calcularlas. Con base en una teoría asintótica para estas ML-estimaciones y el vector *score*, se encuentran aproximaciones para las diferentes desviaciones $-2\log L$, siendo L la función de verosimilitud. A partir de ellas se obtienen estadísticas para distintas pruebas de hipótesis, con distribución asintótica chi-cuadrada. La teoría asintótica se desarrolla para el caso de variables independientes y no idénticamente distribuidas, haciendo las modificaciones necesarias para la conocida situación de variables idénticamente distribuidas. Se hace siempre la distinción entre datos agrupados y no agrupados.

Palabras clave: variable de respuesta binaria, modelo lineal generalizado, teoría asintótica.

Abstract

The logistic models are studied, as a kind of generalized lineal models. A theorem is showed about existence and uniqueness of ML-estimates of the estimation of the logistic regression coefficients and the method in order to calculate it. According to an asymptotic theory for this ML-estimates and the score vector, it has been founded approaching for different deviations $-2\log L$ (in this expression, L is the function of maximum likelihood). In consequence, we have gotten statistics for different hypotheses test which is asymptotically chi-square. The asymptotic theory is developed for the independent variables and no distributed identically variables. It is made the difference between ungrouped and grouped data.

Key words: Binary response, Generalized linear model, Asymptotic theory.

[a]Profesor. E-mail: hllinas@uninorte.edu.co

1. Introducción

Los modelos logísticos son adecuados para situaciones donde se quiere explicar la probabilidad p de ocurrencia de un evento de interés por medio de los valores de ciertas variables "explicativas". Si se asocia al evento de interés una variable dicotómica, entonces esta es una variable de Bernoulli con esperanza condicional p. En cambio, los modelos lineales cubren situaciones completamente diferentes. Estos quieren explicar la esperanza condicional de una variable aleatoria continua. Ambos tipos son casos particulares de los MLG. Además, con base en una teoría asintótica para las ML-estimaciones, se han encontrado aproximaciones para diferentes desviaciones, es decir, para (-2) veces los logaritmos de las ML-funciones. Estas se usan para obtener diferentes pruebas de hipótesis estadísticas que tienen distribuciones asintóticas chi-cuadrado. Por una parte, en algunos libros se mencionan estos resultados dando solo pocos detalles. Esto ocurre en el libro básico sobre MLG de Mc Cullagh & Nelder (1983). En Agresti (1990) ya se encuentran más detalles con mayor enfoque para el caso de variables explicativas categóricas. De todas formas, falta el desarrollo detallado de la teoría asintótica de ML-estimaciones para el caso de variables independientes y no idénticamente distribuidas. En los libros "clásicos" de Estadística Matemática, como Rao (1973) o Zacks (1971), se detalla solo el caso de variables independientes e idénticamente distribuidas. Esto último no se presenta en MLG. Por otra parte, muchos artículos originales, como Wedderburn (1974), Wedderburn (1976) o Mc Cullagh (1983), enfocan el concepto más general de funciones de "cuasi-verosimilitud", las cuales son relevantes para modelos logísticos en casos como problemas con mediciones repetidas. Además, es bien conocido que, para modelos lineales, existe una teoría exacta tanto para las estimaciones como para las pruebas de hipótesis. Por lo tanto, considerando la importancia de los modelos logísticos para muchas aplicaciones, este artículo desarrolla los detalles de los temas arriba mencionados. El artículo va más allá de ser un "estudio descriptivo" acerca de los modelos logísticos porque, como se dijo anteriormente, la teoría que se ha desarrollado no aparece así detallada en la literatura sino, en gran parte, solo esbozada. El artículo está compuesto de cinco secciones en las cuales se presenta un análisis teórico detallado sobre los modelos logísticos 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; se dan los detalles que, así reunidos, no se encuentran comúnmente en la literatura.

2. Modelos logísticos y modelos relacionados

2.1. El modelo de Bernoulli

Supongamos que la variable de interés Y es de Bernoulli. En símbolo, $Y \sim \mathcal{B}(1, p)$, siendo $p := E(Y) = P(Y = 1)$ la probabilidad de que ocurra Y. Haciendo n observaciones independientes de Y, se obtienen los datos $y_i \in \{0, 1\}$,

$i = 1, \ldots, n$, donde y_i es un posible valor de la variable muestral Y_i, las cuales son independientes entre sí. De esta forma, se llega a un modelo estadístico de Bernoulli:

$$Y_i = p_i + e_i \quad \sim \quad \mathcal{B}(1, p_i), \quad i = 1, \ldots, n$$

Fijando $y = (y_1, \ldots, y_n)^T$ obtenemos la función de verosimilitud en el parámetro $p = (p_1, \ldots, p_n)^T$:

$$L(p) = \prod_{i=1}^{n} [p_i^{y_i} (1 - p_i)^{1-y_i}]$$

y el logaritmo de la función de máxima verosimilitud será:

$$\mathcal{L}(p) := \log L(p) = \sum_{i=1}^{n} [y_i \log p_i + (1 - y_i) \log(1 - p_i)] \tag{1}$$

Como $0 \leq f(y, p) \leq 1$, se tiene que $-\infty \leq \mathcal{L}(p) \leq 0$. Hay varias situaciones que se pueden presentar en un modelo de Bernoulli. Se dice que este se puede identificar como alguno de los siguientes modelos: completo, nulo o saturado.

2.2. Los modelos completo y nulo

El *modelo completo* se caracteriza por el supuesto de que todos p_i, $i = 1, \ldots, n$ se consideran como parámetros.

Teorema 1. *En el modelo completo, las ML-estimaciones de p_i son $\hat{p}_i = Y_i$ con valores $\hat{p}_i = y_i$, $\quad i = 1, \ldots, n$. Además, $\mathcal{L}_c := \mathcal{L}(y) = 0$.*

Demostración. Considerando la ecuación (1) se tiene:

$$\mathcal{L}(p) = \sum_{\substack{i \\ y_i = 1}} \log p_i + \sum_{\substack{i \\ y_i = 0}} \log(1 - p_i)$$

Ahora, $\mathcal{L}(p) \stackrel{!}{=} 0$ si y solo si $p_i = y_i$, para cada $i = 1, \ldots, n$.

Esto demuestra la existencia de las ML-estimaciones. Si para algún i se tiene que $p_i \neq y_i$, entonces, $\mathcal{L}(p) < 0$. Esto último demuestra que las ML-estimaciones son únicas porque, si $\widetilde{p}$ es un vector que tiene por lo menos una componente p_i diferente de y_i, entonces, se tendría que $\mathcal{L}(\widetilde{p}) < \mathcal{L}_c$ (ya que al reemplazar $p_i = y_i$ en $\mathcal{L}(p)$ se obtiene que $\mathcal{L}_c = 0$). □

El *modelo nulo* se caracteriza por el supuesto de que todos los p_i, $i = 1, \ldots, n$ se consideran iguales; es decir, se tiene un solo parámetro $p = p_i, i = 1, \ldots, n$. En este caso, (1) será:

$$\mathcal{L}(p) \quad = \quad n[\overline{y} \ \log p + (1 - \overline{y}) \log(1 - p)] \tag{2}$$

Teorema 2. *En el modelo nulo, la ML-estimación de p es $\hat{p} = \overline{Y}$ con valor $\hat{p} = \overline{y}$ Además, $\mathcal{L}_o := \mathcal{L}(\overline{y}) < 0$ si y solo si $0 < \overline{y} < 1$.*

Demostración. De (2) se tiene, para $\overline{y} = 0$, que $\mathcal{L}(p) = 0$ si y solo si $p = 0$ y para $\overline{y} = 1$, que $\mathcal{L}(p) = 0$ si y solo si $p = 1$. Ahora, supongamos que $0 < \overline{y} < 1$.

De la ecuación (2) se puede demostrar que $\hat{p} = \overline{y}$ y que es única. Además, $\log \overline{y}$ y $\log(1 - \overline{y})$ son cantidades negativas, por lo tanto, $\mathcal{L}_o < 0$. □

2.3. El modelo saturado y supuesto

El modelo saturado se caracteriza por los siguientes supuestos:

1. Se supone que:

   a) Se tienen K variables explicativas $\mathbf{X}_1, \ldots, \mathbf{X}_K$ (algunas pueden ser numéricas y otras categóricas) con valores $x_{1i}, \ldots, x_{Ki}$ para $i = 1, \ldots, n$ (fijadas u observadas por el estadístico, según sean variables determinísticas o aleatorias).

   b) Entre las n K-uplas $(x_{1i}, \ldots, x_{Ki}), i = 1, \ldots, n$ de los valores de las variables explicativas $\mathbf{X}_1, \ldots, \mathbf{X}_K$ haya J K-uplas diferentes, definiendo las J poblaciones. Por tanto, $J \leq n$.

   Notación. Para cada población $j = 1, \ldots, J$ se denota:

   - El número de observaciones Y_{ij} en cada población j por n_j, siendo $n_1 + \ldots + n_J = n$;
   - La suma de las n_j observaciones Y_{ij} en j por $Z_j := \sum_{i=1}^{n_j} Y_{ij}$ con valor

$$z_j = \sum_{i=1}^{n_j} y_{ij}, \quad \text{siendo} \quad \sum_{j=1}^{J} z_j = \sum_{j=1}^{J} \sum_{i=1}^{n_j} y_{ij} = \sum_{i=1}^{n} y_i$$

   Para mayor simplicidad en la escritura, se abreviará la j-ésima población $(x_{1j}, \ldots, x_{Kj})$ por el símbolo $\star$.

2. Para cada población $j = 1, \ldots, J$ y cada observación $i = 1, \ldots, n$ en j, se supone que:

   - $(Y_{ij}|\star) \sim \mathcal{B}(1, p_j)$
   - Las variables $(Y_{ij}|\star)$ son independientes entre sí
   - $p_j = P(Y_{ij} = 1|\star) = E(Y_{ij}|\star)$ y $v_j := V(Y_{ij}|\star) = p_j(1 - p_j)$

A continuación se desarrolla el símbolo $\star$.

El supuesto 2 implica:

a) Todos los p_{ij}, $i = 1, \ldots, n$ dentro de cada población j son iguales. Es decir, se tiene como parámetro el vector $p = (p_1, \ldots, p_J)^T$.

b) Para cada población $j = 1, \ldots, J$ se tiene:

- $Z_j \sim \mathcal{B}(n_j, p_j)$
- Las variables Z_j son independientes entre las poblaciones
- $m_j := E(Z_j) = n_j p_j \quad$ y $\quad V_j := V(Z_j) = n_j v_j$

Escrito en forma vectorial, $Z := (Z_1, \ldots, Z_J)^T$ con valores reunidos en $z := (z_1, \ldots, z_J)^T$, tiene:

- $m := E(Z) = (n_1 p_1, \ldots, n_J p_J)^T$
- $V := Cov(Z) = diag\{n_1 v_1, \ldots, n_J v_J\}$, matriz diagonal de tamaño $J \times J$.

En el modelo saturado, el logaritmo de la función de máxima verosimilitud será

$$\begin{aligned} \mathcal{L}(p) &= \sum_{j=1}^{J} \left(\sum_{i=1}^{n_j} [y_{ij} \log p_j + (1 - y_{ij}) \log(1 - p_j)] \right) \\ &= \sum_{j=1}^{J} [z_j \log p_j + (n_j - z_j) \log(1 - p_j)] \end{aligned}$$

Teorema 3. *En el modelo saturado, las ML-estimaciones de* p_j *son* $\tilde{p}_j = \frac{Z_j}{n_j}$, *con valores* $\tilde{p}_j = \dfrac{z_j}{n_j}, \quad j = 1, \ldots, J.$

Además, $\mathcal{L}_s := \mathcal{L}(\tilde{p}) < 0$ *para* $0 < \tilde{p}_j < 1$.

Demostración. Si $0 < \tilde{p}_j < 1$, se tiene que:

$$\frac{\partial \mathcal{L}}{\partial p_j} = \frac{z_j}{p_j} + (-1)\frac{n_j - z_j}{1 - p_j} = 0 \quad \text{si y solo si} \quad \tilde{p}_j = \frac{z_j}{n_j}$$

Por consiguiente, si $0 < z_j < n_j$, se tiene

$$\left.\frac{\partial^2 \mathcal{L}}{\partial p_j^2}\right|_{p_j = \tilde{p}_j} = -\left[\frac{n_j^2}{z_j} + \frac{n_j^2}{n_j - z_j}\right] < 0$$

Falta analizar los dos casos extremos:

- Si $z_j = 0$, entonces $\dfrac{\partial \mathcal{L}}{\partial p_j} = -\dfrac{n_j}{1 - p_j}$ decrece en p_j. En este caso, $\mathcal{L}$ decrece en p_j; es decir, se hace maximal $\mathcal{L}(p)$ para $p_j = 0$.

- Si $z_j = n_j$, entonces $\frac{\partial \mathcal{L}}{\partial p_j} = \frac{n_j}{p_j}$ decrece en p_j. En este caso, $\mathcal{L}$ crece en p_j; es decir, se hace maximal $\mathcal{L}(p)$ para $p_j = 1$.

En el modelo saturado, se puede obtener el valor de $\mathcal{L}$ reemplazando, en la ecuación (3), cada p_j por $\tilde{p}_j, j = 1, \ldots, J$. Por lo tanto:

$$\mathcal{L}_s \quad = \quad \sum_{j=1}^{J} n_j[\tilde{p}_j \log \tilde{p}_j + (1 - \tilde{p}_j) \log(1 - \tilde{p}_j)]$$

Bajo la condición $0 < \tilde{p}_j < 1$ se puede afirmar que $\log \tilde{p}_j$ y $\log(1 - \tilde{p}_j)$ son cantidades negativas. Por consiguiente, la suma del lado derecho de la ecuación anterior es también negativa. □

2.4. El modelo logístico

Se hacen los supuestos 1 y 2 de la sección 2.3, donde adicionalmente se supone que la matriz de diseño

$$C = \begin{bmatrix} 1 & x_{11} & \ldots & x_{K1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1J} & \ldots & x_{KJ} \end{bmatrix}$$

tiene rango completo $Rg(C) = 1 + K \leq J$. Para llegar a un modelo logístico se hace el supuesto adicional

$$\log\left(\frac{p_j}{1 - p_j}\right) \quad = \quad \sum_{k=0}^{K} \beta_k x_{kj}, \quad \text{con} \quad \beta_o = \delta,\ x_{oj} = 1 \tag{3}$$

Sea $\alpha = (\delta, \beta_1, \ldots, \beta_K)^T$ el vector de parámetros en el modelo. Nótese que el supuesto sobre $Rg(C) = 1 + K$ hace identificable al parámetro α.

2.5. Score e información para los modelos saturado y logístico

Teorema 4. *En el modelo saturado se tiene:*

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

$$S(p) := \frac{\partial \mathcal{L}}{\partial p} = \begin{bmatrix} \dfrac{Z_1 - n_1 p_1}{v_1} \\ \vdots \\ \dfrac{Z_J - n_J p_J}{v_J} \end{bmatrix}$$

donde $\mathcal{L} = \mathcal{L}(p)$ es el "logaritmo de la función de máxima verosimilitud" aleatorio; es decir, entran las variables aleatorias Y_i en lugar de los valores y_i. Además, $E(S(p)) = 0$.

b) La matriz de información de la muestra es

$$\Im(p) := Cov(S(p)) = diag\{n_1/v_1, \ldots, n_J/v_J\},$$

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

c) $S(p) = \tilde{\Im} \cdot (\tilde{p} - p)$.

d) $E\left(-\frac{\partial^2 \mathcal{L}}{\partial p^2}\right) = \tilde{\Im}$, *siendo* $\frac{\partial^2 \mathcal{L}}{\partial p^2}$ *la matriz de las segundas derivadas parciales de* $\mathcal{L} = \mathcal{L}(p)$.

Sea $Z^* = (Z_1^*, \ldots, Z_J^*)^T$ *el vector de las variables* $Z_j^* := \frac{Z_j - n_j p_j}{\sqrt{n_j v_j}}$. *Entonces,*

e) $Z^* = V^{-1/2}(Z - m)$

f) $S(p) = \tilde{\Im}^{1/2} Z^* \quad o \quad Z^* = \tilde{\Im}^{-1/2} \cdot S(p)$, *donde* $\tilde{\Im}^{1/2}$ *es la matriz definida por la propiedad* $(\tilde{\Im}^{1/2}) \cdot (\tilde{\Im}^{1/2}) = \tilde{\Im}$, *siendo* $\tilde{\Im}$ *la matriz diagonal de* $J \times J$ *definida en 4 y cuyos elementos diagonales* n_j/v_j *son positivos.*

El j*-ésimo elemento diagonal de* $\tilde{\Im}^{1/2}$ *es* $\sqrt{n_j/v_j}$ *y* $\tilde{\Im}^{-1/2}$ *es la inversa de* $\tilde{\Im}^{1/2}$.

Demostración. Solo debemos aplicar resultados básicos del cálculo y del álgebra lineal y, obviamente, de la estadística (Dobson 2002). □

Teorema 5. *Supóngase que* $\lim_{n_j \to \infty} \frac{n_j}{n \cdot v_j} = \frac{1}{\sigma_j^2} > 0$ *existe para cada* $j = 1, \ldots, J$. *Entonces, para el modelo saturado, valen las siguientes afirmaciones asintóticas* $(n \to \infty)$, *considerando naturalmente* J *fijo:*

a) $\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}}{\partial p^2}\right) \stackrel{a}{=} \frac{1}{n}\tilde{\Im}$, *siendo* $\frac{\partial^2 \mathcal{L}}{\partial p^2}$ *la matriz de las segundas derivadas parciales de* $\mathcal{L} = \mathcal{L}(p)$ *y* $\tilde{\Im}$ *como en la parte d) del teorema 4. Es decir, para cada* $j = 1, \ldots, J$ *se tiene que*

$$\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial p_j^2} - \frac{1}{n}E\left(\frac{\partial^2 \mathcal{L}}{\partial p_j^2}\right) \xrightarrow{P} 0$$

b) $Z^* \xrightarrow{d} \mathcal{N}_J(0, I) \quad y \quad \frac{1}{\sqrt{n}} S(p) \xrightarrow{d} \mathcal{N}_J(0, \tilde{\Xi})$, *siendo*

$$\tilde{\Xi} := \Xi(p) = diag\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_J^2}\right)$$

una matriz definida positiva de tamaño $J \times J$.

Aquí, $\stackrel{a}{=}$ *significa equivalencia asintótica;* $\xrightarrow{P}$, *convergencia en probabilidad y* $\xrightarrow{d}$, *convergencia en distribución.*

Demostración. Considerando el teorema 4d) se tiene, para cada $j = 1, \ldots, J$, que

a)

$$\frac{1}{n}\frac{\partial^2 \mathcal{L}}{\partial p_j^2} - \frac{1}{n}E\left(\frac{\partial^2 \mathcal{L}}{\partial p_j^2}\right) \quad = \quad -\left[\left(\frac{Z_j}{n_j} - p_j\right) \cdot \frac{n_j}{n \cdot v_j} \cdot \frac{1-2p_j}{v_j}\right] \xrightarrow{P} 0$$

ya que, por la ley débil de los grandes números, $\left(\frac{Z_j}{n_j} - p_j\right) \xrightarrow{P} 0$, si $n_j \to \infty$ y, por el supuesto,

$$\frac{n_j}{n \cdot v_j} \cdot \frac{1-2p_j}{v_j} \xrightarrow{n_j \to \infty} \frac{1}{\sigma_j^2} \cdot \frac{1-2p_j}{v_j}$$

b) Como las variables Z_j^* son estandarizadas, entonces $Z_j^* \xrightarrow{d} \mathcal{N}_1(0,1)$, cuando $n_j \to \infty$. Por tanto, $Z^* \xrightarrow{d} \mathcal{N}_J(0, I)$, cuando $n \to \infty$, y J fijo. Ahora, se demostrará la otra parte del inciso b). Considerando el teorema 4f, se tiene que $\frac{1}{\sqrt{n}}S(p) = \left(\frac{1}{n}\tilde{\Im}\right)^{1/2} Z^*$. Entonces, por el supuesto y la parte b) de este teorema,

$$\left(\frac{n_j}{n \cdot v_j}\right)^{1/2} Z_j^* \xrightarrow{d} \mathcal{N}_1\left(0, \frac{1}{\sigma_j^2}\right)$$

Por tanto, $\left(\frac{1}{n}\tilde{\Im}\right)^{1/2} Z^* \xrightarrow{d} \mathcal{N}_J(0, \tilde{\Xi})$, cuando $n \to \infty$, y J fijo. □

Algunas observaciones importantes son las siguientes:

1. El supuesto del teorema 5 implica que

$$\frac{1}{n}\tilde{\Im} \quad \xrightarrow{n \to \infty} \quad \tilde{\Xi} \tag{4}$$

2. Los vectores *score* $S_i(p)$ correspondientes a las observaciones i son independientes, pero sus distribuciones no son idénticas ya que dependen de la población j a la cual i pertenece.

3. La matriz $\tilde{\Im}_i = diag\left\{0, \ldots, 0, \frac{1}{v_j}, 0, \ldots, 0\right\}$, que se refiere a una observación Y_i en j, no es definida positiva. En cambio, la matriz $\tilde{\Im}$ se refiere a toda la muestra y siempre es definida positiva.

4. Cuando se trabaja con el modelo saturado, se tiene el caso de utilizar *datos agrupados* porque las observaciones Y_i, $i = 1, \ldots, n$ se reúnen en J grupos (poblaciones) de tamaño n_j, $j = 1, \ldots, J$. En el caso especial $n_j = 1$, $j = 1, \ldots, J$ (lo que implica que $J = n$) se habla de *datos no agrupados*. La distinción entre datos agrupados y no agrupados es importante por dos razones:

- Algunos métodos de análisis apropiados para datos agrupados no son aplicables a datos no agrupados.
- Las aproximaciones asintóticas para datos agrupados pueden basarse en uno de estos dos casos distintos: $n_j \to \infty$ ($n \to \infty$) o $J \to \infty$. El último caso es apropiado únicamente para datos no agrupados.

5. El supuesto del teorema 5 se puede interpretar de la siguiente manera: "la velocidad" de cada $n_j \to \infty$ debe ser la misma que la de $n \to \infty$. Por ejemplo, en un diseño balanceado todas las n_j son iguales. En este caso, $n_j = \frac{n}{J}$. Por lo tanto, la cantidad $\frac{1}{n} \cdot \frac{n_j}{v_j} = \frac{1}{J \cdot v_j}$ es fija; es decir, no depende de n. Utilizando las notaciones del teorema anterior, se tendría que $\sigma_j^2 = J \cdot v_j$ porque, en este caso, la expresión (4) se convierte en una igualdad de la forma $\frac{1}{n}\tilde{\Im} = \tilde{\Xi}$.

6. En la práctica se puede suponer que el supuesto del teorema 5 siempre se cumple. Pero es importante resaltar que J debe ser fijo. Esta situación se presenta cuando se tienen datos agrupados con J fijo. Por esta razón, debe tomarse como "base" el modelo saturado. Es decir, se empieza con el *score* de la muestra usando los vectores Z_j, donde $j = 1, \ldots, J$.

   Además, si $J \to \infty$ (por ejemplo, si $J = n$), entonces en el modelo saturado no se puede considerar a J como fijo. Obsérvese que esta situación se presenta cuando se tienen datos no agrupados. En este caso, no se puede tomar como "base" el modelo saturado. Ahora se empezaría con el *score* de la muestra utilizando, de una vez, las observaciones Y_i, $i = 1, \ldots, n$. De ahora en adelante, cuando se trabaje con aproximaciones asintóticas para:

   - Datos agrupados y no agrupados, se utilizará únicamente la notación $n \to \infty$.
   - Datos agrupados, esta misma expresión pero acompañada de la expresión J *es fijo*. Esto es con el fin de enfatizar que el tamaño de la población J es fijo.
   - Datos no agrupados, la notación $J \to \infty$ en vez de $n \to \infty$. Esto es con el fin de enfatizar que el tamaño de la población J no es fijo.

Teorema 6. *Considerando las notaciones z, m y V de la sección 2.3 y C de la sección 2.4, se tiene en un modelo logístico:*

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

b) La matriz de información de la muestra es $\Im(\alpha) := Cov\left(\frac{\partial \mathcal{L}}{\partial \alpha}\right) = C^T V C$, matriz de $(1 + K) \times (1 + K)$. Para mayor simplicidad, sea $\Im := \Im(\alpha)$.

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

Para el caso particular de datos no agrupados, en donde $n_j = 1$, $\forall j$ y $J = n$, se tiene que: $Z = Y$, $m = (p_1, \ldots, p_n)^T$, C es la matriz de diseño original de $n \times (1 + K)$ y $V = diag\{v_1, \ldots, v_n\}$.

Demostración. Solo debemos aplicar resultados básicos del cálculo y del álgebra lineal y, obviamente, de la estadística. Cada uno de estos resultados se relaciona con Dobson (2002). □

Teorema 7. *Considerando los supuestos de las secciones 2.3 y 2.4, se tiene:*

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

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

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

Para el caso de datos no agrupados, en donde $J = n$*, el supuesto dado en b) no tiene sentido porque, como se explicó al final de la sección anterior,* J *no es fijo. Por lo tanto, en vez de esa condición, se debe suponer la existencia de una matriz definida positiva* Ξ *tal que* $\frac{1}{J}\Im \xrightarrow{J \to \infty} \Xi$*. De esta forma, se tiene*

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

Demostración.

a) Se sabe que la matriz $\Im$ es de tamaño $(1 + K) \times (1 + K)$.

Entonces,

$$Rg(\Im) = Rg([V^{1/2}C]^T \cdot [V^{1/2}C]) = Rg(V^{1/2}C) = 1 + K$$

Por consiguiente, $\Im$ es de rango completo $1 + K$. Demostremos que $\Im$ es definida positiva. En efecto, para cualquier vector columna $u \neq 0$, de tamaño $(1 + K)$ siempre se cumple que

$$0 \leq (V^{1/2}Cu)^T \cdot (V^{1/2}Cu) = u^T(C^T V C)u = u^T \Im u$$

Ahora, considerando el hecho de que $V^{1/2}$ es invertible y C de rango completo,

$$(V^{1/2}Cu)^T \cdot (V^{1/2}Cu) = 0 \Longleftrightarrow V^{1/2}Cu = 0 \Longleftrightarrow Cu = 0 \Longleftrightarrow u = 0.$$

Esto contradice el hecho de que $u \neq 0$. Por tanto, $\Im$ es definida positiva.

b) Se considera cualquier vector $\lambda := (\lambda_o, \ldots, \lambda_K)^T$ de números reales. Se demostrará, a continuación, que

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

tiene distribución asintótica normal 1-dimensional. Para ello se aplicará el teorema de Lindberg (véase, por ejemplo, Rao 1973 pp. 123(xi), 128(iii), 128(iv))

En efecto,

- $E(\lambda^T \cdot S_i(\alpha)) = 0$
- La matriz de información $\Im_i$ (para una observación Y_i en la población j) correspondiente al modelo logístico viene dada por:

$$\Im_i \;=\; Cov\{S_i(\alpha)\} \;=\; E\{S_i(\alpha)\cdot[S_i(\alpha)]^T\} \;=\; C^*\cdot v_j$$

siendo $v_j = V(Y_i)$ y

$$C^* := \begin{pmatrix} x_{oj}^2 & x_{oj}x_{1j} & \dots & x_{oj}x_{Kj} \\ x_{1j}x_{oj} & x_{1j}^2 & \dots & x_{1j}x_{Kj} \\ \vdots & \vdots & \ddots & \vdots \\ x_{Kj}x_{oj} & x_{Kj}x_{1j} & \dots & x_{Kj}^2 \end{pmatrix}$$

A diferencia del modelo saturado, la matriz de información $\Im_i$ sí es definida positiva puesto que $\Im$ también lo es y $\Im = \sum_i \Im_i$. Por tanto,

$$V(\lambda^T \cdot S_i(\alpha)) = \lambda^T \cdot Cov(S_i(\alpha)) \cdot \lambda = \lambda^T \cdot \Im_i \cdot \lambda > 0$$

- Es fácil verificar que $V = V^*\tilde{\Im}V^*$, donde $V^* = diag\{v_1, \dots, v_J\}$. Por consiguiente, por el teorema 6b, la ecuación (4) y la expresión anterior:

$$\frac{1}{n}\Im = C^T V^* \cdot \frac{1}{n}\tilde{\Im}\cdot V^* C \;\xrightarrow{n\to\infty}\; C^T V^* \cdot \tilde{\Xi}\cdot V^* C$$

siendo $\tilde{\Xi}$ como en el teorema 5b. Es decir,

$$\frac{1}{n}\Im \xrightarrow{n\to\infty} \Xi \tag{5}$$

donde $\Xi := C^T V^* \cdot \tilde{\Xi} \cdot V^* C$. La expresión que aparece en la ecuación (5) se cumple siempre y cuando $\lim_{n_j\to\infty} \frac{n_j}{n\cdot v_j} = \frac{1}{\sigma_j^2} > 0$ exista. La matriz Ξ también es de rango completo porque

$$Rg(\Xi) = Rg(C^T \cdot \underbrace{V^*\tilde{\Xi}V^*}_{D} \cdot C) = Rg([D^{1/2}C]^T[D^{1/2}C]) = 1 + K$$

Mediante un razonamiento análogo a la demostración de la parte a) del teorema 7, se obtiene que Ξ es definida positiva.[1] Considerando la ecuación (5) y sabiendo que Ξ es definida positiva, se tiene que

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

[1] Para el caso no agrupado, debe suponerse en seguida la existencia de una matriz Ξ definida positiva tal que $\frac{1}{J}\Im \xrightarrow{J\to\infty} \Xi$.

- Se verificará la condición de Lindberg. Es decir, para cada $\varepsilon > 0$,

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

donde $E_{\lambda^T \cdot S_i(\alpha)}$ significa que se calcula la esperanza con base en la distribución de $\lambda^T \cdot S_i(\alpha) = \left[\sum_{k=0}^{K} \lambda_k x_{kj}\right](Y_i - p_j)$. Suponiendo que las x_{kj} están acotadas uniformemente con respecto a j (que no parece restricción alguna para la práctica), entonces existe un $N = N(\varepsilon)$ tal que para cada $n \geq N$, $\left[\sum_{k=0}^{K} \lambda_k x_{kj}\right]^2 (y_i - p_j)^2 \leq \varepsilon^2 n$. De esta forma, se cumple la condición de Lindberg, porque

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

ya que la suma anterior no depende de n.

Aplicando el teorema de Lindberg, se tiene

$$\lambda^T \cdot \left(\frac{1}{\sqrt{n}} S(\alpha)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda^T \cdot S_i(\alpha) \xrightarrow{d} \mathcal{N}_1(0, \lambda^T \cdot \Xi \cdot \lambda),$$

Por consiguiente, al aplicar el teorema central del límite multivariado (véase, por ejemplo, Rao 1973 pp. 123(xi), 128(iii), 128(iv)), se concluye que $\frac{1}{\sqrt{n}} S(\alpha) \xrightarrow{d} \mathcal{N}_{1+K}(0, \Xi)$, cuando $n \to \infty$. □

3. Existencia y cálculos de parámetros logísticos

El método que se propone para calcular las ML-estimaciones en un modelo logístico es *el método iterativo de Newton-Raphson*, como se muestra en el siguiente teorema:

Teorema 8 (Teorema de existencia)**.** *Las ML-estimaciones $\hat{\alpha}$ de α existen, son únicas y se calculan según la siguiente fórmula de recursión:*

$$\hat{\alpha}^{(0)} = 0, \qquad \hat{\alpha}^{(t+1)} := \hat{\alpha}^{(t)} + (C^T \hat{V}^{(t)} C)^{-1} \cdot C^T (z - \hat{m}^{(t)})$$

Además, asintóticamente se tiene

$$\sqrt{n} \cdot (\hat{\alpha} - \alpha) \stackrel{a}{=} \left[Cov^{-1}\left(\frac{1}{\sqrt{n}} \cdot \frac{\partial \mathcal{L}}{\partial \alpha}\right)\right] \cdot \left(\frac{1}{\sqrt{n}} \cdot \frac{\partial \mathcal{L}}{\partial \alpha}\right) = \sqrt{n} \cdot (C^T V C)^{-1} \cdot C^T (z - m)$$

Para el caso particular de datos no agrupados, en donde $J = n$, se tiene que: $z = y$ y m, C y V son como se explicó al final del teorema 6.

Demostración. En la parte a) del teorema 7 se demostró que la matriz $\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = -\Im$ es de rango completo $1+K$. Por lo tanto, existen únicamente las ML-estimaciones $\hat{\alpha}$ como soluciones de las $1+K$ ecuaciones $\frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = 0$. O, lo que es lo mismo por la parte a) del teorema 6, teniendo $C^T(z-m) = 0$. Por lo tanto, debe cumplirse que

$$\frac{\partial \mathcal{L}(\hat{\alpha})}{\partial \alpha} = 0 \tag{6}$$

Ahora, del teorema 6, se tiene para $k = 0, \ldots, K$ fijo, que

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = \sum_{j=1}^{J} x_{kj} \left(z_j - n_j \frac{1}{1+\exp\{-\hbar_j\}} \right)$$

donde $\hbar_j$ es la suma del lado derecho de la ecuación (3). Esta última expresión no es lineal en los parámetros β_k. Por tanto, se requiere un método aproximativo que se motiva por la siguiente aproximación de Taylor. Si α_1 es un punto que está entre α y $\hat{\alpha}$, entonces

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

Considerando la ecuación (6), esta expresión se puede reescribir como

$$\hat{\alpha} - \alpha \quad = \quad \left(-\frac{\partial^2 \mathcal{L}(\alpha_1)}{\partial \alpha^2} \right)^{-1} \cdot \frac{\partial \mathcal{L}(\alpha)}{\partial \alpha} = (C^T V_1 C)^{-1} \cdot C^T (z-m)$$

siendo $V_1 := V(\alpha_1)$.

Como α_1 es un punto del segmento de línea que une a α y $\hat{\alpha}$, entonces $\alpha_1 = t\hat{\alpha} + (1-t)\alpha$, para un $t \in [0,1]$. Bajo el supuesto de que $\hat{\alpha}$ es fuertemente consistente para α (es decir, por componentes se cumple que $\hat{\alpha} \xrightarrow{c.s} \alpha$, cuando $n \to \infty$), se tiene que $\alpha_1 \xrightarrow{c.s} \alpha$. Por lo tanto, por componentes, $\alpha_1 \xrightarrow{P} \alpha$ cuando $n \to \infty$.

Esto implica que, por componentes,

$$\left[\underbrace{\left(C^T \cdot \frac{1}{n} V_1 \cdot C \right)^{-1} \cdot \frac{1}{\sqrt{n}} C^T (z-m)}_{= \sqrt{n} \cdot (\hat{\alpha} - \alpha)} - \left(C^T \cdot \frac{1}{n} V \cdot C \right)^{-1} \cdot \frac{1}{\sqrt{n}} C^T (z-m) \right] \xrightarrow{P} 0$$

Por tanto,

$$\underbrace{\left(C^T \cdot \frac{1}{n} V_1 \cdot C \right)^{-1} \cdot \frac{1}{\sqrt{n}} C^T (z-m)}_{= \sqrt{n} \cdot (\hat{\alpha} - \alpha)} \quad \stackrel{a}{=} \quad \underbrace{\left(C^T \cdot \frac{1}{n} V \cdot C \right)^{-1} \cdot \frac{1}{\sqrt{n}} C^T (z-m)}_{= \sqrt{n} \cdot (C^T V C)^{-1} \cdot C^T (z-m)}$$

De este resultado se obtiene una forma aproximada para:

$$\hat{\alpha} \cong \alpha + (C^T V C)^{-1} \cdot C^T (z-m)$$

Reemplazando en el lado derecho α por la t-ésima aproximación $\hat{\alpha}^{(t)}$ de α se obtiene la fórmula de recursión que da la $(t+1)$-ésima aproximación $\hat{\alpha}^{(t+1)}$ de $\hat{\alpha}$, como se indica en la formulación del teorema. □

4. Distribuciones asintóticas

Teorema 9. *Supóngase que* $\lim_{n_j \to \infty} \frac{n_j}{n \cdot v_j} = \frac{1}{\sigma_j^2} > 0$ *existe. Entonces existe una matriz* $\tilde{\Xi} := diag\left\{\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_J^2}\right\}$ *definida positiva tal que*

$$\sqrt{n}(\tilde{p} - p) \xrightarrow{d} \mathcal{N}_J(0, \tilde{\Xi}^{-1}), \quad n \to \infty, \quad J \text{ fijo},$$

siendo $\tilde{p}$ *la estimación de* p *en el modelo saturado.*

Demostración. Considerando el teorema 4c), 4f), $\sqrt{n}(\tilde{p} - p) = \left(\frac{1}{n}\tilde{\Im}\right)^{-1/2} Z^*$.

Ahora, mediante un procedimiento análogo al de la demostración de la parte b) del teorema 5, se tiene que $\left(\frac{1}{n}\tilde{\Im}\right)^{-1/2} Z^* \xrightarrow{d} \mathcal{N}_J(0, \tilde{\Xi}^{-1})$, cuando $n \to \infty$ y J fijo.

Con este resultado y la igualdad anterior, el teorema queda demostrado. □

El teorema 9 no es válido para datos no agrupados, en donde $J = n$. Esto se debe a que, si J no es fijo, no tiene sentido hablar de una aproximación asintóticamente normal (J-dimensional), cuando $J \to \infty$.

Corolario 1. *Para datos no agrupados (aquí se supone la existencia de una matriz* $\Im$ *definida positiva tal que* $\frac{1}{J}\Im \xrightarrow{J \to \infty} \Im$*), son válidas las siguientes afirmaciones:*

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

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

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

d) $\frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{V}(\hat{\beta}_k)}} \xrightarrow{d} \mathcal{N}_1(0, 1), \quad k = 0, 1, \ldots, K, \quad \beta_o = \delta, \quad J \to \infty.$

donde C, m, $\Im$ *son como se describieron en la observación del teorema 6;* $\hat{V}(\hat{\beta}_k)$ *la varianza estimada (por eso* $\hat{V}$*) de* $\hat{\beta}_k$ *y corresponde al* k*-ésimo elemento diagonal de la matriz de covarianzas estimada de* $\hat{\alpha}$, $\hat{C}ov(\hat{\alpha})$. *Recuerde que la expresión "estimada" significa que en* $V(\hat{\beta}_k)$*, los parámetros se reemplazan por estimaciones consistentes.*

Este teorema también es válido para datos agrupados, en donde J *es fijo. En este caso debe suponerse que* $\lim_{n_j \to \infty} \frac{n_j}{n \cdot v_j} = \frac{1}{\sigma_j^2} > 0$ *existe y, además, se tiene que:*

$Y = Z$, $m = (n_1 p_1, \ldots, n_J p_J)^T$, C *es la matriz de diseño de* $J \times (1 + K)$ $\Im = C^T V C$ *con* $V = diag\{n_1 v_1, \ldots, n_J v_J\}$.

Demostración.

a) De la observación del teorema 6, para datos no agrupados (en donde $J = n$), se tiene que

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

Al considerar la ecuación (4), se tiene que

$$\left[\frac{1}{J}\Im\right]^{-1/2} \xrightarrow{J\to\infty} \Xi^{-1/2}$$

Además, por la observación del teorema 7, es válido que

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

Entonces,

$$X^* \xrightarrow{d} \mathcal{N}_{1+K}(0, I), \quad \text{cuando} \quad J \to \infty$$

b) De los teoremas 8 y 6a se tiene que

$$\sqrt{J}(\hat{\alpha} - \alpha) \stackrel{a}{=} \left[\frac{1}{J}\Im\right]^{-1} \left[\frac{1}{\sqrt{J}} S(\alpha)\right]$$

Al considerar la parte b) del teorema 7 y resultados conocidos de la teoría de la probabilidad y de la normal multivariada, se puede concluir, mediante un razonamiento análogo al elaborado en la demostración del teorema 9, que

$$\sqrt{J}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}_{1+K}(0, \Xi^{-1}), \quad \text{cuando} \quad J \to \infty$$

c) Se obtiene inmediatamente de a) y del teorema 8.

d) Se sigue de c), considerando el hecho de que $\dfrac{\hat{\beta}_k - \beta_k}{\sqrt{V(\hat{\beta}_k)}}$ tiene media cero y varianza 1 y que, al reemplazar en $V(\hat{\beta}_k)$ los parámetros por estimaciones consistentes, el resultado queda válido asintóticamente. □

Obsérvese que los resultados que se dieron en el teorema 9 y en el corolario 1b) son bastante similares, lo que significa que tienen la misma interpretación tanto en los modelos logísticos como en los saturados.

Corolario 2. *Supóngase que* $\lim_{n_j \to \infty} \frac{n_j}{n \cdot v_j} = \frac{1}{\sigma_j^2} > 0$ *existe. Entonces, asintóticamente tenemos*

a)

$$(\hat{\alpha}-\alpha)^T \cdot \hat{C}ov^{-1}(\hat{\alpha}) \cdot (\hat{\alpha}-\alpha) \xrightarrow{d} \chi^2(1+K), \quad n \to \infty$$

b) Para cada subparámetro γ de α, de dimensión s,

$$(\hat{\gamma}-\gamma)^T \cdot \hat{C}ov^{-1}(\hat{\gamma}) \cdot (\hat{\gamma}-\gamma) \xrightarrow{d} \chi^2(s), \quad n \to \infty$$

siendo $\hat{C}ov(\hat{\gamma})$ la matriz de covarianzas estimada de la estimación $\hat{\gamma}$.

c)

$$\frac{(\hat{\beta}_k-\beta_k)^2}{\hat{V}(\hat{\beta}_k)} \xrightarrow{d} \chi^2(1), \quad k=0,1,\ldots,K, \quad \beta_o=\delta, \quad n \to \infty$$

Observación. Todos los resultados del teorema se cumplen para datos no agrupados, a pesar de que el supuesto no es válido, como se explicó al final de la observación que aparece antes del teorema 6.

Demostración. Considerando la parte b) del corolario 1, para el caso de datos no agrupados, se tiene que la matriz de covarianzas asintótica de $\sqrt{n} \cdot \hat{\alpha}$ es Ξ^{-1}.

El resultado quedará válido si se pasa a la matriz de covarianzas estimada

$$\hat{C}ov(\sqrt{n} \cdot \hat{\alpha})$$

Sea $B := \hat{C}ov^{-1/2}(\sqrt{n} \cdot \hat{\alpha}) \cdot [\sqrt{n} \cdot (\hat{\alpha}-\alpha)]$, por tanto, del corolario 1b), $B \xrightarrow{d} \mathcal{N}_{1+K}(0, I)$, cuando $n \to \infty$ ya que se obtiene a $\Xi^{1/2} \cdot \Xi^{-1} \cdot \Xi^{1/2} = I$ como matriz de covarianzas asintótica.

Por consiguiente,

$$(\hat{\alpha}-\alpha)^T \cdot \hat{C}ov^{-1}(\hat{\alpha}) \cdot (\hat{\alpha}-\alpha) = B^T B \longrightarrow \chi^2(1+K), \quad n \to \infty,$$

ya que es la suma de los cuadrados de $1+K$ variables normales estandarizadas e independientes de un grado de libertad. De esta forma, se obtiene el resultado a). La parte b) es un caso particular de a) y la c), de b). □

5. 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 desviaciones $-2\mathcal{L}(\hat{\theta})$. 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.

5.1. Comparación de un modelo logístico con el modelo saturado correspondiente

Teorema 10. *Supóngase que* $\lim_{n_j\to\infty} \frac{n_j}{n\cdot v_j} = \frac{1}{\sigma_j^2} > 0$ *existe. Entonces*

$$2[\mathcal{L}(\tilde{p}) - \mathcal{L}(p)] \stackrel{a}{=} \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \cdot \left[Cov^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right)\right] \cdot \left(\frac{\partial \mathcal{L}}{\partial p}\right)$$

donde $\tilde{p}$ *es la ML-estimación de* p *en el modelo saturado.*

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

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

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, de la expresión anterior y de la parte c) del teorema 4, se obtiene que

$$\begin{aligned} 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(p)] &= \left(\tilde{\Im}^{-1} \cdot \frac{\partial \mathcal{L}}{\partial p}\right)^T \cdot \left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p^2}\right) \cdot \left(\tilde{\Im}^{-1} \cdot \frac{\partial \mathcal{L}}{\partial p}\right) \\ &= \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \cdot \tilde{\Im}^{-1} \cdot \left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p^2}\right) \cdot \tilde{\Im}^{-1} \cdot \left(\frac{\partial \mathcal{L}}{\partial p}\right) \end{aligned} \tag{7}$$

Por la ley fuerte de los grandes números, $\tilde{p}_j = \frac{Z_j}{n_j} \stackrel{c.s.}{\longrightarrow} p_j$, cuando $n_j \to \infty$, para cada j. Por el supuesto adicional, vale lo mismo para $n \to \infty$, es decir, $\tilde{p} \stackrel{c.s.}{\longrightarrow} p$, cuando $n \to \infty$, por componentes. Esto implica $p_1 \stackrel{c.s}{\longrightarrow} p$ y, por lo tanto, $p_1 \stackrel{P}{\longrightarrow} p$, cuando $n \to \infty$, por componentes. Por consiguiente, para cada $j = 1, \ldots, J$,

$$\left[\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p_j^2}\right) \quad - \quad \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p)}{\partial p_j^2}\right)\right] \stackrel{P}{\longrightarrow} 0, \qquad n \to \infty, \quad J \text{ fijo.}$$

Esto implica que

$$\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p^2}\right) \stackrel{a}{=} \underbrace{\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p)}{\partial p^2}\right) \stackrel{a}{=} \frac{1}{n}\tilde{\Im}}_{\text{por el teorema 5a}} \tag{8}$$

Por otra parte, por los teoremas 9 y 4,

$$\sqrt{n}\cdot(\tilde{p}-p) = \sqrt{n}\cdot\tilde{\Im}^{-1}\cdot\left(\frac{\partial \mathcal{L}}{\partial p}\right) \stackrel{d}{\longrightarrow} \mathcal{N}_J(0, \tilde{\Xi}^{-1}), \quad n \to \infty.$$

Teniendo en cuenta las ecuaciones (7), (8) y el resultado anterior, se obtiene

$$\underbrace{\left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \sqrt{n}\tilde{\Im}^{-1} \cdot \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(p_1)}{\partial p^2}\right)\sqrt{n}\tilde{\Im}^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right)}_{=2[\mathcal{L}(\tilde{p})-\mathcal{L}(p)]} \stackrel{a}{=}$$

$$\stackrel{a}{=} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \sqrt{n}\tilde{\Im}^{-1} \cdot \frac{1}{n}\tilde{\Im} \cdot \sqrt{n}\tilde{\Im}^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right)}_{=\left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \cdot \left[Cov^{-1}\left(\frac{\partial \mathcal{L}}{\partial p}\right)\right] \cdot \left(\frac{\partial \mathcal{L}}{\partial p}\right)} \qquad \square$$

En el siguiente teorema se encontrará, para datos no agrupados (en donde $J = n$), una expresión que tiene la misma estructura que la del teorema 10. A pesar de ello, el teorema anterior es válido solo cuando J es fijo. Por eso, falla el caso de datos no agrupados.

Teorema 11. *Para los casos de datos agrupados y no agrupados, tenemos:*

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

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

Demostración. Se siguen los pasos de la demostración del teorema 10 sin detallarlos todos. Así, para cualquier par de puntos $\hat{\alpha}$ y α, existe un α_2, que está entre ellos dos, tal que

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

Pero $2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\alpha)] = (\hat{\alpha} - \alpha)^T \cdot \left(-\frac{\partial^2 \mathcal{L}(\alpha_2)}{\partial \alpha^2}\right) \cdot (\hat{\alpha} - \alpha)$ porque $\hat{\alpha}$ es ML-estimación de α. Aquí se necesita hacer el supuesto de que $\hat{\alpha}$ sea consistente para α. Así se cumple que α_2 también es consistente para α. Con lo anterior se obtiene que

$$\frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(\alpha_2)}{\partial \alpha^2}\right) \stackrel{a}{=} \frac{1}{n}\left(-\frac{\partial^2 \mathcal{L}(\alpha)}{\partial \alpha^2}\right) \stackrel{a}{=} \frac{1}{n}\Im$$

Usando ahora el teorema 8 y el corolario 1, con n en lugar de J, se obtiene el resultado en plena analogía con la demostración del teorema 10. $\square$

Teorema 12. *La LR-estadística de prueba (según el método de cocientes de funciones de verosimilitud) para la hipótesis H_0: el modelo logístico (con $\mathbf{X}_1, \ldots, \mathbf{X}_K$), vs. la alternativa H_1: el modelo saturado correspondiente con sus J poblaciones, es equivalente a la llamada <u>desviación</u> que tiene el modelo logístico del modelo saturado:*

$$D^*(M) := 2\log\left(\frac{L(\tilde{p})}{L(\hat{\alpha})}\right) = 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(\hat{\alpha})]$$

con las siguientes características:

a) $D^*(M) \stackrel{a}{=} (Z^*)^T(I_J - P_J)(Z^*)$, *bajo* H_0.

b) $D^*(M) \xrightarrow{d} \chi^2[J-(1+K)]$, *bajo* H_0, $n \to \infty$, J *fijo.*

donde $\tilde{p}$ *y* $\hat{\alpha}$ *son los vectores de las ML-estimaciones de los modelos saturado y logístico, respectivamente;* Z^* *es el vector definido en el teorema 4 y* P_J *es la matriz definida en el teorema 3. Además, se hacen los supuestos de los teoremas 10 y 11.*

Observación. Nótese que aquí se requiere que $J > 1 + K$. Para el caso en que $J = 1+K$, el análisis en un modelo logístico es el mismo que en el modelo saturado. Esta prueba únicamente se cumple para datos agrupados porque J es fijo, lo que no sucede para el caso de datos no agrupados.

Demostración. a) Se puede demostrar que $\frac{\partial \mathcal{L}}{\partial \alpha} = C^T V^{1/2} Z^*$ es un vector columna de tamaño $1 + K$. Ahora, bajo H_0, vale que $\mathcal{L}(p) = \mathcal{L}(\alpha)$. Por lo tanto, bajo H_0, y teniendo en cuenta los teoremas 4b), 4f), 6b), 10 y 11, se tiene que

$$\begin{aligned} D^*(M) &= 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(p)] - 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\alpha)] \\ &\stackrel{a}{=} \left(\frac{\partial \mathcal{L}}{\partial p}\right)^T \tilde{\Im}^{-1} \left(\frac{\partial \mathcal{L}}{\partial p}\right) - \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)^T \Im^{-1} \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right) \\ &= (Z^*)^T \cdot [\underbrace{\tilde{\Im}^{1/2} \tilde{\Im}^{-1} \tilde{\Im}^{1/2}}_{I_J} - \underbrace{V^{1/2} C \cdot \Im^{-1} \cdot C^T V^{1/2}}_{P_J}] \cdot (Z^*) \end{aligned}$$

b) Como la matriz $I_J - P_J$ es una proyección con rango $J - (1 + K)$, entonces (ver Rao (1973, cáp. 1)) existe una matriz ortogonal U de $m \times J$ tal que $I_J - P_J = U^T U$ con $m = J - (1 + K)$. Considerando lo anterior y el teorema 5b) y resultados conocidos relacionados con la normal multivariada, se tiene que $UZ^* \xrightarrow{d} \mathcal{N}_m(0, I)$, cuando $n \to \infty$ y J fijo. Por consiguiente, se tiene $(UZ^*)^T(UZ^*) \xrightarrow{d} \chi^2(m)$, cuando $n \to \infty$ y J fijo, sabiendo que $m = J - (1 + K)$ y que las componentes $(UZ^*)_l$ del vector (UZ^*), con $l = 1, \ldots, m$, son independientes entre sí. La parte b) queda completamente demostrada si se considera que bajo H_0 se cumple que

$$D^*(M) \stackrel{a}{=} (Z^*)^T (\underbrace{I_J - P_J}_{U^T U})(Z^*) = (UZ^*)^T (UZ^*)$$

□

Se espera que la prueba del teorema 12 no rechace H_0 (p-valor alto), o sea que los datos obtenidos no estén en contra del modelo logístico. Es decir, al pasar del modelo saturado al modelo logístico no se pierde información estadísticamente significativa.

5.2. Comparación de un modelo logístico con algún submodelo

Teorema 13. *Para la situación de hacer la hipótesis H_0: un submodelo logístico con $\mathbf{X}_1, \ldots, \mathbf{X}_{\tilde{K}}$, vs. la alternativa H_1: el modelo logístico con $\mathbf{X}_1, \ldots, \mathbf{X}_K$ con $\tilde{K} < K$, son válidas las siguientes afirmaciones:*

a) $\frac{\partial \mathcal{L}}{\partial \alpha_o} = T_o \cdot \frac{\partial \mathcal{L}}{\partial \alpha}$, *siendo* $T_o := \frac{\partial \alpha}{\partial \alpha_o}$ *una matriz de* $(1+\tilde{K}) \times (1+K)$ *de la forma:*

$$\frac{\partial \alpha}{\partial \alpha_o} = \left.\begin{pmatrix} 1 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 1 & 0 & \ldots & 0 \end{pmatrix}\right\} 1+\tilde{K}$$

$$\underbrace{\qquad\qquad}_{1+\tilde{K}} \quad \underbrace{\qquad\qquad}_{K-\tilde{K}}$$

b) $Cov\left(\frac{\partial \mathcal{L}}{\partial \alpha_o}\right) = TT^T$ *es una matriz cuadrada de tamaño* $1+\tilde{K}$,

siendo $T := T_o \Im^{1/2}$ *una matriz de* $(1+\tilde{K}) \times (1+K)$; α *el vector de los* $1+K$ *parámetros en el modelo logístico y* α_o *el vector* α *restringido bajo* $H_0 : \beta_{1+\tilde{K}} = \ldots = \beta_K = 0$.

Demostración. Solo debemos considerar la regla de la cadena y el teorema 6b). □

Teorema 14. *La LR-estadística de prueba para la situación señalada en el teorema 13 es equivalente a la estadística*

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

con las siguientes características:

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

b) $D^*(L) \xrightarrow{d} \chi^2[K-\tilde{K}]$, *bajo* H_0, $J \to \infty$

donde:

- $\hat{\alpha} = (\hat{\delta}, \hat{\beta}_1, \ldots, \hat{\beta}_K)^T$ *es la ML-estimación en el modelo logístico de la alternativa* H_1;
- $\hat{\alpha}_o = (\hat{\delta}_o, \hat{\beta}_{o1}, \ldots, \hat{\beta}_{o\tilde{K}})^T$ *es la ML-estimación en el submodelo logístico de la hipótesis* H_o;
- X^* *es el vector definido en el corolario 1a;*
- $P_{1+K} := T^T \cdot (T \cdot T^T)^{-1} \cdot T$ *y* T *es la matriz definida en el teorema 13.*

Para la situación anterior, una estadística asintóticamente equivalente es la de Wald:

c) $\hat{\gamma}^T \cdot \hat{C}ov^{-1}(\hat{\gamma}) \cdot \hat{\gamma} \xrightarrow{d} \chi^2[K - \tilde{K}]$*, bajo* H_0*,* $J \to \infty$ *donde* $\hat{\gamma}$ *es la estimación de* γ*, que es la parte* $(K - \tilde{K})$*-dimensional del vector* α *que se anula bajo* H_0 *y* $\hat{C}ov(\hat{\gamma})$ *es la matriz de covarianzas estimada de* $\hat{\gamma}$*.*

Observación. Nótese que la hipótesis de la primera parte del teorema es equivalente a la hipótesis $H_o : \gamma = 0$. Es importante señalar que esta prueba que presenta el teorema utiliza datos no agrupados. Aunque también es posible realizarla teniendo en cuenta el modelo saturado. Pero como en la prueba únicamente se considera el modelo logístico, no tiene mucho sentido comparar este con un submodelo teniendo que pasar por el modelo saturado.

Demostración. a) En forma análoga al teorema 11 se tiene que

$$2[\mathcal{L}(\hat{\alpha}_o) - \mathcal{L}(\alpha_o)] \stackrel{a}{=} \left(\frac{\partial \mathcal{L}}{\partial \alpha_o}\right)^T \cdot Cov^{-1}\left(\frac{\partial \mathcal{L}}{\partial \alpha_o}\right) \cdot \left(\frac{\partial \mathcal{L}}{\partial \alpha_o}\right) \tag{9}$$

Por lo tanto,

$$\begin{aligned} D^*(L) &= 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\alpha)] - 2[\mathcal{L}(\hat{\alpha}_o) - \mathcal{L}(\alpha_o)] \\ &= \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)^T \left\{\Im^{-1} - (T_o)^T (TT^T)^{-1} T_o\right\} \cdot \left(\frac{\partial \mathcal{L}}{\partial \alpha}\right) \\ &= \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)^T \Im^{-1/2}}_{(X^*)^T} \cdot [I_{1+K} - P_{1+K}] \cdot \underbrace{\Im^{-1/2}\left(\frac{\partial \mathcal{L}}{\partial \alpha}\right)}_{(X^*)} \end{aligned}$$

b) Se puede demostrar que la matriz $I_{1+K} - P_{1+K}$ es una proyección con rango $K - \tilde{K}$. Por consiguiente (Rao 1973, cáp.1), existe una matriz ortogonal U de $m \times J$ tal que $I_{1+K} - P_{1+K} = U^T U$ con $m = K - \tilde{K}$. Ahora, considerando el corolario 1a), se tiene que $UX^* \xrightarrow{d} \mathcal{N}_m(O, I)$, bajo H_0, cuando $J \to \infty$. Por tanto, $(UX^*)^T(UX^*) \xrightarrow{d} \chi^2(m)$, bajo H_0, cuando $J \to \infty$, sabiendo que $m = K - \tilde{K}$ y que las componentes $(UX^*)_l$ del vector (UX^*), con $l = 1, \ldots, m$, son independientes entre sí. La parte b) queda completamente demostrada si se considera que bajo H_0,

$$D^*(L) \stackrel{a}{=} (X^*)^T (\underbrace{I_{1+K} - P_{1+K}}_{U^T U})(X^*) = (UX^*)^T (UX^*)$$

c) Se sabe que γ es un subparámetro de α con dimensión $K - \tilde{K}$. Entonces, por el corolario 2b), se tiene $(\hat{\gamma} - \gamma)^T \cdot \hat{C}ov^{-1}(\hat{\gamma}) \cdot (\hat{\gamma} - \gamma) \xrightarrow{d} \chi^2[K - \tilde{K}]$, cuando $J \to \infty$. Bajo $H_0 : \gamma = 0$, esta expresión quedará reducida a la que aparece en el teorema y con esto queda demostrada la parte c). □

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 el teorema 14 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.

5.3. Comparación de un modelo logístico con el nulo

Corolario 3. *Para la hipótesis H_0: el modelo nulo (sólo con el intercepto), vs. la alternativa H_1: el modelo logístico (con $\mathbf{X}_1, \ldots, \mathbf{X}_K$), se puede tomar, alternativamente, una de las dos estadísticas de pruebas siguientes:*

a) $2\log\left(\frac{L(\hat{\alpha})}{L(\hat{\delta}_o)}\right) = 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\hat{\delta}_o)] \xrightarrow{d} \chi^2[K], \quad$ *bajo* $H_0, \quad J \to \infty$

donde $\hat{\delta}_o = logit(\overline{Y})$ es la estimación de δ en el modelo nulo.

b) $\hat{\beta}^T \cdot \hat{C}ov^{-1}(\hat{\beta}) \cdot \hat{\beta} \xrightarrow{d} \chi^2[K], \quad$ *bajo* $H_0, \quad J \to \infty,$

donde $\hat{\beta}$ es la ML-estimación de $\beta = (\beta_1, \ldots, \beta_K)^T$ que corresponde a la parte de α que se anula bajo H_0 y $\hat{C}ov(\hat{\beta})$ es la matriz de covarianzas estimada de $\hat{\beta}$.

Observación. Nótese que la hipótesis de la primera parte del teorema es equivalente a la hipótesis $H_o : \beta = 0$. Como esta prueba es un caso particular del teorema 14, también se está considerando el caso de datos no agrupados.

Demostración. La parte a) es un caso particular del teorema 14 con $\tilde{K} = 0$. En este caso, $\hat{\alpha}_o = \hat{\delta}_o$. Ahora se demostrará la parte b). β es un subparámetro de α con dimensión K. Por el corolario 2b), se tiene $(\hat{\beta} - \beta)^T \cdot \hat{C}ov^{-1}(\hat{\beta}) \cdot (\hat{\beta} - \beta) \xrightarrow{d} \chi^2[K]$, cuando $J \to \infty$. Bajo $H_0 : \beta = 0$, esta expresión quedará reducida a la que aparece en el teorema y con esto queda demostrada la parte b). □

En el trabajo práctico se espera que la prueba del teorema 3 sí rechace H_o (p-valor bajo). Es decir, que las K variables explicativas del modelo logístico tienen, en su conjunto, una explicación más informativa que solo el intercepto. En caso contrario, que no es muy común en problemas prácticos, se tendría que chequear otro modelo logístico con más o con otras variables.

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

Corolario 4. *Para la hipótesis H_0: el submodelo con una variable menos (es decir, con $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \ldots, \mathbf{X}_K$), vs. la alternativa H_1: el modelo logístico (con $\mathbf{X}_1, \ldots, \mathbf{X}_K$), se puede tomar, alternativamente, una de las dos estadísticas de pruebas siguientes:*

a) $2\log\left(\frac{L(\hat{\alpha})}{L(\hat{\alpha}_o)}\right) = 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\hat{\alpha}_o)] \xrightarrow{d} \chi^2[1], \quad \textit{bajo } H_0, \quad J \to \infty,$

donde $\hat{\alpha}_o$ *es la estimación bajo* H_o

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

siendo $\hat{\beta}_k$ *la estimación de* β_k*, para cada* $k = 0, 1, \ldots, K$ *en el modelo (bajo* H_1*) con su varianza estimada* $\hat{V}(\hat{\beta}_k)$ *y su error estándar* $SE(\hat{\beta}_k) = \sqrt{\hat{V}(\hat{\beta}_k)}$.

Observación. Como esta prueba es un caso particular del teorema 14, también se está considerando el caso de datos no agrupados.

Demostración. La parte a) es un caso particular del teorema 14b) con $\tilde{K} = K-1$. Es decir, con $K - \tilde{K} = 1$. La parte b) es un caso particular del corolario 2c). □

Observación. Con base en todas las "pruebas parciales" del teorema 4, para cada $k = 0, 1, \ldots, K$ se eliminará la variable explicativa que menor aporte individual tenga en la explicación. Es decir, la variable que tenga p-valor parcial más alto. Así, se sigue eliminando variable tras variable hasta que se rechacen todas las pruebas parciales (todas las que tengan p-valores bajos).

5.5. Análisis de desviaciones (ANODEV)

Con el fin de analizar la bondad de ajuste de un modelo logístico fijo (con variables explicativas $\mathbf{X}_1, \ldots, \mathbf{X}_K$ que definen J poblaciones), se lo compara hacia los dos lados, es decir, con el modelo saturado correspondiente (con estas J poblaciones) y con el modelo nulo (con solo el intercepto). Para esta situación, puede orientarse en la llamada tabla de ANODEV (en inglés: **AN**alysis **O**f **DEV**iance), en analogía a la ANOVA para modelos lineales (tabla 1).

Tabla 1: ANODEV del modelo logístico *vs.* saturado.

Teorema		DF	Estadística
5.3.1	Diferencia de desviaciones	K	$D^*(0) - D^*(M) = 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\hat{\delta}_o)]$
5.1.5	Desviación del modelo logístico	$J - (1 + K)$	$D^*(M) := 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(\hat{\alpha})]$
	Desviación total (del modelo nulo)	$J - 1$	$D^*(0) := 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(\hat{\delta}_o)]$

Una estadística de *desviación relativa* con respecto al modelo saturado es la siguiente, propuesta en la mayor parte de la literatura:

$$RDS := \frac{D^*(0) - D^*(M)}{D^*(0)} = 1 - \frac{D^*(M)}{D^*(0)}$$

Esta desviación relativa la analizaron por Theil (1970) y Goodman (1971) y tiene algunas características:

a) $0 \leq RDS \leq 1$.

b) Cuando el modelo logístico no mejora en ajuste con respecto al modelo nulo: $RDS = 0$ si y solo si $0 < D^*(0) = D^*(M)$ si y solo si $\mathcal{L}(\tilde{p}) > \mathcal{L}(\hat{\delta}_o) = \mathcal{L}(\hat{\alpha})$.

c) Cuando el modelo logístico se ajusta tan bien como el modelo saturado: $RDS = 1$ si y solo si $0 = D^*(M) < D^*(0)$ si y solo si $\mathcal{L}(\tilde{p}) = \mathcal{L}(\hat{\alpha}) > \mathcal{L}(\hat{\delta}_o)$.

d) Al eliminar una variable explicativa (disminuir K) se disminuye el numerador, pero, en general, se disminuye también el denominador (ya que disminuye J). De esta forma, RDS puede disminuir o aumentar al eliminar variables explicativas, siendo esto último el gran defecto que tiene la desviación relativa.

Ahora, si se quiere comparar dos o más modelos logísticos, ya no se puede hacer según lo mencionado en la nota que sigue al teorema 4 por el defecto señalado arriba. Por eso se propone analizar la bondad de ajuste de un modelo logístico fijo (con variables explicativas $\mathbf{X}_1, \ldots, \mathbf{X}_K$ que definen J poblaciones), al compararlo con el modelo completo (que no se basa en poblaciones) y con el modelo nulo (solo con el intercepto). En este caso, la desviación del modelo logístico será

$$D^*(M) := 2[\mathcal{L}(\hat{p}) - \mathcal{L}(\hat{\alpha})] = -2[\mathcal{L}(\hat{\alpha})]$$

ya que $\mathcal{L}(\hat{p}) = 0$. Esta situación se orienta en la tabla 5.5 de ANODEV.

Tabla 2: ANODEV del modelo logístico *vs.* completo.

Teorema		DF	Estadística
5.3.1	Diferencia de desviaciones	K	$D(0) - D(M) = 2[\mathcal{L}(\hat{\alpha}) - \mathcal{L}(\hat{\delta}_o)]$
	Desviación del modelo logístico		$D(M) := 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(\hat{\alpha})]$
	Desviación total (del modelo nulo)		$D(0) := 2[\mathcal{L}(\tilde{p}) - \mathcal{L}(\hat{\delta}_o)]$

Ahora se sugiere como una estadística de desviación relativa (con respecto al modelo completo) la propuesta por C.F & McFadden (1974):

$$RDC := \frac{D(0) - D(M)}{D(0)} = 1 - \frac{D(M)}{D(0)}$$

Esta desviación relativa tiene las características:

a) $0 \leq RDC \leq RDS \leq 1$.

b) $RDC = 0$ si y solo si $0 < D(M) = D(0)$ si y solo si $0 > \mathcal{L}(\hat{\alpha}) = \mathcal{L}(\hat{\delta}_o)$.

c) $RDC = 1$ si y solo si $0 = D(M) < D(0)$ si y solo si $0 = \mathcal{L}(\hat{\alpha}) > \mathcal{L}(\hat{\delta}_o)$.

d) Al eliminar una variable explicativa (disminuir K) se disminuye el numerador, pero el denominador $D(0)$ no cambia de valor porque no depende de las variables explicativas. Por lo tanto, RDC sí disminuye al eliminar variables explicativas. Esto quiere decir que sí se puede comparar las desviaciones relativas RDC entre un modelo logístico y cualquier submodelo.

5.6. Criterio para la escogencia de un buen submodelo logístico

5.6.1. El análisis de un modelo logístico M fijo

Se compara este modelo logístico (con sus $1 + K$ parámetros) con el modelo saturado correspondiente (con J poblaciones/parámetros) y con el modelo nulo (con 1 parámetro, el intercepto) como se hizo en las secciones 5.1 y 5.3, respectivamente. Para que el modelo logístico pueda considerarse como "aceptable", debe estar "cerca" del modelo saturado y "lejos" del modelo nulo. Es decir,

- No debe ser rechazado el modelo logístico *vs.* el saturado según la prueba del teorema 12, y
- Sí debe ser rechazado el modelo nulo *vs.* el logístico según la prueba del teorema 3.

Lo anterior debe reflejar un valor "grande" de la desviación relativa con respecto al modelo saturado, RDS, según se señaló en la sección 5.5. Es importante también recalcar que la desviación relativa RDC es una estadística que tiene un valor "grande" si el modelo logístico, adicionalmente a lo anterior, también está cerca del modelo completo. Todas estas situaciones pueden visualizarse claramente en la figura 1.

5.7. De un modelo logístico M hacia un buen submodelo logístico M_o

Para escoger el mejor submodelo logístico, se compara cada submodelo con:

1. Su modelo saturado (el número de las poblaciones baja) y con el modelo nulo. Como criterio sirve el teorema 4.
2. El modelo inicial, según la hipótesis dada en el teorema 13 y con la estadística señalada en el teorema 14.

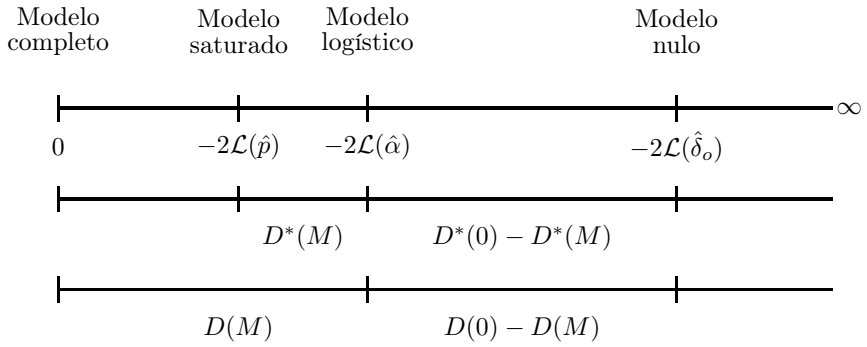

FIGURA 1: Gráfica para el análisis de un modelo logístico M fijo.

3. La sucesión de todos los submodelos anteriores. En este caso, se debe calcular la RDC para cada submodelo. De esta manera, se obtiene una sucesión (finita) decreciente de valores que sirve como un criterio (junto al anterior) para decidir cuándo y por qué se detiene el proceso de eliminación. Es decir, para decidir cuál submodelo puede ser mejor que los anteriores, incluso mejor que el modelo inicial. Esta decisión es siempre un compromiso entre aceptar una pérdida de explicación global, que se expresa en un menor valor RDC (análogamente al coeficiente de ajuste R^2 para modelos lineales), pero que no sea estadísticamente significativa la pérdida; y ganar mejor explicación parcial, que se expresa en el rechazo de todas las hipótesis parciales $H_o : \beta_k = 0$, $k = 0, 1, \ldots, K$, para cada variable que quedó en el modelo final, según el teorema 4. La situación final puede visualizarse en la figura 2.

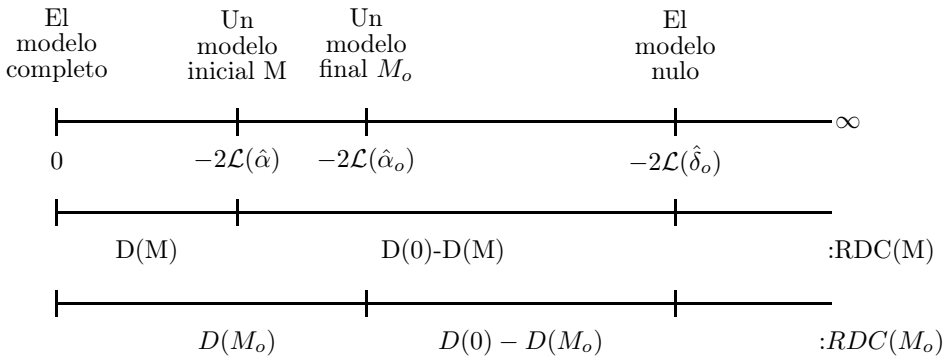

FIGURA 2: Gráfica para escoger el mejor submodelo logístico M_o.

Recibido: Septiembre 2006
Aceptado: Noviembre 2006

Referencias

Agresti, A. (1990), *Categorical Data Analysis*, 2nd edn, John Wiley and Sons, Inc., New York.

C.F, M. & McFadden, D., eds (1974), *Frontiers in Econometrics Applications*, MA: MIT Press, Cambridge, pp. 105–142. "Conditional logit analysis of qualitative choice behavior".

Dobson, A. J. (2002), *An Introduction to Generalized Linear Models*, 2 edn, Chapman & Hall, London.

Goodman, L. (1971), 'The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications', *Technometrics* (13), 33–61.

Mc Cullagh, P. (1983), 'Quasi-likelihood Functions', *Annals of Statistics* (11), 59–67.

Mc Cullagh, P. & Nelder, J. (1983), *Generalized Linear Models*, 2 edn, Chapman and Hall, London.

Rao, C. (1973), *Linear Statistical Inference and its Applications*, 2 edn, John Wiley and Sons, Inc., New York.

Theil, H. (1970), 'On the Estimation of Relationships Involving Qualitative Variables', *Amer. J. Sociol.* (76), 103–154.

Wedderburn, R. (1974), 'Quasi-likelihood Functions, Generalized linear models and the Gauss-Newton Method', *Biometrika* (61), 439–447.

Wedderburn, R. (1976), 'On The Existence and Uniqueness of the Maximum Likelihood Estimates for Certain Generalized Linear Models', *Biometrika* (63), 27–32.

Zacks, S. (1971), *The Theory of Statistical Inference*, 2nd edn, John Wiley and Sons Inc., New York.