

Tema 7. Regresión Logística

Pedro Larrañaga, Iñaki Inza, Abdelmalik Moujahid
Departamento de Ciencias de la Computación e Inteligencia Artificial
Universidad del País Vasco–Euskal Herriko Unibertsitatea

7.1 Introducción

En este tema vamos a introducir las ideas fundamentales subyacentes a un paradigma denominado *regresión logística*. La regresión logística se ha convertido en un paradigma muy usado en Ciencias de la Salud para construir modelos predictores. La razón fundamental de esta popularización es debido al hecho de que los parámetros en los que se basa tienen una interpretación en términos de riesgo.

La regresión logística se basa en la denominada *función logística*:

$$f(z) = \frac{1}{1 + e^{-z}}$$

la cual verifica que: $0 < f(z) < 1$ y puede ser interpretada en términos de probabilidad. Por otra parte la función logística verifica las siguientes propiedades:

$$\begin{aligned}\lim_{z \rightarrow -\infty} \frac{1}{1 + e^{-z}} &= 0 \\ \lim_{z \rightarrow +\infty} \frac{1}{1 + e^{-z}} &= 1 \\ f(0) &= \frac{1}{1 + e^{-0}} = \frac{1}{2}\end{aligned}$$

Este tema se ha estructurado de la manera siguiente: En la Sección 7.2 se introduce el modelo logístico, así como diversos conceptos asociados al mismo, tales como *risk ratio*, *odds ratio*, *formulación logit* y *risk odds ratio*. La Sección 7.3 muestra cómo estimar los parámetros de los que depende el modelo a partir de sus estimaciones máximo verosímiles. En las Secciones 7.3 y 7.4 se introducen respectivamente el *test de la razón de verosimilitud* y el *test de Wald* ambas herramientas útiles para el proceso de *modelización*, el cual es descrito en la Sección 7.6.

7.2 El Modelo Logístico

Si denotamos por C a la variable a predecir, y por X_1, \dots, X_n a las n variables predictoras, el paradigma de regresión logística se expresa de la manera siguiente:

$$P(C = 1|\mathbf{x}) = P(C = 1|X_1 = x_1, \dots, X_n = x_n) = \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}$$

donde \mathbf{x} representa un patrón a clasificar, y $\beta_0, \beta_1, \dots, \beta_n$ son los parámetros, que deben ser estimados a partir de los datos, a fijar para tener determinado un modelo concreto de regresión logística.

Si consideramos que la variable a predecir C es binaria, podemos calcular $P(C = 0|\mathbf{x})$ de la siguiente manera:

$$P(C = 0|\mathbf{x}) = 1 - P(C = 1|\mathbf{x}) = 1 - \frac{1}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}} = \frac{e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}$$

Por ejemplo usando los datos proporcionados por Kleinbaum (1994) donde C representa la variable aleatoria *Enfermedad Coronaria* con posibles valores (1 si, 0 no), X_1 *Nivel de Colesterol* (1 alto, 0 bajo), X_2 *Edad Continua*, y X_3 el *resultado del Electrocardiograma* (1 anormal, 0 normal), supongamos que obtenemos el modelo de regresión logística -obtenido a partir de $N = 609$ casos siguiente:

$$\begin{aligned}\widehat{\beta}_0 &= -3,911 \\ \widehat{\beta}_1 &= 0,652 \\ \widehat{\beta}_2 &= 0,029 \\ \widehat{\beta}_3 &= 0,342\end{aligned}$$

Si quisiéramos comparar el riesgo para dos patrones: $\mathbf{x} = (1, 40, 0)$ y $\mathbf{x}' = (0, 40, 0)$ podríamos comenzar calculando la probabilidad de que $C = 1$ para cada uno de ellos:

$$P(C = 1|\mathbf{x}) = P(C = 1|X_1 = 1, X_2 = 40, X_3 = 0) = \frac{1}{1 + e^{-(-3,911 + 0,652(1) + 0,029(40) + 0,342(0))}} = 0,109$$

$$P(C = 1|\mathbf{x}') = P(C = 1|X_1 = 0, X_2 = 40, X_3 = 0) = \frac{1}{1 + e^{-(-3,911 + 0,652(0) + 0,029(40) + 0,342(0))}} = 0,060$$

para posteriormente utilizar el denominado *risk ratio (RR)* de $X_1 = 1$ frente a $X_1 = 0$ definido de la siguiente manera:

$$RR(\mathbf{x}, \mathbf{x}') = \frac{P(C = 1|\mathbf{x})}{P(C = 1|\mathbf{x}')} = \frac{P(C = 1|X_1 = 1, X_2 = 40, X_3 = 0)}{P(C = 1|X_1 = 0, X_2 = 40, X_3 = 0)} = \frac{0,109}{0,060} = 1,82$$

Es decir, para una persona con 40 años y electrocardiograma normal, el riesgo se multiplica casi por dos al pasar de un nivel de *Colesterol* bajo(0) a uno alto (1).

Otro concepto de interés es el de *odds ratio (OR)* de un determinado patrón \mathbf{x} el cual se denota por $OR(\mathbf{x})$ y se define como el cociente entre la probabilidad de que el patrón pertenezca a la clase 1 entre la probabilidad de que el patrón pertenezca a la clase 0. Es decir:

$$OR(\mathbf{x}) = \frac{P(C = 1|\mathbf{x})}{1 - P(C = 1|\mathbf{x})}$$

Para trabajar con $OR(\mathbf{x})$ de manera ágil, es conveniente expresar el modelo de regresión logística en la manera *logit*. Para ello se efectúa una transformación del modelo, de la manera siguiente:

$$\text{logit} (P(C = 1|\mathbf{x})) = \ln OR(\mathbf{x}) = \ln \left[\frac{P(C = 1|\mathbf{x})}{1 - P(C = 1|\mathbf{x})} \right]$$

Sustituyendo en la fórmula anterior las expresiones correspondientes al modelo logístico obtenemos:

$$\begin{aligned} \text{logit} (P(C = 1|\mathbf{x})) &= \ln \left[\frac{P(C = 1|\mathbf{x})}{1 - P(C = 1|\mathbf{x})} \right] = \frac{\frac{1}{1+e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}}{\frac{e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}{1+e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)}}} = \\ &= \ln e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i\right)} = \beta_0 + \sum_{i=1}^n \beta_i x_i \end{aligned}$$

Tal y como se ha comentado anteriormente, el *odds ratio* (OR) de un individuo con patrón \mathbf{x} se define como el cociente entre la probabilidad de que $C = 1$ dado dicho patrón \mathbf{x} y la probabilidad de que $C = 0$ dado \mathbf{x} . Así un *odds ratio* de $\frac{1}{3}$ para un patrón \mathbf{x} se interpreta diciendo que para dicho patrón la probabilidad de que se dé $C = 1$ es una tercera parte de la probabilidad de que $C = 0$.

Además se tiene que $\ln OR(\mathbf{x}) = \ln \frac{P(C = 1|\mathbf{x})}{1 - P(C = 1|\mathbf{x})} = \beta_0 + \sum_{i=1}^n \beta_i x_i$ y por tanto si $\mathbf{x} = (0, 0, \dots, 0)$ entonces $\ln OR(\mathbf{0}) = \beta_0$.

Siguiendo con el ejemplo anterior relativo a la enfermedad coronaria, si calculásemos el logit para $\mathbf{x} = (1, 40, 0)$ y para $\mathbf{x}' = (0, 40, 0)$ obtenemos:

$$\text{logit} P(C = 1|\mathbf{x}) = \beta_0 + 1 \cdot \beta_1 + 40 \cdot \beta_2 + 0 \cdot \beta_3$$

Así como:

$$\text{logit} P(C = 1|\mathbf{x}') = \beta_0 + 0 \cdot \beta_1 + 40 \cdot \beta_2 + 0 \cdot \beta_3$$

y restando ambos logit se obtiene:

$$\text{logit} P(C = 1|\mathbf{x}) - \text{logit} P(C = 1|\mathbf{x}') = \beta_1$$

Veamos cómo expresar de manera genérica la constatación anterior.

Teorema 1: En un modelo de regresión logística, el coeficiente $\beta_i (i = 1, \dots, n)$ representa el cambio en el logit resultante al aumentar una unidad en la i -ésima variable $X_i (i = 1, \dots, n)$.

Demostración: Sean $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ y $\mathbf{x}' = (x'_1, \dots, x'_i, \dots, x'_n)$ dos patrones

verificando $x_j = x'_j$ para todo $j \neq i$ y $x'_i = x_i + 1$.

Calculando el cambio en el logit obtenemos:

$$\begin{aligned} \text{logit}(\mathbf{x}') - \text{logit}(\mathbf{x}) &= \beta_0 + \sum_{i=1}^n \beta_i x'_i - \left(\beta_0 + \sum_{i=1}^n \beta_i x_i \right) = \beta_i x'_i - \beta_i x_i = \\ &= \beta_i (x_i + 1 - x_i) = \beta_i \end{aligned}$$

□

Otro concepto que resulta de interés es el de *risk odds ratio* (ROR) de \mathbf{x}' frente a \mathbf{x} , el cual mide el riesgo del *odds ratio* de \mathbf{x}' frente al *odds ratio* de \mathbf{x} ($OR(\mathbf{x})$), es decir:

$$ROR(\mathbf{x}', \mathbf{x}) = \frac{OR(\mathbf{x}')}{OR(\mathbf{x})} = \frac{e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x'_i \right)}}{e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i \right)}} = e^{\left(\sum_{i=1}^n \beta_i (x'_i - x_i) \right)}$$

Obviamente $ROR(\mathbf{x}', \mathbf{x})$ se puede expresar de manera alternativa como:

$$ROR(\mathbf{x}', \mathbf{x}) = \prod_{i=1}^n e^{\beta_i (x'_i - x_i)} = e^{\beta_1 (x'_1 - x_1)} \cdot \dots \cdot e^{\beta_n (x'_n - x_n)}$$

7.3 Estimación Máximo Verosímil de los Parámetros

La estimación de los parámetros $\widehat{\beta}_0, \widehat{\beta}_1, \dots, \widehat{\beta}_n$ de un modelo de regresión logística se efectúa por medio del método de estimación por máxima verosimilitud. Según dicho método se obtienen los estimadores máximo verosímiles como funciones de la muestra que hacen que se maximice la función de verosimilitud asociada a la muestra.

Denotando por $L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n)$ a la función de verosimilitud asociada a una muestra de tamaño N , para un modelo de regresión logística con parámetros $\beta_0, \beta_1, \dots, \beta_n$, con una variable clase C dicotómica, se tiene que:

$$\begin{aligned} L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n) &= \\ &= \prod_{j=1}^N P(C = 1 | \mathbf{x}^{(j)})^{c^{(j)}} (1 - P(C = 1 | \mathbf{x}^{(j)}))^{1-c^{(j)}} \end{aligned}$$

Por otra parte, teniendo en cuenta que $\ln(z)$ es una función creciente estrictamente, y por tanto el valor de los parámetros $\beta_0, \beta_1, \dots, \beta_n$ maximizando

$$L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n)$$

coincide con el valor de los parámetros que maximiza

$$\ln L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n)$$

Desarrollando el logaritmo natural de la función de verosimilitud obtenemos:

$$\begin{aligned} \ln L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n) &= \\ &= \sum_{j=1}^N c^{(j)} \ln P(C = 1 | \mathbf{x}^{(j)}) + \sum_{j=1}^N (1 - c^{(j)}) \ln (1 - P(C = 1 | \mathbf{x}^{(j)})) = \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^N c^{(j)} [\ln P(C = 1|\mathbf{x}^{(j)}) - \ln(1 - P(C = 1|\mathbf{x}^{(j)}))] + \sum_{j=1}^N \ln(1 - P(C = 1|\mathbf{x}^{(j)})) = \\
&= \sum_{j=1}^N c^{(j)} \ln \frac{P(C = 1|\mathbf{x}^{(j)})}{1 - P(C = 1|\mathbf{x}^{(j)})} + \sum_{j=1}^N \ln(1 - P(C = 1|\mathbf{x}^{(j)})) =
\end{aligned}$$

Teniendo en cuenta que

$$\ln \frac{P(C = 1|\mathbf{x}^{(j)})}{1 - P(C = 1|\mathbf{x}^{(j)})} = \beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}$$

y que

$$(1 - P(C = 1|\mathbf{x}^{(j)})) = \frac{e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}}{1 + e^{-\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}} = \frac{1}{1 + e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}}$$

Obtenemos

$$\begin{aligned}
&\ln L((\mathbf{x}^{(1)}, c^{(1)}), \dots, (\mathbf{x}^{(N)}, c^{(N)}), \beta_0, \beta_1, \dots, \beta_n) = \\
&= \sum_{j=1}^N c^{(j)} \left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right) - \sum_{j=1}^N \ln \left(1 + e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}\right)
\end{aligned}$$

Los estimadores máximo verosímiles $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_n$ para los parámetros $\beta_0, \beta_1, \dots, \beta_n$ se van a obtener al resolver el siguiente sistema de $n+1$ ecuaciones y $n+1$ incógnitas:

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta_0} &= \sum_{j=1}^N c^{(j)} - \sum_{j=1}^N \frac{e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}}{1 + e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}} = 0 \\
\frac{\partial \ln L}{\partial \beta_1} &= \sum_{j=1}^N c^{(j)} x_1^{(j)} - \sum_{j=1}^N x_1^{(j)} \frac{e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}}{1 + e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}} = 0 \\
&\vdots \\
\frac{\partial \ln L}{\partial \beta_n} &= \sum_{j=1}^N c^{(j)} x_n^{(j)} - \sum_{j=1}^N x_n^{(j)} \frac{e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}}{1 + e^{\left(\beta_0 + \sum_{i=1}^n \beta_i x_i^{(j)}\right)}} = 0
\end{aligned}$$

En el anterior sistema de $n + 1$ ecuaciones y $n + 1$ incógnitas no es posible obtener una fórmula cerrada para los estimadores de los parámetros $\beta_0, \beta_1, \dots, \beta_n$, de ahí que lo habitual sea utilizar técnicas iterativas para llevar a cabo dichas estimaciones. Al utilizar el método de Newton-Raphson para llevar a cabo dichas iteraciones, se obtiene la siguiente fórmula de actualización de los parámetros:

$$\hat{\beta}^{nuevo} = \hat{\beta}^{viejo} + (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t (\mathbf{c} - \hat{\mathbf{p}})$$

donde

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_n)$$

\mathbf{X} matriz cuyas filas son $\mathbf{x}^{(j)}, j = 1, \dots, N$. Es decir, $\mathbf{X} \in M(N, n)$

\mathbf{W} matriz diagonal con elementos $p^{(j)} (1 - p^{(j)}), j = 1, \dots, N$. $\mathbf{W} \in M(N, N)$.

$$\mathbf{W} = \begin{pmatrix} p^{(1)} (1 - p^{(1)}) & \dots & \dots & 0 \\ 0 & p^{(2)} (1 - p^{(2)}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & p^{(N)} (1 - p^{(N)}) \end{pmatrix}$$

$\hat{\mathbf{p}}$ vector cuya componente j -ésima indica la probabilidad estimada en esa iteración.

$$\text{Es decir, } \hat{\mathbf{p}} \in M(N, 1), \text{ con } p^{(j)} = \frac{e^{\left(\mathbf{x}^{(j)} \hat{\beta}^{viejo}\right)}}{1 + e^{\left(\mathbf{x}^{(j)} \hat{\beta}^{viejo}\right)}}$$

\mathbf{c} vector de componentes $c^{(j)}, j = 1, \dots, N$. Por tanto $\mathbf{c} \in M(N, 1)$.

Los criterios de convergencia del método iterativo utilizado para estimar los parámetros pueden ser varios, pero en todos ellos la idea subyacente es que bien $\hat{\beta}^{nuevo} \simeq \hat{\beta}^{viejo}$ o $\ln L(\hat{\beta}^{nuevo}) \simeq \ln L(\hat{\beta}^{viejo})$ o $\hat{\mathbf{p}}^{nuevo} \simeq \hat{\mathbf{p}}^{viejo}$.

7.4 Test de la Razón de Verosimilitud

El test de la razón de verosimilitud se basa en comparar el producto entre -2 y el logaritmo neperiano de un cociente entre verosimilitudes con el percentil correspondiente de una distribución chi-cuadrado. Dicho test de la razón de verosimilitud tiene como objetivo el comparar dos modelos de regresión logística, el denominado *modelo completo* (*full model*) frente al que se conoce como *modelo reducido* (*reduced model*). Este segundo modelo puede verse como un submodelo del modelo completo. La hipótesis nula testada en el test de la razón de verosimilitud establece que los parámetros correspondientes a las variables que forman parte del modelo completo, pero no del modelo reducido, valen cero.

Para ver la manera en la que funciona el test de la razón de verosimilitud vamos a considerar los siguientes tres modelos de regresión logística expresados en su formulación logit:

$$\text{Modelo 1: logit } P_1(C = 1 | \mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2$$

$$\text{Modelo 2: logit } P_2(C = 1 | \mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

$$\text{Modelo 3: logit } P_3(C = 1 | \mathbf{x}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_3 + \beta_5 x_2 x_3$$

Tal y como puede verse, el Modelo 2 es una extensión del Modelo 1, de igual manera

que el Modelo 3 constituye una extensión del Modelo 2. En caso de querer comparar el Modelo 2 frente al Modelo 1, este último jugará el papel de modelo reducido, mientras que el Modelo 2 será el modelo completo. De manera análoga, si quisiésemos comparar el Modelo 3 frente al Modelo 2, dicho Modelo 2 sería el modelo reducido, mientras que el Modelo 3 se interpretará como modelo completo.

Vamos a denotar por $\widehat{L}_1, \widehat{L}_2$ y \widehat{L}_3 los valores de máxima verosimilitud obtenidos respectivamente por el Modelo 1, Modelo 2 y Modelo 3 en relación con un conjunto de N casos previamente determinado. Debido a que cuanto más parámetros tiene un modelo mejor se va ajustando a los datos, y esta es la situación existente con los tres modelos anteriores debido a sus características jerárquicas, se tiene que:

$$\widehat{L}_1 \leq \widehat{L}_2 \leq \widehat{L}_3$$

Pero por otra parte, al ser el logaritmo una función creciente, se obtiene que:

$$\ln \widehat{L}_1 \leq \ln \widehat{L}_2 \leq \ln \widehat{L}_3$$

y por tanto

$$-2\ln \widehat{L}_3 \leq -2\ln \widehat{L}_2 \leq -2\ln \widehat{L}_1$$

siendo esta la relación existente entre los denominados *log likelihood statistics*, a partir de los cuales se va a construir el test de la razón de verosimilitud.

El test de la razón de verosimilitud (LR) tiene en cuenta la resta entre dos *log likelihood statistics*, o lo que es lo mismo, el logaritmo neperiano del cociente entre dos verosimilitudes.

Siguiendo con el ejemplo introducido anteriormente y tratando de comparar el Modelo 2 frente al Modelo 1, el test de la razón de verosimilitud plantea como hipótesis nula el que $\beta_3 = 0$, es decir, que el parámetro de la componente que forma parte del Modelo 2 pero no del Modelo 1 es cero. Por tanto se tiene:

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

La manera en la que funciona el test de razón de verosimilitud es la siguiente: Si la variable X_3 efectúa una gran contribución a la modelización y hace que el Modelo 2 se ajuste mucho mejor a los datos que el Modelo 1, se tendrá que \widehat{L}_2 será mucho mayor que \widehat{L}_1 , y por tanto $\frac{\widehat{L}_1}{\widehat{L}_2} \simeq 0$. Tomando logaritmos neperianos, $\ln \frac{\widehat{L}_1}{\widehat{L}_2} \simeq -\infty$, y de ahí que $-2\ln \frac{\widehat{L}_1}{\widehat{L}_2} \simeq +\infty$. Por tanto cuanto mayor sea el valor de $\text{LR} = -2\ln \frac{\widehat{L}_1}{\widehat{L}_2}$ más en contra estaremos de la hipótesis nula $H_0 : \beta_3 = 0$.

Por otra parte, si la contribución de X_3 es escasa, se tendría que $\frac{\widehat{L}_1}{\widehat{L}_2} \simeq 1$ y por tanto

$$\ln \frac{\widehat{L}_1}{\widehat{L}_2} \simeq 0 \text{ y finalmente } \text{LR} = -2\ln \frac{\widehat{L}_1}{\widehat{L}_2} \simeq 0.$$

Se demuestra teóricamente que $\text{LR} = -2\ln \frac{\widehat{L}_1}{\widehat{L}_2}$ sigue bajo la hipótesis nula H_0 una distribución de probabilidad χ_r^2 cuando N , número de casos en la base de datos,

es suficientemente grande. El número de grados de libertad de la distribución chi-cuadrado, r , es igual al número de parámetros que en el modelo completo deben igualarse a cero para que dicho modelo completo coincida con el modelo reducido.

Nótese que LR verifica $0 \leq LR < +\infty$.

7.5 El test de Wald

El test de Wald constituye otra manera de llevar a cabo test de hipótesis acerca de parámetros sin necesidad de usar el test de la razón de verosimilitud. Sin embargo el test de Wald tan sólo puede ser usado para testar un único parámetro, como por ejemplo ocurre al testar el Modelo 2 frente al Modelo 1. Si tratásemos de testar el Modelo 3 frente al Modelo 2, el test de Wald no sería de aplicación.

Para llevar a cabo el test de Wald hay que tener en cuenta el denominado estadístico de Wald para la variable en cuestión, en este caso denotada por X_j . Para dicha j -ésima variable dicho estadístico de Wald es $\frac{\widehat{\beta}_j}{\widehat{S}_{\beta_j}}$, siendo $\widehat{\beta}_j$ y \widehat{S}_{β_j} las estimaciones máximo verosímiles de β_j y de su correspondiente desviación estándar.

Se verifica que $\frac{\widehat{\beta}_j}{\widehat{S}_{\beta_j}} \rightsquigarrow \mathcal{N}(0, 1)$ o lo que es equivalente, $\left(\frac{\widehat{\beta}_j}{\widehat{S}_{\beta_j}}\right)^2 \rightsquigarrow \chi_1^2$. Esta distribución del estadístico de Wald sirve para aceptar o rechazar la hipótesis nula establecida sobre el j -ésimo parámetro,

$$H_0 : \beta_j = 0$$

$$H_A : \beta_j \neq 0$$

7.6 Modelización

Tanto el test de la razón de verosimilitud como el test de Wald son instrumentos a utilizar para llevar a cabo el proceso de construir un modelo de regresión logística a partir de una base de datos.

Cuando la enumeración completa de todos los modelos posibles resulta computacionalmente costosa, se utilizan estrategias de modelización destinadas a encontrar el mejor subconjunto de variables predictoras.

Las estrategias más extendidas son secuenciales:

- a) *Selección hacia adelante*, en la cual en cada etapa se añade la mejor variable predictora aún no seleccionada.
- b) *Eliminación hacia atrás*, en la cual partiendo del conjunto completo de variables predictoras, se va eliminando en cada etapa la peor variable predictora hasta que las variables que quedan en el modelo son todas ellas pertinentes.
- c) *Modelización paso a paso*, en la cual se combinan las dos estrategias anteriores.

Nótese que mientras que en la estrategia de selección hacia adelante las variables predictoras que son incluidas en el modelo no pueden ser posteriormente eliminadas del mismo y que en la eliminación hacia atrás una variable predictora que ha sido eliminada del modelo no puede más tarde ser incluida en el mismo, en la modelización paso a paso las variables predictoras incluidas en el modelo en una determinada etapa pueden ser excluidas del mismo, al igual que una variable excluida del modelo puede posteriormente ser incluida en el mismo.

Referencias

1. A. Albert, J.A. Anderson (1984). On the Existence of Maximum Likelihood Estimates in Logistic Models. *Biometrika*, **71**, 1-10
2. R. Christensen (1997). *Log-linear Models and Logistic Regression*, Springer
3. D.W. Hosmer, S. Lemeshov (1989) *Applied Logistic Regression*, Wiley
4. D.G. Kleinbaum (1994) *Logistic Regression*, Springer
5. S. Menard (2000). Coefficients of Determination for Multiple Logistic Regression Analysis. *The American Statistician*, **51**, 1, 17-24