INTERVALOS DE CONFIANZA BOOTSTRAP PARA INDICADORES EN REGRESIÓN LOGÍSTICA

Andrés M. Alonso
Departamento de Matemáticas. Universidad Autónoma de Madrid. E.Mail: andres.alonso@uam.es

Resumen
En Silva y Soler (1996) se propone un procedimiento para construir intervalos de confianza de indicadores en regresión logística basados en la desigualdad de Chebyshev. En general, estos intervalos no son exactos y por tanto la cobertura real y la longitud real resultan superiores a las nominales. En este trabajo, presentamos un procedimiento bootstrap para obtener tales intervalos. Presentamos un estudio de Monte Carlo comparando las propiedades en muestras finitas del método bootstrap con el método alternativo.
Palabras claves: Intervalos de confianza, bootstrap, regresión logística.

Abstract
Bootstrap confidence intervals of indicators in logistic regression
Silva and Soler (1996) propose a procedure to build confidence intervals of indicators in logistic regression based on the inequality of Chebyshev. In general, these intervals are not exact and therefore the real coverage and the real length are superior to the nominal. In this work, we present a bootstrap procedure in order to obtain such intervals. We present a Monte Carlo study comparing the finite sample properties of the bootstrap method with the alternative method.

INTRODUCCIÓN

La regresión logística es un procedimiento cuantitativo de gran utilidad para problemas donde la variable dependiente toma valores en un conjunto finito. Ejemplos de sus variados usos son: (*i*) como método descriptivo cuando se desea estudiar la aparición de un determinado evento en un grupo de individuos (v.g. los pacientes de una determinada enfermedad desarrollan un cierto signo propio de ésta, los niños dejan la lactancia materna exclusiva, los licenciados encuentran su primer empleo), (*ii*) como modelo estadístico de pronóstico o clasificación, y (*iii*) como método para la estimación de la razón de disparidad (*odds ratio*) en estudios epidemiológicos. Los libros de Hosmer y Lemeshow (1989), Silva (1994), y Pampel (2000) son excelentes presentaciones de los modelos logísticos en dos de sus principales campos de aplicación: la biometría y la econometría.

En muchas de estas aplicaciones, se proponen y se calculan estimaciones puntuales de índices o indicadores basados en los parámetros estimados de la regresión logística (ver, por ejemplo, Chambless *et al.* (1991), y Silva *et al.* (1992)). Estas estimaciones puntuales presentan un problema evidente: solo damos un valor como estimación del parámetro de interés y éste, en general, no coincide con el valor verdadero del parámetro. Un posible solución es indicar un rango de valores entre los cuales estará el parámetro con una precisión determinada, lo que recibe, usualmente, el nombre de intervalos de confianza. En Silva y Soler (1996) se propone un método general para construir intervalos de confianza para indicadores basados en los parámetros de la regresión logística basados en la desigualdad de Chebyshev. Sin embargo, como es evidente en la expresión (4), estos intervalos no son exactos y por tanto, en general, la cobertura real y la longitud real resultan superiores a los valores nominales.

Un procedimiento alternativo para el cálculo de tales intervalos es el basado en los métodos de remuestreo. Los métodos de remuestreo más populares en la literatura estadística son el jackknife de Quenouille (1949) y Tukey (1958), y el bootstrap de Efron (1979). Los libros de Efron y Tibshirani (1993), Shao y Tu (1995) y Davison y Hinkley (1997) constituyen valiosas referencias sobre los métodos de remuestreo y cubren una amplia gama de aplicaciones.

En el presente trabajo utilizaremos dos procedimientos bootstrap para el cálculo de los intervalos de confianza de los parámetros de la regresión logística y de indicadores basados en éstos. El contenido de este trabajo se divide en las siguientes secciones: en la Sección 2 describimos el procedimiento para construir intervalos de confianza basado en la desigualdad de Chevyshev; en la Sección 3 desarrollamos el procedimiento bootstrap, y mediante un estudio de Monte Carlo lo comparamos con el método propuesto en Silva y Soler (1996). Finalmente, en la Sección 5, ilustramos el comportamiento de los métodos propuestos en un ejemplo con datos reales.

INTERVALOS DE CONFIANZA BASADOS EN LA DESIGUALDAD DE CHEBYSHEV

El modelo logístico que consideraremos en esta sección, establece la siguiente relación entre la probabilidad de que ocurra un suceso (que denotaremos por $Y = 1$) en estudio y un conjunto de p variables independientes ($X_1, X_2, ..., X_p$):

$$\Pr\{Y=1 \mid X_1=x_1, X_2=x_2, \ldots, X_{p1}=x_p\} = \frac{1}{1+\exp(-\beta_0-\beta_1 x_1-\beta_2 x_2-\cdots--\beta_p x_p)} \quad (1)$$

donde $\beta = (\beta_0, \beta_1, ..., \beta_p)$ es un vector de $p + 1$ parámetros reales. Sea $I = f(\beta) = f(\beta_0, \beta_1, ..., \beta_p)$ el indicador de interés basado en los parámetros del modelo logístico, siendo $f(\cdot)$ una función dos veces diferenciable. En Silva y Soler (1996) se propone el siguiente procedimiento para el cálculo de intervalos de confianza (1 - α) %:

Sea $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$ un estimador del vector de parámetros β. Obtener el desarrollo de Taylor de primer orden (despreciando el término de resto) de la función $f(\cdot)$ alrededor del punto β:

$$f(\hat{\boldsymbol{\beta}}) = f(\boldsymbol{\beta}) + \sum_{i=0}^{p} \frac{\partial f}{\partial \beta_i}(\hat{\beta}_i - \beta_i). \quad (2)$$

De la expresión anterior, se obtiene la siguiente aproximación del error cuadrático de $f(\hat{\boldsymbol{\beta}})$ como estimador de $f(\beta)$:

$$ECM(f(\hat{\boldsymbol{\beta}})) = \sum_{i=0}^{p}\sum_{i=0}^{p} \frac{\partial f}{\partial \beta_i}\frac{\partial f}{\partial \beta_j}(\hat{\beta}_i - \beta_i)(\hat{\beta}_j - \beta_j). \quad (3)$$

Utilizando la desigualdad de Chebyshev, se obtiene:

$$\Pr\{f(\hat{\boldsymbol{\beta}}) - q\sqrt{ECM(f(\hat{\boldsymbol{\beta}}))} < f(\boldsymbol{\beta}) < f(\hat{\boldsymbol{\beta}}) + q\sqrt{ECM(f(\hat{\boldsymbol{\beta}}))}\} \geq 1 - \frac{1}{q^2}, \quad (4)$$

donde q es tal que $1 - 1/q^2 = 1 - \alpha$.

Para la implementación práctica del procedimiento anterior bastará con obtener las expresiones de las derivadas $\frac{\partial f(\boldsymbol{\beta})}{\partial \beta_i}$ y sustituirlas en (4) por sus estimadores $\frac{\partial f(\hat{\boldsymbol{\beta}})}{\partial \beta_i}$. En Silva y Soler (1996) se obtienen las expresiones para tres índices relacionados con la lactancia materna.

INTERVALOS DE CONFIANZA BOOTSTRAP

El bootstrap propuesto por Efron (1979) tiene como pieza clave la utilización del principio de analogía (*plug-in*) que constituye uno de los métodos más simples utilizados para obtener un estimador de un parámetro $\beta = \beta(P)$ donde P es el modelo estadístico postulado. Un estimador plug-in o análogo es $\hat{\beta} = \beta(\hat{P})$, donde $\hat{P}$ es un estimador de P. En lo que sigue combinamos el modo en que Efron y Tibshirani (1993), y Shao y Tu (1995) presentan el método bootstrap en una situación general:

> Sea $Z = (Z_1, Z_2, ..., Z_n)$ un conjunto de datos (no necesariamente i.i.d.) generados por el modelo estadístico P, y sea $T(Z)$ el estadístico cuya distribución L(T ; P) deseamos estimar. El método bootstrap propone como estimador de L(T ; P) a la distribución $L^*(T^*; \hat{P}_n)$ del estadístico $T^* = T(Z^*)$, donde Z^* es un conjunto de datos generado por el modelo estimado $\hat{P}_n$.

Notemos que si $\hat{P}_n = P$, entonces las distribuciones L(T ; P) y $L^*(T^*; \hat{P}_n)$ coinciden. De manera que si tenemos un buen estimador de P, es lógico suponer que $L^*(T^*; \hat{P}_n)$ se aproximará a L(T ; P). Concentrándonos en el caso i.i.d. donde el modelo P es la función de distribución F que sigue Xi, la forma en que hemos presentado el bootstrap incluye el *bootstrap clásico* que corresponde a estimar F por la función de distribución empírica $\hat{F}_n = F_n$, el *bootstrap suavizado* para una F continua que utiliza $\hat{F}_n = F_{n,h}$ donde $F_{n,h}$ es la función de distribución asociada a un estimador de la función de densidad f, y el *bootstrap paramétrico* cuando se supone que la función de distribución pertenece a una familia paramétrica F_β y se utiliza $\hat{F}_n = F_{\hat{\beta}}$, donde $\hat{\beta}$ es un estimador de β.

En las secciones 3.1 y 3.2 desarrollamos dos procedimientos bootstrap para el caso de un modelo logístico que pueden interpretarse como la versión clásica y la paramétrica, respectivamente.

Intervalos de confianza bootstrap clásico

En esta sección, consideraremos a $Z = (Y, X) = ((Y_1, X_1), (Y_2, X_2), ..., (Y_n, X_n))$ como independientes e igualmente distribuidos según la función de distribución (modelo) F, y como estadístico de interés a $I = f(\beta)$ cuyo estimador es $\hat{I} = f(\hat{\boldsymbol{\beta}})$. El estimador de F que utilizaremos es la función de distribución empírica $\hat{F}_n$, de esta manera una muestra $Z^* = (Y^*, X^*)$ i.i.d. distribuida según $\hat{F}_n$ corresponderá a un muestreo con reemplazamiento de entre los posibles pares (X_i, Y_i) con $i = 1, 2, \ldots, n$ (este procedimiento recibe también el nombre de *bootstrap por pares*).

El método bootstrap propone como estimador de la distribución de $\hat{I}$ a la distribución del estadístico $\hat{I}^* = f(\hat{\boldsymbol{\beta}}^*)$, donde $\hat{\boldsymbol{\beta}}^*$ es el vector de parámetros estimado en el modelo logístico (1) utilizando el conjunto de datos generados $Z^* = (Y^*, X^*)$. En general no se tienen fórmulas explícitas de los estimadores bootstrap, y el problema se suele abordar mediante aproximaciones de Monte Carlo como la siguiente:

Generar B muestras independientes $Z^{*\,(b)}$ a partir de $\hat{F}_n$, con $b = 1, 2, \ldots, B$.

Obtener $\hat{\boldsymbol{\beta}}^{*(b)}$ para $b = 1, 2, \ldots, B$, utilizando las B muestras generadas en el paso anterior, y calcular el indicador $\hat{I}^{*(b)} = f(\hat{\boldsymbol{\beta}}^{*(b)})$.

Aproximar la distribución $L^*(x)$ por:

$$\hat{L}^*(x) = B^{-1} \sum_{b=1}^{B} \mathbf{1}\{\hat{I}^{*(b)} \leq x\}. \tag{5}$$

Finalmente, obtenemos un intervalo de confianza $(1 - \alpha)$ % para I mediante:

$$[Q^*(\alpha/2), Q^*(1- \alpha/2)] \tag{6}$$

donde $Q^*(\cdot) = \hat{L}^{*-1}(x)$ son los cuantiles de la estimación de la función de distribución bootstrap $\hat{L}^*(x)$. En el Anexo 1, presentamos una rutina implementada en MatLab que desarrolla el algoritmo anterior.

Intervalos de confianza bootstrap paramétrico

En esta sección, especificaremos como modelo generador P de las observaciones (Y, X) a un modelo binomial $B(1, p_x)$ con:

$$\Pr\{Y = y \mid X = x\} = \binom{1}{y} p_x^y (1 - p_x)^{1-y} \tag{7}$$

donde $p_x = \frac{1}{1+\exp(-\beta_0 - \beta_1 x_1 - \ldots - \beta_p x_p)}$ coincide con la relación logística (1). De esta manera, una vez que se obtienen los estimadores $\hat{\boldsymbol{\beta}}$ del vector de parámetros, podemos obtener muestras (Y^*, X^*) del modelo estimado $\hat{P}$ mediante el siguiente procedimiento:

Sea $\boldsymbol{X}^* = (X_1^*, X_2^*, \ldots, X_n^*)$ una muestra de la función de distribución empírica de $\hat{F}_X(x) = n^{-1} \sum_{i=1}^{n} 1\{X_i \leq x\}$.

Para cada X_i^* con $i = 1, 2, \ldots, n$, obtener una observación Y_i^* que sigue una distribución binomial $B(1, p_{X_i^*})$.

En ocasiones las variables independientes X corresponden a un diseño fijo, y en tal caso omitimos el primer paso del procedimiento anterior. El resto del cálculo de intervalos de confianza para el indicador I es similar al presentado en la Sección 3.1. En el Anexo 2, presentamos una rutina implementada en MatLab que desarrolla el algoritmo anterior.

RESULTADOS DE SIMULACIÓN

En esta sección, presentamos los resultados de un estudio de simulación, en el que comparamos los intervalos de confianza bootstrap con los intervalos propuestos por Silva y Soler (1996). En la comparación, utilizaremos el siguiente modelo motivado por el Ejemplo 1 desarrollado en el libro de Hosmer y Lemeshow (1989):

$$\Pr\{Y = 1 \mid X = x\} = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)}, \tag{8}$$

donde Y es el estado respecto a enfermedades coronarias, X es la edad, y $(\beta_0, \beta_1) = (5.31, -0.11)$. Para X consideraremos dos casos: (A) una distribución uniforme en $\{20, 21, \ldots, 70\}$ y (B) los valores reportados en la Tabla 1.1 de Hosmer y Lemeshow (1989). Consideraremos los siguientes dos indicadores: $I_1 = \beta_1$, e $I_2 = \Pr\{Y = 1 \mid X = 40\}$, es decir, I_1 es el parámetro asociado a la variable edad, e I_2 es la probabilidad de tener una enfermedad coronaria en un individuo de 40 años. Utilizaremos el ejemplo estudiado en Silva y Soler (1996) sobre lactancia materna en Cuba, que corresponde con el ejemplo 14.2 del libro de Silva (1994). En ese estudio se obtiene una muestra de 1000 niños menores de un año y se registra la edad X en días y su status Y respecto a la lactancia materna exclusiva ($Y = 1$ si en el momento de la encuesta el niño mantenía la lactancia materna). En primer lugar, calculamos los intervalos de confianza basados en la aproximación normal, en la desigualdad de Chebyshev y los intervalos bootstrap por pares para las probabilidades $\Pr\{Y = 1 \mid X = x\}$ con $x \in \{0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330, 360\}$. En la Tabla 3, aparecen los resultados obtenidos con los tres métodos. Las principales conclusiones son: (a) los intervalos bootstrap son similares a los construidos mediante la distribución (asintótica) normal (en la Figura 1 ilustramos la similitud entre la distribución bootstrap de $I^{*}=\Pr^{*}\{Y=1 \mid X=0\}$ con la correspondiente distribución normal) y (*b*) los intervalos basados en la desigualdad de Chebyshev son mayores que los intervalos gaussianos y bootstrap y en ocasiones tienen extremos negativos (una característica no deseable dada la naturaleza del indicador considerado).

Tabla 1: Resultados de simulación para el indicador I_1.

A	Cobertura media (e.e.)	Cobertura (i./s.)	Longitud media (e.e.)
Teórico	95%	2.50%/2.50%	0.09
Chebyshev	99.44 (0.01)	0.56/0.00	0.20 (0.0004)
Bootstrap I	92.50 (0.17)	3.66/3.84	0.12 (0.0002)
Bootstrap II	92.91 (0.16)	3.56/3.53	0.11 (0.0002)
B	Cobertura media (e.e.)	Cobertura (i./s.)	Longitud media (e.e.)
Teórico	95%	2.50%/2.50%	0.10
Chebyshev	99.43 (0.01)	0.56/0.01	0.22 (0.0003)
Bootstrap I	92.58 (0.17)	3.81/3.61	0.11 (0.0002)
Bootstrap II	93.63 (0.16)	3.38/2.99	0.11 (0.0002)

Tabla 2: Resultados de simulación para el indicador I_2.

A	Cobertura media (e.e.)	Cobertura (i./s.)	Longitud media (e.e.)
Teórico	95%	2.50%/2.50%	0.24
Chebyshev	98.19 (0.04)	1.25/0.54	0.52 (0.0004)
Bootstrap I	92.71 (0.17)	3.78/3.50	0.27 (0.0003)
Bootstrap II	92.60 (0.16)	3.88/3.52	0.26 (0.0002)
B	Cobertura media (e.e.)	Cobertura (i./s.)	Longitud media (e.e.)
Teórico	95%	2.50%/2.50%	0.23
Chebyshev	98.81 (0.01)	0.62/0.56	0.51 (0.0002)
Bootstrap I	94.07 (0.14)	2.99/2.94	0.25 (0.0002)
Bootstrap II	94.41 (0.14)	2.76/2.83	0.25 (0.0001)

Tabla 3: Intervalos de confianza para $\Pr\{Y = 1 \mid X = x\}$.

x	Intervalo gaussiano	Intervalo Chebyshev	Intervalo bootstrap
0	0.6092 0.7367	0.5301 0.8223	0.6085 0.7522
30	0.4968 0.6087	0.4252 0.6816	0.5021 0.6136
60	0.3801 0.4686	0.3225 0.5249	0.3845 0.4702
90	0.2687 0.3413	0.2208 0.3867	0.2687 0.3382
120	0.1744 0.2408	0.1299 0.2813	0.1675 0.2390
150	0.1058 0.1663	0.0643 0.2019	0.0960 0.1682
180	0.0613 0.1127	0.0255 0.1415	0.0523 0.1150
210	0.0346 0.0753	0.0058 0.0968	0.0286 0.0748
240	0.0193 0.0498	-0.0026 0.0648	0.0145 0.0498
270	0.0106 0.0327	-0.0053 0.0426	0.0077 0.0343
300	0.0058 0.0213	-0.0054 0.0277	0.0041 0.0226
330	0.0032 0.0139	-0.0045 0.0178	0.0020 0.0136
360	0.0017 0.0090	-0.0035 0.0114	0.0011 0.0090

A continuación, consideramos el cálculo de los intervalos de confianza basados en la desigualdad de Chebyshev y los intervalos bootstrap para los tres índices relacionados con la lactancia materna estudiados en Silva y Soler (1996): (*i*) el percentil de la edad de abandono de la lactancia $M_\omega = \frac{\ln(\omega)-\ln(1-\omega)-\beta_0}{\beta_1}$ (en particular, la mediana con $\omega = 0.5$), (*ii*) el índice de deserción $ID_t = \frac{\Pr\{Y=1|X=0\}-\Pr\{Y=1|X=x\}}{\Pr\{Y=1|X=0\}}$ (en particular para $x = 90$), y (*iii*) el índice de lactancia

acumulada $ILA = \frac{\ln(1-\Pr\{Y=1|X=0\})-\ln(\Pr\{Y=1|X=120\})}{120\beta_1}$. Los resultados se presentan en la Tabla 4. Como en la Tabla 3 y en los resultados de simulación, los intervalos basados en desigualdad de Chebyshev son más conservadores que los intervalos bootstrap.

La aplicación práctica de los intervalos de confianza bootstrap es la usual, es decir, servirán como complemento a la estimación puntual. Por ejemplo, tenemos que la estimación de la mediana de la edad de abandono de la lactancia materna es 42.32 días y un intervalo de confianza del 95% es [29.58 53.15] lo que nos informa de que el 50% de los niños no es alimentado con leche materna a partir de, aproximadamente, entre 30 y 53 días.
Una ventaja del método bootstrap es que además del cálculo de los intervalos de confianza para un parámetro, como $M_{0.5}$, podemos dar una estimación de distribución muestral. En la Figura 2, representamos distribución bootstrap de $M^*_{0.5}$, podemos reconocer algunas características de esta distribución, por ejemplo, su asimetría a la izquierda, -0.412 con un intervalo de confianza del 95% [-0.547 -0.263].

Tabla 4: Intervalos de confianza para $M_{0.5}$, ID_{120} e *ILA*.

Indicador	Intervalo Chebyshev	Intervalo bootstrap
$M_{0.5}$	15.62 69.03	29.58 53.15
ID_{120}	0.301 0.800	0.494 0.613
ILA	0.336 0.524	0.391 0.469

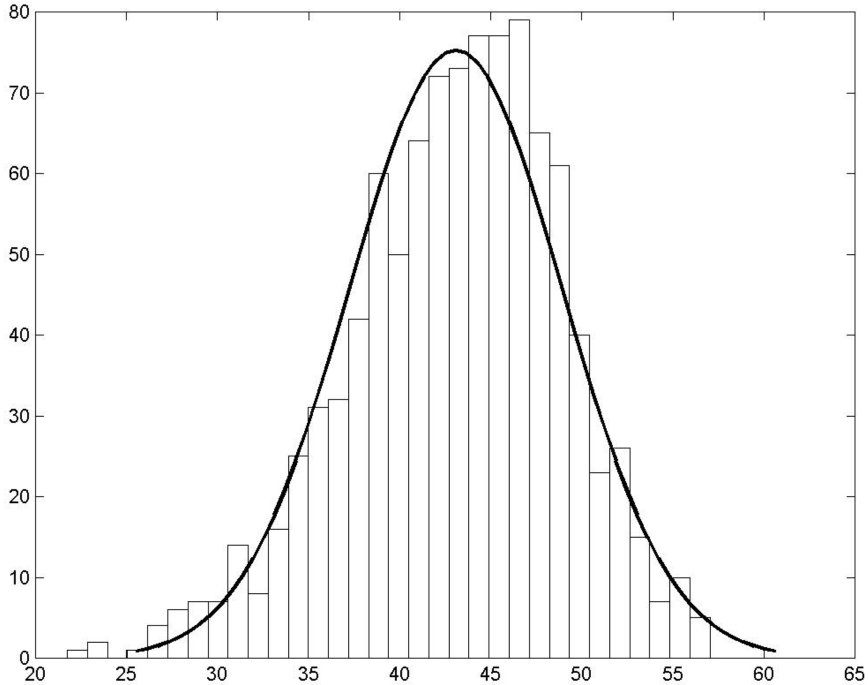

Figura 2: Distribución bootstrap de $M^*_{0.5}$.

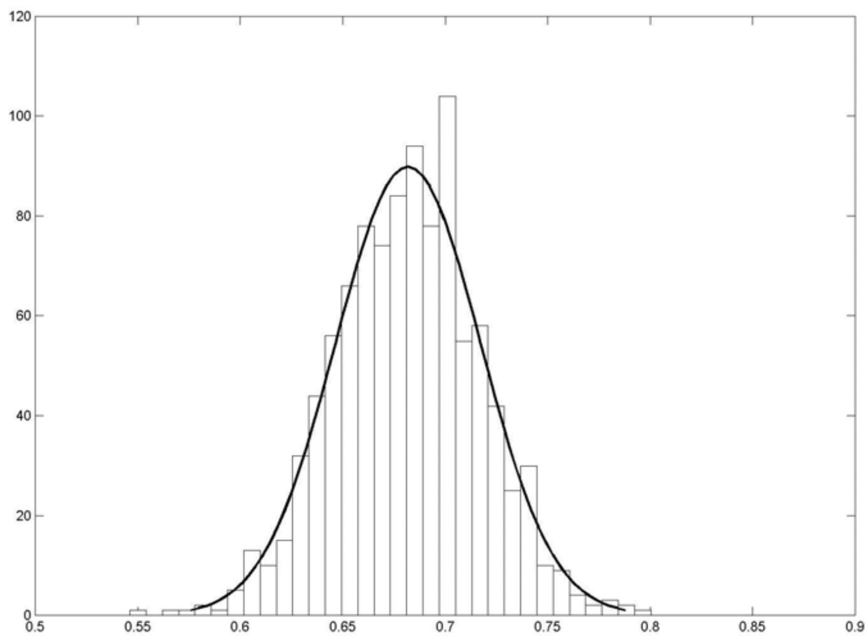

Figura 1: Distribución bootstrap de $I^* = \Pr^*\{Y = 1 \mid X = 0\}$.

REFERENCIAS

Chambless LE, Dobson A, Patterson CC, Raines B. 1991. On the use of a logistic risk score in predicting risk of coronary disease. *Statisitics in Medicine, 9:* 385 - 396.

Davison AC, Hinkley DV. 1997. *Bootstrap Methods and their Applications*. Cambridge University Press. Cambridge.

Efron B. 1979. Bootstrap methods: Another look at the jackknife. *Annals of Statistic.* 7: 1 - 26.
Efron B, Tibshirani RJ. 1993. *Introduction to the bootstrap.* Chapman & Hall. New York.
Hosmer DW, Lemeshow. 1989. *Applied Logistic Regression.* John Wiley & Sons. New York.
Pampel FC. 2000. *Logistic regression. A primer*, Series on Quantitative Applications in the Social Sciences, Sage Publications. Thousand Oaks. USA.
Quenouille M. 1949. Approximation test of correlation in time series. *Journal of the Royal Statistical Society. Series B*, 11: 18 - 84.
Shao J, Tu D. 1995. *The Jackknife and Bootstrap*. Springer-Verlag. New York.
Silva LC. 1994. *Excursión a la regresión logística en ciencias de la salud.* Editorial Díaz de Santos. Madrid.
Silva LC, Amador M, Valdés F. 1992. Discontinuity indices: A tool for epidemiological studies on breast feeding. *International Journal of Epidemiolog.*, 24: 965 - 969.
Silva LC, Soler S. 1996. Intervalos de confianza para indicadores basados en los parámetros de la regresión logística, *Cuadernos de Bioestadística y sus Aplicaciones Informáticas*. **14:** 47 - 61.
Tukey J. 1958. Bias and confidence in not quite large samples, *Annals of Mathematical Statistics*. 29: 614.

ANEXO 1

```
function [Ri, Rs, Rp] = bcinterval(Yresponse, Xdesign, Xdata, B, Alpha)

%
% Rutina para obtener los intervalos de confianza bootstrap de la probabilidad
% p = 1/(1 + exp(-Xdata*Bparameters)), ver Fórmula (8).
%
% Se implementa el procedimiento bootstrap por pares de la Subsección 3.1.
%
% Input:
% ------
% Yresponse : valores de la variable dependiente (binario n x 1).
% Xdesign : valores de las variables independientes (real n x p).
% Xdata : valores de las variables independientes a evaluar (real 1 x p).
% B : número de muestras bootstrap (entero 1 x 1).
% Alpha : (1 - Alpha) nivel de confianza (real 1 x 1).
%
% Output:
% -------
% [Ri, Rs] : intervalo de confianza (1 - alpha)% (real 1 x 2).
% Rp : Estadísticos bootstrap (real B x 1).
%


[n p] = size(Xdesign);
x = Xdesign(:,2);
y = Yresponse;
isx = 1;
t = ones(n,1);
[dev0,idf0,b0,irank0,se0,cov0,v0,ifail0] = g02gbf(x,isx,y,t);

% Fijando los valores por defecto de los parámetros opcionales de la función
% g02gbf que estima el modelo logístico.
link = 'g';
wt = zeros(n,1);
mean1 = 'm';
offset = 'n';
ip = g02gbf10(isx,mean1);
v = zeros(n,ip+7);
```

```
tol = sqrt(eps);
maxit = 100;

% Inicializando los vectores bootstrap.

Bstar = zeros(p, 1);
pstar = zeros(B,1);
Ystar = zeros(n,1);
Xstar = zeros(n,1);

% Selección de los índices que definen las muestras bootstrap.
Irandom = unidrnd(n, n, B);

for i = 1:B
  % Notemos que al coincidir los índices en x e y, se seleccionan pares de
  % observaciones. Paso 1 del algoritmo de la Subseccción 3.1.
  Ystar = y(Irandom(:,i), 1);
  Xstar = x(Irandom(:,i), 1);

  % Cálculo de los parámetros bootstrap de la regresión logística. Paso 2 del
  % algoritmo de la Subseccción 3.1.
  [dev1,idf1,Bstar,irank1,se1,cov1,v1,ifail1] = g02gbf(Xstar,isx,Ystar,t, ...
    link,wt,mean1,offset,v,tol,maxit);
  % Calculo de los estadísticos bootstrap.

  pstar(i) = 1/(1 + exp(-Bstar(1)-Xdata(2)*Bstar(2)));
end

pstar = sort(pstar);


% Obtención del intervalo de confianza. Fórmula (6) del artículo.
Ri = pstar(floor(0.5*Alpha*B));
Rs = pstar(floor((1-0.5*Alpha)*B));

% Estadísticos bootstrap ordenados.
Rp = pstar;
```

ANEXO 2

```
function [Ri, Rs, Rp] = pbcinterval(Yresponse, Xdesign, Xdata, B, Alpha)

%
% Rutina para obtener los intervalos de confianza bootstrap de la probabilidad
% p = 1/(1 + exp(-Xdata*Bparameters)), ver Fórmula (8).
%
% Se implementa el procedimiento bootstrap paramétrico de la Subsección 3.2.
%
% Input:
% ------
% Yresponse : valores de la variable dependiente (binario n x 1).
% Xdesign : valores de las variables independientes (real n x p).
% Xdata : valores de las variables independientes a evaluar (real 1 x p).
% B : número de muestras bootstrap (entero 1 x 1).
% Alpha : (1 - Alpha) nivel de confianza (real 1 x 1).
```

```
%
% Output:
% -------
% [Ri, Rs] : intervalo de confianza (1 - alpha)% (real 1 x 2).
% Rp : Estadísticos bootstrap (real B x 1).
%
% Función auxiliar:
% -----------------
% sample.m
%

[n p] = size(Xdesign);
x = Xdesign(:,2);
y = Yresponse;
isx = 1;
t = ones(n,1);
[dev0,idf0,b0,irank0,se0,cov0,v0,ifail0] = g02gbf(x,isx,y,t);

% Fijando los valores por defecto de los parámetros opcionales de la función
% g02gbf que estima el modelo logístico.
link = 'g';
wt = zeros(n,1);
mean1 = 'm';
offset = 'n';
ip = g02gbf10(isx,mean1);
v = zeros(n,ip+7);
tol = sqrt(eps);
maxit = 100;

% Inicializando los vectores bootstrap.

Bstar = zeros(p, 1);
pstar = zeros(B,1);
Ystar = zeros(n,1);
Xstar = zeros(n,1);

% Selección de los índices que definen las muestras bootstrap.
Irandom = unidrnd(n, n, B);


for i = 1:B
  % Selección de los valores de las variables independientes que forman la
  % muestra bootstrap. Paso 1 del algoritmo de la Subsección 3.2.

  Xstar = Xdesign(Irandom(:,i), 2);

  % Obtención del valor de la variable dependiente para el valor obtenido en
  % el paso anterior. Paso 2 del algoritmo de la Subsección 3.2.
  Ystar = sample(b0, [ones(n,1) Xstar]);

  % Cálculo de los parámetros bootstrap de la regresión logística.
  [dev1,idf1,Bstar,irank1,se1,cov1,v1,ifail1] = g02gbf(Xstar,isx,Ystar,t, ...
    link,wt,mean1,offset,v,tol,maxit);
```

```
   % Cálculo de los estadísticos bootstrap.

   pstar(i) = 1/(1 + exp(-Bstar(1)-Xdata(2)*Bstar(2)));
end

pstar = sort(pstar);

% Obtención del intervalo de confianza. Fórmula (6) del artículo.
Ri = pstar(floor(0.5*Alpha*B));
Rs = pstar(floor((1-0.5*Alpha)*B));

% Estadísticos bootstrap ordenados.
Rp = pstar;


function R = sample(Bparameters, Xdesign);

%
% Función auxiliar a la rutina pbcinterval() que genera una muestra aleatoria
% de una distribución binomial con probabilidad dada por
% p = 1/(1 + exp(-Xdesign*Bparameters)).
%
% Input:
% ------
% Bparameters : Parámetros de la función logística (real p x 1).
% Xdesign : valores de las variables independientes (real n x p).
%
% Output:
% -------
% R : vector de respuestas (binario n x 1).
%

p = max(size(Bparameters));
[n p1] = size(Xdesign);

p = 1./(1 + exp(-Xdesign*Bparameters));
n = ones(n, 1);
R = binornd(n, p);
```

<u>Comentario</u>: Notemos que las rutinas bcinterval() y pbcinterval() pueden modificarse fácilmente para construir intervalos de confianza de otros indicadores basados en los parámetros de la regresión logística, únicamente será necesario modificar la línea donde se calculan los estadísticos bootstrap, esto es, cambiar la línea de comandos pstar(i) = 1/(1 + exp(-Bstar(1)-Xdata(2)*Bstar(2))); por la correspondiente expresión del nuevo indicador de interés. Así, por ejemplo, para los indicadores del ejemplo las expresiones son:

Indicador	Expresión en Matlab
M_{ω}	pstar(i) = (log(omega) – log(1-omega) – Bstar(1))/Bstar(2);
ID_t	pstar(i) = (exp(-Bstar(1)-Xdata(2)*Bstar(2)) – exp(-Bstar(1))) / (1 + exp(-Bstar(1)-Xdata(2)*Bstar(2)));
ILA	pstar(i) = log(exp(-Bstar(1))*(1+exp(-Bstar(1)-Xdata(2)*Bstar(2))) / (1+exp(-Bstar(1))))/(120*Bstar(2));