

**Entregar dos archivos:** uno conteniendo un informe reportando los resultados correspondientes a cada ítem y otro archivo con el código con el que los obtuvo de manera de poder reproducir las salidas.

1. Los datos **birthwt** (library MASS) corresponden a un estudio sobre los factores de riesgo asociados con el bajo peso de niños al nacer. El data frame **birthwt** fue registrado en el Baystate Medical Center, Springfield durante 1986. Tiene los datos correspondientes a 189 niños y 10 columnas en las que se registran las variables:

low: indicator of birth weight less than 2.5 kg.

age: mother's age in years.

lwt: mother's weight in pounds at last menstrual period.

race: mother's race (1 = white, 2 = black, 3 = other).

smoke: smoking status during pregnancy.

ptl: number of previous premature labours.

ht: history of hypertension.

ui: presence of uterine irritability.

ftv: number of physician visits during the first trimester.

bwt: birth weight in grams.

- a) Ajustar un modelo de regresión logística para predecir el bajo peso al nacer usando las variables: lwt, smoke y ht. ¿Son los coeficientes significativos al 5%?
- b) A partir del ajuste anterior construir un intervalo de confianza de nivel asintótico 0.95 para cada uno de los coeficientes del modelo ajustado.
- c) Construir la *tabla de confusión* y computar el porcentaje de aciertos, es decir la *precisión de predicción*.

	Predichos	
Observados	0	1
0		
1		

- d) ¿Cuál sería el valor estimado de la probabilidad de que un niño nazca con bajo peso si su madre tuvo un peso antes de la última menstruación de 120 libras, no fumó durante y tiene antecedentes de hipertensión?

- e) Hallar un intervalo de confianza de nivel aproximado 0.95 para la probabilidad pedida en el ítem anterior con un método que le garantice que los límites del intervalos estarán comprendidos en el  $(0, 1)$ .
- f) Hallar un intervalo de confianza de nivel aproximado 0.95 para la probabilidad pedida en el ítem anterior basado en el método Delta.
- g) Calcular un intervalo de bootstrap normal de nivel 0.95 para la probabilidad solicitada con  $Nboot = 1000$ .
- h) Calcular un intervalo de bootstrap percentil de nivel 0.95 para la probabilidad solicitada con  $Nboot = 1000$ .
- i) Comparar los intervalos obtenidos en los 4 últimos ítems.

## 2. Simulación y Bootstrap

- a) Simulación: El **nivel de cubrimiento empírico** de un procedimiento es la proporción de veces que los intervalos construidos con dicho procedimiento (utilizando datos simulados) contiene al parámetro de interés  $\theta$ , en las  $Nrep = 500$  replicaciones del experimento.

El objetivo de este ejercicio es estudiar el nivel de cubrimiento empírico y la longitud promedio de distintos métodos. Para ello se computarán intervalos de nivel aproximado 0.95 para la probabilidad en  $x_o = 1$ ,  $p(x_o)$ , mediante los dos procedimientos de aproximación normal (el basado en la monotonía y en el método delta), los dos procedimientos bootstrap.

Para ello, generaremos datos  $(Y_1, X_1), \dots, (Y_n, X_n)$  según un modelo logístico de manera tal que las variables de respuesta  $Y_i$  son dicotómicas,

$$p(x, \beta_0, \beta_1) = E(Y_i | X_i = x) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x)},$$

$X_i \sim N(0, 1)$  y  $(\beta_0, \beta_1) = (1/2, 1/2)$  y  $n = 50$ .

En cada replicación computar un intervalo para  $p(x_o)$

- 1) con un método que le garantice que los límites del intervalos estarán comprendidos en el  $(0, 1)$
- 2) basado en el método delta
- 3) usando el método bootstrap normal
- 4) usando el método bootstrap percentil

Completar la siguiente tabla con el nivel de cubrimiento empírico (NCE) y la longitud promedio en las  $N_{rep}$  replicaciones:

	$IC \subset (0, 1)$	IC Delta	IC Boot Normal	IC Boot Percentil
NCE				
Long. Promedio				