Aplicación del Método Bootstrap para la Estimación del Error Estándar de la Media Muestral

Autor: Lic. Dindo Valdez Blanco

Introducción

En este artículo se presenta una aplicación del método de estimación denominado Bootstrap, específicamente aplicado en la estimación de la desviación estándar del estimador de la media y la mediana. Para realizar la aplicación se desarrolla una macro realizada en el programa Minitab versión 15.0 en español para Windows.

En la mayoría de los estudios estadísticos se utilizan los estimadores clásicos como la media muestral, y en base a la misma se calculan intervalos de confianza para la media poblacional, el problema radica cuando se desea utilizar otro estimador como la mediana muestral, surge entonces la pregunta ¿es posible calcular un intervalo de confianza para la mediana poblacional?, justamente este artículo propone el uso de la metodología Bootstrap como una alternativa de respuesta.

El método Bootstrap

En estadística, el método Bootstrap, es aplicado para estimar la medida de precisión de los estimadores muestrales denominado el error estándar del estimador. Esta técnica permite la estimación de la distribución muestral de cualquier estimador utilizando únicamente un método de remuestreo. Lo anterior es útil para calcular intervalos de confianza y pruebas de hipótesis de estadísticos como la mediana muestral, la correlación muestral, etc.

El método Bootstrap es una buena opción para realizar estimaciones cuando los estadísticos son difíciles de manipular algebraicamente y no se puede encontrar una expresión matemática de sus estadísticas como el error estándar del estimador de la mediana, en vista que dicho estimador está relacionado con los estadísticos de orden, su función de probabilidad depende del tamaño de la muestra y de la función de distribución de donde es extraída la muestra aleatoria.

Estimación del error estándar de la mediana por el método Bootstrap

Para desarrollar el concepto, se generan 50 datos aleatorios de una ley de probabilidad exponencial con media $\beta = 15$, la figura 1 muestra el histograma de los datos obtenidos por simulación.

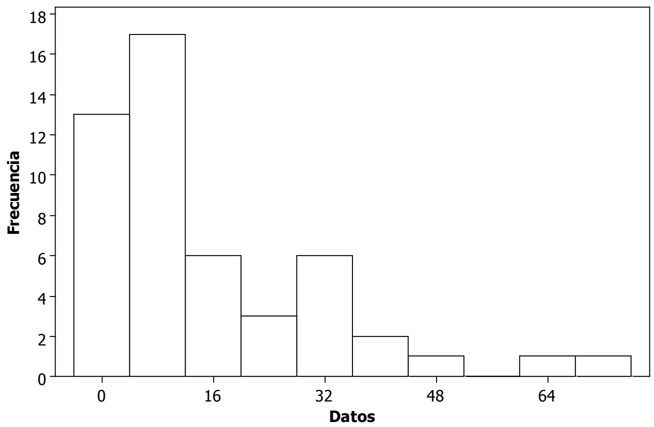

Figura 1. Histograma de la muestra aleatoria exponencial

Considerando este conjunto de datos como una muestra aleatoria de tamaño $n = 50$, se calcula una media muestral $\bar{x} = 15.98$ y la desviación estándar de la muestra $S = 16.31$. De esta manera el estimador de la media de la distribución de probabilidad exponencial $f(x) = \frac{1}{\beta} e^{-x/\beta}$, a partir del método de momentos o del método de máxima verosimilitud es: $\hat{\mu} = \hat{\beta} = \bar{x} = 15.98$.

Luego el error estándar estimado del estadístico $\bar{x}$ está dado por:

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{16.31}{\sqrt{50}} = 2.3066$$

Utilizando el método Bootstrap para realizar tal estimación del error estándar de la media, es necesario emplear el siguiente proceso:

1º. Se selecciona una muestra aleatoria con reemplazo de la muestra original con el mismo tamaño $n = 50$ y se calcula la media de la muestra $\bar{x}$.

2º. Se repite el paso 1º un número de veces determinado digamos *B*, y se calcula media muestral en cada caso: $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_B$

3º. Se calcula la desviación estándar de los B promedios estimados en el paso 2º y se toma este valor como la estimación del error estándar de $\bar{x}$.

En el método Bootstrap es necesario que las muestras generadas en el paso 2º sean del mismo tamaño *n* de la muestra original. Si el tamaño de las muestras es menor que *n*, entonces el procedimiento tiende a sobreestimar el error estándar de $\bar{x}$, de igual manera si el tamaño de las muestras es mayor que *n*, entonces el procedimiento tiende a subestimar la desviación estándar.

Para realizar la aplicación de la metodología Bootstrap con nuestro ejemplo se ha desarrollado una macro[1] en el programa Minitab versión 15.0 para Windows en español, la misma tiene las siguientes sentencias:

Figura 2. Macro que ejecuta el método Bootstrap con la muestra aleatoria

```
Macro                         # Inicio de la macro
bootstrap c1                  # En este paso se nombra la macro como boot1
mcolumn c1 c10 c11            # Se definen las columnas donde se almacenarán los datos
mconstant k1 k2 k3 k4 k5      # Se definen las constantes del programa
erase c10                     # Se borran los datos de la columna c10
let k1 = n(c1)                # Se hace que la consante k1 sea igual al tamaño de la muestra
let k2 = 300                  # Se define el número de muestras B que se generarán
do k4=1:k2                    # Inicio del contador k4 que va de 1 a k2
sample k1 c1 c11;             # Se genera una muestra aleatoria con reemplazo de tamaño
replace.                      # k1 de la columna c1 y se almacena en la columna c11
let k5 = mean(c11)            # Se calcula la media muestral y se almacena en k5
stack c10 k5 c10              # Se almacena la media muestral en la columna c10
enddo                         # Fin del contador
let k3 = stdev(c10)           # Se calcula la desviación estándar de las medias muestrales
print k3                      # Se imprime el error estándar estimado de la media muestral
endmacro                      # Fin de la macro
```

Haciendo correr la macro con el método Bootstrap de la figura 2 se obtiene un valor estimado del error estándar de la media muestral igual a $\hat{\sigma}_{\bar{x}} = 2.3098$, el mismo es un valor cercano al calculado con la clásica fórmula del error estándar de la media. El valor de las remuestras $B = 300$ es suficiente para estimar el error estándar de la media, sin embargo se pueden realizar más replicas en el método de Bootstrap[2]. La metodología Bootstrap puede ser utilizado con cualquier tamaño de muestra sin embargo es evidente que con muestras muy pequeñas el método no es tan confiable, para verificar esto se hizo correr la macro con $n = 10$ y se obtuvo un valor estimado del error estándar igual a 4.8782, de tal forma que esta estimación tiene un margen de error muy alto.

1 - Para ejecutar la macro se debe habilitar el editor de comandos del programa Minitab escribiendo el comando %bootstrap c1.

2 - Efron y Tibshirani indican que muy rara vez es necesario utilizar un número de réplicas mayor a 300.

En este caso no era necesario estimar el error estándar de la media $\bar{x}$ con este método en vista que se tiene la fórmula de la misma. El método Bootstrap es útil cuando no se tiene una fórmula explícita de la desviación estándar del estimador, por ejemplo no existe una fórmula general para calcular el error estándar de la mediana muestral. En este caso si se desea estimar la desviación estándar de la mediana muestral por el método Bootstrap simplemente hay que calcular las medianas de cada réplica en lugar de los promedios (reemplazar la línea de código let k5 = mean(c11) por let k5 = median(c11) en la macro de Minitab), realizando el cambio y haciendo correr la macro se obtiene un valor estimado del error estándar de la mediana muestral igual a $\hat{\sigma}_{me} = 2.1956$. Este resultado es de confiar que se aproxime bastante al valor exacto.

Conclusión

El método de Bootstrap desarrollado para estimar la desviación típica de un estimador $\hat{\theta}$ de un parámetro θ se puede decir que es un método no paramétrico de estimación. En vista que las muestras Bootstrap son obtenidas por muestras repetidas con reemplazamiento de los mismos datos y no se realiza ninguna suposición respecto a la distribución de probabilidad del estimador.

El método Bootstrap es una alternativa eficiente a la inferencia clásica basada en supuestos paramétricos donde es imposible calcular una expresión de la desviación estándar del estimador o tiene una fórmula muy complicada para el cálculo de su error estándar.

Bibliografía

1. Efron, B., Tibshirani, R. (1986). “Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy”. Revista electrónica Statistical Science, 1(1), páginas 54-77.
2. Bradley Efron (1979). “Bootstrap methods: Another look at the jackknife”. Revista electrónica The Annals of Statistics, 7, páginas 1-26.
3. Bradley Efron (1981). “Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods”, Biometrika, 68, páginas 589-599

“Nuestro ánimo se inclina a confiar en aquellos a quienes no conocemos por esta razón: porque todavía no nos han traicionado.”.
Simón Bolívar