

# Métodos de Bootstrapping: Fundamentos y Aplicaciones

David Valcárcel

2025-07-17

## Tabla de contenidos

1. Introducción	1
2. Fundamento teórico	1
3. Bootstrapping no paramétrico	2
3.1. Fundamentos conceptuales . . . . .	2
3.2. Procedimiento ampliado . . . . .	2
3.2.1. 1. Muestra bootstrap . . . . .	2
3.2.2. 2. Cálculo del estadístico . . . . .	2
3.2.3. 3. Réplicas bootstrap . . . . .	2
3.2.4. 4. Estimación de la distribución . . . . .	3
3.3. Implementación en R . . . . .	3
3.4. Referencias recomendadas . . . . .	5

## 1. Introducción

- El **bootstrapping** es un método de remuestreo propuesto por Efron (1979).
- Su objetivo es aproximar la distribución de un estimador estadístico utilizando remuestras de los datos observados.
- Se utiliza para construir **intervalos de confianza**, realizar **contrastos de hipótesis**, y estimar la **precisión** de estadísticas complejas.

---

## 2. Fundamento teórico

- Sea  $\hat{\theta} = s(\mathbf{X})$  un estimador basado en la muestra  $\mathbf{X} = (X_1, \dots, X_n)$ .
- El problema: no conocemos la distribución de  $\hat{\theta}$ .
- Solución: estimarla mediante remuestreo de los datos.

**Principio del bootstrapping:** > La muestra observada es representativa de la población →

simular nuevas muestras “como si” provinieran de ella.

---

### 3. Bootstrapping no paramétrico

#### 3.1. Fundamentos conceptuales

El bootstrapping no paramétrico es un método de remuestreo que permite estimar propiedades de distribuciones (sesgo, error estándar, intervalos de confianza) **sin asumir una forma paramétrica subyacente**. Se basa en el principio de que la muestra observada es la mejor representación disponible de la población subyacente.

**Supuesto clave:** Las observaciones son independientes e idénticamente distribuidas (i.i.d.)

#### 3.2. Procedimiento ampliado

##### 3.2.1. 1. Muestra bootstrap

Dada una muestra original  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ : - Se genera  $\mathbf{X}^*$  mediante muestreo con reemplazo - Cada  $X_i^*$  se selecciona independientemente de  $\mathbf{X}$  - Tamaño de  $\mathbf{X}^* =$  tamaño de  $\mathbf{X}$  ( $n$  observaciones) - **Propiedad clave:** En promedio, el 63.2% de los datos originales aparecen en cada muestra bootstrap ( $1 - (1 - 1/n)^n \approx 1 - e^{-1}$ )

##### 3.2.2. 2. Cálculo del estadístico

- $\hat{\theta}^* = s(\mathbf{X}^*)$  es la versión bootstrap del estadístico
- Ejemplos comunes:
  - Media:  $\bar{X}^* = \frac{1}{n} \sum X_i^*$
  - Mediana
  - Coeficientes de regresión
  - Estadísticos de correlación

##### 3.2.3. 3. Réplicas bootstrap

- Número de iteraciones  $B$  típicos:
  - $B \geq 1000$  para errores estándar
  - $B \geq 10,000$  para intervalos de confianza
- Las réplicas  $\hat{\theta}^{*(1)}, \dots, \hat{\theta}^{*(B)}$  forman una distribución empírica

### 3.2.4. 4. Estimación de la distribución

- Error estándar bootstrap:

$$\widehat{se}_{boot} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B \left( \hat{\theta}^{*(b)} - \bar{\hat{\theta}}^* \right)^2}$$

donde  $\bar{\hat{\theta}}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*(b)}$

- Sesgo estimado:

$$\widehat{sesgo}_{boot} = \bar{\hat{\theta}}^* - \hat{\theta}$$

- Intervalos de confianza (métodos comunes):

1. Percentil:  $[\theta_{\alpha/2}^*, \theta_{1-\alpha/2}^*]$
2. BCa (bias-corrected and accelerated)
3. Normal:  $\hat{\theta} \pm z_{\alpha/2} \cdot \widehat{se}_{boot}$

### 3.3. Implementación en R

```
library(boot)

# 1. Función para calcular el estadístico
media_boot <- function(data, indices) {
  muestra <- data[indices] # Remuestreo con índices
  return(mean(muestra))
}

# 2. Datos originales (ejemplo: 100 observaciones)
datos <- rnorm(100, mean = 10, sd = 2)

# 3. Configurar bootstrap (B = 2000 réplicas)
set.seed(123)
resultados_boot <- boot(
  data = datos,
  statistic = media_boot,
  R = 2000
)
```

```
# 4. Resultados
print(resultados_boot)
```

ORDINARY NONPARAMETRIC BOOTSTRAP

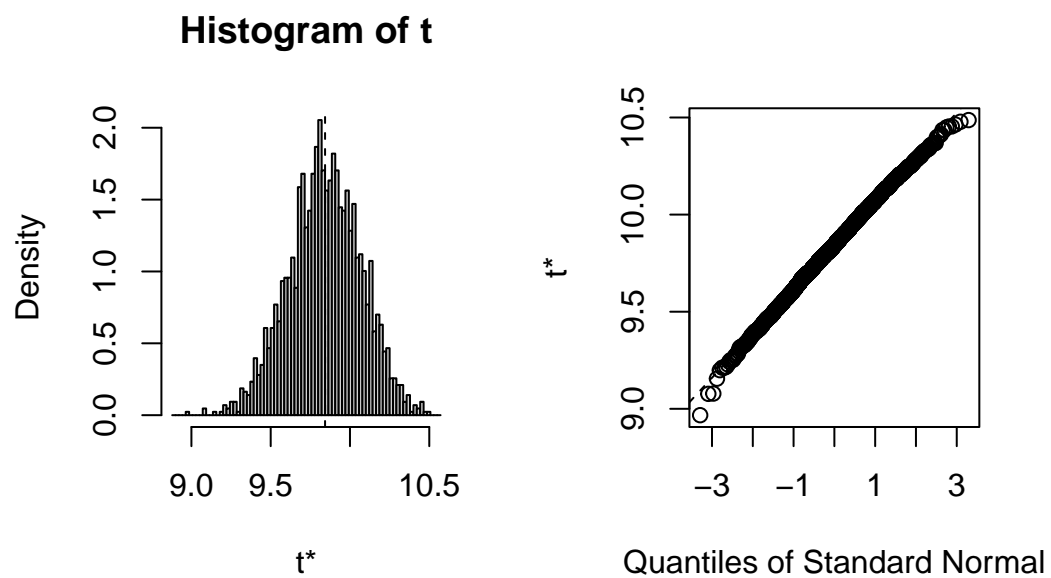
Call:

```
boot(data = datos, statistic = media_boot, R = 2000)
```

Bootstrap Statistics :

	original	bias	std. error
t1*	9.842843	-0.0006870407	0.227421

```
plot(resultados_boot)
```



```
# 5. Intervalo de confianza percentil 95%
boot.ci(resultados_boot, type = "perc")
```

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS

Based on 2000 bootstrap replicates

CALL :

```
boot.ci(boot.out = resultados_boot, type = "perc")
```

Intervals :

Level	Percentile
-------	------------

95%	( 9.393, 10.272 )
-----	-------------------

Calculations and Intervals on Original Scale

### 3.4. Referencias recomendadas

- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall. Efron y Tibshirani (1993)
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.
- Canty, A., & Ripley, B. (2021). boot: Bootstrap R (S-Plus) Functions. CRAN Package.

Efron, Bradley, y Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Springer US. <https://doi.org/10.1007/978-1-4899-4541-9>.