Problemas sobre bootstrap: soluciones

- Media recortada
- Error típico de la varianza

Media recortada

Consideremos la siguiente muestra de tamaño $n = 10$:

Sea $\hat{\theta}$ la media recortada al 40% que se obtiene al eliminar los dos mayores y los dos menores datos y calcular el promedio de los 6 datos restantes. Sea $\hat{\sigma}_R$ el estimador bootstrap de la desviación típica de $\hat{\theta}$ basado en R remuestras. Calcula $\hat{\sigma}_R$ para $R = 10, 100, 1000, 2000$ y usando 10 conjuntos independientes de R remuestras. ¿A qué valor parecen converger los valores obtenidos? ¿Qué número de remuestras te parece suficiente?

Solución:

Escribimos una función que calcula el error típico bootstrap (basado en R remuestras) de la media recortada:

```
media_recortada <- function(x){
  mean(sort(x)[3:8])
}

et_media_recortada <- function(x, R){
  media_original <- media_recortada(x)
  n <- length(x)
  muestras_bootstrap <- sample(x, n*R, rep = TRUE)
  muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)
  medias_bootstrap <- apply(muestras_bootstrap, 2, media_recortada)
  et_media <- sd(medias_bootstrap)
  return(et_media)
}
```

Aplicamos la función a la muestra del enunciado, con $R = 1000$:

```
muestra_original <- c(1, 2, 3.5, 4, 7,  7.3,  8.6,  12.4, 13.8, 18.1)
et_media_recortada(muestra_original, 1000)
```

```
## [1] 2.089302
```

Repetimos 10 veces para $R = 1000$ y para $R = 10$ y comparamos los resultados:

```
veces <- 10

R <- 1000
estimador1000 <- replicate(veces, et_media_recortada(muestra_original, R))
sd(estimador1000)
```

```
## [1] 0.03017246
```

```
mean(estimador1000)
```

```
## [1] 2.021527
```

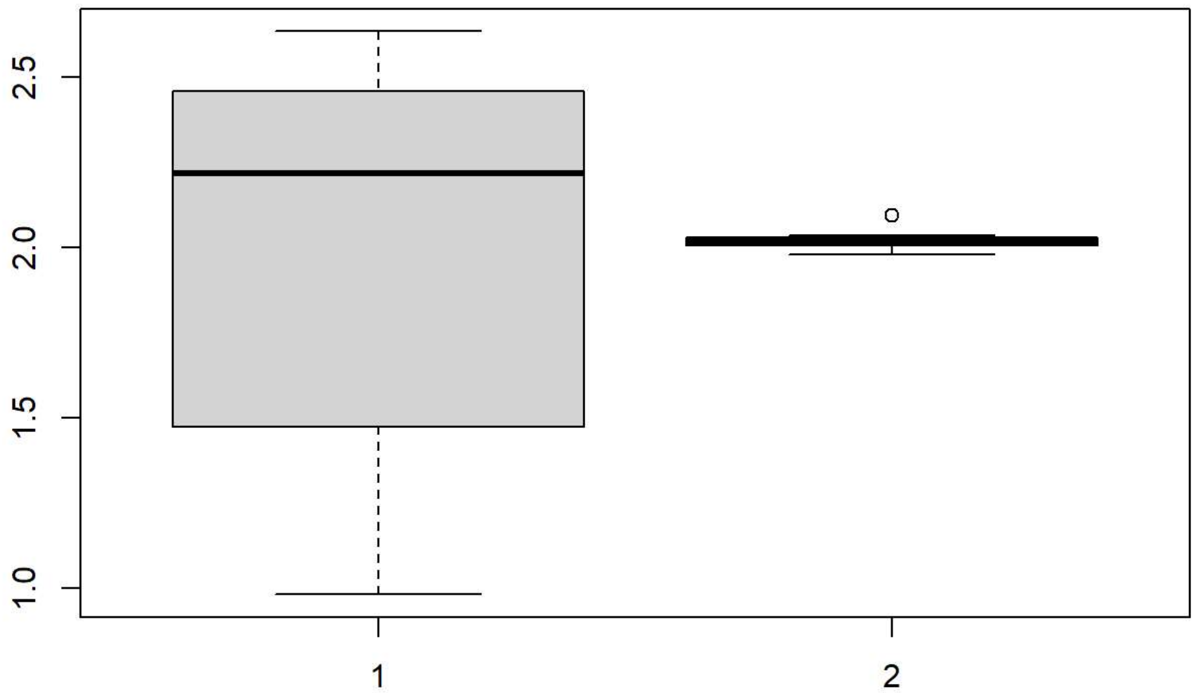
```
R <- 10
estimador10 <- replicate(veces, et_media_recortada(muestra_original, R))
sd(estimador10)
```

```
## [1] 0.5693009
```

```
mean(estimador10)
```

```
## [1] 2.014339
```

```
boxplot(estimador10, estimador1000)
```



Error típico de la varianza

Sea S^2 la varianza muestral de una muestra de vaiid $X_1, \dots X_n$ de una distribución con varianza σ^2.

- Para la muestra del problema anterior se tiene $S^2 \approx 30.84$. Usa bootstrap para determinar el error típico de este estimador de σ^2.
- Compara el resultado con el error típico que darías si, por ejemplo, supieras que los datos proceden de una distribución normal.
- Calcula un intervalo de confianza para σ^2 usando el método bootstrap híbrido. Fija $1 - \alpha = 0.95$.

Solución:

```r
muestra_original <- c(1, 2, 3.5, 4, 7,  7.3,  8.6,  12.4, 13.8, 18.1)
# muestra_original <- rnorm(100)

var_original <- var(muestra_original)
var_original
```

```
## [1] 30.84233
```

```r
n <- length(muestra_original)
R <- 1000

muestras_bootstrap <- sample(muestra_original, n*R, rep = TRUE)
muestras_bootstrap <- matrix(muestras_bootstrap, nrow = n)
var_bootstrap <- apply(muestras_bootstrap, 2, var)

# Error típico bootstrap
sd(var_bootstrap)
```

```
## [1] 10.48629
```

```r
# Error típico basado en normalidad (aplicando Lema de Fisher)
sqrt(2*var_original^2 / (n-1))
```

```
## [1] 14.53922
```

```r
# IC bootstrap híbrido
alfa <- 0.05
T_bootstrap <- sqrt(n) * (var_bootstrap - var_original)
ic_min <- var_original -  quantile(T_bootstrap, 1-alfa/2)/sqrt(n)
ic_max  <- var_original -  quantile(T_bootstrap, alfa/2)/sqrt(n)

c(ic_min, ic_max)
```

```
##    97.5%     2.5%
## 13.47620 53.30013
```

```r
# IC bajo normalidad
(n-1)*var(muestra_original)* c(1/qchisq(1-alfa/2,n-1), 1/qchisq(alfa/2,n-1))
```

```
## [1]  14.59204 102.79295
```