

Bootstrap para las Calificaciones

Armando Sánchez López

08/08/2024

Bootstrap para las Calificaciones

Planteamiento

La pregunta de investigación es determinar si la diferencia entre calificaciones es estadísticamente significativa suponiendo que controlamos solo por tiempo. Utilizaremos el método bootstrap para estimar esta diferencia.

Datos

```
# Calificaciones y grupo
cali <- c(84.06, 79.38, 56.56, 82.19,
         67.19, 63.44, 85.00, 64.38,
         98.75, 76.56, 80.94, 49.06,
         73.69, 70.31, 51.25, 54.06)

grupo <- c(1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 0)

cali_exp <- cali[grupo == 1] # tiempo 90 minutos
cali_cont <- cali[grupo == 0] # tiempo 60 minutos

mean(cali_exp)

## [1] 73.4775

mean(cali_cont)

## [1] 68.625
```

Distribución del Estadístico por Bootstrap

1. Distribución Bootstrap de las Medias

```
# Función para calcular el promedio bootstrap
boot_mean <- function(x){
  return(mean(sample(x, size = length(x), replace = TRUE)))
}

N <- 1000
paso_boot_exp <- replicate(N, boot_mean(cali_exp))
paso_boot_cont <- replicate(N, boot_mean(cali_cont))

mean(paso_boot_exp > paso_boot_cont)
```

```
## [1] 0.778
```

2. Intervalo de Confianza Bootstrap

```
# Función para calcular intervalo de confianza bootstrap
bootstrap_ci <- function(x, alpha = 0.05, B = 1000) {
  boot_samples <- replicate(B, sample(x, replace = TRUE))
  boot_means <- apply(boot_samples, 2, mean)
  ci_lower <- quantile(boot_means, alpha / 2)
  ci_upper <- quantile(boot_means, 1 - alpha / 2)
  return(list(ci_lower = ci_lower, ci_upper = ci_upper))
}

# Intervalo de confianza bootstrap para cali_exp y cali_cont
ci_exp <- bootstrap_ci(cali_exp)
ci_cont <- bootstrap_ci(cali_cont)
```

3. Prueba de Hipótesis

```
# H0: La diferencia entre las calificaciones no es significativa
# H1: La diferencia entre las calificaciones es significativa

# Diferencia entre las medias de los dos grupos
diff_means <- mean(cali_exp) - mean(cali_cont)

# Bootstrap para la diferencia de medias
bootstrap_diff <- replicate(1000, {
  boot_sample_exp <- sample(cali_exp, replace = TRUE)
  boot_sample_cont <- sample(cali_cont, replace = TRUE)
  mean(boot_sample_exp) - mean(boot_sample_cont)
})

# Calculamos el p-valor
p_value <- mean(bootstrap_diff >= diff_means)
```

Resultados

```
cat("Intervalo de confianza para cali_exp:", ci_exp$ci_lower, "-", ci_exp$ci_upper, "\n")

## Intervalo de confianza para cali_exp: 66.98963 - 79.53375
cat("Intervalo de confianza para cali_cont:", ci_cont$ci_lower, "-", ci_cont$ci_upper, "\n")

## Intervalo de confianza para cali_cont: 58.26975 - 80.07
cat("Diferencia de medias:", diff_means, "\n")

## Diferencia de medias: 4.8525
cat("p-valor de la prueba de hipótesis:", p_value, "\n")

## p-valor de la prueba de hipótesis: 0.526
```

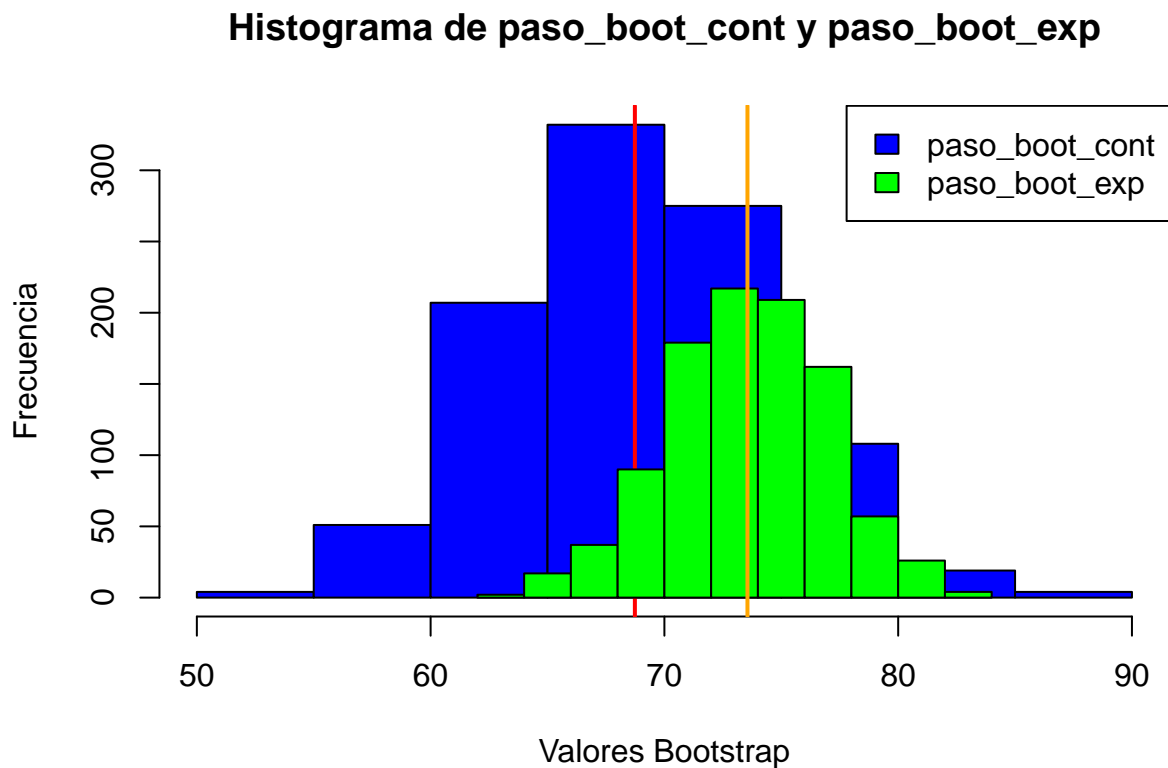
4. Visualización

```
mean_cont <- mean(paso_boot_cont)
mean_exp <- mean(paso_boot_exp)

hist(paso_boot_cont, col = "blue", main = "Histograma de paso_boot_cont y paso_boot_exp",
     xlab = "Valores Bootstrap", ylab = "Frecuencia")
abline(v = mean_cont, col = "red", lwd = 2)

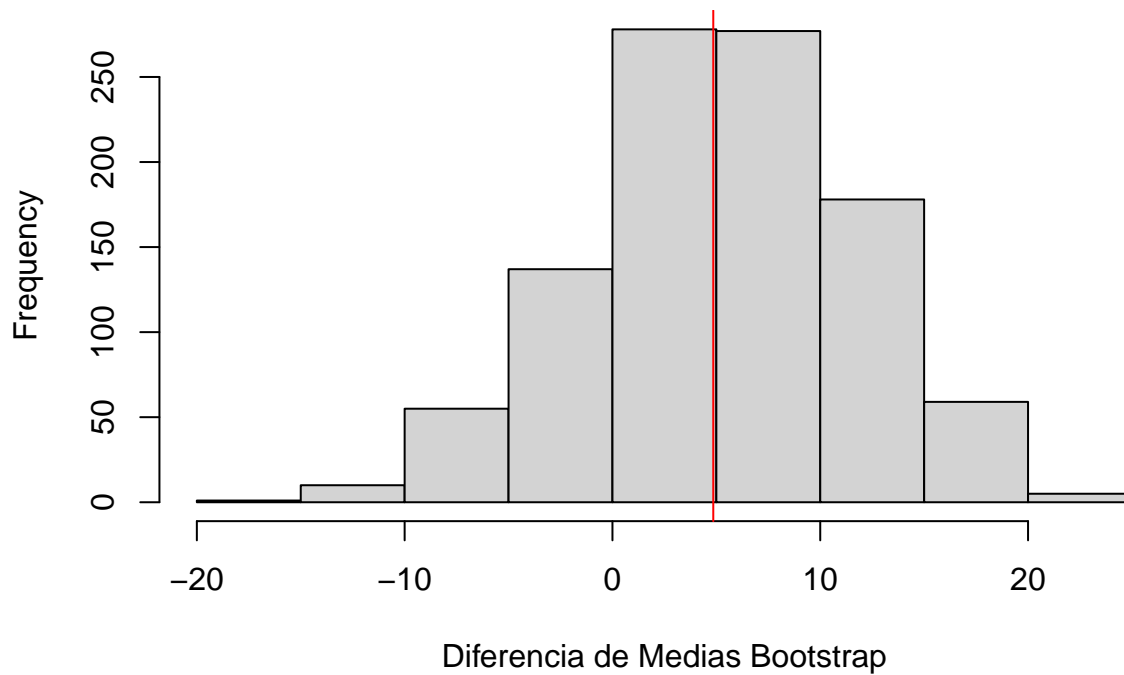
hist(paso_boot_exp, col = "green", add = TRUE)
abline(v = mean_exp, col = "orange", lwd = 2)

legend("topright", legend = c("paso_boot_cont", "paso_boot_exp"), fill = c("blue", "green"))
```



```
hist(bootstrap_diff, main = "Histograma de la Diferencia Bootstrap", xlab = "Diferencia de Medias Bootstrap",
     ylab = "Frecuencia")
abline(v = diff_means, col = "red")
```

Histograma de la Diferencia Bootstrap



5. Combinaciones y p-valor

```
library(gtools)

# Generar todas las combinaciones de 16 elementos en grupos de 8
comb <- combinations(n = 16, r = 8, v = 1:16, set = TRUE, repeats.allowed = FALSE)

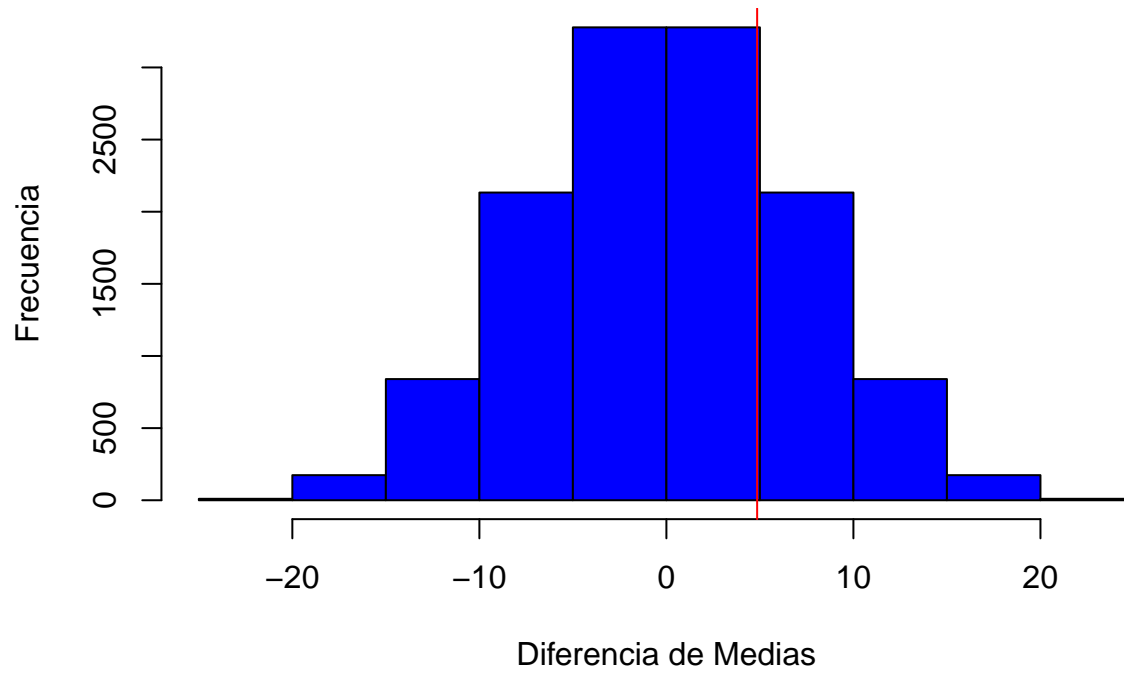
diff_est <- rep(NA, 12870)

for (ii in 1:12870){
  diff_est[ii] <- mean(cali[comb[ii, ]]) - mean(cali[-comb[ii, ]])
}

# Calcular el p-valor
p_val_combinaciones <- mean(abs(diff_est) >= diff_means)

# Visualización
hist(diff_est, col = "blue", main = "Histograma de las Diferencias de Medias Combinadas",
      xlab = "Diferencia de Medias", ylab = "Frecuencia")
abline(v = diff_means, col = "red")
```

Histograma de las Diferencias de Medias Combinadas



Resultados Finales

```
cat("p-valor de las combinaciones:", p_val_combinaciones, "\n")
```

```
## p-valor de las combinaciones: 0.5028749
```