

Teste de permutações

Diferença de médias

2026

Uma grande cadeia de lojas de varejo divide sua área de atuação em 30 regiões. Em 10 regiões foi desenvolvida uma campanha de anúncios publicitários, enquanto que nas demais 20 regiões não houve campanha. Pode ser afirmado que a campanha de anúncios publicitários acarretou um aumento nas vendas?

Os dados abaixo mostram a variação percentual nas vendas em cada região em relação ao mês anterior. Os dados encontram-se em Savage, S. L. (2009), *The Flaw of Averages*, Wiley: Hoboken, Chapter 20.

```
# Com campanha
comcamp <- c(6.6, 9.8, 8.6, -1.3, 4.9, 14.1, 11.4, 13.9, 6.6, -1.1)

# Sem campanha
semcamp <- c(-2.8, 1.9, -7.9, 8.5, 4.1, 3, 3.3, 1.7, 6.6, 7, 3.4, 5,
            8.2, 6.6, 4.1, 10.1, -3.5, -1.2, 10.6, 9.9)
```

```
cat("n1 =", n1 <- length(comcamp), "\n")
```

```
## n1 = 10
```

```
cat("n2 =", n2 <- length(semcamp), "\n")
```

```
## n2 = 20
```

```
n <- n1 + n2
```

```
cat("\n log10(número de permutações) =", log10(choose(n, n1)), "\n")
```

```
##
```

```
## log10(número de permutações) = 7.477772
```

O número de permutações é da ordem de 10^7 (10 milhões).

A diferença entre as médias é usada como estatística de teste.

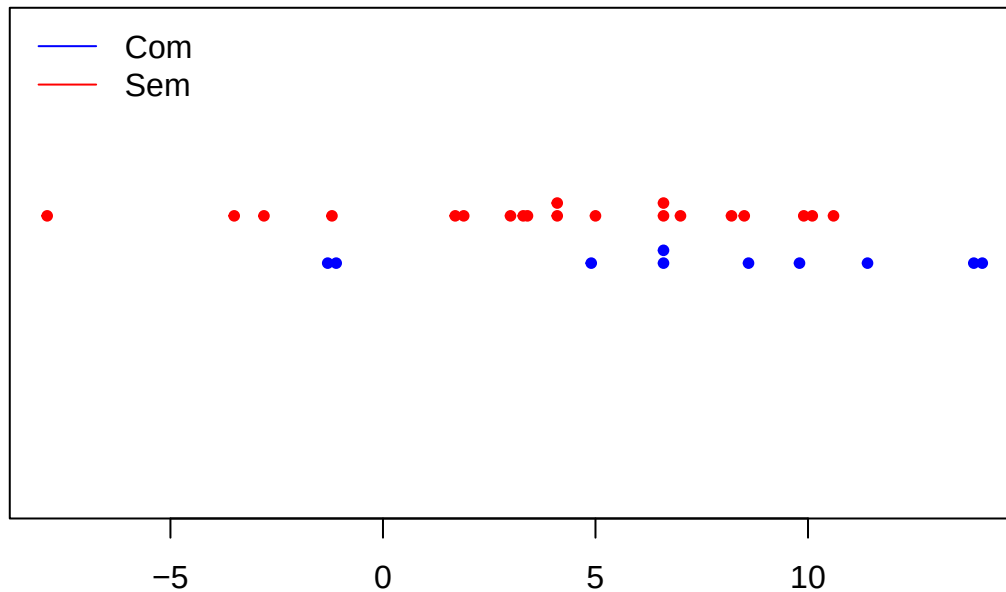
```
dados <- c(comcamp, semcamp) # amostra combinada
dobs <- mean(comcamp) - mean(semcamp)
cat("\n Diferença das médias (%) =", dobs, "\n")
```

```
##
```

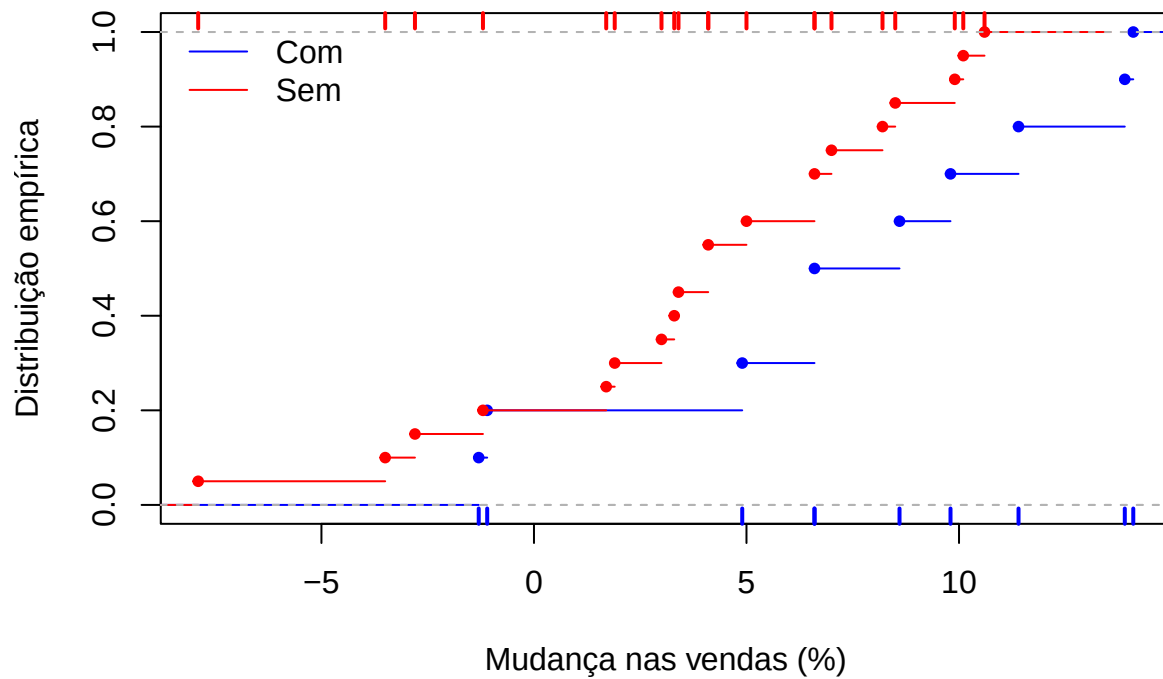
```
## Diferença das médias (%) = 3.42
```

```
# Gráfico de pontos
```

```
stripchart(comcamp, method = "stack", pch = 20, col = "blue", xlim = range(dados))
stripchart(semcamp, method = "stack", pch = 20, col = "red", add = TRUE,
           at = 1.2, xlab = )
legend("topleft", c("Com", "Sem"), lty = 1, col = c("blue", "red"),
      bty = "n")
```



```
# Distribuição empírica
plot(ecdf(comcamp), main = "", pch = 20, col = "blue",
     xlab = "Mudança nas vendas (%)", ylab = "Distribuição empírica",
     xlim = range(dados))
lines(ecdf(semcamp), pch = 20, col = "red")
rug(comcamp, col = "blue", lwd = 2)
rug(semcamp, col = "red", lwd = 2, side = 3)
legend("topleft", c("Com", "Sem"), lty = 1, col = c("blue", "red"),
     bty = "n")
```



Nota 1. Os dois gráficos acima indicam que as duas distribuições diferem apenas em suas médias? Em seguida são calculadas as diferenças de médias nas amostras de permutação.

```

R <- 9999
dl <- c()
set.seed(212)
for (j in 1:R) {
  w <- sample(dados)
  x1 <- w[1:n1]      # x'
  y1 <- w[(n1 + 1):n] # y'
  dl[j] <- mean(x1) - mean(y1)
}

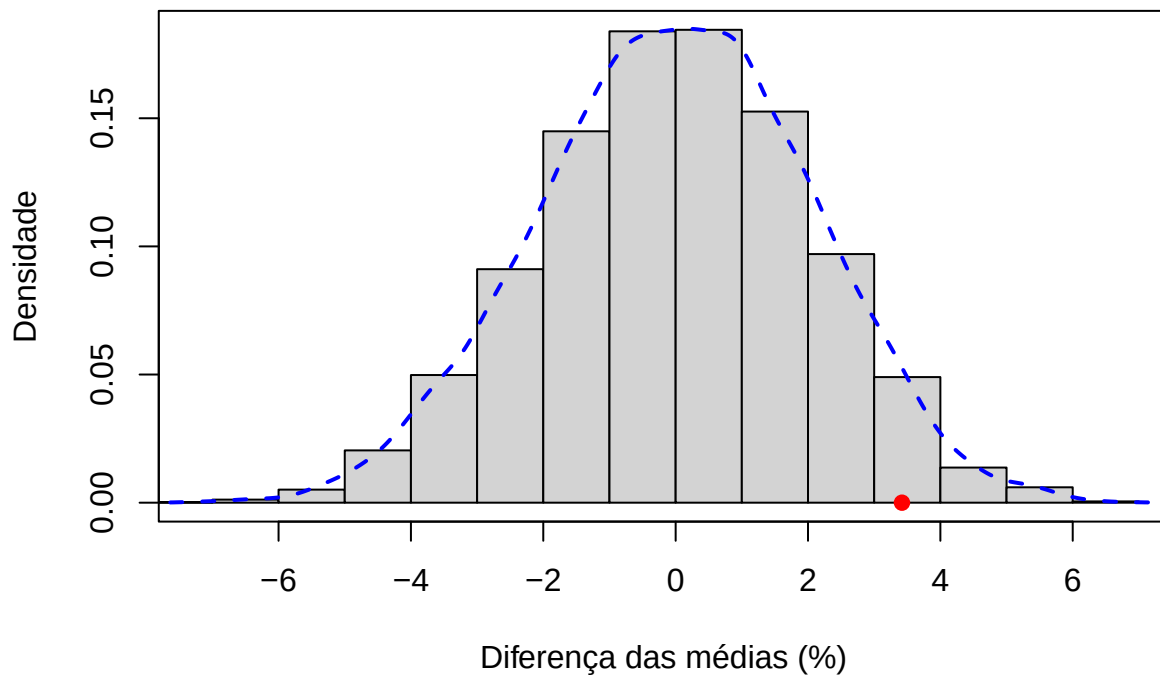
```

As figuras seguintes representam a distribuição da diferença de médias das 9999 permutações (supondo que as duas distribuições são as mesmas).

```

hist(dl, freq = FALSE, main = "", xlab = "Diferença das médias (%)",
      ylab = "Densidade", xlim = range(dl, dobs))
lines(density(dl), col = "blue", lty = 2, lwd = 2)
points(dobs, 0, pch = 19, col = "red")
box()

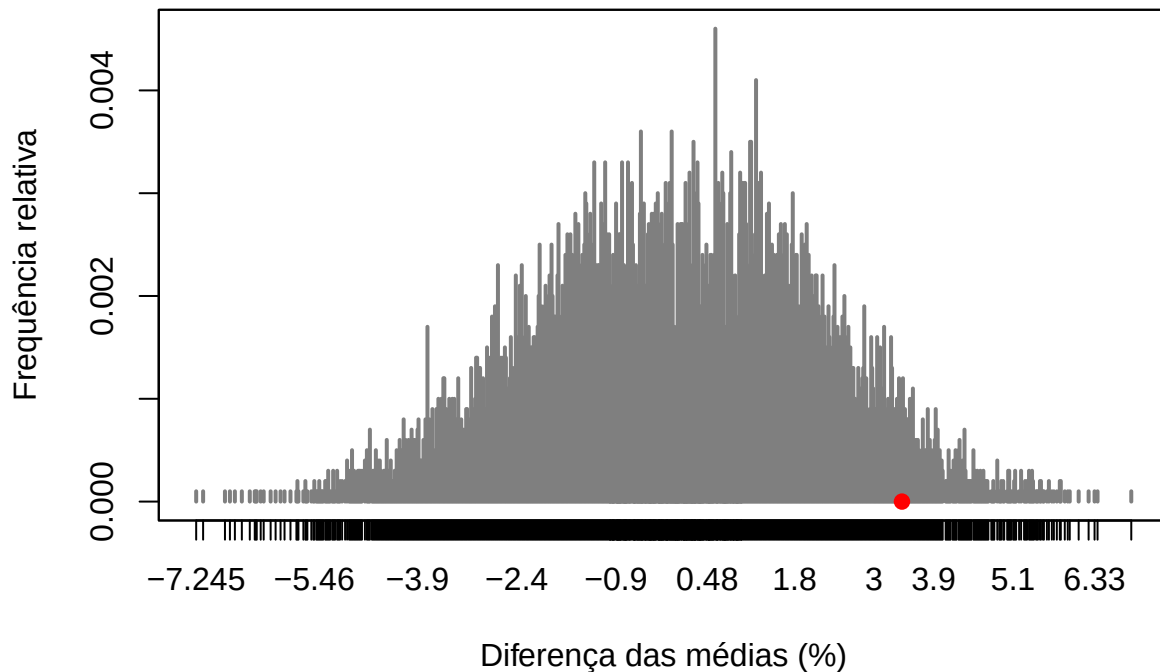
```



```

plot(table(dl) / R, col = "gray50", xlab = "Diferença das médias (%)",
      ylab = "Frequência relativa", xlim = range(dl, dobs))
points(dobs, 0, pch = 19, col = "red")
box()

```



Finalmente, é calculada uma aproximação do valor- p do teste.

```
# H1: média(comcamp) > média(semcamp)
k <- sum(dl >= dobs)
cat("\n valor-p aprox. =", (k + 1) / (R + 1), "\n")
```

```
##
```

```
## valor-p aprox. = 0.0436
```

Nota 2. Com base no resultado acima, pode ser afirmado que a campanha de anúncios publicitários acarretou um aumento nas vendas?

Nota 3. Refaça o exemplo utilizando outros números de permutações (R) na simulação de Monte Carlo.

Nota 4. Pode ser provado que o resultado com a estatística do teste t de Student para amostras independentes com variâncias iguais seria o mesmo; vide Wald and Wolfowitz (1944) Tests based on permutations of the observations, *The Annals of Mathematical Statistics* 15, 358–372 (section 7). Este resultado é mencionado em Ernst, M. D. (2004) Permutation methods: a basis for exact inference, *Statistical Science* 19, 676–685 (section 3.3).