

# Aplicações do Método Bootstrap

ESTAT0090 – Estatística Computacional

Prof. Dr. Sadraque E. F. Lucena

sadraquelucena@academico.ufs.br

# Bootstrap

- Bootstrap é um método que resolve um problema clássico: muitas vezes temos um **estimador** (média, mediana, correlação, coeficiente de regressão) e queremos saber **quão confiável ele é**. Para isso, precisamos do **erro padrão** ou de um **intervalo de confiança**.
- Só que às vezes:
  - não existe fórmula teórica (mediana, quantis).
  - a fórmula é complicada ou exige suposições fortes (normalidade, independência).
  - temos amostras pequenas e não dá pra confiar muito em aproximações assintóticas.
- É aí que entra o bootstrap: a gente **simula** a distribuição do estimador **a partir da própria amostra**.
  - Como? Reamostrando com reposição.
- Nos Exercícios a seguir use sempre a semente `set.seed(123)`.

# Atividade 18.1 – Erro padrão e viés da média

Gere amostras de tamanho  $n = 10, 30, 50$  de:

1.  $X \sim N(0, 1)$  (distribuição simétrica)
2.  $X \sim \text{Gama}(\text{shape} = 2, \text{scale} = 5)$  (distribuição assimétrica)

Para cada amostra:

- a. estime, por bootstrap ( $R = 10.000$ ), o erro padrão e o viés da média amostral;
- b. compare o erro padrão bootstrap com o erro padrão teórico ( $\sqrt{1/n}$  e  $\sqrt{50/n}$ );
- c. O que acontece com o erro padrão quando o tamanho da amostra aumenta?
- d. O erro padrão se comporta diferente para a Normal e para a Gama?

# Atividade 18.2 – Mediana em dados reais

Faça

```
library(boot)
x <- faithful$waiting
```

A variável `x` contém o tempo (em minutos) entre erupções do gêiser Old Faithful no Parque Nacional Yellowstone, Wyoming, EUA.

Calcule a mediana do tempo entre as erupções e apresente erro padrão e o intervalo de confiança bootstrap BCa (que costuma ser o mais confiável, pois corrige viés e assimetria).

- Dica: use a função `boot.ci()` para obter o intervalo bootstrap BCa.

# Atividade 18.3 – Teste de hipóteses para diferença de médias

Verifique se existe diferença significativa ao nível de 5% no consumo de carros manuais e automáticos usando o data frame `mtcars`. O consumo dos veículos foi registrado na variável `mpg`.

```
fit <- lm(mpg ~ wt, data = mtcars)
```

## Atividade 18.4 – Regressão

Ajuste um modelo de regressão linear para `mpg` vs. `wt`. Obtenha o erro padrão do parâmetro associado a `wt` e obtenha um intervalo de confiança de 95% para ele.

```
library(dplyr)
df <- mtcars %>%
  mutate(am = factor(am, labels = c("auto", "manual")))
```

# Fim