

# Recursos para programação eficiente em R e C++

Prof. Wagner Hugo Bonat

1. A média harmônica é dada pela seguinte equação  $Mh = \frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}}$ . Implemente uma função R para o seu cálculo usando a estrutura de repetição `for`. Não use nenhum recurso do R que use vetorização. Gere uma amostra de tamanho 100 com o seguinte código

```
set.seed(123)
y <- rpois(100, lambda = 10)
```

Baseado nesta amostra calcule a média harmonica usando a sua função.

2. Reimplemente a função do exercício 1, porém agora usando tudo que conseguir de recursos de vetorização já disponíveis no R. Faça uma comparação do tempo computacional necessário por cada uma das abordagens. Use o package `bench`.
3. Reimplemente a função do exercício 1, porém agora usando C++ e o pacote `Rcpp`. Novamente compare o tempo computacional entre as três abordagens usando o pacote `bench`.
4. Em estatística testes qui-quadrado são muito populares. O R tem uma função para realizar tal teste para diversas situações. Uma situação comum é para verificar a associação em tabelas de contingência. O exemplo abaixo retirado da documentação tem o objetivo de avaliar a associação entre genero e perfil político.

```
M <- as.table(rbind(c(762, 327, 468), c(484, 239, 477)))
dimnames(M) <- list(gender = c("F", "M"),
                    party = c("Democrat", "Independent", "Republican"))
M
```

```
##      party
## gender Democrat Independent Republican
##      F      762      327      468
##      M      484      239      477
```

O código abaixo executa o teste

```
chisq.test(M)

##
##  Pearson's Chi-squared test
##
## data:  M
## X-squared = 30.07, df = 2, p-value = 2.954e-07
```

Use uma ferramenta de debug para investigar como a estatística de teste é calculada. Reimplemente apenas a estatística de teste em uma função própria. Compare a sua função com a do R em termos de tempo computacional. Suponha que o interesse é apenas obter a estatística de teste.

5. Comparar dois grupos é uma atividade popular em estatística. Considere o conjunto de dados `iris` disponível no R. Suponha que desejamos testar se a variável `Sepal.Length` é em média diferente entre as espécies setosa e versicolor. Para isso vamos usar um teste de aleatorização. Sob a hipótese nula o tamanho médio das sepalas é igual. Isso significa que tanto faz o grupo ao qual a flor pertence. Para

simular desta situação é suficiente juntarmos os dados das duas espécies e sortear aleatoriamente a qual espécie a flor pertence. Para medir a diferença calculamos a média de cada espécie e fazemos a diferença. Vamos repetir esse processo um grande número de vezes e ver como é a distribuição da estatística diferença. Após isso, basta calcular a diferença observada na amostra e identificar qual o percentual de vezes que ocorreu ela ou uma estatística mais extrema para obter o chamado p-valor. Implemente esse procedimento de forma sequencial e em paralelo. Compare o tempo computacional.