

Segunda Lista de Exercícios: Fundamentos de R

Exercício 1. Crie os seguintes vetores:

- (a) $(1, 2, 3, \dots, 19, 20)$;
- (b) $(20, 19, 18, \dots, 2, 1)$;
- (c) $(1, 2, 3, \dots, 19, 20, 19, 18, \dots, 2, 1)$;

Use a função `help` para descobrir o funcionamento da função `rep`. Em seguida, tente utilizar esta função para resolver os seguintes itens:

- (d) $(4, 6, 3, 4, 6, 3, \dots, 4, 6, 3)$, em que há 12 ocorrências do número 4;
- (e) $(4, 6, 3, 4, 6, 3, \dots, 4, 6, 3, 4)$, em que há 13 ocorrências do número 4, 12 ocorrências do número 6 e 12 ocorrências do número 3;
- (f) Crie um vetor cujas as entradas são dadas pelo valor de $e^x \cos(x)$ nos pontos $x = 3, 3.1, 3.2, 3.3, \dots, 6$.

Exercício 2. Calcule:

- (a) $\sum_{i=10}^{100} (i^3 + 4i^2)$;
- (b) $\sum_{i=10}^{25} \left(\frac{2^i}{i} + \frac{3^i}{i^2} \right)$.

Exercício 3. A função `sample` é utilizada para realizar sorteios de um dado conjunto. Exemplo: `sample(x = 1:5, size = 10, replace = TRUE)` sorteará 10 (`size = 10`) números do conjunto/vetor $\{1, 2, 3, 4, 5\}$ (`x = 1:5`) com reposição (`replace = TRUE`). Caso se deseje realizar um sorteio sem reposição deve-se utilizar `replace = FALSE`.

Execute as seguintes linhas de comando:

```
xVec <- sample(0:999, 250, replace=T)
yVec <- sample(0:999, 250, replace=T)
```

Suponha que `xVec` seja representado por (x_1, x_2, \dots, x_n) e que o vetor `yVec` seja representado por (y_1, x_2, \dots, y_n) em que $n = 250$.

- (a) Crie um vetor com as entradas ímpares de `xVec`;
- (b) Crie o vetor $(y_2 - x_1, \dots, y_n - x_{n-1})$;
- (c) Crie o vetor $\left(\frac{\sin(y_2)}{\cos(x_1)}, \frac{\sin(y_3)}{\cos(x_2)}, \dots, \frac{\sin(y_n)}{\cos(x_{n-1})} \right)$;
- (d) Crie o vetor $(x_1 + 2x_2 - x_3, x_2 + 2x_3 - x_4, \dots, x_{n-2} + 2x_{n-1} - x_n)$;

Exercício 4. Este exercício ainda utilizará os vetores `xVec` e `yVec` definidos no Exercício 4. Tente utilizar, a seguir, algumas das seguintes funções: `sort`, `order`, `mean`, `sqrt`, `sum`, `abs`, `max`, `which`.

- (a) Quais são os índices do vetor `yVec` que correspondem a valores maiores do que 600?

- (b) Selecione os valores de `yVec` que sejam maiores do que 600.
- (c) Quais valores do vetor `xVec` correspondem às entradas do vetor `yVec` que são maiores do que 600?
- (d) Crie o vetor $(|x_1 - \bar{x}|^{1/2}, |x_2 - \bar{x}|^{1/2}, \dots, |x_n - \bar{x}|^{1/2})$, em que \bar{x} representa a média do vetor `xVec`.
- (e) Quantos valores do vetor `yVec` estão a uma distância de até 200 do valor máximo deste vetor?
- (f) Quantos números de `xVec` são divisíveis por 2?
- (g) Selecione os valores de `yVec` que estão nas posições 1, 4, 7, 10,...

Exercício 5. Use um loop para escrever uma função chamada `num.impar` cuja entrada seja um vetor `v` e cuja saída retorne quantos números ímpares existem no vetor `v`. Em seguida, escreva uma função `num.impar2` que faça a mesma coisa sem usar um loop.

Exercício 6. Os dois primeiros termos da sequência de Fibonacci são iguais a 1. Os termos subsequentes da sequência são encontrados somando os dois termos imediatamente anteriores. Escreva uma função chamada `fibonacci` cuja entrada seja `n` e cuja saída retorne os primeiros `n` termos da sequência de Fibonacci para qualquer $n \geq 3$. Exemplo:

```
> fibonacci(6)
[1] 1 1 2 3 5 8
```

Exercício 7. Escreva uma função chamada `tamanho.seq` que simule o lançamento de uma moeda com probabilidade p de cara até que a primeira coroa seja vista. A função deve retornar a quantidade de caras que foram obtidas até a obtenção da primeira coroa. A entrada da função deverá ser a probabilidade p de sair cara. Por fim, replique a execução desta função 100 vezes com uma moeda honesta e encontre a maior sequência de caras obtidas nas 100 realizações do experimento.

Exemplo: supondo que coroa seja 0 e que cara seja 1, se a sequência obtida da realização de um experimento foi 1,1,1,1,0 então a função retornará o valor 4.

Exercício 8. Na clássica formulação da sequência de Fibonacci um par de coelhos em idade reprodutiva pode originar apenas outro par novos coelhos. É possível generalizar o problema para que cada par de coelhos em idade reprodutiva possa originar k pares de novos coelhos. Para esse exercício, crie uma função cuja entrada seja n (número de meses) e k (número de novos pares por casal) e cuja saída seja a quantidade total de coelhos no mês n . As intruções sobre a formulação do problema podem ser lidas em: <http://rosalind.info/problems/fib/>

Exercício 9. Para este exercício, você irá precisar das funções `str_detect` e `str_replace` do pacote `stringr`.

- (a) Instale o pacote `stringr`. Em seguida, conheça o funcionamento das funções `str_detect` e `str_replace` a partir dos comandos `?str_detect` e `?str_replace`.
- (b) Uma sequência de DNA é uma série de letras representando a estrutura primária de uma molécula ou cadeia de DNA com a capacidade de carregar informações. As sequências de DNA são construídas a partir do alfabeto {A, C, G, T}, cujos símbolos representam as bases adenina, citosina, guanina e timina. Como exemplo, AAGATGCCGT é uma cadeia de DNA de comprimento 10. Crie uma função cuja entrada seja uma sequência X de DNA e que sua saída seja um vetor de tamanho 4 em que a primeira entrada deste vetor seja o número vezes em que o símbolo A aparece na sequência X, a segunda entrada seja o número vezes em que o símbolo C aparece na sequência X e o mesmo coisa para G e T nas posições 3 e 4 respectivamente. Por fim, aplique a sequência

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGGATT

na função criada.

- (c) Uma sequência de RNA é formada pelo alfabeto {A, C, G, U}. Pode-se obter uma sequência Y de RNA a partir de uma sequência X de DNA trocando o símbolo T na sequência X pelo símbolo U. Assim, se a sequência X de DNA é dada por AAGATGCCGT, então a sequência Y de RNA será AAGAUGCCGU. Crie uma função que transforma uma sequência de DNA em uma sequência de RNA. Por fim, aplique a sequência de DNA do item anterior nesta função.

Exercício 10. Instale o pacote `dslabs` e em seguida leia o conjunto de dados `murders`:

```
install.packages(dslabs)
library(dslabs)
data(murders)
```

- (a) Calcule a média da variável `population`. Qual o estado mais populoso? E qual o estado menos populoso?
- (b) Crie uma nova coluna no data frame `murders` para representar a taxa por estado de mortes por arma de fogo por 10 mil habitantes.
- (c) Imprima o nome dos estados por ordem crescente de taxa de mortalidade.
- (d) Quais estados apresentam uma taxa de mortalidade inferior a 0.06? E superior a 0.5?

- (e) Qual estado possui a maior taxa de mortalidade e qual estado possui a menor taxa de mortalidade?
- (f) Divida o data frames por região. Em seguida, calcule a média e o desvio padrão da taxa de morte para cada região. Qual região possui a maior taxa de mortalidade? E a menor?
- (g) Apresente simultaneamente os boxplots da taxa de mortalidade por região. Interprete os resultados.