

Exercícios 3

Funções

Em **R**, as funções são também objectos e podem ser manipuladas como tal. Têm três principais propriedades:

argumentos Lista dos parâmetros que a função vai receber, em que alguns podem ter um valor por defeito e ser opcionais. Uma forma rápida de consultar os argumentos de uma função é através da função **args**.

corpo Ao conjunto de expressões avaliado na execução da função chama-se corpo da função. É possível visualizar o corpo de uma função através da função **body**.

ambiente Uma função é definida num determinado ambiente (*environment*) activo, o que irá determinar os objectos visíveis e acessíveis pela função.

1. Crie uma função que faça a conversão de graus Fahrenheit (°F) para graus Celsius (°C), sabendo que a sua relação é dada pela Equação 1.

$$^{\circ}C = (^{\circ}F - 32) \div 1,8 \quad (1)$$

Teste a sua função para saber quantos graus Celsius correspondem a 159,3 °F, uma das temperaturas registadas mais altas de sempre na superfície da Terra, no deserto de Lut, no Irão.

2. Expanda a sua função para, opcionalmente, também fazer a conversão de graus Celsius para Fahrenheit. A relação inversa é dada pela Equação 2.

$$^{\circ}F = (^{\circ}C \times 1.8) + 32 \quad (2)$$

3. Atente ao seguinte bloco de código:

```
x = 2
z = 56.3
f = function(x) {
  a = 34
  y = x / 4 * a * z
  y
}
f(21)
a
```

Qual o valor das variáveis **a**, **x** e **z** dentro do ambiente da função? O que aconteceu ao valor da variável **a** após a execução da função?

4. Escreva três funções que recebam um **data.frame** e devolvam, por cada coluna:
 - o máximo;
 - o mínimo;
 - a média.



Figura 1: *Iris sibirica* (fonte: wikimedia.org).

Escreva duas versões para cada caso: uma recorrendo a um ciclo; e outra sem recorrer (explicitamente) a ciclos. Defina ainda o argumento opcional `first.n`, onde o utilizador poderá escolher para encontrar os referidos valores apenas considerando os primeiros n elementos de cada coluna. Teste as suas funções com o conjunto de dados `iris`.

- Assuma que gosta de coleccionar flores de Iris (Figura 1). Dispõe de uma tabela onde anota algumas das propriedades destas flores, registando, para cada exemplar, o valor de cada uma destas propriedades: comprimento e largura da pétala; comprimento e largura da sépala. A Tabela 1 mostra um excerto dessa tabela.

Número do exemplar	Comprimento da Sépala	Largura da Sépala	Comprimento da Pétala	Largura da Pétala	Espécie
1	5,1	3,5	1,4	0,2	Setosa
51	7,0	3,2	4,7	1,4	Versicolor
101	6,3	3,3	6,0	2,5	Virginica

Tabela 1: Excerto da tabela com os registos das flores Iris, disponível em **R** na variável `iris`.

Após uma ida ao campo, uma nova amostra foi colhida. O novo exemplar tem as características indicadas na Tabela 2.

Número do exemplar	Comprimento da Sépala	Largura da Sépala	Comprimento da Pétala	Largura da Pétala
Nova amostra	4,7	3,2	1,7	0,3

Tabela 2: Características da nova amostra de flor Iris.

O problema consiste em descobrir a que espécie pertence.

Este é um problema típico de classificação¹. Noutra disciplina, aprenderá mais sobre os métodos de classificação. Por agora, pode resolver este problema com uma implementação simples do algoritmo dos k -vizinhos mais próximos²:

- Considere que cada característica tem o mesmo peso;
- Cada uma das características pode ser vista como um eixo (ou variável), então é como se tivesse quatro dimensões;
- A nova amostra será classificada como sendo da mesma espécie que o vizinho mais próximo ($k = 1$) no conjunto de dados.
- Considere a distância como a distância Euclidiana³.

Resolva este problema aplicando uma função sobre o `data.frame` `iris`.

¹https://en.wikipedia.org/wiki/Statistical_classification

²https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm

³https://en.wikipedia.org/wiki/Euclidean_distance