

## Folha 1: Conceitos Básicos do R

---

### Vetores e Funções

1. O R possui vários conjuntos de dados que vêm incluídos em diferentes *packages*. Um desses *packages* é o *datasets* que oferece o conjunto de dados *women*.

- (a) Comece por carregar a *package* *datasets* com o comando: `library(datasets)`. Se obter erro a indicar que a *package* não está instalada, use o comando `install.packages(datasets)`.
- (b) Veja o conteúdo do conjunto de dados *women*, escrevendo `women` na consola.
- (c) Para saber a que se refere este conjunto dados, procure *women* na caixa de ajuda do RStudio. Em alternativa, pode executar o comando `help(women)` na consola.
- (d) Crie um vetor com a informação dos pesos executando o comando `pesos <- women[,2]`.
- (e) Sobre o vetor *pesos* obtenha:
  - a média;
  - os pesos superiores 130lb;
  - os pesos superiores à média;
  - o número de mulheres com peso superior à média;
- (f) Altere o vetor *pesos* de forma a que os pesos inferiores a 130lb seja reduzido em 5lb.
- (g) Qual o resultado do comando `range(pesos)` ?
- (h) Sabendo que  $1lb = 0.453592kg$ , escreva uma função `lb2kg` que converta um peso em *lb* para *kg*.
- (i) Teste a sua função com os seguintes comandos: `lb2kg(100)` e `lb2kg("100")`. O que acontece?
- (j) Aplique a sua função ao vetor *pesos*.
- (k) Altere a sua função para que o peso em *kg* seja arredondado a um valor inteiro.

## Data Frames

2. Aceda ao conteúdo do conjunto de dados `Pima.tr` disponível na *package* `MASS`. Sobre estes dados responda às seguintes questões.

- (a) Que informação está representada neste conjunto de dados?
- (b) Quais são as mulheres identificadas como diabéticas?
- (c) Qual é a percentagem de mulheres diabéticas? (Sugestão: use a função `nrow()`.)
- (d) Quais são as mulheres diabéticas e com idade inferior a 40?
- (e) Qual é a concentração de glucose e o IMC (`bmi`) das mulheres com um IMC normal, isto é no intervalo  $]19.1, 25.8]$ ?
- (f) Qual é a concentração de glucose e o IMC (`bmi`) das mulheres com um IMC abaixo ou acima do peso normal?
- (g) Crie uma função que, dada uma idade, devolve o conjunto de mulheres cuja idade é superior à indicada. Por omissão, a idade dada deve ser 40.

## Funções e Manipulação Básica de Dados

3. Considerando o *dataset* `women` da pergunta 1, Crie um ficheiro em R com o nome `'IMC.R'` com a instrução `library(datasets)`. Adicione ao ficheiro o conteúdo solicitado abaixo. Sempre que precise de executar o ficheiro deve escrever na consola do R `source('IMC.R')` ou então carregar no botão `Source` e `Run`.

- (a) Crie uma função `lb2kg` que dado um peso em *lb* devolve a peso em *kg* arredondado a um inteiro ( $1lb = 0.453592kg$ ).
- (b) Crie uma função `in2m` que dado uma altura em *inches* devolve a altura em *metros* arredondada a 2 casas decimais ( $1in = 0.0254m$ ).
- (c) A partir do *dataset* `women` e das duas funções acima, crie uma nova matriz `m` com a mesma informação contida em `women` mas com métricas diferentes.
- (d) Calcule o valor do IMC ( $peso/altura^2$ ), arredondado a uma casa decimal, para cada uma das linhas da matriz.
- (e) Utilize a função `cbind()` para criar uma nova matriz `m1` que resulta de anexar à matriz `m` uma 3a coluna com o cálculo anterior.

- (f) Crie uma função `calcIMC` que dado um peso em kg e uma altura em metros, calcula o valor de IMC e retorna uma resposta de acordo com o que é apresentado na tabela abaixo. Teste a sua função na linha de comandos.

| IMC nas Mulheres | Condição   |
|------------------|------------|
| < 19.1           | pesoAbaixo |
| 19.1 – 25.8      | pesoNormal |
| > 25.8           | pesoAcima  |

- (g) Usando as funções `cbind()`, `sapply()`, crie uma matriz `m2` que resulta de anexar a `m1` uma quarta coluna com a informação do peso definida pela função da alínea anterior.
- (h) O que acontece ao conteúdo da matriz `m2`?
- (i) A partir da matriz `m1`, crie um data frame `df` com as 4 colunas seguintes: altura, peso, imc e info (com o resultado da função `calcIMC`).
- (j) Execute a função `summary()` sobre a matriz `m2` e o data frame `df`. Qual é a diferença?
- (k) Execute o comando `df$info` e verifique que se trata de um *factor* com apenas uma categoria. Altere essa coluna do data frame de modo a que seja um factor com as 3 categorias possíveis. Atenção: o conteúdo da coluna não deve ser alterado, apenas o conjunto das categorias possíveis.
- (l) Execute novamente a função `summary()` sobre o data frame `df`. Qual é a diferença?
- (m) Qual o efeito dos comandos `df[-1,]` e `df[, -ncol(df)]`?
- (n) Adicione ao final do data frame duas linhas: uma com a altura = 1.74 e o peso = 50 e outra com altura = 1.50 e o peso = 74. Para tal, aceda ou índice da linha que está adicionar com `nrow(df) + 1`. Em alternativa, pode também usar a função `rbind()`.
- (o) Usando uma estratégia semelhante, adicione uma nova coluna ao data frame com a informação do sexo dos indivíduos a que estes dados dizem respeito: "f". Em alternativa, pode também usar a função `cbind()`.