

Atividade nº 3 – Estruturas de Dados em R

Instruções para entrega da lista:

- O relatório de respostas da lista (desenvolvimento, comandos, resultados, saídas gráficas e **comentários**) deve estar apresentado em documento com extensão `.html` ou `.pdf`, **gerada em R Markdown**, com todos os comandos utilizados na solução da presente lista.
- Os arquivos com o relatório de respostas deverá ser denominado `064-241_At03-SEUNOME-SEUSOBRENOME.pdf` (ou `.html`).
- Não esqueça de **se identificar no preâmbulo do arquivo**, além de rotular corretamente as questões cujos comandos e resultados você estará apresentando.
- Apresente todos os comandos (todos os comandos que funcionaram!) que utilizou para obter os resultados solicitados.
- Preserve a ordem** das questões e apresente brevemente suas justificativas e comentários.
- O upload do relatório (extensão `.pdf` ou `.html`) deverá ser efetuado exclusivamente no Moodle, até a data marcada.
- Não hesite em procurar o **Fórum de Dúvidas** do Moodle ou a monitoria, caso tenha alguma dúvida com relação à solução da presente lista de exercícios. Caso não resolva, acione o professor. Acostume-se a interagir para obter sugestões de solução das dúvidas.

- Matriz.* Use o R e resolva matricialmente o seguinte sistema de equações:

$$\begin{cases} x + 2y + 3z = 20 \\ 2x + 5y + 9z = 100 \\ 5x + 7y + 8z = 200 \end{cases}$$

Verifique previamente se o sistema tem solução única, é indeterminado ou é impossível. Como realizar esta verificação?

- Fatores.* Exercícios expressos:

- Se `z <- factor(c("p", "q", "p", "r", "q"))` e os níveis de `z` forem `"p", "q", "r"`, escreva um comando em R que troque o nível `"p"` para `"w"`, de maneira que `z` seja igual a: `"w", "q", "w", "r", "q"`.
- Execute os comandos abaixo e crie os fatores `s1` e `s2` por meio de um sorteio com reposição das letras do alfabeto. Antes verifique o que o comando `sample()` faz. Brinque um pouco com ele.

```
set.seed(666)
s1 <- factor(sample(letters, size=5, replace=TRUE))
s2 <- factor(sample(letters, size=5, replace=TRUE))
```

Use um código em R para concatenar `s1` e `s2` em um único fator com 10 elementos.

- Considere o conjunto de dados `iris` e use a função `cut()` e categorize a variável `Sepal.Length`, criando o seguinte fator com cinco níveis, referentes aos intervalos: `(4.3, 5.02]`, `(5.02, 5.74]`,

(5.74, 6.46], (6.46, 7.18] e (7.18, 7.9], com frequência em cada intervalo de 32, 41, 42, 24 e 11, respectivamente. Confira.

- d. Considere novamente o conjunto de dados iris. Gere uma tabela de contingência com duas linhas e três colunas. As linhas devem estar relacionadas com a variável Sepal.length (< 5: TRUE ou FALSE) e as colunas, com o fator Species. A saída deverá ser a seguinte:

	setosa	versicolor	virginica
FALSE	30	49	49
TRUE	20	1	1

- e. Considere o fator respostas <- factor(c("Concordo", "Concordo", "Concordo fortemente", "Discordo", "Concordo")), com a seguinte saída:

```
[1] Concordo          Concordo          Concordo
[4] Desconcordo      Concordo
Levels: Concordo Concordo fortemente Discordo
```

Mais tarde percebeu-se que existe um novo nível, “Discordo fortemente”. Use comando em R para incluir “Discordo fortemente” como um novo nível de tributo do fator, fornecendo a seguinte saída:

```
[1] Concordo          Concordo          Concordo fortemente
[4] Discordo          Concordo
Levels: Concordo fortemente Concordo Discordo
Discordo fortemente
```

- f. Se x <- factor(c("alto", "baixo", "médio", "alto", "alto", "baixo", "médio")), use comando em R para os valores numéricos dos vários níveis de x, apresentando a seguintes saída:

```
levels value
1  alto     1
2  baixo    2
3  médio    3
```

3. Listas. Exercícios expressos:

- Se: p <- c(2, 7, 8), q <- c("A", "B", "C") e x <- list(p, q), qual o valor de x[2]?
- Se: w <- c(2, 7, 8), v <- c("A", "B", "C") e x <- list(w, v), qual código em R substituirá "A" em x por "K"?
- Se: a <- list("x" = 5, "y" = 10, "z" = 15), qual código em R dará a soma de todos os elementos de a?
- Se: nova <- list(a = 1:10, b = "Férias", c = "Olá"), escreva código em R que adicionaria 1 a cada elemento do primeiro vetor em nova?

- e. Se: `b <- list(a = 1:10, c = "Olá", d = "AA")`, escreva código em R que dê todos os elementos de `b`, exceto o segundo elemento, do primeiro vetor de `b`?
- f. Seja `x <- list(a = 5:10, c = "Olá", d = "AA")`, escreva código em R para acrescentar um novo item `z` à lista `x`, com `z = "Novo item"`?
- g. Considere `y <- list(1:10, "string", 666)`, escreva código em R que atribua novos nomes, "um", "dois", "três", aos elementos de `y`.
- h. Se: `x <- list(y = 1:10, t = "Olá", f = "TT", r = 5:20)`, escreva código em R que dê o comprimento do vetor `r`, de `x`.

4. *Data frame*. Crie os *data frames* indicados abaixo:

```
df.A <- data.frame(var1 = 1:5, var2 = LETTERS[1:5])  
df.B <- data.frame(var1 = 6:10, var2 = LETTERS[6:10])  
var3 <- 11:20
```

- a. Crie o *data frame* `df.res` a partir da vinculação por linhas de `df.A` e `df.B`.
- b. Apresente a classe de cada coluna (por exemplo, número inteiro, caractere, fator etc.).
- c. Construa `df.A2` conforme indicado abaixo e o vincule por linhas com `df.B`. Comente o resultado obtido. Por que?

```
df.A2 <- data.frame(var1 = 1:6, var2 = LETTERS[1:6])
```

- d. Use o comando `cbind()` para adicionar o vetor `var3` como uma nova variável de `df.res`.
- e. Rearranje colunas desse *data frame*, de seguinte maneira que elas fiquem na seguinte ordem: `var1, var3, var2`.

5. *Data frame*. Considere o conjunto de dados `state.x77{datasets}`.

- a. Verifique se o objeto `state.x77` é um *data frame*, caso contrário, transforme-o para um *data frame*.
- b. Determine quantos estados têm uma renda inferior a 4300.
- c. Determine qual é o estado com maior renda.
- d. Construa o *data frame* `state` com os conjuntos de dados `state.abb`, `state.area`, `state.division`, `state.name`, `state.region`. Os nomes das linhas devem ser os nomes dos estados.
- e. Construa o *data frame* `estados`, adicionando, por coluna, o *data frame* `state.x77` ao objeto `states`.
- f. Renomeie os nomes das variáveis utilizando apenas as 3 primeiras letras após o ponto final dos nomes originais (por exemplo, `States.abb` será denominada `abb`).
- g. Remova a variável `div`.

- h. Construa `estados.a`, um subconjunto de `estados`, através da remoção das variáveis `Life Exp`, `HS Grad`, `Frost`, `abbe` e `are`.
 - i. Categorize a variável `Illiteracy`, em níveis de analfabetismo: baixo, para valores no intervalo $[0, 1)$; algum, no intervalo $[1, 2)$ e alto, no intervalo $[2, \infty)$. Adicione essa variável ao `data frame` `estados`.
 - j. Determine o estado com maior renda na região oeste e que tenha baixo analfabetismo. Qual é o valor dessa renda?
6. Seja o `array` `Titanic{datasets}`, com o qual já trabalhamos em sala. Sempre que julgar conveniente, use transformação e visualização de dados para responder as questões abaixo:
- a. Segundo a *Wikipedia*, havia cerca de 2.224 passageiros e tripulantes a bordo do Titanic quando ele afundou. Para quantos deles temos informações neste conjunto de dados? Das pessoas para as quais temos dados, quantas sobreviveram e quantas não? Qual é a taxa de sobrevivência global?
 - b. Quantos passageiros do Titanic eram homens e quantos eram mulheres? O que você encontra quando tabula por classe de bilhete?
 - c. Quantos passageiros de cada sexo sobreviveram e quantos deles não sobreviveram? Qual é a taxa de sobrevivência para passageiros por sexo?
 - d. Mostre a distribuição de idade por classe de bilhete, por sexo. Comente o que você encontrou?
 - e. Como o sexo, a classe do bilhete e a idade de um passageiro afetaram sua chance de sobrevivência? Tente usar um gráfico para responder a esta pergunta.