

Lista 02 de exercícios

Fundamentos Estatística para Ciência dos Dados

Luís Gustavo Silva

Disciplina ministrada por Renato M. Assunção, DCC, UFMG

Março de 2017

Exercício 1 Escreva uma função que calcule a média geométrica de um vetor $\mathbf{x} = (x_1, \dots, x_n)$ e retorne uma mensagem de aviso se algum elemento de \mathbf{x} é negativo. A média geométrica de \mathbf{x} é:

$$f(\mathbf{x}) = \left(\prod_i^n x_i \right)^{\frac{1}{n}}$$

Dica: `?warning`

Exercício 2 • Aplique a função criada no item anterior em cada uma das colunas do dataframe `data`.

- Calcule o desvio-padrão para cada uma das colunas.
- Calcule o total para cada uma das linhas.
- Selecione as linhas em que variável `X.1 > 3` e `X.20 < 3`. Quantas linhas foram selecionadas?
- Substitua o nome das variáveis para `Var1`, `Var2`, ..., `Var25`.

```
> set.seed(123)
```

```
> data <- data.frame(matrix(rnorm(10000, mean=3), ncol=25, dimnames=list(NULL, paste("X", 1:25, sep="."))))
```

Exercício 3 Leia o conjunto de dados `iris.csv` e faça:

- Conheça os atributos do dataframe com `?iris`.
- Faça o gráfico `boxplot` para cada variável: `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, `Species` numa mesma janela gráfica.
- Faça o gráfico `boxplot` para cada variável: `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, `Species` numa mesma janela gráfica e pelo grupo `Species`.
- Faça o gráfico `histograma` para cada variável: `Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`, `Species` numa mesma janela gráfica.

Dica: `par(mfrow=c(...))`

Exercício 4 Leia o conjunto de dados `attitude.csv` e faça:

- Conheça os atributos do dataframe com `?attitude`.
 - Experimente fazer `plot(attitude)`.
 - Calcule a média para cada uma das colunas.
 - Transforme a variável `complaints` em uma variável discreta com as seguintes categorias: `[0 – 60] = "bad"`, `[61 – 80] = "okay"` e `[81 – 100] = "good"`. Dica: `?cut`.
 - Faça o gráfico `boxplot` da variável `rating` pela variável `complaints` categorizada no item anterior.
 - Faça um gráfico de barras da variável `complaints` categorizada.
 - Modifique as categorias para `"Ruim"`, `"Normal"`, `"Bom"` e refaça os gráficos.
-