

ME115 - Linguagem R

Turma B - Prof. Rafael Pimentel Maia

Prova 1 - 20/04/2023 - 1S2023

NOME: _____ RA: _____

Carregando pacotes exigidos: knitr

Instruções

- Edite o cabeçalho do arquivo Rmd colocando seu nome e RA.
- Lembre-se de alterar a opção `eval = FALSE` no chunk acima para que os chunks subsequentes sejam avaliados.
- Use caminhos relativos para que eu possa executar o seu código no meu computador sem erros e sem ter que editá-lo.
- Para todos os problemas abaixo, escreva o código para responder cada uma das questões.
- Tenha certeza de que o seu arquivo .Rmd compila sem erros, gerando um arquivo HTML ou pdf.

Questão 1 (3,0 pontos)

Mostrar os comandos que podem ser usados para criar os objetos e/ou executar as instruções a seguir e os resultados obtidos.

- (a) Construa uma matriz 3×50 e preencha cada coluna com uma amostra aleatória de uma distribuição normal com média 5, 6 e 6, respectivamente, e desvio padrão igual 1,3. Adicione os seguintes nomes às colunas da matriz **Nota1**, **Nota2** e **Nota3**, respectivamente. (0,75 ponto)

Solução:

```
m <- mapply(rnorm, n = 50, mean = c(5,6,6), sd = 1.3)
colnames(m) <- c("Nota1", "Nota2", "Nota3")
```

- (b) Supondo que cada coluna da matriz do item (a) represente a nota em uma dada avaliação e que cada linha represente um aluno que realizou a prova: transforme a matriz em um *dataframe* e adicione uma coluna com a nota média do aluno e outra coluna indicando “Aprovado” se a nota média foi maior ou igual a 5 e “Reprovado” caso contrário. (0,75 ponto)

Solução:

```
notas <- data.frame(m)
notas$media <- apply(notas, 1, mean)
notas$situacao <- ifelse(notas$media >= 5, "Aprovado", "Reprovado")
```

- (c) Apresente estatísticas descritivas de posição e de dispersão da nota média dos alunos. Qual o número e a porcentagem de alunos aprovados? (0,75 ponto)

Solução:

```
cv <- function(x) sd(x)/mean(x)
mysum <- function(x) c(media = mean(x), mediana = median(x), desvio_padrao = sd(x), minimo = min(x), ma
```

```
mysum(notas$media)

##          media          mediana desvio_padrao          minimo          maximo
##    5.5388756    5.4297829    0.7528683    4.3441151    7.4034272
##          c_var
##    0.1359244

p <- mean(notas$situacao=="Aprovado")*100
```

A porcentagem de alunos aprovados na turma é de 70%

- (c) Calcule a média geométrica das notas dos alunos dada por $\sqrt[3]{X_1 \times X_2 \times X_3}$ em que X_1 , X_2 e X_3 representam as notas na avaliação 1, 2 e 3, respectivamente. Apresente estatísticas descritivas de posição e de dispersão da média geométrica dos alunos. (0,75 ponto)

Solução:

```
notas$geom <- apply(notas[,1:3], 1, function(x) (prod(x))^(1/3))
mysum(notas$geom)

##          media          mediana desvio_padrao          minimo          maximo
##    5.4220898    5.3193829    0.7792688    4.2033316    7.3038755
##          c_var
##    0.1437211
```

Questão 2

(2,5 pontos)

Escrever uma função que converta o valor da temperatura de graus Celsius para Fahrenheit e vice-versa, de Celsius para Kelvin e vice-versa, e de Fahrenheit para Kelvin e vice-versa, de acordo com o solicitado pelo usuário. A função deve ter três argumentos: x que é o valor da temperatura na escala original, *entrada* que é unidade de medida da temperatura na entrada (C para Celsius, F para Fahrenheit e K para Kelvin) e *saida* que é unidade de medida da temperatura na conversão (C para Celsius, F para Fahrenheit e K para Kelvin). A função deve retornar o valor da temperatura na nova unidade de medida e a unidade de medida. Teste sua função convertendo de $30C$ para Kelvin, e de $34K$ para Fahrenheit.

Observação: $F = C \times 1,8 + 32$, $K = C + 273$, $F = (K - 273) \times 1,8 + 32$

Solução

```
converte <- function(x, entrada, saida){
  r <- ifelse(entrada == "C" & saida == "F", x*1.8 + 32,
    ifelse(entrada == "C" & saida == "K", x + 273,
      ifelse(entrada == "F" & saida == "C", (x - 32)/1.8,
        ifelse(entrada == "F" & saida == "K", (x - 32)/1.8 + 273,
          ifelse(entrada == "K" & saida == "C", x - 273,
            (x - 273)*1.8 + 32 # entrada K e saida F
          ))))
  return(r)
}

converte(30, entrada = "C", saida = "K")

## [1] 303

converte(34, entrada = "K", saida = "F")

## [1] -398.2
```

Questão 3

(3,0 pontos)

Nessa questão, exploraremos o conjunto de dados `data(swiss)` do pacote `datasets` (da base do *R*). O conjunto de dados contém medidas de fertilidade padronizada e indicadores socioeconômicos para cada uma das 47 províncias de língua francesa da Suíça por volta de 1888. As variáveis presentes no banco de dados são:

- **Fertility**: medida de fertilidade padronizada comum,
- **Agriculture**: % de homens que trabalhavam na agricultura,
- **Examination**: % de recrutas recebendo a nota mais alta no exame do exército,
- **Education**: % de educação além da escola primária para recrutas,
- **Catholic**: % 'católico' (em oposição a 'protestante'),
- **Infant.Mortality**: % nascidos vivos que vivem menos de 1 ano.

- (a) Use funções apropriadas para descrever esse conjunto de dados. Cite o número de variáveis e o número de observações. Mostre um tabela com os nomes das variáveis e suas respectivas classes. (0,75 ponto)

Solução

```
data(swiss)
str(swiss)

## 'data.frame': 47 obs. of 6 variables:
## $ Fertility      : num  80.2 83.1 92.5 85.8 76.9 76.1 83.8 92.4 82.4 82.9 ...
## $ Agriculture    : num  17 45.1 39.7 36.5 43.5 35.3 70.2 67.8 53.3 45.2 ...
## $ Examination    : int   15 6 5 12 17 9 16 14 12 16 ...
## $ Education      : int   12 9 5 7 15 7 7 8 7 13 ...
## $ Catholic       : num   9.96 84.84 93.4 33.77 5.16 ...
## $ Infant.Mortality: num   22.2 22.2 20.2 20.3 20.6 26.6 23.6 24.9 21 24.4 ...

nvar <- ncol(swiss)
nobs <- nrow(swiss)

classes <- sapply(swiss, class)
kable(as.data.frame(classes))
```

	classes
Fertility	numeric
Agriculture	numeric
Examination	integer
Education	integer
Catholic	numeric
Infant.Mortality	numeric

O conjunto de dados possui 6 colunas e 47 observações.

- (b) Apresente uma tabela que contenha os valores de média, mediana, desvio padrão, mínimo, máximo e coeficiente de variação para cada uma das variáveis do banco de dados. (0,75 ponto)

Solução

```
kable(round(as.table(apply(swiss, 2, mysum)),3))
```

	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
media	70.143	50.660	16.489	10.979	41.144	19.943
mediana	70.400	54.100	16.000	8.000	15.140	20.000
desvio_padrao	12.492	22.711	7.978	9.615	41.705	2.913

	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
minimo	35.000	1.200	3.000	1.000	2.150	10.800
maxino	92.500	89.700	37.000	53.000	100.000	26.600
c_var	0.178	0.448	0.484	0.876	1.014	0.146

- (c) Crie uma variável que separe as provincias em quatro grupos: até 50% de católicos e até 50% de trabalhadores agrícolas; mais de 50% de católicos e até 50% de trabalhadores agrícolas; até 50% de católicos e mais de 50% de trabalhadores agrícolas; e, mais de 50% de católicos e mais de 50% de trabalhadores agrícolas. Apresenta uma tabela com a proporção de provincias por grupo (dica, utilize a função *prop.table*) (0,75 ponto)

Solução

```
swiss$grupos <- with(swiss,
  ifelse(Catholic > 50 & Agriculture > 50, "> 50% catolico e > 50% agricola",
  ifelse(Catholic > 50 & Agriculture <= 50, "> 50% catolico e <= 50% agricola",
  ifelse(Catholic <= 50 & Agriculture > 50, "<= 50% catolico e > 50% agricola",
  "<= 50% catolico e <= 50% agricola"))))

kable(prop.table(table(swiss$grupo)), digits = 3)
```

Var1	Freq
> 50% catolico e > 50% agricola	0.255
<= 50% catolico e <= 50% agricola	0.319
<= 50% catolico e > 50% agricola	0.298
> 50% catolico e <= 50% agricola	0.128

- (d) Calcule, em uma única linha de comando, estatísticas descritivas como a média, mediana e desvio-padrão da taxa de fertilidade padronizada segundo o grupo. Apresente os resultados. (Dica: utilizar uma função da família *apply*) (0,75 ponto)

Solução

```
tapply(swiss$Fertility, swiss$grupo, mysum)

## $` > 50% catolico e > 50% agricola`
##      media      mediana desvio_padrao      minimo      maxino
## 79.5166667 79.3500000  8.6787969 65.0000000 92.4000000
##      c_var
## 0.1091444
##
## $`<= 50% catolico e <= 50% agricola`
##      media      mediana desvio_padrao      minimo      maxino
## 66.7800000 67.6000000 12.5456424 35.0000000 85.8000000
##      c_var
## 0.1878653
##
## $`<= 50% catolico e > 50% agricola`
##      media      mediana desvio_padrao      minimo      maxino
## 65.62142857 65.30000000  5.36888164 56.60000000 74.20000000
##      c_var
## 0.08181598
##
## $`> 50% catolico e <= 50% agricola`
```

```
##          media          mediana desvio_padrao          minimo          maximo
##  70.3500000  79.5000000   21.2634663   42.8000000   92.5000000
##          c_var
##    0.3022525
```

Questão 4

(1,5 pontos)

Escreva uma função chamada *soma_pares* que recebe um vetor de números inteiros como argumento e retorna a soma e a quantidade apenas dos números pares no vetor. Aplique a função no vetor (1, 2, 3, 5, 6, 7, 8, NA, 10, 11, 12, 14, 15)

Solução

```
soma_pares <- function(x){
  x <- x[!is.na(x)]
  resto <- x%%2 # resto da divisão por dois
  pares <- x[resto==0]
  soma <- 0
  for(i in pares) soma <- soma + i
  return(soma)
}

teste <- c(1, 2, 3, 5, 6, 7, 8, NA, 10, 11, 12, 14, 15)
soma_pares(teste)

## [1] 52
```