

Introdução ao programa R

Exercícios: Lista 1

Ronald Targino, DEMA-UFC

Parte 1

Usando os dados da *Tabela 2.1 Estatística Básica*:

1. Remova todos os objetos da seção de trabalho do R.
2. Crie vetores para armazenar os dados de cada uma das colunas.

```
# Questão 2: complete esse script
V1 = 1:36
s = "solteiro"
k = "casado"
V2 = c(s,k,k,s,s,k,s,s,k,s,k,s,s,k,k,s,k,k,s,s,k,k,k,s,k,k,s,k,k)
f = 'ensino fundamental'
m = 'ensino médio'
V3 = c(f,f,f,m,f, ...)
V4 = c(NA,1,2,NA,NA,0,NA, ...)
V5 = c(400, 456, 525, ... ) / 100
V6 = c(26, 32, 36, ...)
V7 = c(3, 10, 5, 10, ...)
i = "interior"
p = "capital"
o = "outra"
V8 = c(i, p, p, o, o, ...)
```

3. Crie o *data frame* TEB formado por todos os vetores da questão 2.

```
# Questão 3: complete esse script
TEB = data.frame(V1, V2, ...)
```

4. Verifique a estrutura do *data frame* TEB.
5. Qual a dimensão (número de linhas e de colunas) do *data frame* TEB.
6. Renomeie as variáveis (cabeçalho) para: “Identificação”, “Estado Civil”, “Grau de Instrução”, “No. de Filhos”, “Salário”, “IdadeA”, “IdadeM”, “Região de Procedência”.
7. Obtenha a tabela de frequência para a variável No. de Filhos.
8. Calcule média, mediana, desvio padrão e variância do salário.
9. Crie o vetor salario1, dado pelo salário acrescido de 20%.
10. Crie o vetor salario2, dado pelo salário adicionado de 1,5.
11. Crie o vetor salario3, dado pelo salário subtraído de 0,8.
12. Crie o vetor salario4, dado pelo salário multiplicado por 2.
13. Crie o vetor salario5, dado pelo salário dividido por 3.
14. Calcule média, mediana, desvio padrão e variância para salario1, salario2, salario3, salario4 e salario5.
15. Os resultados da questão anterior estão de acordo com as propriedades das estatísticas?
16. Salve o *data frame* dados com o nome “DadosTEB.csv”.

Parte 2

17. Remova todos os objetos da seção de trabalho do R.
18. Leia o arquivo Tabela1.csv e atribua-o a um objeto (*data frame*) denominado *tabela1*.
19. Renomeie o cabeçalho para: “Identificação”, “Estado Civil”, “Grau de Instrução”, “No. de Filhos”, “Salário”, “IdadeA”, “IdadeM”, “Região de Procedência”.
20. Crie a variável Idade (em meses) a partir das variáveis IdadeA e IdadeM.
21. Adicione a variável Idade ao *data frame* *tabela1*.
22. Elimine as colunas IdadeA e IdadeM do *data frame* *tabela1*.
23. Verifique a estrutura do *data frame* *tabela1*.
24. Obtenha a dimensão do *data frame* *tabela1*.
25. Obtenha a lista de objetos ativos na seção de trabalho do R.
26. Salve o *data frame* *tabela1* com o nome “Tabela1Nova.csv”.
27. Elimine todos os objetos da seção de trabalho do R.

Parte 3

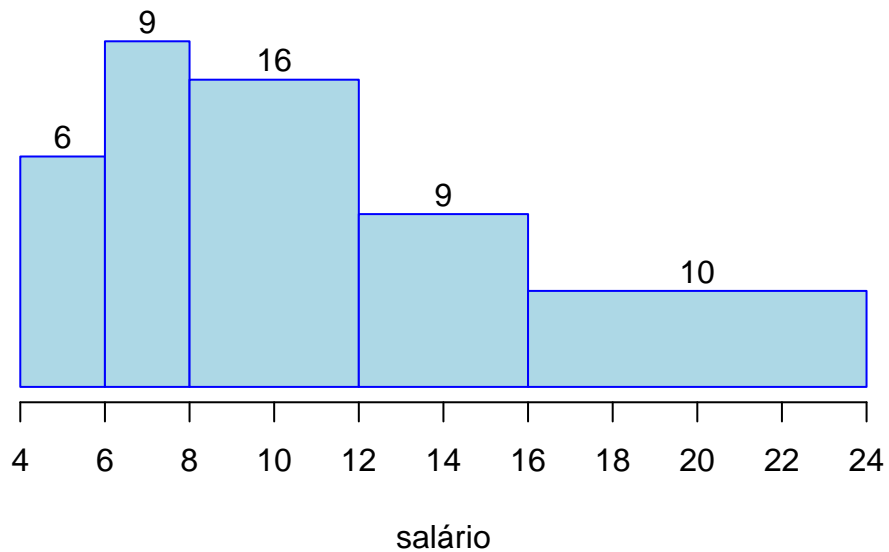
28. Remova todos os objetos da seção de trabalho do R.
29. Leia o arquivo Tabela2.csv e atribua-o ao objeto (*data frame*) *tabela2*. Não altere os nomes das variáveis, a menos que seja solicitado em alguma questão.

variável	definição
nr	número de registro do empregado (identificação)
ec	estado civil
gi	grau de instrução
nf	número de filhos
sa	salário (em salários mínimos)
rp	região de procedência
id	idade em anos

30. Apresente
 - a) os registros dos empregados de identificação 12 e 32 (ver função *subset*).
 - b) apenas os registros das variáveis nr, nf e sa dos empregados de identificação 10 a 15.
 - c) todos os registros dos funcionários com salário superior a 16 sm.
 - d) média, mediana, mínimo, máximo, variância, desvio padrão, coeficiente de variação, boxplot e histograma para a variável salário.
 - e) a tabela de frequência da variável número de filhos.
 - f) ordene a base de dados *tabela2* pelas variáveis ec, gi e rp, sequencialmente e em ordem crescente.
 - g) ordene a base de dados *tabela2* pelas variáveis ec, gi e rp, sequencialmente nas ordens crescente, decrescente e crescente.
31. Crie um novo *data frame* contendo apenas os registros dos empregados com ensino médio. Nomeie o *data frame* de *res1*.
32. Crie o *data frame* *res2* contendo apenas os registros de empregados casados com ensino superior.
33. Crie o *data frame* *res3* contendo apenas as variáveis grau de instrução, salário e região de procedência para os empregados com salário acima de 16 sm.
34. Qual o salário médio dos empregados que são casados e tem ensino médio?
35. Quais os salários médios para todas as combinações possíveis das variáveis ec, gi e rp?
36. Quais empregados são solteiros?
37. Apresente apenas os registros dos empregados que ganham menos de 5 sm?
38. Quantos empregados ganham de 8 a 10 sm?
39. Renomear a variável id para idm (idade em meses).
40. Crie o vetor ida (idade em anos) a partir da variável idm. Adicione-o a base de dados *tabela2*.

41. Obtenha a tabela de contingência para as variáveis região de procedência e grau de instrução (ver função *xtabs*).
42. Obtenha a tabela de frequência para a variável estado civil.
43. Obtenha a média, o desvio padrão e o coeficiente de variação para a variável número de filhos.
44. Ordene o *data frame* *tabela2* em ordem decrescente pela variável salário.
45. Construa os seguintes histogramas (use a função *text* e os argumentos *breaks*, *yaxt*, *xaxp*, *right* na função *hist*; veja também a função *par*):

Número de empregados segundo as classe salariais



Número de empregados segundo as classe salariais

