

Introdução ao programa R

Ronald Targino

Lista de Exercícios: 2

1. Dados simulados para uma empresa. Ao iniciar, use a função `rm(list=ls())` para eliminar todos os objetos da seção de trabalho do R. Todos os vetores devem conter **N=1000** registros. Use amostragem com reposição (`replace = TRUE`) apenas se o número de possíveis “valores” distintos para o vetor for menor do que **N**. Use a função `set.seed` (com valores diferentes para seu argumento) antes de gerar cada vetor de dados para tornar seus resultados reprodutíveis.

- a) Construa os vetores valor, produto, quantidade, margem de lucro (**porcentagem**), mês e ano da seguinte forma:

vetor(variável)	dados	use a função	probabilidades
valor	500, 600, 700, ..., 4000	sample	iguais
produto	tv, notebook, impressora, smartphone	sample	{0,3;0,2;0,1;0,4}
quantidade	10, 11, 12, ..., 4000	sample	iguais
mes	1 a 12	sample	iguais
margem	1,0 a 35,0	runif	iguais
ano	2017, 2018, 2019	sample	{0,50; 0,30; 0,20}

```
# construção das variáveis mes, margem e ano
N = 1000
set.seed(150)
mes = sample(1:12, size = N, replace = TRUE)
set.seed(235)
margem = round(runif(N, min = 1, max = 35), 1)
set.seed(450)
ano = sample(c(2017, 2018, 2019), size = N, prob = c(0.50, 0.30, 0.20), replace = TRUE)
```

- b) Construa a base de dados (*data frame*) vendas com as seguintes colunas: ano, mes, produto, quantidade, valor, margem; Elimine, com exceção do objeto vendas, todos os demais objetos da seção de trabalho do R. Use a função `ls()` para verificar que está disponível/ativo apenas o objeto vendas.
- c) Construa a variável lucro dada por $\text{lucro} = \text{quantidade} \times \text{valor} \times \text{margem}$ (use a função `with` ou a função `attach` para anexar o objeto *vendas* e usar diretamente as variáveis.);
- d) Adicione, como última coluna, a variável lucro à base de dados vendas; elimine o objeto lucro, deixando-o apenas na base de dados vendas (veja a função `cbind`);
- e) Calcule as estatísticas mínimo, média, mediana, máximo e desvio padrão para a variável lucro em cada um dos 3 anos (construa uma função que retorne todas as estatísticas e use a função `tapply`);
- f) Organize em uma matriz os resultados do item anterior. Nomeie linhas e colunas;
- g) Compare graficamente (histograma e boxplot) a variável lucro nos 3 anos(use o argumento `mfrow` da

- função *par* para criar uma única figura com os 3 histogramas e os 3 boxplots);
- h) Construa uma tabela para a variável produto com as frequências absolutas e respectivos percentuais (veja as funções *table*, *prop.table* e *margin.table*);
 - i) Construa uma tabela de contingência para as variáveis ano e produto (veja a função *xtabs*);
 - j) Calcule o lucro médio em cada ano e em cada mês/ano (veja a função *tapply*);
 - k) Obtenha as estatísticas da função *summary* para o lucro em cada ano;
 - l) Construa um gráfico de linhas para expressar o lucro médio mensal em cada ano (veja as funções *plot* e *lines*; opção: usar resultado do item j);
 - m) Determine o valor total (valor \times quantidade) para cada um dos 4 produtos em cada um dos 3 anos;
 - n) Construa o *data frame* vendas33 que contenha apenas os registros do *data frame* vendas com margem superiores a 33,00 (veja a função *subset*);
 - o) Com base no *data frame* vendas, construa o *data frame* MaioresMargens que contenha apenas as variáveis ano, mês, produto e margem superior a 34,00;
 - p) Com base no *data frame* vendas, construa o *data frame* MenoresMargens que contenha apenas as variáveis ano, mês, produto e margem inferior a 2,00;
 - q) Combine em um só *data frame* denominado MargensExtremas os *data frames* MaioresMargens e MenoresMargens (veja *rbind*);
 - r) Crie o vetor id (identificação) com os inteiros de 1 a N em ordem crescente e adicione-o como primeira coluna no *data frame* vendas;
 - s) Determine a identificação e o nome dos produtos com margem superior a 32,00(veja a função *subset*);
 - t) Apresente, do *data frame* vendas, apenas as linhas com a identificação determinada no item anterior;
 - u) Quantas impressoras foram vendidas com margem superior a 20,00? (veja as funções *subset* e *dim*)
 - v) Salve o *data frame* vendas em um arquivo de nome “VendasAtualizado.csv”;
 - w) Crie o *data frame* vendasaleatorias20 contendo as colunas ano, mês, produto e lucro e apenas 20 linhas selecionadas aleatoriamente do *data frame* vendas;
 - x) Adicione duas novas linhas ao *data frame* vendas: registros {N+1, 2020, 1, tv, 50, 2000, 15} e {N+2, 2020, 1, notebook, 350, 2700, 20} (veja *rbind*);
 - y) Renomeie para “TV” todos os registros “tv” (veja a função *recode* do pacote *dplyr*);
 - z) Salve o *data frame* vendas em um arquivo de nome “VendasAtualizado2.csv”.