Lista 2 de exercícios para prática de R

# Introdução à Análise de Dados - FACE/UFMG

Robert Iquiapaza 2024-04-19

# Apresentação

São propostas diferentes atividades, que vão desde as mais simples até mais avançadas. As mesmas podem ser resolvidas ao longo do desenvolvimento da disciplina.

O propósito é a prática em si, dessa forma a entrega da resolução é voluntária, e poderá impactar a nota de *participação*.

# Estruturas do controle

1. Suponha que você está trabalhando com um banco de dados simulado contendo in- formações sobre transações financeiras. Cada linha representa uma transação com as seguintes colunas:

* ID: Número de identificação do cliente
* Valor: Valor monetário da transação.
* Tipo: Indica se a transação é uma entrada (E) ou saída (S), consideradas como positivo e negativo, respectivamente.

# simulação do banco de dados para 80 clientes

# sample sorteia aleatoriamente com substituição

# runif gera numeros aleatórios da distribuição uniforme, no intervalo df = data.frame(ID= sample(1:10,80,replace = T),

Valor= runif(80,1210,5500),

Tipo = sample(c("E","S"),80,replace = T))

Você deseja calcular o saldo acumulado para cada ID, considerando todas as transações até o momento. O saldo acumulado é a soma dos valores das transações anteriores para o mesmo ID, podendo ser negativo.

* 1. Escreva um loop for que percorra cada linha do banco de dados e atualize o Saldo\_Parcial acumulado no banco de dados, para cada ID. Ao terminar imprima os saldos finais de cada ID e responda quantos IDs, e qual a proporção, terminam com saldo negativo.
  2. Faça a mesma tarefa anterior usando a função sapply.
  3. Explique como cada abordagem funciona e discuta as vantagens e desvantagens de usar

for versus sapply para resolver esse problema.

1. Cenário: Suponha que você está trabalhando com um banco de dados que armazena informações sobre alunos e suas notas em diferentes disciplinas. O banco de dados possui as seguintes colunas:

* AlunoID: Identificador único do aluno.
* Nome: Nome completo do aluno.
* Disciplina: Nome da disciplina (por exemplo, Matemática, História, Ciências).
* Nota: Nota obtida pelo aluno na disciplina (em uma escala de 0 a 10). Exercício: Escreva um programa em R que realiza o seguinte:
* Carrega os dados do banco de dados simulado (você pode criar um data.frame fictício

com alguns registros).

* Calcula a média das notas para cada disciplina.
* Identifica o aluno com a maior nota em cada disciplina.
* Imprime os resultados na tela.

# Lidando com erros

1. - Considere que você está desenvolvendo um pacote no R para análise de dados financeiros. Uma das funções do seu pacote é calcular o Valor Presente Líquido (VPL) de uma série de fluxos de caixa futuros. No entanto, para garantir a robustez da função, você precisa implementar um sistema de condições que capture e trate os seguintes cenários:
   1. Se a série de fluxos de caixa contiver valores não numéricos, a função deve lançar um erro com uma mensagem explicativa.
   2. Se a taxa de desconto fornecida for negativa, a função deve lançar um aviso, pois uma taxa de desconto negativa não é convencional.
   3. Se a série de fluxos de caixa for um vetor vazio, a função deve lançar um erro indicando que não há dados para calcular o VPL.

Escreva o código em R para essa função, incluindo o sistema de condições de erros e alertas mencionados acima. Além disso, explique como o sistema de tratamento de erros e alertas contribui para a confiabilidade e usabilidade da função em um ambiente de produção.

1. - Analise o código e proponha alternativas para aprimorar a função de forma que a sua execução seja mais robusta e ofereça mensagens e alertas mais informativos quando for o caso.

mult\_matrizes <- function(A,B){ # Multiplicação de matrizes

C <- A %\*% B # como garantir a compatibilidade das dimensões # Exibindo a matriz resultante C

print(C)

}

#Exemplo de aplicação

# Definindo as matrizes A e B

A <- matrix(c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12), nrow = 3, byrow = TRUE)

B <- matrix(c(1, 2, 3, 4, 5, 6, 7, 8, 9), nrow = 3, byrow = TRUE)

mult\_matrizes(A, t(A)) mult\_matrizes(A, B) # quais dimensões?

# Manipulação de Dados:

1. Importe um conjunto de dados do tipo CSV com o comando read.csv(). Por ex- emplo, use banco\_empresas.csv da pasta Documentos. Especifique fileEncoding = "WINDOWS-1252", check.names = F, na.strings = "-".
2. Limpe e prepare os dados para análise. Comente sobre o formato dos dados, vantagens e desvantagens para realizar analises.
3. Utilize o pacote dplyr para realizar operações de manipulação de dados. Por exemplo, calcular estatísticas descritivas por setores
4. Transforme o banco de dados para formato long usando a função pivot\_longer(-(1:3),names\_sep

="\_",names\_to = c(".value", "Ano")). Para isso antes de executar esse código será necessário alterar os nomes das colunas de interesse para o formato “NomeVariavel\_Ano” (i.e: “Ativo.Tot\_2015”)

1. Considere o Retorno EBIT sobre Ativos (uma *proxy* de retorno sobre o investimento), e analise as estatísticas descritivas, para todo o banco de dados e por setores.
2. Encontre as 10 empresas com melhores e piores retornos.
3. Elabore um texto interpretando os resultados dos cálculos em 5 e 6.

# Visualização de Dados:

1. Crie gráficos de dispersão com o comando plot(). Use o banco de dados da questão anterior banco\_empresas.csv em formato longpara as questões desta seção.
2. Crie histogramas com o comando hist().
3. Crie boxplots com o comando boxplot().
4. Utilize o pacote ggplot2 para criar gráficos mais personalizados.
5. Responda:

* Como é a distribuição das variáves ?
* Qual o tipo de relação entre as variáveis ?

**Lembre-se:**

* A prática leva à perfeição. Não desista se você encontrar dificuldades. Comece com exercícios simples e avance gradualmente para exercícios mais complexos.
* Utilize a ajuda online (help()) e a documentação dos pacotes para aprender sobre as

funções disponíveis.

* Pratique com frequência para desenvolver suas habilidades em R.
* A comunidade R é muito ativa e está sempre disposta a ajudar. Participe de comunidades online de usuários do R para obter ajuda e trocar ideias.
* https://stackoverflow.com/questions/tagged/r
* https://pt.stackoverflow.com/questions/tagged/r