

Exercícios de R com R Markdown

Insira seu nome aqui

Insira a data

Descrição da Atividade

Nesta atividade você irá realizar algumas análises de dados na forma de um documento reproduzível com R Markdown. Você deve usar o RStudio para editar este arquivo .Rmd e processá-lo em um arquivo HTML.

Algumas recomendações:

- Se você não estiver habituado com R Markdown, acostume-se a processar com frequência o documento, usando o botão **Knit**. Isso permitirá que eventuais erros no documento ou no código R sejam identificados rapidamente, pouco depois de terem sido cometidos, o que facilitará sua correção. Na verdade, é uma boa ideia você fazer isso **agora**, para garantir que seu ambiente esteja configurado corretamente. Se você receber uma mensagem de erro do tipo `Error in library(foo)`, isso significa que o pacote `foo` não está instalado. Para instalar um pacote, execute o comando `install.packages("foo")` no Console, ou clique em *Tools -> Install Packages*.
- **Não esqueça** de colocar seu nome e a data no cabeçalho do arquivo .Rmd (onde diz “Insira seu nome aqui” e “Insira a data”).
- As variáveis usadas no código R devem ser usadas para tornar o relatório mais reproduzível, eliminando uma fonte de erros. Em vez de copiar valores manualmente, atribua os resultados a variáveis e exiba-os usando código R *inline*. Por exemplo, nas respostas você pode escrever algo como

```
Os tempos de execução ficam entre `r texec.min` e `r texec.max` ms.
```

- Após concluir a atividade, você deverá submeter no Moodle um arquivo ZIP contendo:
 - o arquivo fonte .Rmd;
 - a saída processada (PDF ou HTML) do arquivo .Rmd;
 - outros arquivos necessários ao processamento do arquivo .Rmd (se houver).

Análise de Tempos de Execução

O primeiro exercício consiste em analisar os dados de 92 tempos de execução de uma sub-rotina contidos no arquivo `texec-iops.dat`. Cada linha do arquivo representa uma execução, para a qual foram coletadas duas métricas:

- `texec`: tempo de execução da sub-rotina, em ms;
- `iops`: quantidade de operações de entrada/saída realizadas pela sub-rotina.

Analise os dados e responda às seguintes questões:

1. Qual o maior e o menor tempo de execução registrados?
2. Qual o tempo médio de execução?
3. Qual é a mediana do número de operações de E/S?
4. Em quantas execuções não foi realizada nenhuma operação de E/S?
5. Quantas execuções levaram 300 ms ou mais para executar? Qual o número médio de operações de E/S dessas execuções?

6. Gere um gráfico de dispersão do tempo de execução em função do número de operações de E/S (ou seja, o eixo x é o número de operações e o eixo y é o tempo de execução). De que maneira o número de operações de E/S influencia o tempo de execução?

Neste exercício, o código R está parcialmente escrito abaixo. Como ele tem algumas lacunas que precisam ser preenchidas, ele não pode ser executado como está. Você só precisa fazer duas coisas: (1) completar as lacunas no código e (2) mudar `eval` no cabeçalho do bloco (*chunk*) para `TRUE` para que ele seja executado quando você processar o documento com **Knit**.

Código e saídas do R:

```
# le os dados do arquivo para o data frame dados
dados <- read.table("texec-iops.dat", header = TRUE)

# 1
# texec.min é o menor elemento da coluna texec
texec.min <- ____ (dados$texec)
# texec.max é o maior elemento da coluna texec
texec.max <- ____ (____)

# 2
# texec.avg é a média de texec
texec.avg <- ____ (____)
# arredonda para uma casa decimal
texec.avg <- round(texec.avg, 1)

# 3
# iops.med é a mediana da coluna iops
iops.med <- ____ (____)

# 4
# iops.0 é o número de valores 0 em iops
iops.0 <- ____ (dados$iops == 0)

# 5
# dados.300 contém apenas as linhas com texec mínimo de 300
# iops.300.avg é a média de iops em dados.300
dados.300 <- _____
iops.300.avg <- ____ (dados.300$iops)
iops.300.avg <- round(iops.300.avg, 1) # arredonda para uma casa decimal

# 6
with(dados, plot(____, ____))
```

Respostas:

1. Resposta da questão 1
2. Resposta da questão 2
3. ...

Análise do conjunto `mtcars`

Diversos conjuntos de dados podem ser encontrados em qualquer instalação de R, e podem ser acessados diretamente pelo seu nome.¹ Um deles é `mtcars`, que traz 11 características de 32 carros publicadas na

¹O comando `data()` mostra uma lista dos conjuntos disponíveis.

revista americana *Motor Trend*. Você pode visualizar o conjunto com `View(mtcars)`, e consultar a descrição de cada variável usando o comando `?mtcars`. Neste exercício você irá analisar algumas dessas variáveis.

Usando o conjunto `mtcars`, responda às seguintes questões:

1. A variável `hp` representa a potência do motor. Qual carro tem o motor mais potente, e qual é a sua potência? E o motor menos potente?
2. A variável `am` indica o tipo de transmissão (0=automática, 1>manual). Qual carro com transmissão automática é o mais potente?
3. A variável `mpg` representa o consumo de combustível (em milhas por galão, 1 mpg = 0,425 km/l). É possível afirmar que o consumo médio de carros automáticos é maior que, menor que, ou igual ao consumo de carros com transmissão manual?
4. Quais carros possuem consumo igual ou melhor que 12 km/l?
5. A variável `qsec` representa o tempo (em segundos) para percorrer 1/4 de milha, e a variável `vs` indica a disposição dos cilindros do motor (0=cilindros em V, 1=cilindros em linha). Com base nos dados, é possível afirmar que motores com cilindros em V têm melhor desempenho (em termos de `qsec`) do que motores com cilindros em linha?
6. A variável `wt` representa o peso (em milhares de libras). Com base nos dados, como você diria que o peso afeta o consumo? (Dica: faça um gráfico de dispersão de consumo x peso.)
7. A variável `gear` representa o número de marchas. Com base nos dados, é possível afirmar que carros com mais marchas são mais (ou menos) potentes?

Código e saídas do R:

```
# seu código R vai aqui
```

Respostas:

1. Resposta da questão 1
2. Resposta da questão 2
3. ...

Análise do conjunto `WWWusage`

O conjunto `WWWusage` contém uma série de 100 medições do número de usuários conectados à Internet através de um servidor a cada minuto. Usando esse conjunto, responda às seguintes questões:

1. Em que minuto o servidor teve a maior carga, e qual foi essa carga?
2. Em que minuto o servidor teve a menor carga, e qual foi essa carga?
3. Em que minuto o servidor teve o maior aumento de carga? Qual foi esse aumento (de quantos para quantos usuários)?
4. Em que minuto o servidor teve a maior redução de carga? Qual foi essa redução (de quantos para quantos usuários)?
5. Quantas vezes o número de usuários manteve-se constante em duas medições consecutivas?
6. Determine o número de usuários para o minuto 101, supondo que um eventual aumento ou redução de carga não ultrapasse os valores máximos obtidos nas questões 3 e 4. (Dica: você deverá estimar uma faixa de valores.)

Código e saídas do R:

```
# seu código R vai aqui
```

Respostas:

1. Resposta da questão 1
2. Resposta da questão 2
3. ...