

UNIVERSIDADE DA REGIÃO DE JOINVILLE - UNIVILLE

Bacharelado em Engenharia de Software (BES)

Estatística para computação

Professora Priscila Ferraz Franczak

Engenheira Ambiental - UNIVILLE

Mestre em Ciência e Engenharia de Materiais - UDESC

Doutora em Ciência e Engenharia de Materiais - UDESC

priscila.franczak@gmail.com

Plano de Aula

1. Operações com vetores e matrizes

1.1 Alguns comandos úteis

1.2 Usando alguns tipos de combinação de operações

2. Entrando com dados

2.1 Uso do comando `scan()`

2.2 Uso do comando `edit()`

2.3 Lendo dados de um arquivo-texto

2.4 Lendo dados de uma planilha eletrônica

2.5 Carregando dados contidos em pacotes

3. Salvando e abrindo arquivos do R

4. Exercícios

1. Operações com vetores e matrizes

- Existem diferenças entre as operações feitas considerando os elementos de um vetores ou matriz e as operações considerando todo o objeto.

```
> v1<-c(2,1,0)  #criando vetor v1  
> v2<-c(1,3,1)  #criando vetor v2  
> v1*v2  #multiplicando elemento a elemento  
[1] 2 3 0
```

- O resultado da multiplicação é um vetor do mesmo tamanho.

- O mesmo é válido para matrizes:

```
> v3<-c(1,1,2)  #criando o vetor 3  
> A<-cbind(v1,v2,v3)  #criando a matriz A  
> A
```

	v1	v2	v3
[1,]	2	1	1
[2,]	1	3	1
[3,]	0	1	2

```
> B<-t(A)  #B recebe a transposta de A  
> B
```

	[,1]	[,2]	[,3]
v1	2	1	0
v2	1	3	1
v3	1	1	2

- Multiplicando elemento a elemento de $A*B$:

```
> A*B
```

	v1	v2	v3
[1,]	4	1	0
[2,]	1	9	1
[3,]	0	1	4

- Como multiplicar matrizes como em Álgebra?
- Basta trocar o sinal * por %*%

```
> A%*%B
```

	[,1]	[,2]	[,3]
[1,]	6	6	3
[2,]	6	11	5
[3,]	3	5	5

- Produto entre os vetores v_1 e v_2 :

```
> v1%*%v2  
      [,1]  
[1,]      5
```


- Podemos encontrar o determinante da matriz A:

```
> det(A)  
[1] 9
```

1.1. Alguns comandos úteis

Símbolo	Significado
<code>==</code>	Igual
<code>!=</code>	Diferente
<code><, ></code>	Menor que, maior que
<code><=, >=</code>	Menor ou igual a, maior ou igual a
<code>max()</code> , <code>min()</code> , <code>range()</code>	Máximo, mínimo, amplitude
<code>which.max(x)</code>	Índice do maior valor de x
<code>which.min(x)</code>	Índice do menor valor de x
<code>sum(x)</code>	Soma dos elementos de x
<code>prod(x)</code>	Produto dos elementos de x

Símbolo	Significado
<code>mean(x)</code> , <code>var(x)</code>	Média aritmética de x, variância amostral de x
<code>median(x)</code>	Mediana de x
<code>order(x)</code>	Vetor contendo as posições ordenadas crescentes de x
<code>rank(x)</code>	Vetor com o ranqueamento de x
<code>sort(x)</code>	Versão ordenada crescente de x
<code>rev(x)</code>	Vetor x com a ordem inversa
<code>cor(x,y)</code>	Correlação entre os vetores x e y
<code>cov(x,y)</code>	Covariância dos elementos de x e y
<code>colSums(A)</code>	Retorna a soma das colunas da matriz A
<code>rowSums(A)</code>	Retorna a soma das linhas da matriz A

1.2 Usando alguns tipos de combinação de operações

- Vamos criar um vetor com uma sequência de 0 a 10 e somar essa sequência:

```
>x<-1:10
```

```
>x
```

```
[1]  1  2  3  4  5  6  7  8  9 10
```

```
>sum(x)
```

```
[1] 55
```

- Qual a soma dos valores menores que 5?

```
>sum(x[x<5])  
[1] 10
```

- Qual a resposta lógica para $x < 5$?

```
>x<5  
[1] TRUE TRUE TRUE TRUE FALSE FALSE FALSE  
[8] FALSE FALSE FALSE
```

- Se os valores **falsos** fossem representados por 0 e os **verdadeiros** por 1, podemos multiplicar esses valores lógicos por números:

```
>x<5
```

```
[1] TRUE TRUE TRUE TRUE FALSE FALSE FALSE
```

```
[8] FALSE FALSE FALSE
```

```
>1*(x<5)
```

```
[1] 1 1 1 1 0 0 0 0 0 0
```


2. Entrando com dados

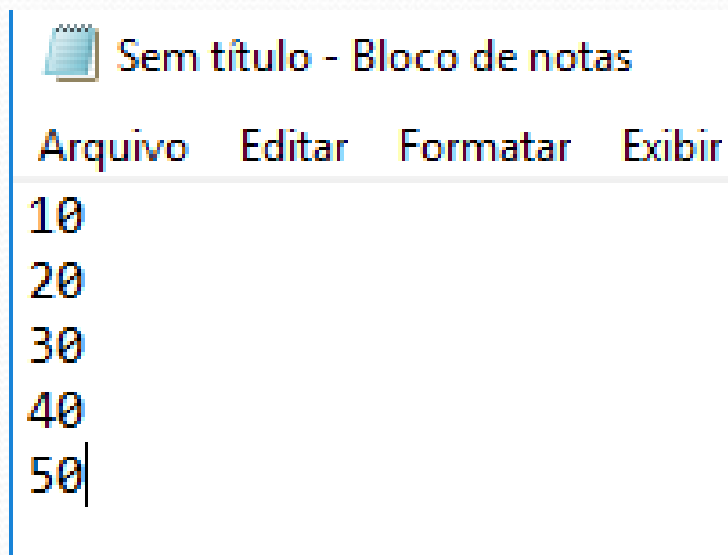
- Entrar de maneira correta com o conjunto de dados é etapa fundamental de qualquer análise.
- Esse conjunto pode ser digitado diretamente no console do R, ou lido de arquivos externos.

2.1 Uso do comando `scan()`

- Tem como finalidade a entrada de dados do console ou de um arquivo para armazená-lo em formato de vetor ou lista.
- Dessa maneira o **R** permite o uso do *prompt* para entrada de dados em um vetor sem a separação por vírgula.

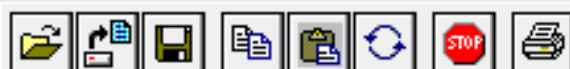
Exemplo

- Dados copiados de um editor de texto:



```
Sem título - Bloco de notas
Arquivo  Editar  Formatar  Exibir
10
20
30
40
50|
```

- Após copiar os dados do editor, basta digitar `scan()` dentro do console do R (e não o RStudio), atribuindo no objeto teste, apertar *Enter*, depois colar o texto selecionado e apertar *Enter* novamente:



R é um software livre e vem sem GARANTIA ALGUMA.
Você pode redistribuí-lo sob certas circunstâncias.
Digite 'license()' ou 'licence()' para detalhes de distribuição.

R é um projeto colaborativo com muitos contribuidores.
Digite 'contributors()' para obter mais informações e
'citation()' para saber como citar o R ou pacotes do R em publicações.

Digite 'demo()' para demonstrações, 'help()' para o sistema on-line de ajuda,
ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

```
> teste<-scan()
```

```
1: 10
```

```
2: 20
```

```
3: 30
```

```
4: 40
```

```
5: 50
```

```
6:
```

```
Read 5 items
```

```
> teste
```

```
[1] 10 20 30 40 50
```

```
> |
```

2.2 Uso do comando `edit()`

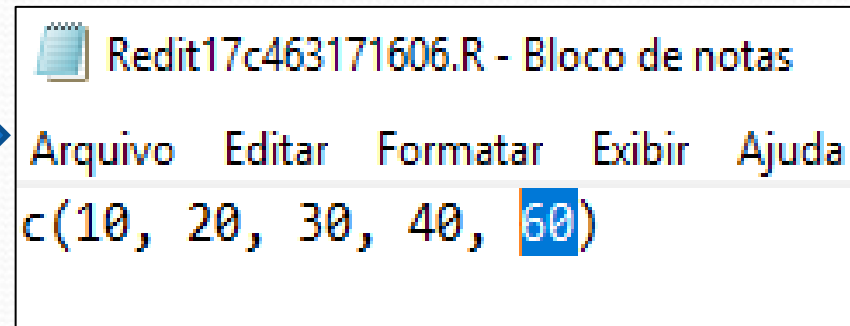
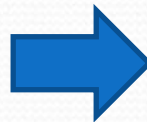
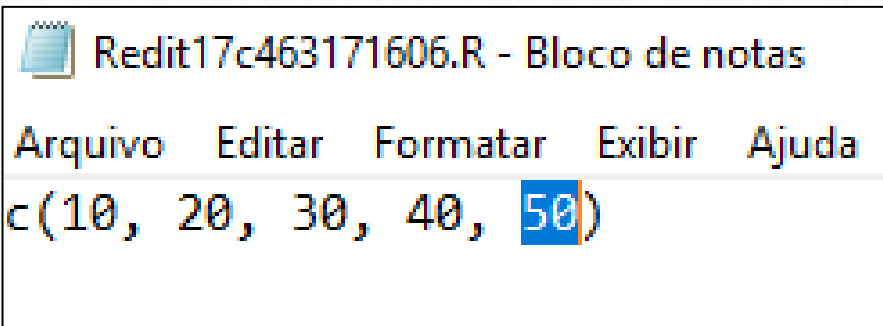
- Comando útil na edição de objetos já existentes, sejam eles vetores, data.frames, funções, etc.

- Exemplo: criar o vetor teste:

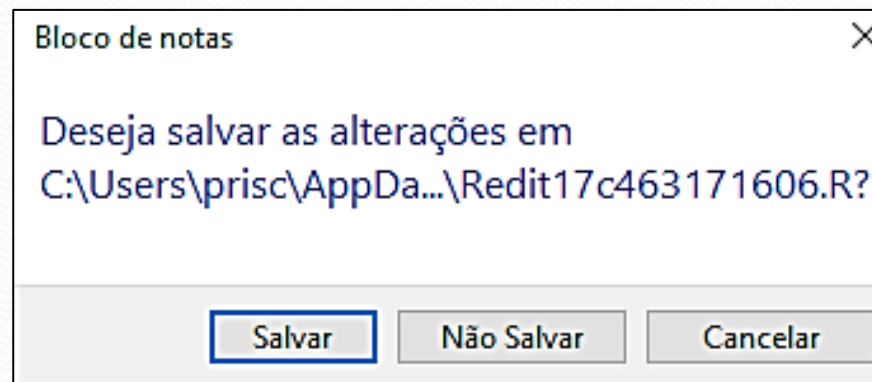
```
>teste<-c(10,20,30,40,50)
>teste
[1] 10 20 30 40 50
```

- Porém, houve um erro: o último elemento deveria ser 60.
- Não precisamos criar novamente o objeto teste se usarmos o comando edit():


```
> teste2<-edit(teste)
```



- Depois de alterar o vetor na janela do editor, basta fechar essa janela e salvar.




O novo objeto **teste2** será atribuído com a edição do **teste**:

```
>teste2
```

```
[1] 10 20 30 40 60
```

2.3 Lendo dados de um arquivo-texto

- Pode-se ter dados tabulados em um arquivo-texto ou em outros formatos que permitam a conversão para dados texto.
- Os dados geralmente estão separados por espaços ou vírgulas.

- 
- O comando `read.table()` é útil para ler dados de um arquivo-texto no formato de um `data.frame`.

- Considere os dados armazenados no arquivo chamado “**dados.txt**”.
- A primeira linha do arquivo corresponde ao cabeçalho, ou seja, nomes das colunas



dados.txt - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

Porosidade_ha, Porosidade_hasi

54.25, 47.7

51.26, 45.15

40.19, 31.95

65.05, 32.76

49.56, 31.2

- Usamos `read.table()` para ler esses dados e armazená-lo em um objeto.
- Fornecemos o endereço e nome do arquivo (entre aspas).
- Devemos informar também que a primeira linha do arquivo corresponde ao cabeçalho.

- Nesse comando é necessário informar o caminho completo do arquivo.
- O R não utiliza a barra invertida (\), mas a barra normal (/).

```
> dados<-read.table("C:/Users/prisc/Desktop/tese/dados.txt",
+                   header=TRUE, sep=",")
> dados
  Porosidade_ha Porosidade_hasi
1          54.25          47.70
2          51.26          45.15
3          40.19          31.95
4          65.05          32.76
5          49.56          31.20
```



dados.txt - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

Porosidade_ha, Porosidade_hasi

54.25, 47.7

51.26, 45.15

40.19, 31.95

65.05, 32.76

49.56, 31.2

Se as colunas forem separadas por espaços ou tabulações no arquivo-texto, o argumento `sep` não é necessário.

- O R cria o objeto dados como um data.frame:

```
> is.data.frame(dados)  
[1] TRUE
```

- O R entende como padrão o ponto como separador decimal.
- Caso os números sejam separados por vírgulas como decimal, é preciso deixar isso explícito no argumento `dec` do comando `read.table()`

`dec = “,”`

- Os arquivos no formato CSV (Coma Separated Values) são arquivos de texto separados por vírgulas e podem ser lidos pelo R no comando `read.csv()`
- O comando `read.table()` também funciona para lê-los, basta acrescentar o argumento `sep=","`

- O R pode ler outros formatos e acessar banco de dados SQL.
- Porém, os comandos necessários à realização de algumas dessas operações não se encontram nos pacotes incluídos na instalação do R, exigindo a instalação de pacotes adicionais.

2.4 Lendo dados de uma planilha eletrônica

- Microsoft Excel → formatos .xls e .xlsx
- Podemos salvar a planilha eletrônica no formato texto para ler no R usando o `read.table`, mas também é possível ler os dados diretamente da planilha do Excel.

- O comando `read_excel()` lê o conteúdo de uma planilha eletrônica para o R com a estrutura de dados de um `data.frame`.
- O pacote não consta na instalação base do R e deve ser instalado antes. Porém esse processo pode gerar alguns problemas durante a instalação.
- `install.packages("readxl")`

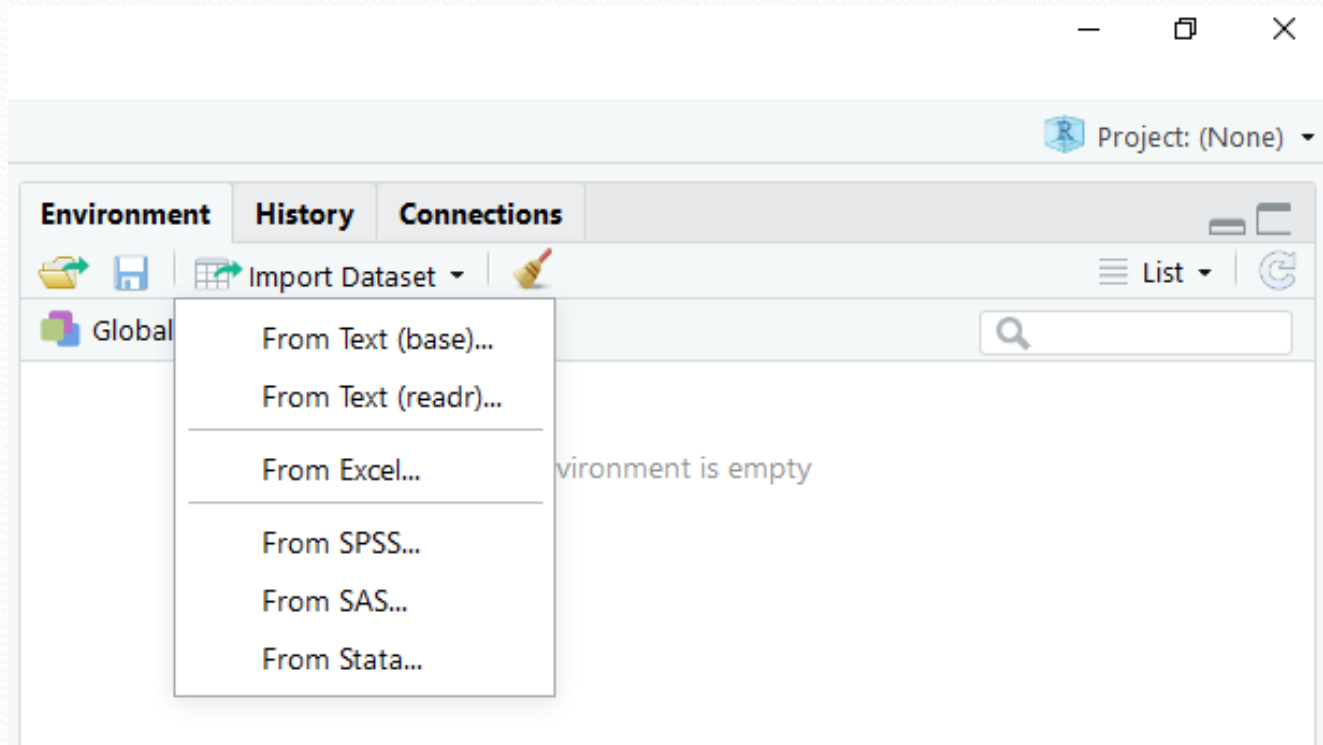

```

> library(readxl)
> Exercicios <- read_excel("C:/Users/prisc/OneDrive/Aulas/Univille/Aulas/Análise de dados com R/Aulas/Exercícios/aula 4/Exercicios.xlsx",
+                           col_types = c("text", "numeric", "text",
+                                         "numeric"))
> View(Exercicios)

```

	Nome	Idade	Sexo	NF
1	José Santos	17	M	92
2	Angela Dias	17	F	75
3	Aline Souza	16	F	81
4	Mayara Costa	15	F	87
5	Lara Lins	15	F	90
6	Nicolas Barros	13	M	88

- Há outro caminho diretor para importar tabelas do Excel:



Import Excel Data

File/Url:

C:/Users/prisc/Desktop/tese/dados.xlsx

Browse...

Data Preview:

porosidade ha <small>(character)</small>	porosidade hasi <small>(character)</small>
54.25	47.7
51.26	45.15
40.19	31.95
65.05	32.76

Previewing first 50 entries.

Import Options:

Name: dados

Max Rows:

☒ First Row as Names

Sheet: Default

Skip:

0

☒ Open Data Viewer

Range: A1:D10

NA:

Code Preview:

```
library(readxl)
dados <- read_excel("C:/Users/prisc
/Desktop/tese/dados.xlsx")
view(dados)
```

? Reading Excel files using readxl

Import

Cancel

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ - + R [Icons] Go to file/function Addins

Untitled1 x dados x

Filter

	porosidade ha	porosidade hasi
1	54.25	47.7
2	51.26	45.15
3	40.19	31.95
4	65.05	32.76
5	49.56	31.2

Showing 1 to 5 of 5 entries

Console ~/ ↗

Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

```
> library(readxl)
Warning message:
package 'readxl' was built under R version 3.4.3
> dados <- read_excel("C:/Users/prisc/Desktop/tese/dados.xlsx")
> View(dados)
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

+ - +R [Icons] Go to file/function [Icons] Addins

Untitled1* x dados x

Source on Save [Icons] Run [Icons] Sou

```
1 dados
2
```

2:1 (Top Level) ↕

Console ~/ ↗

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

```
> library(readxl)
Warning message:
package 'readxl' was built under R version 3.4.3
> dados <- read_excel("C:/Users/prisc/Desktop/tese/dados.xlsx")
> View(dados)
> dados
# A tibble: 5 x 2
  `porosidade ha` `porosidade hasi`
  <chr>          <chr>
1 54.25         47.7
2 51.26         45.15
3 40.19         31.95
4 65.05         32.76
5 49.56         31.2
```

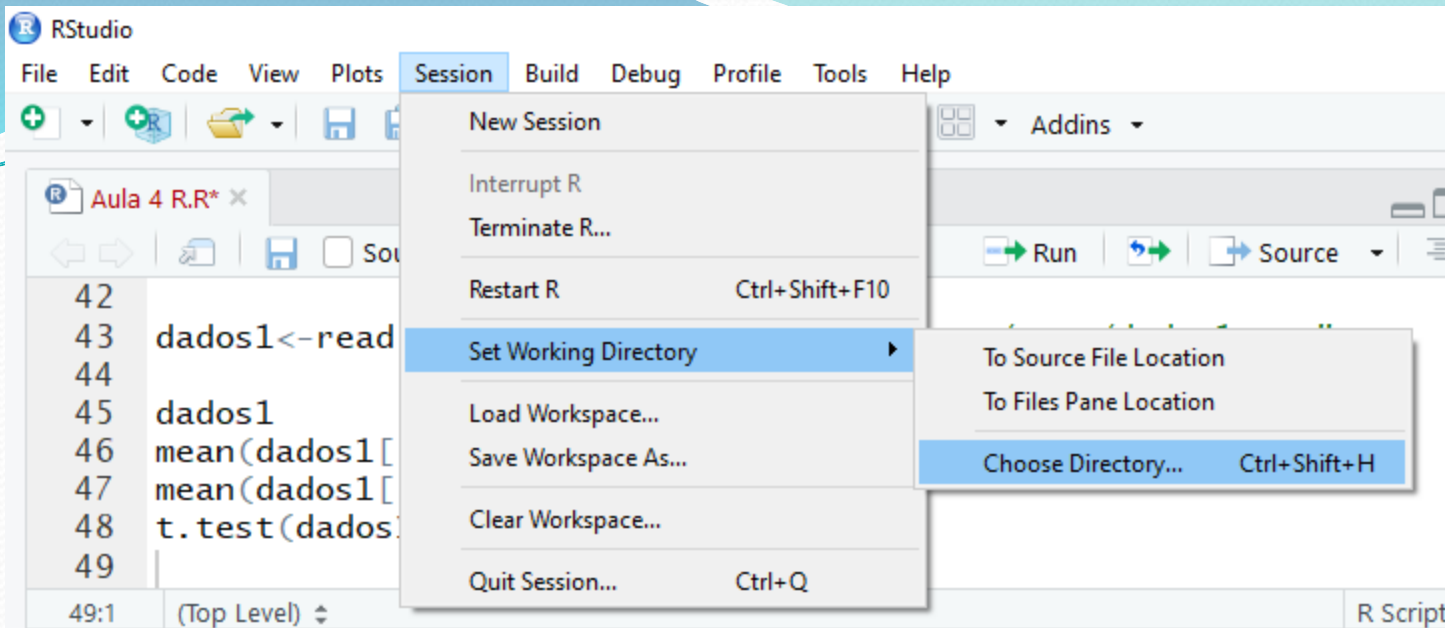
- Usando o comando `read.table()`
- Abrir a planilha `dados1.xls` e salvar como texto separado por tabulações.


```
> dados1<-read.table("C:/Users/prisc/Desktop/tese/dados1.txt",  
+                      header=TRUE)
```

```
> dados1
```

	ha	hasi
1	54.25	47.70
2	51.26	45.15
3	40.19	31.95
4	65.05	32.76
5	49.56	31.20

- Em alguns momentos o R pode estar lendo o arquivo no lugar errado e então dá erro na abertura da tabela.
- Verifica-se o diretório de onde ele está lendo através do comando `getwd()`
- Caso haja necessidade, é só mudar o diretório através do menu Session → Set Working Directory → Choose Directory



```
Console ~/  
[1] 52.062
> mean(dados1[,2])
[1] 37.752
> t.test(dados1)

One Sample t-test

data: dados1
t = 12.913, df = 9, p-value = 4.113e-07
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 37.03999 52.77401
sample estimates:
mean of x
 44.907
```

```
> getwd()
[1] "C:/Users/prisc/Documents"
> |
```



```
> dados1<-read.table("C:/Users/prisc/Desktop/tese/dados1.txt",  
+                      header=TRUE)
```

```
> dados1
```

	ha	hasi
1	54.25	47.70
2	51.26	45.15
3	40.19	31.95
4	65.05	32.76
5	49.56	31.20

- Podemos calcular a média de cada coluna:

```
> mean(dados1[,1])  
[1] 52.062  
> mean(dados1[,2])  
[1] 37.752
```

- Podemos obter a estatística básica das duas composições:

```
> summary(dados1)
```

	ha	hasi
Min.	:40.19	Min. :31.20
1st Qu.	:49.56	1st Qu.:31.95
Median	:51.26	Median :32.76
Mean	:52.06	Mean :37.75
3rd Qu.	:54.25	3rd Qu.:45.15
Max.	:65.05	Max. :47.70

- Podemos realizar o teste de comparação de médias entre as duas composições:

```
> t.test(dados1)
```

One Sample t-test

```
data: dados1
```

```
t = 12.913, df = 9, p-value = 4.113e-07
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
```

```
37.03999 52.77401
```

```
sample estimates:
```

```
mean of x
```

```
44.907
```


3. Salvando e abrindo arquivos do R

Há basicamente três tipos de arquivos no R:

- Projetos, que são arquivos *.Rdata, onde são salvos os objetos criados durante uma sessão do R.

3. Salvando e abrindo arquivos do R

- Arquivos-texto: *.Rhistory, contendo o histórico com todos os comandos digitados em uma sessão do R.
- Arquivos-texto: *.R, contendo funções implementadas ou rotinas de análises.

—

□

×

Exercícios ▾

EnvironmentHistoryConnections

📁📄📊 Import Dataset ▾🧽

☰ List ▾🔄

🌐 Global Environment ▾🔍

Environment is empty

FilesPlotsPackagesHelpViewer

📁 New Folder🗑 Delete➡ Rename⚙ More ▾🔄

📁 OneDrive > Aulas > Univille > Aulas > Análise de dados com R > Aulas > Exercícios 🌐 ...

	▲ Name	Size	Modified
	📁 ..		
<input type="checkbox"/>	📄 .Rhistory	825 B	Feb 20, 2018, 9:38 PM
<input type="checkbox"/>	📁 .Rproj.user		
<input type="checkbox"/>	📁 Arquivos em excel		
<input type="checkbox"/>	📁 aula 1		
<input type="checkbox"/>	📁 aula 2		
<input type="checkbox"/>	📁 aula 3		
<input type="checkbox"/>	📁 aula 4		
<input type="checkbox"/>	🌐 Exercícios.Rproj	218 B	Mar 18, 2018, 3:30 PM
<input type="checkbox"/>	🌐 Operações aritméticas.R	1.5 KB	Feb 5, 2018, 4:10 PM



4. Exercícios