

Ensino de Estatística

Lupércio França Bessegato
Ronaldo Rocha Bastos
Departamento de Estatística/UFJF

Roteiro Geral

1. Tratamento da informação
2. Produção de dados
- 3. Introdução ao R**
4. Probabilidade
5. Inferência estatística
6. Referências

Ensino de Estatística - 2017

2

Introdução ao R

Roteiro do Módulo

1. Introdução ao R:
 - a) Fundamentos da linguagem R
 - b) Introdução ao R Commander
 - c) Visualização gráfica no R Commander
 - d) Análise exploratória de dados
 - e) Exploração de conjunto de dados

Ensino de Estatística - 2017

4

Fundamentos da Linguagem R

Introdução

- Ferramenta de análise estatística de dados com muitos recursos e possibilidades;
- Constantemente atualizado e revisado incluindo novas técnicas e métodos;
- Programa de domínio público:
√ <http://cran.r-project.org>

Download do Programa

The screenshot shows the 'Download and Install R' section of the CRAN website. It lists operating systems: Linux, Mac OS X, and Windows. A box highlights the 'Download R 3.3.2 for Windows (62 megabytes, 32/64 bit)' link. Below this, there are links for 'Installation and other instructions' and 'New features in this version'. A 'Frequently asked questions' section is also visible at the bottom.

Comandos

- Para solicitar uma tarefa do R podemos digitar uma linha de comando no console

```
R Console
R version 2.11.0 (2010-04-22)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

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 publica

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

> sqrt(4)
[1] 2
> |
```

- Todas as funções do R devem ser digitadas em letras **minúsculas**
 √ O R é sensível a letras maiúsculas e minúsculas.
- Todas as palavras-chaves do R estão em letras minúsculas
- R usa um ponto “.” em vez de virgula “,” quando há números com casas decimais.

Alguns Comandos

- Soma → +
- Subtração → -
- Divisão → /
- Multiplicação → *
- Potenciação → ** ou ^
- Raiz quadrada de n → $\text{sqrt}(x)$
- Logaritmo natural → $\log(x)$
- Logaritmo decimal → $\log_{10}(x)$

- Consulte Table of useful R commands
 √ http://www.ufjf.br/lupercio_bessegato

Table of Useful R commands

Command	Purpose	Command	Purpose
help()	Obtain documentation for a given R command	plot()	Produces a scatterplot
example()	View some examples on the use of a command	qqplot()	Lattice command for producing a qqplot
rl, read()	Enter data manually as a vector in R	lm()	Determine the least square regression line
seq()	Make arithmetic progression vector	anova()	Analysis of variance (use on results of lm())
rep()	Make vector of repeated values	predict()	Obtain predicted values from linear model
data()	Load (data into a data.frame) built-in dataset	glm()	Generalized linear model
View()	View dataset in a spreadsheet-type format	read.csv()	Read data from a CSV file
str()	Display internal structure of an R object	read.table()	Read data from a table
read.csv(), read.table()	Load data into a data.frame from existing data file	library(), require()	Make available an R add-on package
dim()	See dimensions (if of matrix) of data.frame	length()	Give length of a vector
is.na()	Identify missing values	is.na()	Identify missing values
rm()	Remove an item from memory	is.null()	Identify null values
names()	List names of variables in a data.frame	is.vector()	Identify vector values
hist()	Command for producing a histogram	is.data.frame()	Identify data.frame values
histogram()	Lattice command for producing a histogram	is.factor()	Identify factor values
stem()	Make a stem plot	is.numeric()	Identify numeric values
table()	List all values of a variable with frequencies	is.integer()	Identify integer values
xtable()	Cross tabulation tables using formulas	is.double()	Identify double values
marginplot()	Make a margin plot	is.character()	Identify character values
cut()	Group values of a variable into larger bins	is.logical()	Identify logical values
mean(), median()	Identify "center" of distribution	is.list()	Identify list values
sd()	Apply function to a column within factors	is.environment()	Identify environment values
summary()	Display 5-number summary and more	is.function()	Identify function values
var(), sd()	Find variance, sd of values in a vector	is.package()	Identify package values
sort()	Sort all values in a vector	is.call()	Identify call values
quantile()	Find the position of a quantile in a dataset	is.symbol()	Identify symbol values
barplot()	Produce a bar graph	is.expression()	Identify expression values
barchart()	Lattice command for producing bar graphs	is.promise()	Identify promise values
boxplot()	Produce a boxplot	is.nullenv()	Identify null environment values
boxplot()	Lattice command for producing boxplots	is.newenv()	Identify new environment values

Help

- Help no R é bastante completo e autosuficiente.
- Se não conhecemos o comando de alguma técnica podemos chamar a função *help*:
 √ $\text{help}(\text{palavra-chave})$ ou $?\text{palavra-chave}$

Alguns Comandos para Ajuda

- Sempre procure ajuda no Google!

```
help.start()      # abre página de ajuda geral
help(hist)       # ajuda sobre a função hist
?hist           # idem
help.search("variance") # ajuda para localizar string
??variance      # idem
apropos("test")  # lista funções contendo a string test
example(hist)   # mostra um exemplo da função hist
```

```
# procura por hist em manuais de ajuda e mailing lists
RSiteSearch("hist")
```

```
# vignettes de pacotes instalados e em uso
vignette()      # mostra vignettes disponíveis
vignette("foo") # mostra vignette específica
```

```
# ajuda sobre conjunto de dados de pacotes instalados e em uso
help(datasetname) # para conjunto de dados datasetname
help(faithful)    # para conjunto de dados faithful
```

Ensino de Estatística - 2017

13

Help na Web

- R Project
√ <http://www.r-project.org>
- Tutorial de Introdução ao R
√ <http://ecologia.ib.usp.br/bie5782/doku.php?id=start>
- R Graphical Manual
√ http://www.imsbio.co.jp/RGM/R_image_list?page=1&sort=-

Ensino de Estatística - 2017

14

Armazenando dados

- Tipos de dados:
 - √ Numéricos
 - √ Caracteres: compostos por letras ou palavras.
 - √ Lógicos
- Quando os dados são armazenados, eles são chamados de objetos.
- Para armazenar um objeto basta utilizar o símbolo “<-” ou “=”.

Ensino de Estatística - 2017

15

Tipos de Objetos

- Variável
- Vetor
- Matriz
- Data Frame
- Array
- Lista
- Fatores

Ensino de Estatística - 2017

16

R Commander

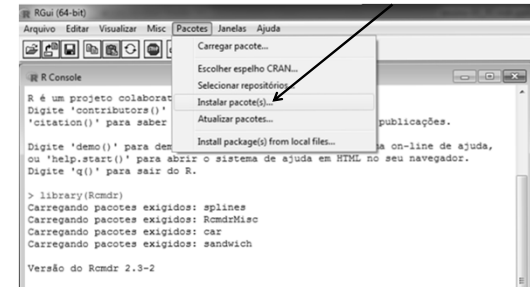
- Pacote *R Commander*
 - √ Interface gráfica desenvolvida por John Fox que facilita o uso do R.
 - √ Informações: <http://socserv.socsci.mcmaster.ca/jfox/Misc/Rcmdr/>

Ensino de Estatística - 2017

17

Instalação do R Commander

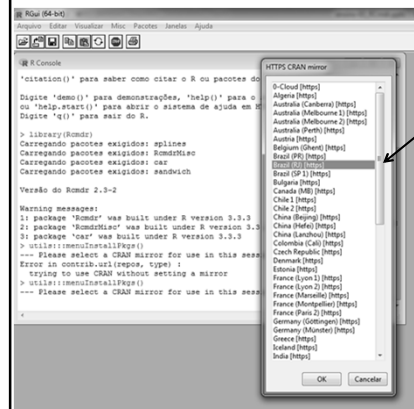
- 1º passo:
 - √ Na barra de menu, clique em “Pacotes” e, em seguida, “Instalar pacote(s)...”.



Ensino de Estatística - 2017

18

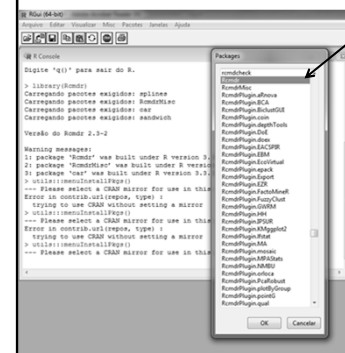
- 2º passo:
 - √ Escolha a localidade mais próxima de você



Ensino de Estatística - 2017

19

- 3º passo:
 - √ Selecione na lista o pacote “Rcmdr”
 - √ Aguarde enquanto o pacote estiver sendo instalado no R.



Ensino de Estatística - 2017

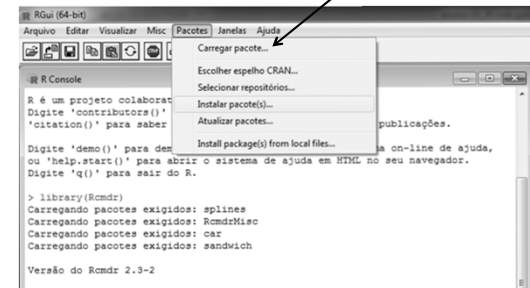
20

- 4º passo:
 - ✓ Aguarde o download do pacote.
 - ✓ Prossiga assim que o console do R indicar que o pacote foi instalado com sucesso

Ensino de Estatística - 2017

21

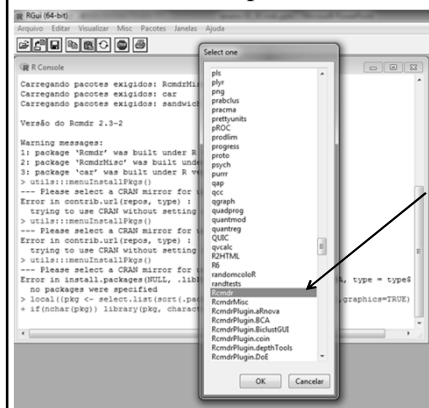
- 5º passo: Carregamento do pacote
 - ✓ Selecione novamente Pacotes na barra de menu e clique em “Carregar pacote...”



Ensino de Estatística - 2017

22

- 6º passo:
 - ✓ Selecione o pacote “Rcmdr” e aperte o botão OK



Ensino de Estatística - 2017

23

- 7º passo:
 - ✓ Janela avisará que o R Commander requer outros pacotes
 - ✓ Clique em “Sim”.



Ensino de Estatística - 2017

24

- 8º passo:
 - ✓ Mantendo a opção CRAN selecionada, clique em OK

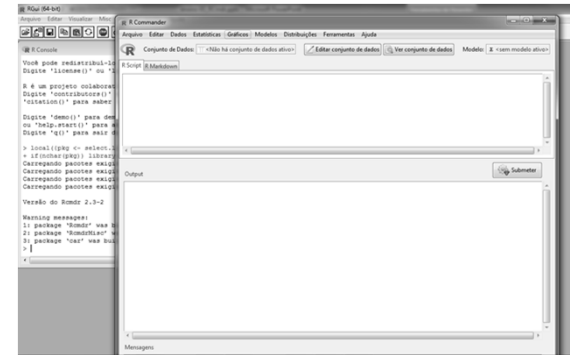


- ✓ Aguarde a instalação de todos os pacotes (demora de 5 a 10 minutos).

Ensino de Estatística - 2017

25

- 9º passo:
 - ✓ Após o término das instalações, aparecerá a janela do R Commander

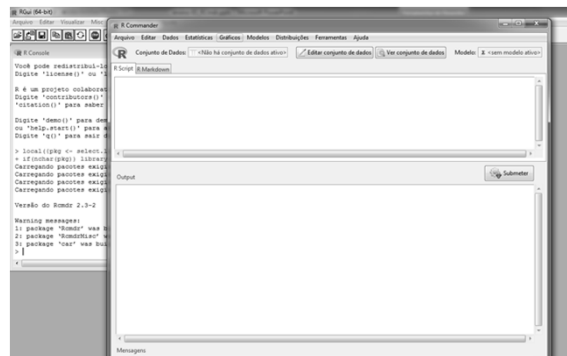


Ensino de Estatística - 2017

26

- Comandos de carregamento do pacote:

```
> library(Rcmdr)
> require(Rcmdr)
```



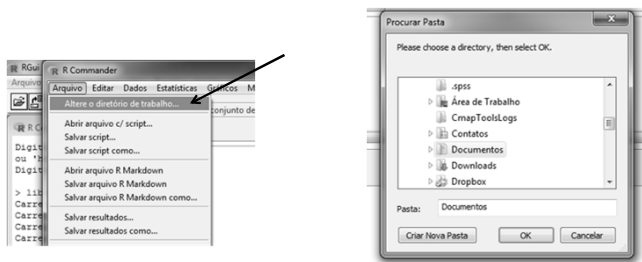
Ensino de Estatística - 2017

27

Introdução ao R Commander

Mudança do Diretório de Trabalho

- ✓ Clicar em “Arquivo” e selecionar “Altere o diretório de trabalho ...”
- ✓ Escolha seu diretório de trabalho



Ensino de Estatística - 2017

29

- Arquivos de formato livre mais usados:
 - ✓ Extensão .txt
 - ✓ Extensão .csv
- Cuidados com a importação dos dados:
 - ✓ Separador de campos:
 - Vírgula ou ponto-e-vírgula ou Tab ou outro
 - ✓ Separador de decimais
 - Ponto ou vírgula

Ensino de Estatística - 2017

30

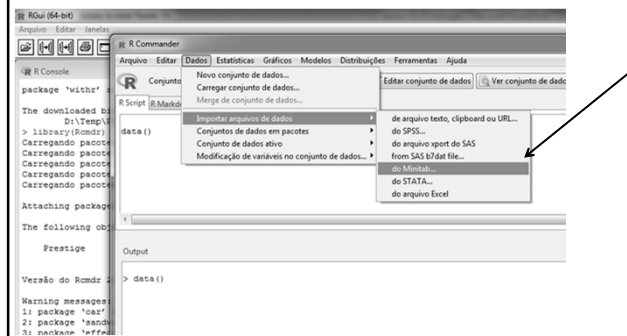
- Em geral, arquivos .txt no formato local:
 - ✓ Separador de campos:
 - Tab
 - ✓ Separador de decimal:
 - Vírgula
- Em geral, arquivos .txt no formato USA:
 - ✓ Separador de campos:
 - “;”
 - ✓ Separador de decimal:
 - Ponto

Ensino de Estatística - 2017

31

Importação de Conjunto de Dados

- 1º passo:
 - ✓ Clicar em Dados e selecionar “Importar arquivo de dados”



Ensino de Estatística - 2017

32

- 2º passo – Arquivo extensão .txt:
 ✓ Seleccionar “de arquivo texto, clipboard ou ...”

i) Definir o nome do conjunto de dados
 ii) Definir separador de campos
 iii) Definir separador de decimais

Ensino de Estatística - 2017 33

- 3º passo: Seleccionar o arquivo desejado:

Ensino de Estatística - 2017 34

Janelas

- Janela do Script:
 - ✓ Nela são salvas as sintaxes dos comandos
 - ✓ Serve para:
 - enviar as linhas de comandos
 - Executar o script em outra ocasião (CTRL + R)

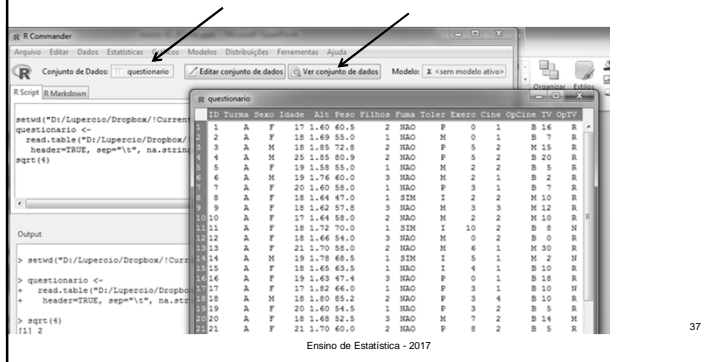
Ensino de Estatística - 2017 35

- Janela de Saída:
 - ✓ Resultados dos comandos dados

Ensino de Estatística - 2017 36

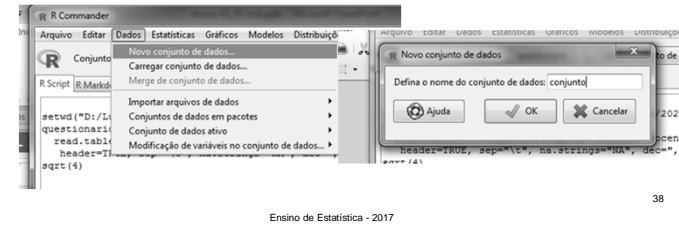
Verificando o Conjunto de Dados

- Selecionar o conjunto de dados
- Clicar em “Ver conjunto de dados”



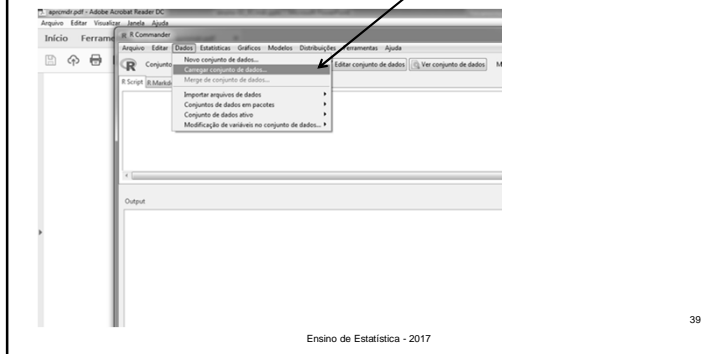
Criando Conjunto de Dados

- 1º passo:
 - ✓ Clicar em “Dados”
 - ✓ Selecionar “Novo conjunto de dados”
 - ✓ Nomear o novo conjunto de dados

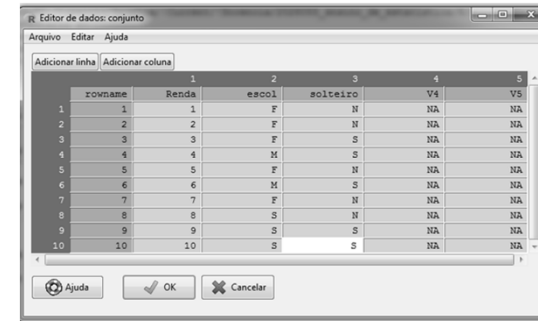


Carregamento de Dados

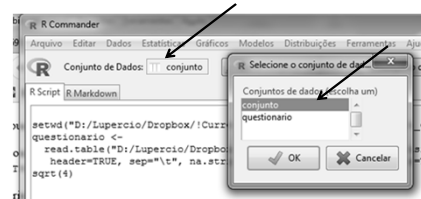
- 1º passo:
 - ✓ Clicar em Dados e selecionar “Carregar conjunto de dados ...”



- 2º passo:
 - ✓ Digitar os dados
 - ✓ Inserir colunas (TAB) quando necessário
 - ✓ Inserir linhas (ENTER) quando necessário



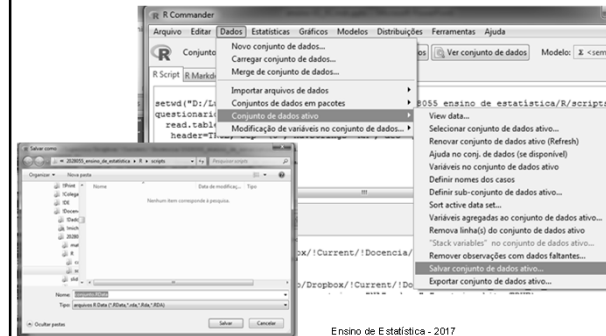
- Conjunto de dados está no R Commander
 - ✓ Clicar em conjuntos de dados
 - ✓ Selecionar o conjunto de interesse



Ensino de Estatística - 2017

41

- Salvamento do conjunto de dados:
 - ✓ Clique em Dados:
 - ✓ Selecionar “Conjunto de dados ativo” e “Salvar conjunto de dados ativo ...”



Ensino de Estatística - 2017

42

Conjuntos de Dados Internos do R

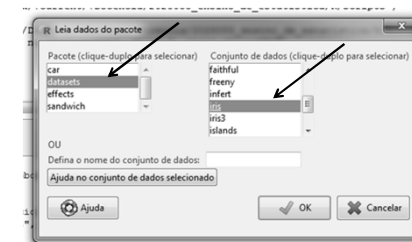
- Dados que vêm junto com o programa ou em seu pacotes
- Ver os variados tipos de conjunto de dados
 - ✓ Clicar em “Dados”
 - ✓ Selecionar “Conjunto de dados em pacotes” e “Listar conjuntos de dados em pacotes”



Ensino de Estatística - 2017

43

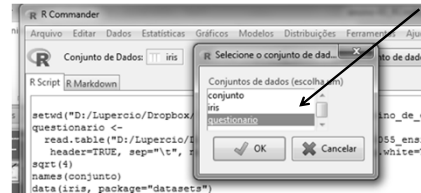
- Carregamento conjunto de dados de pacote:
 - ✓ Clicar em “Dados”
 - ✓ Selecionar “Conjunto de dados em pacotes” e “Ler dados de pacote attachado ...”
 - ✓ Escolher conjunto de dados de interesse



Ensino de Estatística - 2017

44

- Selecionar conjunto de dados de interesse:
 ✓ Escolher conjunto de dados “questionario”



Ensino de Estatística - 2017

45

- Conjunto de dados carregado

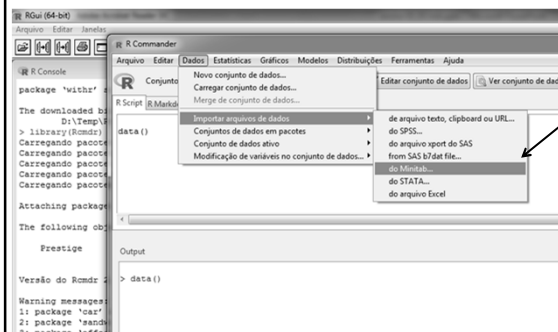
Id	Turma	Sexo	Idade	Alt	Peso	Filhos	Fuma	Toler	Exerc	Cine	OpCine	TV	OpTV
1	A	F	17	1,60	60,6	2	NAC	P	0	1	B	16	R
2	A	F	18	1,69	66,0	1	NAC	M	0	1	B	7	R
3	A	M	18	1,85	72,8	2	NAC	P	5	2	M	15	R
4	A	M	25	1,85	80,9	2	NAC	P	5	2	B	20	R
5	A	F	19	1,68	65,0	1	NAC	M	2	2	B	5	R
6	A	M	19	1,76	60,0	3	NAC	M	2	1	B	2	R
7	A	F	20	1,60	69,0	1	NAC	P	3	1	B	7	R
8	A	F	18	1,64	47,0	1	SIM	I	2	2	M	10	R
9	A	F	18	1,62	67,8	3	NAC	M	3	3	M	12	R
10	A	F	17	1,64	68,0	2	NAC	M	2	2	M	10	R
11	A	F	18	1,72	70,0	1	SIM	I	10	2	B	8	N
12	A	F	18	1,66	64,0	3	NAC	M	0	2	B	0	R
13	A	F	21	1,70	68,0	2	NAC	M	6	1	M	30	R
14	A	M	19	1,78	66,6	1	SIM	I	6	1	M	2	N
15	A	F	18	1,66	63,6	1	NAC	I	4	1	B	10	R
16	A	F	19	1,63	47,4	3	NAC	P	0	1	B	18	R
17	A	F	17	1,82	66,0	1	NAC	P	3	1	B	10	N
18	A	M	18	1,80	85,2	2	NAC	P	3	4	B	10	R
19	A	F	20	1,60	64,6	1	NAC	P	3	2	B	5	R
20	A	F	18	1,68	62,6	3	NAC	M	7	2	B	14	M
21	A	F	21	1,70	60,0	2	NAC	P	9	2	B	5	R
22	A	F	18	1,66	69,6	1	NAC	M	0	3	B	5	R
23	A	F	16	1,57	49,2	1	SIM	I	5	4	B	10	R
24	A	F	20	1,66	48,0	1	SIM	I	0	1	M	28	R
25	A	F	20	1,69	61,6	2	NAC	P	8	5	M	4	N
26	A	F	19	1,64	67,0	2	NAC	I	6	2	B	5	R
27	B	F	23	1,62	63,0	2	NAC	M	8	2	M	5	R
28	B	F	18	1,62	62,0	1	NAC	P	1	1	M	10	R
29	B	F	19	1,57	49,0	2	NAC	P	3	1	B	12	R

Ensino de Estatística - 2017

46

Importação de Conjunto de Dados – Extensão .csv

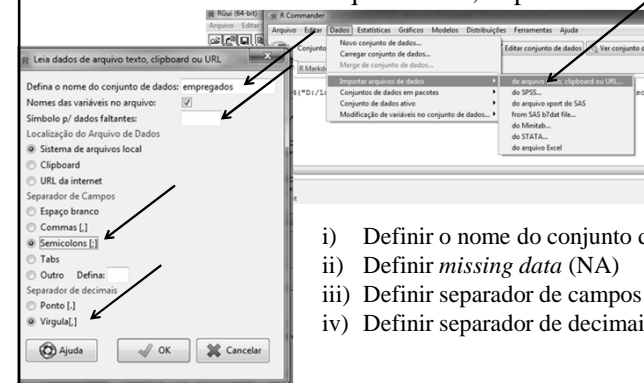
- 1º passo:
 ✓ Clicar em Dados e selecionar “Importar arquivo de dados”



Ensino de Estatística - 2017

47

- 2º passo – Arquivo extensão .csv:
 ✓ Selecionar “de arquivo texto, clipboard ou ...”

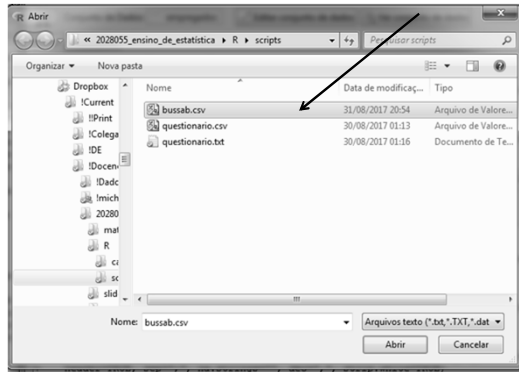


Ensino de Estatística - 2017

48

- i) Definir o nome do conjunto de dados
- ii) Definir missing data (NA)
- iii) Definir separador de campos
- iv) Definir separador de decimais

- 3º passo: Selecionar o arquivo desejado:



Ensino de Estatística - 2017

49

- Conjunto de dados carregados:

A screenshot of an R console window showing a data frame named 'empregados'. The data frame has 30 rows and 8 columns. The columns are: ID, Estado.Civil, Instru.º, Filhos, Salario, Anos, Meses, and Proced.nóia. The data is as follows:

ID	Estado.Civil	Instru.º	Filhos	Salario	Anos	Meses	Proced.nóia
1	1	0	0	NA	4.00	26	3
2	2	1	0	1	4.56	32	10
3	3	1	0	2	5.25	36	5
4	4	0	1	NA	5.73	20	10
5	5	0	0	NA	6.26	40	7
6	6	1	0	0	6.66	28	0
7	7	0	0	NA	6.86	41	0
8	8	0	0	NA	7.39	43	4
9	9	1	1	1	7.59	34	10
10	10	0	1	NA	7.44	23	6
11	11	1	1	2	8.12	33	6
12	12	0	0	NA	8.46	27	11
13	13	0	1	NA	8.74	37	5
14	14	1	0	3	8.95	44	2
15	15	1	1	0	9.13	30	5
16	16	0	1	NA	9.55	38	8
17	17	1	1	1	9.77	31	7
18	18	1	0	2	9.80	39	7
19	19	0	2	NA	10.53	25	8
20	20	0	1	NA	10.76	37	4
21	21	1	1	1	11.06	30	9
22	22	0	1	NA	11.59	34	2
23	23	0	0	NA	12.00	41	0
24	24	1	2	0	12.79	26	1
25	25	1	1	2	13.23	32	5
26	26	1	1	2	13.60	35	0
27	27	0	0	NA	13.85	46	7
28	28	1	1	0	14.69	29	8
29	29	1	1	5	14.71	40	6
30	30	1	1	2	15.99	35	10

Ensino de Estatística - 2017

50

Manipulação dos Dados

Preparação dos Dados

- Valores perdidos (*missing data*):
 - √ Devem ser codificados por NA
- Verificar fatores:
 - √ Necessário garantir que o R reconheça as variáveis categóricas e sua ordem (variáveis ordinais)
 - √ Pode ser adequado renomear os fatores para facilitar sua compreensão.

Ensino de Estatística - 2017

52

- Verificação da estrutura do banco:
 - √ Digitar comando no script e compilar :
 - Ou direto no prompt do R

```
> str(empregados)
'data.frame': 36 obs. of 8 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Estado.Civil: int  0 1 1 0 0 1 0 0 1 0 ...
 $ Instru.Æo : int  0 0 0 1 0 0 0 0 1 1 ...
 $ Filhos   : int  NA 1 2 NA NA 0 NA NA 1 NA ...
 $ Salario  : num  4 4.56 5.25 5.73 6.26 6.66 6.86 7.39 7.59 7.44 ...
 $ Anos     : int  26 32 36 20 40 28 41 43 34 23 ...
 $ Meses    : int  3 10 5 10 7 0 0 4 10 6 ...
 $ Proced.ncia : Factor w/ 3 levels "capital","interior",...: 2 1 1 3 3 2 ...
```

√ Os dados estão tecnicamente corretos?

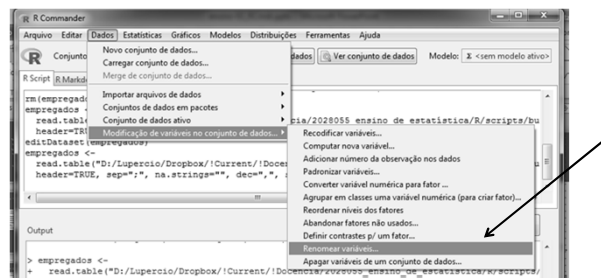
- Manipulação do conjunto de dados:

```
> str(empregados)
'data.frame': 36 obs. of 8 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Estado.Civil: int  0 1 1 0 0 1 0 0 1 0 ...
 $ Instru.Æo : int  0 0 0 1 0 0 0 0 1 1 ...
 $ Filhos   : int  NA 1 2 NA NA 0 NA NA 1 NA ...
 $ Salario  : num  4 4.56 5.25 5.73 6.26 6.66 6.86 7.39 7.59 7.44 ...
 $ Anos     : int  26 32 36 20 40 28 41 43 34 23 ...
 $ Meses    : int  3 10 5 10 7 0 0 4 10 6 ...
 $ Proced.ncia : Factor w/ 3 levels "capital","interior",...: 2 1 1 3 3 2 ...
```

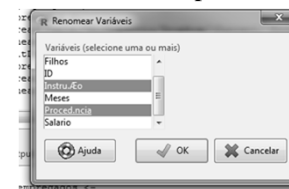
- √ NA's reconhecidos
- √ Variáveis com denominação errada:
 - Instrução
 - Procedência
- √ Variáveis não reconhecidas como fatores:
 - Estado civil (nominal)
 - Instrução (ordinal)

Renomear Variáveis

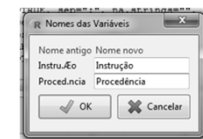
- 1º passo:
 - √ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Renomear variáveis ...”



- 2º passo:
 - √ Selecionar as variáveis para renomeação



- 3º passo:
 - √ Renomear variáveis



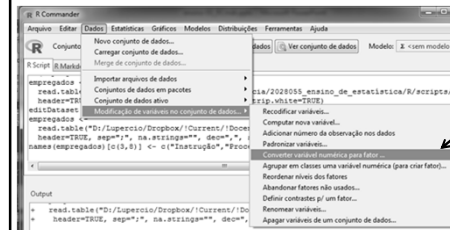
- Conjunto de dados com cabeçalho corrigido

ID	Estado.Civil	Instrução	Filhos	Salario	Anos	Meses	Procedência
1	0	0	NA	4.00	26	3	interior
2	1	0	1	4.56	32	10	capital
3	1	0	2	5.25	36	5	capital
4	0	1	NA	5.73	20	10	outra
5	0	0	NA	6.26	40	7	outra
6	1	0	0	6.66	28	0	interior
7	0	0	NA	6.86	41	0	interior
8	0	0	NA	7.39	43	4	capital
9	1	1	1	7.59	34	10	capital
10	0	1	NA	7.44	23	6	outra
11	1	2	8.12	33	6	interior	
12	0	0	NA	8.46	27	11	capital
13	0	1	NA	8.74	37	5	outra
14	1	0	3	8.95	44	2	outra
15	1	1	0	9.13	30	5	interior
16	0	1	NA	9.35	38	8	outra
17	1	1	1	9.77	31	7	capital
18	1	0	2	9.80	39	7	outra
19	0	2	NA	10.53	25	8	interior
20	0	1	NA	10.76	37	4	interior
21	1	1	1	11.06	30	9	outra
22	0	1	NA	11.59	34	2	capital
23	0	0	NA	12.00	41	0	outra
24	1	2	0	12.79	26	1	outra
25	1	1	2	13.23	32	5	interior
26	1	1	2	13.60	35	0	outra
27	0	0	NA	13.85	46	7	outra
28	1	1	0	14.69	29	8	interior
29	1	1	5	14.71	40	6	interior
30	1	1	2	15.99	35	10	capital

Ensino de Estatística - 2017

57

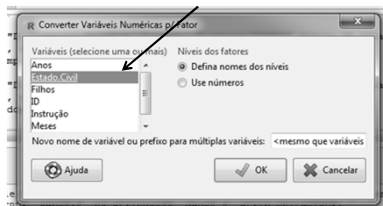
- Conversão e recodificação Estado.Civil:
 $\sqrt{1 = \text{casado}; 0 = \text{não}}$
- 1º passo:
 $\sqrt{\text{Clicar em "Dados" e selecionar "Modificação de variáveis no conjunto de dados" e "Converter variável numérica para fator ..."}}$



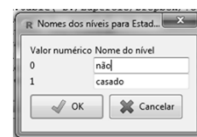
Ensino de Estatística - 2017

58

- 2º passo:
 $\sqrt{\text{Selecionar variável numérica de interesse}}$



- 3º passo:
 $\sqrt{\text{Modificar nomes dos níveis}}$



Ensino de Estatística - 2017

59

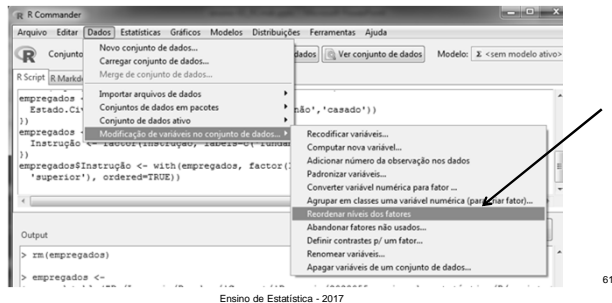
- Conversão da variável Instrução para fator e renomeação de seus níveis:
 $\sqrt{0 = \text{fundamental}}$
 $\sqrt{1 = \text{médio}}$
 $\sqrt{2 = \text{superior}}$
- Repetir o procedimento!
- Os níveis estão ordenados??

Ensino de Estatística - 2017

60

Reordenar Fator

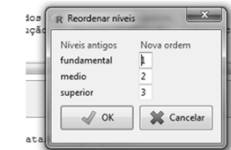
- 1º passo:
 - ✓ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Reordenar níveis dos fatores”



- 2º passo:
 - ✓ Selecionar variável:



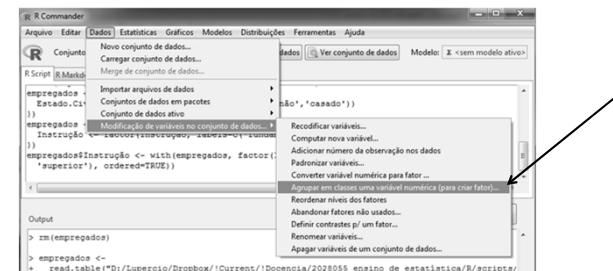
- 3º passo:
 - ✓ Estabelecer ordem dos níveis



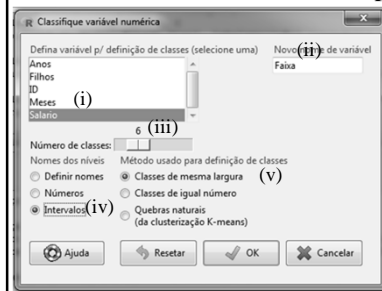
Discretização de Variável

- Discretizar variável Renda:
 - ✓ Quantidade de faixas de renda: 6
 - ✓ Amplitude faixas: a mesma
 - ✓ Nome dos níveis: a partir dos máximos e mínimos

- 1º passo:
 - ✓ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Agrupar em classes uma variável numérica (para criar fator) ...”



- 2º passo:
 - i. Selecionar variável
 - ii. Definir novo nome
 - iii. Definir quantidade de classes
 - iv. Definir nome dos níveis
 - v. Definir método da partição



Ensino de Estatística - 2017

65

- Conjunto de dados com nova variável:

ID	Estado.Civil	Instrução	Filhos	Salario	Anos	Meses	Procedência	Faixa
1	não casado	fundamental	NA	4.00	26	3	interior	(3.98,7.22]
2	casado	fundamental	1	4.56	32	10	capital	(3.98,7.22]
3	casado	fundamental	2	5.25	36	5	capital	(3.98,7.22]
4	não casado	medio	NA	5.73	20	10	outra	(3.98,7.22]
5	não casado	fundamental	NA	6.26	40	7	outra	(3.98,7.22]
6	casado	fundamental	0	6.66	28	0	interior	(3.98,7.22]
7	não casado	fundamental	NA	6.86	41	0	interior	(3.98,7.22]
8	não casado	fundamental	NA	7.39	43	4	capital	(7.22,10.4]
9	casado	medio	1	7.59	34	10	capital	(7.22,10.4]
10	não casado	medio	NA	7.44	23	6	outra	(7.22,10.4]
11	casado	medio	2	8.12	33	6	interior	(7.22,10.4]
12	não casado	fundamental	NA	8.46	27	11	capital	(7.22,10.4]
13	não casado	medio	NA	8.74	37	5	outra	(7.22,10.4]
14	casado	fundamental	3	8.95	44	2	outra	(7.22,10.4]
15	casado	medio	0	9.13	30	5	interior	(7.22,10.4]
16	não casado	medio	NA	9.35	38	8	outra	(7.22,10.4]
17	casado	medio	1	9.77	31	7	capital	(7.22,10.4]
18	casado	fundamental	2	9.80	39	7	outra	(7.22,10.4]
19	não casado	superior	NA	10.53	25	8	interior	(10.4,13.6]
20	não casado	medio	NA	10.76	37	4	interior	(10.4,13.6]
21	casado	medio	1	11.06	30	9	outra	(10.4,13.6]
22	não casado	medio	NA	11.59	34	2	capital	(10.4,13.6]
23	não casado	fundamental	NA	12.00	41	0	outra	(10.4,13.6]
24	casado	superior	0	12.79	26	1	outra	(10.4,13.6]
25	casado	medio	2	13.23	32	5	interior	(10.4,13.6]
26	casado	medio	2	13.60	35	0	outra	(10.4,13.6]
27	não casado	fundamental	NA	13.85	46	7	outra	(13.6,16.9]
28	casado	medio	0	14.69	29	8	interior	(13.6,16.9]
29	casado	medio	5	14.71	40	6	interior	(13.6,16.9]
30	casado	medio	2	15.99	35	10	capital	(13.6,16.9]

Ensino de Estatística - 2017

66

- Verificação do conjunto de dados:

```
> str(empregados)
'data.frame': 36 obs. of 9 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Estado.Civil: Factor w/ 2 levels "não","casado": 1 2 2 1 1 2 1 1 2 1 ...
 $ Instrução  : Ord.factor w/ 3 levels "fundamental"<.: 1 1 1 2 1 1 1 1 ...
 $ Filhos    : int  NA 1 2 NA NA 0 NA NA 1 NA ...
 $ Salario   : num  4 4.56 5.25 5.73 6.26 6.66 6.86 7.39 7.59 7.44 ...
 $ Anos      : int  26 32 36 20 40 28 41 43 34 23 ...
 $ Meses     : int  3 10 5 10 7 0 0 4 10 6 ...
 $ Procedência: Factor w/ 3 levels "capital","interior",,: 2 1 1 3 3 2 ...
 $ Faixa     : Factor w/ 6 levels "(3.98,7.22]",...: 1 1 1 1 1 1 1 2 2 2 ...
```

√ Podemos agora iniciar a exploração dos dados

Ensino de Estatística - 2017

67

Calcular Nova Variável

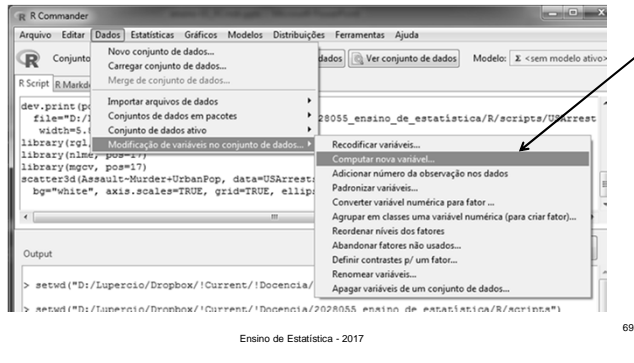
- Criar a variável:

$$\sqrt{\text{Renda}} = \text{Salario} * 350$$

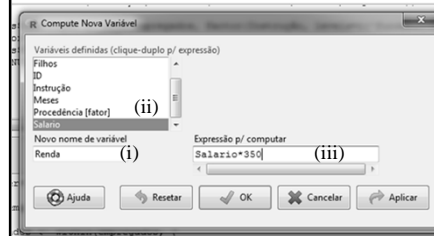
Ensino de Estatística - 2017

68

- 1º passo:
 - ✓ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Computar nova variável ...”



- 2º passo:
 - Definir nome da nova variável
 - Selecionar variável que será utilizada
 - Expressão a ser calculada



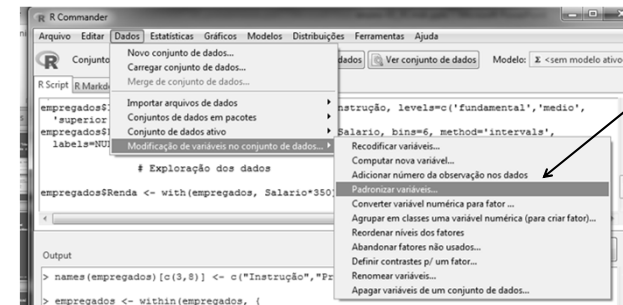
Padronização de Variável

- Criar nova variável centrada e na escala de desvios padrão:

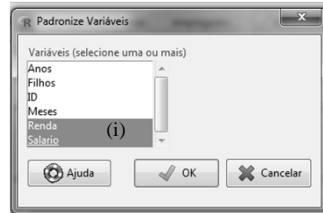
$$Z_i = \frac{x_i - \bar{x}}{s}$$

- ✓ x_i : i-ésima observação da variável X
- ✓ z_i : i-ésima observação padronizada da variável X
- ✓ \bar{x} : média da variável X
- ✓ s : desvio padrão da variável X

- 1º passo:
 - ✓ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Padronizar variáveis ...”



- 2º passo:
 - i. Selecionar as variáveis a serem padronizadas



- Verifique as novas variáveis e comente o resultado

Ensino de Estatística - 2017

73

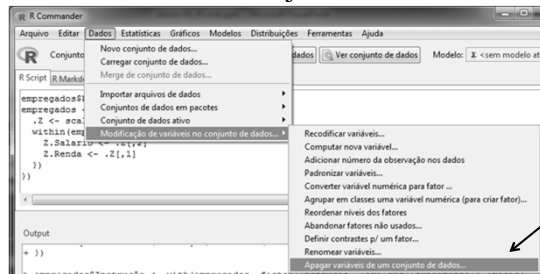
- Variáveis padronizadas são iguais:
 - √ Renda é combinação linear de salário!
 - √ Pode-se deletar uma das variáveis

Ensino de Estatística - 2017

74

Remoção de Variável

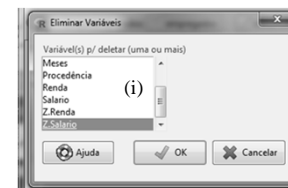
- 1º passo:
 - √ Clicar em “Dados” e selecionar “Modificação de variáveis no conjunto de dados” e “Apagar variáveis de um conjunto de dados ...”



Ensino de Estatística - 2017

75

- 2º passo:
 - i. Selecionar as variáveis a serem descartadas

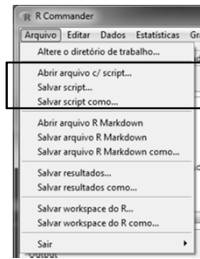


Ensino de Estatística - 2017

76

Script

- Script:
 - √ Arquivo que o R Commander salva os comandos efetuados



Ensino de Estatística - 2017

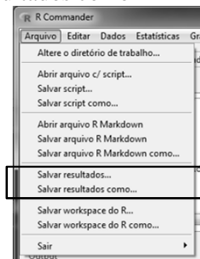
77

- Salvar Script:
 - √ Salva o que tiver sido feito no Script que você estiver trabalhando no momento
- Salvar script como
 - √ Salva script com outro nome
- Abrir arquivo com script
 - √ Abre scripts criados em outros momentos
- #
 - √ Possibilita comentários no R
 - √ R para interpretação da linha

Ensino de Estatística - 2017

78

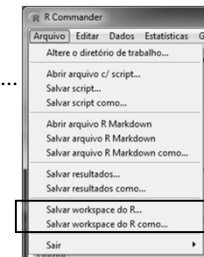
- Resultados:
 - √ O que aparece na janela Output
 - √ Comandos:
 - Salvar resultados
 - Salvar resultados como



Ensino de Estatística - 2017

79

- R workspace:
 - √ Salva o que foi feito ou inserido no R Commander na memória
 - √ Comandos:
 - Salvar workspace do R ...
 - Salvar do workspace do R como ...



Ensino de Estatística - 2017

80

Exploração dos Dados

COMANDOS

Ensino de Estatística - 2017

82

Análise Exploratória

Associações Bivariadas

- Quantitativa vs. Qualitativa
- Quantitativa vs. Quantitativa
- Qualitativa vs. Qualitativa

Ensino de Estatística - 2017

84

Variável Quantitativa vs. Variável Qualitativa

Variável Qualitativa vs. Quantitativa

Objetivo:

- representar graficamente as duas variáveis combinadas;
- definir e calcular uma medida de associação entre as variáveis.

Ensino de Estatística - 2017

86

Box-plot

- Pode ser utilizado para comparações entre diferentes grupos de dados
 - √ Variável quantitativa vs. variável categórica

87

Ensino de Estatística - 2017

Exemplo – Questionário

- Dados sobre:
 - √ Sexo
 - √ Idade
 - √ Altura
 - √ Peso
 - √ Qte filhos na família
 - √ etc.
 - √ Planilha: *BD_elementos_131.xls/questionario*

Ensino de Estatística - 2017

88

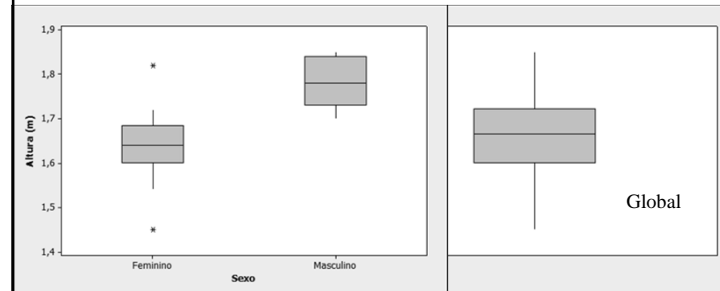
Análise de Peso e Altura

- Objetivo:
 - Analisar o comportamento do peso e da altura com relação ao sexo
 - √ Análise gráfica
 - √ Análise de medidas resumo

Ensino de Estatística - 2017

89

• Box-plot



- √ O padrão de altura é diferente para cada sexo?
- √ Sexo ajuda a explicar a variabilidade de altura?
- √ Simetria? Locação? Variabilidade? *Outliers*?

Ensino de Estatística - 2017

90

• Medidas resumo por Sexo:

Descriptive Statistics: Altura

Variable	Sexo	N	Mean	StDev	CoefVar	Median	IQR
Alt	F	37	1,6338	0,0665	4,07	1,6400	0,0850
	M	13	1,7808	0,0541	3,04	1,7800	0,1100

• Medidas resumo – Global:

Descriptive Statistics: Altura

Variable	N	Mean	StDev	CoefVar	Median	IQR
Alt	50	1,6720	0,0906	5,42	1,6650	0,1225

- √ Simetria?
- √ Locação?
- √ Variabilidade?

Ensino de Estatística - 2017

91

Comentário

- É possível perceber, a partir dessas medidas e gráficos, uma dependência entre altura e sexo?

Ensino de Estatística - 2017

92

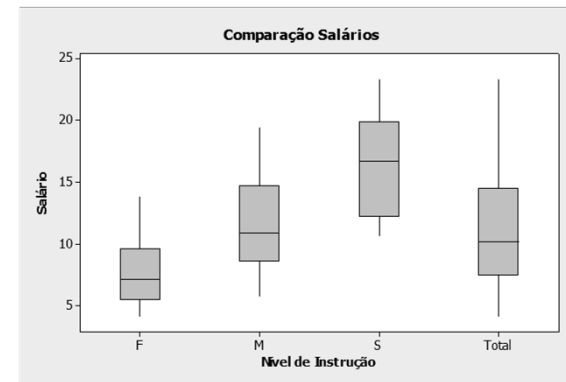
Exemplo – Dados de Empregados

- Dados sobre:
 - √ grau de instrução
 - F(Ensino Fundamental), M(Ensino Médio) e S(Ensino Superior)
 - √ salário
 - fração do Salário Mínimo
 - √ Região de procedência
- Planilha: *BD_elementos_131.xls/ciaMB*
- √ Fonte: Bussab e Morettin, Cap. 2

Ensino de Estatística - 2017

93

- Box-plots para comparação global de salários



Ensino de Estatística - 2017

94

- Medidas resumo comparando-se com o Total

Descriptive Statistics: sal_1								
Variable	instrucao_1	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
sal_1	F	7,837	2,956	4,000	5,503	7,125	9,588	13,850
	M	11,528	3,715	5,730	8,585	10,910	14,695	19,400
	S	16,48	4,50	10,53	12,23	16,74	19,89	23,30
	Total	11,122	4,567	4,000	7,476	10,165	14,480	23,300

Ensino de Estatística - 2017

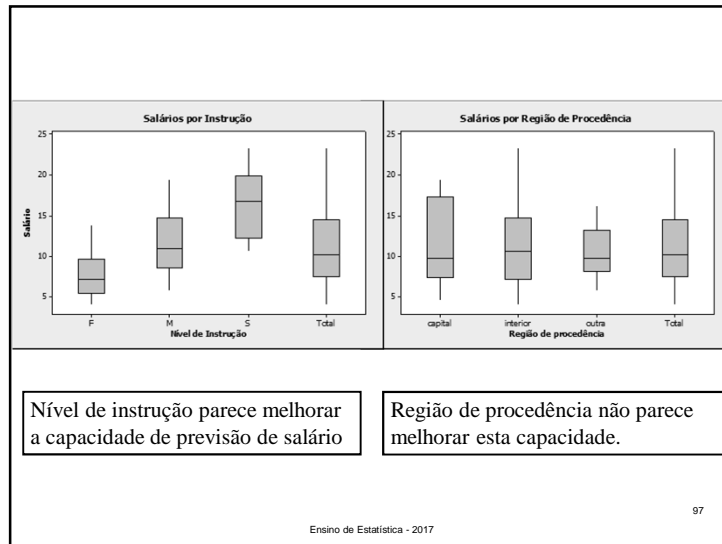
95

Comentário

- Há indícios de uma associação entre salário e nível de instrução?
 - √ Há indícios amostrais de aumento do salário com o nível de escolaridade do empregado

Ensino de Estatística - 2017

96



• Salários vs. Escolaridade:

√ Cálculo das variâncias:

```
MTB > Describe 'sal_1';
SUBC> By 'instrucao_1';
SUBC> Variance.

Descriptive Statistics: sal_1

Variable  instrucao_1  Variance
sal_1    F             8,741
          M            13,802
          S            20,27
          Total        21,045
```

As variâncias DENTRO de cada nível são menores que a variância global

• Salários vs. Região de Procedência:

√ Cálculo das variâncias:

```
MTB > Describe 'sal_1';
SUBC> By 'rp_1';
SUBC> Variance.

Descriptive Statistics: sal_1

Variable  rp_1      Variance
sal_1    capital  29,99
          interior 28,05
          outra   9,894
          Total  21,045
```

As variâncias DENTRO de cada nível não são menores que a variância global

Referências

Bibliografia Recomendada

- AGRESTI, A.; FINLAY, B. *Métodos estatísticos para as ciências sociais*. Penso, 2012.
- MAGALHÃES, M.N.; LIMA, A.C.P.L. *Noções de Probabilidade e Estatística*. Edusp, 2011.
- MOORE, D. S.; MCCABE, G. P. *Introdução à prática da estatística*. LTC, 2002.
- WILD, C.J. E SEBER, G.A.F. *Encontros com o acaso: um primeiro curso de análise de dados e inferência*. LTC, 2000.