



Introdução ao R 4. Análise univariada 1/31

Fúlvio Nedel SPB/UFSC

Objetivos d análise Análise univariada Variáveis numéricas Variáveis

## Introdução ao uso do



## em Ciências da Saúde

4 - Análise de dados univariada

### Fúlvio Borges Nedel

Departamento de Saúde Pública – SPB Centro de Ciências da Saúde – CCS Universidade Federal de Santa Catarina – UFSC

Grups de Recerca d'Amèrica i Àfrica Llatines – GRAAL http://graal.uab.cat

25 de outubro de 2017

## Roteiro

#### Introdução ao R 4. Análise univariada 2/31

Fúlvio Nedel SPB/UFSC

Introdução Objetivos análise

univariad

Variáveis numéricas Variáveis categóricas

- 1 Introdução
  - Objetivos da análise

- 2 Análise univariada
  - Variáveis numéricas
  - Variáveis categóricas



# Introdução Relembrando...

#### Introdução ao R 4. Análise univariada 3/31

Fúlvio Nedel SPB/UFSC

Introdução

Objetivos da análise

Análise univariada Variáveis numéricas Variáveis categóricas

## Objetivos da análise

- Descrever o Índice de Massa Corporal (IMC) e analisar fatores associados à sua média.
- Descrever a frequência de categorias do estado nutricional e analisar fatores possivelmente associados à obesidade:
  - 1 sexo
  - 2 idade
  - 3 condição socioeconômica
  - 4 participação em grupos de promoção da saúde

## Roteiro

#### Introdução ao R 4. Análise univariada 4/31

Fúlvio Nedel SPB/UFSC

Introdução Objetivos d análise

Análise univariada

numéricas Variáveis categóricas

- 1 Introdução
  - Objetivos da análise

- 2 Análise univariada
  - Variáveis numéricas
  - Variáveis categóricas

## Análise univariada

Introdução ao R 4. Análise univariada 5/31

Fúlvio Nedel SPB/UFSC

Introdução Objetivos d análise

Análise univariada

Variáveis numéricas Variáveis

- Toda análise deve começar com uma descrição univariada
  - completa e
  - adequada ao tipo de variável (contínua, categórica, ...)
- Eventualmente queremos calcular intervalos de confiança para médias e proporções
- Crie uma nova linha comentada em seu arquivo de sintaxe. Algo como:

```
# Análise dos dados
# ------
### Análise univariada
```

## Descrição - análise univariada



Introdução ao R 4. Análise univariada 6/31

Fúlvio Nedel SPB/UFSC

Introdução Objetivos o análise

Análise univariada

- Inicie com uma sessão vazia e carregue o arquivo de dados
- Verá que ele agora tem dois objetos,
  - o banco de dados com as variáveis selecionadas antes da transformação e
  - o banco com as variáveis transformadas, para a análise.

```
rm(list=ls())
load('cursoR.RData')
ls()
[1] "cursoR" "cursoR2"
class(cursoR)
[1] "data.frame"
class(cursoR2)
[1] "data.frame"
```



Introdução ao

4. Análise

## Descrição – análise univariada

Uma primeira olhada no banco

3 Variables 300 Observations

56 65 74 80

n missing distinct 300 0

lowest : 23 25 26 29 34, highest: 89 91 93 94 95



library(Hmisc) describe(cursoR2[1:3]) cursoR2[1:3]

idade

Oferece um resumo um pouco mais detalhado e de muito melhor leitura que summary

univariada 7/31 Fúlvio Nedel

Análise univariada

Frequency 198 Proportion 0.66 0.34

imc

Value

24.22 27.18 30.48 34.76

sexo : u8. Sexo:

Feminino Masculino

.25 .50 .75 .90 .95 37.21

lowest: 17.30104 17.90123 18.35937 18.48977 18.66201 highest: 40.62500 40.81633 41.20708 42.09964 46.13610

n missing distinct Info Mean Gmd .05 .10 300 0 60 0.999 64.57 14.49 43 49 .25 .50 .75 .90 .95

85



### Descrição – análise univariada Uma primeira olhada no banco

R A A L

```
Introdução ao
R
4. Análise
univariada 8/31
```

Fúlvio Nedel SPB/UFSC

Objetivos

Análise univariada

Variáveis numéricas Variáveis imccat : Estado nutricional
n missing distinct

3 Variables 300 Observations

Value normal sobrepeso obesidade

describe(cursoR2[4:6])

cursoR2[4:6]

299

obeso

Frequency 99 116 84 Proportion 0.331 0.388 0.281

abep2 : Classificação ABEP agrupada n missing distinct

n missing distinct 299 1 2

Value sim não Frequency 84 215 Proportion 0.281 0.719

282 18 3

Value A/B C D/E
Frequency 73 178 31
Proportion 0.259 0.631 0.110



## Descrição – análise univariada Uma primeira olhada no banco



```
Introdução ao
  4. Análise
univariada 9/31
```

Fúlvio Nedel

Análise univariada

```
id
```

```
describe(cursoR2[7:8])
cursoR2[7:8]
  Variables 300 Observations
grupo : Participa em grupo de hipertensos ou diabéticos
      n missing distinct
    298
Value
       Não Sim
Frequency 285 13
Proportion 0.956 0.044
      n missing distinct
    300
                     300
lowest: 1 10 100 101 102, highest: 95 96 97 98 99
```



# Descrição – análise univariada Uma primeira olhada no banco

grupo : Participa em grupo de hipertensos ou diabéticos

lowest: 1 10 100 101 102, highest: 95 96 97 98 99



```
Introdução ao R
4. Análise
univariada 9/31
Fúlvio Nedel
SPB/UFSC
Introdução
Observas di
análise
univariada
```

```
2 Variables 300 Observations
```

n missing distinct

Value Não Sim Frequency 285 13 Proportion 0.956 0.044

describe(cursoR2[7:8])

cursoR2[7:8]

298

id
n missing distinct
300 0 300

class(cursoR2\$id)
[1] "character"

describe(as.numeric(cursoR2\$id))

as.numeric(cursoR2\$id)

n missing distinct Info Mean Gmd .05 300 0 300 1 150.5 100.3 15.95

.10 .25 .50 .75 .90 .95 30.90 75.75 150.50 225.25 270.10 285.05

## Alguns pacotes em R para estudos epidemiológicos

Introdução ao R 4. Análise univariada 10/31

Fúlvio Nedel SPB/UFSC

Objetivos

Análise univariada há vários pacotes, além do Hmisc, para facilitar a análise e apresentação de dados epidemiológicos: Epi, epitools, epibasix, gmodels...

■ instale e carregue o pacote epiDisplay

```
install.packages("epiDisplay", dep=T)

library(epiDisplay)

Loading required package: foreign

Loading required package: MASS

Loading required package: nnet

Attaching package: 'epiDisplay'
The following object is masked from 'package:ggplot2':
    alpha
The following object is masked from 'package:lattice':
    dotplot
```

#### Introdução ao R 4. Análise

4. Análise univariada 11/31

Fúlvio Nedel SPB/UFSC

ntrodução Objetivos análise

univariada Variáveis

numéricas

Variaveis categóricas

### Medidas de

- tendência central (média, mediana, moda)
- posição (quantis)
- dispersão (amplitude, desvio-padrão, coeficiente de variação, intervalo interquartílico)
- forma (assimetria e curtose)
- já vimos algumas funções (mean, sd)

### **Gráficos**

- histograma
- boxplot
- densidade
- polígono de frequência acumulada
- de barras, para variáveis discretas
- de pontos, ramo-e-folhas...





```
Introdução ao
R
4. Análise
univariada
12/31
```

Fúlvio Nedel SPB/UFSC

Introdução
Objetivos da análise

Variáveis numéricas

```
categóricas
```

```
attach(cursoR2)
```

### Média

```
mean(imc)
```

[1] NA

### summary(imc)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's 17.30 24.22 27.18 27.77 30.48 46.14 1

```
Introdução ao
R
```

4. Análise univariada 12/31

Fúlvio Nedel SPB/UFSC

Objetivos análise Análise

Variáveis numéricas

Variáveis categóricas

```
Média
```

```
mean(imc, na.rm = T)
[1] 27.76622
```

#### Mediana

```
median(imc, na.rm = T)
[1] 27.1809
```

### Amplitude

```
range(imc, na.rm = T)
[1] 17.30104 46.13610
```

### Desvio-padrão e coeficiente de variação

```
sd(imc, na.rm = T)
[1] 4.971806
round(sd(imc, na.rm = T)/mean(imc, na.rm = T)*100, 2)
[1] 17.91
```

```
Introducão ao
 4 Análise
 univariada
```

Fúlvio Nedel

Variáveis numéricas

```
Assimetria
```

```
e1071::skewness(imc, na.rm = TRUE)
[1] 0.6374804
```

#### Curtose

```
e1071::kurtosis(imc, na.rm = T)
[1] 0.2573438
```

#### Quantiles

```
quantile(imc, na.rm = T)
     0% 25% 50% 75% 100%
17.30104 24.21875 27.18090 30.47595 46.13610
quantile(imc, p = c(.025, .975), na.rm = T)
   2.5% 97.5%
19.79113 38.60288
```

### Intervalo de confiança para a média

```
ci(imc)
       mean sd se lower95ci upper95ci
299 27.76622 4.971806 0.287527 27.20038
                                     28.33206
```





### Introdução ao

4. Análise univariada 14/31

Fúlvio Nedel SPB/UFSC

ntrodução Objetivos ( análise

univariada Variáveis

numéricas

categóricas

### Moda

- A moda é uma característica tanto de variáveis numéricas como categóricas
- É o valor mais frequente, ou, em variáveis contínuas, o de maior densidade de frequência.
- Não há entre as funções básicas do R, uma que calcule a moda.



ntrodução Objetivos análise Análise

univariada Variáveis

numéricas

categór

#### Moda

- A moda é uma característica tanto de variáveis numéricas como categóricas
- É o valor mais frequente, ou, em variáveis contínuas, o de maior densidade de frequência.
- Não há entre as funções básicas do R, uma que calcule a moda.
- Podemos construí-la:

```
moda <- function(x) {
    d = density(x, na.rm = T)
    i <- which.max(d$y)
    d$x[i]
}
moda(imc)
[1] 25.40029</pre>
```

## Variáveis numéricas – Gráficos

```
Introdução ao
R
```

4. Análise univariada 15/31

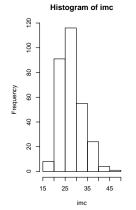
Fúlvio Nedel SPB/UFSC

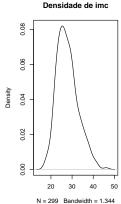
Objetivos da análise Análise

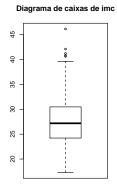
Variáveis numéricas

categóricas









Introdução ao R 4. Análise univariada 16/31

Fúlvio Nedel SPB/UFSC

ntrodução Objetivos análise

Variáveis numérica

Variáveis categóricas

- frequência absoluta e relativa
- frequências acumuladas
- gráficos de barras
- intervalos de confiança para proporções
  - distribuição binomial
  - distribuição multinomial
- já vimos algumas funções (table, cumsum)

#### Introdução ao R 4. Análise univariada 17/31

Fúlvio Nedel SPB/UFSC

Objetivos análise Análise

Variáveis categóricas

```
Estado nutricional
```

Com as funções básicas

### Frequência absoluta

```
table(imccat)
imccat
  normal sobrepeso obesidade
    99    116    84
```

addmargins(table(imccat))

### Acrescentar o total

```
imccat
normal sobrepeso obesidade
99 116 84
```

Sum

299

#### Frequência acumulada

```
cumsum(table(imccat))
  normal sobrepeso obesidade
    99 215 299
```

```
Introdução ao
R
4. Análise
univariada
18/31
```

Fúlvio Nedel SPB/UFSC

```
Objetivos
análise
Análise
```

Análise univariada Variáveis numéricas

Variáveis categóricas

## **Estado nutricional**

Com as funções básicas (cont.)

### Frequência relativa

```
prop.table(table(imccat))
imccat
    normal sobrepeso obesidade
0.3311037 0.3879599 0.2809365
```

### Porcentagem, arredondada para um decimal

#### Porcentagem acumulada

```
cumsum(round(prop.table(table(imccat))*100, 1))
  normal sobrepeso obesidade
    33.1   71.9   100.0
```





### Introdução ao

4. Análise univariada

Fúlvio Nedel SPB/UFSC

Objetivos d análise Análise univariada

numérica: Variáveis

Variáveis categóricas

### Mas...

■ um comando pra cada coisa?!





#### Introdução ao R 4. Análise

4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada

Variáveis categóricas

## Mas...

- um comando pra cada coisa?!
- até pras proporções?!!

#### Introdução ao R 4. Análise

univariada 19/31 Fúlvio Nedel

SPB/UFS

Análise

Análise

univariada

Variáveis

Variáveis categóricas

## Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

#### Introdução ao R 4. Análise univariada

Fúlvio Nedel

SPB/UFS

Objetivos d análise Análise univariada Variáveis

Variáveis categóricas

## Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

#### Introdução ao R 4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada Variáveis numéricas

Variáveis categóricas

### Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

### Bom...

■ o que pode parecer limitação

#### Introdução ao R 4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada Variáveis numéricas

Variáveis categóricas

### Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

- o que pode parecer limitação
- representa versatilidade

#### Introdução ao R 4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada Variáveis

Variáveis categóricas

### Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

- o que pode parecer limitação
- representa versatilidade
- o usuário pode construir sua tabela

Introdução ao R 4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada Variáveis

Variáveis categóricas

### Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

- o que pode parecer limitação
- representa versatilidade
- o usuário pode construir sua tabela
- ou construir funções com o modelo de tabela que deseja

Introdução ao R 4. Análise univariada 19/31

Fúlvio Nedel SPB/UFSC

Objetivos o análise Análise univariada Variáveis numéricas

Variáveis categóricas

### Mas...

- um comando pra cada coisa?!
- até pras proporções?!!
- até pro total da tabela?!!!

- o que pode parecer limitação
- representa versatilidade
- o usuário pode construir sua tabela
- ou construir funções com o modelo de tabela que deseja
- ou usar funções de outros pacotes

#### Introdução ao R 4. Análise univariada 20/31

Fúlvio Nedel SPB/UFSC

Objetivos análise Análise univariada

Variáveis categóricas

```
Estado nutricional
```

Com as funções básicas (cont.)

```
x <- table(imccat)
px <- round(prop.table(table(imccat))*100, 1)
cumx <- cumsum(x)
cumpx <- cumsum(px)
miolo <- cbind(Freq=x, '%'=px, Freq.acum=cumx, '%acum'=cumpx)
Total <- c(sum(x), sum(px), sum(x), sum(px))
tab.imccat <- rbind(miolo, Total)</pre>
```

```
tab.imccat
                                 tab.imccat[,c(1,3,2,4)]
        Freq
               % Freq.acum %acum
                                                         % %acum
                                          Freq Freq.acum
         99 33.1
                       99 33.1
normal
                                 normal
                                           99
                                                   99
                                                       33.1 33.1
sobrepeso 116 38.8 215 71.9
                                 sobrepeso 116
                                                   215
                                                       38.8
                                                           71.9
             28.1
obesidade 84
                  299 100.0
                                 obesidade 84
                                                   299
                                                       28.1 100.0
Total 299 100.0
                  299 100.0
                                 Total 299
                                                   299 100.0 100.0
```

 $\verb"t(tab.imccat[,1:2]") \textit{ \# note os decimais nas frequências absolutas :} ($ 

```
normal sobrepeso obesidade Total
Freq 99.0 116.0 84.0 299
% 33.1 38.8 28.1 100
```

## Construir uma função para apresentar as frequências

```
Introdução ao
R
4. Análise
univariada
21/31
```

Fúlvio Nedel SPB/UFSC

Introdução Objetivos o análise Análise

Variáveis numéricas

categóricas

```
tabua <- function(x, ...) {
  tab <- table(x)
  ptab <- round(prop.table(tab)*100, 1)
  cumtab <- cumsum(tab)
  cumptab <- cumsum(ptab)
  miolo <- cbind(Freq=tab, '%'=ptab, Freq.acum=cumtab, '%acum'=cumptab)
  Total <- c(sum(tab), sum(ptab), sum(tab), sum(ptab))
  tabela <- rbind(miolo, Total)
  return(tabela)
}</pre>
```

#### tabua(imccat)

```
Freq % Freq.acum %acum
normal 99 33.1 99 33.1
sobrepeso 116 38.8 215 71.9
obesidade 84 28.1 299 100.0
Total 299 100.0 299 100.0
```

## Funções em outros pacotes



```
Introdução ao
R
4. Análise
univariada
```

Fúlvio Nedel SPB/UFSC

ntrodução Objetivos di análise Análise Inivariada

Variáveis categóricas

```
Já vimos a função describe, de Hmisc
```

```
describe(imccat)
imccat : Estado nutricional
    n missing distinct
299 1 3
```

Value	normal	sobrepeso	obesidade
Frequency	99	116	84
Proportion	0.331	0.388	0.281

A função tab1, do pacote epiDisplay (antigo epicalc) apresenta a contagem de 'missings' (que table não inclui por padrão), as proporções e a frequência acumulada (que describe não apresenta), além de incluir um gráfico com os resultados da tabela.

```
tab1(imccat)
tab1(imccat, cum.percent = T)
tab1(imccat, cum.percent = T, missing = F)
```



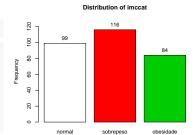
Introdução ao R 4. Análise univariada 23/31

Fúlvio Nedel SPB/UFSC

Introdução
Objetivos o
análise
Análise
univariada
Variáveis

Variáveis categóricas

#### tab1(imccat, missing = F) imccat : %(NA+) Frequency %(NA-) normal 99 33.0 33.1 sobrepeso 116 38.7 38.8 obesidade 84 28.0 28.1 NA's 0.3 0.0 Total 300 100.0 100.0



O argumento missing=FALSE faz com que não sejam incluídos no gráfico.

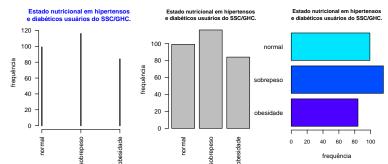
## Também podemos fazer nossos próprios gráficos

```
Introdução ao
R
4. Análise
univariada
24/31
```

Fúlvio Nedel SPB/UFSC

Objetivos d análise Análise univariada Variáveis

Variáveis categóricas





## Guardar os gráficos em arquivo



Introdução ao R 4. Análise univariada 25/31

Fúlvio Nedel SPB/UFSC

Objetivos análise Análise univariada Variáveis numéricas

Variáveis categóricas

ordinary text without R code

```
No Rstudio, gráficos podem facilmente ser salvos como figuras, PDF
      ou copiados para a área de transferência:
File Edit Code View Plots Session Build Debug Profile Tools Help
Q - | 😅 - | 🔒 🙉 🔠
 vodados.R × Pibivariada.Ru
                                            ntitled4* x P Untitled2* x P Untitled2* x P Untitled3* x C cursoft-4.mw x >> ---
                    Zoom Plot...
  00 2 8 7 8
                                                                                                                                                   ☐ List • (6)
                    Save as Image...
                                              Replace All
                                                                                                        6 Global Environment +
                    Copy to Clipboard...
                                                                                                                       300 obs. of 9 variables
                                                                                                        cursoR
  562 \end{frame}
                    Remove Plot...
                                                                                                        o cursoR2
                                                                                                                       300 obs. of 9 variables
                                                                                                          miolo
                                                                                                                       num [1:3, 1:4] 99 116 84 33,1 38,...
  564 \begin{frame, clear All...
                                                                                                          tab imccat
                                                                                                                       num [1:4, 1:4] 99 116 84 299 33,1...
      \smaller
                                                                                                        Values
  566 \begin{columns}
                                                                                                          CHRDX
                                                                                                                       Named num [1:3] 33.1 71.9 100
  567 \column{.55\textwidth}
                                                                                                                       Named int [1:3] 99 215 299
  568 - <<eval=FALSE,echo=TRUE>>=
  569 tabl(imccat, missing = F)
                                                                                                         Files Plots Packages Help Viewer
                                                                                                         🗇 🍮 🔎 Zoom 🝱 Export + 🧿 🥑
                                                                                                                                                S Publish → (6)
  571 - <<echo=FALSE>>=
  572 tab1(imccat, graph = F)
                                                                                                                          Distribution of imccat
  574 \column{.45\textwidth}
  575 - <echo=FALSE results='hide'.out.width='1.1\\textwidth'.fig.height=5.fig.width=6>>=
  576 tabl(imccat, missing = F)
                                                                                                            20
                                                                                                                                  116
  577 ⊚
  578 \end{columns}
                                                                                                                      99
  579 \larger[.5]\redescuro
                                                                                                            8
  580 O argumento \texttt{missing=FALSE} faz com que não sejam incluídos no gráfico.
  581 \end{frame}
  582
 Console Compile PDF x
                                                                                                            8
 CursoR/cursoR-4.rnw
                                                                                                            9
  ordinary text without R code
 label: unnamed-chunk-1 (with options)
 List of 1
  S include: logi FALSE
                                                                      3%
                                                                                                                    normal
                                                                                                                                sobrepeso
                                                                                                                                             obesidade
```

1 4%



#### Guardar os gráficos em arquivo



Introdução ao R 4. Análise univariada 25/31

Fúlvio Nedel SPB/UFSC

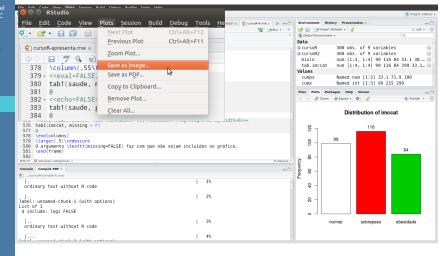
análise

Análise
univariada

Variáveis
numéricas

Variáveis
categóricas

No Rstudio, gráficos podem facilmente ser salvos como figuras, PDF ou copiados para a área de transferência:





#### Guardar os gráficos em arquivo

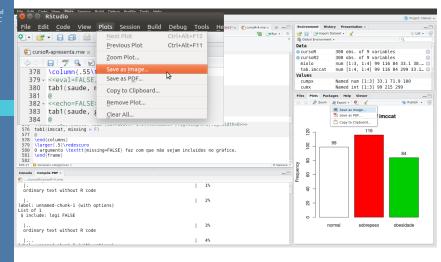


Introdução ao R 4. Análise univariada 25/31

Fúlvio Nedel SPB/UFSC

Introdução
Objetivos d
análise
Análise
univariada
Variáveis
numéricas

Variáveis categóricas No Rstudio, gráficos podem facilmente ser salvos como figuras, PDF ou copiados para a área de transferência:





Introdução ao R

> 4. Análise univariada 26/31

Fúlvio Nedel SPB/UFSC

Introdução Objetivos

Análise univariada Variáveis

Variáveis categóricas ■ Para buscar uma função que os calcule, comece com

??"confidence interval"

- Temos (no mínimo)
  - $\blacksquare \ \mathsf{prop.test} \{ \mathsf{stats} \} \to \mathsf{aproxima} \\ \mathsf{ç} \\ \mathsf{ão} \ \mathsf{normal} \ \mathsf{e}$
  - $\blacksquare$  ci{epiDisplay}  $\rightarrow$  método exato (binomial)

Variáveis dicotômicas

table(obeso)

0.2809365



#### Introdução ao R 4. Análise univariada

Fúlvio Nedel

Introdução

análise Análise univariada

```
Trabalhar, de momento, com a variável obeso, que é dicotômica
```

```
obeso
sim não
84 215

prop.test(table(obeso))
1-sample proportions test with continuity correction

data: table(obeso), null probability 0.5
X-squared = 56.522, df = 1, p-value = 5.558e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.2314582 0.3361532
sample estimates:

p
```



Variáveis dicotômicas



#### Introdução ao R 4. Análise

4. Análise univariada 27/31

Fúlvio Nedel SPB/UFSC

Objetive análise

> Análise Inivariada Variáveis

```
Trabalhar, de momento, com a variável obeso, que é dicotômica
```

```
table(obeso)
obeso
 im não
 84 215
prop.test(table(obeso))
1-sample proportions test with continuity correction
data: table(obeso), null probability 0.5
X-s quared = 56.522, df = 1, p-value = 5.558e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.2314582 0.3361532
samble estimates:
0.2809365
```



Variáveis dicotômicas

table(obeso)



```
Introdução ao
R
4. Análise
univariada
```

Fúlvio Nedel

```
,
Introdução
```

análise Análise univariada

```
Trabalhar, de momento, com a variável obeso, que é dicotômica
```

```
obeso
 im não
 84 215
prop.test(table(obeso))
1-sample proportions test with continuity correction
data: table(obeso), null probability 0.5
X-s quared = 56.522, df = 1, p-value = 5.558e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.2314582 0.3361532
samble estimates:
0.2809365
```

```
ci(obeso)
```

```
events total probability se exact.lower95ci exact.upper95ci 215 299 0.7190635 0.02599274 0.6644214 0.769295
```



Variáveis dicotômicas

215



0.769295

```
Introdução ao
R
4. Análise
univariada
```

Fúlvio Nedel

```
Introdução
```

Análise univariada

Variáveis categóricas

```
Trabalhar, de momento, com a variável obeso, que é dicotômica
table(obeso)
obeso
  op.test(table(obeso))
1-sample proportions test with continuity correction
data: table(obeso), null probability 0.5
X-squared = 56.522, df = 1, p-value = 5.558e-14
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
0.2314582 0.3361532
sample estimates:
0.2809365
ci(ceso)
 events total probability se exact.lower95ci exact.upper95ci
```

299 0.7190635 0.02599274 0.6644214

Variáveis dicotômicas



```
Introdução ao
R
4. Análise
univariada
28/31
```

Fúlvio Nedel SPB/UFSC

```
Introdução
Objetivos
análise
Análise
```

univariada Variáveis numéricas

```
Apresentar os resultados numa tabela bonita ;)
```

```
x <- ci(obeso)
\# x[5:6]
IC95 \leftarrow matrix(c(as.numeric(rev(1-x[5:6])), as.numeric(x[5:6])),
             nrow = 2, byrow = T)
colnames(IC95) <- c('IC95.inf', 'IC95.sup')</pre>
tab.obeso <- cbind(tabua(obeso)[-3,1:2], round(IC95*100, 2))
options(OutDec = ',') # marcador decimal como vírqula
tab.obeso
    Freq % IC95.inf IC95.sup
    84 28,1 23,07 33,56
sim
não 215 71,9 66,44 76,93
```

# Intervalos de confiança para proporções Variáveis politômicas

```
Introdução ao
R
4. Análise
univariada
29/31
```

Fúlvio Nedel SPB/UFSC

Objetivos ( análise Análise univariada

Variáveis categóricas

#### De volta a estado nutricional, que é politômica

```
tabua(imccat)[,-3]

Freq % %acum
normal 99 33,1 33,1
sobrepeso 116 38,8 71,9
obesidade 84 28,1 100,0
Total 299 100,0 100,0
```

O que se faz muitas vezes é calcular o IC da probabilidade da categoria contra seu complemento, i.e., todas as outras categorias em conjunto.

# Intervalos de confiança para proporções Variáveis politômicas

Introdução ao R 4. Análise univariada 30/31

Fúlvio Nedel

Introdução
Objetivos o análise
Análise
univariada
Variáveis

Variáveis categóricas Entretanto, isso não é muito adequado, porque a distribuição de probabilidades da categoria de uma variável politômica segue uma distribuição multinomial, que é um pouco mais complexa que a binomial.

```
library(MultinomialCI)
round(multinomialCI(table(imccat), .05)*100, 2)
      [,1] [,2]
[1,] 27,09 39,46
[2,] 32,78 45,14
```

[3,] 22,07 34,44

# Intervalos de confiança para proporções Variáveis politômicas

G R A A

Introdução ao R 4. Análise univariada 30/31

Fúlvio Nedel SPB/UFSC

Objetivos análise Análise univariada

Variáveis categóricas Entretanto, isso não é muito adequado, porque a distribuição de probabilidades da categoria de uma variável politômica segue uma distribuição multinomial, que é um pouco mais complexa que a

```
Claro que a tabela pode ser melhorada:
```

#### Apêndice Comparar os intervalos de confiança



Introdução ao R 4. Análise univariada 31/31

Fúlvio Nedel SPB/UFSC

ntrodução Objetivos análise

Análise univariada Variáveis

	%	IC normal	IC binomial	IC multinomial
normal	33,1	27,86 - 38,8	27,8 - 38,76	27,09 - 39,46
sobrepeso	38,8	33,29 - 44,6	33,24 - 44,58	32,78 - 45,14
obesidade	28,1	23,15 - 33,62	23,07 - 33,56	22,07 - 34,44

