

MAT02018 - Estatística Descritiva

Distribuição de Frequências: R, construção de tabelas e um minuto de história

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA

Porto Alegre, 2022

ComplementaR

ComplementaR

- ▶ Nesta aula são apresentadas algumas poucas funções em R relacionadas a discussão das nossas aulas de **Distribuição de Frequências**.
- ▶ Para tal, vamos utilizar o exemplo original de **Bussab e Morettin (2010)**¹ sobre os dados dos empregados da seção de orçamentos da Companhia MB.

¹Bussab, W. O e Morettin, P. A. **Estatística Básica**, Saraiva, 2010.

ComplementaR

- ▶ A planilha eletrônica correspondente encontra-se no arquivo `companhia_mb.xlsx` ([no Moodle](#)).
- ▶ Vamos começar **carregando os dados** para o R.
 - ▶ Existem **várias formas** de se carregar diferentes **arquivos de dados** no R.
 - ▶ Como arquivo de interesse encontra-se no formato do **Excel (xlsx)**, vamos utilizar a função `read_excel` do pacote `readxl`.
 - ▶ Caso você não tenha o pacote, instale-o:
`install.packages("readxl")`.

ComplementaR

```
# install.packages("readxl")
library(readxl)

dados <- read_excel(path = "companhia_mb.xlsx")

class(dados) # classe do objeto dados

## [1] "tbl_df"     "tbl"        "data.frame"
dim(dados) # dimensão do objeto dados

## [1] 36  7
```

ComplementaR

- ▶ Note que o objeto dados é uma tabela de dados bruto.

```
head(dados) # apresenta as primeiras linhas do objeto dados
```

```
## # A tibble: 6 x 7
##       N `Estado Civil` `Grau de Instrução` `N de Filhos` `Salario` (x Sa~` Id
##   <dbl> <chr>          <chr>                  <dbl>           <dbl> <dbl>
## 1     1 solteiro      ensino fundamental        NA             4
## 2     2 casado        ensino fundamental        1             4.56
## 3     3 casado        ensino fundamental        2             5.25
## 4     4 solteiro      ensino médio            NA             5.73
## 5     5 solteiro      ensino fundamental        NA             6.26
## 6     6 casado        ensino fundamental         0             6.66
```

ComplementaR

- A função `table` retorna contagens dos valores de cada variável, e portanto, podemos utilizar esta função para a apuração dos dados, bem como para computar as frequências.

```
table(dados$`Estado Civil`)
```

```
##  
##      casado solteiro  
##        20       16
```

```
table(dados$`Grau de Instrução`)
```

```
##  
## ensino fundamental      ensino médio      superior  
##                 12                  18                  6
```

```
table(dados$`N de Filhos`)
```

```
##  
## 0 1 2 3 5  
## 4 5 7 3 1
```

ComplementaR

- ▶ A função `cut` pode ser utilizada para criar uma nova variável que expressa a antiga variável em classes.

```
dados$Idade.classes <- cut(x = dados$Idade,  
                           breaks = c(20, 29, 39, 49),  
                           include.lowest = TRUE,  
                           right = FALSE)
```

```
table(dados$Idade.classes)
```

```
##  
## [20,29] [29,39] [39,49]  
##       7      18      11
```

- ▶ Para casa: especifique de diferentes formas os argumentos `breaks`, `include.lowest`, `right` da função `table` e a avalie os seus resultados.

ComplementaR

- ▶ Uma forma de calcular as frequências relativas é dividindo o **vetor** de frequências pelo tamanho da amostra (ou conjunto de dados).

```
table(dados$`Estado Civil`) / 36
```

```
##  
##      casado    solteiro  
## 0.5555556 0.4444444
```

```
table(  
  dados$`Grau de Instrução`)/length(dados$`Grau de Instrução`)  
)
```

```
##  
## ensino fundamental      ensino médio      superior  
##            0.3333333        0.5000000       0.1666667
```

ComplementaR

- ▶ Uma outra forma de se obter as frequências relativas é utilizando a função `prop.table`.

```
prop.table(x = table(dados$`N de Filhos`))
```

```
##  
##      0      1      2      3      5  
## 0.20 0.25 0.35 0.15 0.05
```

ComplementaR

- ▶ Para obter as porcentagens, basta multiplicar as frequências relativas por 100.

```
prop.table(x = table(dados$Idade.classes)) * 100
```

```
##  
## [20,29] [29,39] [39,49]  
## 19.44444 50.00000 30.55556
```

- ▶ Se você quiser, pode arredondar os resultados com a função round.

```
round(x = prop.table(x = table(dados$Idade.classes)) * 100,  
      digits = 2)
```

```
##  
## [20,29] [29,39] [39,49]  
## 19.44 50.00 30.56
```

ComplementaR

- ▶ As frequências acumuladas podem ser obtidas com uma função de somas cumulativas, a função `cumsum`.

```
cumsum(x = table(dados$`Grau de Instrução`))
```

```
## ensino fundamental      ensino médio      superior
##                      12                  30                  36
```

```
cumsum(x = prop.table(x = table(dados$`Grau de Instrução`)))
```

```
## ensino fundamental      ensino médio      superior
##          0.3333333        0.8333333        1.0000000
```

ComplementaR

```
cumsum(x = prop.table(x = table(dados$`N de Filhos`)) * 100)

##    0     1     2     3     5
##   20   45   80   95  100

cumsum(
  round(
    prop.table(x = table(dados$Idade.classes)) * 100,
    digits = 2))

## [20,29] [29,39] [39,49]
##    19.44    69.44   100.00
```

ComplementaR

- Você pode juntar as frequências absolutas, relativas, acumuladas e porcentagens em um `data.frame` para apresentar em forma de tabela.

```
df.freq <- data.frame(  
  Idade = unique(dados$Idade.classes),  
  Freq = as.numeric(table(dados$Idade.classes)),  
  FreqRel = as.numeric(prop.table(table(dados$Idade.classes))),  
  Porcentagem = as.numeric(prop.table(table(dados$Idade.classes)) * 100),  
  FreqAcumulada = as.numeric(cumsum(table(dados$Idade.classes))),  
  FreqRelAcumulada = as.numeric(cumsum(prop.table(table(dados$Idade.classes))))
```

ComplementaR

```
df.freq
```

```
##      Idade Freq   FreqRel Porcentagem FreqAcumulada FreqRelAcumulada
## 1 [20,29)    7 0.1944444    19.44444        7       0.1944444
## 2 [29,39)   18 0.5000000    50.00000       25       0.6944444
## 3 [39,49]   11 0.3055556    30.55556       36       1.0000000
```

ComplementaR

- ▶ Existem outras formas de construir uma tabela de frequências.
 - ▶ Uma delas é utilizando a função `freq` do pacote `summarytools`.

```
# install.packages("summarytools")
```

```
summarytools::freq(dados$`Grau de Instrução`)
```

	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
ensino fundamental	12	33.33	33.33	33.33	33.33
ensino médio	18	50.00	83.33	50.00	83.33
superior	6	16.67	100.00	16.67	100.00
<NA>	0			0.00	100.00
Total	36	100.00	100.00	100.00	100.00

Construção de tabelas

Construção de tabelas

- ▶ Nesta seção são apresentadas as normas que devem ser seguidas na elaboração de tabelas.
- ▶ Tais normas devem ser empregadas em relatórios, documentos acadêmicos, científicos, etc.
- ▶ Estas normas são apresentadas na publicação **Normas de apresentação tabular (IBGE, 1993)**, que encontra-se no Moodle.

Tabelas

- ▶ **Tabela** é uma forma não discursiva de apresentar informações, das quais o dado numérico se destaca como informação central.

Tabela 4.1 População residente segundo o sexo. Censo Demográfico 2010

Sexo	População residente	Porcentagem
Homens	93.406.990	48,97%
Mulheres	97.348.809	51,03%
Total	190.755.799	100,00%

Fonte: IBGE, Censo Demográfico 2010.

Principais elementos das tabelas

- ▶ **Número** é o identificador numérico da tabela, em um conjunto de tabelas.
 - ▶ Deve ser dado em algarismos arábicos e deve suceder a palavra **Tabela**.
- ▶ **Título** explica o tipo de dado que a tabela contém.
 - ▶ Deve ser colocado acima dos dados, sem ponto final.
- ▶ **Cabeçalho** especifica o conteúdo de cada coluna.
- ▶ **Indicador de linha** especifica o conteúdo de cada linha.
- ▶ **Dado numérico** é a quantificação do fato observado.
- ▶ **Moldura** é o conjunto de traços que dão estrutura aos dados numéricos e aos termos necessários à sua compreensão.

Principais elementos das tabelas

- ▶ Tabelas devem ser delimitadas, no alto e embaixo, por traços horizontais.
- ▶ Tabelas **não** devem ser delimitadas, à direita e à esquerda, por traços verticais.
- ▶ O cabeçalho deve ser delimitado por traços horizontais.
- ▶ Para maior clareza, podem ser feitos traços verticais no interior da tabela separando as colunas.
- ▶ Podem ser feitos traços verticais no interior do cabeçalho para separar o que as colunas contém.

Principais elementos das tabelas

Principais elementos das tabelas

- ▶ **Célula** é o espaço na tabela resultante do cruzamento de uma linha com uma coluna.
 - ▶ Nenhuma célula deveria ficar em branco. Quando, por algum motivo, a célula não apresenta um dado numérico, esta não deve ficar em branco. Neste caso, são recomendados os sinais convencionais propostos pelo IBGE, tais como:
 - ▶ “..” Não se aplica dado numérico.
 - ▶ “...” Dado numérico não disponível.
 - ▶ “x” Dado omitido a fim de evitar a individualização da informação.
 - ▶ “0”; “0,0”; “0,00” Dado numérico igual a zero resultante de arredondamento de um dado numérico originalmente positivo.

Principais elementos das tabelas

Número



Título



Tabela 4.2 População residente segundo o sexo. Censo Demográfico 2010

Cabeçalho	Sexo	População residente	Porcentagem
	Homens	93.406.990	48,97%
Indicador de linha	Mulheres	97.348.809	51,03%
	Total	190.755.799	100,00%



Célula

Elementos eventuais

- ▶ **Fonte** identifica a pessoa física ou jurídica responsável pelos dados.
 - ▶ A fonte deve ser colocada na primeira linha do rodapé da tabela, precedida da palavra **Fonte**.
 - ▶ **Não** se indica a fonte nos casos em que os dados foram obtidos pelo pesquisador, ou pelo grupo de pesquisadores, ou pela instituição que apresenta a tabela.
- ▶ **Notas** são informações de natureza geral que servem para esclarecer o conteúdo das tabelas ou para explicar o método utilizado no levantamento dos dados.
 - ▶ As notas devem ser colocadas no rodapé da tabela, logo após a fonte e numeradas, se for o caso.

Elementos eventuais

Tabela 4.3 População brasileira presente, com 15 anos ou mais, segundo a alfabetização. Censo Demográfico 1950

Alfabetização	População presente
Sabem ler e escrever	14.916.779
Não sabem ler e escrever	15.272.632
Sem declaração	60.012
Total	30.249.423

Fonte: IBGE. Anuário Estatístico do Brasil. Rio de Janeiro: IBGE, 1977. v. 38.

Nota: Exclusive 31.960 pessoas recenseadas nos Estados de Minas Gerais (10.461), São Paulo (7.588) e Paraná (13.911), cujas declarações não foram apuradas por extravio de material de coleta.

Um minuto de história

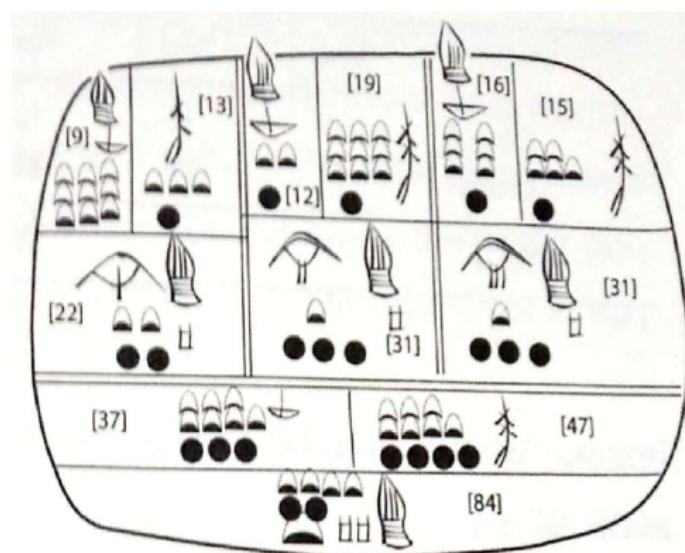
Um minuto de história

- ▶ No livro “**The Seven Pillars of Statistical Wisdom**”, Stephen M. Stigler apresenta **sete ideias básicas de estatística** que foram revolucionárias quando introduzidas, e que permanecem como um avanço conceitual importante e profundo.
 - ▶ Segundo o autor, estas ideias básicas articulam o **núcleo intelectual central do raciocínio estatístico**.
- ▶ A primeira das sete ideias básicas apresentada no livro é a **agregação**, ou seja, a **compactação/redução** dos dados para gerar informação. O que o autor nos mostra é que este princípio (o resumo estatístico) é tão antigo quanto a escrita.

Um minuto de história

- ▶ A seguir são apresentadas uma **tabuleta de argila suméria datada de cerca de 3000 aC**, uma reconstrução de uma destas tabuletas com números modernos.
 - ▶ A tabuleta apresenta contagens (frequências) de dois tipos de mercadoria, possivelmente o rendimento de duas safras em três anos.

Um minuto de história



Um minuto de história

- Versão moderna da tabuleta de argila suméria.

	Year 1	Year 2	Year 3	Total
Crop A	9	12	16	37
Crop B	13	19	15	47
Total	22	31	31	84

Para casa

1. Resolver os exercícios do Capítulo 4.6 do livro **Fundamentos de Estatística**² (disponível no Sabi+).
2. Para o seu levantamento estatístico, monte um pequeno relatório de pesquisa apresentando tabelas de resultados. Utilize os passos de construção de tabelas apresentados nesta aula. Compartilhe no Fórum Geral do Moodle.

²Vieira, S. **Fundamentos de Estatística**, Atlas, 2019, pg. 39-40.

Próxima aula

- ▶ Avaliação I.

Por hoje é só!

Bons estudos!

