

LCE0216

Introdução à Bioestatística Florestal

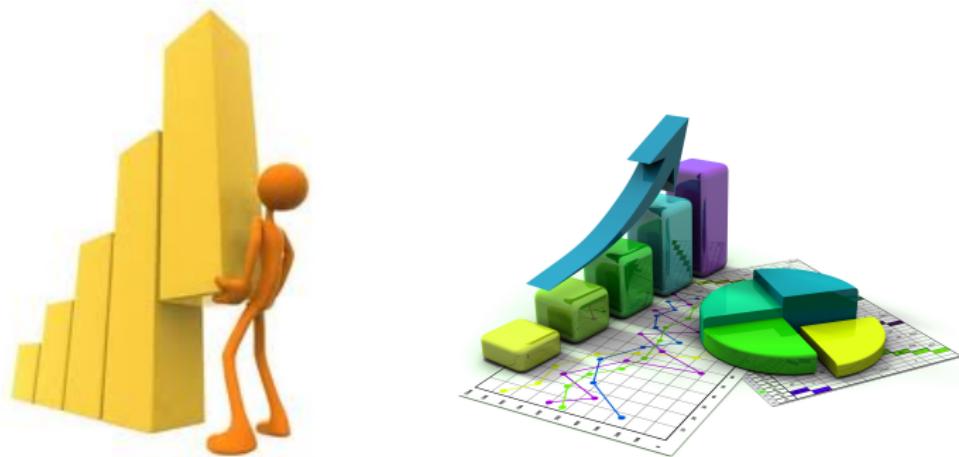
1. Estatística Descritiva

Profa. Dra. Clarice Garcia Borges Demétrio

Escola Superior de Agricultura “Luiz de Queiroz”
Universidade de São Paulo

Piracicaba, 27 de fevereiro de 2018

Estatística é a ciência que fornece os princípios e os métodos para coleta, organização, resumo, análise e interpretação dos dados.



População é um conjunto de indivíduos (ou unidades) com pelo menos uma característica em comum, sobre os quais desejamos obter informações.

População é um conjunto de indivíduos (ou unidades) com pelo menos uma característica em comum, sobre os quais desejamos obter informações.

Amostra é um subconjunto de uma população.

População é um conjunto de indivíduos (ou unidades) com pelo menos uma característica em comum, sobre os quais desejamos obter informações.

Amostra é um subconjunto de uma população.



Dados e variáveis

Variável é uma condição ou característica das unidades da população. A variável pode assumir valores diferentes em diferentes unidades.

Dados e variáveis

Variável é uma condição ou característica das unidades da população. A variável pode assumir valores diferentes em diferentes unidades.

Dados são os valores da variável em estudo, obtidos por meio de uma amostra.

Variáveis: Exemplos

Denotamos com letra maiúscula a variável de interesse. E.g.:

- ▶ X = número de árvores com cancro em uma amostra de 10 árvores.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Variáveis: Exemplos

Denotamos com letra maiúscula a variável de interesse. E.g.:

- ▶ X = número de árvores com cancro em uma amostra de 10 árvores.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

- ▶ M = taxa de mortalidade das árvores em uma floresta nativa.

$$M \in [0, 1]$$

Variáveis: Exemplos

Denotamos com letra maiúscula a variável de interesse. E.g.:

- ▶ X = número de árvores com cancro em uma amostra de 10 árvores.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

- ▶ M = taxa de mortalidade das árvores em uma floresta nativa.

$$M \in [0, 1]$$

- ▶ A = ângulo de dispersão das sementes de ipê-roxo a partir de uma árvore matriz.

$$A \in [0, 2\pi]$$

Variáveis: Exemplos

Denotamos com letra maiúscula a variável de interesse. E.g.:

- ▶ X = número de árvores com cancro em uma amostra de 10 árvores.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

- ▶ M = taxa de mortalidade das árvores em uma floresta nativa.

$$M \in [0, 1]$$

- ▶ A = ângulo de dispersão das sementes de ipê-roxo a partir de uma árvore matriz.

$$A \in [0, 2\pi]$$

- ▶ E = espécie arbórea numa floresta.

$$E \in \{\text{Castanheira, Seringueira, Mogno, ...}\}$$

Variáveis: Exemplos

Denotamos com letra maiúscula a variável de interesse. E.g.:

- ▶ X = número de árvores com cancro em uma amostra de 10 árvores.

$$X \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

- ▶ M = taxa de mortalidade das árvores em uma floresta nativa.

$$M \in [0, 1]$$

- ▶ A = ângulo de dispersão das sementes de ipê-roxo a partir de uma árvore matriz.

$$A \in [0, 2\pi]$$

- ▶ E = espécie arbórea numa floresta.

$$E \in \{\text{Castanheira, Seringueira, Mogno, ...}\}$$

- ▶ C = conceito atribuído a um produto.

$$C \in \{\text{Bom, Médio, Ruim}\}$$

Tipos de variáveis

As variáveis são classificadas em dois tipos:

- ▶ **quantitativas ou numéricas;**
- ▶ **qualitativas ou categorizadas.**

Variáveis quantitativas

As **variáveis quantitativas** se referem a quantidades ou a medições. Desse modo, elas podem ser classificadas em:

Variáveis quantitativas

As **variáveis quantitativas** se referem a quantidades ou a medições. Desse modo, elas podem ser classificadas em:

- ▶ **discretas**: quando assumem apenas valores inteiros.
Exemplo: número de árvores por hectare.

Variáveis quantitativas

As **variáveis quantitativas** se referem a quantidades ou a medições. Desse modo, elas podem ser classificadas em:

- ▶ **discretas**: quando assumem apenas valores inteiros.
Exemplo: número de árvores por hectare.
- ▶ **contínuas**: quando assumem valores em intervalos ou na reta real.
Exemplo: diâmetro de árvores numa floresta plantada.

Variáveis qualitativas

As **variáveis qualitativas** referem-se a qualidades. As variáveis categorizadas podem ser classificadas como:

Variáveis qualitativas

As **variáveis qualitativas** referem-se a qualidades. As variáveis categorizadas podem ser classificadas como:

- ▶ **nominais:** as categorias são independentes.
Exemplo: procedência de uma espécie.

Variáveis qualitativas

As **variáveis qualitativas** referem-se a qualidades. As variáveis categorizadas podem ser classificadas como:

- ▶ **nominais:** as categorias são independentes.
Exemplo: procedência de uma espécie.

- ▶ **ordinais:** as categorias são dependentes.
Exemplo: grau de infestação de cupim em uma estante.

Variáveis

i Categorizadas (Qualitativas)

- ▶ Nominais
- ▶ Ordinais

ii Numéricas (Quantitativas)

- ▶ Discretas
- ▶ Contínuas

Variáveis Multidimensionais

Variáveis multidimensionais resultam da combinação de duas ou mais variáveis. Exemplos:

Variáveis Multidimensionais

Variáveis multidimensionais resultam da combinação de duas ou mais variáveis. Exemplos:

- ▶ Localização de uma árvore em um fragmento florestal de 500×1000 metros. A localização é dada por $S = (X, Y)$, em que X e Y são variáveis quantitativas que representam as coordenadas em um plano cartesiano.
 $X \in [0; 500]; Y \in [0; 1000]$

Variáveis Multidimensionais

Variáveis multidimensionais resultam da combinação de duas ou mais variáveis. Exemplos:

- ▶ Localização de uma árvore em um fragmento florestal de 500×1000 metros. A localização é dada por $S = (X, Y)$, em que X e Y são variáveis quantitativas que representam as coordenadas em um plano cartesiano.
 $X \in [0; 500]; Y \in [0; 1000]$
- ▶ Atributos relacionados a cada uma das árvores em uma amostra. $A = (D, H, S)$, em que
 D é o diâmetro da árvore (variável quantitativa contínua),
 H é a altura (variável quantitativa contínua) e
 S é a espécie (variável qualitativa nominal).

A estatística se divide em duas grandes áreas:

- i a descriptiva, a qual descreve os dados em forma de tabelas, gráficos e medidas-resumo, e ,
- ii a inferencial, que objetiva obter conclusões sobre uma população baseado em amostras.

O enfoque nesta etapa será dado no item (i).

1. Variáveis Quantitativas Discretas

Como o número de valores repetidos em uma variável discreta pode ser grande, uma representação tabular para este tipo de variável é considerar os valores que a variável assume e o número de vezes que cada valor aparece (**frequência**). Temos então, o que chamamos de uma **tabela de frequências**.

Representação Tabular de Variáveis Discretas

Esquematicamente:

Tabela: Título da Tabela

Valores da variável	Frequências
x_1	f_1
x_2	f_2
\vdots	\vdots
x_k	f_k
$\sum_{i=1}^k f_i = n$	

- **Frequência relativa:** $f'_i = \frac{f_i}{n}$
- **Frequência acumulada:** F_i
- **Frequência relativa acumulada:** F'_i

Representação Tabular de Variáveis Discretas

Exemplo: Considere os dados a seguir referentes ao número de brotos deixados em cepas de *Eucalyptus grandis* após o primeiro corte.

2	1	2	2	0	3	3	2	1	1	2	2	0	1	1	3	1	2	1	1	1
2	0	0	3	2	1	2	2	3	0	2	3	3	0	3	2	2	0	1	1	1

Representação Tabular de Variáveis Discretas

Exemplo: Considere os dados a seguir referentes ao número de brotos deixados em cepas de *Eucalyptus grandis* após o primeiro corte.

2	1	2	2	0	3	3	2	1	1	2	2	0	1	1	3	1	2	2	1	1
2	0	0	3	2	1	2	2	3	0	2	3	3	0	3	2	2	0	1	1	

Tabela: Distribuição de frequências do número de brotos deixados em cepas de *Eucalyptus grandis* após o primeiro corte

Número de brotos	f_i	f'_i	F_i	F'_i
0				
1				
2				
3				
Total				

Representação Tabular - Variáveis Discretas

Tabela: Distribuição de frequências do número de brotos deixados em cepas de *Eucalyptus grandis* após o primeiro corte.

Número de brotos	f_i	f'_i	F_i	F'_i
0	7	0,175	7	0,175
1	11	0,275	18	0,450
2	14	0,350	32	0,800
3	8	0,200	40	1,000
Total	40	1,000		

Representação gráfica - Variáveis Discretas

A representação gráfica de uma variável quantitativa discreta cujos valores estão dispostos em uma tabela de frequências é um **diagrama de barras**.

Representação gráfica - Variáveis Discretas

A representação gráfica de uma variável quantitativa discreta cujos valores estão dispostos em uma tabela de frequências é um **diagrama de barras**.

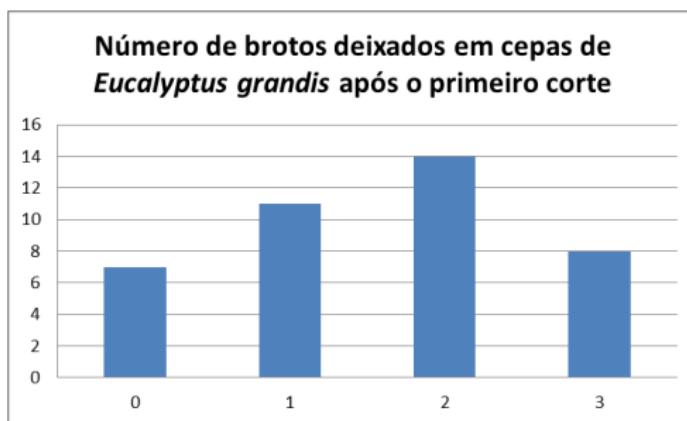


Figura: Representação gráfica de variáveis quantitativas discretas.

Importância da representação gráfica

O objetivo da representação gráfica é observar o **comportamento** da variável de interesse.

Alguns comportamentos podem ser:

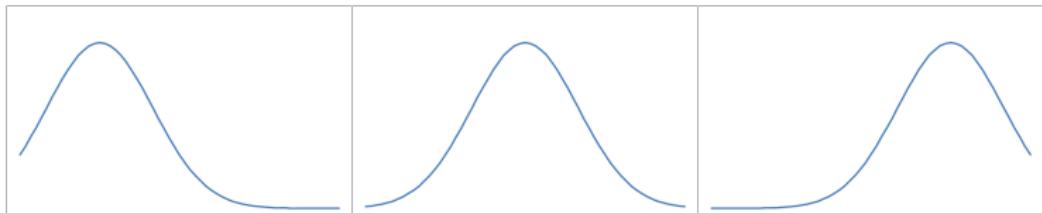


Figura: Da esquerda para a direita tem-se: assimétrica à direita, simétrica, assimétrica à esquerda.

2. Variáveis Quantitativas Contínuas

Devido a natureza da variável, considera-se o intervalo de variação dos dados (**amplitude**) de forma a dividí-los em intervalos menores chamados **intervalos de classe**. Para a construção de uma tabela em classes de frequências deve-se calcular:

Número de intervalos classes (K)

A escolha do número de classes é arbitrária, existindo, no entanto, duas regras que podem ser utilizadas:

- ▶ $K = \sqrt{n}$
- ▶ Fórmula de Sturges:

$$K = 1 + \log_2 n = 1 + 3,32 \log_{10} n.$$

n corresponde ao número de dados.

Representação Tabular e Gráfica

Amplitude total (A)

Limite inferior: $X_{\text{mínimo}}$

Limite superior: $X_{\text{máximo}}$

Amplitude total: $A = X_{\text{máximo}} - X_{\text{mínimo}}$

Amplitude das classes (h)

$$h = \frac{A}{K}.$$

Assim, uma representação tabular para este tipo de variável é considerar os intervalos de classe e contar quantos valores pertencem aos intervalos, obtendo as respectivas frequências.

Representação Tabular - Variáveis Contínuas

Esquematicamente:

Tabela: Título da Tabela

Classes	Frequências
$x_1 \vdash x_2$	f_1
$x_2 \vdash x_3$	f_2
\vdots	\vdots
$x_{k-1} \vdash x_k$	f_k
$\sum_{i=1}^k f_i = n$	

Analogamente ao caso discreto, pode-se definir a **frequência relativa** como $f'_i = \frac{f_i}{n}$, bem como as frequências acumulada (F_i) e relativa acumulada (F'_i).

Representação Tabular - Variáveis Contínuas

Exemplo: Construir a tabela de distribuição de frequências para a variável diâmetro das árvores em uma floresta.

Dados observados

10,2	10,3	11,6	12,0	12,6	12,6	13,0	13,2	13,5	14,0
15,2	15,8	16,2	18,1	18,3	18,4	18,7	19,6	19,9	20,3
20,3	21,9	22,4	23,5	24,6	24,9	31,7	33,1	40,0	40,7
48,3	50,0	50,8	52,4	53,2	61,0	63,2	72,4	78,8	92,5

Representação Tabular - Variáveis Contínuas

Tabela: Distribuição de frequências para diâmetro das árvores em uma floresta

Diâmetro	f_i	f'_i	F_i	F'_i
10,2 – 22,0	22	0,550	22	0,550
22,0 – 33,8	6	0,150	28	0,700
33,8 – 45,6	2	0,050	30	0,750
45,6 – 57,4	5	0,125	35	0,875
57,4 – 69,2	2	0,050	37	0,925
69,2 – 81,0	2	0,050	39	0,975
81,0 – 92,8	1	0,025	40	1,000
Total	40	1,00		

2. Variáveis Quantitativas Contínuas

- ▶ Gráfico de pontos
- ▶ Diagrama de ramos e folhas
- ▶ Histograma e polígono de frequência
- ▶ Ogiva

2. Variáveis Quantitativas Contínuas

Gráfico de pontos

- ▶ Para pequenos conjuntos de dados
- ▶ Fornece uma boa ideia da dispersão dos dados e da existência de dados discrepantes

Representação Gráfica de Variáveis Quantitativas

Gráfico de pontos

Considere os valores observados para a altura (m) dos alunos da disciplina LCE0216 em 2017:

1,69	1,78	1,73	1,56	1,59	1,75	1,72	1,70
1,65	1,70	1,64	1,65	1,67	1,80	1,90	1,72
1,56	1,93	1,75	1,74	1,65	1,80	1,57	1,71

Representação Gráfica - Variáveis Contínuas

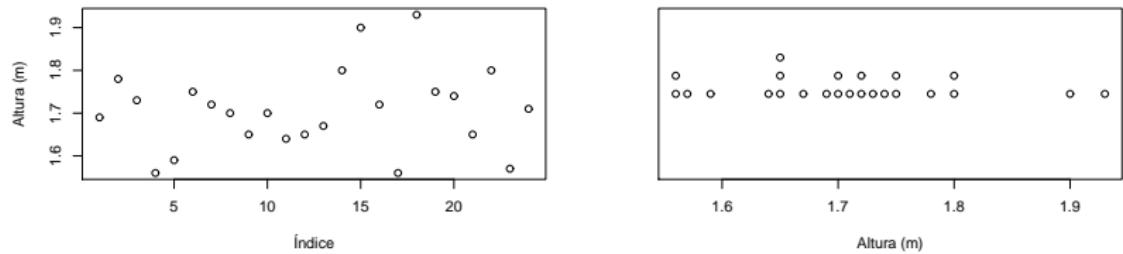


Figura: Gráficos de pontos.

Diagrama de ramos e folhas

- ▶ Pode ser utilizado para grandes conjuntos de dados;
- ▶ Fornece uma boa ideia da distribuição;
- ▶ Não se perde informação sobre os dados;
- ▶ Permite a detecção de valores discrepantes.

Diagrama de ramos e folhas

- ▶ Pode ser utilizado para grandes conjuntos de dados;
- ▶ Fornece uma boa ideia da distribuição;
- ▶ Não se perde informação sobre os dados;
- ▶ Permite a detecção de valores discrepantes.

Ideia básica para a construção: dividir cada observação em duas partes, a primeira corresponde ao ramo, colocado à esquerda de uma linha vertical e a segunda, a folha, colocado à direita.

Representação Gráfica - Variáveis Contínuas

Diagrama de ramos e folhas

Considere os valores observados para a altura (m) dos alunos da disciplina LCE0216 em 2017:

1,56	1,56	1,57	1,59	1,64	1,65	1,65	1,65
1,67	1,69	1,70	1,71	1,72	1,72	1,73	1,74
1,75	1,75	1,76	1,78	1,80	1,80	1,90	1,93

Representação Gráfica - Variáveis Contínuas

Diagrama de ramos e folhas

Considere os valores observados para a altura (m) dos alunos da disciplina LCE0216 em 2017:

1,56	1,56	1,57	1,59	1,64	1,65	1,65	1,65
1,67	1,69	1,70	1,71	1,72	1,72	1,73	1,74
1,75	1,75	1,76	1,78	1,80	1,80	1,90	1,93

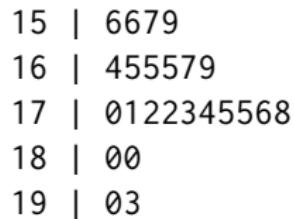


Figura: Diagrama de ramos e folhas para a variável altura dos alunos da disciplina LCE0216 em 2017.

Histograma e polígono de frequências

São utilizados para representar as distribuições de frequências, e fornecem uma boa ideia a respeito da distribuição da variável.

Representação Gráfica de Variáveis Quantitativas

Histograma e polígono de frequências

São utilizados para representar as distribuições de frequências, e fornecem uma boa ideia a respeito da distribuição da variável.

Histograma

É uma sequência de retângulos postos lado a lado onde cada retângulo tem como base a amplitude da classe e como altura a frequência.

Representação Gráfica de Variáveis Quantitativas

Histograma e polígono de frequências

São utilizados para representar as distribuições de frequências, e fornecem uma boa ideia a respeito da distribuição da variável.

Histograma

É uma sequência de retângulos postos lado a lado onde cada retângulo tem como base a amplitude da classe e como altura a frequência.

Polígono de frequência

É um gráfico que se obtém unindo por uma poligonal os pontos correspondentes às frequências das diversas classes, centradas nos respectivos pontos médios. Para se obter as interseções dos polígonos com o eixo, cria-se, em cada extremo do diagrama uma classe de frequência nula.

Representação Gráfica de Variáveis Quantitativas

Histograma

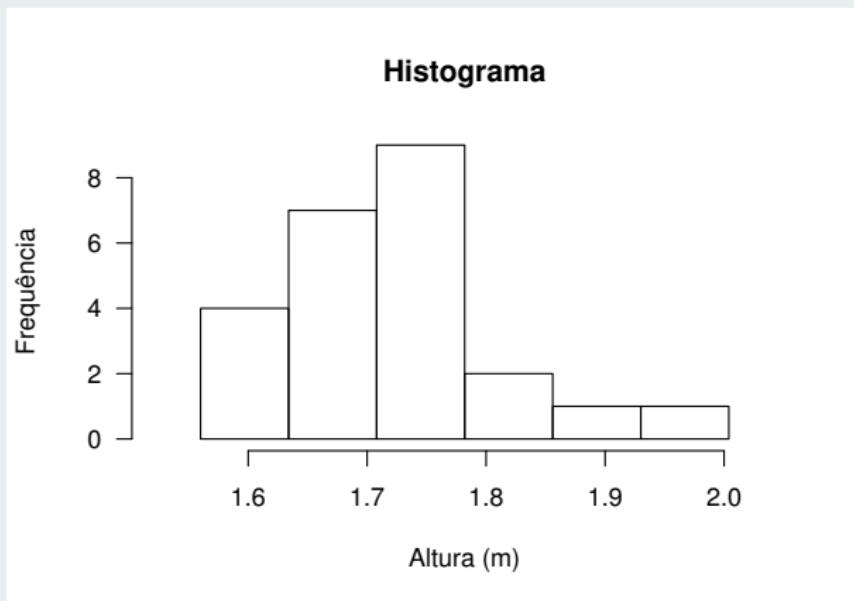


Figura: Histograma para a variável altura.

Representação Gráfica de Variáveis Quantitativas

Histograma e polígono de frequências

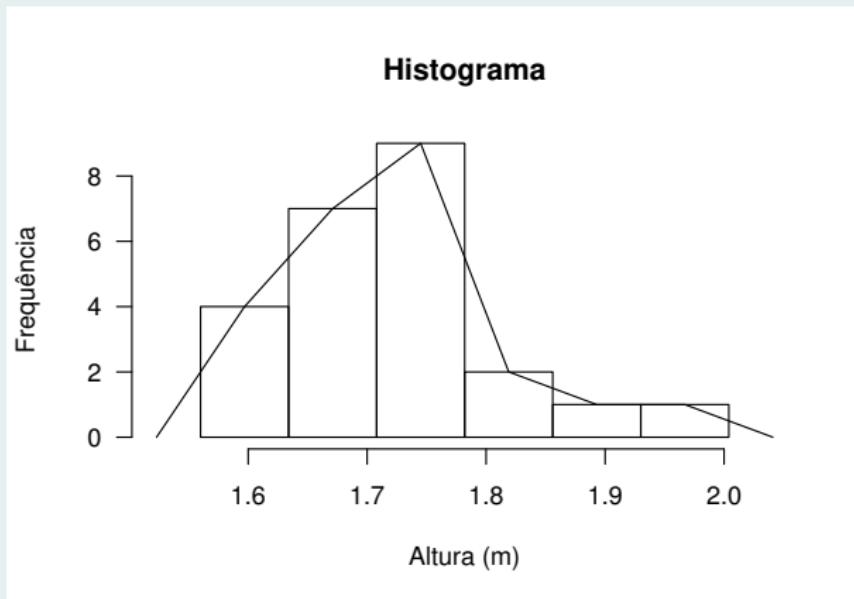


Figura: Histograma e polígono de frequência para a variável altura.

Representação Gráfica de Variáveis Quantitativas

Ogiva de Galton

Gráfico representativo de uma distribuição acumulada de frequências. Trata-se de uma poligonal ascendente formada ligando-se os pontos de coordenadas $(LS_i; F_i)$ ou (LS_i, F'_i) , onde LS_i é o limite superior da classe i .

Representação Gráfica de Variáveis Quantitativas

Ogiva

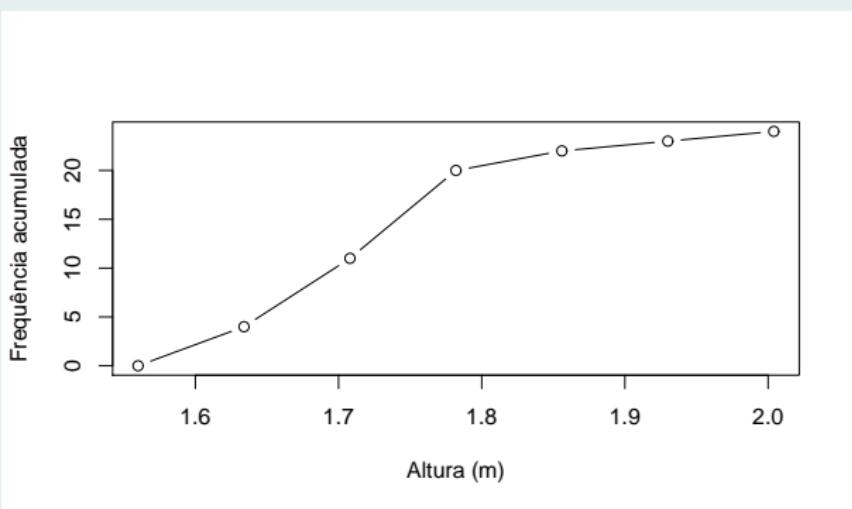


Figura: Ogiva para a variável altura.

Representação Gráfica - Variáveis Contínuas

Exemplo: Construção de um histograma.

Tabela: Distribuição de frequências para diâmetro das árvores em uma floresta

Diâmetro	f_i	f'_i	F_i	F'_i
10,2 \vdash 22,0	22	0,550	22	0,550
22,0 \vdash 33,8	6	0,150	28	0,700
33,8 \vdash 45,6	2	0,050	30	0,750
45,6 \vdash 57,4	5	0,125	35	0,875
57,4 \vdash 69,2	2	0,050	37	0,925
69,2 \vdash 81,0	2	0,050	39	0,975
81,0 \vdash 92,8	1	0,025	40	1,000
Total	40	1,00		

Representação Gráfica - Variáveis Contínuas

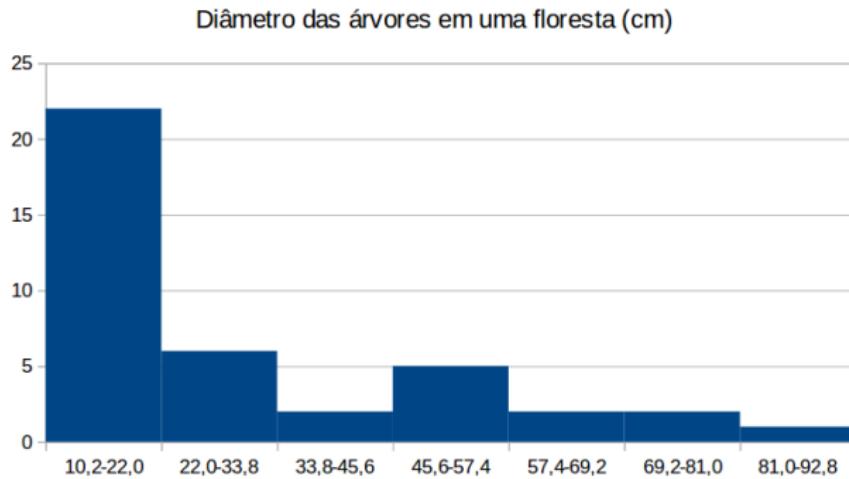


Figura: Histograma.

Representação Tabular e Gráfica

1. Variáveis Qualitativas Nominais (categorias das variáveis são independentes)

Uma representação tabular pode ser feita por meio da **contagem (frequência)** de ocorrências das categorias. A representação gráfica da tabela de frequências deve ser feita por meio de um **diagrama de barras** ou em **gráfico de setores**.

Representação Tabular - Variáveis Qualitativas

Considere que as espécies arbóreas de uma floresta nativa foram classificadas de acordo com a síndrome de regeneração, cuja tabela de frequências é apresentada a seguir.

Síndrome de regeneração	Número de espécies
Heliófitas	4
Oportunistas de clareira	11
Tolerantes	15
Total	30

Representação Gráfica - Variáveis Qualitativas

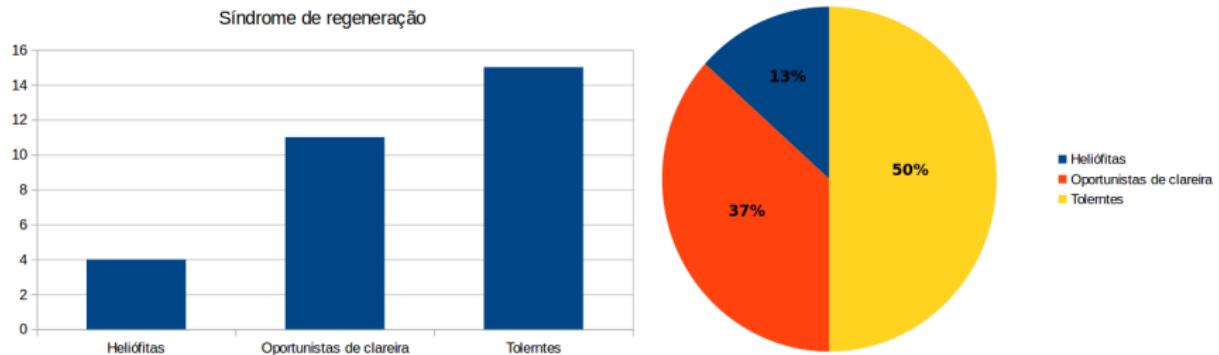


Figura: À esquerda, diagrama de barras. À direita, gráfico de setores.

Representação Tabular e Gráfica

2. Variáveis Qualitativas Ordinais (categorias das variáveis são dependentes)

Neste caso, a construção de tabelas e gráficos são feitas de forma análoga ao caso das variáveis qualitativas nominais (**tabela de frequência e diagramas de barra ou gráfico de setores**).



R é uma linguagem de programação e ambiente para computação estatística e gráfica, disponível sob os termos da licença GNU de software livre. R é a excelência em software para

- ▶ Manipulação de dados;
- ▶ Análise estatística;
- ▶ Visualização gráfica;
- ▶ Pesquisa reproduzível;

Software estatístico: R

Instalação: <https://cran.r-project.org/>

The Comprehensive R Archive Network



Download and Install R
Precompiled binary distributions of the base system and contributed packages. **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms
Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Monday 2016-10-31, Sincere Pumpkin Patch) [R-3.3.2.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

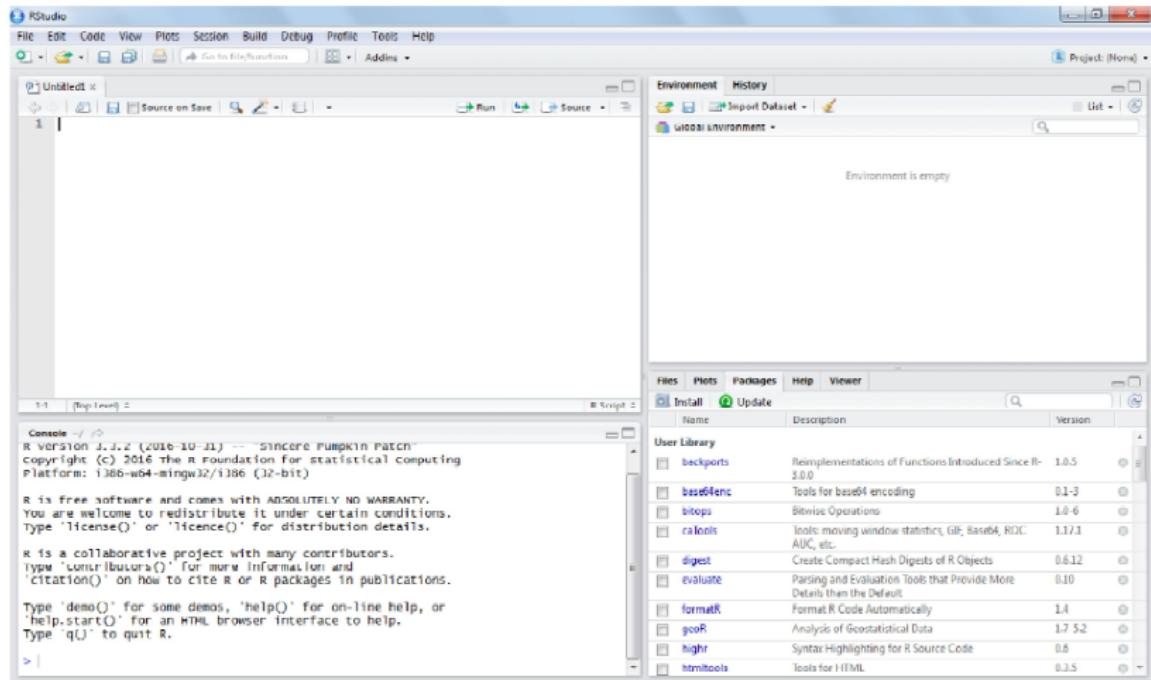
What are R and CRAN?

Ambiente de desenvolvimento RStudio

Instalação: <https://www.rstudio.com/>

The screenshot shows the official RStudio website. At the top, there's a navigation bar with links for 'rstudio::conf', 'Products', 'Resources', 'Pricing', 'About Us', 'Blogs', and a search icon. Below the navigation is a large banner featuring the 'RStudio' logo and the text 'Open source and enterprise-ready professional software for R'. To the right of the banner is a sidebar with four buttons: 'Download RStudio', 'Discover Shiny', 'shinyapps.io Login', and 'Discover RStudio Connect'. At the bottom of the page, there are three sections: a screenshot of the RStudio IDE interface showing the 'Environment', 'Plots', and 'Console' panes; a map of the United States with a 'ZIP explorer' overlay; and a row of six hexagonal icons representing R packages: 'markdown' (red), 'knitr' (red), 'Shiny' (blue), 'ggplot2' (black), 'gridExtra' (grey), and 'tidyverse' (orange).

Ambiente de desenvolvimento RStudio



Inserindo os dados no software

Digitando diretamente

```
altura <- c(10.3, 20.3, 33.4) # Se for quantitativa  
cor <- c("amarela", "azul")    # Se for qualitativa
```

Software estatístico: R

Inserindo os dados no software

Digitando diretamente

```
altura <- c(10.3, 20.3, 33.4) # Se for quantitativa  
cor <- c("amarela", "azul") # Se for qualitativa
```

Importando um conjunto de dados

```
dados <- read.csv(file = "//.../ex1.csv",  
                  header = TRUE, dec = ".", sep = ",")
```

Software estatístico: R

Exemplo: Classificação de espécies arbóreas de acordo com a síndrome de regeneração

OP	H	OP
H	OP	T
T	T	T
OP	T	T
T	T	T
OP	OP	OP
T	H	OP
T	T	OP
T	T	OP
OP	T	H

- ▶ OP: Oportunista de clareira;
- ▶ H: Heliófita;
- ▶ T: Tolerantes.

Códigos

```
# Lendo os dados
dados <- read.csv("dados/sindrome.csv", header = FALSE)
str(dados)

## 'data.frame': 30 obs. of 1 variable:
## $ V1: Factor w/ 3 levels "H","OP","T": 2 1 2 1 2 3 3 3 3 2 ...

# Tabela de frequências
freq <- table(dados)
freq

## dados
## H OP T
## 4 11 15
```

Códigos

```
# Configura janela gráfica
par(mfrow = c(1, 2), mar = c(4, 4, 2, 2))

# Gráfico de barras
barplot(freq, col = "green",
        ylab = "Número de espécies",
        xlab = "Síndrome de regeneração")

# Gráfico de setores
porcentagem <- round(100 * freq / sum(freq), 0)
pie(freq, labels = porcentagem,
    col = c("red", "green", "black"))
legend("topright", c("H", "OP", "T"),
       fill = c("red", "green", "black"))
```

Output

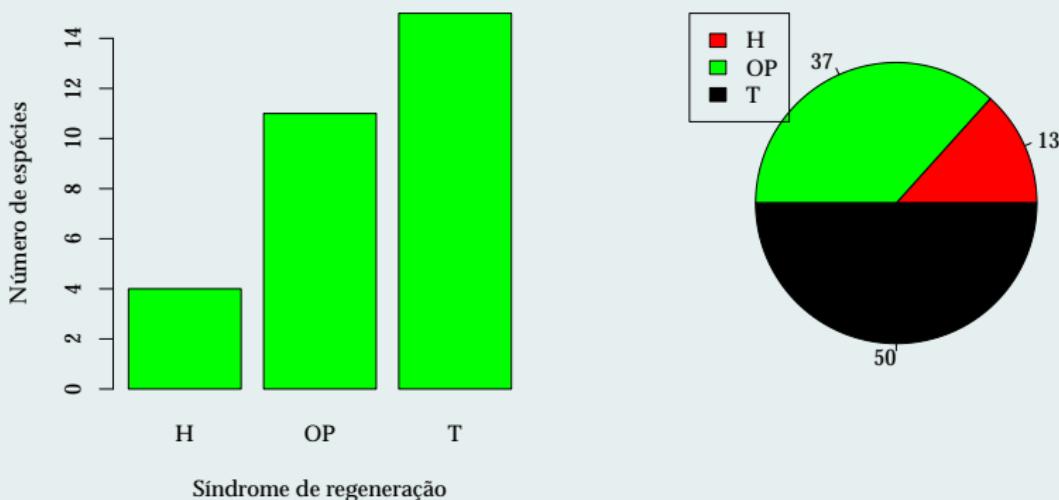


Figura: À esquerda, diagrama de barras. À direita, gráfico de setores para a variável síndrome de regeneração.

Software estatístico: R

Exemplo: Valores observados para a altura (m) dos alunos da disciplina LCE0216:

1,56	1,56	1,57	1,59	1,64	1,65	1,65	1,65
1,67	1,69	1,70	1,71	1,72	1,72	1,73	1,74
1,75	1,75	1,76	1,78	1,80	1,80	1,90	1,93

No R

```
# Inserindo as alturas dos alunos de 2017
altura <- c(1.69, 1.78, 1.73, 1.56, 1.59, 1.75, 1.72, 1.76, 1.65, 1.70,
           1.64, 1.65, 1.67, 1.80, 1.80, 1.90, 1.72, 1.56, 1.93, 1.75, 1.74,
           1.65, 1.80, 1.57, 1.71)
```

Códigos

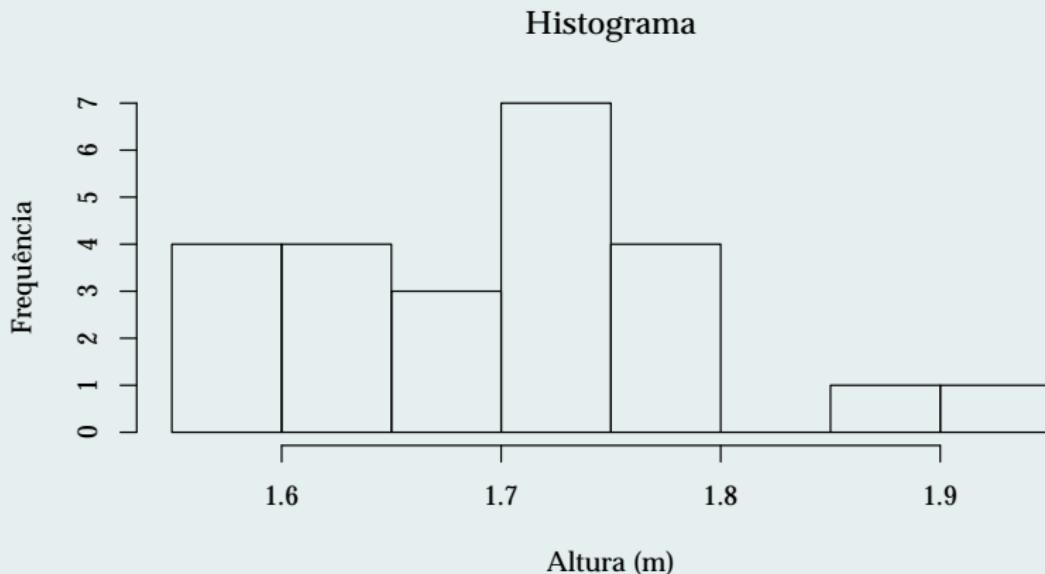
```
# Representação em ramo e folhas
stem(altura)

##
##      The decimal point is 1 digit(s) to the left of the |
##
##      15 | 6679
##      16 | 455579
##      17 | 0122345568
##      18 | 00
##      19 | 03
```

Software estatístico: R

Códigos

```
# Histogram  
hist(altura, xlab = "Altura (m)", ylab = "Frequência", main = "Histograma")
```



Software estatístico: R

Códigos

```
# Gráfico de frequência em pontos  
stripchart(altura, method = "stack", offset = 1,  
           xlab = "Altura (m)", pch = 1)
```

