

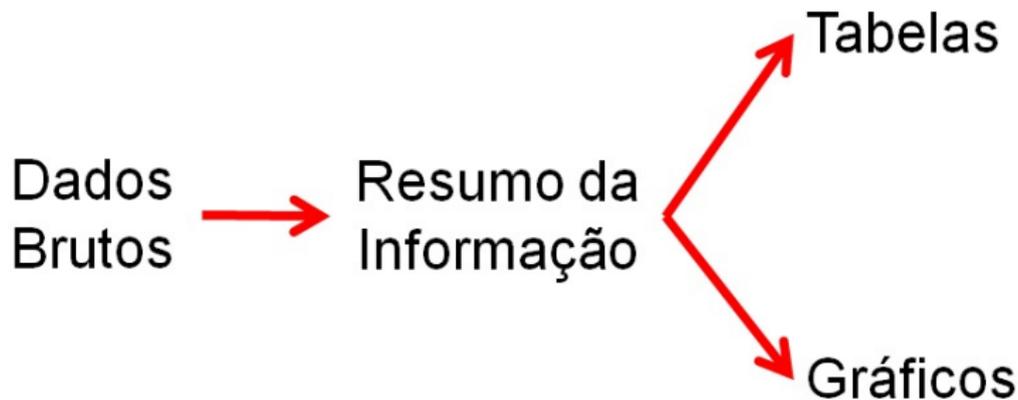
Escola Superior de Agricultura “Luiz de Queiroz” Universidade de São Paulo

Estatística Descritiva

Professora Renata Alcarde Sermarini

Piracicaba
março 2017

Descrição tabular e gráfica



Dados brutos

são dados na forma como foram obtidos no campo.

Descrição tabular

Uma tabela deve conter:

- Título
- Corpo
- Cabeçalho
- Coluna indicadora

O **título** explica o **conteúdo** da tabela

O corpo da tabela é composto pelos dados, organizados em linhas e colunas que se cruzam

Célula é a interseção de uma linha com uma coluna

Linha é uma série de células organizadas **horizontalmente**

Coluna é uma série de células organizadas **verticalmente**

O **cabeçalho** especifica o conteúdo das **colunas**

A **coluna indicadora** especifica o conteúdo das **linhas**

Descrição tabular

Tabela: Título da tabela

rótulo da 1ª col	rótulo da 2ª col
...	...

Descrição tabular

Tabela de distribuição de frequências

é constituída, basicamente, de duas colunas, uma contendo os possíveis valores que a variável assume e outra contendo o número de vezes (frequência) que cada um desses valores ocorre.

Exemplo: Conceito em relação à qualidade da comida do Rucas

Tabela: Distribuição de frequências da variável conceito em relação à qualidade da comida do Rucas

Conceito (x_i)	frequência (f_i)
Ruim	
Razoável	
Boa	
Ótima	
Total	

Descrição tabular

Nas tabelas de distribuição de frequências, é usual fornecer, além das frequências, as frequências relativas.

$$\text{Frequência relativa} = \frac{\text{Frequência}}{\text{Total}}.$$

Tabela: Distribuição de frequências da variável conceito em relação à qualidade da comida do Rucas

Conceito (x_i)	frequência (f_i)	freq relativa (f'_i)
Ruim		
Razoável		
Boa		
Ótima		
Total		

Descrição tabular

Perguntas:

- Qual o conceito mais frequente?
- Qual é a porcentagem dos alunos que acham a comida ruim ou razoável?
- Qual é a porcentagem dos alunos que acham a comida pelo menos boa?

Tabela: Distribuição de frequências da variável conceito em relação à qualidade da comida do Rucas

Conceito (x_i)	frequência (f_i)	freq relativa (f'_i)	freq acumulada (F_i)	freq acum rel (F'_i)
Ruim				
Razoável				
Boa				
Ótima				
Total				

Descrição tabular

Exercício: 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
2	0	0	3	2	1	2	2	3	0	2	3	3	0	3	2	2	0	1	1

Tabela: Distribuição de frequências para a variável 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				

Número de valores diferentes
muito grande



Agrupamento dos dados em classes

Características:

- as classes devem abranger todas as observações
- o extremo superior de uma classe é o extremo inferior da classe subsequente
- cada valor observado deve estar presente em apenas uma classe

Descrição tabular

Procedimento para a construção de uma tabela em classes de frequências:

Número de 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.

Descrição tabular

Procedimento para a construção de uma tabela em classes de frequências:

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)

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

Descrição tabular

Exemplo:

Tabela: Distribuição de frequências para uma variável contínua

Altura	f_i	f'_i	F_i	F'_i
$x_{min} \vdash x_{min} + h$				
$x_{min} + h \vdash x_{min} + 2h$				
$x_{min} + 2h \vdash x_{min} + 3h$				
...				
Total				

Descrição tabular

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

Rol

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

Descrição gráfica

Variável qualitativa

- Gráfico de barras
- Gráfico de setores
- Diagrama de Pareto

Descrição gráfica

Gráfico de barras

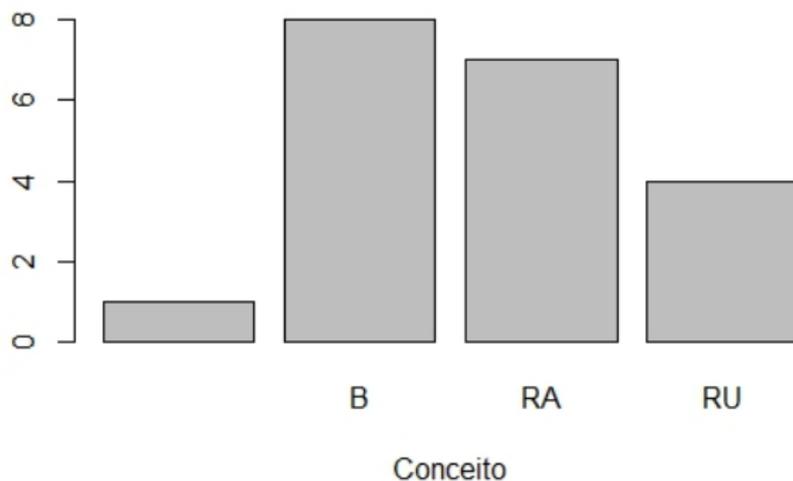


Figura: Gráfico de barras para a variável conceito a respeito da comida do Rucas

Descrição gráfica

Gráfico de barras horizontais

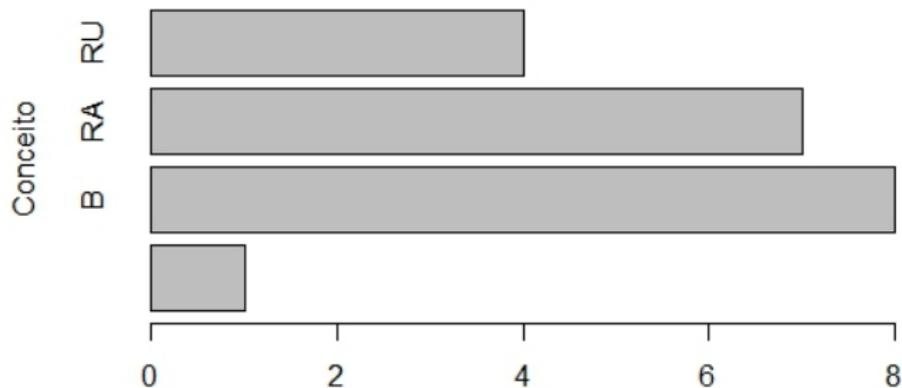


Figura: Gráfico de barras horizontais para a variável conceito a respeito da comida de Bucar

Descrição gráfica

Gráfico de setores

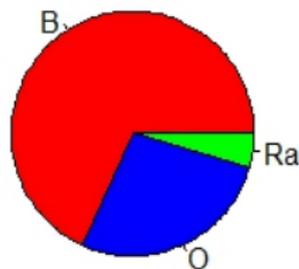


Figura: Gráfico de setores para a variável conceito a respeito da comida do Rucas

Descrição gráfica

Diagrama de Pareto

- Gráfico de Barras ordenadas (da mais alta para a mais baixa);
- Muito utilizado em controle de qualidade (busca por erros mais comuns);
 - Barras mais baixas → erros de menor ocorrência;
 - Ordem em que os erros devem ser sanados, reduzidas as perdas ou solucionadas as reclamações;
- Identificar causas de sucesso.

Descrição gráfica

Diagrama de Pareto: Exemplo

As reclamações no PROCON, em determinado ano, foram enquadradas em algumas áreas, conforme pode ser observado na Tabela 20.

Tabela: Reclamações no PROCON por área em um determinado ano

Área	Frequência	Frequência relativa
Alimentos	15	0,01
Assuntos financeiros	345	0,23
Habitação	120	0,08
Produtos	315	0,21
Saúde	105	0,07
Serviços	600	0,40
Total	1500	1,00

Fonte: Vieira, Sônia; 2012

Descrição gráfica

Diagrama de Pareto: Exemplo

As reclamações no PROCON, em determinado ano, foram enquadradas em algumas áreas, conforme pode ser observado na Tabela 20.

Tabela: Reclamações no PROCON por área em um determinado ano

Área	Frequência	Frequência relativa
Serviços	600	0,40
Assuntos financeiros	345	0,23
Produtos	315	0,21
Habitação	120	0,08
Saúde	105	0,07
Alimentos	15	0,01
Total	1500	1,00

Fonte: Vieira, Sônia; 2012

Descrição gráfica

Diagrama de Pareto

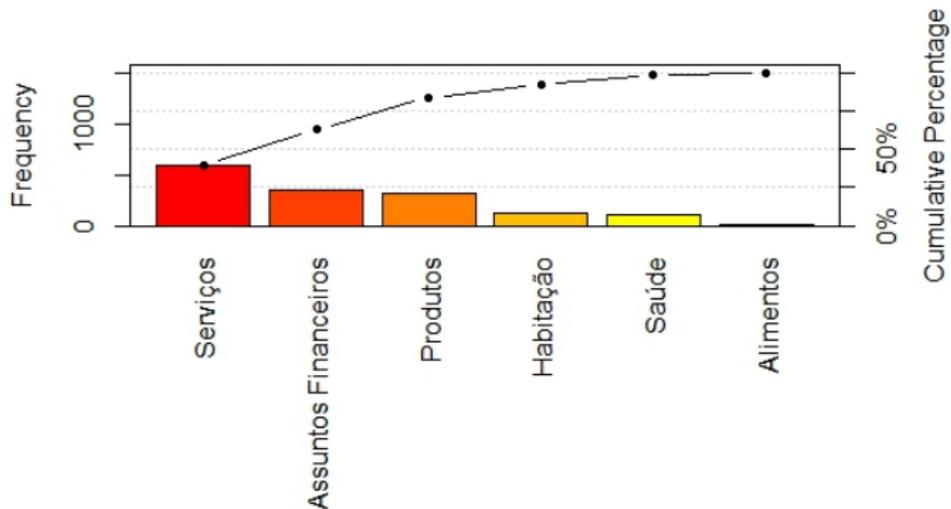


Figura: Reclamações no PROCON por área em determinado ano

Descrição gráfica

Exercício: Construir os gráficos de barra e de setores para a variável Toler.

Descrição gráfica

Variável quantitativa

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

Descrição gráfica

Gráfico de pontos

- para pequenos conjuntos de dados
- fornece uma boa ideia da dispersão dos dados e da existência de dados discrepantes

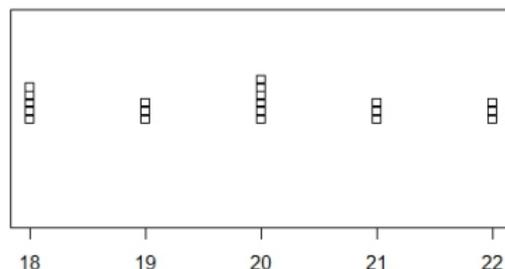


Figura: Gráfico de pontos para a variável Idade

Descrição gráfica

Diagrama de ramos e folhas

- pode ser utilizado para grandes conjuntos de dados;
- fornece uma boa ideia do formato da distribuição dos dados (moda, simetria, ...);
- 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.

Descrição gráfica

Diagrama de ramos e folhas

The decimal point is at the |

```
18 | 00000
19 | 000
20 | 000000
21 | 000
22 | 000
```

Figura: Diagrama de ramos e folhas para a variável Idade

Descrição gráfica

Exercício: Considere os dados apresentados a seguir, correspondentes ao rendimento médio, em kg ha^{-1} , de 32 híbridos de milho. Apresente o diagrama de ramos e folhas para a variável em questão.

3973	4500	4550	4552	4614	4660	4680	4685
4760	4769	4770	4778	4849	4960	4975	4980
4993	5056	5063	5110	5117	5166	5172	5202
5230	5403	5513	5823	5889	6047	6166	6388

Descrição gráfica

Histograma e polígono de frequências

- são utilizados para representar as distribuições de frequências, e
- fornecem uma boa ideia a respeito do formato da distribuição dos dados (simetria, moda, ...)

Histograma

é uma sequência de retângulos postos lado a lado em que 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 obterem as interseções dos polígonos com o eixo, cria-se, em cada extremo do diagrama uma classe de frequência nula.

Descrição gráfica

Histograma e polígono de frequências

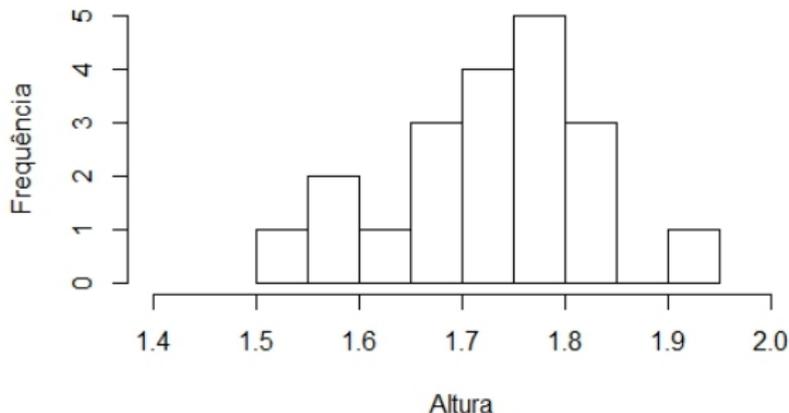


Figura: Histograma para a variável Altura

Descrição gráfica

Histograma e polígono de frequências

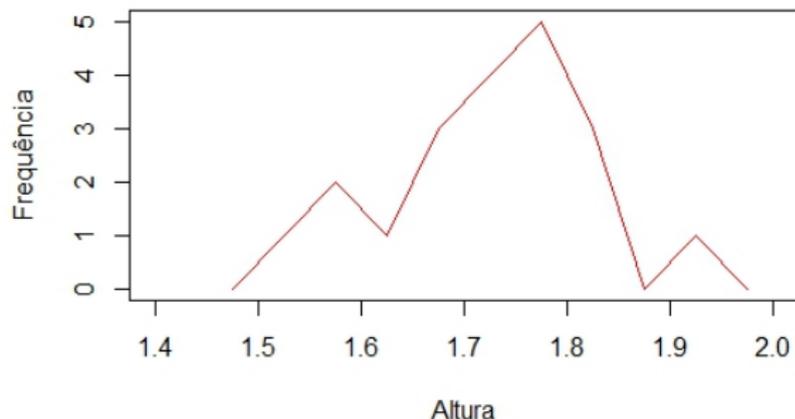


Figura: Polígono de frequências para a variável Altura

Descrição gráfica

Histograma e polígono de frequências

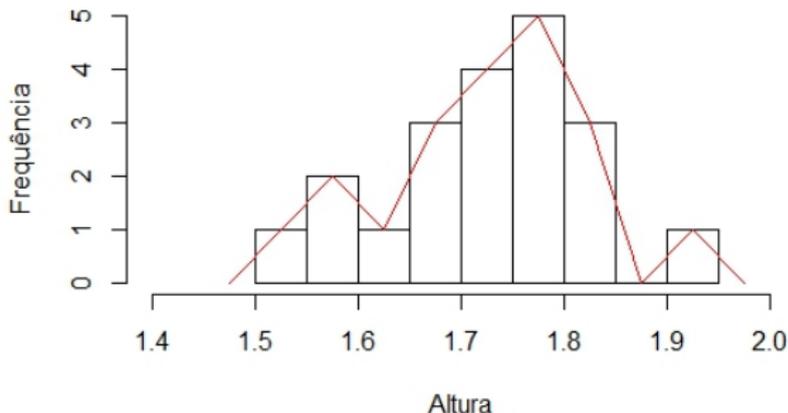


Figura: Histograma e polígono de frequências para a variável Altura

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 $(LS_i; F'_i)$, em que LS_i é o limite superior da classe i .

Descrição gráfica

Exemplo: Construir uma ogiva de Galton para a variável Altura

Descrição gráfica

Exemplo: Construir uma ogiva de Galton para a variável Altura

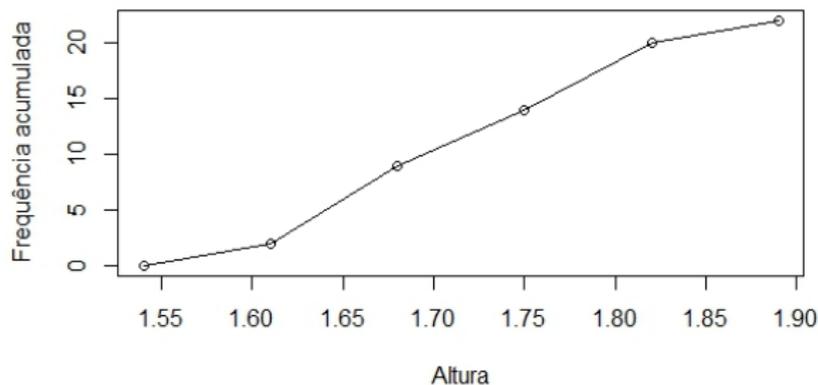


Figura: Ogiva de Galton para a variável Altura

Descrição gráfica

Histograma com amplitudes desiguais de classes

Atenção: o histograma deve ser construído com as densidades de frequências e não com as frequências.

$$\text{densidade}_i = \frac{f_i}{h_i}.$$

Descrição gráfica

Exemplo

Construir um histograma para a distribuição das rendas das pessoas com 10 anos de idade ou mais na região sudeste do Brasil em 1995.

Tabela: Rendimento médio mensal, em número de salários mínimos, de pessoas de 10 anos ou mais de idade na região sudeste do Brasil em 1995

Renda	Frequência (f_i)	Amplitude (h_i)	Densidade de freq (f_i/h_i)	Frequência relativa (f_i')	Densidade de freq rel (f_i'/h_i)
0 † 1/2	1,09				
1/2 † 1	5,62				
1 † 2	7,25				
2 † 3	5,04				
3 † 5	5,55				
5 † 10	5,02				
10 † 20	2,33				
≥ 20	1,25				
Total	33,15				

Descrição gráfica

Exemplo

Construir um histograma para a distribuição das rendas das pessoas com 10 anos de idade ou mais na região sudeste do Brasil em 1995.

Tabela: Rendimento médio mensal, em número de salários mínimos, de pessoas de 10 anos ou mais de idade na região sudeste do Brasil em 1995

Renda	Frequência (f_i)	Amplitude (h_i)	Densidade de freq (f_i/h_i)	Frequência relativa (f'_i)	Densidade de freq rel (f'_i/h_i)
0 † 1/2	1,09	0,5	2,18	0,033	0,066
1/2 † 1	5,62	0,5	11,24	0,170	0,339
1 † 2	7,25	1	7,25	0,219	0,219
2 † 3	5,04	1	5,04	0,152	0,152
3 † 5	5,55	2	2,775	0,167	0,084
5 † 10	5,02	5	1,004	0,151	0,030
10 † 20	2,33	10	0,233	0,070	0,007
≥ 20	1,25	∞	≈ 0	0,038	≈ 0
Total	33,15				

Descrição gráfica

Códigos em R

```
dados<- read.csv2("dados_alunos1a.csv")
dados
str(dados)
attach(dados)
#distribuicao de frequencias
table(OpRucas)
sum(table(OpRucas))
table(OpRucas)/20

#grafico de barras
barplot(table(OpRucas), xlab="Conceito")
barplot(table(OpRucas), xlab="Conceito", col="red",
        border = "blue")
barplot(table(OpRucas), ylab="Conceito", horiz=TRUE)
```

Descrição gráfica

Códigos em R

```
#grafico de setores  
pie(table(OpRucas), col=c(2,3,4,5))
```

```
#outro exemplo  
a<- c(2,1,2,2,0,3,3,2,1,1,2,2,0,1,1,3,1,2,1,1,  
      2,0,0,3,2,1,2,2,3,0,2,3,3,0,3,2,2,0,1,1)  
table(a)  
table(a)/length(a)  
cumsum(table(a))  
cumsum(table(a)/length(a))
```

Descrição gráfica

Códigos em R

```
#grafico de pontos  
stripchart(Idade, method="stack")
```

```
#diagrama de ramos e folhas  
stem(Idade)
```

Descrição gráfica

Códigos em R

```
altura<- c(1.65,1.75,1.75,1.70,1.80,1.66,  
          1.70,1.65,1.60,1.76,1.61,1.68,  
          1.61,1.77,1.82,1.85,1.72,1.65,  
          1.72,1.65,1.54,1.75)  
altura[order(altura)]  
#histograma  
hist(altura)  
hist(altura, breaks=c(1.54, 1.61, 1.68, 1.75, 1.82, 1.89),  
      right=F, freq=T)  
hist(altura, breaks=c(1.54, 1.61, 1.68, 1.75, 1.82, 1.89),  
      right=F, freq=T, main="", ylab="Frequência",  
      xlab="Altura", xlim=c(1.5, 2))
```

Descrição gráfica

Códigos em R

```
diametro<- c(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)
```

```
#histograma
```

```
hist(diametro)
```

```
histo<-hist(diametro, main="", ylab="Frequência",  
            xlab="Diâmetro", xlim=c(4, 99),  
            breaks = c(10.2,22.0,33.8,45.6,  
                       57.4,69.2,81.0,92.8))
```

Descrição gráfica

Códigos em R

```
#poligono de frequencias
plot(c(0,histo$counts,0) ~
      c(10.2-11.8/2,histo$mids,92.8+11.8/2),
      col=2, type="l", xlim=c(4, 99),
      xlab="Diâmetro", ylab="Frequência")

#histograma e poligolo de frequencias
hist(diametro, main="", ylab="Frequência", xlab="Diâmetro",
      breaks = c(10.2,22.0,33.8,45.6,
                 57.4,69.2,81.0,92.8),
      xlim=c(4, 99))

lines(c(0,histo$counts,0) ~
      c(10.2-11.8/2,histo$mids,92.8+11.8/2),
      col=2)
```

Descrição gráfica

Códigos em R

```
#ogiva de Galton  
plot(c(0, cumsum(histo$counts)) ~ histo$breaks,  
type="b", xlab="altura", ylab=expression(F[i]))
```