

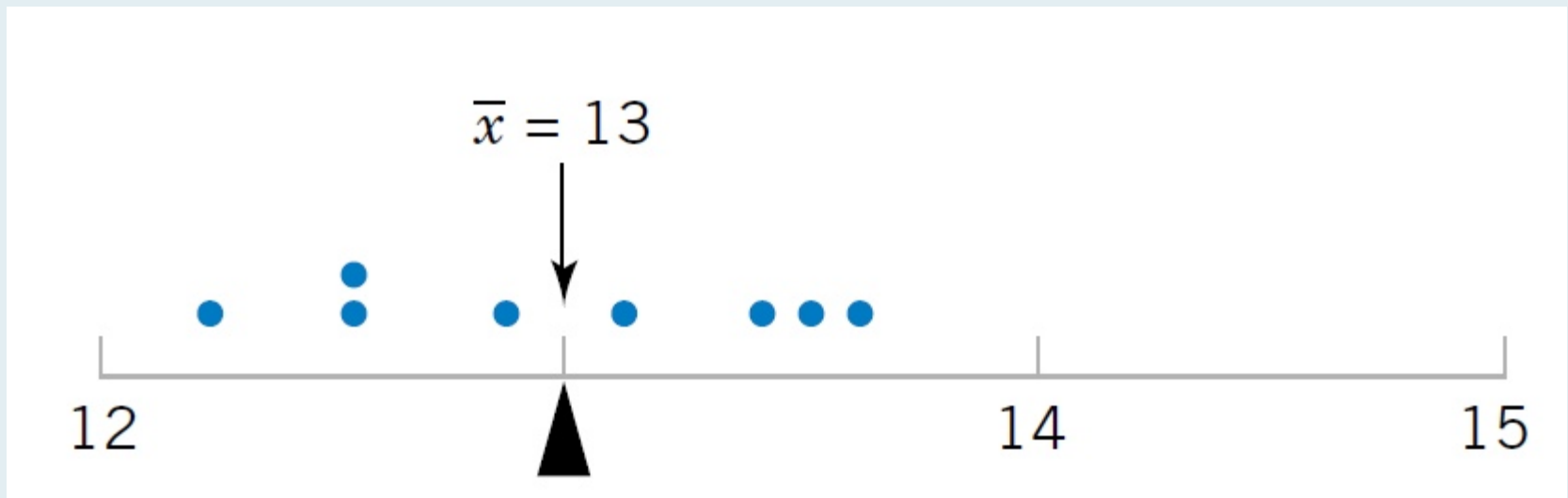
Probabilidade e Estatística
CCT - UDESC

Apresentação de Dados

Fernando Deeke Sasse
Departamento de Matemática
CCT-UDESC

Média amostral

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$



Média populacional

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

Variância amostral

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Fórmula eficiente para cálculo da variância amostral

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^n (x_i^2 + \bar{x}^2 + 2\bar{x}x_i)}{n - 1} = \frac{\sum_{i=1}^n x_i^2 + n\bar{x}^2 - 2\bar{x} \sum_{i=1}^n x_i}{n - 1}$$

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n - 1}$$

Variância Populacional

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

Amostragem aleatória

Amostragem aleatória

População: totalidade das observações de interesse

Amostragem aleatória

População: totalidade das observações de interesse

Distribuição de probabilidade: modelo para uma população

Amostragem aleatória

População: totalidade das observações de interesse

Distribuição de probabilidade: modelo para uma população

Amostra: subconjunto de observações selecionadas de uma população

Amostragem aleatória

População: totalidade das observações de interesse

Distribuição de probabilidade: modelo para uma população

Amostra: subconjunto de observações selecionadas de uma população

Amostra aleatória: resultado de algum mecanismo de chance.

$$X_1, X_2, \dots, X_n$$

As variáveis aleatórias X_1, X_2, \dots, X_n

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

1. Todos os

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

1. Todos os X_i

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

1. Todos os X_i são variáveis aleatórias independentes

As variáveis aleatórias X_1, X_2, \dots, X_n

são uma amostragem aleatória de tamanho n se:

1. Todos os X_i são variáveis aleatórias independentes
2. Todos os

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

1. Todos os X_i são variáveis aleatórias independentes
2. Todos os X_i

As variáveis aleatórias X_1, X_2, \dots, X_n
são uma amostragem aleatória de tamanho n se:

1. Todos os X_i são variáveis aleatórias independentes
2. Todos os X_i têm a mesma distribuição de probabilidade.

As variáveis aleatórias X_1, X_2, \dots, X_n

são uma amostragem aleatória de tamanho n se:

1. Todos os X_i são variáveis aleatórias independentes
2. Todos os X_i têm a mesma distribuição de probabilidade.

Estatística: qualquer função das observações de uma amostragem aleatória

Diagramas Ramo e Folha (Stem-and-leaf)

Capacidade compressiva (psi) de 80 elementos de liga alumínio-lítio

105	221	183	186	121	181	180	143
97	154	153	174	120	168	167	141
245	228	174	199	181	158	176	110
163	131	154	115	160	208	158	133
207	180	190	193	194	133	156	123
134	178	76	167	184	135	229	146
218	157	101	171	165	172	158	169
199	151	142	163	145	171	148	158
160	175	149	87	160	237	150	135
196	201	200	176	150	170	118	149

Pergunta: qual é a porcentagem de elementos correspondentes a valores menores que 120 psi?

Construção de um diagrama Ramo-e-Folha

Divida cada número x_i em duas partes: (a) uma folha

(a) **um ramo** consistindo de um ramo com dois ou mais dígitos iniciais ;

(b) uma folha consistindo dos dígitos remanescentes.

Ramo	Folha	Frequência
7	6	1
8	7	1
9	7	1
10	5 1	2
11	5 8 0	3
12	1 0 3	3
13	4 1 3 5 3 5	6
14	2 9 5 8 3 1 6 9	8
15	4 7 1 3 4 0 8 8 6 8 0 8	12
16	3 0 7 3 0 5 0 8 7 9	10
17	8 5 4 4 1 6 2 1 0 6	10
18	0 3 6 1 4 1 0	7
19	9 6 0 9 3 4	6
20	7 1 0 8	4
21	8	1
22	1 8 9	3
23	7	1
24	5	1

Tamanho dos ramos

Stem	Leaf
6	1 3 4 5 5 6
7	0 1 1 3 5 7 8 8 9
8	1 3 4 4 7 8 8
9	2 3 5

Poucos ramos: pouca informação

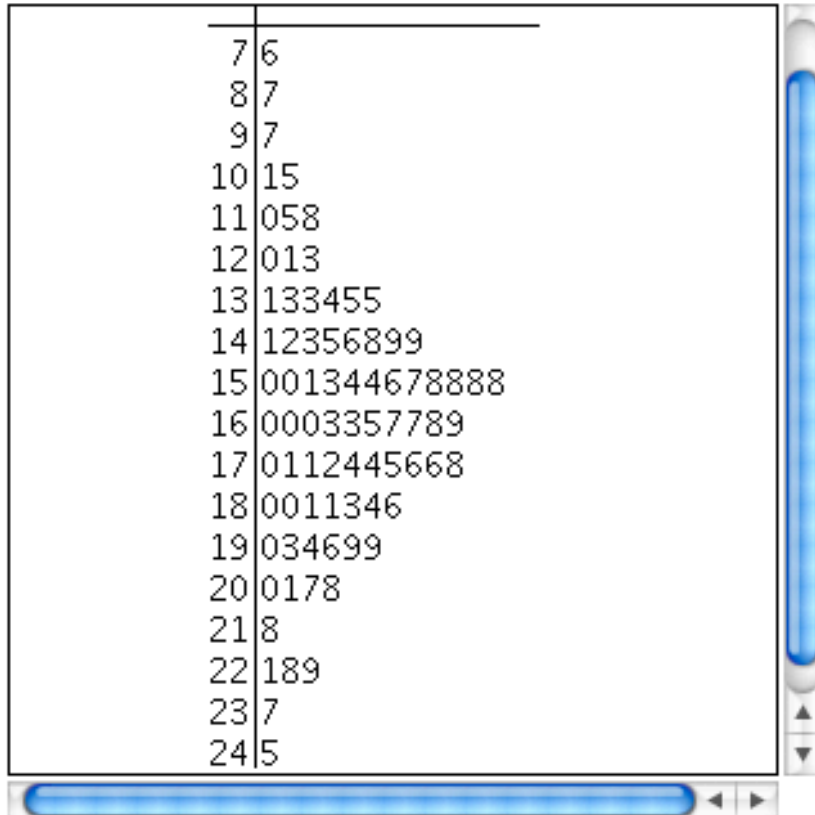
Stem	Leaf
6L	1 3 4
6U	5 5 6
7L	0 1 1 3
7U	5 7 8 8 9
8L	1 3 4 4
8U	7 8 8
9L	2 3
9U	5

Número adequado de ramos

Número excessivo de ramos

Stem	Leaf
6z	1
6t	3
6f	4 5 5
6s	6
6e	
7z	0 1 1
7t	3
7f	5
7s	7
7e	8 8 9
8z	1
8t	3
8f	4 4
8s	7
8e	8 8
9z	
9t	2 3
9f	5
9s	
9e	

Stem and Leaf Plot



© Shodor

Title:

View Guess Difficulty:

Enter data:

Calculate these values:

Mean:

Median:

Mode:

Distribuições de frequência e histogramas

Distribuição de frequência

- Domínio de dados deve ser dividido em intervalos
- Preferencialmente os intervalos devem ser de igual largura para facilitar a visualização
- 5 a 20 intervalos é bom número
- \sqrt{n} onde n é o número de observações, é razoável

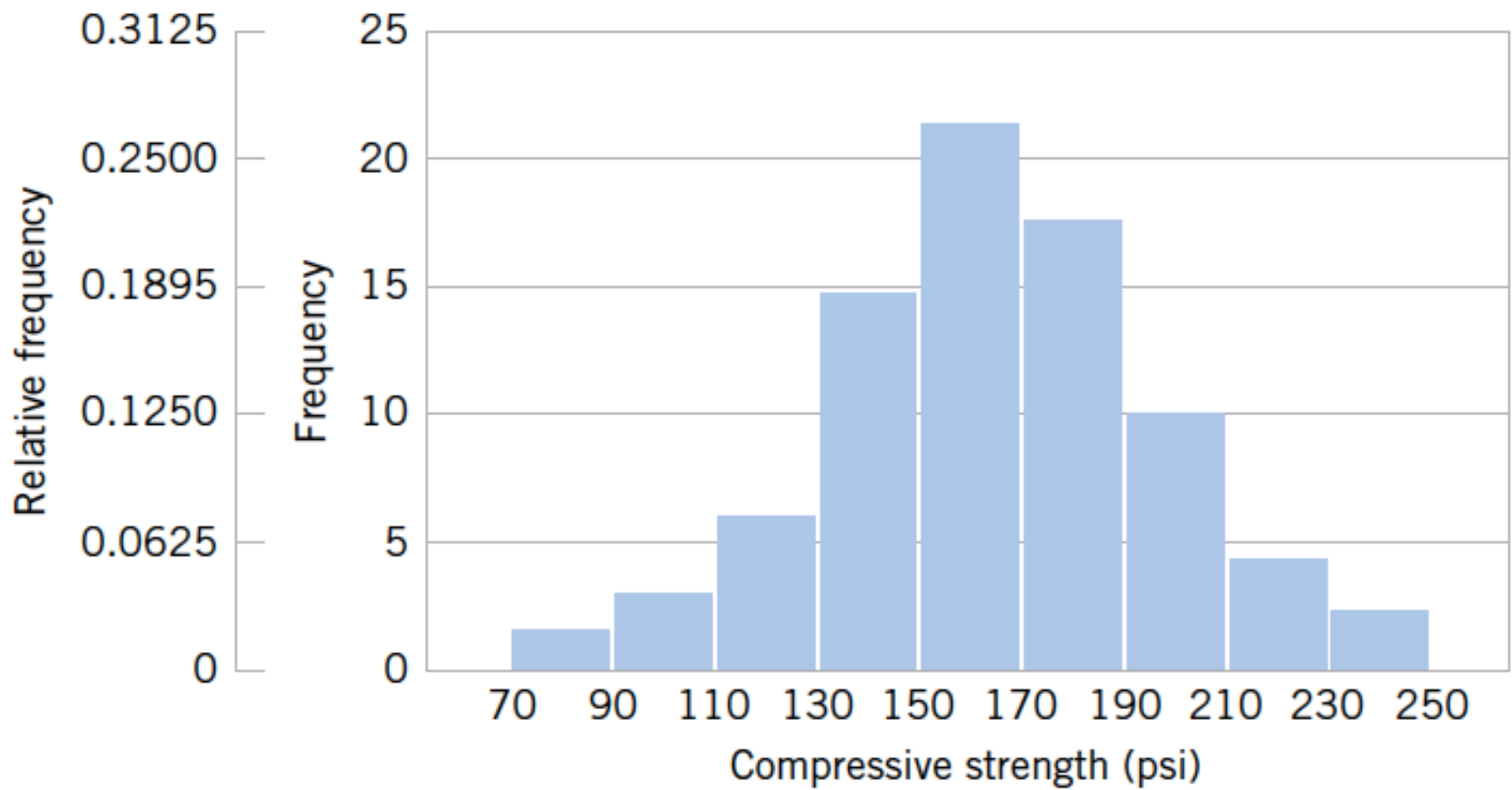
Table 6-4 Frequency Distribution for the Compressive Strength Data in Table 6-2

Class	$70 \leq x < 90$	$90 \leq x < 110$	$110 \leq x < 130$	$130 \leq x < 150$	$150 \leq x < 170$	$170 \leq x < 190$	$190 \leq x < 210$	$210 \leq x < 230$	$230 \leq x < 250$
Frequency	2	3	6	14	22	17	10	4	2
Relative frequency	0.0250	0.0375	0.0750	0.1750	0.2750	0.2125	0.1250	0.0500	0.0250
Cumulative relative frequency	0.0250	0.0625	0.1375	0.3125	0.5875	0.8000	0.9250	0.9750	1.0000

Histograma: apresentação visual da distribuição de frequências

Passos para a construção de um histograma com intervalos iguais:

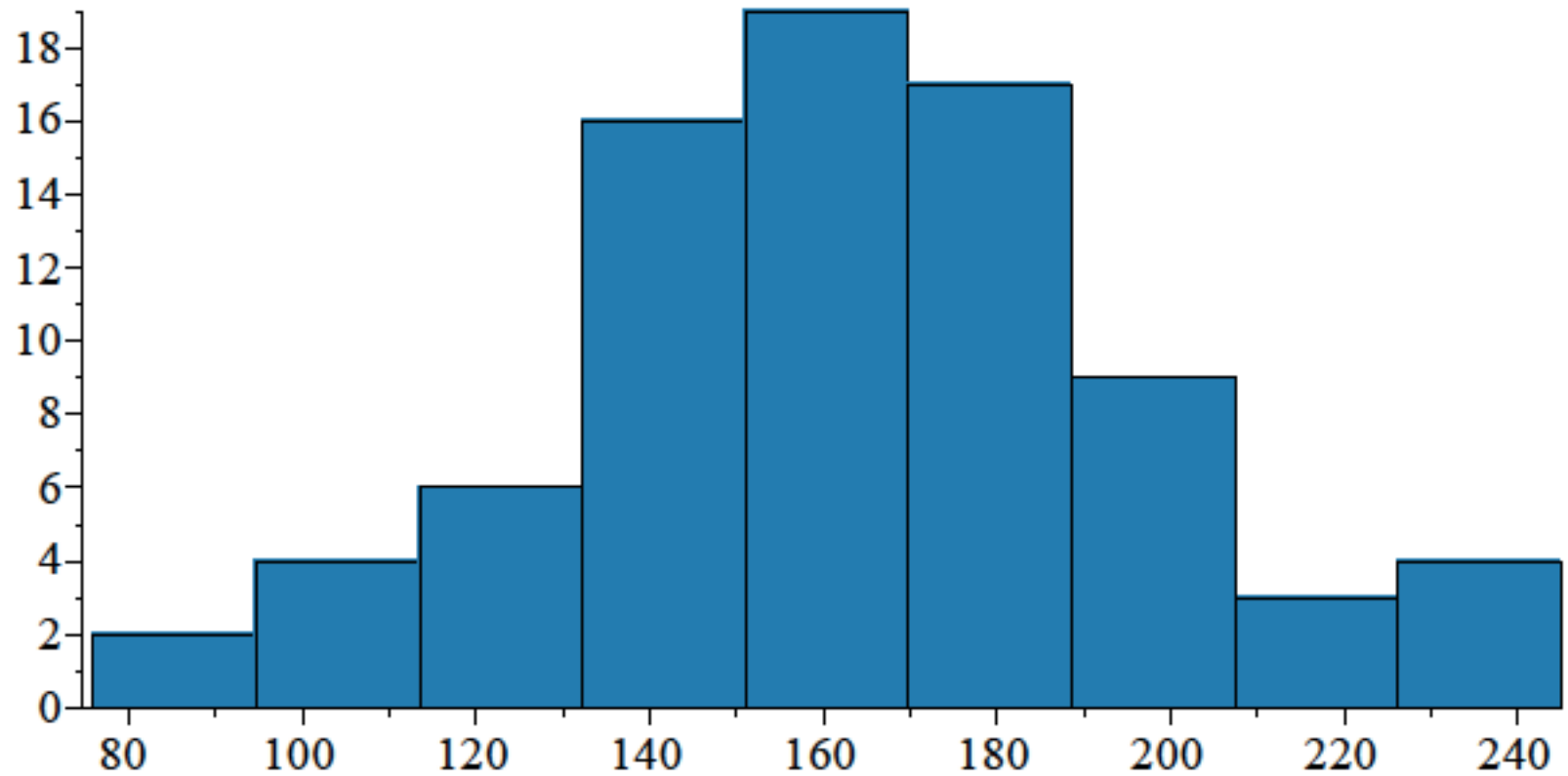
1. Marque os limites dos intervalos na escala horizontal.
2. Marque as frequências ou frequências relativas na escala vertical.
3. Acima de cada intervalo desenhe um retângulo onde a altura é igual à frequência (ou frequência relativa) correspondente a este intervalo.



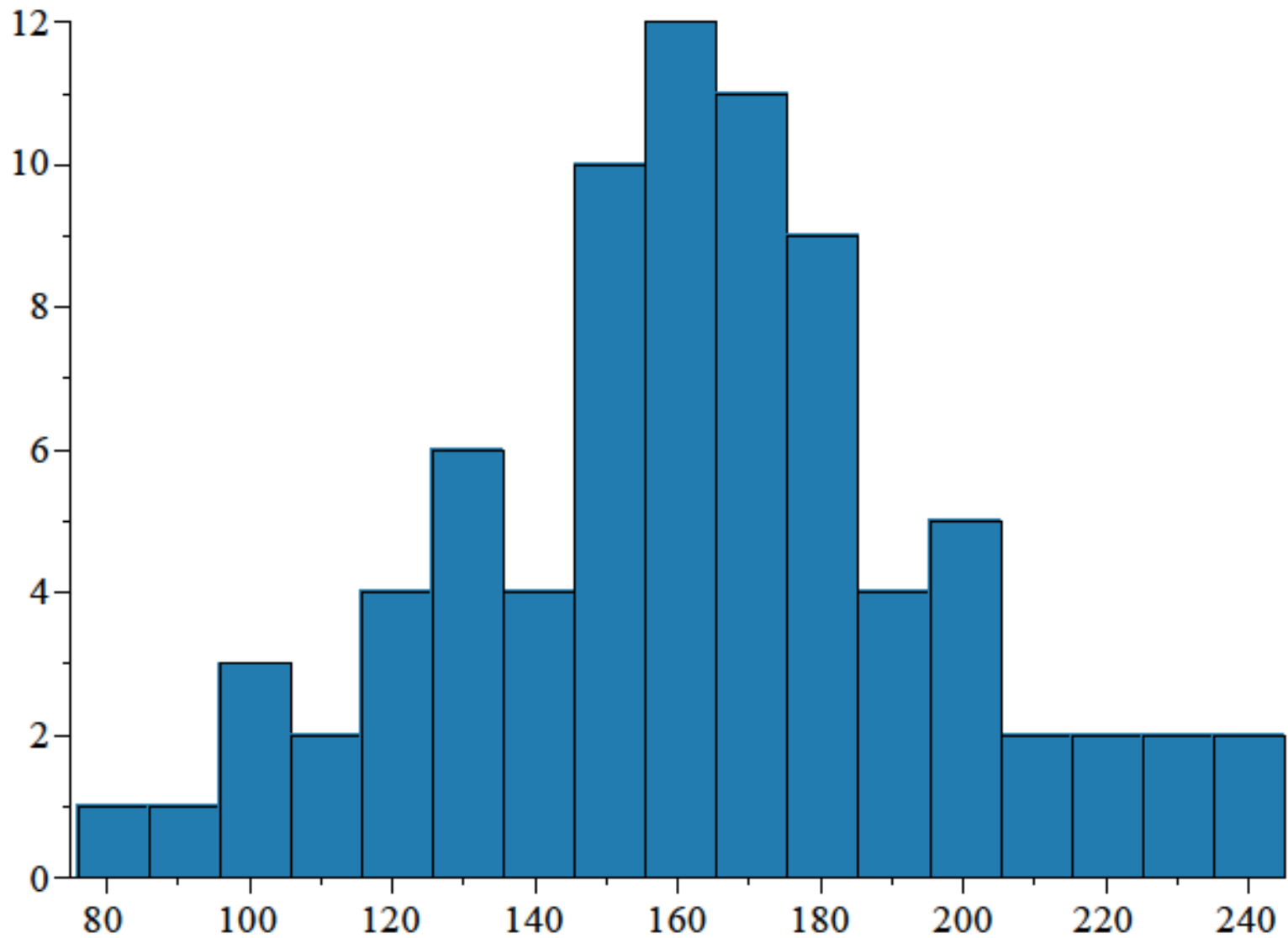
Histogramas no Maple

```
> with (Statistics) :  
> L := [105, 221, 183, 186, 121, 181,  
180, 143, 97, 154, 153, 174, 120, 168,  
167, 141, 245, 228, 174, 199, 181, 158,  
176, 110, 163, 131, 154, 115, 160, 208,  
158, 133, 207, 180, 190, 193, 194, 133, 156,  
123, 134, 178, 76, 167, 184, 135, 229, 146, 218,  
157, 101, 171, 165, 172, 158, 169, 199, 151,  
142, 163, 145, 171, 148, 158, 160,  
175, 149, 87, 160, 237, 150, 135, 196, 201,  
200, 176, 150, 170, 118, 149] :
```

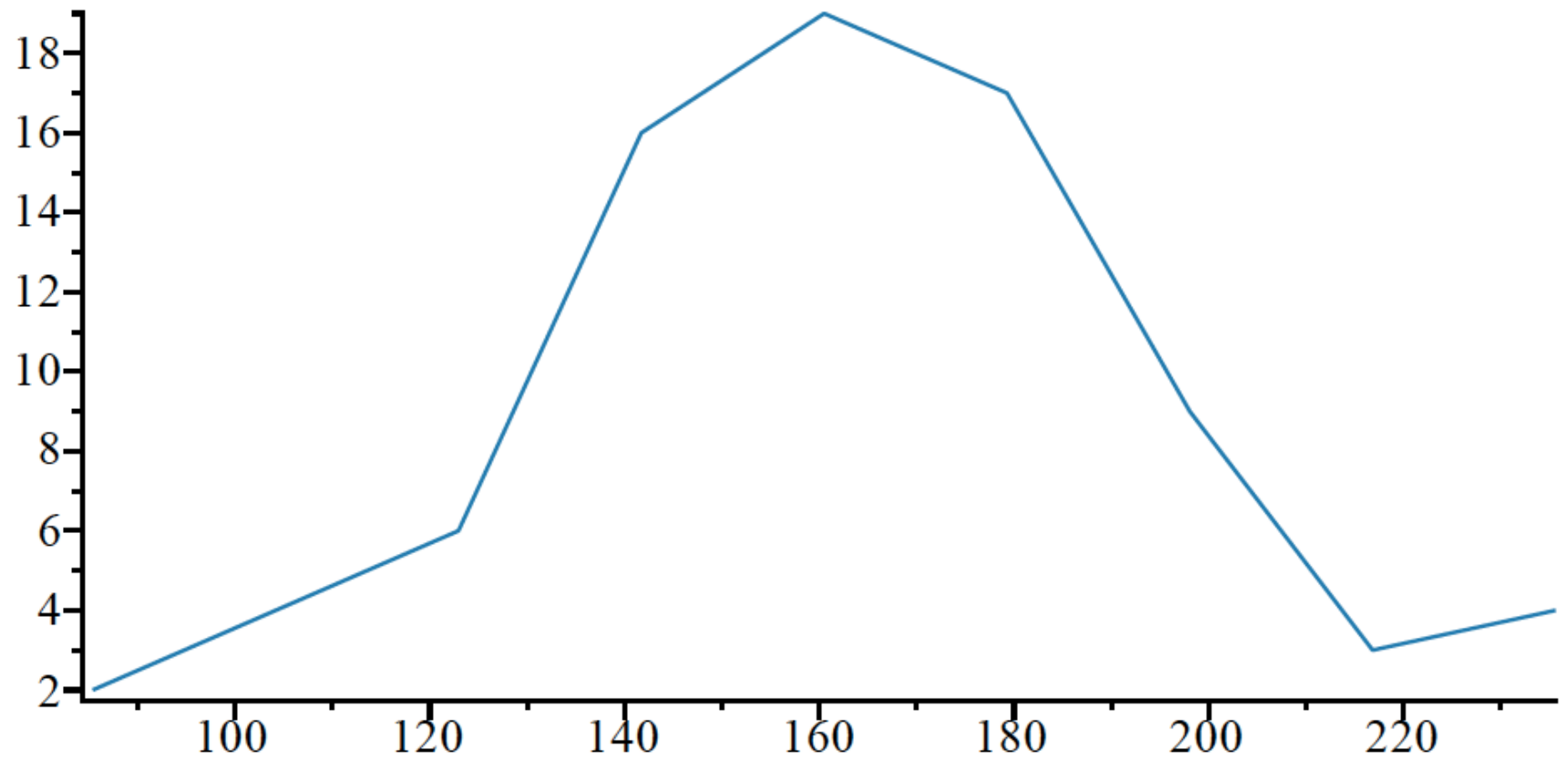
```
> Histogram(L, frequencyscale = absolute, bincount = 9)
```



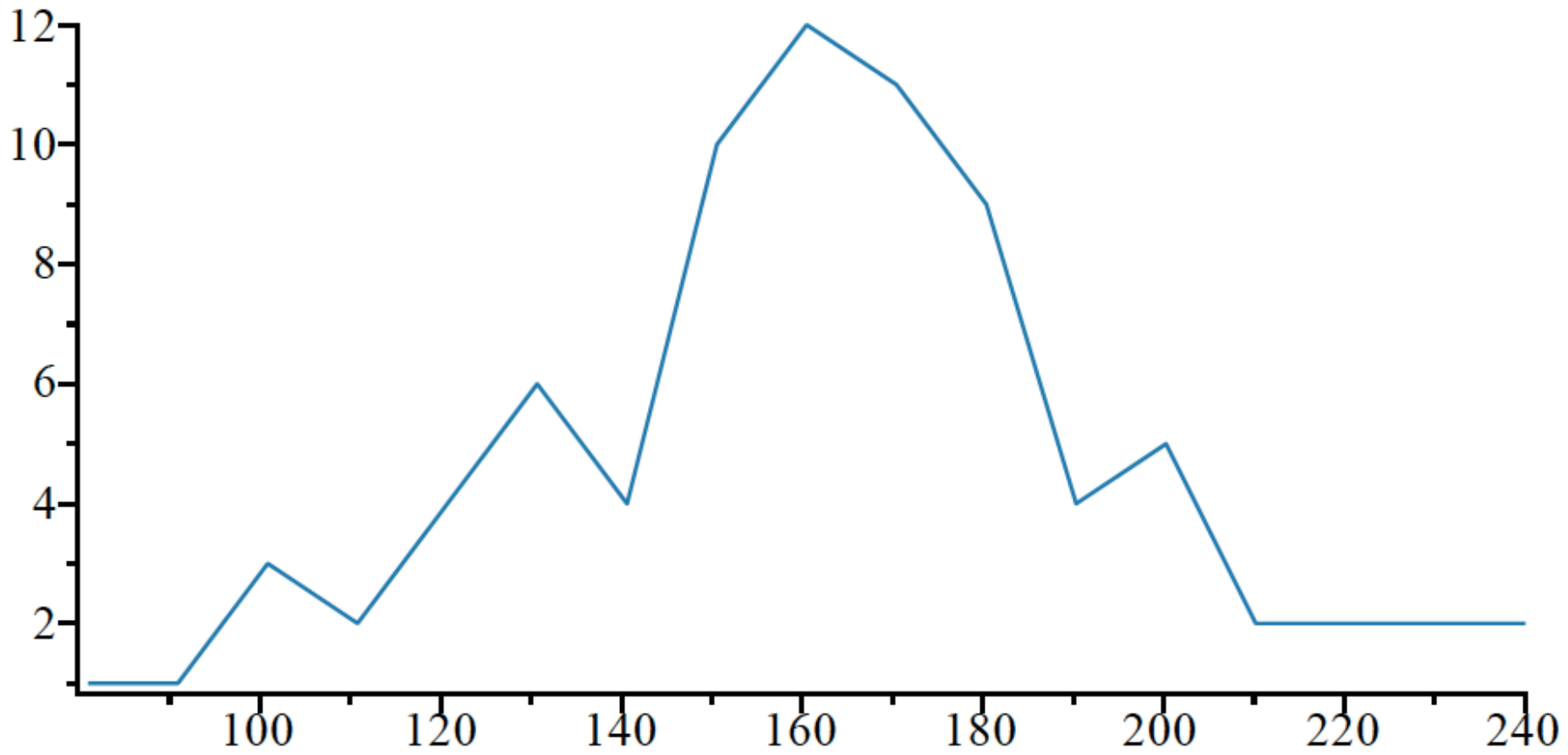
```
> Histogram(L, frequencyscale = absolute, bincount = 17)
```



```
> FrequencyPlot(L, frequencyscale= absolute, bincount= 9)
```



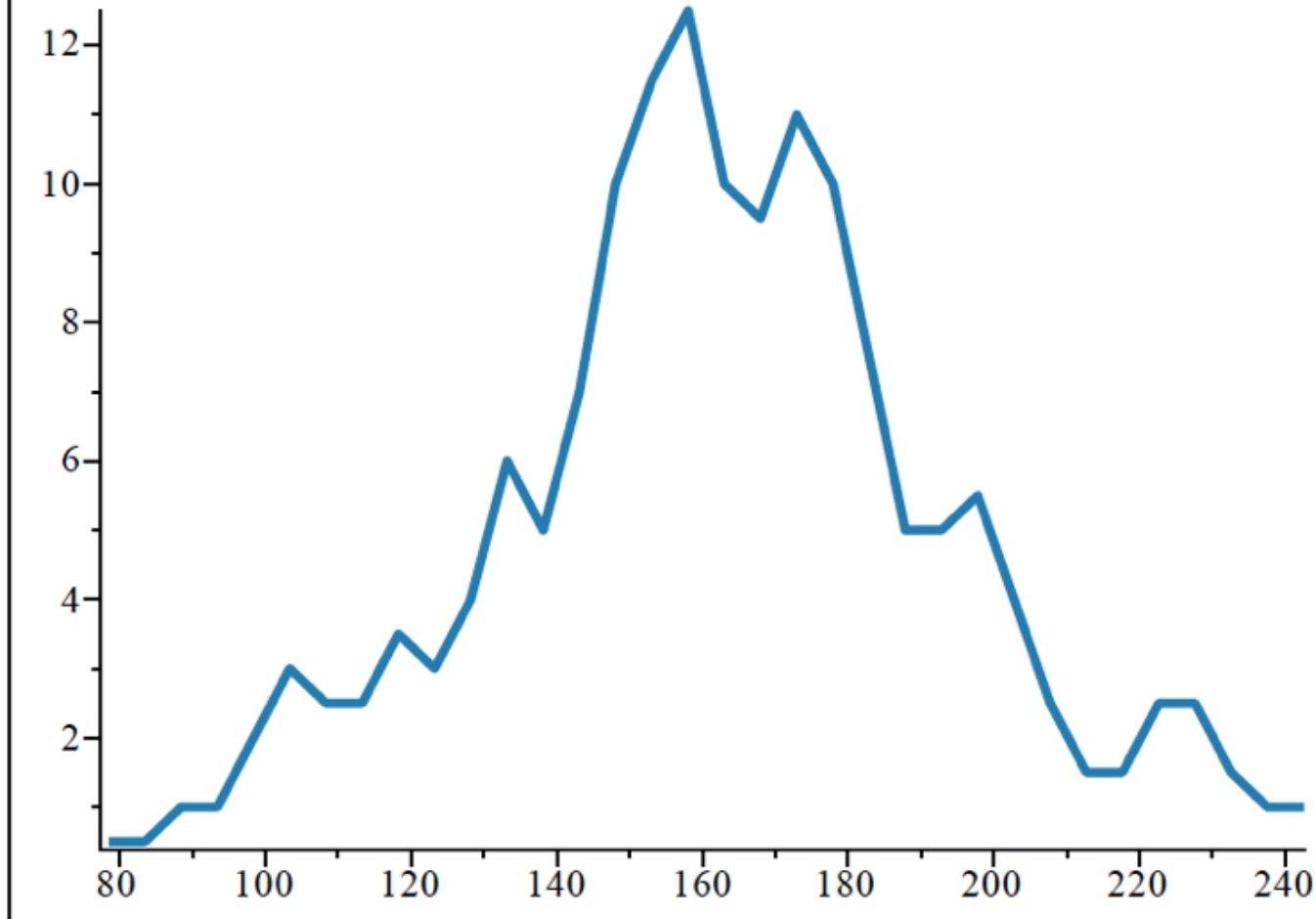
```
> FrequencyPlot(L, frequencyscale= absolute, bincount= 17)
```



Esta opção especifica quanto plots de frequência são usados para computar a média O default é 1.



```
> FrequencyPlot(L, thickness = 3, averageshifted = 2, frequencyscale = absolute, bincount = 17)
```



Um plot de frequência deslocada média é obtido calculando m plots de frequência com o mesmo tamanho de intervalo, deslocando-os por w/m , e tomando a média.

Plots de frequência deslocada média são mais suaves que plots de frequência ordinários e são menos dependentes da escolha da origem.