Parte 02

Estatística Descritiva

Prof. Dr. Renato de Oliveira Moraes

Técnicas de amostragem

- **Amostragem probabilística**
  Todos os os elementos da população têm uma probabilidade conhecida e diferentes de zero de pertencer à amostra

- **Amostragem não probabilística**
  Alguns elementos da população têm probabilidade zero de pertencer à amostra

Técnicas de amostragem probabilística

- Amostragem casual simples
- Amostragem Sistemática
- Amostragem estratificada
- Amostragem por meio de conglomerados
- Amostragem múltipla (vários estágios)

Amostragem probabilística

✓ Amostragem Casual Simples

Todos elementos da população têm a mesma probabilidade de participar da amostra. Os indivíduos da população são numerados e a seleção é feita através de um sorteio.

Formas mais comuns:

- Sorteio
- Tabela de números aleatórios
- Programas que geram números randômicos.

Amostragem probabilística

- ✓ **Amostragem Sistemática**

  Os elementos da amostra são divididos em grupos de mesmo tamanho que estão próximos no tempo ou no espaço. De cada grupo é escolhido um elemento para compor a amostra.

Amostragem probabilística

- ✓ **Amostragem Casual Estratificada**

  A população é divida em grupos, chamados estratos, onde existe uma grande homogeneidade dentro dos grupos e uma heterogeneidade entre eles. Dentro de cada grupo são escolhidos, aleatoriamente, elementos para compor a amostra.

Amostragem probabilística

✓ Amostragem por Conglomerados

A população é divida em grupos, chamados conglomerados, onde existe uma heterogeneidade dentro dos grupos e uma homogeneidade entre eles – ao contrário dos estratos. Inicialmente são escolhidos os conglomerados que vão fornecer elementos para amostra. Não é necessário colher elementos de todos os grupos, já que se supõe que cada um deles seja representativo da população. Nos conglomerados selecionados pode-se realizar um censo (ACC-I), ou colher dentro deles amostras probabilísticas (ACC-II).

Estatística Descritiva

- Medidas de posição
  - Média amostral – $\overline{x}$
  - Mediana amostral
  - Moda amostral
- Medidas de posição
  - Variância amostral – s^2
  - Desvio padrão amostral – s

Exemplo

Dada a amostra de 7 elementos:

1,2 1,2 1,3 1,5 1,7 1,9 2,0

mediana = 1,5

moda = 1,2

média = 1,4

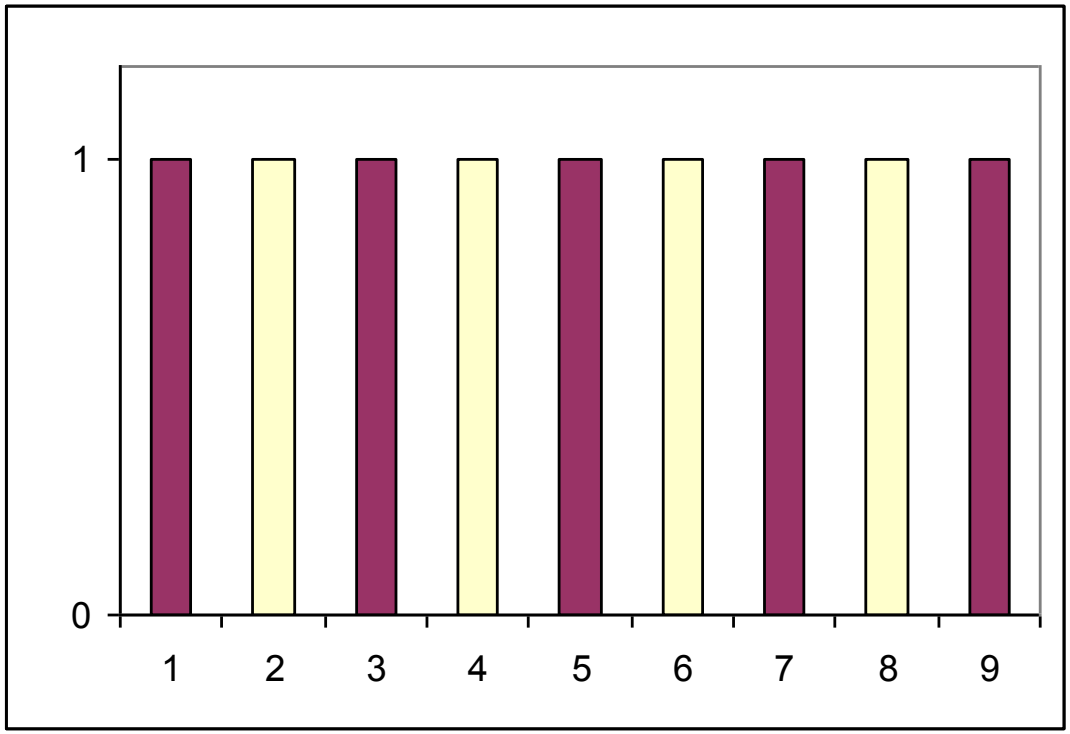
Exemplo

Amostra 1: 1 3 5 7 9

média = 5

Amostra 2: 2 4 6 8

média = 5



Medidas de posição não são suficientes para uma boa caracterização dos dados de uma amostra

Medidas de Dispersão

$$\text{Média amostral}: \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\text{Variância amostral}: s^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}$$

$$\text{Desvio padrão amostral}: s = \sqrt{s^2}$$

Exemplo

Amostra 1: 1 3 5 7 9

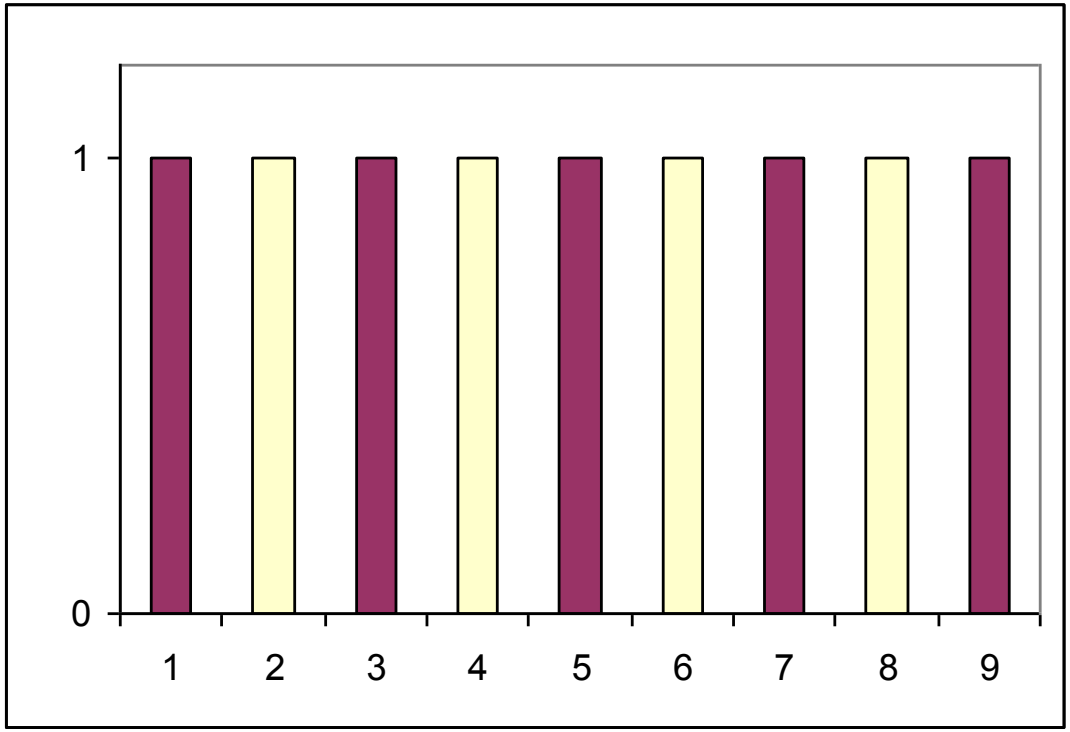
média = 5

S^2=10

Amostra 2: 2 4 6 8

média = 5

S^2=6,67



Exercício 2

Um paciente fez 6 exames de sangue em 6 meses consecutivos para medir seu nível de fosfato por decilitro de sangue. Os resultados obtidos foram:

5.6 5.2 4.6 4.9 5.7 6.4

Calcule a média, mediana e o desvio padrão manualmente.

Exercício 3

Os salários de 20 funcionários de um banco, ordenados em ordem crescente são: $1400, 1400, 1400, 1400, 1400, 1400, 1400, 1400, 1550, 1550, 1650, 1650, 1800, 1800, 1900, 2000, 2050, 2250, 2300, 2400.

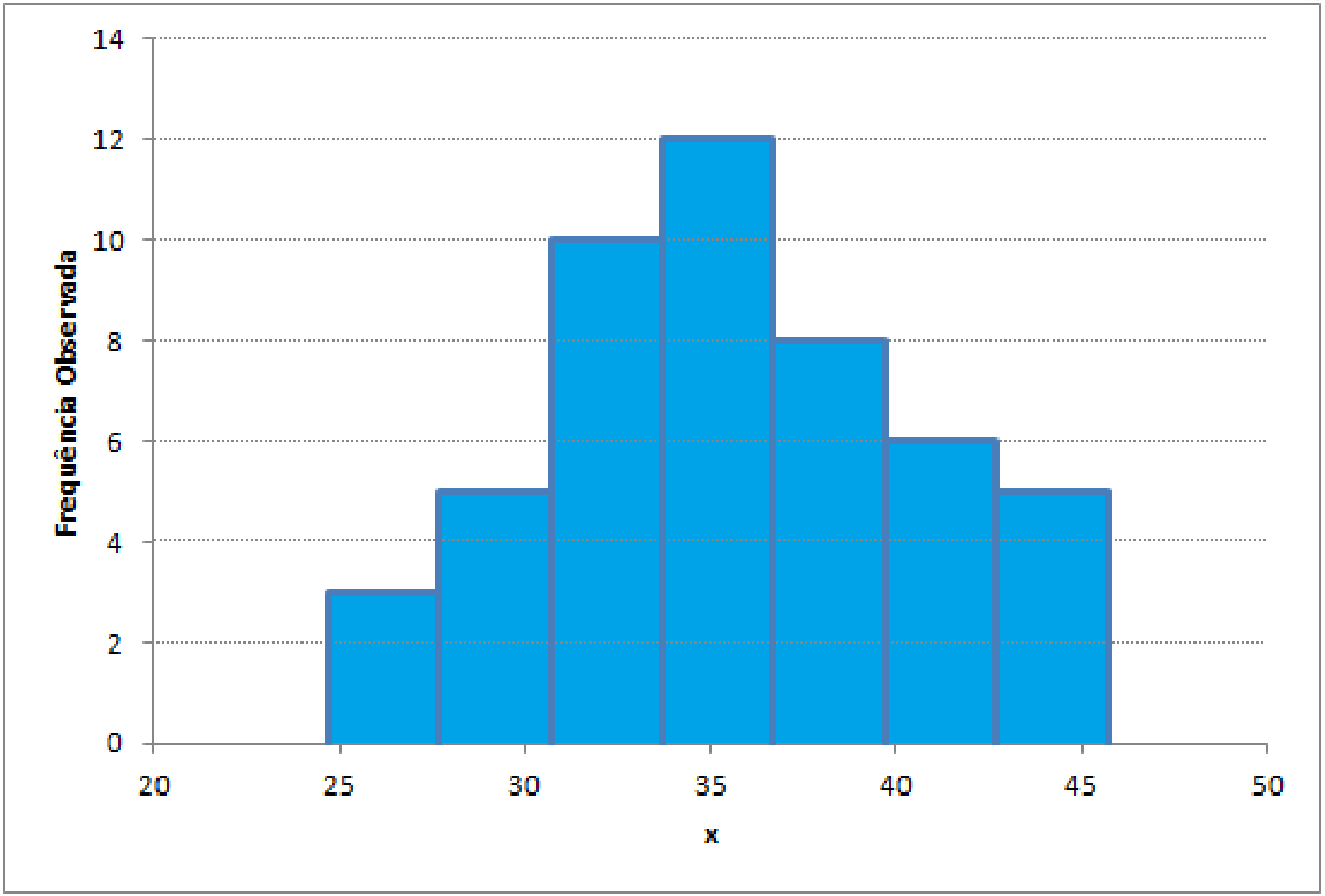
Calcule média, mediana e moda para estes dados.

Descritiva no MS Excel

- Média =media(range)
- Moda =modo(range)
- Mediana =med(range)
- Quartil =quartil(range;quarto)
  - Quarto:
    - 0: Mínimo
    - 1: Primeiro Quartil
    - 2: Segundo Quartil (Mediana)
    - 3: Terceiro Quartil
    - 4: Máximo
- Variância =vara(range)
- Desvio padrão =desvpada(range)

Histograma - Dados

33,7	29,6	35,6	31,3	36,7	34,4	32,1
34,0	33,6	26,7	28,5	33,5	29,1	36,2
39,1	32,1	29,8	43,6	39,7	40,3	30,6
43,1	36,4	41,5	31,2	43,1	41,9	37,7
35,5	35,8	33,8	32,8	39,5	31,4	40,3
31,4	36,0	44,0	30,6	38,3	40,1	35,6
44,1	43,7	27,1	41,2	24,7	35,6	38,8


14
12
10
8
6
4
2
0
Frequência Observada
20
25
30
35
40
45
50
x

Dados originais

N = 49
Mínimo = 24,7
Máximo = 44,1
Amplitude (R) = 19,4
Número de classes (k) = 7

$$k = \sqrt{N}$$

$$ou$$

$$k \cong 1 + 3{,}322 \log_{10}(N)$$

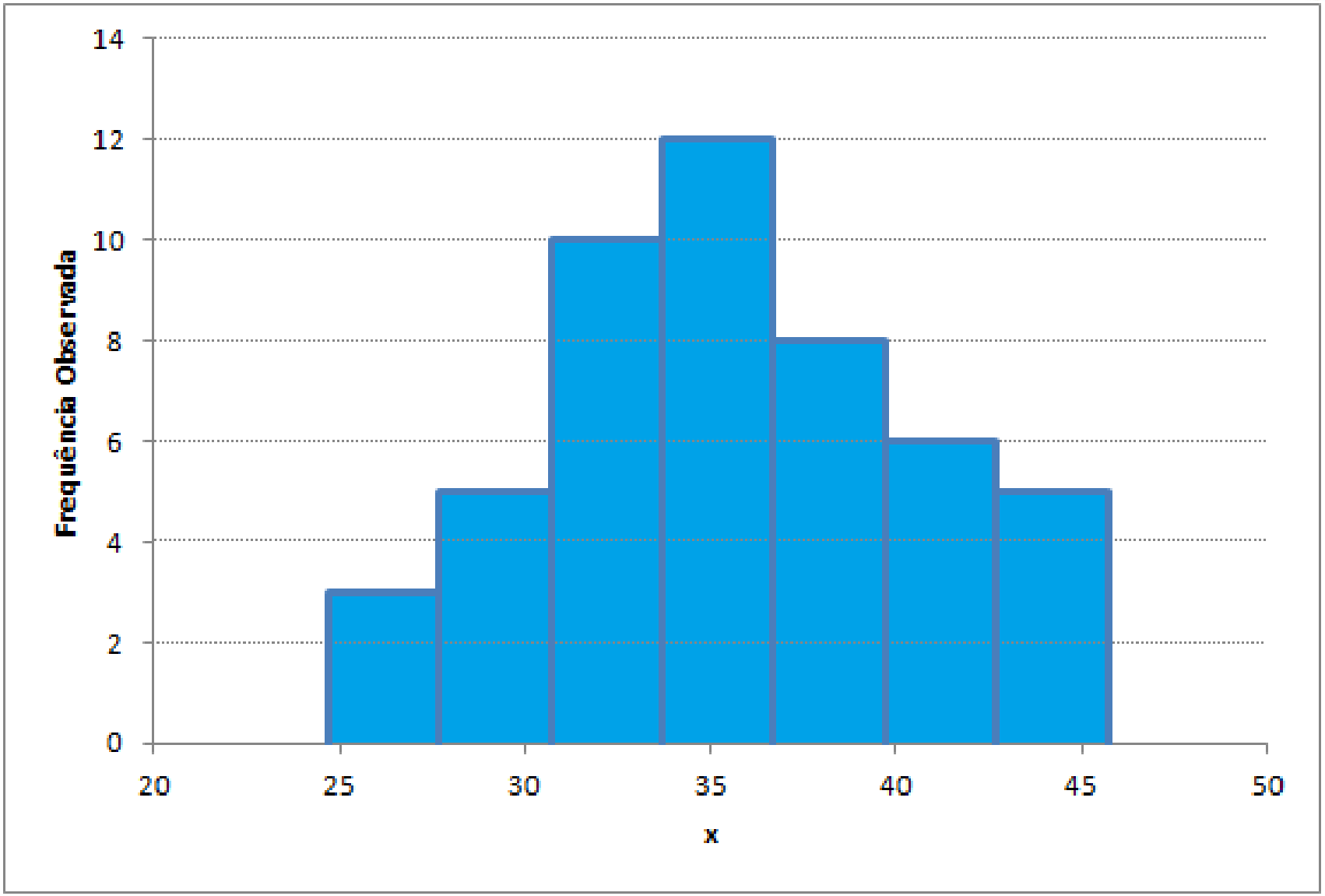
Regra de Sturges

Amplitude das classes (h) = R/k = 2,77 = 3

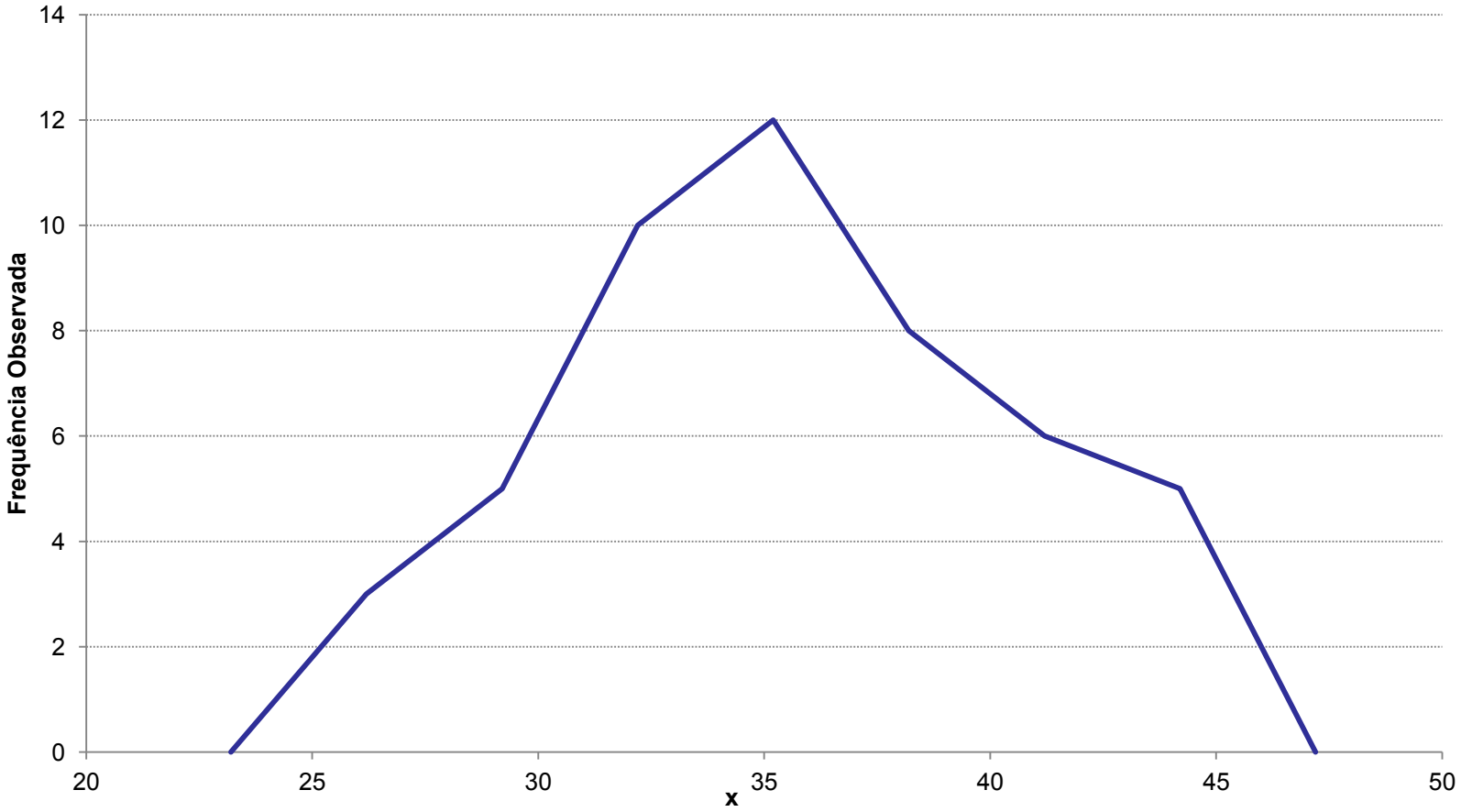
Classes de freqüência

Classes	Limites	f	F	p	P
1	24,7|--27,7	3	3	6,1%	6,1%
2	27,7|--30,7	5	8	10,2%	16,3%
3	30,7|--33,7	10	18	20,4%	36,7%
4	33,7|--36,7	12	30	24,5%	61,2%
5	36,7|--39,7	8	38	16,3%	77,6%
6	39,7|--42,7	6	44	12,2%	89,8%
7	42,7|--45,7	5	49	10,2%	100,0%

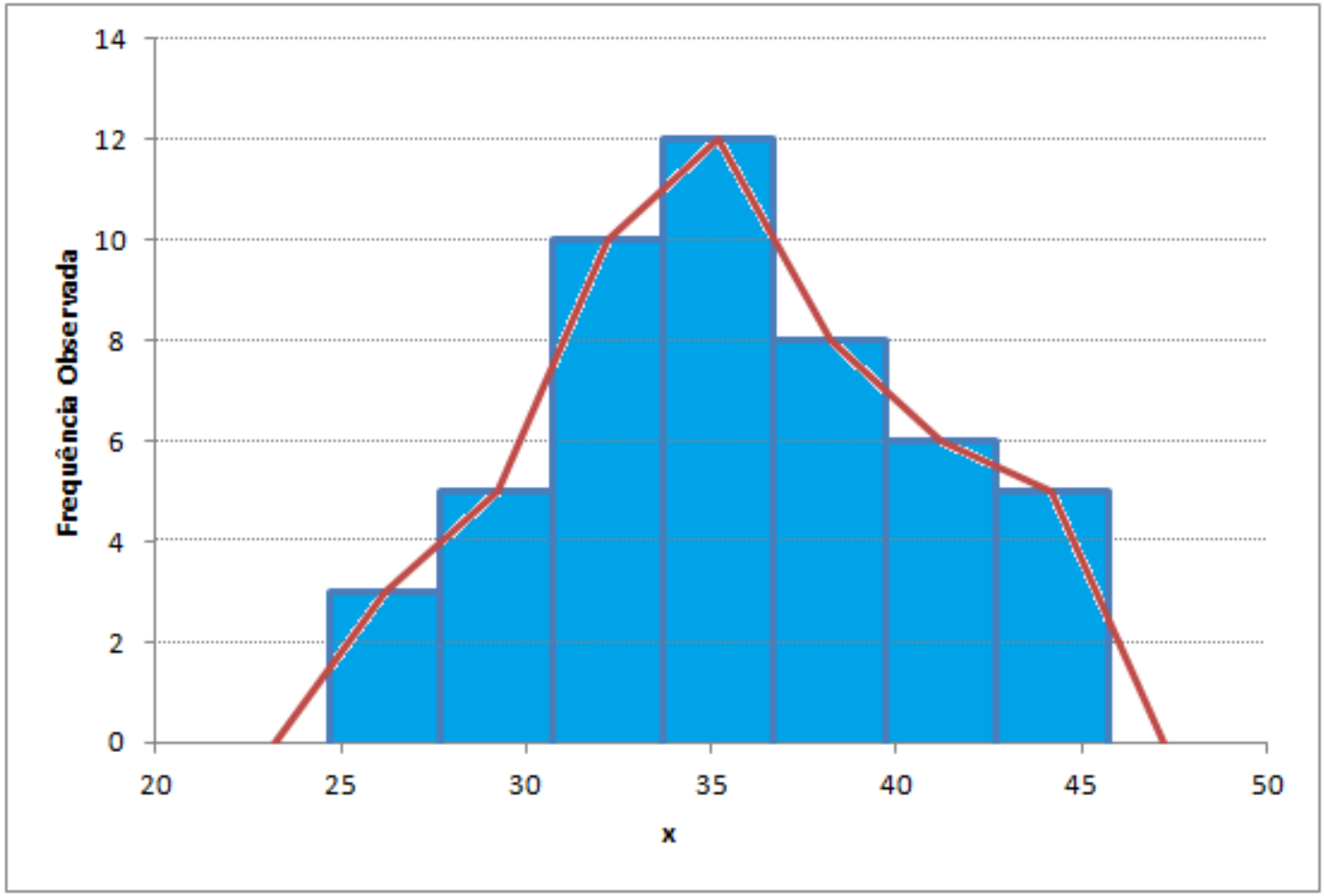
$\Sigma =$ 49 $\Sigma =$ 100,0%


14
12
10
8
6
4
2
0
Frequência Observada
20
25
30
35
40
45
50
x

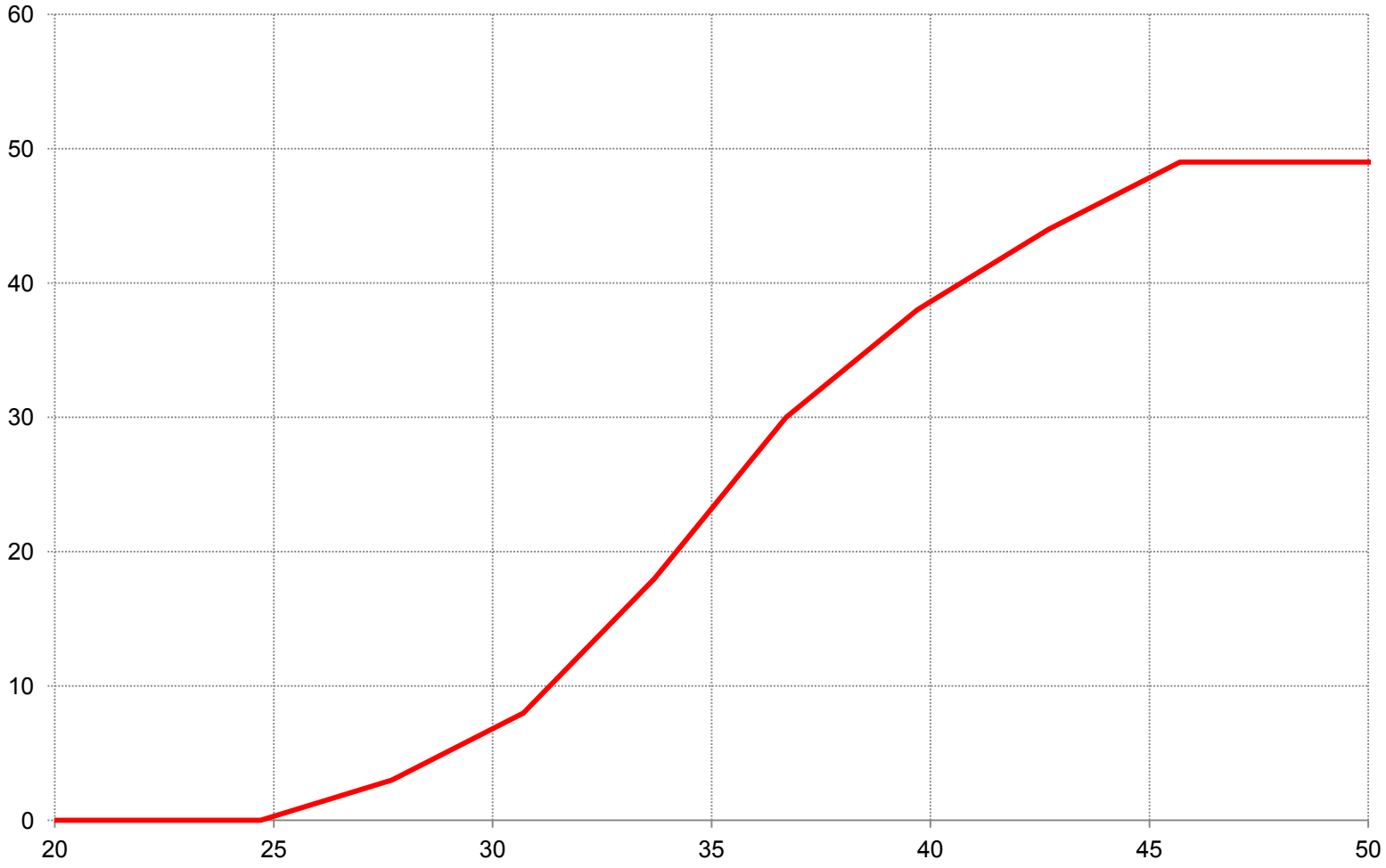
Polígono de frequências



Polígono de frequências



Frequência Acumulada



Exercício 4

Calcule média, mediana e moda da distribuição de freqüências dada abaixo

Classes	Freqüências
150-179	3
180-209	8
210-239	10
240-269	13
270-299	33
300-329	40
330-359	35
360-389	30
390-419	16
420-449	12

Medidas de posição em dados agrupados

- Média
- Mediana
- Moda

Dados

Classes	xi	f	F
12|—16	14	2	2
16|—20	18	5	7
20|—24	22	11	18
24|—28	26	18	36
28|—32	30	10	46
32|—36	34	7	53
36|—40	38	3	56
	Σ	56	

Média

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i f_i}{\sum f_i}$$

Onde:

x_i é o ponto central da classe ***i***

n é a quantidade de elementos – tamanho da amostra

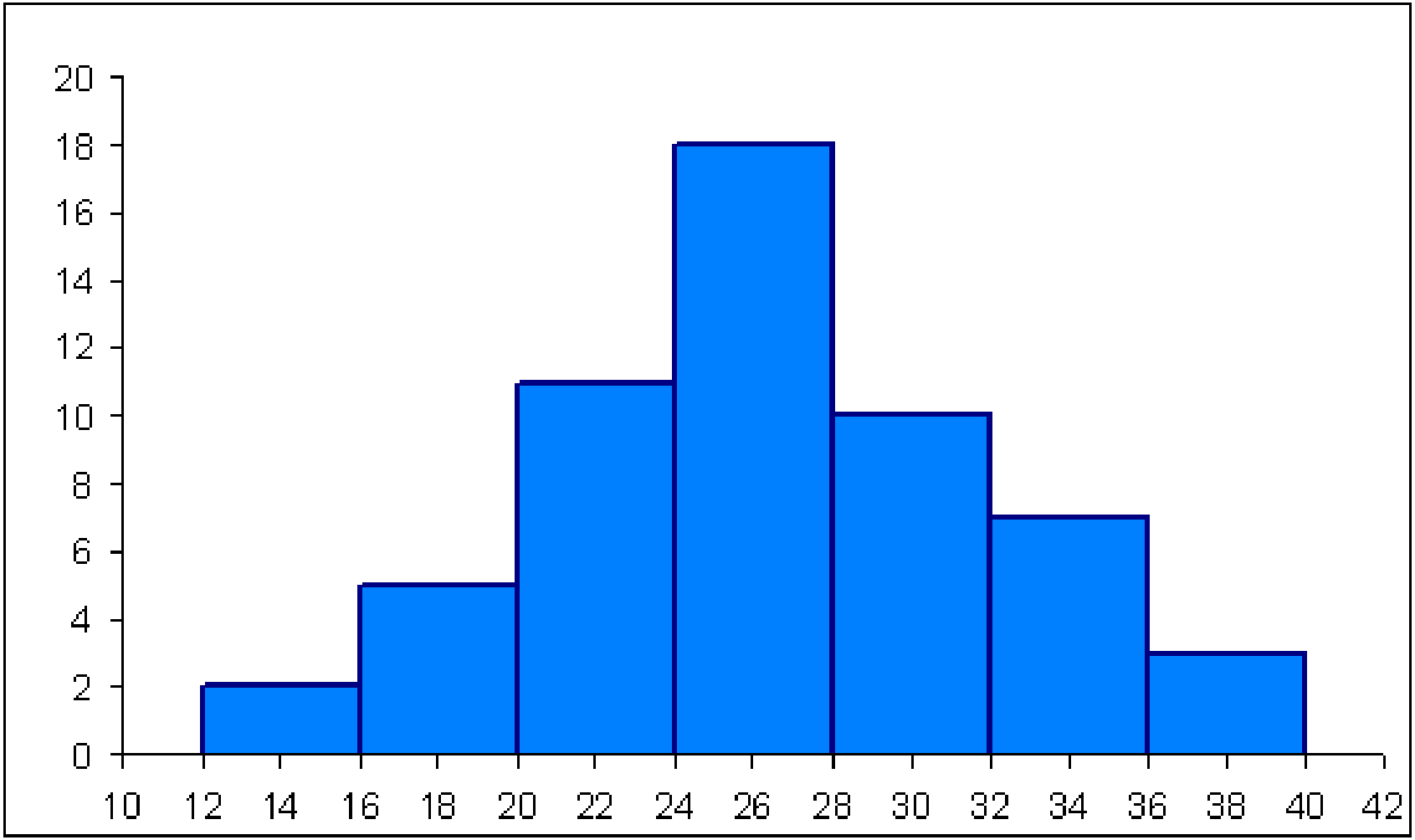
f_i é a freqüência observada da classe ***i***

Média

Classes	x_i	f	x_i f
12|—16	14	2	28
16|—20	18	5	90
20|—24	22	11	242
24|—28	26	18	468
28|—32	30	10	300
32|—36	34	7	238
36|—40	38	3	114
	Σ	56	1.480

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i f_i}{\sum f_i} = \frac{1.480}{56} = 26{,}43$$

Média


20
18
16
14
12
10
8
6
4
2
0
10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42

Mediana

$$md = L_i + \frac{\left(n/2 - F_a\right)}{f_{md}} h_{md}$$

Onde:

L_i é o limite inferior da classe mediana

n é a quantidade de elementos – tamanho da amostra

F_a é a freqüência acumulada da classe imediatamente anterior

f_{med} é a freqüência observada da classe mediana

h é a amplitude da classe mediana

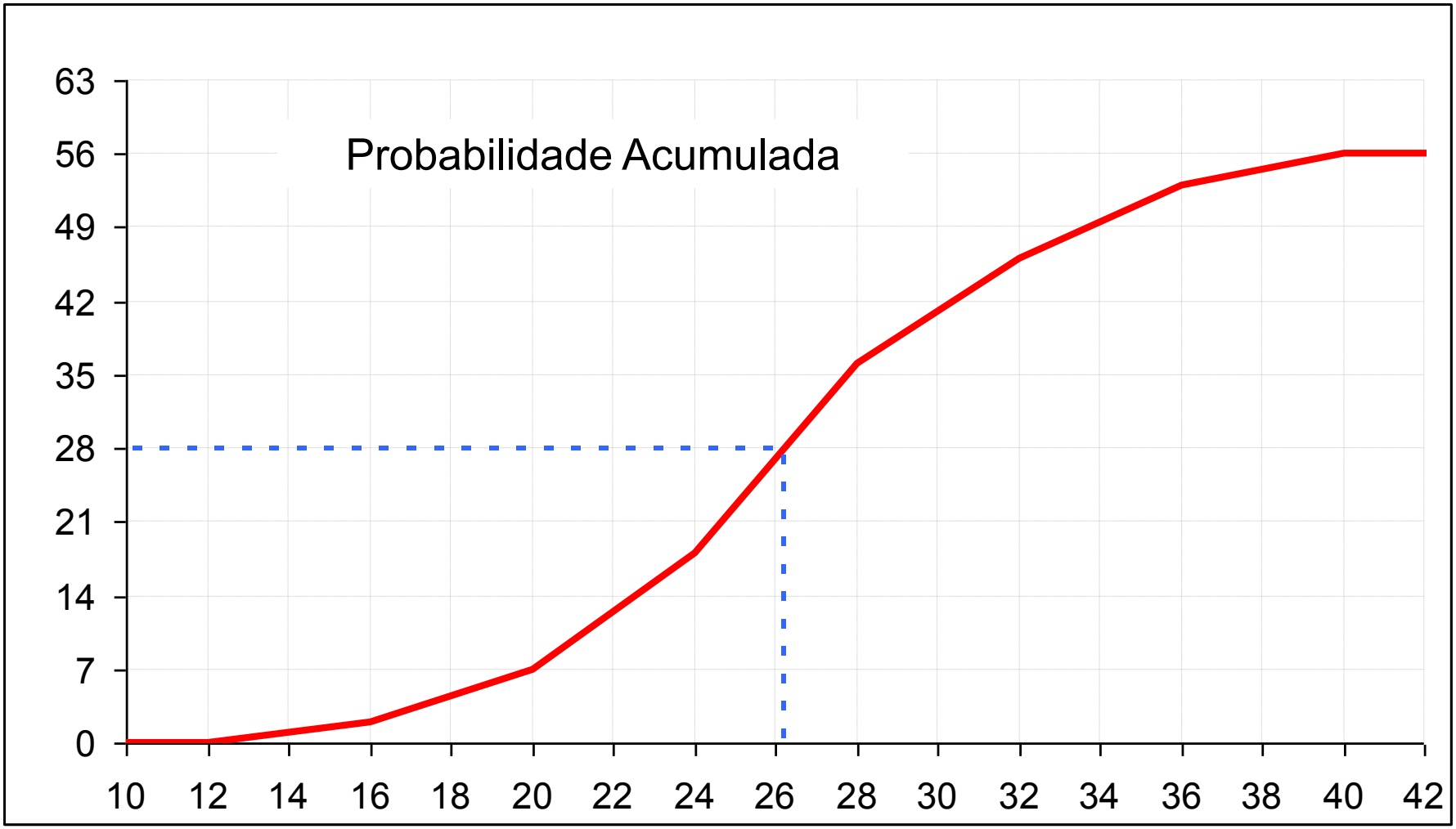
Mediana

Classes	x_i	f	F
12|—16	14	2	2
16|—20	18	5	7
20|—24	22	11	18
24|—28	26	18	36
28|—32	30	10	46
32|—36	34	7	53
36|—40	38	3	56
	Σ	56	

Classe da mediana

$$md = L_i + \frac{\left(n/2 - F_a\right)}{f_{md}} h_{md} = 24 + \frac{\left(56/2 - 18\right)}{18} 4 = 26{,}22$$

Mediana


Probabilidade Acumulada
63
56
49
42
35
28
21
14
7
0
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42

Moda

$$m_o = L_i + \frac{d_1}{(d_1 + d_2)} h$$

Onde:

L_i é o limite inferior da classe modal

d_1 é a diferença entre a freqüência da classe modal e da classe imediatamente anterior

d_2 é a diferença entre a freqüência da classe modal e da classe imediatamente posterior

h é a amplitude da classe modal

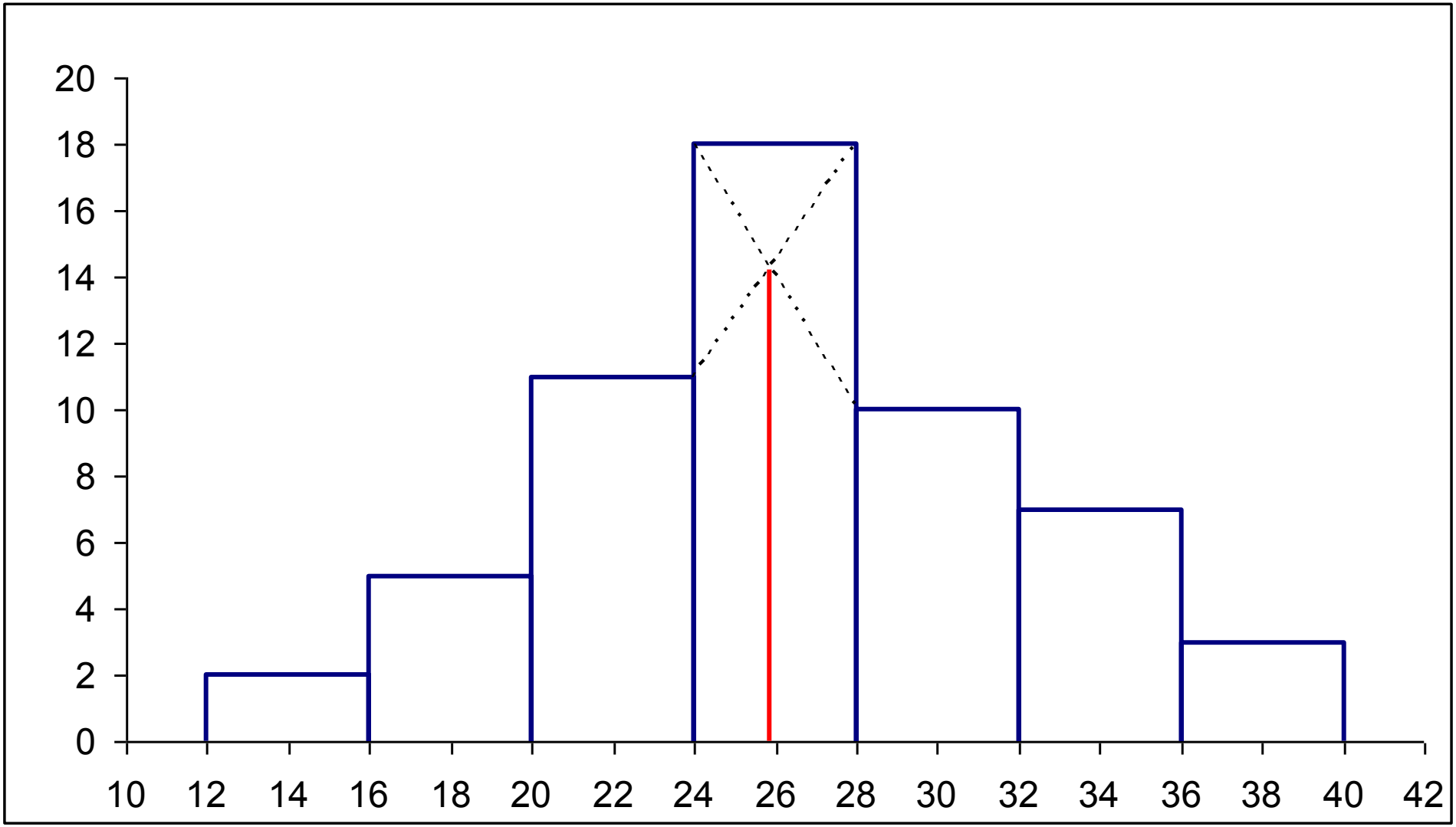
Moda

Classes	x_i	f	F
12|—16	14	2	2
16|—20	18	5	7
20|—24	22	11	18
24|—28	26	18	36
28|—32	30	10	46
32|—36	34	7	53
36|—40	38	3	56
	Σ	56	

Classe da moda

$$m_o = L_i + \frac{d_1}{(d_1 + d_2)} h = 24 + \frac{7}{(7+8)} 4 = 25{,}87$$

Moda


20
18
16
14
12
10
8
6
4
2
0
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
42

Exercício

Classes	x_i	f	x_i*f	F
150-179	164,5	3	493,5	3
180-209	194,5	8	1556	11
210-239	224,5	10	2245	21
240-269	254,5	13	3308,5	34
270-299	284,5	33	9388,5	67
300-329	314,5	40	12580	107
330-359	344,5	35	12057,5	142
360-389	374,5	30	11235	172
390-419	404,5	16	6472	188
420-449	434,5	12	5214	200
		200	**64.550**	

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i \times f_i}{n} = \frac{64.550}{200} = 322{,}75$$

$$md = L_i + \frac{\left(n/2 - F_a\right)}{f_{md}} h_{md} = 300 + \frac{(100-67)}{40} 30$$

$$md = 300 + \frac{33}{40} 30 = 300 + 24{,}75 = 324{,}75$$

$$m_o = L_i + \frac{d_1}{(d_1 + d_2)} h$$

$$m_o = 300 + \frac{7}{(5+7)} 30$$

$$m_o = 317{,}5$$

Exercício

Considere a amostra a seguir e construa:

i. Distribuição em classes de freqüência
ii. Histograma
iii. A partir das classes de freqüência, determine:
   a) Média
   b) Moda
   c) Mediana
   d) Variância
   e) Desvio Padrão
   f) Coeficiente de Variação

Dados da amostra

34,5	35,2	39,9	29,2	20,9	39,5	45,6	35,1
41,5	37,8	30,6	28,2	31,3	30,2	33,0	36,8
30,9	39,4	41,0	34,0	34,0	41,2	38,0	40,2
33,3	42,5	35,7	28,5	35,0	35,9	32,6	35,8
37,6	33,4	32,5	37,0	41,1	30,4	31,7	33,1
40,3	37,8	38,8	39,5	33,6	30,1	36,1	32,0
38,5	34,0	39,3	39,6	27,2	38,7	41,6	33,8
42,2	38,8	25,4	36,5	27,3	38,9	34,8	31,8
33,2	41,8	32,3	32,5	31,9	35,7	33,3	41,6
23,2	38,0	37,7	44,4	37,2	34,6	33,2	34,3

Classe	Limites	f	F	p	P
1	20|–-23	1	1	1,3%	1,3%
2	23|–-26	2	3	2,5%	3,8%
3	26|–-29	4	7	5,0%	8,8%
4	29|–-32	10	17	12,5%	21,3%
5	32|–-35	21	38	26,3%	47,5%
6	35|–-38	16	54	20,0%	67,5%
7	38|–-41	15	69	18,8%	86,3%
8	41|–-44	9	78	11,3%	97,5%
9	44|–-47	2	80	2,5%	100,0%
		80		100,0%	

Exercício

Considere a amostra a seguir e construa:

i. Distribuição em classes de freqüência
ii. Histograma
iii. A partir das classes de freqüência, determine:
   a) Média
   b) Moda
   c) Mediana
   d) Variância
   e) Desvio Padrão
   f) Coeficiente de Variação

Dados da amostra

45	46	48	51	47	46	45	50
39	45	45	45	45	47	43	47
47	46	49	43	46	46	44	44
45	48	45	47	46	43	44	49
49	45	46	46	44	46	43	47
46	44	48	44	45	45	44	43
44	49	44	46	46	47	45	46
45	43	41	45	43	47	44	46
45	47	42	44	43	47	45	45
40	48	47	42	47	42	47	47