

PRO3200 - Lista de Exercícios - Amostragem e Estatística Descritiva

1. Para cada uma das amostras abaixo, informar o tipo do processo de amostragem:

P - Amostragem probabilística

NP - Amostragem não probabilística

Para uma pesquisa sobre os hábitos dos estudantes. Construí uma amostra com o seguinte procedimento:

- (NP) Todos os meus colegas da faculdade (tenho telefone e email de todos eles)
A amostra não é probabilística, já que nem todos os elementos da população (estudantes) tem probabilidade diferente de zero de pertencer à amostra. No caso, os estudantes que não são colegas tem probabilidade zero de pertencer à amostra.
- (NP) Fiquei na única porta de entrada da escola abordando todos os meus conhecidos
A amostra não é probabilística, já que nem todos os elementos da população (estudantes) tem probabilidade diferente de zero de pertencer à amostra. No caso, os estudantes que não são conhecidos tem probabilidade zero de pertencer à amostra.
- (P) Fiquei na única porta de entrada da escola e cada 12 pessoas que entravam, eu abordava uma
A amostra é probabilística, já que todas as pessoas tem probabilidade conhecida e diferente de zero de pertencer à amostra.
- (P) Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista
A amostra é probabilística, já que todos os alunos das escolas tem probabilidade conhecida e diferente de zero de pertencer à amostra.
- (P) Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados
A amostra é probabilística, já que todos os alunos das escolas tem probabilidade conhecida e diferente de zero de pertencer à amostra. Este caso se assemelha ao anterior.

2. Um estudo sobre o desempenho dos vendedores de uma grande cadeia de lojas de varejo está sendo planejado. Para tanto, deve ser colhida uma amostra probabilística dos vendedores.

Classifique cada uma das amostras abaixo conforme a seguinte tipologia:

(A) Amostragem casual simples

(B) Amostragem Sistemática

(C) Amostragem estratificada

(D) Amostragem por meio de conglomerados

- (B) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei todos vendedores que ocupavam posições múltiplas de 15 (15ª posição, 30ª posição, 45ª posição, 60ª posição, 75ª posição, 90ª posição, 105ª posição, etc)
A amostra é sistemática, já que a retirada dos elementos da amostra é periódica, ou seja, um vendedor é retirado a cada 15 presentes na lista.
- (D) Escolhi casualmente 3 lojas da rede. A amostra foi composta de todos os vendedores que atualmente em cada uma destas 3 lojas.
A amostragem é por meio de conglomerados, já que a população (vendedores) apresenta uma subdivisão em grupos (lojas), e 3 grupos foram casualmente escolhidos (sorteados) para compor a amostra.

- (C) Em cada uma das lojas, identifiquei todos os vendedores (lista de vendedores por loja). Selecionei aleatoriamente k vendedores da loja. Onde k é um número inteiro proporcional à quantidade de vendedores da loja

A amostragem é estratificada, já que a amostra leva em consideração a presença de estratos (lojas) na composição da população (vendedores). Neste caso, a amostragem estratificada é proporcional, pois o número de elementos selecionados em cada estrato é proporcional à quantidade de elementos existentes no estrato.

- (A) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei aleatoriamente N vendedores

A amostragem é casual simples, pois todos os elementos da população (vendedores) tem igual probabilidade de pertencer à amostra.

3. Com a amostra ao lado, determine:

- (a) Média
(b) Moda
(c) Mediana
(d) Variância
(e) Desvio padrão
(f) Coeficiente de variação

13	11	11
12	12	12
12	12	11
13	13	13

(a) $\bar{x} = \frac{3 \cdot 11 + 5 \cdot 12 + 4 \cdot 13}{12} = 12,08$

(b) Moda é o(s) valor(es) de máxima frequência: $m_o = 12$

(c) Mediana para um conjunto par de n valores ordenados é definida como o valor médio entre os valores de ordem n/2 e (n+1)/2: $md = 12$

(d) $S^2 = \frac{3 \cdot (11-12,08)^2 + 5 \cdot (12-12,08)^2 + 4 \cdot (13-12,08)^2}{11} = 0,6288$

(e) $S = \sqrt{S^2} = \sqrt{0,6288} = 0,793$

(f) $c_v = \frac{S}{\bar{x}} = \frac{0,793}{12,08} = 0,0656 = 6,56\%$

4. Com a amostra ao lado, determine:

- (a) Média
(b) Moda
(c) Mediana
(d) Variância
(e) Desvio padrão
(f) Coeficiente de variação

6	9	10	9	6
7	6	8	8	6
9	9	9	6	8
5	9	9	6	8

(a) $\bar{x} = \frac{1 \cdot 5 + 6 \cdot 6 + 1 \cdot 7 + 4 \cdot 8 + 7 \cdot 9 + 1 \cdot 10}{20} = 7,65$

(b) Moda é o(s) valor(es) de máxima frequência: $m_o = 9$

(c) Mediana para um conjunto par de n valores ordenados é definida como o valor médio entre os valores de ordem n/2 e (n+1)/2: $md = 8$

(d) $S^2 = \frac{1 \cdot (5-7,65)^2 + 6 \cdot (6-7,65)^2 + 1 \cdot (7-7,65)^2 + 4 \cdot (8-7,65)^2 + 7 \cdot (9-7,65)^2 + 1 \cdot (10-7,65)^2}{19} = 2,239$

(e) $S = \sqrt{S^2} = \sqrt{2,239} = 1,496$

(f) $c_v = \frac{S}{\bar{x}} = \frac{1,496}{7,65} = 0,1956 = 19,56\%$

5. Com a amostra abaixo, construa um histograma.

8,9	8,3	9,5	8,9	8,8	9,6
9,8	8,3	8,1	8,7	8,9	9,1
9,0	9,4	8,8	9,9	9,0	10,1
10,0	8,3	9,1	9,9	10,2	9,9
8,5	9,0	9,0	8,4	8,9	8,0

N = número total de dados disponíveis = 30

Máximo = 10,2

Mínimo = 8

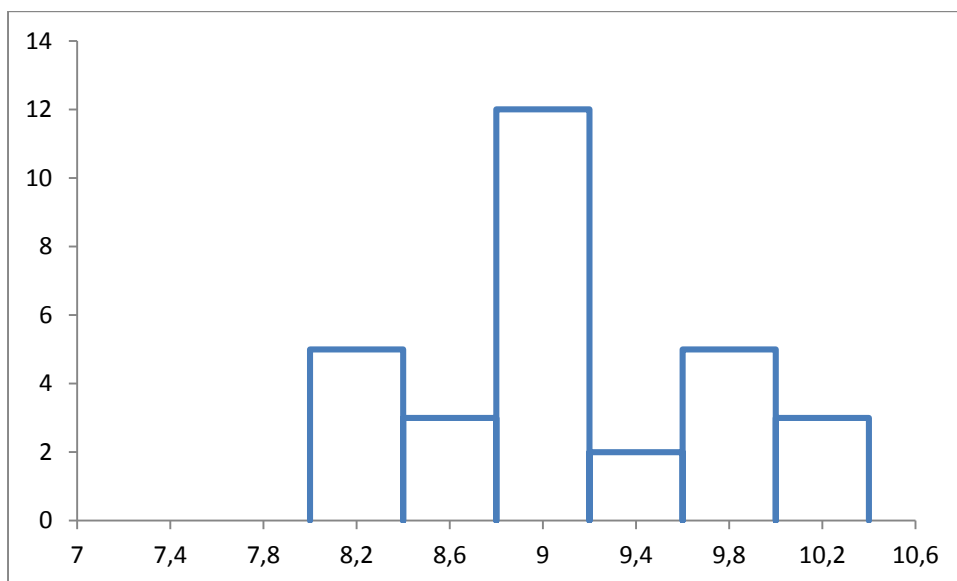
R = amplitude do conjunto de dados = máximo – mínimo = 2,2

k = número de classes. Para defini-la, podemos utilizar a Regra de Sturges: $k = 1 + 3,322(\log_{10} N)$

$k = 1 + 3,322(\log_{10} 30) = 1 + 3,322 \cdot 1,477 = 5,907 \approx 6$

h = amplitude das classes $\approx \frac{R}{k} = \frac{2,2}{6} = 0,37$. Para uma melhor visualização, adotou-se $h = 0,4$.

Classe	x_i	f_i
8,0 --- 8,4	8,2	5
8,4 --- 8,8	8,6	3
8,8 --- 9,2	9,0	12
9,2 --- 9,6	9,4	2
9,6 --- 10,0	9,8	5
10,0 --- 10,4	10,2	3



6. Com a amostra abaixo, construa um histograma e polígono de frequência. (Classes de frequência)

20,9	21,8	30,5	23,3	22,0	21,8	25,5	20,7
26,2	21,6	20,9	24,9	20,0	24,0	29,7	25,5
22,4	24,0	23,0	25,1	24,5	20,9	22,9	21,4
19,6	28,0	24,5	24,8	23,2	22,3	24,2	24,1
21,9	24,1	20,4	22,5	23,5	21,6	23,0	26,0
21,2	22,2	22,6	27,1	23,9	22,0	22,7	26,7
20,9	25,1	27,0	22,8	22,1	20,8	19,9	21,3

N = número total de dados disponíveis = 56

Máximo = 30,5

Mínimo = 19,6

R = amplitude do conjunto de dados = máximo – mínimo = 10,9

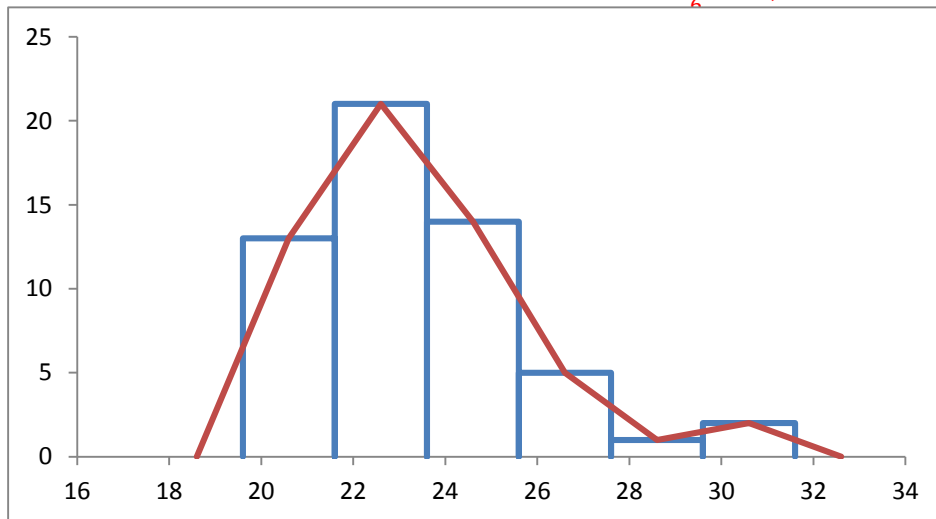
k = número de classes. Para defini-la, podemos utilizar a Regra de Sturges: $k = 1 + 3,322(\log_{10} N)$

$k = 1 + 3,322(\log_{10} 56) = 1 + 3,322 \cdot 1,748 = 6,807 \approx 7$

h = amplitude das classes $\approx \frac{R}{k} = \frac{10,9}{7} = 1,6$. Para uma melhor visualização, adotou-se h = 2.

Classe	x_i	f_i	F_i
19,6 --- 21,6	20,6	13	13
21,6 --- 23,6	22,6	21	34
23,6 --- 25,6	24,6	14	48
25,6 --- 27,6	26,6	5	53
27,6 --- 29,6	28,6	1	54
29,6 --- 31,6	30,6	2	56

$h = \frac{13,6}{6} = 2,267$. Para uma melhor visualização,



7. Com a amostra ao lado, construa um histograma, polígono de frequência e gráfico de frequência acumulada. (Classes de frequência)

N = número de dados disponíveis = 36

Máximo = 26,2

Mínimo = 12,6

R = amplitude do conjunto de dados

R = máximo – mínimo = 13,6

k = número de classes. Para defini-la, podemos utilizar a Regra de Sturges:

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

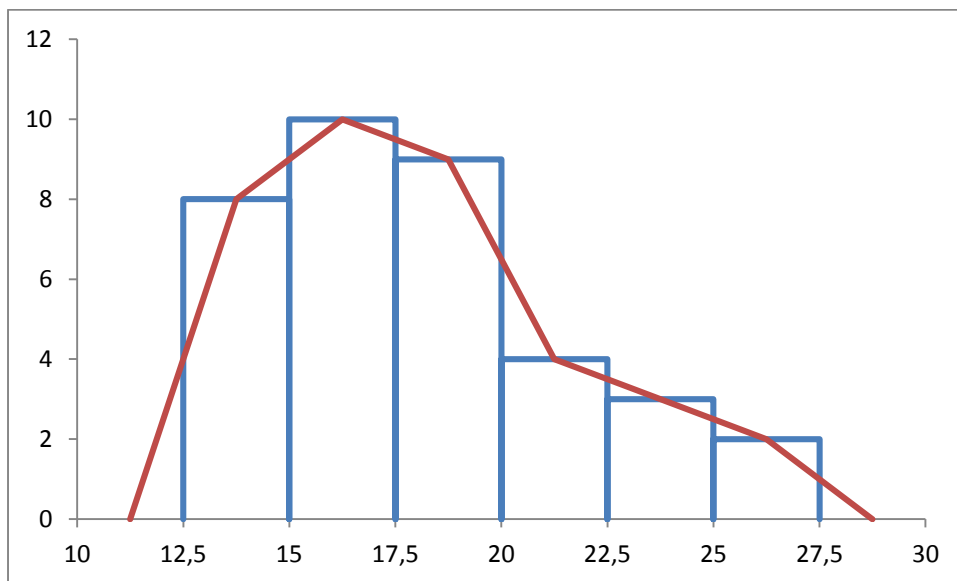
$k = 1 + 3,322(\log_{10} 36)$

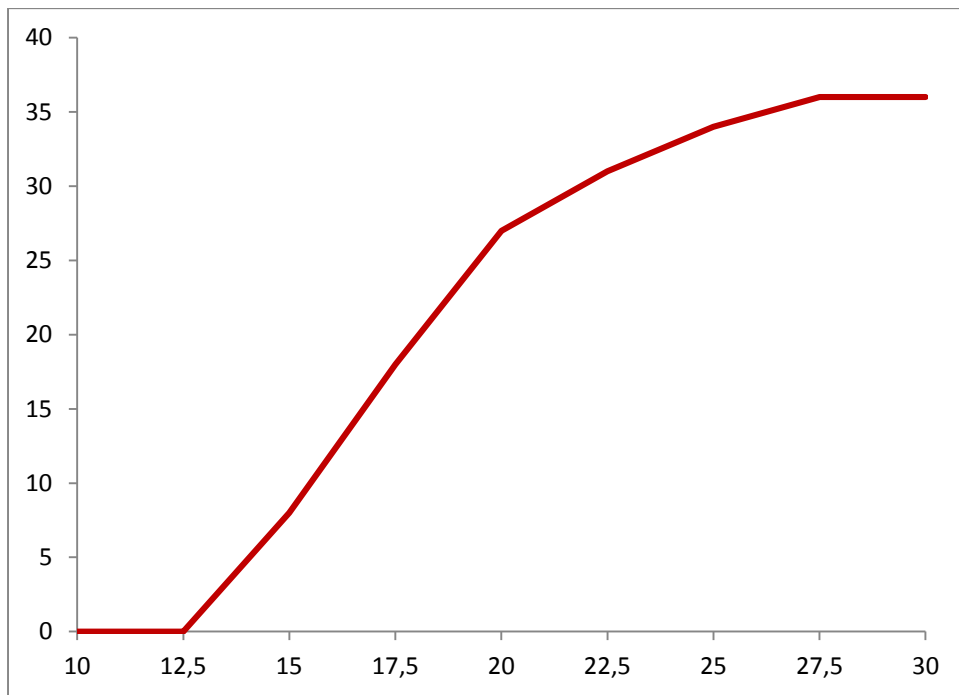
$k = 1 + 3,322 \cdot 1,556 = 6,169 \approx 6$

h = amplitude das classes $\approx \frac{R}{k}$

18,9	15,8	17,0	12,6
24,2	18,9	16,1	16,4
17,8	15,9	17,6	12,9
26,2	18,6	18,7	17,6
21,7	17,8	14,0	20,3
23,5	16,0	23,5	14,1
22,3	16,1	18,1	14,4
25,0	13,4	16,8	12,7
12,6	15,2	20,2	15,2

Classe	x_i	f_i	F_i
12,5 --- 15,0	13,75	8	8
15,0 --- 17,5	16,25	10	18
17,5 --- 20,0	18,75	9	27
20,0 --- 22,5	21,25	4	31
22,5 --- 25,0	23,75	3	34
25,0 --- 27,5	26,25	2	36





Classe	x_i	f_i	F_i
1 9,6 --- 11,8	10,7	5	5
2 11,8 --- 14,0	12,9	13	18
3 14,0 --- 16,2	15,1	11	29
4 16,2 --- 18,4	17,3	17	46
5 18,4 --- 20,6	19,5	13	59
6 20,6 --- 22,8	21,7	11	70
7 22,8 --- 25,00	23,9	1	71
Total		71	

8. Considere o resumo dos dados de uma amostra ao lado (classes de frequência). Determine:
- Média
 - Moda
 - Mediana
 - Variância
 - Desvio padrão
 - Coefficiente de variação

$$(a) \bar{x} = \frac{10,7 \cdot 5 + 12,9 \cdot 13 + 15,1 \cdot 11 + 17,3 \cdot 17 + 19,5 \cdot 13 + 21,7 \cdot 11 + 23,9 \cdot 1}{71} = 16,9$$

(b) Classe modal: 16,2 |---| 18,4

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

L_i = limite inferior da classe modal = 16,2

d_1 = diferença entre a frequência da classe modal e a da classe anterior = 17 - 11 = 6

d_2 = diferença entre a frequência da classe modal e a da classe seguinte = 17 - 13 = 4

h = amplitude das classes = 2,2

$$m_o = 16,2 + \frac{6}{6+4} \cdot 2,2 = 17,5$$

(c) Classe que contém a mediana: 16,2 |--- 18,4

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

L_i = limite inferior da classe que contém a mediana = 16,2

n = número de elementos do conjunto de dados = 71

F_a = soma das frequências das classes anteriores à que contém a mediana = 29

f_{md} = frequência da classe que contém a mediana = 17

h_{md} = amplitude da classe que contém a mediana = 2,2

$$m_d = 16,2 + \frac{35,5 - 29}{17} \cdot 2,2 = 17,0$$

(d) Cálculo do S^2 :

$$\frac{5 \cdot (10,7 - 16,9)^2 + 13 \cdot (12,9 - 16,9)^2 + 11 \cdot (15,1 - 16,9)^2 + 17 \cdot (17,3 - 16,9)^2 + 13 \cdot (19,5 - 16,9)^2 + 11 \cdot (21,7 - 16,9)^2 + 1 \cdot (23,9 - 16,9)^2}{70}$$

$$S^2 = 11,84$$

$$(e) S = \sqrt{S^2} = \sqrt{11,84} = 3,441$$

$$(f) c_v = \frac{S}{\bar{x}} = \frac{3,441}{16,9} = 0,204 = 20,4\%$$

9. Considere o resumo dos dados de uma amostra abaixo (classes de frequência). Determine

- Média
- Moda
- Mediana
- Variância
- Desvio padrão
- Coeficiente de variação

Classe	x_i	f_i	F_i
1 4,7 --- 6,9	5,8	1	1
2 6,9 --- 9,1	8,0	4	5
3 9,1 --- 11,3	10,2	16	21
4 11,3 --- 13,5	12,4	41	62
5 13,5 --- 15,7	14,6	72	134
6 15,7 --- 17,9	16,8	100	234
7 17,9 --- 20,1	19,0	74	308
8 20,1 --- 22,3	21,2	43	351
9 22,3 --- 24,5	23,4	10	361
10 24,5 --- 26,7	25,6	3	364
		364	

$$(a) \bar{x} = \frac{5,8 \cdot 1 + 8,0 \cdot 4 + 10,2 \cdot 16 + 12,4 \cdot 41 + 14,6 \cdot 72 + 16,8 \cdot 100 + 19,0 \cdot 74 + 21,2 \cdot 43 + 23,4 \cdot 10 + 25,6 \cdot 3}{364} = 16,67$$

(b) Classe modal: 15,7 |--- 17,9

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

L_i = limite inferior da classe modal = 15,7

d_1 = diferença entre a frequência da classe modal e a da classe anterior = $100 - 72 = 28$

d_2 = diferença entre a frequência da classe modal e a da classe seguinte = $100 - 74 = 26$

h = amplitude das classes = 2,2

$$m_o = 15,7 + \frac{28}{28 + 26} \cdot 2,2 = 16,84$$

(c) Classe que contém a mediana: 15,7 |--- 17,9

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

L_i = limite inferior da classe que contém a mediana = 15,7

n = número de elementos do conjunto de dados = 364

F_a = soma das frequências das classes anteriores à que contém a mediana = 134

f_{md} = frequência da classe que contém a mediana = 100

h_{md} = amplitude da classe que contém a mediana = 2,2

$$m_d = 15,7 + \frac{182 - 134}{100} \cdot 2,2 = 16,76$$

(d) Cálculo do S^2 :

$$\frac{1 \cdot (5,8 - 16,67)^2 + 4 \cdot (8,0 - 16,67)^2 + 16 \cdot (10,2 - 16,67)^2 + 41 \cdot (12,4 - 16,67)^2 + 72 \cdot (14,6 - 16,67)^2 + 100 \cdot (16,8 - 16,67)^2 + 74 \cdot (19,0 - 16,67)^2 + 43 \cdot (21,2 - 16,67)^2 + 10 \cdot (23,4 - 16,67)^2 + 3 \cdot (25,6 - 16,67)^2}{363}$$

$$S^2 = 11,36$$

$$(e) S = \sqrt{S^2} = \sqrt{11,36} = 3,370$$

$$(f) c_v = \frac{S}{\bar{x}} = \frac{3,370}{16,67} = 0,2021 = 20,21\%$$

10. Aborde 40 pessoas divididas em 2 grupos de 20. Para o primeiro grupo faça as seguintes perguntas:

(a) O Rio Amazonas tem mais ou menos de 400 Km?

(b) Qual a extensão do Rio Amazonas?

Tabule as respostas da pergunta. Calcule a média, mediana, variância, desvio padrão e coeficiente de variação. Construa também um histograma e polígono de frequências.

Para o segundo grupo faça as seguintes perguntas:

(a) O Rio Amazonas tem mais ou menos de 4.000 Km?

(b) Qual a extensão do Rio Amazonas?

Tabule as respostas da pergunta. Calcule a média, mediana, variância, desvio padrão e coeficiente de variação. Construa também um histograma e polígono de frequências.

Compare os resultados obtidos nos dois grupos. Existe alguma diferença (aparentemente) relevante? O que a explicaria?

Gabarito

Q1

- (NP) Todos os meus colegas da faculdade (tenho telefone e email de todos eles)
- (NP) Fiquei na única porta de entrada da escola abordando todos os meus conhecidos
- (P) Fiquei na única porta de entrada da escola e cada 12 pessoas que entravam, eu abordava uma
- (P) Consegui uma lista de todos os alunos das escolas com uma ordenação aleatória, e selecionei os 20 primeiros da lista
- (P) Consegui uma lista de todos os alunos das escolas em ordem alfabética. Gerei 20 números aleatórios. Selecionei da lista de alunos aqueles que ocupavam posições equivalentes aos números aleatórios gerados

Q2

- (B) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei todos vendedores que ocupavam posições múltiplas de 15 (15ª posição, 30ª posição, 45ª posição, 60ª posição, 75ª posição, 90ª posição, 105ª posição, etc)
- (D) Escolhi casualmente 3 lojas da rede. A amostra foi composta de todos os vendedores que atualmente em cada uma destas 3 lojas.
- (C) Em cada uma das lojas, identifiquei todos os vendedores (lista de vendedores por loja). Selecionei aleatoriamente k vendedores da loja. Onde k é um número inteiro proporcional à quantidade de vendedores da loja
- (A) Lista de todos os vendedores (que atuam em todas as lojas da rede). Selecionei aleatoriamente N vendedores

Q3

Média	12,08
Moda	12
Mediana	12
Variância	0,63
Desvio padrão	0,79
Coeficiente de variação	6,56%

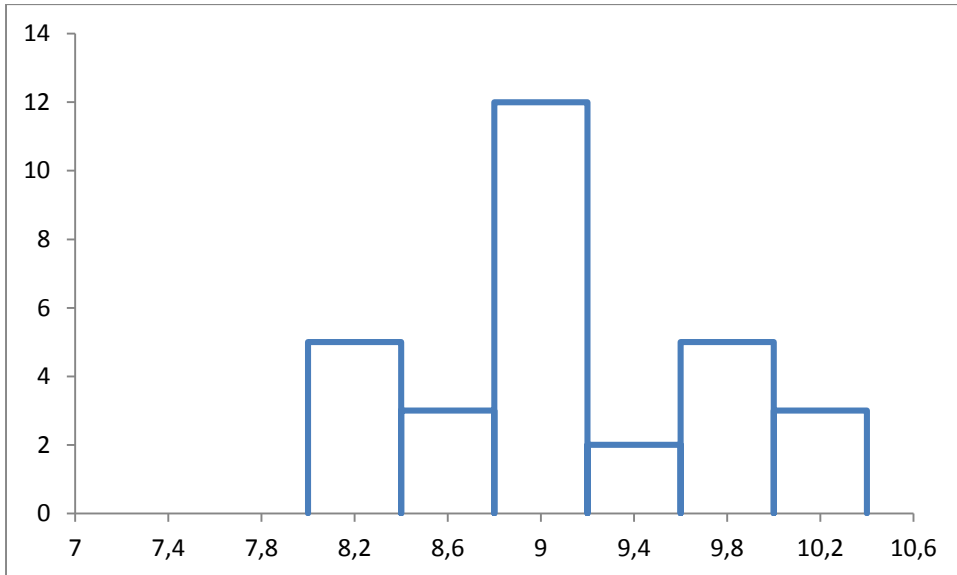
Q4

Média	7,65
Moda	9
Mediana	8
Variância	2,24
Desvio padrão	1,50
Coeficiente de variação	19,56%

Q5

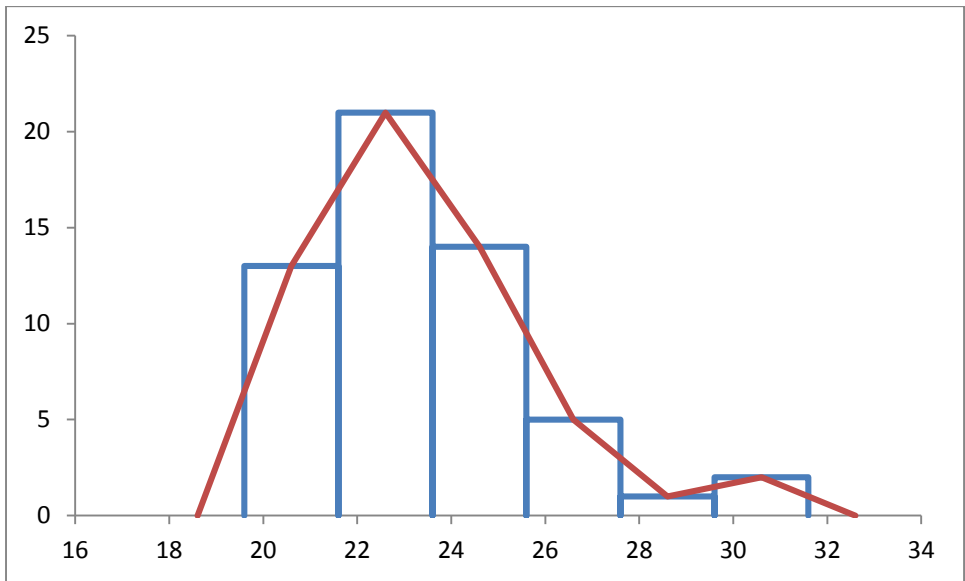
N	30
Máximo	10,2
Mínimo	8
Amplitude	2,2
k	6 (regra de Sturges)
h	0,4 0,37

Classe	x_i	f_i
8,0 ---8,4	8,2	5
8,4 ---8,8	8,6	3
8,8 ---9,2	9	12
9,2 ---9,6	9,4	2
9,6 ---10	9,8	5
10,0 ---10,4	10,2	3



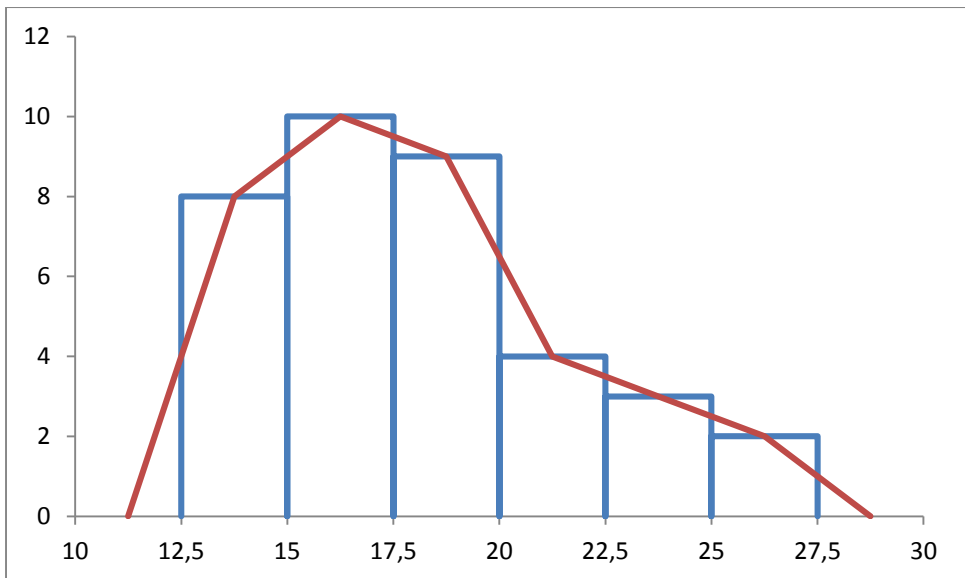
Q6

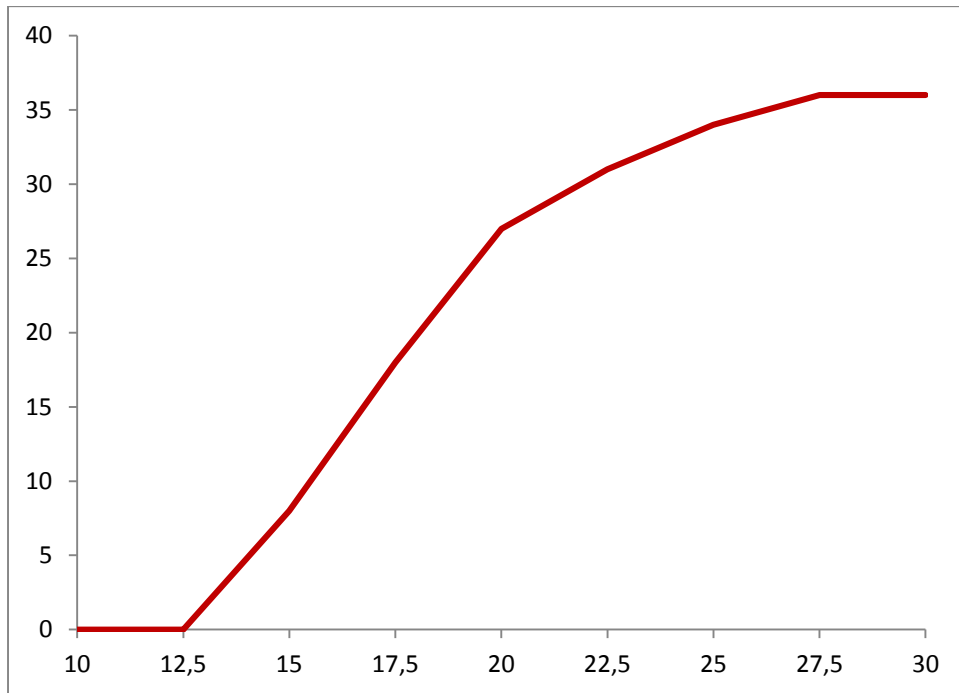
Classe	x_i	f_i	F_i
1 19,6 ---21,6	20,6	13	13
2 21,6 ---23,6	22,6	21	34
3 23,6 ---25,6	24,6	14	48
4 25,6 ---27,6	26,6	5	53
5 27,6 ---29,6	28,6	1	54
6 29,6 ---31,6	30,6	2	56



Q7

Classe	x_i	f_i	F_i
1 12,5 ---15,0	13,75	8	8
2 15,0 ---17,5	16,25	10	18
3 17,5 ---20,0	18,75	9	27
4 20,0 ---22,5	21,25	4	31
5 22,5 ---25,0	23,75	3	34
6 25,0 ---27,5	26,25	2	36





Q8

Média	16,9
Moda	17,0
Mediana	17,5
Variância	11,84
Desvio Padrão	3,441
Coefficiente de Variação	20,40%

Q9

Média	16,67
Moda	16,76
Mediana	16,84
Variância	11,36
Desvio Padrão	3,370
Coefficiente de Variação	20,21%