

Primeira Lista de Exercícios

- 1- Em 1997, um aluno da UFBA, João, decidiu fazer um levantamento da proporção de estudantes segundo a cor: para isso ele entrevistou 1000 alunos nos campi da UFBA e pediu para o estudante declarar a sua cor. João obteve a tabela de distribuição de frequência da tabela 1. Naquele ano, o IBGE divulgou o resultado do PNAD (pesquisa nacional por amostragem de domicílio) com a tabela de distribuição de frequência da tabela 2. Construa o gráfico de barras para a variável Cor e comente os resultados obtidos por João.

Cor	frequência	frequência relativa	porcentagem
Branca	508	0,508	50,800%
Preta	80	0,080	8,000%
Parda	346	0,346	34,600%
Amarela	30	0,030	3,000%
Indígena	36	0,036	3,600%
Total	1000	1,000	100,000%

Tabela 1: Tabela de distribuição de frequência obtida por João.

Cor	porcentagem
Branca	21,0%
Preta	19,9%
Parda	58,5%
Amarela	0,4%
Indígena	0,2%
Total	100,0%

Tabela 2: Tabela de distribuição de frequência obtida pelo IBGE no PNAD de 1997.

- 2- Um pesquisador com interesse em estudar o perfil socioeconômico da seção de orçamentos da companhia MB, coletou uma amostra com 36 funcionários e obteve a amostra armazenada no arquivo `companhia_MB.xlsx`. Construa a tabela de distribuição de frequência a variável “Procedência” e construa o gráfico de barras. Interprete os resultados.
- 3- Vinte e uma pacientes de uma clínica médica tiveram o seu nível de potássio no plasma medido. Os resultados estão na tabela 3.
- (a) Construa o histograma e interprete.
- (b) Qual a porcentagem dos valores que estão acima do nível 3?
- 4- Foram feitas medidas em operários da construção civil a respeito da taxa de hemoglobina no sangue (em gramas/cm³). Os resultados estão na tabela 4.
- (a) Organize os dados em faixas de tamanho 1 a partir de 11.
- (b) Construa o histograma.

Nível	Frequência
2,25 — — — — 2,55	1.00
2,55 — — — — 2,75	3.00
2,75 — — — — 2,95	2.00
2,95 — — — — 3,15	4.00
3,15 — — — — 3,35	5.00
3,35 — — — — 3,65	6.00

Tabela 3: Nível de potássio no plasma.

11.1	12.5	14.4	12.6	12.6	13.2	15.8	12.7	15.4	12.3
12.2	13.9	13.6	11.3	13.4	13.0	14.7	12.3	16.3	13.7
11.7	12.3	12.7	11.7	15.2	16.9	13.5	13.5	15.2	14.1

Tabela 4: Taxa de hemoglobina no sangue.

(c) Taxas abaixo de 12 ou acima de 16 são consideradas alteradas e requerem acompanhamento médico. Obtenha a tabela de distribuição de frequências da variável acompanhamento médico com duas opções: sim ou não.

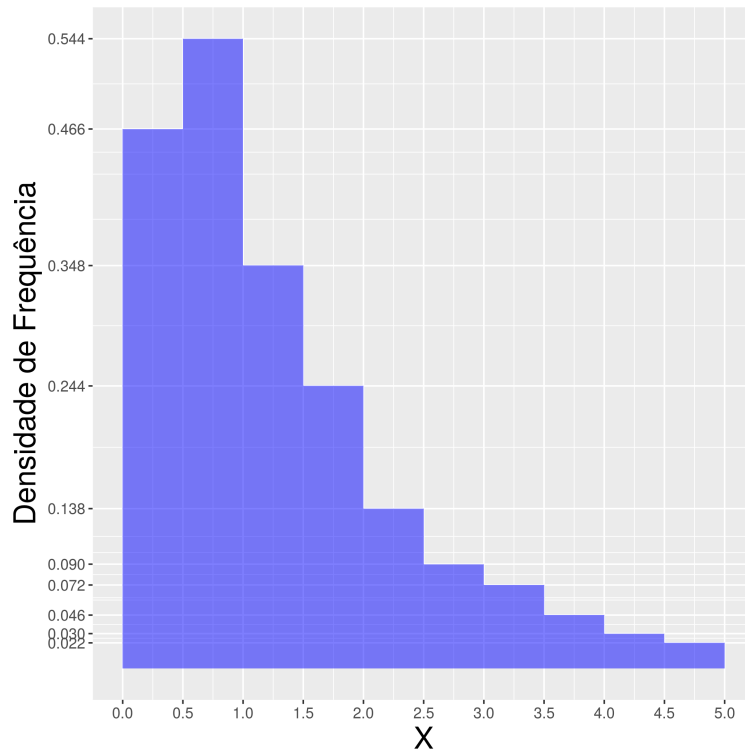
- 5- Considere as notas finais (X) da turma 1 de Estatística Aplicada à Saúde: 4,4; 5,2; 5,3; 5,6; 6,1; 6,4; 7,6; 7,6; 8,0; 8,1; 8,2; 8,9; 9,0; 9,1; 9,8. Calcule o primeiro quartil, o segundo quartil, o terceiro quartil e o quantil de ordem 23%. Construa o diagrama de caixa e calcule o coeficiente de Bowley. Você julga que esta turma tem assimetria? Justifique a sua resposta.
- 6- Construa o diagrama de caixa e calcule o coeficiente de Bowley para as 15 maiores cidades do Brasil segundo o IBGE (em 10.000 habitantes). Existe algum ponto exterior nesse conjunto de dados? Você julga que a variável população é assimétrica? Justifique a sua resposta.

Município	População
São Paulo	1125,4
Rio de Janeiro	632
Salvador	267,6
Brasília	257
Fortaleza	245,2
Belo Horizonte	235,5
Manaus	180,2
Curitiba	175,2
Recife	153,8
Porto Alegre	140,9
Belém	139,3
Goiânia	130,2
Guarulhos	122,2
Campinas	108
São Gonçalo	100

7- Considere uma amostra com 1000 indivíduos de uma determinada população. Nessa amostra, coletamos a variável quantitativa contínua X e construímos o histograma apresentado na Figura 1.

- (a) Construa a tabela de distribuição de frequências.
- (b) Construa o diagrama de caixa e o coeficiente de Bowley. Você julga que existe assimetria na variável quantitativa contínua X .

Figura 1: Histograma para a variável X.



- (c) Calcule o quantil de ordem 15%.
- 8- Considere os alunos de duas turmas de Estatística Básica em um dado semestre da UFBA. O gráfico da Figura 2 mostra o gráfico de barras da variável “Idade”.
- Construa a tabela de distribuição de frequências para a variável quantitativa discreta “Idade”.
 - Construa o diagrama de caixa e o coeficiente de Bowley. Você julga que existe assimetria na variável quantitativa discreta “Idade”.
 - Qual a idade classifica um aluno entre os 20% mais jovens.
 - Qual a idade classifica um aluno entre os 15% mais velhos.
- 9- Alunos da Escola de Educação Física foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiram correr sem parar. Os dados estão apresentados na tabela 5.

Tabela 5: Resistência dos alunos da Escola de Educação Física.

Faixas	Frequências
0 — — — 4	438
4 — — — 8	206
8 — — — 12	125
12 — — — 16	22
16 — — — 20	9

- Qual é a variável em estudo?
 - Construa o histograma.
 - Determine o primeiro, o segundo e o terceiro quartil.
 - Construa o diagrama de caixa e calcule o coeficiente de Bowley. Você diria que a variável do item (a) é simétrica?
- 10- Vinte e uma pacientes de uma clínica médica tiveram o seu nível de potássio no plasma medido. Os resultados estão descritos na tabela 6.

Figura 2: Idade para os alunos das duas turmas de Estatística Básica Básica.

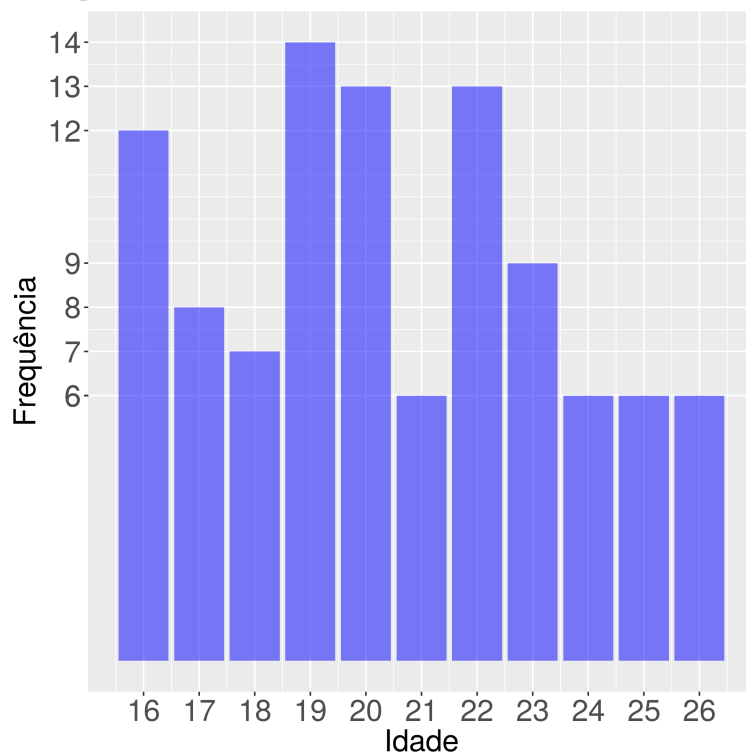


Tabela 6: Nível de potássio no plasma.

Nível	Frequência
2,25 — — — 2,55	1
2,55 — — — 2,75	3
2,75 — — — 2,95	2
2,95 — — — 3,15	4
3,15 — — — 3,35	5
3,35 — — — 3,65	6

- Construa o histograma.
- Determine o primeiro, o segundo e o terceiro quartil.
- Construa o diagrama de caixa e calcule o coeficiente de Bowley. Você diria que existe assimetria nesta variável? Justifique a sua resposta.
- Qual a porcentagem de valores que estão acima do nível 3.

- 11- Foram feitas medições em operários da construção civil a respeito da taxa de hemoglobina no sangue (em gramas/cm³) com resultados na tabela 7.

Tabela 7: Taxa de hemoglobina nos operários.

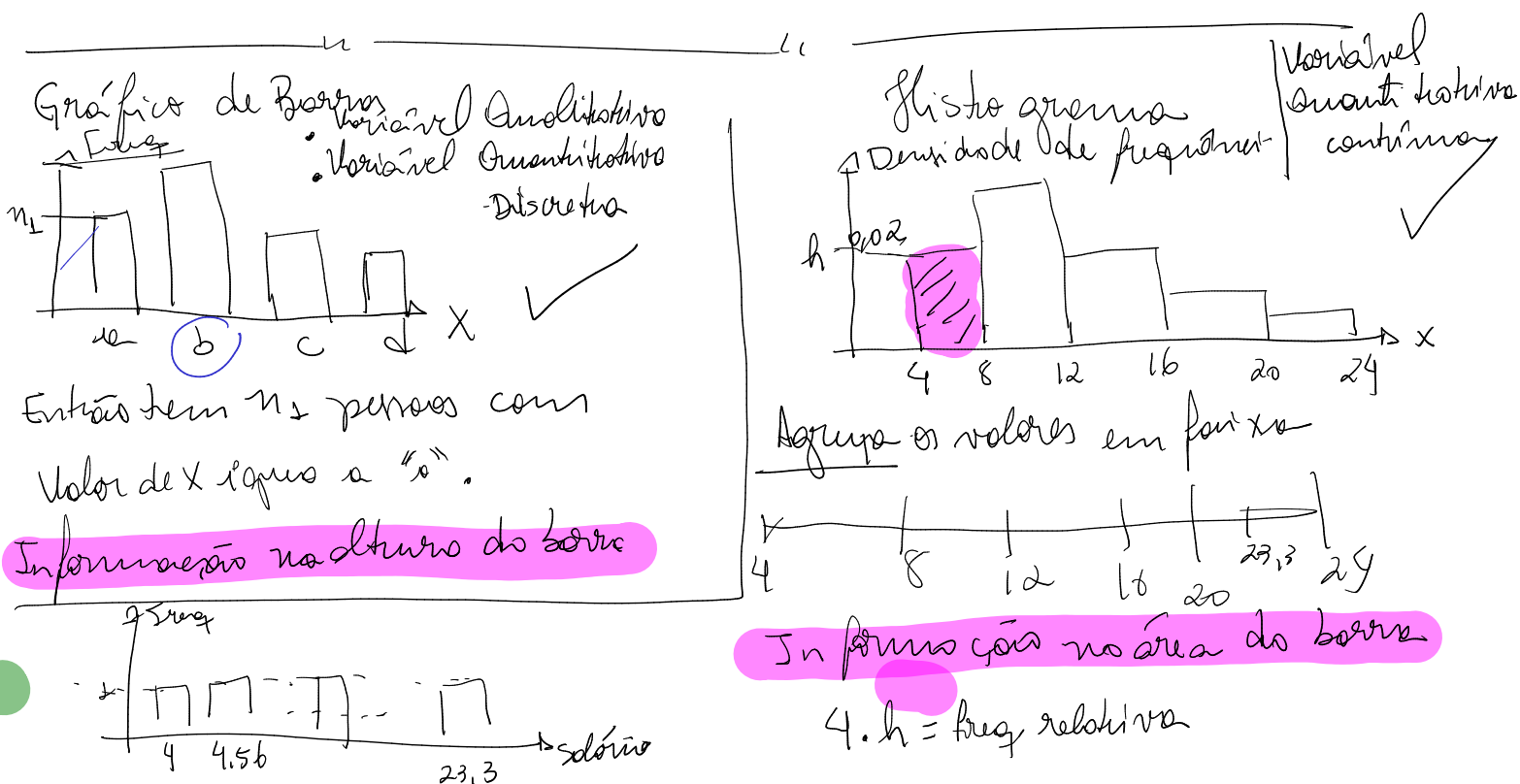
11,1	12,2	11,7	12,5	13,9	12,3	14,4	13,6	12,7	12,6
11,3	11,7	12,6	13,4	15,2	13,2	13,0	16,9	15,8	14,7
13,5	12,7	12,3	13,5	15,4	16,3	15,2	12,3	13,7	14,1

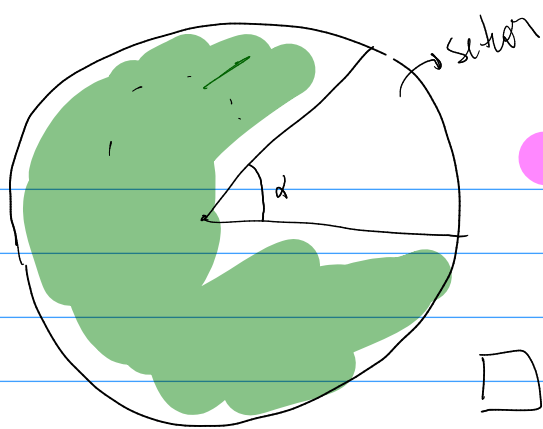
- (a) Organize os dados em faixas de tamanho 1 a partir de 11.
 (b) Construa o histograma.
 (c) Determine o primeiro, o segundo e o terceiro quartil. (Use os dados brutos)
 (d) Construa o diagrama de caixa e calcule o coeficiente de Bowley. Você diria que a taxa de hemoglobina nos operários tem assimetria? Justifique a sua resposta.
- 12- A tabela 8 apresenta as frequências relativas de ocorrências de faixas de altura (em cm) para uma amostra de 100 crianças de 12 anos de idade.

Tabela 8: Altura de 100 crianças com 12 anos.

Faixas	Frequência relativa
100 — — — — 110	0,10
110 — — — — 120	0,25
120 — — — — 130	0,30
130 — — — — 140	0,25
140 — — — — 160	0,10

- (a) Construa o histograma.
 (b) Construa o diagrama de caixa e calcule o coeficiente de Bowley. Você julga que a altura das crianças tem assimetria? Justifique a sua resposta.
 (c) Desejando-se separar os 15% mais altos, qual seria o ponto de corte?



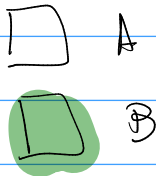


$$h = \frac{\text{freq. Relativa}}{4}$$

Informações no abertura do setor

$$360 \cdot \text{freq. Relativa} = \alpha$$

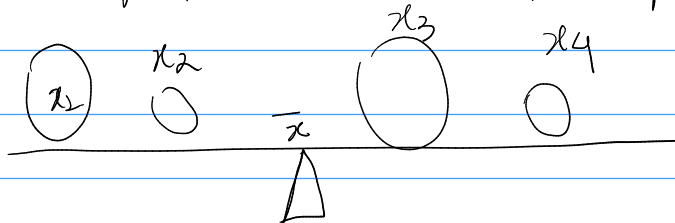
(Gráfica de setores)



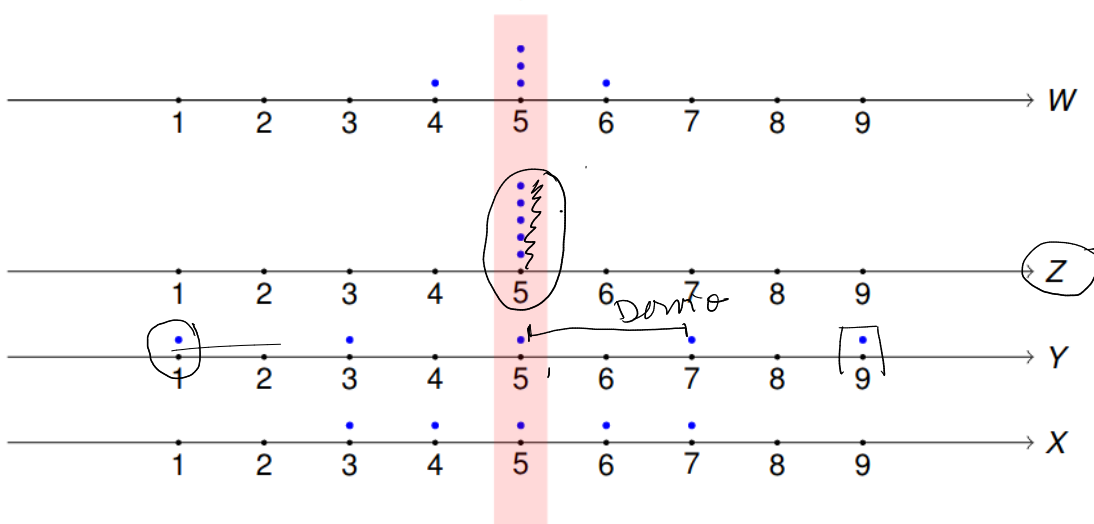
Média: $x_1 + \dots + x_n$

n

medidas de posição: "um valor que representa todos numos"



Média, Moda e Mediana



5 pessoas fizeram os testes x, z, y, w

$\text{Var}(x) = 2$
 $\text{DP}(x) = 1, 4$

$\text{Var}(y) = 8$
 $\text{DP}(y) = 2, 8$

$\text{Var}(z) = 0$
 $\text{DP}(z) = 0$

$\text{Var}(w) = 0, 4$
 $\text{DP}(w) = 0, 6$

Olhar a média não basta! ~ Média | Mediana | Moda não captura a distribuição dos valores

Varianças:

- 1) Calcular os desvios: $(x_i - \bar{x})$
- 2) Média dos desvios ao quadrado / "Problema do sinal"

Ex:
$$\begin{array}{c|c} 3 & 5 & 7 \\ \hline \bar{x} = 5 \end{array} \quad \left. \begin{array}{l} (3-5) = -2 \\ (5-5) = 0 \\ (7-5) = 2 \end{array} \right\} \quad \frac{-2+0+2}{3} = 0$$

Média dos Desvios sempre é 0 !!

$$\begin{array}{l} (3-5)^2 = 4 \\ (5-5)^2 = 0 \\ (7-5)^2 = 4 \end{array} \quad \frac{4+0+4}{3} = 2,67$$

$$Var(X) = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \rightarrow \frac{\text{cm}^2}{\text{unidade}^2}$$

$$dp(X) = \sqrt{Var(X)} = \sqrt{\text{cm}^2} = \text{cm} // \rightarrow \text{no mesmo unidade de dos dados}$$

Interpretação: $dp(X)$ é pequena \Rightarrow os dados são + homogêneos
 $dp(X)$ é grande \Rightarrow os dados são + heterogêneos

\rightarrow Medida de posição.

Modo: Valor que tem maior frequência "ponto de máximo".

\rightarrow Variável Qualitativa ou Variável Quantitativa Discreta

Mediana: Medida de posição ("um valor que representa bem td mundo").

\rightarrow Valor no meio de sequência ordenada.

$$\frac{3+1}{2} = 2$$

$$Md(X) = x_{(2)} = 3$$

Ex: 3; 1; 25

$$\frac{1}{x_{(1)}}$$

$\downarrow x_{(2)}$
3
9
Mediana //

$x_{(3)}$
25

tamanho da amostra
ímpar

Ex: 7, 18, 11, 3

$$\begin{array}{c} \text{---} \bullet x_{(2)} \\ 3 \quad 7 \quad 11 \quad 18 \\ \text{---} \bullet x_{(3)} \end{array} \quad \left| \quad \frac{7+11}{2} = 9 = Md(X) \right.$$

tamanho da amostra é par

$$\frac{(n+1)}{2} = \frac{4+1}{2} = 2,5$$

$$\lfloor 2,5 \rfloor = 2; \lceil 2,5 \rceil = 3$$

$$\frac{7+11}{2} = 9$$

Amostra: x_1, x_2, \dots, x_n

Notação: $x_{(1)}$: Menor n^o

$x_{(2)}$: Menor n^o sem contar $x_{(1)}$

$x_{(3)}$: Menor n^o sem contar $x_{(1)}$ e $x_{(2)}$

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)}$$

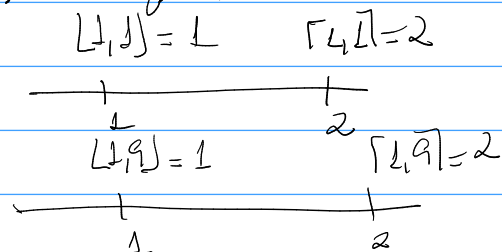
↳ Estatística de ordem 1

$$Md(x) = \begin{cases} x_{(\frac{n+1}{2})}, & \text{se } n \text{ é ímpar} \end{cases}$$

$$\frac{x_{(\lfloor \frac{n+1}{2} \rfloor)} + x_{(\lceil \frac{n+1}{2} \rceil)}}{2}, \text{ se } n \text{ é par}$$

$\lfloor \rfloor$: arredondando para baixo

$\lceil \rceil$: arredondando para cima



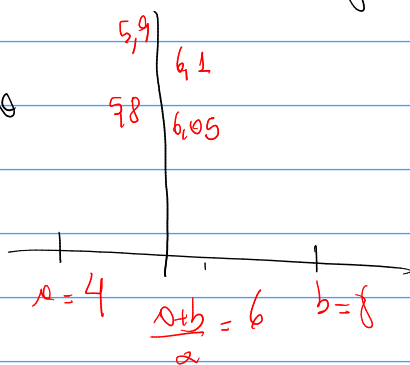
X	Densidade de frequência	Delta	Frequência relativa	Porcentagem	Frequência
[0; 0,5)	0,466	0,5	0,233	23,30%	233
[0,5; 1)	0,544	0,5	0,272	27,20%	272
[1; 1,5)	0,348	0,5	0,174	17,40%	174
[1,5; 2)	0,244	0,5	0,122	12,20%	122
[2; 2,5)	0,138	0,5	0,069	6,90%	69
[2,5; 3)	0,09	0,5	0,045	4,50%	45
[3; 3,5)	0,072	0,5	0,036	3,60%	36
[3,5; 4)	0,046	0,5	0,023	2,30%	23
[4; 4,5)	0,03	0,5	0,015	1,50%	15
[4,5; 5]	0,022	0,5	0,011	1,10%	11
total	--	--	1	100,00%	1000

Altura da barra no histograma Base da barra no histograma

Se tenho apenas a tabela de distribuição de frequência?

x freq f_{rel} $\frac{100 \cdot f_{rel}}{a}$ $\frac{a+b}{2}$
 $[a, b)$ n_i f_i $\%$ $\frac{a+b}{2}$
 $\rightarrow n_i$ pessoas com valor igual a $\frac{a+b}{2}$

Suposição Assumindo



Mediana de x :

$$x_{(1)} = x_{(2)} = x_{(3)} = \dots = x_{(233)} = 0,25$$

$$x_{(234)} = x_{(235)} = \dots = x_{(233+272)} = 0,75 = x_{(505)}$$

$$x_{(506)} = x_{(507)} = \dots = x_{(505+174)} = 1,25 = x_{(679)}$$

$$x_{(680)} = x_{(681)} = \dots = x_{(679+122)} = 1,75 = x_{(801)}$$

$$x_{(802)} = x_{(803)} = \dots = x_{(801+69)} = 2,25 = x_{(870)}$$

$$X_{(871)} = \dots = X_{(870+45)} = 2,75 = X_{(915)}$$

$$X_{(915)} = 3,25 = \dots = X_{(915+36)} = X_{(951)}$$

$$X_{(951)} = 3,75 = \dots = X_{(951+23)} = X_{(974)}$$

$$X_{(974)} = 4,25 = \dots = X_{(974+15)} = X_{(989)}$$

$$X_{(989)} = 4,75 = \dots = X_{(1000)}$$

~~~~~

$$X_{(1)} = \dots = X_{(232)} = 0,25$$

$$X_{(234)} = 0,75 = \dots = X_{(506)}$$

$$X_{(506)} = 1,25 = \dots = X_{(679)}$$

$$X_{(679)} = 1,75 = \dots = X_{(801)}$$

$$X_{(801)} = 2,25 = \dots = X_{(870)}$$

$$X_{(870)} = 2,75 = \dots = X_{(915)}$$

$$X_{(915)} = 3,25 = \dots = X_{(951)}$$

$$X_{(951)} = 3,75 = \dots = X_{(974)}$$

$$X_{(974)} = 4,25 = \dots = X_{(989)}$$

$$X_{(989)} = 4,75 = \dots = X_{(1000)}$$

$$n = 100 \text{ exps}$$

$$\frac{(1000+1)}{2} = 500,5$$

$$\lfloor 500,5 \rfloor = 500; \lceil 500,5 \rceil = 501$$

$$M_d(X) = \frac{X_{(500)} + X_{(501)}}{2} = \frac{0,75 + 0,75}{2}$$

$$M_d(X) = 0,75$$