ESTATÍSTICA

Michelle Hanne Soares de Andrade
michellehanne.andrade@gmail.com

Estatística Descritiva

Exercício

- Um borracheiro anotou a vida útil dos pneus dos carros de seus clientes

2,4	5,2	3,5	4,9	3,5	6
4,5	3,2	3,7	2,9	4,7	5,2
3,8	5,4	6,2	5,5	4,5	6

- Faça um Diagrama de Ramos e Folhas, calcule a Amplitude e faça a Tabela de Frequência.

Exercício

Diagrama de Ramo e de Folha

Ramo	Folha	Freq
2	4 9	2
3	2 5 5 7 8	5
4	5 5 7 9	4
5	2 2 4 5	4
6	0 0 2	3

Exercício

Tabela de Frequência

t	Freq Absoluta	Freq Relativa
$2{,}4 \vdash 3{,}375$	6	0,33
$3{,}375 \vdash 4{,}35$	1	0,06
$4{,}35 \vdash 5{,}325$	6	0,33
$5{,}325 \vdash 6{,}3$	5	0,28
Total	18	1

- $K = \sqrt{18} \cong 4{,}243$
- $Amplitude = \frac{6{,}3-2{,}4}{4} = 0{,}975$

Moda (Mo)

- Moda é o valor que ocorre com maior frequência em uma distribuição de valores.
- Ex 1: A série 7, 8, 9, 10, 10, 10, 11, 12, 13, 15 tem Mo=10.
- Ex 2: 3, 5, 8, 10, 12, 13 é amodal (não tem moda).
- Ex 3: 2, 3, 4, 4, 4, 5, 6, 7, 7, 7, 8, 9 é bimodal (tem duas modas).

Moda (Mo)

- Moda de dados agrupados sem intervalo de classe:

Nº DE MENINOS	f_i
0	2
1	6
2	10
3	12
4	4
	$\Sigma = 34$

Mo = 3,

(pois 3 tem frequência 12)

Moda (Mo)

- Moda de dados agrupados com intervalo de classe:
- A classe de maior frequência é chamada classe modal. A moda será, então, o ponto médio desta classe.

$$\boxed{Mo = \frac{\ell^* + L^*}{2}}$$

i	ESTATURAS (cm)	f_i
1	150 ⊢ 154	4
2	154 ⊢ 158	9
3	158 ⊢ 162	11 ←
4	162 ⊢ 166	8
5	166 ⊢ 170	5
6	170 ⊢ 174	3
		$\Sigma = 40$

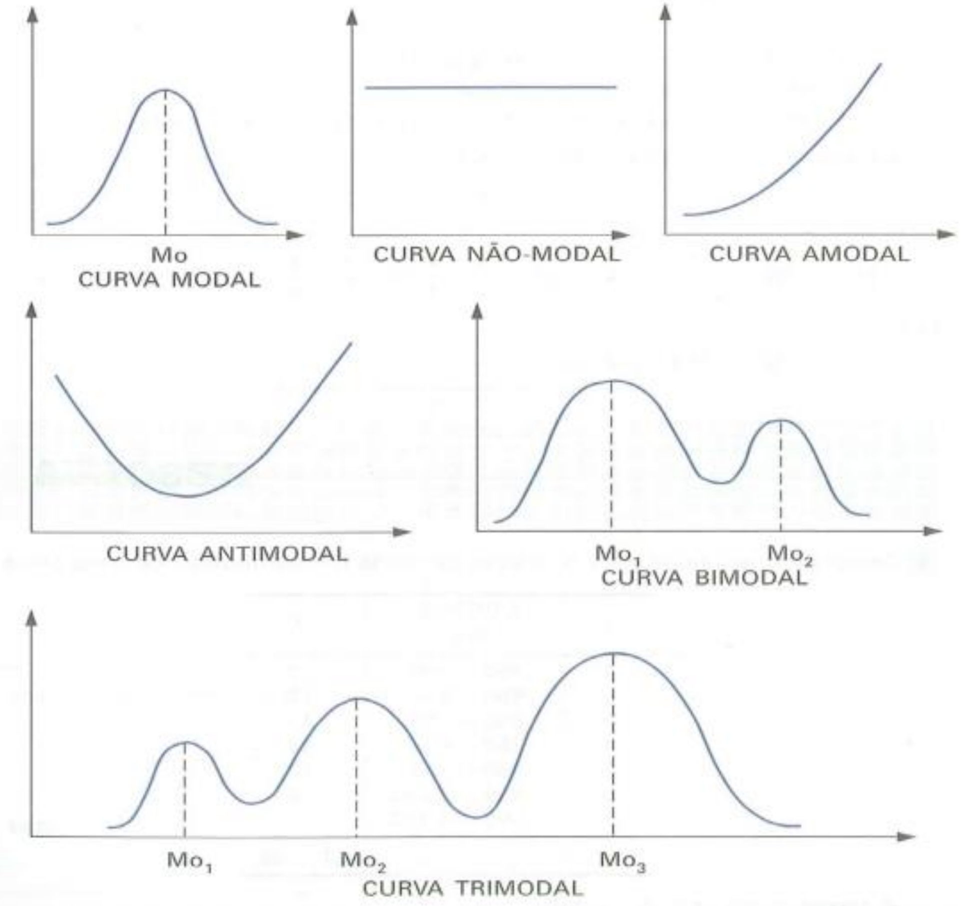
$$Mo = \frac{\ell^* + L^*}{2}$$

$$Mo = \frac{158 + 162}{2} = \frac{320}{2} = 160$$

$$Mo = 160 \text{ cm}$$

Expressões Gráficas da Moda

- Na curva de frequência, a moda é o valor que corresponde, no eixo das abcissas, ao ponto de ordenada máxima. Assim, temos:



Média Variáveis Descritivas

- Variáveis descritivas podem, em alguns casos, ser associadas a valores discretos a fim de tornar plausível a realização de operações aritméticas. Baseando apenas na opinião dos colegas, qual deve ser a tendência do resultado?

classe	freq
B	5
M	10
A	15
Total	30

Média com Variáveis Descritivas

- Considerando B = 0; M = 1 e A = 2, temos que:

$$\overline{x} = \frac{5 \cdot 0 + 1 \cdot 10 + 15 \cdot 2}{30} = 1,\overline{3}.$$

- Neste caso, temos uma tendência mais favorável a classe M.

Mediana de dados agrupados sem intervalo de classe

- O valor que divide a distribuição de frequências em 2 grupos com mesmo número de elementos estará na posição dada por:

$$\frac{\sum f_i}{2}$$

- Neste caso basta identificar a frequência acumulada imediatamente superior à metade da soma das frequências:

Mediana de dados agrupados sem intervalo de classe

Nº DE MENINOS	f_i	F_i
0	2	2
1	6	8
2	10	18
3	12	30
4	4	34
	$\Sigma = 34$	

$$\frac{\Sigma f_i}{2} = \frac{34}{2} = 17$$

- **Mediana = 2 meninos**

Mediana de dados agrupados sem intervalo de classe

- No caso de existir uma frequência acumulada (Fi), tal que:

$$F_i = \frac{\sum f_i}{2},$$

- A mediana será dada por:

$$Md = \frac{x_i + x_{i+1}}{2},$$

- Isto é, a mediana será a média aritmética entre o valor da variável correspondente a essa frequência acumulada e a seguinte.

Mediana de dados agrupados sem intervalo de classe

Exemplo:

TABELA 6.8

x_i	f_i	F_i
12	1	1
14	2	3
15	1	4
16	2	6
17	1	7
20	1	8
	$\Sigma = 8$	

Temos:

$$\frac{8}{2} = 4 = F_3$$

Logo:

$$Md = \frac{15 + 16}{2} = \frac{31}{2} = 15{,}5$$

donde:

$$Md = 15{,}5$$

Mediana de dados agrupados com intervalo de classe

- 1º. Determinar as frequências acumuladas
- 2º Calcular $\frac{\sum f_i}{2}$
- 3º Marcar a classe correspondente à frequência acumulada imediatamente superior à $\frac{\sum f_i}{2}$ (classe Mediana), em seguida utilizar a fórmula

$$\mathrm{Med} = \mathrm{li} + \left[\frac{\left(\frac{n}{2}\right) - fa_{[i-1]}}{fc} \right] . h$$

li é o limite inferior da classe mediana;
fa (i-1) é a frequência acumulada da classe anterior à classe da mediana;
fc é a frequência simples da classe mediana;
h é a amplitude do intervalo da classe mediana

Mediana de dados agrupados com intervalo de classe

i	ESTATURAS (cm)	f_i	F_i
1	150 ⊢ 154	4	4
2	154 ⊢ 158	9	13
3	158 ⊢ 162	11	24 ← classe mediana
4	162 ⊢ 166	8	32
5	166 ⊢ 170	5	37
6	170 ⊢ 174	3	40
		$\Sigma = 40$	

$$\frac{\Sigma f_i}{2} = \frac{40}{2} = 20$$

$$Md = \ell^* + \frac{\left[\frac{\Sigma f_i}{2} - F(ant)\right] h^*}{f^*}$$

$$Md = 158 + \frac{(20 - 13)4}{11} = 158 + \frac{28}{11} = 158 + 2{,}54 = 160{,}54,$$

$$Md = 160{,}5 \text{ cm}$$

Exemplo Média, Mediana e Moda

- Suponha que os resumos a seguir correspondem aos salários (em milhares de reais) pagos por duas companhias, digamos A e B.

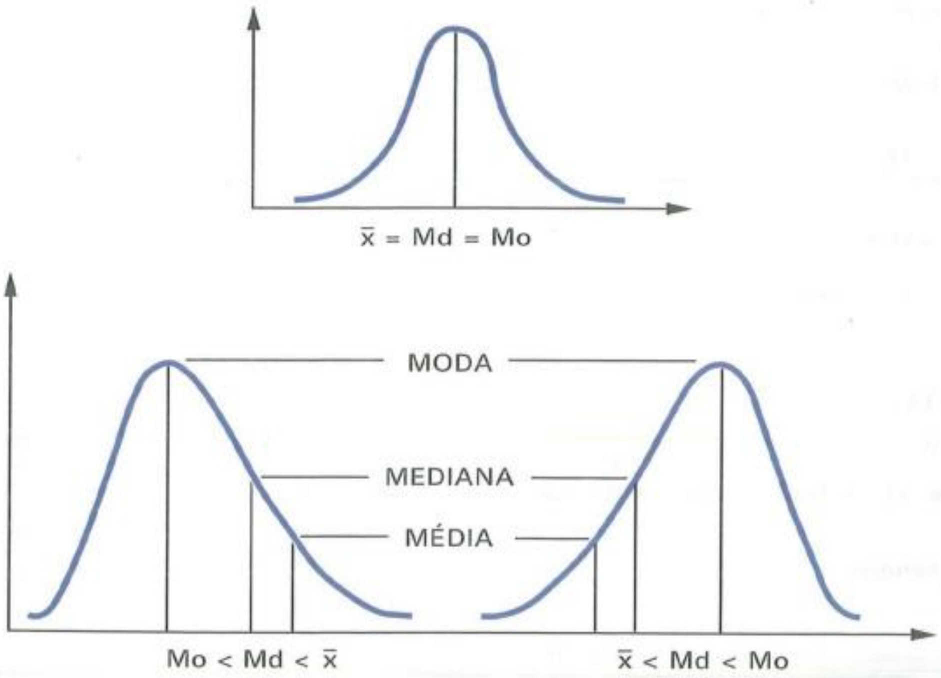
Companhia	A	B
média	2,5	2,0
mediana	1,7	1,9
moda	1,5	1,9

- Considerando apenas os dados indicados no resumo acima, qual empresa e mais atrativa?

Posição Relativa da Média, Mediana e Moda

- Quando uma distribuição é simétrica, as três medidas coincidem.
- Quando a distribuição é assimétrica, torna-se diferentes essas medidas e a diferença é tanto maior quanto maior é a assimetria.
- $\bar{X} = Md = Mo$, no caso da curva simétrica
- $Mo < Md < \bar{X}$, no caso da curva assimétrica positiva
- $\bar{X} < Md < Mo$, no caso da curva assimétrica negativa

Posição Relativa da Média, Mediana e Moda



Medidas separatrizes

- Medidas separatrizes dividem a sequência ordenada dos dados em partes que contêm a mesma quantidade de elementos.
- A mediana trata de um caso particular pois separa a amostra em duas porções as quais contêm 50% da informação cada. Considerando, por exemplo, o conjunto de dados 1,2,5,5,5,8,10,11,12,12,13,15, temos:

$$\underbrace{1, 2, 5, 5, 5, 8}_{\text{50\% das observações}} \quad \Big| \quad \underbrace{10, 11, 12, 12, 13, 15}_{\text{50\% das observações}}$$

$$\downarrow$$

$$\text{mediana: } \frac{8 + 10}{2} = 9$$

- **As regras de calculo são análogas as adotadas no caso da mediana; basta adapta-las para a porção desejada dos dados.**

Quartis – Medidas de Tendência Central

- Ao dividir os dados **ordenados em 4 partes** de mesmo tamanho, cada um resumira 25% da informação.
- Os elementos Q1, Q2 e Q3 que separam tais grupos são denominados os quartis amostrais.
- **Em particular, o segundo quartil Q2 corresponde a mediana**

$$\underbrace{1, 2, 5}_{25\%\ \text{dos dados}} \quad | \quad \underbrace{5, 5, 8}_{25\%\ \text{dos dados}} \quad | \quad \underbrace{10, 11, 12}_{25\%\ \text{dos dados}} \quad | \quad \underbrace{12, 13, 15}_{25\%\ \text{dos dados}}$$

$$Q_1 = \frac{5+5}{2} = 5 \quad Q_2 = \frac{8+10}{2} = 9 \quad Q_3 = \frac{12+12}{2} = 12$$

Quartis

- O primeiro quartil (Q1), separa a sequência ordenada deixando 25% de seus valores a esquerda e 75% de seus valores a direita.
- O segundo quartil (Q2) separa a sequencia ordenada deixando 50% de seus valores a esquerda e 50% de seus valores a direita. Note que o Q2 é a Mediana da série.
- O terceiro quartil (Q3) obedece a mesma regra dos anteriores.

Q1	Q2	Q3	Q4
25%	25%	25%	25%

Quartis

- O primeiro quartil (Q1), separa a sequência ordenada deixando 25% de seus valores a esquerda e 75% de seus valores a direita.
- O segundo quartil (Q2) separa a sequencia ordenada deixando 50% de seus valores a esquerda e 50% de seus valores a direita. Note que o Q2 é a Mediana da série.
- O terceiro quartil (Q3) obedece a mesma regra dos anteriores.

Q1	Q2	Q3	Q4
25%	25%	25%	25%

Quartis - Exemplo

- Dados não Agrupados

3, 4, 5, 5, 6, 8, 9, 11, 15, 21

$Q_2 = 7$

$Q_1 = 5$

$Q_3 = 11$

Quartis - Exemplo

1) Cálculo da mediana para os 3 quartis. Na realidade serão calculadas "**3 medianas**" em uma mesma série.

- Calcule os quartis da série: { 5, 2, 6, 9, 10, 13, 15 }
- Ordenação (crescente ou decrescente) dos valores: { 2, 5, 6, 9, 10, 13, 15 }
- O valor que divide a série acima em duas partes **iguais e igual a 9**, logo a **Md** = 9 que será = Q2.
- Temos: **{2, 5, 6 } e {10, 13, 15 }** outros dois grupos de valores iguais proporcionais pela mediana (quartil 2).

Quartis - Exemplo

2) Cálculo do quartil 1 e 3 → calcular as medianas das partes iguais provenientes da verdadeira Mediana da série (quartil 2).

Temos { 2, 5, 6 } a mediana é = 5. Ou seja: será o Q1

E {10, 13, 15 } a mediana é =13. Ou seja: será o Q3

Quartis - Exemplo

- Calcule os quartis da serie: { 1, 1, 2, 3, 5, 5, 6, 7, 9, 9, 10, 13 }
- Ordenar e Calcular o Quartil 2 = Md = (5+6)/2 = 5,5
- O quartil 1 será a mediana da série à esquerda de Md: { 1, 1, 2, 3, 5, 5 }
- → Q1 = (2+3)/2 = 2,5
- O quartil 3 será a mediana da série à direita de Md: {6, 7, 9, 9, 10, 13 }
- → Q3 = (9+9)/2 = 9

Quartis - para dados agrupados em classes

- Determine a classe que contém o valor quartil a ser calculado.
- A identificação da classe é feita por meio do termo da ordem calculada pela expressão:

$$P_{QK} = \frac{K \sum f_i}{4} \ (onde\ K = 1{,}2\ ou\ 3)$$

- Essa expressão determina a posição do referente quartil ou classe que contém o quartil.

Quartis - para dados agrupados em classes

- Assim, temos:

$$Q_k = l_{qk} + \left[\frac{\frac{K \cdot \sum fi}{4} - F_{ant}}{f_{QK}} \right] \cdot a_{Qk}$$

Sendo:

l_{Qk} = limite inferior da classe do quartil considerado.
F_{ant} = frequência acumulada da classe anterior à classe do quartil considerado.
a_{QK} = amplitude do intervalo de classe do quartil considerado.
f_{QK} = frequência simples da classe do quartil considerado.

Quartis - para dados agrupados em classes

- Para o calculo dos quartis de dados agrupados com intervalos de classe, consideramos a distribuição dos pesos de um grupo de turistas que visita um parque temático em Fortaleza/CE/Julho/06.

i	Pesos (kg)	Frequência (f_i)	acumulada (Fa)
1	10 ⊢ 30	10	10
2	30 ⊢ 50	24	34
3	50 ⊢ 70	57	91
4	70 ⊢ 90	44	135
5	90 ⊢ 110	29	164
6	110 ⊢ 130	16	180

Quartis - para dados agrupados em classes

Primeiro, calcula-se a classe a que pertence o quartil Q_1 (k=1), ou seja, a posição:

$$PQ1 = \frac{1 \cdot \sum f_i}{4} = \frac{180}{4} = 45$$

Observando a **coluna de freqüência acumulada**, verificamos que o quadragésimo quinto **termo** pertence à terceira classe (a freqüência acumulada da teceria classe abrange do 35º termo ao 91º termo).

Sabendo que a classe do primeiro quartil é a terceira classe, podemos verificar qual o **valor numérico** do primeiro quartil utilizando a expressão:

$$Q_1 = l_{Q_1} + \left[\frac{\frac{1\sum f_1}{4} - F_{ant}}{f_{Q_1}} \right] \cdot a_{Q_1} = 50 + \left[\frac{45 - 34}{57} \right] \cdot 20 \cong 53{,}9\text{kg}$$

Quartis - para dados agrupados em classes

2° quartil $\rightarrow \dfrac{2\sum f_i}{4} = \dfrac{2\times 180}{4} = 90$ (o segundo quartil pertence à terceira classe).

$$Q_2 = l_{Q_2} + \left[\frac{\frac{2\sum f_i}{4} - F_{ant}}{f_{Q_2}} \right] \bullet a_{Q_2} = 50 + \left[\frac{90 - 34}{57} \right] \cdot 20 \cong 69{,}7 \text{ kg}$$

3° quartil $\rightarrow \dfrac{3\sum f_i}{4} = \dfrac{3\times 180}{4} = 135$ (o terceiro quartil pertence à quarta classe)

$$Q_3 = l_{Q_3} + \left[\frac{\frac{3\sum f_i}{4} - F_{ant}}{f_{Q_3}} \right] \bullet a_{Q_3} = 70 + \left[\frac{135 - 91}{44} \right] \cdot 20 = 90{,}0 \text{ kg}$$

Assim temos: $Q_1 = 53{,}9$ kg; $Q_2 = 69{,}7$ kg e $Q_3 = 90{,}0$ kg

Generalização dos Quantis Amostrais

- Os quatro elementos $K_1; K_2; K_3$ e K_4 que separam os dados em **cinco partes iguais** são denominados **quintis** amostrais.
- Analogamente:

  i : 10 partes iguais - **decis** amostrais;

  ii : 100 partes iguais - **percentis** amostrais;

  iii : q partes iguais - **quantis** amostrais (q e um inteiro não-negativo qualquer)

Percentis

- Os percentis dividem uma distribuição de frequência em cem partes iguais. P1(1%), P2(2%), P3(3%), . . . , P99(99%)
- Para determinarmos a classe que contém o i-ésimo percentil, devemos calcular a posição do elemento correspondente ao percentil desejado.
- A fórmula para o cálculo dos i-ésimo percentil (i=1,2,...,99) é:

$$P_i = L_P + \left(\frac{\frac{i}{100}.n - F_{ant}}{f_P} \right).h$$

L_P = limite inferior da classe que contém o percentil
i = número do percentil a ser calculado (1,2,..., 99)
n = tamanho da amostra
F_{ant} = frequência acumulada anterior à classe que contém o percentil
f_P = frequência simples (ou absoluta) da classe que contém o percentil
h = amplitude da classe que contém o percentil

Percentis

- Se observarmos que ***os quartis e decis*** são múltiplos dos ***percentis***, então basta estabelecer a fórmula de cálculo dos **percentis**. Todas as outras medidas podem ser identificadas como percentis. A fórmula utilizada é a mesma usada para o cálculo da mediana.

$Q_1 = P_{25}$
$Q_2 = P_{50} = Md$
$Q_3 = P_{75}$

$D_1 = P_{10}$
$D_2 = P_{20}$
$D_3 = P_{30}$
$D_4 = P_{40}$
$D_5 = P_{50} = Md$

$D_6 = P_{60}$
$D_7 = P_{70}$
$D_8 = P_{80}$
$D_9 = P_{90}$

Percentis - Exemplo

- Considere uma tabela de custos:

Custos R$	Freqüência f_i	F_i	posições
450 |— 550	8	8	1ª a 8ª
550 |— 650	10	18	9ª a 19ª
650 |— 750	11	29	20ª a 29ª
750 |— 850	16	45	30ª a 45ª
850 |— 950	13	58	46ª a 58ª
950 |— 1050	5	63	59ª a 63ª
1050 |— 1150	1	64	64ª
Total	64	--	--

Percentis - Exemplo

a) Q_1

A posição ocupada pelo primeiro quartil é $\frac{64}{4} = 16^{a}$ posição, que corresponde a classe 550 |— 650. Aplicando a fórmula:

$$Q_1 = 550 + \left(\frac{\frac{1}{4}.64 - 8}{10} \right).100 = 630 \text{ reais.}$$

Percentis - Exemplo

b) Q_3

A posição ocupada pelo terceiro quartil é $\frac{3.64}{4} = 48^{a}$ posição, que corresponde a classe 850 |— 950. Aplicando a fórmula:

$$Q_3 = 850 + \left(\frac{\frac{3}{4}.64 - 45}{13} \right).100 = 873{,}08 \text{ reais.}$$

c) D_9

A posição ocupada pelo nono decil é $\frac{9.64}{10} = 57{,}6 \sim 58^{a}$ posição, que corresponde a classe 850 |— 950. Aplicando a fórmula:

$$D_9 = 850 + \left(\frac{\frac{9}{10}.64 - 45}{13} \right).100 = 946{,}92 \text{ reais.}$$

Percentis - Exemplo

d) P_{38}

A posição ocupada pelo 38º percentil é $\frac{38.64}{100} = 24{,}32 \sim 24^a$ posição, que corresponde a classe 650 |— 750. Aplicando a fórmula:

$$P_{38} = 650 + \left(\frac{\frac{38}{100}.64 - 18}{11} \right).100 = 707{,}45 \text{ reais.}$$

e) P_{25}

Lembre-se que o 25º percentil corresponde ao primeiro quartil, que calculamos anteriormente. Assim: $P_{25} = Q_1 = 630$ reais.