

NOTAS DE AULA

Medidas Descritivas

Prof.: IDEMAURO ANTONIO RODRIGUES DE LARA

PRINCIPAIS MEDIDAS DESCRITIVAS

1. Medidas de tendência central

1.1 Média

1.2 Mediana

1.3 Moda

2. Medidas de dispersão

2.1 Amplitude total

2.2 Desvio médio

2.3 Variância

2.4 Desvio padrão

2.5 Coeficiente de Variação

3. Medidas que caracterizam a forma da distribuição

3.1 Assimetria

3.2 Curtose

1. MEDIDAS DE TENDÊNCIA CENTRAL

1.1 Média aritmética:

A média aritmética é dada por:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

em que n corresponde ao tamanho da amostra e x_i ao i -ésimo valor observado.

Exemplo: Com o objetivo de avaliar a produção de leite, em kg, foram observadas as produções médias diárias de 10 produtores rurais atendidos por um plano governamental, cujos valores são apresentados a seguir:

9,80	9,90	9,95	10,00	8,78
9,90	9,34	10,34	11,75	15,00

Produção média de leite:

$$\begin{aligned}\bar{x} &= \frac{9,80 + 9,90 + \dots + 15,00}{10} \\ &= \frac{104,76}{10} = 10,48 \text{ kg.}\end{aligned}$$

Observações:

Média sem considerarmos o maior valor observado (15,00):

$$\bar{x} = \frac{9,80 + 9,90 + \dots + 11,75}{9} = \frac{89,76}{9} = 9,97 \text{ kg.}$$

- A média é bastante afetada por valores extremos;
- Deve-se ter cuidado ao usar a média, quando a distribuição dos dados é assimétrica.

Média para dados agrupados em tabelas de frequências

X_i	f_i
x_1	f_1
x_2	f_2
\vdots	\vdots
x_k	f_k
Total	n

$$n = \sum_{i=1}^k f_i$$

$$\hat{\mu} = \bar{x} = \frac{f_1 x_1 + f_2 x_2 + \dots + f_k x_k}{f_1 + f_2 + \dots + f_k}.$$

Exemplo: Encontrar o número médio de recipientes, em que se encontram larvas do mosquito da dengue, em uma amostra de 45 domicílios.

Tabela 1: Distribuição de frequências

X_i	f_i
0	6
1	11
2	13
3	7
4	5
5	1
6	2
Total	45

Cálculo da média:

$$\begin{aligned}
 \bar{x} &= \frac{0 \times 6 + 1 \times 11 + \dots + 6 \times 2}{45} \\
 &= \frac{95}{45} \\
 &= 2,11 \text{ recipientes/domicílio.}
 \end{aligned}$$

Média para dados agrupados em tabelas de classes de frequências

X_i	m_i	f_i
$\ell_1 \vdash L_1$	m_1	f_1
$\ell_2 \vdash L_2$	m_2	f_2
\vdots	\vdots	\vdots
$\ell_k \vdash L_k$	m_k	f_k
Total	n	

$$m_i = \frac{\ell_i + L_i}{2}$$

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^k m_i f_i}{\sum_{i=1}^k f_i}.$$

Exemplo: Considere a distribuição a seguir.

Tabela 2: Distribuição dos comprimentos das asas de uma espécie de inseto, em mm.

X_i	m_i	f_i
2 – 4	3	10
4 – 6	5	16
6 – 8	7	10
8 – 10	9	2
10 – 12	11	1
12 – 14	13	1
Total		40

$$\begin{aligned}
 \bar{x} &= \frac{3 \times 10 + \dots + 13 \times 1}{40} \\
 &= \frac{222}{40} \\
 &= 5,55 \text{ mm.}
 \end{aligned}$$

1. MEDIDAS DE TENDÊNCIA CENTRAL

1.2 Mediana:

A mediana é o valor que ocupa a posição central em um conjunto de dados em ordem crescente (Rol). Logo, esta medida de tendência central é pouco afetada por valores extremos ou discrepantes.

Sendo $x_{[1]}, x_{[2]}, \dots, x_{[n]}$ as estatísticas de ordem da amostra, a mediana é dada por:

$$\text{md} = \tilde{x} = \begin{cases} x_{[\frac{n+1}{2}]}, & \text{se } n \text{ for ímpar} \\ \frac{x_{[n/2]} + x_{[n/2+1]}}{2}, & \text{se } n \text{ for par} \end{cases}$$

Exemplo: Para os valores observados de produção média diária de leite, tem-se o seguinte rol:

8,78	9,34	9,80	9,90	9,90	9,95	10,00	10,34	11,75	15,00
------	------	------	------	------	------	-------	-------	-------	-------

$$\begin{aligned}
 \text{md} = \tilde{x} &= \frac{x_{[5]} + x_{[6]}}{2} = \frac{9,90 + 9,95}{2} = \frac{19,85}{2} \\
 &= 9,925 \text{ kg}
 \end{aligned}$$

Mediana para dados agrupados em tabelas de frequências

Exemplo: Considerando a distribuição da Tabela 2.

Tabela 1: Distribuição de frequências

X_i	f_i	F_i
0	6	6
1	11	17
2	13	30
3	7	37
4	5	42
5	1	43
6	2	45

$$\text{md} = \tilde{x} = x_{[23]} = 2$$

Mediana para dados agrupados em tabelas de classes de frequências

X_i	f_i	F_i
$\ell_1 \vdash L_1$	f_1	F_1
$\ell_2 \vdash L_2$	f_2	F_2
\vdots	\vdots	\vdots
$\ell_{md} \vdash L_{md}$	f_{md}	F_{md}
\vdots	\vdots	\vdots
$\ell_k \vdash L_k$	f_k	F_k
Total	n	

Sendo: $h_{md} = L_{md} - \ell_{md}$ a amplitude da classe mediana e F_{md-1} , a frequência acumulada anterior à classe mediana, tem-se:

$$md = \tilde{x} = \ell_{md} + \left(\frac{\frac{n}{2} - F_{md-1}}{f_{md}} \right) h_{md}$$

Exemplo: Considere a distribuição a seguir.

Tabela 2: Distribuição dos comprimentos das asas de uma espécie de inseto, em mm.

X_i	f_i	F_i
2 \vdash 4	10	10
4 \vdash 6	16	26
6 \vdash 8	10	36
8 \vdash 10	2	38
10 \vdash 12	1	39
12 \vdash 14	1	40

A mediana amostral é estimada por:

$$md = 4 + \left(\frac{\frac{40}{2} - 10}{16} \right) 2 = 5,25\text{mm}$$

1. MEDIDAS DE TENDÊNCIA CENTRAL

1.3 Moda

Moda é o atributo ou o valor de maior frequência em uma distribuição.

Exemplo: Para os valores observados de produção média diária de leite, tem-se:

8,78	9,34	9,80	9,90	9,90
9,95	10,00	10,34	11,75	15,00

$$m_o = \ddot{x} = 9,90 \text{ kg}$$

Moda para dados agrupados em tabelas de frequências

Exemplo: Considerando a distribuição da Tabela 2.

Tabela 1: Distribuição de frequências

X_i	f_i
0	6
1	11
2	13
3	7
4	5
5	1
6	2

A moda do número de recipientes encontrados nas residências é:

$$mo = \ddot{x} = 2$$

Moda para dados agrupados em tabelas de classes de frequências

Em geral, basta identificar a classe modal, que é a classe de maior frequência (para intervalos regulares) ou de maior densidade de frequência (para intervalos não regulares). Embora, também possa ser calculada pelo critério gráfico de Czuber:

$$mo = \ell_{mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right) h_{mo},$$

em que: ℓ_{mo} é o limite inferior da classe modal;

$$\Delta_1 = f_{mo} - f_{mo-1};$$

$$\Delta_2 = f_{mo} - f_{mo+1};$$

h_{mo} é amplitude da classe modal.

Exemplo: Considere a distribuição a seguir.

Tabela 2: Distribuição dos comprimentos das asas de uma espécie de inseto, em mm.

X_i	f_i
2 \vdash 4	10
4 \vdash 6	16
6 \vdash 8	10
8 \vdash 10	2
10 \vdash 12	1
12 \vdash 14	1

A classe modal é 4 \vdash 6. Então:

$$\text{mo} = \frac{4 + 6}{2} = 5 \text{ mm}$$

é a moda bruta. A moda pelo processo de Czuber é:

$$\text{mo} = 4 + \left(\frac{6}{6 + 6} \right) 2 = 5 \text{ mm},$$

2. MEDIDAS DE DISPERSÃO

Exemplo Foram observadas quatro amostras de duas colheitadeiras de milho, quanto a porcentagem de quebra de sementes, conforme a tabela a seguir:

Amostra	Colheitadeira	
	A	B
1	5	0
2	4	3
3	5	7
4	6	10
média	5	5

2. MEDIDAS DE DISPERSÃO

As medidas de dispersão são estatísticas descritivas que visam fornecer o grau de variabilidade das observações em relação a um valor central (geralmente a média aritmética). As mais usuais são:

- Amplitude total
- Desvio Médio
- **Variância**
- **Desvio Padrão**
- Coeficiente de Variação

2. MEDIDAS DE DISPERSÃO

2.1 Amplitude total

$$A_t = \max(x) - \min(x)$$

Exemplo: Considerando-se o exemplo de porcentagem de quebra de sementes de milho, temos:

Colheitadeira	Amplitude
A	$6-4 = 2$
B	$10-0 = 10$

Agora, se considerarmos os conjuntos

$A = \{3, 4, 9, 11, 19, 20\}$ e $B = \{13, 20, 21, 21, 21, 30\}$, ambos têm a mesma amplitude $A_t = 17$, no entanto, o conjunto B parece ser mais “homogêneo”.

2. MEDIDAS DE DISPERSÃO

2.2 Desvio médio

i. Desvio de uma observação em relação a uma constante:

$$d_i = x_i - k$$

ii. Desvio de uma observação em relação à média aritmética:

$$e_i = x_i - \bar{x}$$

iii. Soma dos desvios em relação à média

$$\sum_{i=1}^n e_i = 0$$

iv. **Desvio Médio**

$$Dm_x = \frac{1}{n} \sum_{i=1}^n |e_i| = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

Exemplo: Considerando-se o exemplo de porcentagem de quebra de sementes de milho, temos:

Colheitadeira	Desvios				$\sum_{i=1}^4 e_i /4$
A	0,0	-1,0	0,0	1,0	$2,0/4 = \mathbf{0,5}$
B	-5,0	-2,0	2,0	5,0	$14,0/4 = \mathbf{3,5}$

2. MEDIDAS DE DISPERSÃO

2.3 Variância

Variância é a média dos quadrados dos desvios em relação à média aritmética.

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^n e_i^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

Estimador da variância populacional:

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

Exemplo:

Colheitadeira A:

$$S_{X_1}^2 = \frac{\sum_{i=1}^4 (x_i - \bar{x})^2}{4 - 1}$$

$$= \frac{(5 - 5)^2 + (4 - 5)^2 + (5 - 5)^2 + (6 - 5)^2}{3} = 0,67(\%)^2$$

Colheitadeira B:

$$S_{X_2}^2 = \frac{\sum_{i=1}^4 (x_i - \bar{x})^2}{4 - 1}$$

$$\frac{(0 - 5)^2 + (3 - 5)^2 + (7 - 5)^2 + (10 - 5)^2}{3} = 19,33(\%)^2$$

Variância para dados agrupados em tabelas de frequências

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^k f_i e_i^2 = \frac{1}{n-1} \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

em que k corresponde ao número de diferentes valores para a variável e $n = \sum_{i=1}^k f_i$

Ou ainda,

$$S_X^2 = \frac{1}{n-1} \left[\sum_{i=1}^k f_i x_i^2 - \frac{\left(\sum_{i=1}^k f_i x_i \right)^2}{n} \right]$$

Exemplo: Foram observadas 76 goiabas quanto ao número de pintas pretas, cujos dados são apresentados na tabela a seguir:

X_i	f_i
0	21
1	33
2	13
3	4
4	2
6	2
8	1
Total	76

$$\begin{aligned}\bar{x} &= \frac{0 \times 21 + 1 \times 33 + \dots + 8 \times 1}{76} \\ &= 1,30 \text{ pintas}\end{aligned}$$

$$\begin{aligned}S_X^2 &= \frac{21(0 - 1,30)^2 + \dots + 1(8 - 1,30)^2}{76 - 1} \\ &= 2,13 \text{ pintas}^2\end{aligned}$$

Variância para dados agrupados em tabelas de classes de frequências

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^k f_i e_i^2 = \frac{1}{n-1} \sum_{i=1}^k f_i (m_i - \bar{x})^2$$

em que k corresponde ao número de diferentes valores para a variável e $n = \sum_{i=1}^k f_i$

Ou ainda,

$$S_X^2 = \frac{1}{n-1} \left[\sum_{i=1}^k f_i m_i^2 - \frac{\left(\sum_{i=1}^k f_i m_i \right)^2}{n} \right]$$

Exemplo: Considere a distribuição a seguir.

Tabela 2: Distribuição dos comprimentos das asas de uma espécie de inseto, em mm.

X_i	m_i	f_i	$m_i \times f_i$	$m_i^2 \times f_i$
2 \vdash 2	3	10	30	30
4 \vdash 6	5	16	80	400
6 \vdash 8	7	10	70	490
8 \vdash 10	9	2	18	162
10 \vdash 12	11	1	11	121
12 \vdash 14	13	1	13	169
Total		40	222	1372

A variância amostral é dada por:

$$\begin{aligned}
 S_X^2 &= \frac{1}{40 - 1} \left[1372 - \frac{(222)^2}{40} \right] \\
 &= 3,58 \text{ mm}^2
 \end{aligned}$$

2. MEDIDAS DE DISPERSÃO

2.4 Desvio padrão

O **desvio padrão** corresponde à raiz quadrada positiva da variância, ou seja:

$$S_X = \sqrt{S_X^2}$$

Assim sendo, o desvio padrão tem a mesma unidade dos dados originais.

Para o exemplo do comprimento das asas dos insetos, o desvio padrão é:

$$S_X = \sqrt{3,58} = 1,89 \text{ mm}$$

2. MEDIDAS DE DISPERSÃO

2.5 Coeficiente de variação

O **coeficiente de variação** é uma medida de dispersão relativa e é dada por

$$CV_X = \frac{S_X}{\bar{x}} 100(\%)$$

Como é uma medida pura, pode ser usada na comparação de dispersões quando as unidades de medidas são diferentes.

Orientação

$$\begin{aligned} CV &\leq 10\% \Rightarrow \text{baixo} \\ 10\% < CV &\leq 20\% \Rightarrow \text{médio} \\ 20\% < CV &\leq 30\% \Rightarrow \text{alto} \\ CV &> 30\% \Rightarrow \text{muito alto} \end{aligned}$$

Exemplo. Comparar a variabilidade das idades com a variabilidade dos tempos de reação de uma substância administrada em indivíduos similares.

Estatísticas	Idade (em anos)	Tempo (em minutos)
média	37,57	10,22
desvio padrão	9,83	3,57

CONSIDERAÇÕES SOBRE MOMENTOS DE UMA DISTRIBUIÇÃO

Definição 1 Chama-se de r -ésimo momento amostral à estatística:

$$m'_r = \frac{1}{n} \sum_{i=1}^n x_i^r, \quad r \in \mathbb{N}^*$$

Definição 2 Chama-se de r -ésimo momento amostral central à estatística:

$$m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r, \quad r \in \mathbb{N}^*$$

Observações:

- i. O primeiro momento (ao redor de zero) é a média amostral, ou seja, $m'_1 = \bar{x}$;
- ii. O primeiro momento central é nulo, ou seja, $m_1 = 0$;
- iii. A variância é dada pelo segundo momento central, ou seja, $m_2 = \hat{\sigma}_x^2$.

3. MEDIDAS DE ASSIMETRIA E CURTOSE

Estas medidas descritivas estão relacionadas à forma da distribuição.

3.1 Coeficientes de Assimetria

a. Coeficientes de Pearson

$$As = \frac{3(\bar{x} - md)}{S_x} \qquad As = \frac{\bar{x} - mo}{S_x}$$

a. Coeficiente do momento

$$\hat{\alpha}_3 = \frac{m_3}{s_x^3}$$

Orientação:

$As = 0$ indica que a distribuição é simétrica.

$As < 0$ indica que a distribuição é assimétrica negativa (ou à esquerda).

$As > 0$ indica que a distribuição é assimétrica positiva (ou à direita).

3.2 Coeficiente de Curtose

$$\hat{\alpha}_4 = \frac{m_4}{s_x^4} = \frac{m_4}{m_2^2}$$

Assim se:

$$\hat{\alpha}_4 < 3 \text{ (platicúrtica)} \quad \hat{\alpha}_4 = 3 \text{ (mesocúrtica)} \quad \hat{\alpha}_4 > 3 \text{ (leptocúrtica)}$$

Também é possível considerar a medida $\hat{\alpha}^*_4 = \hat{\alpha}_4 - 3$ que quando positiva sinaliza para uma distribuição leptocúrtica, se negativa indica uma distribuição platicúrtica e nula uma distribuição normal ou mesocúrtica.

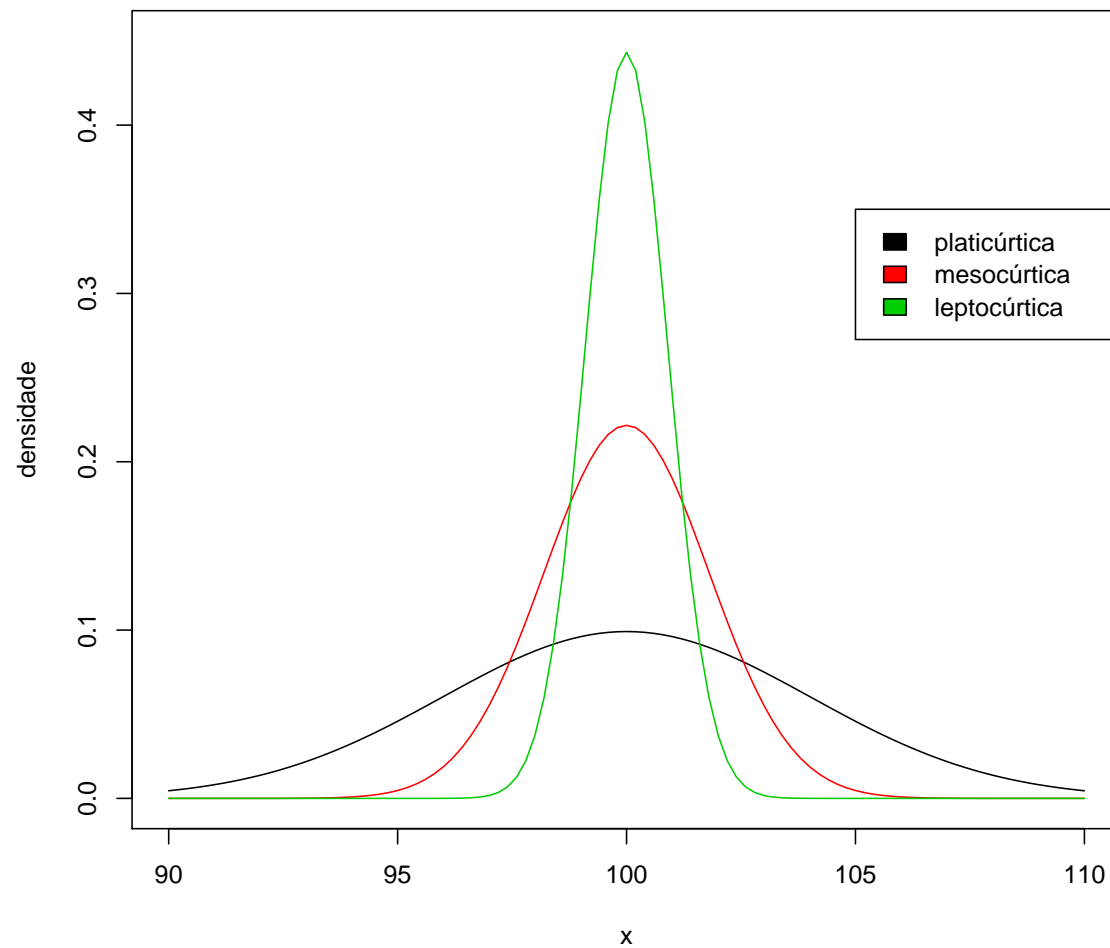


Figura 1: Três curvas normais com graus de curtose diferentes

REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDRADE, D.F.; OGLIARI, P.J. **Estatística para as Ciências Agrárias e Biológicas com noções de experimentação**. Editora da UFSC, Florianópolis, 2007.
2. BARBETA, P.A.; REIS, M.M.; BORNIA, A.C. **Estatística para Cursos de Engenharia e Informática**. São Paulo, Atlas, 2010.
3. BUSSAB, W.O. ; P.A. MORETIN, **Estatística Básica**, 5ª edição. Editora Saraiva, 2002.
4. MAGALHÃES, M.N.; LIMA, A.C. P de. **Noções de Probabilidade e Estatística**. 6ª ed. São Paulo: EDUSP, 2007.