

Unidade II - Estatística descritiva

- 2.1. Apresentação de dados
 - 2.1.1 Séries estatísticas
 - 2.1.2 Tabelas
 - 2.1.3 Gráficos
- 2.2. Distribuições de freqüências e gráficos
 - 2.2.1 Tabelas de classificação simples
 - 2.2.2 Tabelas de classificação cruzada
- 2.3. Medidas descritivas
 - 2.3.1 Medidas de localização ou tendência central
 - 2.3.2 Medidas separatrizes
 - 2.3.3 Medidas de variação ou dispersão
 - 2.3.4 Medidas de formato
- 2.4. Análise exploratória de dados

Profa. Clause Piana

1

Medidas descritivas

⇒ São funções de valores de uma variável numérica

Objetivo → reduzir um conjunto de dados numéricos a um pequeno grupo de valores que deve fornecer toda a informação relevante a respeito desses dados

⇒ Podem ser classificadas em quatro grupos:

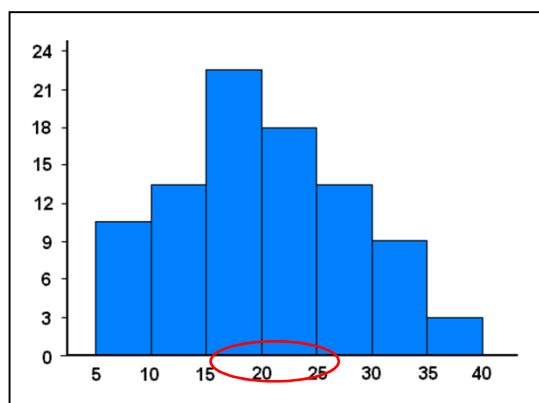
- ◆ Medidas de localização ou tendência central
- ◆ Medidas separatrizes
- ◆ Medidas de variação ou dispersão
- ◆ Medidas de formato

Profa. Clause Piana

2

Medidas de localização ou tendência central:
descrevem o centro da distribuição.

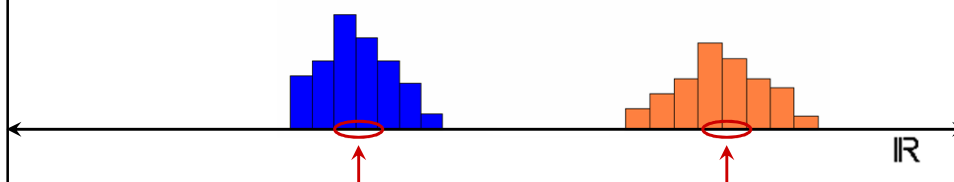
Expectativa: centro contém a maior parte das observações



Profa. Clause Piana

3

Medidas de localização ou tendência central:
descrevem valores que caracterizam o centro da distribuição.

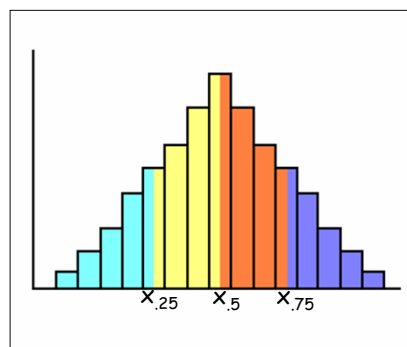
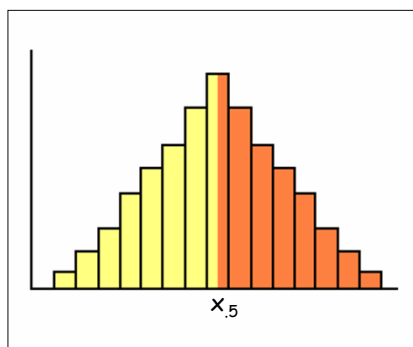


Dão ideia da localização das distribuições na reta do reais

Profa. Clause Piana

4

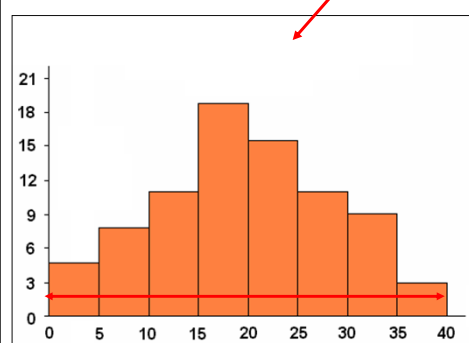
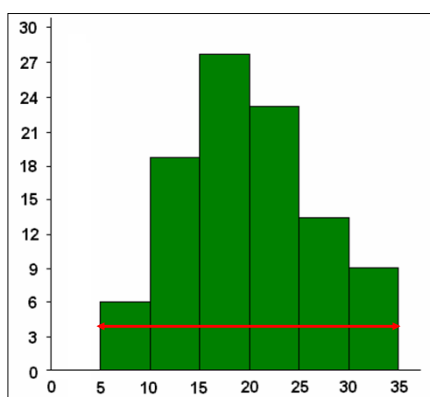
Medidas separatrizes: indicam limites para proporções de observações em um conjunto de dados ordenado



Profa. Clause Piana

5

Medidas de variação ou dispersão: informam quanto os valores de um conjunto de dados diferem entre si.



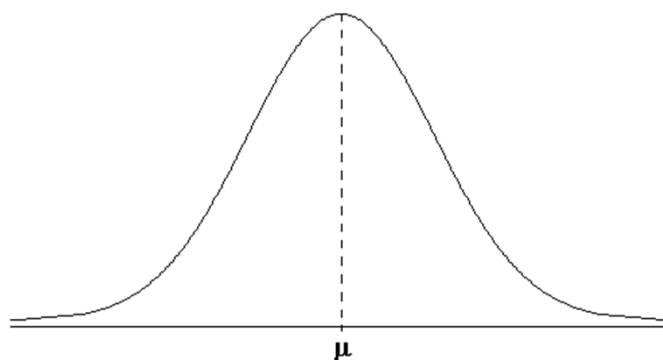
Profa. Clause Piana

6

Medidas de formato

Informam sobre a assimetria e a curtose da distribuição em relação à curva normal

Curva normal

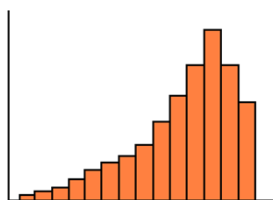


Profa. Clause Piana

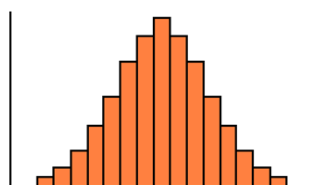
7

Medidas de formato

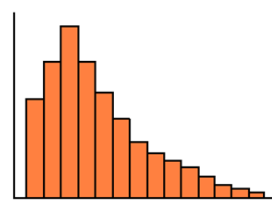
Assimetria: simetria da distribuição em relação à curva normal. Como se distribuem as observações em torno da média?



assimétrica negativa



simétrica



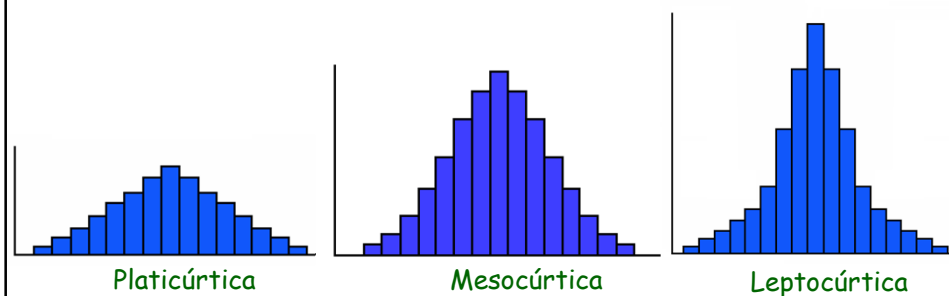
assimétrica positiva

Profa. Clause Piana

8

Medidas de formato

Curtose: achatamento da distribuição em relação à curva normal. Como se concentram as observações no centro ou nas caudas da distribuição?



Plat(i)- [Do gr. *platys*]: 'chato', 'plano', 'largo'

Lept(o)- [Do gr. *leptós*]: 'delgado', 'miúdo', 'magro', 'alongado'

Medidas descritivas

- ⇒ Existe uma grande variedade de medidas descritivas, muitas delas competidoras entre si
- ⇒ Considerações para a escolha da medida mais adequada:
 - ◆ A medida é de fácil interpretação? É intuitiva?
 - ◆ Existem valores atípicos que podem afetá-la exageradamente?
 - ◆ O propósito da análise é meramente descritivo ou planeja-se fazer inferências?

Valores atípicos ou discrepantes

Variável: estatura (m)

	Jogadores de basquete	Jóqueis	
	2,08	1,60	
típico →	1,98	1,62	
	1,95	1,65	
	2,03	1,55	
	2,10	1,57	
	1,93	1,58 ← típico	
atípico →	1,58	1,98 ← atípico	

A caracterização de um valor como **atípico** depende do contexto. Um valor considerado atípico num determinado contexto pode ser típico em outro contexto.

Algumas medidas são afetadas pela presença de valores atípicos no conjunto de dados, outras não.

1. Medidas de localização ou tendência central

Medidas de localização ou tendência central

Objetivo → indicar ou representar o **centro** de uma distribuição

Expectativa: centro contém a maior parte das observações

Medidas de localização mais utilizadas:

- ◆ Média aritmética
- ◆ Mediana
- ◆ Moda

Média aritmética

Medida mais conhecida e utilizada:

- ✓ facilidade de cálculo e de compreensão
- ✓ propriedades matemáticas e estatísticas

Definição: Média aritmética é uma combinação linear de todas as observações

Conjunto de observações: x_1, x_2, \dots, x_n

Conjunto de coeficientes: c_1, c_2, \dots, c_n

Combinação linear: $c_1x_1 + c_2x_2 + \dots + c_nx_n = \sum c_i x_i$

Média aritmética

Combinação linear: $c_1x_1 + c_2x_2 + \dots + c_nx_n = \sum c_i x_i$

Na média aritmética os coeficientes são pesos (p_i).

$$p_1x_1 + p_2x_2 + \dots + p_nx_n = \sum p_i x_i$$

Média aritmética

- Simples** → todos os pesos são iguais ($p_i=p$)
 - $p_1 = p_2 = \dots = p_n = p$
 - $px_1 + px_2 + \dots + px_n = \sum px_i$
- Ponderada** → os pesos são diferentes
 - $p_1x_1 + p_2x_2 + \dots + p_nx_n = \sum p_i x_i$

Profa. Clause Piana

15

Média aritmética simples (\bar{x})

Para um conjunto de n valores:

$$x_i = x_1, x_2, \dots, x_n \quad e \quad p_i = p = \frac{1}{n}$$

$$\bar{x} = \sum \frac{1}{n} x_i = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

soma de todos os valores
total de valores somados

Interpretação: se tivéssemos um total fixo e quiséssemos dividi-lo em n partes iguais, esse valor constante seria a média aritmética simples. A média é a quantidade comum a todos.

Exemplo:

X = peso (kg)

$x_i = 9, 7, 4, 5, 10$

$$\bar{x} = \frac{9+7+4+5+10}{5} = \frac{35}{5} = 7 \text{ kg}$$

16

Média aritmética ponderada (\bar{x}_p)

Temos um conjunto de valores e um conjunto de pesos:

$$x_i = x_1, x_2, \dots, x_n$$

$$p_i = p_1, p_2, \dots, p_n, \text{ sendo } p_i > 0 \text{ e } \sum p_i = 1$$

$$\bar{x}_p = \frac{\sum x_i p_i}{\sum p_i}$$

← soma de produtos de valores e pesos

← soma dos pesos

Exemplo:

X = nota

$$x_i = 9, 4,5, 5, 10$$

$$p_i = 10, 10, 5, 5$$

$$\bar{x}_p = \frac{9 \times 10 + 4,5 \times 10 + 5 \times 5 + 10 \times 5}{10 + 10 + 5 + 5}$$

$$\bar{x}_p = \frac{210}{30} = 7$$

Profa. Clause Piana

17

Propriedades algébricas da média aritmética

1ª propriedade: A média de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é a própria constante.

Verificação numérica:

$$x_i = 4, 4, 4, 4, 4$$

$$\bar{x} = 4$$

Profa. Clause Piana

19

2ª propriedade: Ao somar (ou subtrair) uma constante c por todos os valores de um conjunto de dados, sua média também é somada (ou subtraída) por esta constante.

Verificação numérica:

$$x_i = 9, 7, 5, 10, 4 \quad \bar{x} = 7$$

1. Somar $c=2$

$$x_i+2 = 11, 9, 7, 12, 6$$

$$\bar{x}_{x+2} = \frac{\sum(x_i+2)}{n} = \frac{11+9+7+12+6}{5} = \frac{45}{5} = 9$$

$$\begin{aligned} \bar{x}_{x+2} &= 7 + 2 \\ \bar{x}_{x+c} &= \bar{x} + c \end{aligned}$$

Profa. Clause Piana

20

Demonstração:

$$\begin{aligned}\bar{x}_{x+c} &= \frac{\sum(x_i + c)}{n} \\ &= \frac{\sum x_i + \sum c}{n} \\ &= \frac{\sum x_i + nc}{n} \\ &= \frac{\sum x_i}{n} + \frac{nc}{n} = \frac{\sum x_i}{n} + c = \bar{x} + c\end{aligned}$$

$$\boxed{\bar{x}_{x+c} = \bar{x} + c}$$

Profa. Clause Piana

21

3ª propriedade: Ao multiplicar (ou dividir) uma constante c por todos os valores de um conjunto de dados, sua média também é multiplicada (ou dividida) por esta constante.

Verificação numérica:

$$x_i = 9, 7, 5, 10, 4 \quad \bar{x} = 7$$

Multiplicar por $c=2$

$$2x_i = 18, 14, 10, 20, 8$$

$$\bar{x}_{2x} = \frac{\sum 2x_i}{n} = \frac{18+14+10+20+8}{5} = \frac{70}{5} = 14$$

$$\boxed{\begin{aligned}\bar{x}_{2x} &= 2 \times 7 \\ \bar{x}_{cx} &= c\bar{x}\end{aligned}}$$

Profa. Clause Piana

22

Demonstração:

$$\begin{aligned}\bar{x}_{cx} &= \frac{\sum cx_i}{n} \\ &= \frac{c \sum x_i}{n} \\ &= c \frac{\sum x_i}{n} = c\bar{x}\end{aligned}$$

$$\boxed{\bar{x}_{cx} = c\bar{x}}$$

Profa. Clause Piana

23

4ª propriedade: A soma de todos os desvios em relação à média de um conjunto de valores é nula.

$$\sum (x_i - \bar{x}) = 0$$

desvio

diferença entre a observação e a média aritmética

Verificação numérica:

$$x_i = 9, 7, 5, 10, 4$$

$$\bar{x} = 7$$

i	x_i	$(x_i - \bar{x})$
1	9	2
2	7	0
3	5	-2
4	10	3
5	4	-3
Σ	35	0

Demonstração:

$$\begin{aligned}
 \sum (x_i - \bar{x}) &= \sum x_i - \sum \bar{x} \\
 &= \sum x_i - n\bar{x} \\
 &= \sum x_i - n \frac{\sum x_i}{n} \\
 &= \sum x_i - \sum x_i = 0
 \end{aligned}
 \qquad
 \bar{x} = \frac{\sum x_i}{n}$$

$$\boxed{\sum (x_i - \bar{x}) = 0}$$

Profa. Clause Piana

25

5ª propriedade: A soma dos quadrados dos desvios em relação a uma constante c é mínima quando $c = \bar{x}$.

$$\sum (x_i - c)^2 \leftarrow \boxed{\text{é mínima quando } c = \bar{x}}$$

Verificação numérica:

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - 5)^2$	$(x_i - 10)^2$
1	9	2	4	16	1
2	7	0	0	4	9
3	5	-2	4	0	25
4	10	3	9	25	0
5	4	-3	9	1	36
Σ	35	0	26	46	71

$$n = 5 \quad \bar{x} = 7$$

Profa. Clause Piana

26

Demonstração:

Somar e subtrair uma constante não altera a equação

Rearranjar os termos não altera a equação

$$\begin{aligned} \sum (x_i - c)^2 &= \sum (x_i - c + \bar{x} - \bar{x})^2 \\ &= \sum [(x_i - \bar{x}) + (\bar{x} - c)]^2 \\ &= \sum [(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2] \\ &= \sum (x_i - \bar{x})^2 + \sum 2(x_i - \bar{x})(\bar{x} - c) + \sum (\bar{x} - c)^2 \\ &= \sum (x_i - \bar{x})^2 + 2(\bar{x} - c) \sum (x_i - \bar{x}) + n(\bar{x} - c)^2 \end{aligned}$$

Este termo é igual a zero

quando $c = \bar{x} \rightarrow n(\bar{x} - c)^2 = 0$

quando $c \neq \bar{x} \rightarrow n(\bar{x} - c)^2 > 0$

Profa. Clause Piana

27

Aplicação:

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - 5)^2$	$(x_i - 10)^2$
1	9	2	4	16	1
2	7	0	0	4	9
3	5	-2	4	0	25
4	10	3	9	25	0
5	4	-3	9	1	36
Σ	35	0	26	46	71

Expressão geral: $\sum (x_i - c)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - c)^2$

Para $c = \bar{x}$ $\sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - \bar{x})^2$
 $= 26 + 5(7 - 7)^2 = 26$

Para $c = 5$ $\sum (x_i - 5)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - 5)^2$
 $= 26 + 5(7 - 5)^2 = 46$

Para $c = 10$ $\sum (x_i - 10)^2 = \sum (x_i - \bar{x})^2 + n(\bar{x} - 10)^2$
 $= 26 + 5(7 - 10)^2 = 71$

Resumo: Propriedades da média aritmética

1ª propriedade: A média de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é a própria constante.

2ª propriedade: Ao somar (ou subtrair) uma constante c por todos os valores de um conjunto de dados, sua média também é somada (ou subtraída) por esta constante. $\bar{x}_{x+c} = \bar{x} + c$

3ª propriedade: Ao multiplicar (ou dividir) uma constante c por todos os valores de um conjunto de dados, sua média também é multiplicada (ou dividida) por esta constante. $\bar{x}_{cx} = c\bar{x}$

4ª propriedade: A soma de todos os desvios em relação à média de um conjunto de valores é nula. $\sum(x_i - \bar{x}) = 0$

5ª propriedade: A soma dos quadrados dos desvios em relação a uma constante c é mínima quando $c = \bar{x}$.

Profa. Clause Piana $\sum(x_i - c)^2 \leftarrow$ é mínima quando $c = \bar{x}$

29

Propriedades estatísticas da média aritmética

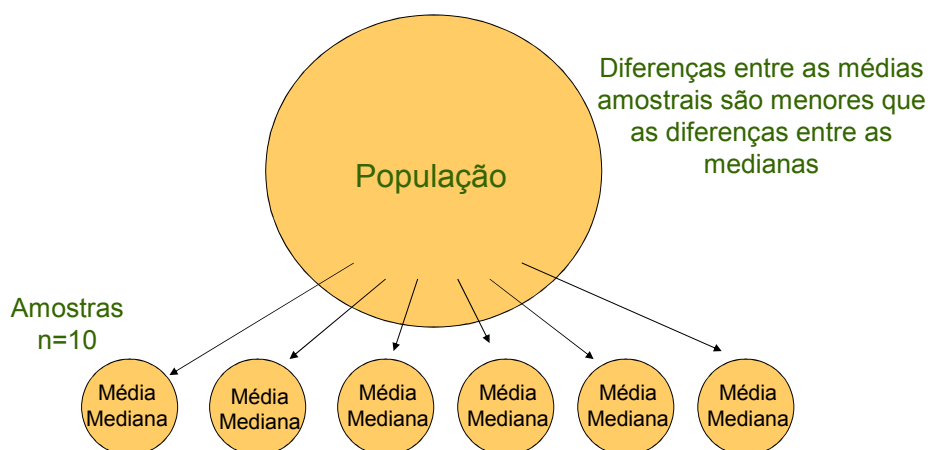
Propriedades estatísticas da média aritmética

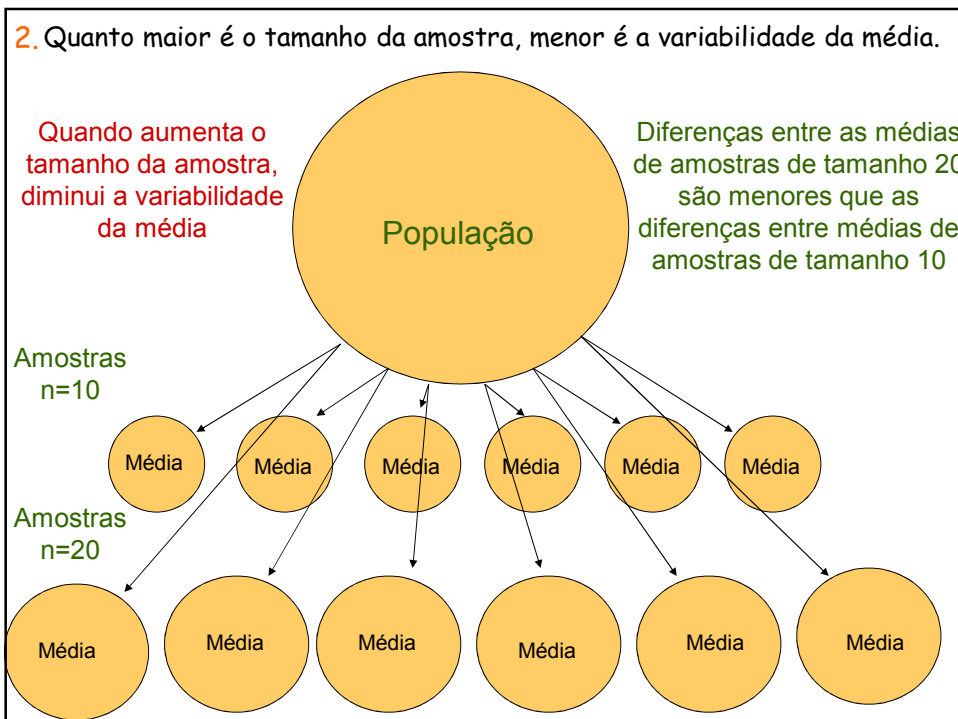
⇒ As propriedades estatísticas da média aritmética necessitam de um conhecimento mais avançado para serem adequadamente compreendidas.

Podemos destacar as duas propriedades mais importantes:

1. Se for considerado um grande número de amostras de dados de mesmo tamanho e obtidos na mesma população, a média aritmética varia menos de uma amostra para a outra do que outras medidas de localização.
2. Quanto maior for o tamanho da amostra de valores utilizada para o cálculo da média, menor é a sua variabilidade.

1. A média aritmética varia menos de uma amostra para a outra do que outras medidas de localização.



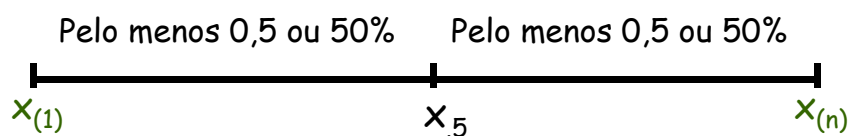


Mediana (M_d ou $x_{.5}$)

É a medida que divide um conjunto de dados **ordenado** em duas partes aproximadamente iguais.

A proporção de, pelo menos, 0,5 dos valores é menor ou igual à mediana e, pelo menos, 0,5 é maior ou igual.

Conjunto de dados ordenado: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$



Para obter a mediana:

1. Ordenar os dados
2. Determinar a posição da mediana

Dois casos:

n ímpar → um valor central

$$\text{posição} \rightarrow p = \frac{n+1}{2} \quad x_{.5} = x_{(p)}$$

n par → dois valores centrais

$$\text{posições} \left\{ \begin{array}{l} p = \frac{n+1}{2} \rightarrow p_1 \\ \phantom{p = \frac{n+1}{2}} \rightarrow p_2 \end{array} \right. \quad x_{.5} = \frac{x_{(p_1)} + x_{(p_2)}}{2}$$

Quando p não é inteiro, tomamos os dois inteiros mais próximos.

Exemplo:

$X = \text{Peso (kg)}$

$x_i = 5, 9, 7, 4, 12, 10$

1. Ordenar os dados

$$x_{(i)} = 4, 5, \boxed{7, 9}, 10, 12$$

2. Determinar a posição (p) da mediana

$n = 6$ (par)

$$p = \frac{n+1}{2} = \frac{6+1}{2} = 3,5 \begin{array}{l} \rightarrow p_1 = 3 \\ \rightarrow p_2 = 4 \end{array} \quad \begin{array}{l} x_{.5} = \frac{x_{(3)} + x_{(4)}}{2} \\ x_{.5} = \frac{7+9}{2} = 8 \end{array}$$

$$\boxed{x_{.5} = 8 \text{ kg}}$$

Moda (Mo)

- ⇒ É o valor de maior ocorrência num conjunto de dados.
- ⇒ É a única medida que pode não existir e, existindo, pode não ser única.

Exemplos:

Interpretação para variáveis discretas:

X = número de unidades defeituosas em um lote de produção

1. $x_i = 5, 5, 3, 7, 9, 5, 4, 1$ $Mo = 5$
2. $x_i = 9, 5, 4, 5, 7, 1, 2, 2$ $Mo = 2$ e 5 (conjunto bimodal)
3. $x_i = 1, 3, 8, 4, 5, 9, 2, 11$ não existe Mo (conjunto amodal)

Interpretação para variáveis contínuas:

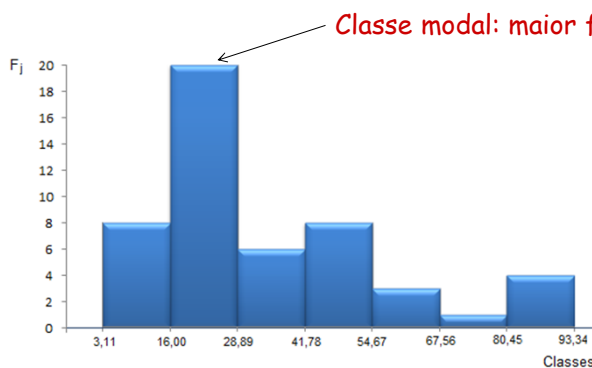
Valores gastos (em reais) pelas primeiras 50 pessoas que entraram em um determinado Supermercado, no dia 01/01/2000.

3,11 8,88 9,26 10,81 12,69 13,78 15,23 15,62 17,00 17,39
 18,36 18,43 19,27 19,50 19,54 20,16 20,59 22,22 23,04 24,47
 24,58 25,13 26,24 26,26 27,65 28,06 28,08 28,38 32,03 36,37
 38,64 38,98 39,16 41,02 42,97 44,08 44,67 45,40 46,69 48,65
 50,39 52,75 54,80 59,07 61,22 70,32 82,70 85,76 86,37 93,34

Valores de **variáveis contínuas**, em geral, não se repetem.

A moda tem importância apenas conceitual.

Ela está relacionada com o pico da distribuição

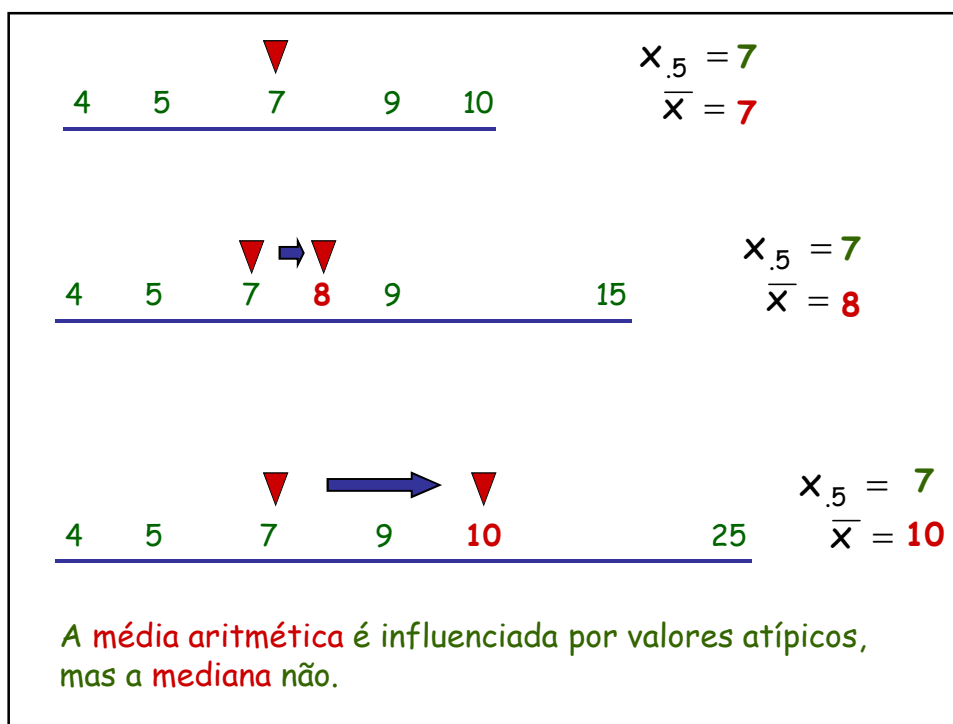
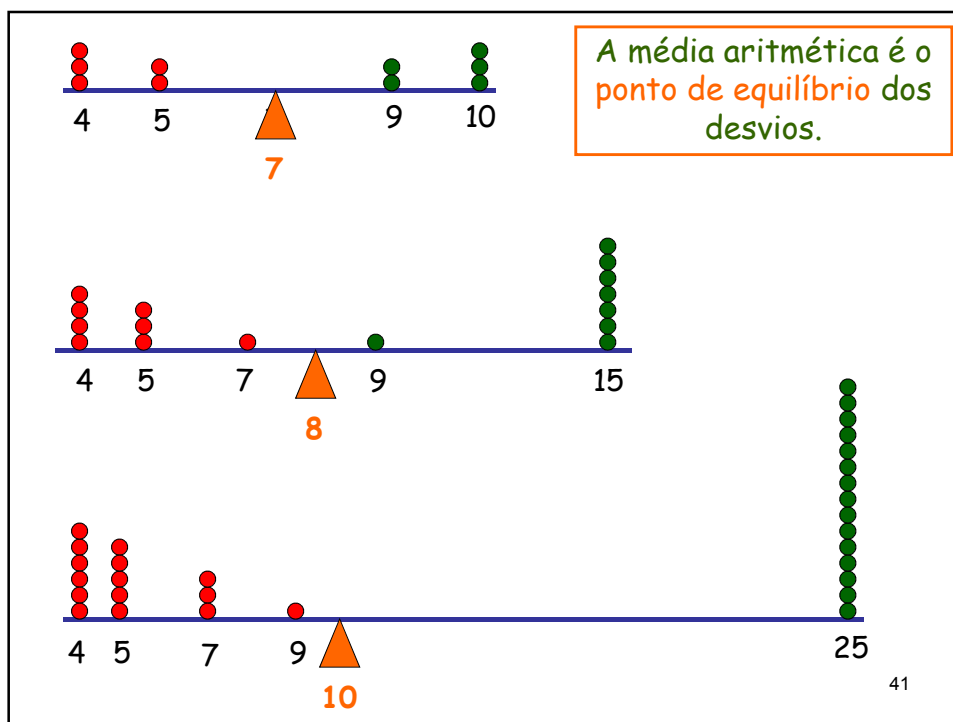


Moda é o ponto da distribuição onde temos a maior frequência.

Principais características das três medidas de localização

Média aritmética

- ⇒ No cálculo da média participam todos os valores observados.
 - ⇒ É uma medida de fácil interpretação e presta-se muito bem a tratamentos estatísticos adicionais.
 - ⇒ É uma medida que sempre existe e é rígida e unicamente determinada.
 - ⇒ É um valor típico do conjunto de dados, podendo substituir todos os seus valores sem alterar o total.
 - ⇒ É o ponto de equilíbrio das observações.
 - ⇒ **Representativa:** quando a distribuição é simetria e têm maior frequência no centro.
- Desvantagem:** É uma medida altamente influenciada por valores atípicos ou discrepantes (não resistente).



Mediana

- ⇒ Define exatamente o centro de uma distribuição, mesmo quando os valores se distribuem assimetricamente em torno da média.
- ⇒ Pode ser determinada mesmo quando não se conhece todos os valores do conjunto de dados.
- ⇒ É uma medida que sempre existe e é única.
- ⇒ Pode ser utilizada para definir o meio de um número de objetos, propriedades ou qualidades que possam de alguma forma ser ordenados.
- ⇒ É uma medida resistente, ou seja, não sofre influência de valores discrepantes.

Desvantagem: É uma medida que não se presta a cálculos matemáticos.

43

Moda

- ⇒ É uma medida que têm existência real dentro do conjunto de dados e em grande número de vezes.
- ⇒ Não exige cálculo, apenas uma contagem.
- ⇒ Pode ser determinada também para variáveis qualitativas nominais.

Desvantagens:

Deixa sem representação todos os valores do conjunto de dados que não forem iguais a ela.

Não se presta a cálculos matemáticos.

Pode não existir.

Profa. Clause Piana

44

Variável categórica
qualitativa nominal

Tipo de veículo	F_j	
Carro de passageiro	248	← Classe Modal
Minivan	62	
Caminhão de 2 eixos	42	
Caminhão de multieixo	12	
Moto	55	
Barco a motor	9	

Profa. Clause Piana

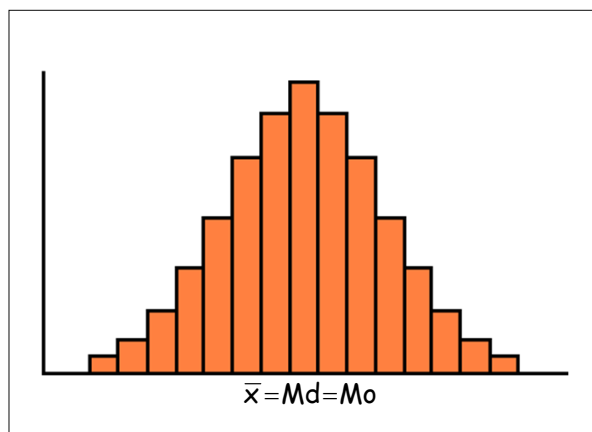
45

**Relação entre as três medidas
segundo o formato da distribuição**

Profa. Clause Piana

46

Relação entre as três medidas

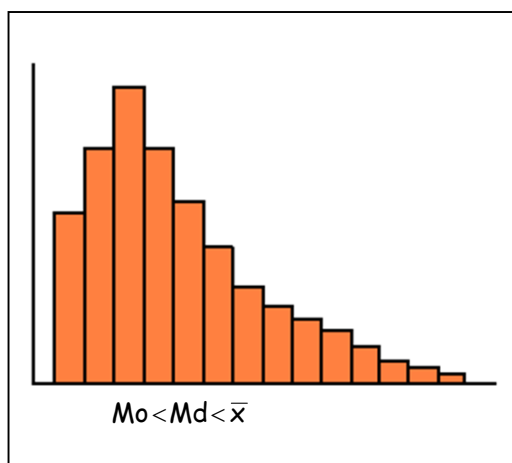


Distribuição unimodal simétrica

Profa. Clause Piana

47

Relação entre as três medidas

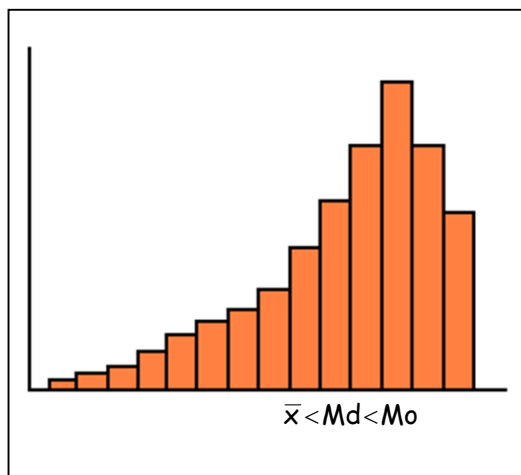


Distribuição unimodal e assimétrica, com cauda para a direita

Profa. Clause Piana

48

Relação entre as três medidas



Distribuição unimodal e assimétrica, com cauda para a esquerda

Profa. Clause Piana

49

Outras médias:

Média geométrica

Média harmônica

Profa. Clause Piana

50

Média geométrica

⇒ Definida como a n-ésima raiz do produto de n valores

Simples: $\bar{x}_g = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod x_i} = \left(\prod x_i\right)^{\frac{1}{n}}$

Ponderada: $\bar{x}_{gp} = \sqrt[\sum p_i]{x_1^{p_1} \times x_2^{p_2} \times \dots \times x_n^{p_n}} = \sqrt[\sum p_i]{\prod x_i^{p_i}} = \left(\prod x_i^{p_i}\right)^{\frac{1}{\sum p_i}}$

Sequência de valores em **progressão geométrica**:

7, 21, 63 → 21 é a média geométrica de 7 e 63

- ⇒ A **média geométrica** é definida somente para $x_i > 0$, para toda unidade de observação i.
- ⇒ Apropriada para calcular médias de razões, taxas de variação, índices econômicos e taxas de crescimento de microrganismos

51

Exemplo:

Se um investimento rende 10% no primeiro ano e 30% no segundo ano, qual o rendimento médio desse investimento?

Seja 100 reais o montante aplicado inicialmente.

Após esses dois anos o montante será: $100 \times 1,10 \times 1,30 = 143$

Tomando a **média aritmética**, temos $\bar{x} = \frac{1,10 + 1,30}{2} = 1,20$

Calculando o montante ao final dos dois anos:

$$100 \times 1,20 \times 1,20 = 144$$

Tomando a **média geométrica**, temos $\bar{x}_g = \sqrt{1,10 \times 1,30} = 1,1958$

Calculando o montante ao final dos dois anos:

Profa. Clause Piana $100 \times 1,1958 \times 1,1958 = 142,99$

52

Média harmônica

⇒ Definida como o inverso da média aritmética dos inversos dos n valores

Simples:
$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

Ponderada:
$$\bar{x}_{hp} = \frac{p_1 + p_2 + \dots + p_n}{\frac{1}{x_1} p_1 + \frac{1}{x_2} p_2 + \dots + \frac{1}{x_n} p_n} = \frac{\sum p_i}{\sum \frac{1}{x_i} p_i}$$

- ⇒ A **média harmônica** é definida somente para $x_i \neq 0$, para toda unidade de observação i .
- ⇒ Utilizada em teoria musical (frequências musicais) e em outras situações especiais (Física e Economia).

Profa. Clause Piana

53

Exemplo:

 **mais heterogêneo**
Notas: 3, 6 e 9

$$\bar{x} = \frac{3+6+9}{3} = 6$$

$$\bar{x}_h = \frac{3}{\frac{1}{3} + \frac{1}{6} + \frac{1}{9}} = 4,9$$


Notas: 4, 7 e 7

$$\bar{x} = \frac{4+7+7}{3} = 6$$

$$\bar{x}_h = \frac{3}{\frac{1}{4} + \frac{1}{7} + \frac{1}{7}} = 5,6$$

O candidato que tem notas mais homogêneas leva vantagem.

Exemplos de aplicação da média harmônica

Em algumas situações, é a média harmônica que provê a correta noção de média.

Exemplo 1.

Se metade da **distância** de uma viagem é feita a 40 km por hora e a outra metade da distância a 60 km por hora, então a velocidade média para a viagem é dada pela média harmônica:

$$\bar{x}_h = \frac{2}{\frac{1}{40} + \frac{1}{60}} = 48$$

Isso significa que o tempo total para a viagem seria o mesmo, se toda a viagem fosse feita a 48 quilômetros por hora.

Note-se, entretanto, que se se tivesse viajado por metade do **tempo** em uma velocidade e a outra metade na outra velocidade, a média aritmética, nesse caso 50 km por hora, proveria a correta noção de média.

Fonte: http://pt.wikipedia.org/wiki/M%C3%A9dia_harm%C3%B4nica)

Profa. Clause Piana

56

Exemplo 2.

Se um circuito elétrico contém duas resistências conectadas em paralelo, uma com uma resistência de 40 ohm e outra com 60 ohm, então, a média das duas resistências é:

$$\bar{x}_h = \frac{2}{\frac{1}{40} + \frac{1}{60}} = 48 \text{ ohm}$$

Isso significa que a resistência do circuito é a mesma que a de duas resistências de 48 ohm conectadas em paralelo.

Note-se que isso não deve ser confundido com sua resistência equivalente, 24 ohm, que é a resistência necessária para substituir as duas resistências em paralelo. A resistência equivalente é igual a metade do valor da média harmônica de duas resistências em paralelo.

Fonte: http://pt.wikipedia.org/wiki/M%C3%A9dia_harm%C3%B4nica)

Profa. Clause Piana

57

Exemplo 3.

Em finanças, a média harmônica é usada para calcular o custo médio de ações compradas durante um período.

Se um investidor compra 1000 reais em ações todo mês durante três meses e os preços na hora de compra são de 8 reais, 9 reais e 10 reais, então, o preço médio que o investidor pagou por ação é:

$$\bar{x}_h = \frac{3}{\frac{1}{8} + \frac{1}{9} + \frac{1}{10}} = 8,926 \text{ reais}$$

Note-se, entretanto, que se um investidor comprasse 1000 **ações** por mês, a média aritmética seria usada.

Fonte: http://pt.wikipedia.org/wiki/M%C3%A9dia_harm%C3%B4nica)

Profa. Clause Piana

58

Relação entre as três médias

$$\bar{X}_h \leq \bar{X}_g \leq \bar{X}$$

Exemplo:

$X = \text{peso (kg)}$ $x_i = 9, 7, 4, 5, 10$

$$\bar{x} = \frac{9+7+4+5+10}{5} = 7 \text{ kg}$$

$$\bar{x}_g = \sqrt[5]{9 \times 7 \times 4 \times 5 \times 10} = 6,608 \text{ kg}$$

$$\bar{x}_h = \frac{5}{\frac{1}{9} + \frac{1}{7} + \frac{1}{4} + \frac{1}{5} + \frac{1}{10}} = 6,219 \text{ kg}$$

A igualdade entre as três médias só se verifica quando todos os valores do conjunto são iguais.

59

2. Medidas separatrizes

Medidas separatrizes

São medidas descritivas que buscam dividir um conjunto de dados ordenado em proporções essencialmente iguais.

Mediana → divide o conjunto ordenado em **duas** partes

Quartis → dividem o conjunto ordenado em **quatro** partes

Decis → dividem o conjunto ordenado em **dez** partes

Percentis → dividem o conjunto ordenado em **cem** partes

Essas medidas delimitam proporções de valores no conjunto de dados ordenado.

Profa. Clause Piana

61

Quantil → é uma medida que delimita uma proporção qualquer de valores no conjunto ordenado. É denotado por

x_p , sendo p = proporção

Mediana → delimita a proporção 0,5

$Md = x_{0,5} \rightarrow p = 0,5$

Quartis → delimitam as proporções 0,25, 0,5 e 0,75

$Q_1 = x_{0,25} \rightarrow p = 0,25$

$Q_2 = x_{0,5} \rightarrow p = 0,5$

$Q_3 = x_{0,75} \rightarrow p = 0,75$

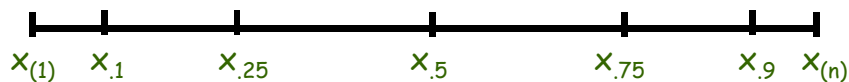
Decis → delimitam as proporções 0,1, 0,2, ... até 0,9

$D_1 = x_{0,1} \rightarrow p = 0,1 \quad \dots \quad D_5 = x_{0,5} \rightarrow p = 0,5 \quad \dots \quad D_9 = x_{0,9} \rightarrow p = 0,9$

Percentis → delimitam as proporções 0,01, 0,02, ..., até 0,99

$P_1 \rightarrow p = 0,01 \quad \dots \quad P_{50} \rightarrow p = 0,5 \quad \dots \quad P_{99} \rightarrow p = 0,99$

Quantis (x_p)



$x_{.1}$: pelo menos 0,10 dos valores são menores ou iguais a $x_{.1}$

$x_{.25}$: pelo menos 0,25 dos valores são menores ou iguais a $x_{.25}$

$x_{.5}$: pelo menos 0,50 dos valores são menores ou iguais a $x_{.5}$

$x_{.75}$: pelo menos 0,75 dos valores são menores ou iguais a $x_{.75}$

$x_{.9}$: pelo menos 0,90 dos valores são menores ou iguais a $x_{.9}$

Determinação do quantil (x_p) pelo método de inversão da função de distribuição empírica (FDE):

1. Ordenar os dados
2. Multiplicar a proporção (p) pelo número de observações (n)
3. Fazer $np = j + f$, onde j = parte inteira e f = parte decimal

$$\left\{ \begin{array}{l} \text{Se } f = 0 \Rightarrow x_p = \frac{x_{(j)} + x_{(j+1)}}{2} \\ \text{Se } f > 0 \Rightarrow x_p = x_{(j+1)} \end{array} \right.$$

Nem sempre há um valor do conjunto para a proporção desejada.

Exercício proposto:

Foram registrados os tempos de frenagem (em décimos de segundos) para 21 motoristas que dirigiam a 30 milhas por hora. Os valores obtidos foram:

69 57 70 80 46 61 65 74 75 55 67
56 71 72 61 66 58 68 70 68 59

Para o conjunto de valores, calcule os quartis e interprete esses valores.

1 milha = 1,61 km

Medidas para dados agrupados em classe

Distribuições de variáveis discretas - medidas exatas

A partir da distribuição de frequências, calcule as medidas de localização (média, mediana e moda).

Número de erros em um conjunto de caracteres (*string*) de 1.000 bits.

j	Classe	F_j	F'_j	f_j	f'_j
1	0	55	55	0,1571	0,1571
2	1	60	115	0,1714	0,3286
3	2	112	227	0,32	0,6486
4	3	82	309	0,2343	0,8829
5	4	31	340	0,0886	0,9714
6	5	8	348	0,0229	0,9943
7	6	2	350	0,0057	1,0000
	Σ	350	-	1,0000	-

Moda → Classe 2
 Classe modal → Classe 3
 Classe mediana → Classe 4

Classe que contém a mediana → Classe 3
 Classe que contém o terceiro quartil → Classe 4

Média ponderada $\bar{x}_p = \frac{\sum c_j F_j}{\sum F_j} = 2,017$

Distribuições de variáveis contínuas - medidas aproximadas

A partir da distribuição de frequências, calcule as medidas de localização (média, mediana e moda).

Valores gastos (em reais) pelas primeiras 50 pessoas que entraram em um determinado Supermercado, no dia 01/01/2000.

j	Classe	c_j	F_j	f_j	f'_j
1	3,11 — 16,00	9,56	8	0,16	0,16
2	16,00 — 28,89	22,45	20	0,4	0,56
3	28,89 — 41,78	35,34	6	0,12	0,68
4	41,78 — 54,67	48,23	8	0,16	0,84
5	54,67 — 67,56	61,12	3	0,06	0,9
6	67,56 — 80,45	74,01	1	0,02	0,92
7	80,45 — 93,34	86,90	4	0,08	1
	Σ	-	50	1	-

Classe modal → Classe 2
 Classe mediana → Classe 3

Classe que contém a mediana e o primeiro quartil → Classe 2
 Classe que contém o terceiro quartil → Classe 4

Média ponderada $\bar{x}_p = \frac{\sum c_j F_j}{\sum F_j} = 37,57$ reais

3. Medidas de variação ou dispersão

Observando os três conjuntos de dados abaixo, verificamos que uma medida de tendência central não é suficiente para diferenciá-los. **Que característica dos dados poderia evidenciar que os conjuntos são diferentes?**

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4

A variabilidade !!!

Medidas de variação ou dispersão

Objetivo → indicar quanto os valores diferem entre si ou quanto eles se afastam da média

⇒ Complementam as medidas de tendência central

Medidas de variação mais utilizadas:

- ◆ Amplitude total
- ◆ Amplitude interquartílica
- ◆ Variância
- ◆ Desvio padrão
- ◆ Coeficiente de variação

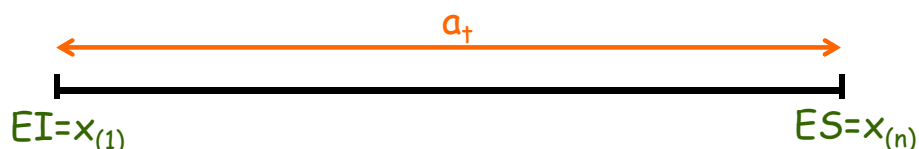
Profa. Clause Piana

71

Amplitude total (a_t)

⇒ Fornece uma idéia rudimentar de variação

⇒ É obtida pela diferença entre o maior valor e o menor valor de um conjunto de dados



$$a_t = ES - EI$$

ES: extremo superior do conjunto de dados ordenado

EI: extremo inferior do conjunto de dados ordenado

Profa. Clause Piana

$$a_t = x_{(n)} - x_{(1)}$$

72

Exemplo:

$X = \text{peso (kg)}$

$$x_i = 9, 7, 4, 5, 10$$

$$a = ES - EI = 10 - 4 = 6 \text{ kg}$$



Significado: todos os valores do conjunto de dados diferem, no máximo, em 6kg

Desvantagens

- ◆ pouco precisa
- ◆ extremamente influenciada por valores discrepantes

Profa. Clause Piana

73

Cálculo da amplitude total no exemplo:

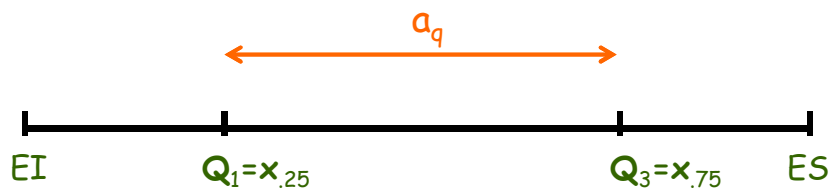
i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4
a_t	0	4	7

Profa. Clause Piana

74

Amplitude interquartílica (a_q)

⇒ É obtida pela diferença entre o terceiro e o primeiro quartis



$$a_q = Q_3 - Q_1$$

Q_1 : primeiro quartil

Q_3 : terceiro quartil

Profa. Clause Piana

75

Exemplo:

$X = \text{peso (kg)}$

$x_i = 3, 3, 4, 6, 7, 9, 9, 11, 12$

$Q_1 = 4 \text{ kg}$ e $Q_3 = 9 \text{ kg}$

$$a_q = Q_3 - Q_1 = 9 - 4 = 5 \text{ kg}$$



Significado: 50% dos valores mais centrais do conjunto, diferem, no máximo, em 5 kg

Vantagem

- ♦ medida resistente (não é afetada por valores discrepantes)

Profa. Clause Piana

76

Variância (s^2)

- ⇒ Medida de variação mais utilizada:
 - ◆ facilidade de compreensão
 - ◆ propriedades matemáticas e estatísticas
- ⇒ Considera o desvio da média como unidade básica da variação:

Desvio: $(x_i - \bar{x})$

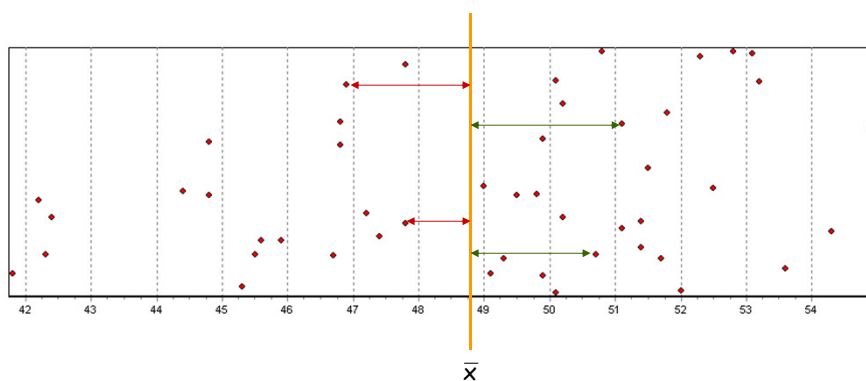
mede quanto cada valor varia em relação à média

Profa. Clause Piana

77

n=48

Desvio: $(x_i - \bar{x})$



Os desvios medem a variação (ou afastamento) de cada observação em relação à média aritmética

Profa. Clause Piana

78

Exemplo:

$$x_i = 4, 5, 7, 9, 10$$

$$\bar{x} = 7$$

$$(x_i - \bar{x}) \begin{cases} 4 - 7 = -3 \\ 5 - 7 = -2 \\ 7 - 7 = 0 \\ 9 - 7 = 2 \\ 10 - 7 = 3 \end{cases}$$

variação de cada x_i em relação à média

Média dos desvios → variação média do conjunto de valores

$$\begin{array}{l} \text{soma de todos os desvios} \rightarrow \sum (x_i - \bar{x}) = 0 \\ \text{total de desvios somados} \rightarrow n \end{array}$$

$$3^{\text{a}} \text{ propriedade da média} \rightarrow \sum (x_i - \bar{x}) = 0$$

79

Solução: elevar os desvios ao quadrado → desvios negativos ficam positivos e podem ser somados

$$\begin{array}{l} \text{soma dos quadrados dos desvios} \rightarrow \sum (x_i - \bar{x})^2 \\ \text{total de desvios somados} \rightarrow n \end{array}$$

Média dos quadrados dos desvios

Variância: definida como a média dos quadrados dos desvios

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

número de graus de liberdade ou desvios independentes

Profa. Clause Piana

80

Por que utilizar n-1 como denominador?

Porque este denominador confere à variância melhores propriedades estatísticas (importante na inferência estatística).

- ⇒ Quando o objetivo for apenas **descrever a variação de um conjunto de valores**, podemos usar o denominador n.

$$s_n^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

- ⇒ Quando o objetivo for **estimar a variação de uma população** por meio da variação de um conjunto de valores (amostra), **devemos** usar o denominador n-1.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

Profa. Clause Piana

81

Derivação de outra fórmula da variância:

$$\begin{aligned}
 s^2 &= \frac{\sum (x_i - \bar{x})^2}{n-1} \longrightarrow \sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
 &= \sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2 \\
 &= \sum x_i^2 - 2 \frac{\sum x_i}{n} \sum x_i + n \left(\frac{\sum x_i}{n} \right)^2 \\
 &= \sum x_i^2 - 2 \frac{(\sum x_i)^2}{n} + n \frac{(\sum x_i)^2}{n^2} \\
 &= \sum x_i^2 - 2 \frac{(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n} \\
 \sum (x_i - \bar{x})^2 &= \sum x_i^2 - \frac{(\sum x_i)^2}{n}
 \end{aligned}$$

Profa. Clause Piana

82

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \leftarrow \text{Fórmula de definição}$$



$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} \quad \leftarrow \text{Fórmula prática}$$

Profa. Clause Piana

83

Exemplo:

X = peso (kg)

$$x_i = 9, 7, 5, 10, 4 \rightarrow \bar{x} = 7 \text{ kg}$$

$$\begin{aligned} s^2 &= \frac{\sum (x_i - \bar{x})^2}{n-1} \\ &= \frac{(9-7)^2 + (7-7)^2 + (5-7)^2 + (10-7)^2 + (4-7)^2}{5-1} \\ &= \frac{4+0+4+9+9}{4} = \frac{26}{4} = 6,5 \end{aligned}$$

$$s^2 = 6,5 \text{ kg}^2 \quad \leftarrow \text{unidade de medida fica elevada ao quadrado}$$

Profa. Clause Piana

84

Propriedades algébricas da variância

1ª propriedade: A variância de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é zero.

$$x_i = 4, 4, 4, 4, 4$$

$$s^2 = \frac{(4-4)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2 + (4-4)^2}{5-1}$$

$$s^2 = 0$$

2ª propriedade: Se somarmos uma constante c a todos os valores de um conjunto de dados, a variância destes dados não se altera.

Verificação numérica:

$$x_i = 9, 7, 5, 10, 4 \begin{cases} \bar{x} = 7 \\ s^2 = 6,5 \end{cases}$$

Somar $c=2$

$$x_{i+2} = 11, 9, 7, 12, 6 \begin{cases} \bar{x}_{x+2} = 9 \\ s^2_{x+2} = ? \end{cases} \rightarrow \boxed{\bar{x}_{x+c} = \bar{x} + c}$$

$$\begin{aligned} s^2_{x+2} &= \frac{(11-9)^2 + (9-9)^2 + (7-9)^2 + (12-9)^2 + (6-9)^2}{5-1} \\ &= \frac{4+0+4+9+9}{4} = \frac{26}{4} = 6,5 \text{ kg}^2 \end{aligned}$$

Profa. Clause Piana

87

Demonstração: $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

$$s^2_{x+c} = \frac{\sum [(x_i + c) - (\bar{x} + c)]^2}{n-1}$$

$$= \frac{\sum (x_i - \bar{x} + c - c)^2}{n-1}$$

$$= \frac{\sum (x_i - \bar{x})^2}{n-1} = s^2$$

$$\boxed{s^2_{x+c} = s^2}$$

Profa. Clause Piana

88

3ª propriedade: Se multiplicarmos todos os valores de um conjunto de dados por uma constante c , a variância destes dados fica multiplicada pelo quadrado desta constante.

Verificação numérica:

$$x_i = 9, 7, 5, 10, 4 \begin{cases} \bar{x} = 7 \\ s^2 = 6,5 \end{cases}$$

Multiplicar por $c=2$

$$2x_i = 18, 14, 10, 20, 8 \begin{cases} \bar{x}_{2x} = 14 \rightarrow \boxed{\bar{x}_{cx} = c\bar{x}} \\ s_{2x}^2 = 26 \rightarrow \boxed{s_{cx}^2 = c^2 s^2} \end{cases}$$

$$s_{2x}^2 = \frac{(18-14)^2 + (14-14)^2 + (10-14)^2 + (20-14)^2 + (8-14)^2}{5-1}$$

$$= \frac{16+0+16+36+36}{4} = \frac{104}{4} = 26 \text{ kg}^2 = 2^2 \times 6,5$$

Profa. Clause Piana

89

Demonstração:

$$s_{xc}^2 = \frac{\sum (x_i c - \bar{x} c)^2}{n-1}$$

$$= \frac{\sum [c(x_i - \bar{x})]^2}{n-1}$$

$$= \frac{\sum c^2 (x_i - \bar{x})^2}{n-1}$$

$$= \frac{c^2 \sum (x_i - \bar{x})^2}{n-1}$$

$$= c^2 \frac{\sum (x_i - \bar{x})^2}{n-1} = c^2 s^2$$

$$\boxed{s_{cx}^2 = c^2 s^2}$$

Profa. Clause Piana

90

Resumo: Propriedades da variância

1ª propriedade: A variância de um conjunto de dados que não varia, ou seja, cujos valores são uma constante, é zero.

2ª propriedade: Se somarmos uma constante c a todos os valores de um conjunto de dados, a variância destes dados não se altera.

$$s_{x+c}^2 = s^2$$

3ª propriedade: Se multiplicarmos todos os valores de um conjunto de dados por uma constante c , a variância destes dados fica multiplicada pelo quadrado desta constante.

$$s_{cx}^2 = c^2 s^2$$

Profa. Clause Piana

91

Cálculo da variância no exemplo:

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4
Amplitude total	0	4	7
Variância	0	2,5	7,5

Profa. Clause Piana

92

Desvantagens da variância:

1. Como a variância é calculada a partir da média, é uma medida pouco resistente, ou seja, muito influenciada por valores atípicos.
2. Como a unidade de medida fica elevada ao quadrado, a interpretação da variância se torna mais difícil.

Para solucionar o problema de interpretação da variância surge outra medida: o **desvio padrão**.

Profa. Clause Piana

93

Desvio padrão (s)

⇒ É definido como a raiz quadrada positiva da variância

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

Exemplo:

X = peso (kg)

$x_i = 9, 7, 5, 10, 4$

$\bar{x} = 7 \text{ kg}$

$s^2 = 6,5 \text{ kg}^2$

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

$$s = \sqrt{6,5 \text{ kg}^2}$$

$$s = 2,55 \text{ kg}$$

Profa. Clause Piana

94

Apresentação do desvio padrão:

$$\bar{x} \pm s$$

$$7 \pm 2,55$$

Peso médio de 7 kg com uma variação média de 2,55 kg acima e abaixo da média.

Significado: variação média em torno da média aritmética

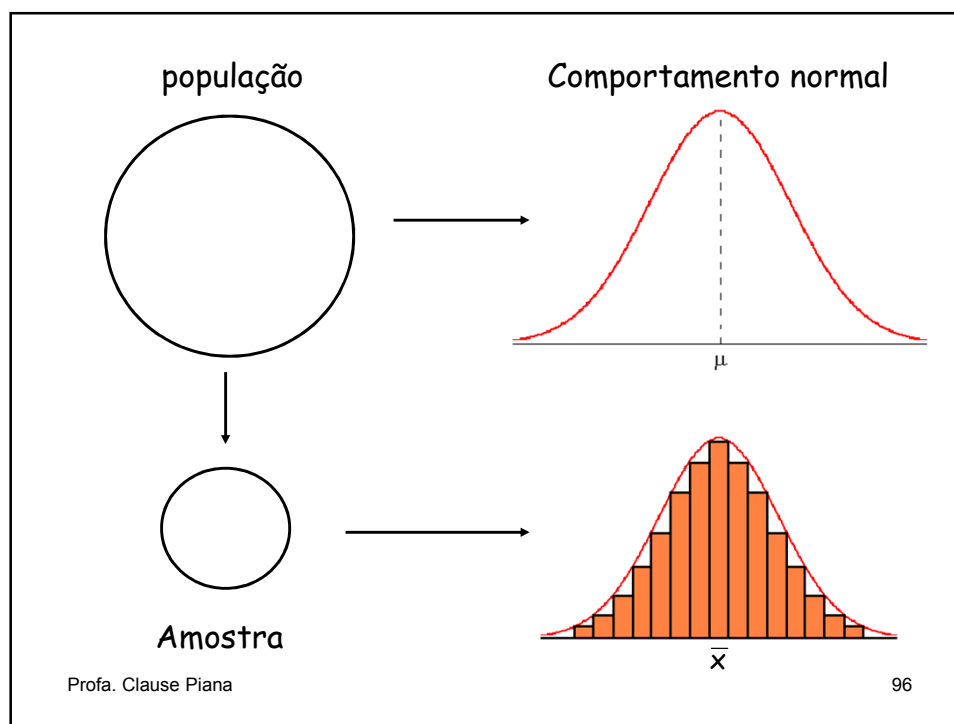
Vantagens

- ⇒ Facilidade de interpretação
- ⇒ É possível associar proporções de valores a intervalos entre a média e o desvio padrão

Numa distribuição **simétrica e unimodal**

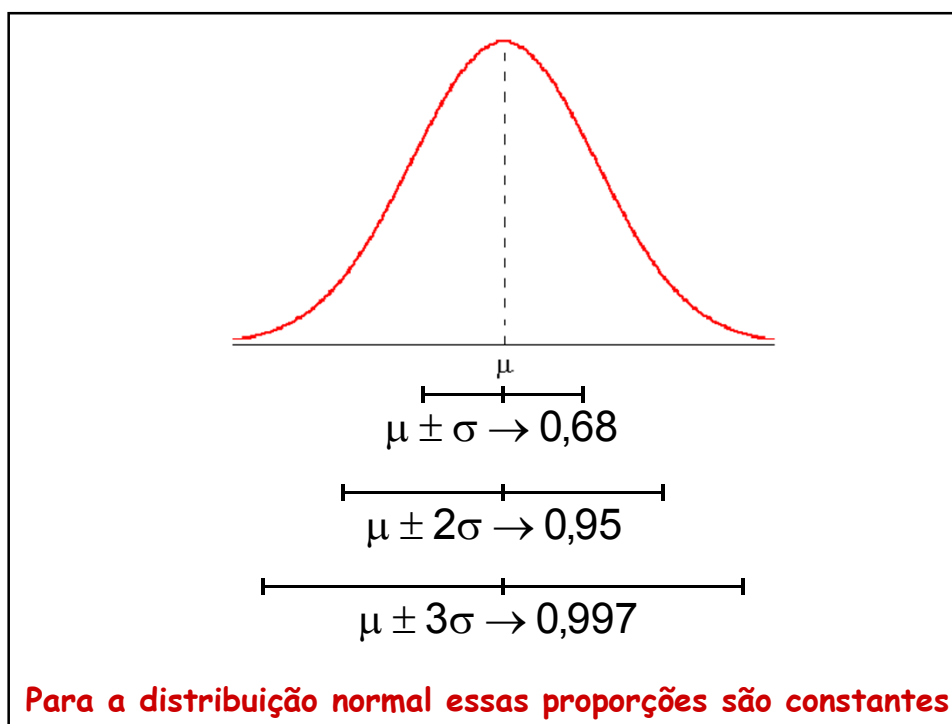
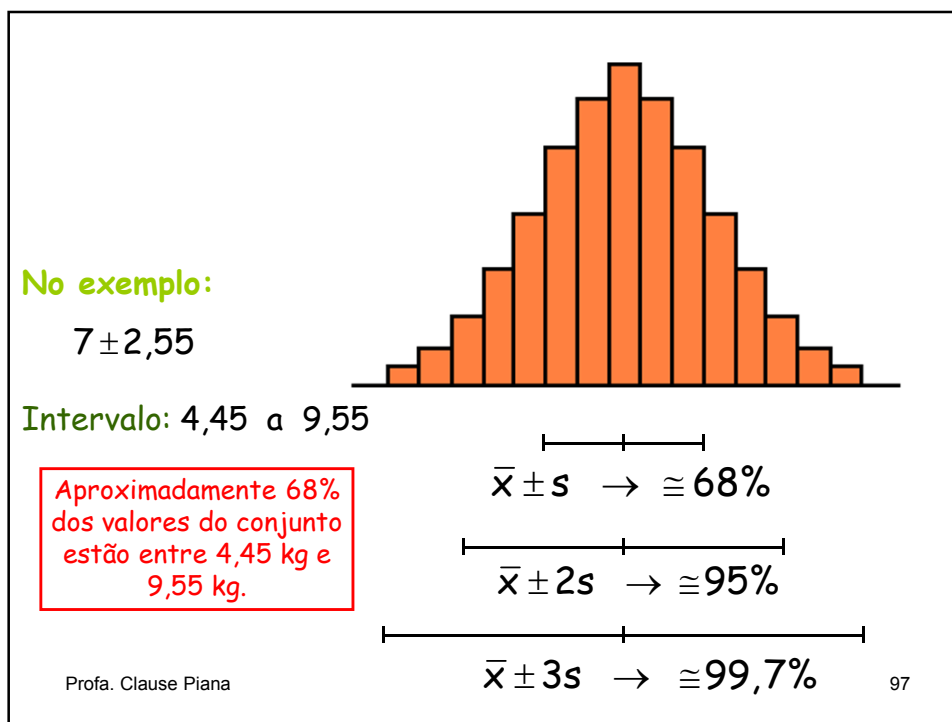
$$\bar{x} \pm s \rightarrow \text{proporção aproximada: } 0,68$$

95



Prof. Clause Piana

96



Cálculo do desvio padrão no exemplo:

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4
Amplitude total	0	4	7
Variância	0	2,5	7,5
Desvio padrão	0	1,581	2,739

Profa. Clause Piana

99

Coeficiente de Variação (CV)

⇒ O coeficiente de variação é definido como a proporção (ou percentual) da média representada pelo desvio padrão.

$$CV = \frac{s}{\bar{x}} 100$$

⇒ É uma medida que somente tem sentido para variáveis mensuradas em *escala de razão* e cuja média não é zero nem está próxima de 0.

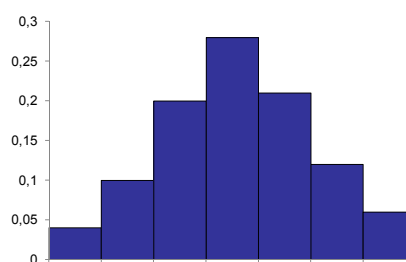
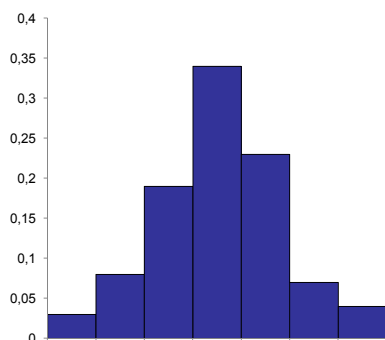
Exemplo: $X = \text{peso (kg)}$ $\bar{x} = 7 \text{ kg}$

$$x_i = 9, 7, 5, 10, 4 \quad \left\{ \begin{array}{l} \bar{x} = 7 \text{ kg} \\ s = 2,55 \text{ kg} \end{array} \right.$$

$$CV = \frac{s}{\bar{x}} 100 = \frac{2,55 \text{ kg}}{7 \text{ kg}} 100 = 36,4\%$$

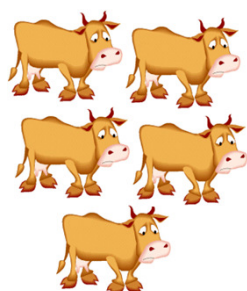
100

⇒ O CV é a medida mais utilizada para comparar variabilidades de diferentes conjuntos de dados



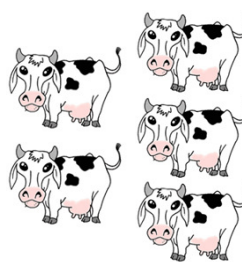
Exemplo 1:

Consideremos que x_{1i} e x_{2i} são conjuntos de valores referentes a **produção diária de leite** (em kg) de vacas das raças Jersey e Holandesa, para os quais foram obtidas as seguintes medidas:



Jersey (X_1): $\bar{x}_1 = 13$ kg
 $s_1 = 3,4$ kg

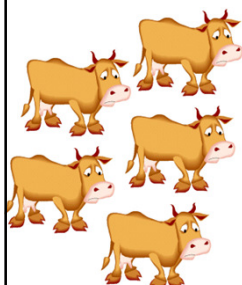
Profa. Clause Piana



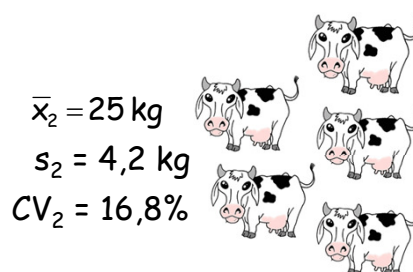
Holandesa (X_2): $\bar{x}_2 = 25$ kg
 $s_2 = 4,2$ kg

102

Qual grupo varia mais em relação à produção de leite?



$$\begin{aligned}\bar{x}_1 &= 13 \text{ kg} \\ s_1 &= 3,4 \text{ kg} \\ CV_1 &= 26,2\%\end{aligned}$$



$$\begin{aligned}\bar{x}_2 &= 25 \text{ kg} \\ s_2 &= 4,2 \text{ kg} \\ CV_2 &= 16,8\%\end{aligned}$$

O maior desvio padrão, quando comparado à sua média, representou menor variação.

Quando as médias são diferentes, devemos usar o CV.

Profa. Clause Piana

103

Exemplo 2:

Consideremos, agora, que x_i e y_i são conjuntos de valores referentes a **alturas** (em cm) e **pesos** (em kg) de um grupo de estudantes, para os quais foram obtidas as seguintes medidas:

$$\text{Altura (X): } \bar{x} = 178 \text{ cm} \quad s_x = 28 \text{ cm} \quad CV_x = 15,73\%$$

$$\text{Peso (Y): } \bar{y} = 62 \text{ kg} \quad s_y = 10 \text{ kg} \quad CV_y = 16,13\%$$

Peso e altura não são grandezas comparáveis.

Quando as unidades de medida são diferentes, devemos usar o CV.

Profa. Clause Piana

104

- ⇒ Para a comparação de variabilidades o uso do CV é particularmente recomendado em duas situações:
- ◆ quando as médias dos conjuntos são diferentes
 - ◆ quando as unidades de medida são diferentes

Se a média e a unidade de medida dos conjuntos são iguais, qualquer medida de variação pode ser utilizada para a comparação.

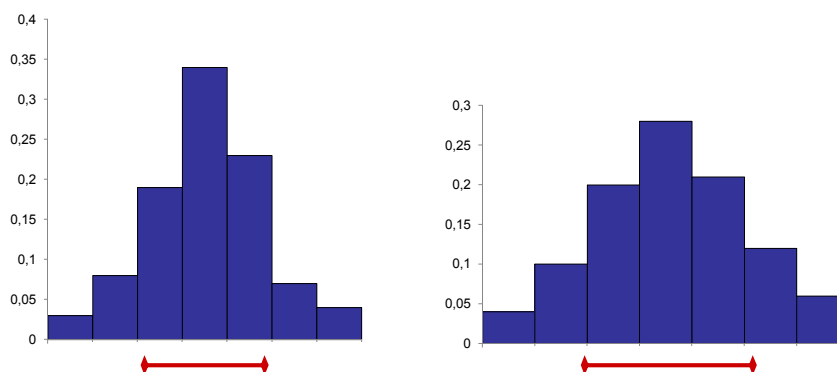
Vantagens:

- É desprovido de unidade de medida (expresso em percentagem)
- É uma medida relativa, pois relaciona o desvio padrão com a sua respectiva média aritmética

Desvantagem

- O CV pode ter seu valor grandemente alterado enquanto a variabilidade permanece constante.

105



Em distribuições simétricas e unimodais observa-se que quanto menor é o desvio padrão em relação a média:

- ⇒ menor é o intervalo para a mesma proporção de valores
- ⇒ menor é a amplitude do histograma
- ⇒ menor é o coeficiente de variação

Cálculo do coeficiente de variação no exemplo:

i	x_i (kg)	y_i (kg)	z_i (kg)
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4
Amplitude total	0	4	7
Variância	0	2,5	7,5
Desvio padrão	0	1,581	2,739
CV	0	39,53 %	68,46 %

Profa. Clause Piana

107

Exercícios propostos:

- Um professor fez cinco provas, com os pesos de 3, 6, 8, 4 e 7, para as quais um aluno obteve as notas 10, 9, 8, 7, e 6, respectivamente. Calcule a média aritmética simples e a média aritmética ponderada para conjunto de notas.
- Contou-se o número de vendas de determinado produto durante os sete dias de uma semana, com os seguintes resultados:

14 20 20 20 15 16 18

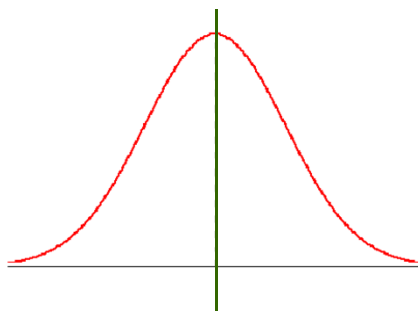
 - Determine a mediana e a moda.
 - Calcule a variância, o desvio padrão e o coeficiente de variação.

108

4. Medidas de formato

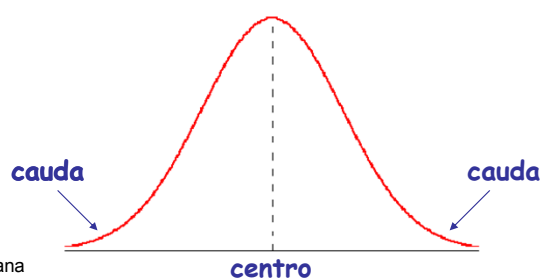
Formato de uma distribuição

- ⇒ O formato é um aspecto importante de uma distribuição. Está relacionado com as idéias de **simetria** e **curtose**.
- ⇒ A **simetria** em torno de um eixo indica que o formato da distribuição à esquerda e à direita desse eixo é o mesmo.



Formato de uma distribuição

- ⇒ O formato é um aspecto importante de uma distribuição. Está relacionado com as idéias de **simetria** e **curtose**.
- ⇒ A **simetria** em torno de um eixo indica que o formato da distribuição à esquerda e à direita desse eixo é o mesmo.
- ⇒ A **curtose** está relacionada com o grau de concentração das observações no centro e nas caudas da distribuição.



Profa. Clause Piana

111

Formato de uma distribuição

- ⇒ O formato é um aspecto importante de uma distribuição. Está relacionado com as idéias de **simetria** e **curtose**.
- ⇒ A **simetria** em torno de um eixo indica que o formato da distribuição à esquerda e à direita desse eixo é o mesmo.
- ⇒ A **curtose** está relacionada com o grau de concentração das observações no centro e nas caudas da distribuição.
- ⇒ A curtose não tem interpretação tão intuitiva quanto a simetria. Em geral, a curtose é discutida apenas para distribuições simétricas.
- ⇒ Existem várias formas de medir a assimetria e a curtose de uma distribuição.
- ⇒ Algumas dessas medidas se baseiam em quantidades denominadas **momentos**.

112

Momentos (m_r)

⇒ Quantidade fundamental na caracterização da distribuição de uma variável.

$$m_r = \frac{\sum (x_i - a)^r}{n}$$

Média dos desvios em relação à constante a na potência r

⇒ Pode ser definido como a média aritmética de uma potência qualquer (r) dos desvios dos valores de um conjunto em relação a uma constante (a) escolhida como origem.

Profa. Clause Piana

113

$$m_r = \frac{\sum (x_i - a)^r}{n}$$

Momentos {

Centrados na origem (ordinários) → $a = 0$

$$\frac{\sum (x_i - 0)^r}{n} = \frac{\sum x_i^r}{n} = m'_r$$

Centrados na média → $a = \bar{x}$

$$\frac{\sum (x_i - \bar{x})^r}{n} = m_r$$

Profa. Clause Piana

114

Momentos centrados na origem (ordinários) $m'_r = \frac{\sum x_i^r}{n}$

Para $r = 1$:

$$m'_1 = \frac{\sum x_i}{n} \leftarrow \text{Média de } X$$

Para $r = 2$:

$$m'_2 = \frac{\sum x_i^2}{n} \leftarrow \text{Média dos quadrados de } X$$

Para $r = 3$:

$$m'_3 = \frac{\sum x_i^3}{n} \leftarrow \text{Média dos cubos de } X$$

Profa. Clause Piana

115

Momentos centrados na média $m_r = \frac{\sum (x_i - \bar{x})^r}{n}$

Para $r = 1$: Média dos desvios

$$m_1 = \frac{\sum (x_i - \bar{x})}{n} = 0$$

Para $r = 2$: Média dos quadrados dos desvios

$$m_2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

Profa. Clause Piana

116

Para $r = 3$: **Média dos cubos dos desvios**

$$m_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$

Para $r = 4$: **Média dos desvios na potência quatro**

$$m_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

Profa. Clause Piana

117

Medidas de assimetria

- ⇒ Informam se a maioria dos valores se localiza à esquerda, ou à direita, ou se estão distribuídos uniformemente em torno da média aritmética.
- ⇒ Uma das medidas de assimetria mais precisas é o **coeficiente de assimetria**, calculado a partir do segundo e do terceiro momentos centrados na média:

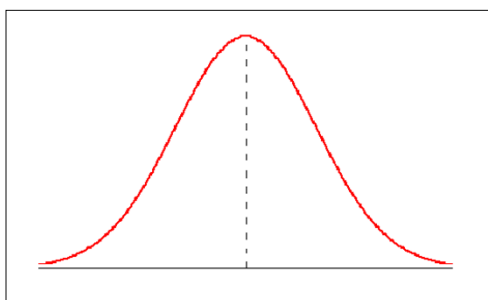
$$a_3 = \frac{m_3}{m_2 \sqrt{m_2}}$$

- ⇒ Indica o grau e o sentido do afastamento da simetria.

Profa. Clause Piana

118

Distribuição normal → Simétrica



$$a_3 = 3$$

Profa. Clause Piana

119

Classificação quanto à simetria

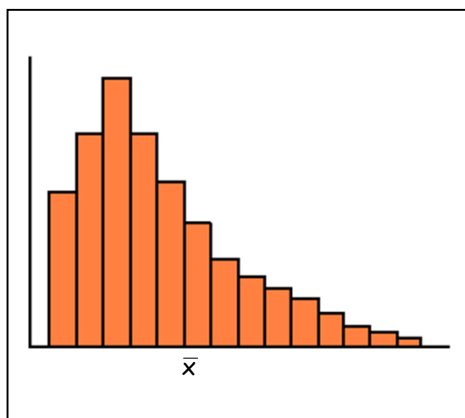
Se $a_3=0$, a distribuição é classificada como **simétrica**, indicando os valores estão uniformemente distribuídos em torno da média.



Profa. Clause Piana

120

Se $a_3 > 0$, a distribuição é classificada como **assimétrica positiva**, indicando que a maioria dos valores são menores ou se localizam à esquerda da média.



Profa. Clause Piana

121

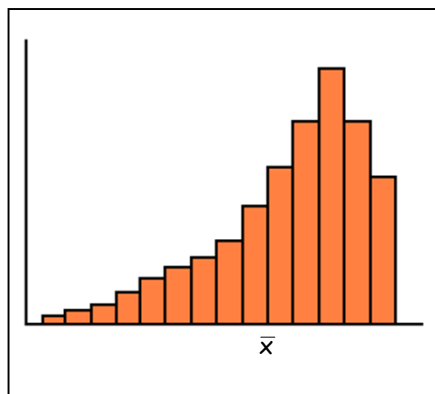
i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
1	4,5	-6	36	-216
2	5	-5,5	30,25	-166,375
3	5,1	-5,4	29,16	-157,464
4	5,1	-5,4	29,16	-157,464
5	5,5	-5	25	-125
6	5,5	-5	25	-125
7	5,7	-4,8	23,04	-110,592
8	6,4	-4,1	16,81	-68,921
9	6,5	-4	16	-64
10	7,5	-3	9	-27
11	7,7	-2,8	7,84	-21,952
12	7,9	-2,6	6,76	-17,576
13	7,9	-2,6	6,76	-17,576
14	8	-2,5	6,25	-15,625
15	8,4	-2,1	4,41	-9,261
16	8,5	-2	4	-8
17	8,9	-1,6	2,56	-4,096
18	9,5	-1	1	-1
19	9,6	-0,9	0,81	-0,729
20	9,6	-0,9	0,81	-0,729
21	10,1	-0,4	0,16	-0,064
22	12,4	1,9	3,61	6,859
23	12,7	2,2	4,84	10,648
24	13,1	2,6	6,76	17,576
25	13,1	2,6	6,76	17,576
26	14,1	3,6	12,96	46,656
27	14,4	3,9	15,21	59,319
28	14,7	4,2	17,64	74,088
29	16,9	6,4	40,96	262,144
30	17,1	6,6	43,56	287,496
31	18,9	8,4	70,56	592,704
32	19,2	8,7	75,69	658,503
33	27	16,5	272,25	4492,125
Soma	346,5	0	851,58	5211,27
Média	10,5			

X - quantidade de cádmio em peixes marinhos, observados em diferentes locais do Atlântico Norte

A maioria dos desvios tem sinal negativo, ou seja, a maioria dos valores é menor do que a média!

122

Se $\alpha_3 < 0$, a distribuição é classificada como **assimétrica negativa**, indicando que a maioria dos valores são maiores ou se localizam à direita da média aritmética.



Profa. Clause Piana

123

Interpretação teórica → **populações**

- ♦ Se $\alpha_3 < 0$ → **assimétrica negativa**
- ♦ Se $\alpha_3 = 0$ → **simétrica**
- ♦ Se $\alpha_3 > 0$ → **assimétrica positiva**

Interpretação prática → **amostras**

- ♦ Se $\alpha_3 < -0,5$ → **assimétrica negativa**
- ♦ Se $-0,5 \leq \alpha_3 \leq 0,5$ → **simétrica**
- ♦ Se $\alpha_3 > 0,5$ → **assimétrica positiva**

Profa. Clause Piana

124

Uma distribuição simétrica possui muitas vantagens:

- ⇒ Não há ambigüidades na indicação do centro. Numa distribuição unimodal a simetria implica que a média, mediana e moda coincidam ou, em termos amostrais, estejam muito próximas.
- ⇒ Em geral, a interpretação e as aplicações são mais simples.
- ⇒ Muitos procedimentos usuais pressupõem uma distribuição normal, que é uma distribuição simétrica.
- ⇒ Em muitas situações onde o modelo não é normal, é suficiente que a distribuição seja simétrica.

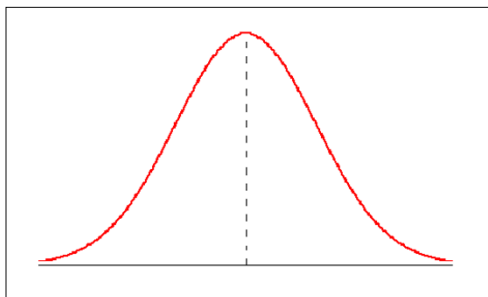
Medidas de curtose

- ⇒ Indicam a concentração de valores no centro ou nas caudas de uma distribuição.
- ⇒ O **coeficiente de curtose** é calculado a partir do segundo e do quarto momentos centrados na média.

$$a_4 = \frac{m_4}{m_2^2}$$

- ⇒ A classificação é feita tendo por base a curtose que ocorre na distribuição normal, classificada como mesocúrtica.

Distribuição normal → Mesocúrtica



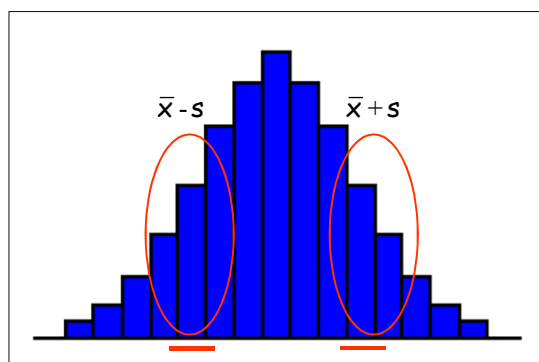
$$a_4=3$$

Profa. Clause Piana

127

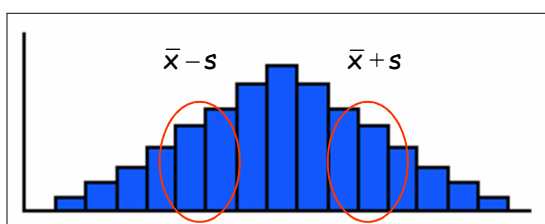
Classificação quanto à curtose

Se $a_4=3$, a distribuição é classificada como **mesocúrtica**, indicando que a concentração das observações ocorre de forma semelhante à da distribuição normal.



128

Se $a_4 < 3$, a distribuição é classificada como **platicúrtica**, indicando que ocorre baixa concentração de valores no centro, tornando a distribuição mais achatada que a distribuição normal.

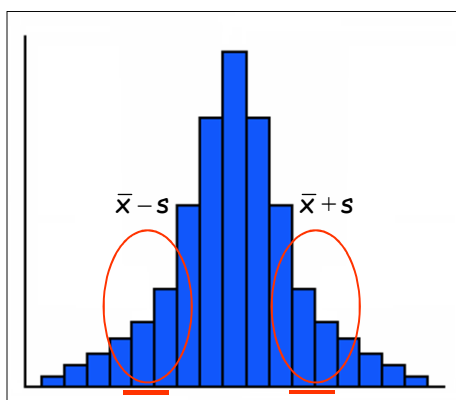


A concentração de valores nos eixos **média mais ou menos o desvio padrão** é maior que na distribuição normal.

Profa. Clause Piana

129

Se $a_4 > 3$, a distribuição é classificada como **leptocúrtica**, indicando que ocorre alta concentração de valores no centro e nas caudas, o que provoca um pico maior que o da distribuição normal.



A concentração de valores em torno dos eixos **média mais ou menos o desvio padrão** é menor do que na distribuição normal.

Profa. Clause Piana

130

Calcule os coeficientes de assimetria e curtose para as variáveis Y e Z.

i	x_i	y_i	z_i
1	4	2	1
2	4	5	8
3	4	4	5
4	4	6	4
5	4	3	2
Σ	20	20	20
Média	4	4	4

Profa. Clause Piana

131

Y = Peso (kg) Média = 4 kg

y_i	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^4$
2	-2			
5	1			
4	0			
6	2			
3	-1			
Σ	0			

$$m_2 = \frac{\sum (y_i - \bar{y})^2}{n}$$

$$m_3 = \frac{\sum (y_i - \bar{y})^3}{n}$$

$$m_4 = \frac{\sum (y_i - \bar{y})^4}{n}$$

$$a_3 = \frac{m_3}{m_2 \sqrt{m_2}}$$

$$a_4 = \frac{m_4}{m_2^2}$$

132

Z = Número de filhos		Média = 4 filhos		
z_i	$(z_i - \bar{z})$	$(z_i - \bar{z})^2$	$(z_i - \bar{z})^3$	$(z_i - \bar{z})^4$
1	-3			
8	4			
5	1			
4	0			
2	-2			
Σ	0			

$$m_2 = \frac{\sum (z_i - \bar{z})^2}{n}$$

$$m_3 = \frac{\sum (z_i - \bar{z})^3}{n}$$

$$m_4 = \frac{\sum (z_i - \bar{z})^4}{n}$$

$$a_3 = \frac{m_3}{m_2 \sqrt{m_2}}$$

$$a_4 = \frac{m_4}{m_2^2}$$

133

i	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
1	4,5	-6	36	-216	1296,00
2	5	-5,5	30,25	-166,375	915,06
3	5,1	-5,4	29,16	-157,464	850,31
4	5,1	-5,4	29,16	-157,464	850,31
5	5,5	-5	25	-125	625,00
6	5,5	-5	25	-125	625,00
7	5,7	-4,8	23,04	-110,592	530,84
8	6,4	-4,1	16,81	-68,921	282,58
9	6,5	-4	16	-64	256,00
10	7,5	-3	9	-27	81,00
11	7,7	-2,8	7,84	-21,952	61,47
12	7,9	-2,6	6,76	-17,576	45,70
13	7,9	-2,6	6,76	-17,576	45,70
14	8	-2,5	6,25	-15,625	39,06
15	8,4	-2,1	4,41	-9,261	19,45
16	8,5	-2	4	-8	16,00
17	8,9	-1,6	2,56	-4,096	6,554
18	9,5	-1	1	-1	1,000
19	9,6	-0,9	0,81	-0,729	0,6561
20	9,6	-0,9	0,81	-0,729	0,6561
21	10,1	-0,4	0,16	-0,064	0,02560
22	12,4	1,9	3,61	6,859	13,03
23	12,7	2,2	4,84	10,648	23,43
24	13,1	2,6	6,76	17,576	45,70
25	13,1	2,6	6,76	17,576	45,70
26	14,1	3,6	12,96	46,656	167,96
27	14,4	3,9	15,21	59,319	231,34
28	14,7	4,2	17,64	74,088	311,17
29	16,9	6,4	40,96	262,144	1677,72
30	17,1	6,6	43,56	287,496	1897,47
31	18,9	8,4	70,56	592,704	4978,71
32	19,2	8,7	75,69	658,503	5728,98
33	27	16,5	272,25	4492,125	74120,06
Soma	346,5	0	851,58	5211,27	95789,63
Média	10,5				

Como as medidas de formato estão baseadas nos desvios ao cubo e na quarta potência, o peso de desvios de grande magnitude pode ser desproporcional nos coeficientes.

Calcule os coeficientes de assimetria e curtose para os dados:

$$m_2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$m_3 = \frac{\sum (x_i - \bar{x})^3}{n}$$

$$m_4 = \frac{\sum (x_i - \bar{x})^4}{n}$$

$$a_3 = \frac{m_3}{m_2 \sqrt{m_2}}$$

$$a_4 = \frac{m_4}{m_2^2}$$

Bibliografia

FERREIRA, D.F. Estatística básica. Lavras: Editora UFLA, 2005.

SILVEIRA JUNIOR, P. ; MACHADO, A.A. ; ZONTA, E.P.;
SILVA, J.B. da. Curso de Estatística v.1. Pelotas:
Universidade Federal de Pelotas, 1992, 135p.

Sistema Galileu de Educação Estatística. Disponível em:
<http://www.galileu.esalq.usp.br/topico.html>