Universidade Federal de São Paulo

Medidas de tendência central (ou de posição), Quantis e Medidas de dispersão

Professor

Julio Cezar

AULA DE HOJE

- Medidas de tendência central ou de posição;
- Quantis;
- Medidas de dispersão.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Notação para variáveis e para a soma

O valores de uma variável **x** em um conjunto de dados pode ser representado simbolicamente por:

$$x_1, x_2, x_3, \dots, x_n$$

onde n é o número de medidas sobre a variável **x** no conjunto dos dados.

Os valores $x_1, x_2, x_3, \dots, x_n$ representam a 1ª observação, a 2ª observação, ..., *n*-ésima observação.

UNIFESP
Universidade Federal de São Paulo
1933

Notação para variáveis e para a soma

Notação de somatória $\sum$:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

soma de todos os **x** com *i* variando de 1 até n**.**

Exemplo: Se $x_1 = 2{,}5; x_2 = 2{,}0; x_3 = 3{,}2; x_4 = 2{,}5; x_5 = 3.$

$$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3\,. \text{ Então } \sum_{i=1}^{3} x_i = 7{,}7.$$

$$\sum_{i=1}^{4} (2 - x_i) = (2 - x_1) + (2 - x_2) + (2 - x_3) + (2 - x_4). \text{ Então } \sum_{i=1}^{4} x_i = -2{,}2.$$

UNIFESP

Universidade Federal de São Paulo

1933

PROPRIEDADES

Propriedades básicas de somatória:

Se a e b são constantes

$$\sum_{i=1}^{n} \mathrm{a} = n\mathrm{a}.$$

$$\sum_{i=1}^{n} \mathrm{b}x_i = \mathrm{b}\sum_{i=1}^{n} x_i \,.$$

$$\sum_{i=1}^{n} (\mathrm{b}x_i + \mathrm{a}) = \mathrm{b}\sum_{i=1}^{n} x_i + n\mathrm{a}.$$

$$\sum_{i=1}^{n} (x_i - \mathrm{a})^2 = \sum_{i=1}^{n} x_i^2 - 2\,\mathrm{a}\sum_{i=1}^{n} x_i + n\mathrm{a}^2.$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Observação

Se a somatória se estender a **todos os valores dos dados** é comum não colocar o índice na variável e os limites da soma, ou seja,

$$\sum_{i=1}^{n} x_i = \sum \mathbf{x}.$$

Se n for o número de elementos do conjunto de dados.

Medidas de tendência central (ou de posição)

UNIFESP
Universidade Federal de São Paulo
1933

MÉDIA ARITMÉTICA

A **média aritmética** é uma medida que indica onde está o "centro" da sua amostra ou da população. A **média amostral** é representada por $\bar{x}$ e a **média populacional** é representada por μ. As fórmulas de cálculo são apresentadas a seguir:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

MÉDIA ARITMÉTICA

Formalizando: Se $x_1, x_2, \dots, x_n$ são os n valores distintos ou não da variável X, a média aritmética ou simplesmente média de X pode ser escrita como

$$\bar{\mathrm{x}} = \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Exemplo: variável número de filhos:

$$1, 2, 0, 1, 2, 3, 0, 1, 2, 1, 0, 2, 2, 0, 5, 2, 1, 3, 2, 3$$

$$\boxed{\bar{\mathrm{x}} = \frac{1+2+0+1+2+3+0+1+2+1+0+2+2+0+5+2+1+3+2+3}{20} = \frac{33}{20} = 1{,}65}$$

UNIFESP

UNIVERSIDADE FEDERAL DE SÃO PAULO

1933

MÉDIA ARITMÉTICA

Com repetições de valores: Se x_1 possui f_1 observações iguais, x_2 possui f_2 observações iguais, etc. até, x_k possui f_k observações. Assim, a média de X pode ser escrita como

$$\bar{\mathrm{x}} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_k x_k}{n} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$$

Note que, $fr_i = \frac{f_i}{n}$ representa a frequência relativa (Já vimos!!!)

$$\bar{\mathrm{x}} = \sum_{i=1}^{k} fr_i x_i$$

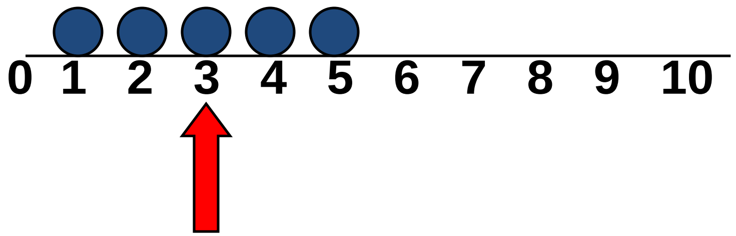
$$\boxed{\bar{\mathrm{x}} = 0{,}20 \times 0 + 0{,}25 \times 1 + 0{,}35 \times 2 + 0{,}15 \times 3 + 0{,}05 \times 5 = 1{,}65}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

MÉDIA ARITMÉTICA

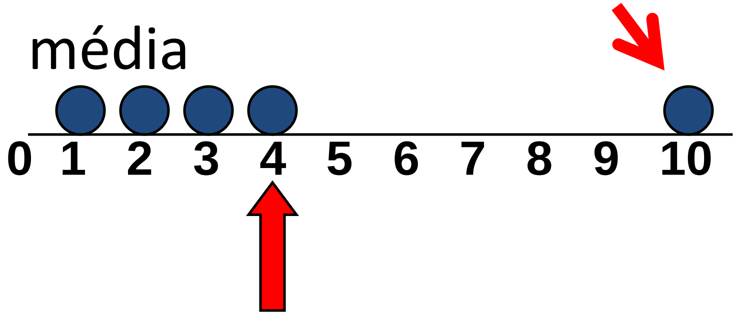
- Média = soma dos valores dividida pelo total de termos;
- Afetada por valores extremos.

Valores extremos afetam a média



Média = 3

$$\bar{x} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$$



Média = 4

$$\bar{x} = \frac{1+2+3+4+10}{5} = \frac{20}{5} = 4$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Média para dados agrupados em classes

Neste caso utiliza-se a frequências das classes e o valor xi da variável é substituido pelo ponto médio PMi de cada classe, ou seja,

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i \,.\, \mathrm{PM}_i}{\sum_{i=1}^{n} f_i}$$

•*k*: número de classes

•f_i: frequência na classe *i*

•PM_i: ponto médio da classe *i*

Ou em termos da frequencia relativa fr_i:

$$\bar{x} = \sum_{i=1}^{k} fr_i \,.\, \mathrm{PM}_i$$

UNIFESP
Universidade Federal de São Paulo
1933

Média para dados agrupados em classes

Exemplo: vamos usar os dados sobre os diâmetros de árvores (utilizados nas aulas anteriores).

classes	PM_i	f_i
10 |- 20	15	3
20 |- 30	25	6
30 |- 40	35	5
40 |- 50	45	4
50 |- 60	55	2
total		20

Temos 5 classes, então:

$$\bar{x} = \frac{\sum_{i=1}^{5} f_i \,.\mathrm{PM}_i}{20} = \frac{3\mathrm{x}15+6\mathrm{x}25+5\mathrm{x}35+4\mathrm{x}45+2\mathrm{x}55}{20}$$

$$\boxed{\bar{x} = 33{,}5}$$

Note que se calcularmos a média usando a soma sobre todos os valores vamos obter: $\bar{x}$ = 32,4.

A diferença ocorre pois ao agruparmos os dados em classes perdemos informação sobre os valores nas classes, que são substituidas pelo ponto médio.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

OUTRAS MEDIDAS DE CENTRO:

Média Geométrica:

$$\overline{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

Média Harmônica:

$$\overline{X}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$$

Ponto Médio (PM):

$$PM = \frac{\text{maior valor} + \text{menor valor}}{2}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

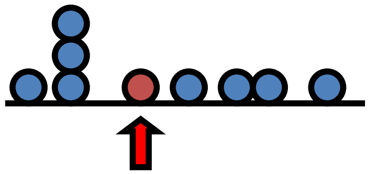
Medidas de Tendência Central

Tendência Central

- **Média Aritmética**

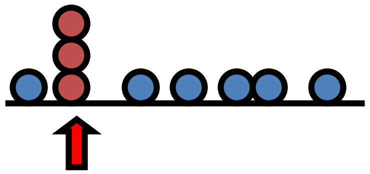
  $$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- **Mediana**

  Valor central dos valores ordenados

- **Moda**

  Valor observado com maior frequência

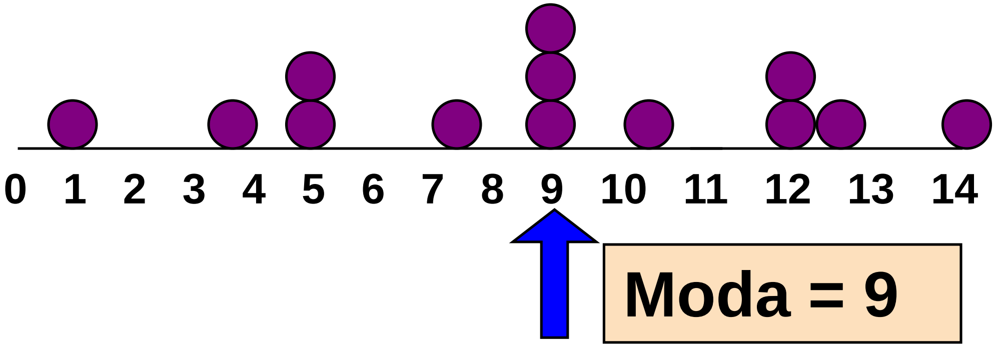
- **Média Geométrica**

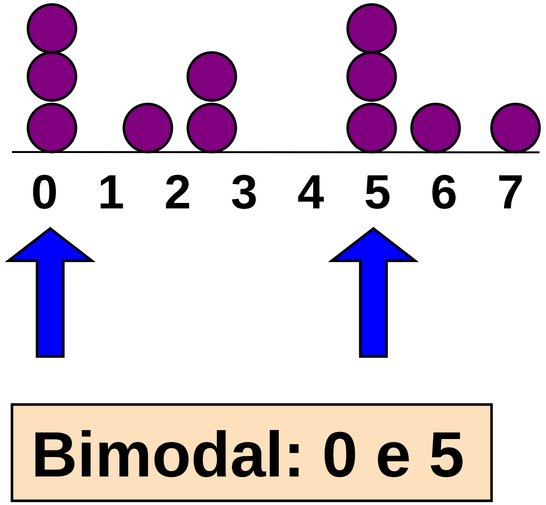
  $$\bar{X}_G = (X_1 \times X_2 \times \cdots \times X_n)^{1/n}$$

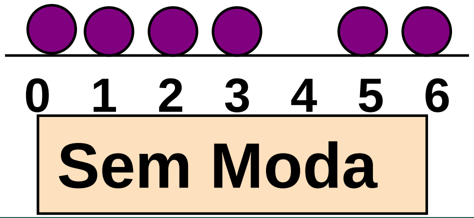
UNIFESP
Universidade Federal de São Paulo
1933

Moda: Valor que ocorre com mais frequência

- Não é afetada por valores extremos;
- Utilizada para dados numéricos ou categóricos;
- Pode ser que não exista moda;
- Pode ser que existam várias modas (multimodal).







UNIFESP
Universidade Federal de São Paulo
1933

Moda

- **Exemplo:** variável número de filhos;

Tabela 2.5: Freqüências e porcentagens dos empregados da seção de orçamentos da Companhia MB, segundo o número de filhos.

Nº de filhos z_i	Freqüência n_i	Porcentagem $100\,f_i$
0	4	20
1	5	25
2	7	35
3	3	15
5	1	5
Total	20	100

Mo=2, corresponde a realização com maior frequência.

- **Exemplo:** variável salários;

Tabela 2.6: Distribuição de freqüências da variável S, salário dos empregados da seção de orçamentos da Companhia MB.

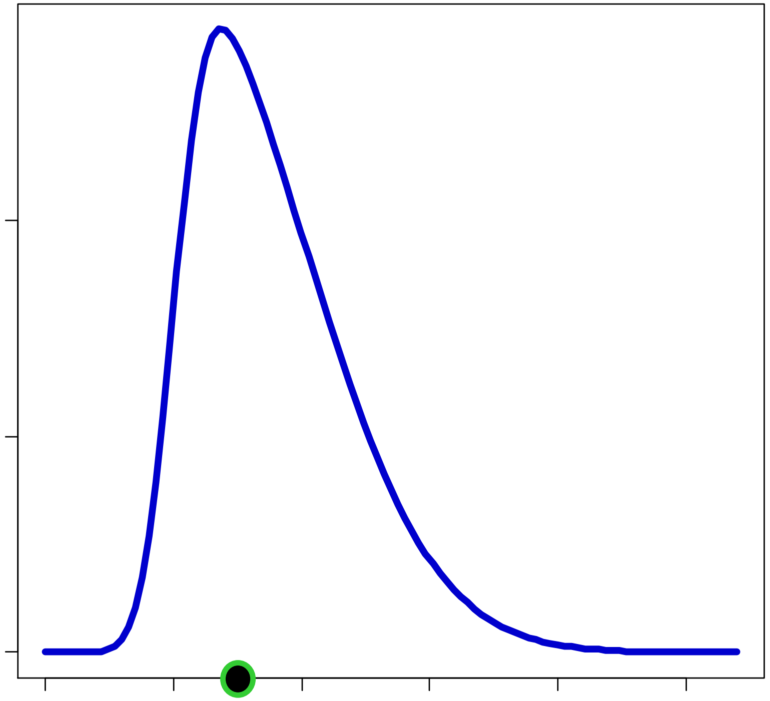
Classes de salários	Ponto médio s_i	Freqüência n_i	Porcentagem $100\,f_i$
4,00 ⊢ 8,00	6,00	10	27,78
8,00 ⊢ 12,00	10,00	12	33,33
12,00 ⊢ 16,00	14,00	8	22,22
16,00 ⊢ 20,00	18,00	5	13,89
20,00 ⊢ 24,00	22,00	1	2,78
Total	—	36	100,00

Mo=10, corresponde a realização com maior frequência..

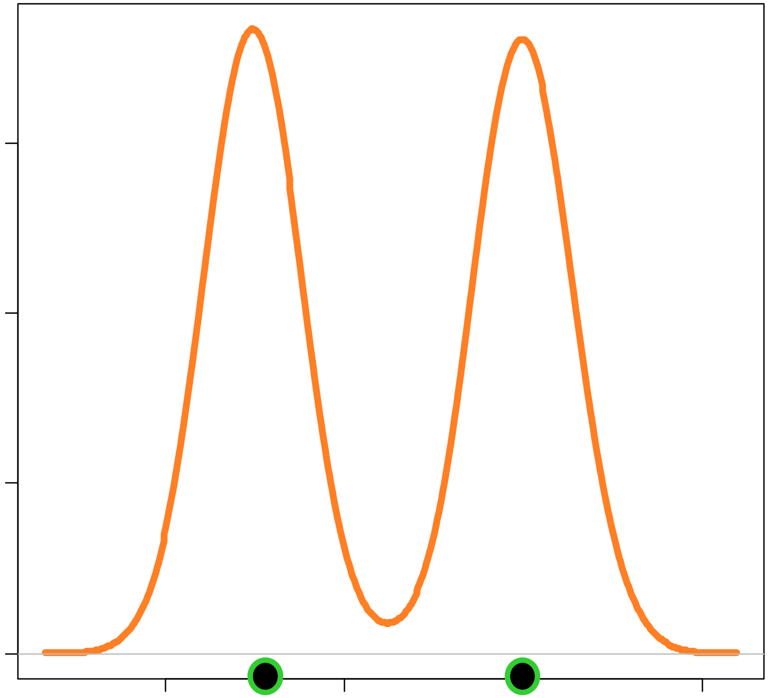
UNIFESP
Universidade Federal de São Paulo
1933

Moda

Unimodal



Bimodal



UNIFESP
Universidade Federal de São Paulo
1933

Estatísticas de Ordem

Considere agora a amostra $x_1, x_2, x_3, \ldots, x_n$ e suponha que ordene a amostra, de tal forma que $x_{(1)}$ é o menor elemento da amostra, $x_{(2)}$ é o segundo menor elemento, . . ., $x_{(n)}$ é o maior elemento da amostra. Os valores $x_{(1)}, x_{(2)}, x_{(3)}, \ldots, x_{(n)}$ são chamados de **estatísticas de ordem** da amostra. Outras medidas de tendência central e de dispersão serão definidas a partir das estatísticas de ordem.

Estatísticas de Ordem

Então,

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}$$

Exemplo: $x_1 = 1, x_2 = -2, x_3 = 5, x_4 = 1, x_5 = 2$

$$-2 \leq 1 \leq 1 \leq 2 \leq 5$$

$$x_{(1)} = -2, x_{(2)} = 1, x_{(3)} = 1, x_{(4)} = 2, x_{(5)} = 5$$

em que $x_{(1)} = -2$ representa o valor mínimo da amostra e $x_{(5)} = 5$ representa o valor máximo da amostra.

UNIFESP
Universidade Federal de São Paulo
1933

MEDIANA

Com base na estatística de ordem, a mediana de uma variável X pode ser definida como:

$$Md(X) = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & \text{se o } n \text{ é ímpar;} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se o } n \text{ é par;} \end{cases}$$

Exemplos: (i) $x_{(1)} = -2, x_{(2)} = 1, \boxed{x_{(3)} = 1}, x_{(4)} = 2, x_{(5)} = 5$

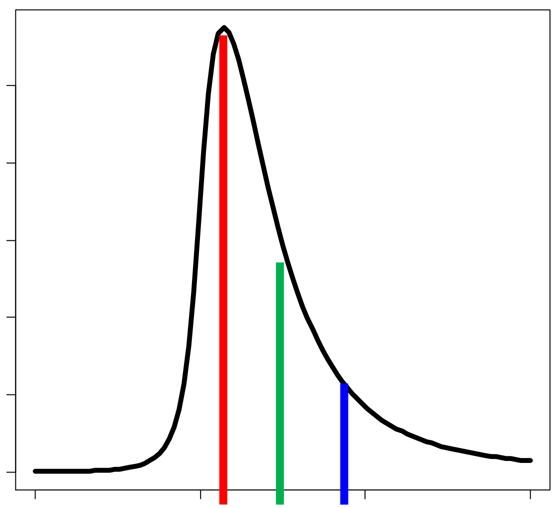
(ii) $x_{(1)} = -2, x_{(2)} = 1, \boxed{x_{(3)} = 1, x_{(4)} = 2}, x_{(5)} = 5, x_{(6)} = 7$

(i) $Md(X) = x_{(3)} = 1$ e (ii) $Md(X) = \dfrac{X_{(3)}+X_{(4)}}{2}$ =1,5

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

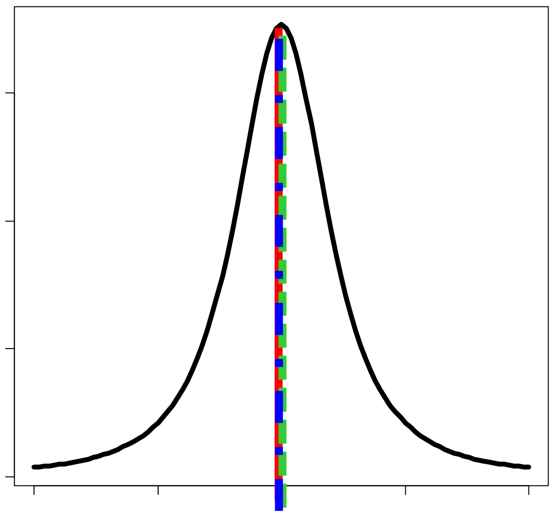
MODA, MÉDIA E MEDIANA

Assimetria à direita



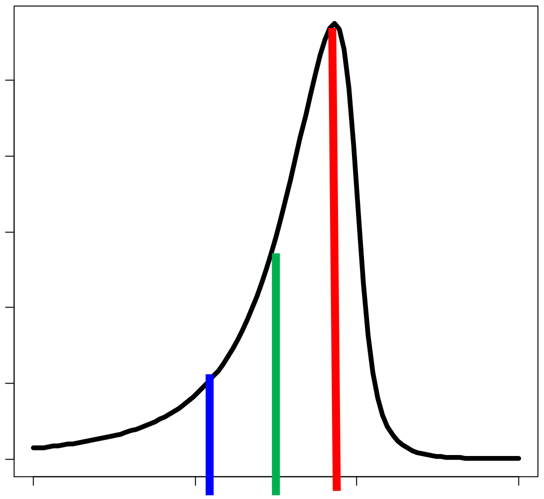
Mo < Md < $\overline{x}$

Simetria



Mo = Md = $\overline{x}$

Assimetria à esquerda



$\overline{x}$ < Md < Mo

Medidas Separatrizes

São números reais que dividem a sequência ordenada de dados em partes que contêm a mesma quantidade de elementos da série.

Desta forma, a mediana que divide a sequência ordenada em dois grupos, cada um deles contendo 50% dos valores da sequência, é também uma medida separatriz. Além da mediana, as outras medidas separatrizes que destacaremos são os percentis e os quartis.

PERCENTIL E QUARTIL

O ***percentil*** de ordem **k** (onde **k** é qualquer valor entre **0** e **100**), denotado por $\mathbf{P_k}$, é o valor tal que **K%** dos valores do conjunto de dados são menores ou iguais a ele. Assim, o percentil de ordem **10**, o $\mathbf{P_{10}}$, é o valor da variável tal que **10%** dos valores são menores ou iguais a ele; o percentil de ordem **65** deixa **65%** dos dados menores ou iguais a ele, etc.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

PERCENTIL E QUARTIL

O ***percentil*** de ordem **k** (onde **k** é qualquer valor entre **0** e **100**), denotado por $\mathbf{P_k}$, é o valor tal que **K%** dos valores do conjunto de dados são menores ou iguais a ele. Assim, o percentil de ordem **10**, o $\mathbf{P_{10}}$, é o valor da variável tal que **10%** dos valores são menores ou iguais a ele; o percentil de ordem **65** deixa **65%** dos dados menores ou iguais a ele, etc.

Os percentis de ordem **10**, **20**, **30**, . . ., **90** dividem o conjunto de dados em **dez partes** com mesmo número de observações e são chamados de ***decis***.

Percentil e Quartil

Os percentis de ordem **25**, **50** e **75** dividem o conjunto de dados em **quatro partes** com o mesmo número de observações. Assim, estes três percentis recebem o nome de ***quartis***.

- **primeiro quartil (Q1)**
- **segundo quartil (Q2)**
- **terceiro quartil (Q3)**

UNIFESP
Universidade Federal de São Paulo
1933

Percentil e Quartil

Os percentis de ordem **25**, **50** e **75** dividem o conjunto de dados em **quatro partes** com o mesmo número de observações. Assim, estes três percentis recebem o nome de ***quartis***.

- **primeiro quartil (Q1)**
- **segundo quartil (Q2)**
- **terceiro quartil (Q3)**

O segundo quartil é a já conhecida **mediana**.

UNIFESP
Universidade Federal de São Paulo
1933

PERCENTIL E QUARTIL

Existem vários processos para calcular os ***percentis***. Vamos focar com um método mais simples encontrado em Anderson *et al.*, (2002). As diferenças serão muito pequenas e desaparecerão à medida que aumenta o número de dados.

O seguinte procedimento pode ser usado para calcular o *p*-ésimo percentil:

Percentil e Quartil

1º) Arranje os dados na ordem ascendente (ordem de classificação do menor valor para o maior).

2º) Calcule um índice *i*:

$$i = \left(\frac{k}{100}\right)n$$

3º) Se *i* **não for um número inteiro**, arredonde para cima. O próximo inteiro maior que *i* denota a posição do *p*-ésimo percentil. Se *i* **for um inteiro**, o *p*-ésimo percentil é a média dos valores de dados nas posições *i* e *i* + 1.

UNIFESP
Universidade Federal de São Paulo
1933

QUANTIS

Alguns Quantis q(p) para 0<p<1:

$q(0{,}25) = 1º\ Quartil = 25º\ Percentil$

$q(0, 50) = 2\ º\ Quartil = \ Mediana(Md) = 5º\ Decil = 50º\ Percentil$

$q(0{,}75) = 3º\ Quartil = 75º\ Percentil$

$q(0{,}40) = 4º\ Decil$

$q(0{,}95) = 95º\ Percentil$

Exemplo: $x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 8, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

P_{20}=?

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Quantis

Alguns Quantis q(p) para 0<p<1:

$q(0{,}25) = 1º\ Quartil = 25º\ Percentil$

$q(0{,}50) = 2\ º\ Quartil = \ Mediana(Md) = 5º\ Decil = 50º\ Percentil$

$q(0{,}75) = 3º\ Quartil = 75º\ Percentil$

$q(0{,}40) = 4º\ Decil$

$q(0{,}95) = 95º\ Percentil$

Exemplo: $x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 8, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

P_{20}=? Então, $i = \left(\frac{k}{100}\right) n = \left(\frac{20}{100}\right) \text{x}\ 9 = 1{,}8.$

UNIFESP
Universidade Federal de São Paulo
1933

Quantis

Alguns Quantis q(p) para 0<p<1:

$q(0{,}25) = 1º\ Quartil = 25º\ Percentil$

$q(0,50) = 2\ º\ Quartil = \ Mediana(Md) = 5º\ Decil = 50º\ Percentil$

$q(0{,}75) = 3º\ Quartil = 75º\ Percentil$

$q(0{,}40) = 4º\ Decil$

$q(0{,}95) = 95º\ Percentil$

Exemplo: $x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 8, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

$\mathbf{P_{20}}$=? Então, $i = \left(\frac{k}{100}\right) n = \left(\frac{20}{100}\right) x\ 9 = 1{,}8$. Assim arrendando para **2**

UNIFESP

UNIVERSIDADE FEDERAL DE SÃO PAULO

1933

QUANTIS

Alguns Quantis q(p) para 0<p<1:

$q(0{,}25) = 1º\ Quartil = 25º\ Percentil$

$q(0,50) = 2\ º\ Quartil = \ Mediana(Md) = 5º\ Decil = 50º\ Percentil$

$q(0{,}75) = 3º\ Quartil = 75º\ Percentil$

$q(0{,}40) = 4º\ Decil$

$q(0{,}95) = 95º\ Percentil$

Exemplo: $x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 8, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

Então, i $= \left(\frac{k}{100}\right) n = \left(\frac{20}{100}\right)$ x 9 = 1,8. Assim arrendando para **2**. $\mathbf{P_{20}}$= **3**

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Observações

Não existe uma medida de centro que seja melhor que a outra. O problema em estudo é que poderá definir qual a mais adequada. Em muitas situações, elas podem ser complementares.

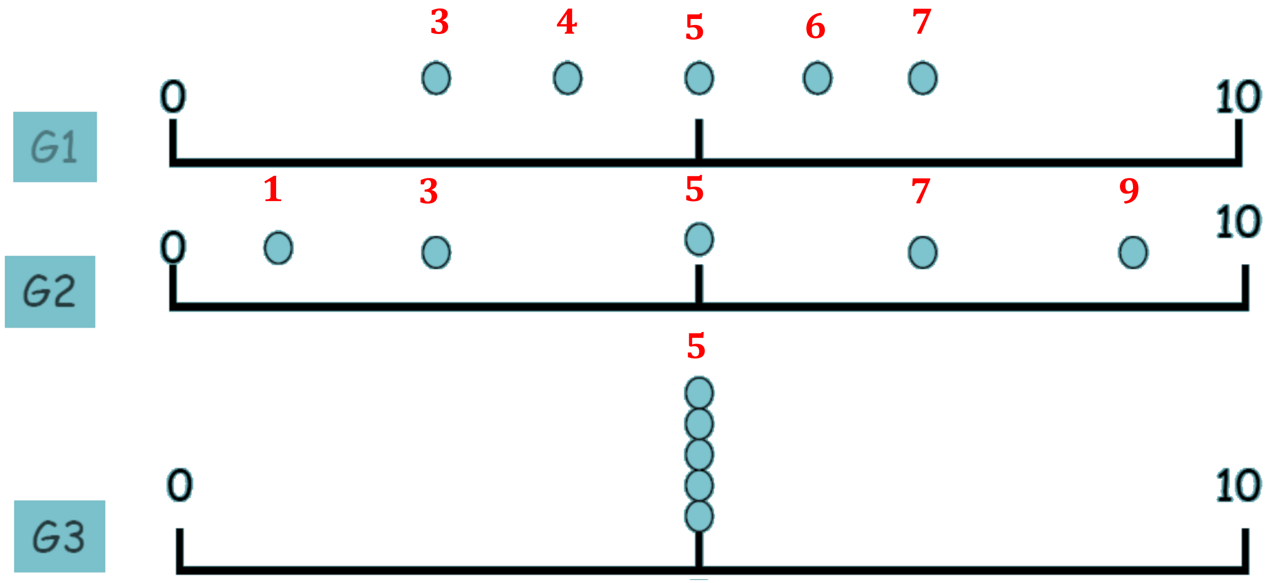
- A **Média** é a mais utilizada, mas pode não ser adequada em casos nos quais existem valores extremos.
- A **Mediana** pode ser mais adequada quando existem valores extremos.

Note que: localizar valores extremos (discrepantes) é importante para verificar se estes são erros de medida ou de digitação nos dados. Neste caso podemos eliminá-los.
Caso contrário, devemos mantê-los como parte dos dados serem estudados. Neste caso, pode ser útil medidas menos sensíveis a eles, tal como a mediana.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Exemplo

Considere as notas de um teste aplicado em três grupos de alunos. **G1:** 3,4,5,6,7, **G2:** 1,3,5,7,9 e **G3:** 5,5,5,5,5.



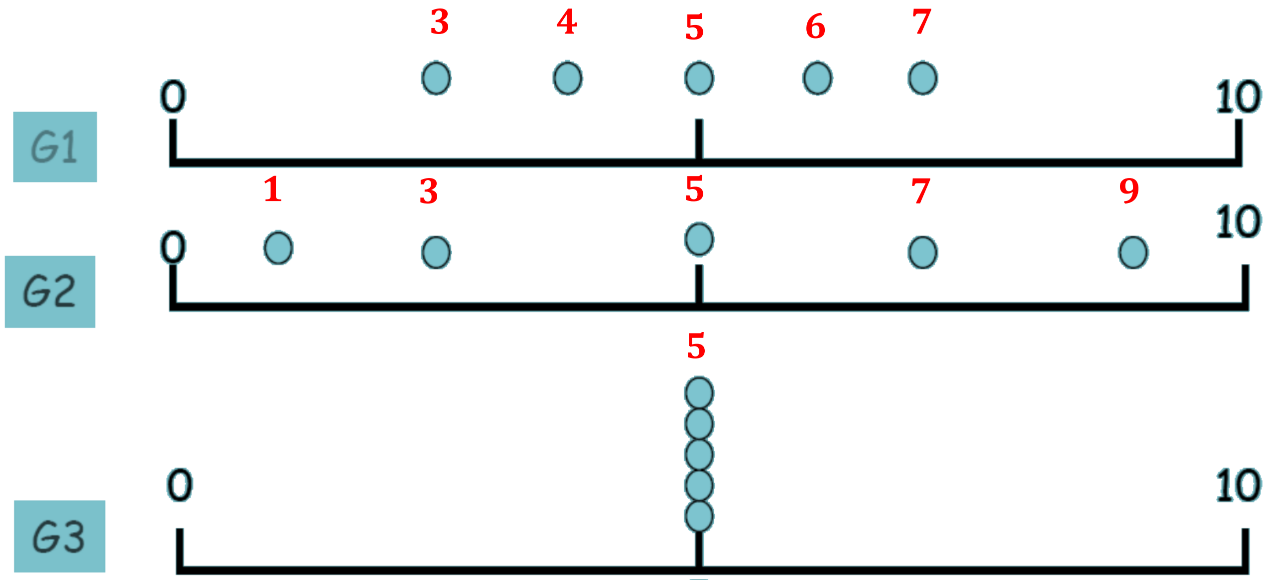
$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$$

$$Md_1 = Md_2 = Md_3 = 5$$

UNIFESP
Universidade Federal de São Paulo
1933

Exemplo

Considere as notas de um teste aplicado em três grupos de alunos. **G1:** 3,4,5,6,7, **G2:** 1,3,5,7,9 e **G3:** 5,5,5,5,5.



Já que as medidas centrais são a mesmas, como diferenciar os bancos? Talvez olhando para a dispersão (variação) dos dados!

$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$$

$$Md_1 = Md_2 = Md_3 = 5$$

UNIFESP
Universidade Federal de São Paulo
1933

Medidas de Posição

- **Moda (Mo):** valor ou atributo que ocorre com maior frequência;
- **Média ($\overline{x}$):** soma de todos os valores da variável em estudo dividido pelo número de observações;
- **Mediana (Md):** valor que deixa 50% das observações à esquerda;
- **Mínimo (mín):** menor observação;
- **Máximo (max):** maior observação;
- **Quantil:** quartil, percentil e decil.

UNIFESP
Universidade Federal de São Paulo
1933

MEDIDAS DE DISPERSÃO

- **Dispersão ou Variação**
  - Amplitude
  - Amplitude Interquartil
  - Variância
  - Desvio absoluto
    Desvio Padrão
  - Coeficiente de Variação

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Medidas de Dispersão

- **Amplitude (A):** diferença entre o valor máximo e o valor mínimo;
- **Variância (S^2):** média dos quadrados dos desvios em relação à média aritmética;
- **Desvio padrão (S):** mede a variabilidade independente do número de observações e com a mesma unidade de medida da média;
- **Coeficiente de variação (CV):** é uma medida de dispersão relativa; exprime a variabilidade em relação a média; útil para comparar duas ou mais variáveis (grupos);
- **Distância interquartil (d_q):** uma medida alternativa ao desvio padrão.

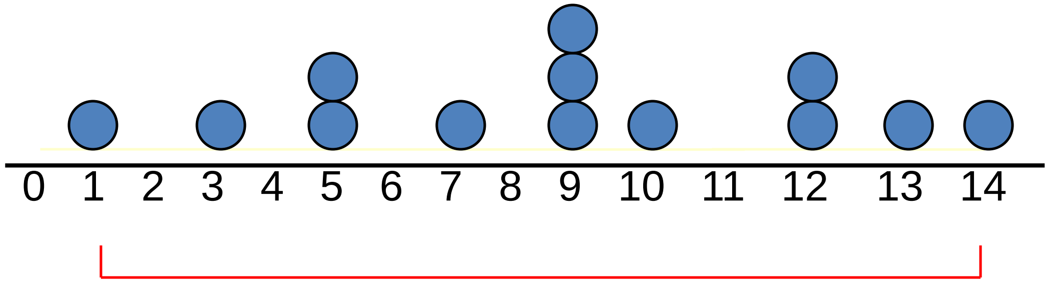
UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Amplitude

- A = máximo – mínimo ou $A = x_{(n)} - x_{(1)}$
- Medida de variação mais simples
- Diferença entre o maior e o menor valor de um conjunto de dados:

$$\boxed{\text{Amplitude} = X_{maior} - X_{menor}}$$

Exemplo:

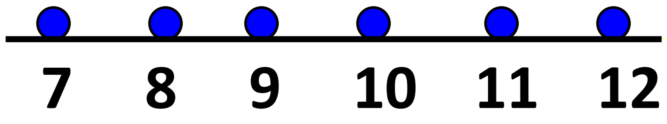


Amplitude = 14 - 1 = 13

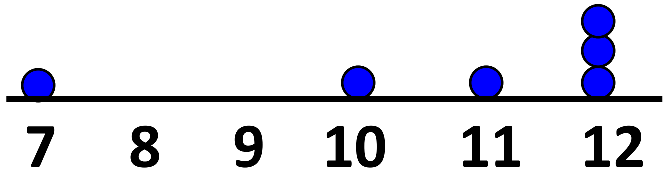
UNIFESP
Universidade Federal de São Paulo
1933

Desvantagens da Amplitude

Ignora a forma como os dados são distribuídos.



Amplitude = 12 - 7 = 5



Amplitude = 12 - 7 = 5

Sensivel a valores extremos.

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**5**

Amplitude = 5 - 1 = 4

1,1,1,1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3,4,**120**

Amplitude = 120 - 1 = 119

UNIFESP
Universidade Federal de São Paulo
1933

Desvio médio absoluto (DMA)

Usa a soma dos módulos dos desvios.

$$DM = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

Entretanto, o valor absoluto dos desvios é mais difícil de tratamento matemático (no uso de derivadas, por exemplo)

UNIFESP
Universidade Federal de São Paulo
1933

DESVIO MÉDIO ABSOLUTO (DMA)

Exemplo: Amostra (X_i) : 10 12 14 15 17 18 18 24

$n = 8 \qquad \bar{X} = 16$

$$DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n} = \frac{|10-16| + |12-16| + |14-16| + \cdots + |24-16|}{8}$$

$$DM = \frac{26}{8} = 3{,}25$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Medidas de Dispersão

	Amostral	Populacional
Variância:	$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$	$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$
Desvio Padrão:	$S = \sqrt{S^2}$	$\sigma = \sqrt{\sigma^2}$

UNIFESP
Universidade Federal de São Paulo
1933

VARIÂNCIA

Observação: usa-se a soma de cada desvio ao quandrado

A variância

- Mede a dispersão em torno da media;
- Variância de uma amostra com n elementos,

onde: $\overline{X}$= média

n = tamanho da amostra

X_i = i-ésimo valor de X

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

VARIÂNCIA

A fórmula para a variância pode ser reescrita como:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n\text{-}1} = \boxed{\frac{n\sum_{i=1}^{n}X_i^{\,2} - \left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1)}}$$

Utiliza apenas os valores dos dados

Note que se a variável x tem uma unidade de medida, por exemplo quilograma, metro, reais,..., a variância tem o inconveniente de ter a unidade da variável ao quadrado. Então usamos o desvio padrão tomando a raiz quadrada da variância.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

DESVIO-PADRÃO AMOSTRAL

- Medida de variação mais utilizada;
- Mostra a variação em torno da média;
- Raiz quadrada da variância;
- Tem a **mesma unidade dos dados originais.**

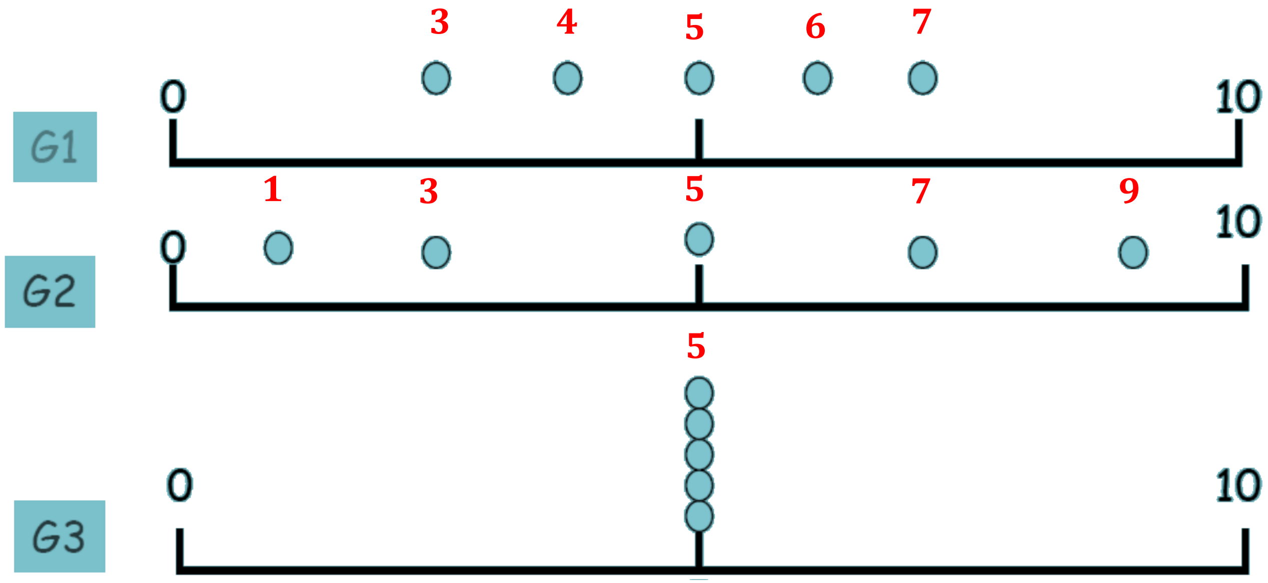
Desvio-Padrão da amostra:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

EXEMPLO

Considere as notas de um teste aplicado em três grupos de alunos. **G1:** 3,4,5,6,7, **G2:** 1,3,5,7,9 e **G3:** 5,5,5,5,5

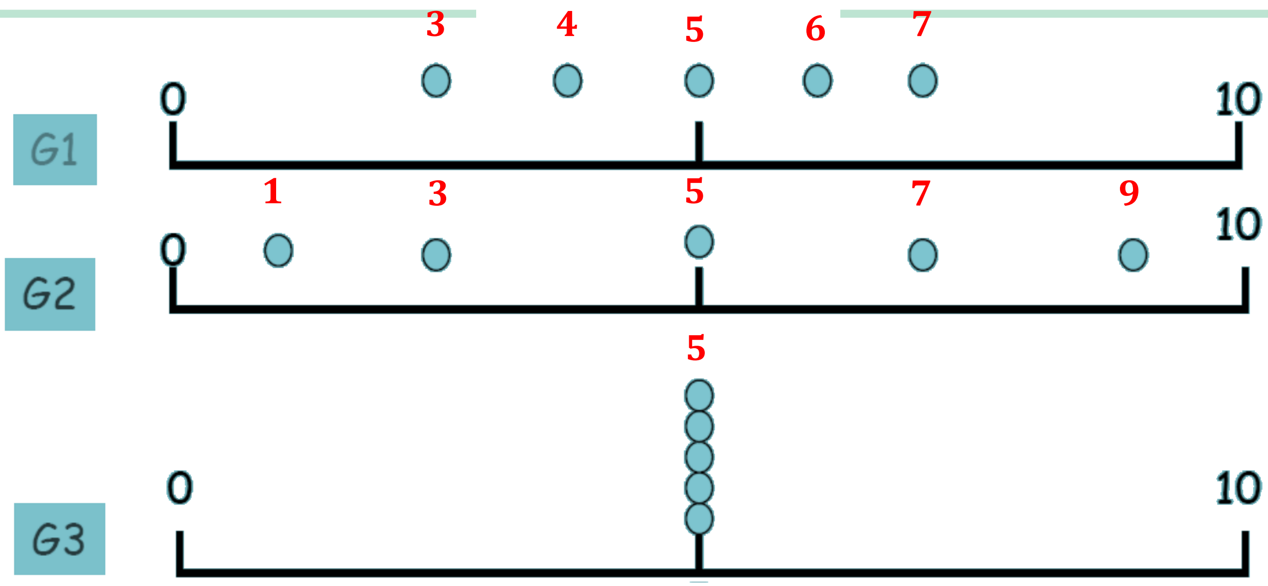


$$\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = 5$$

$$Md_1 = Md_2 = Md_3 = 5$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Exemplo



$$\boxed{A_1 = \mathbf{4},\ A_2 = \mathbf{8} \text{ e } A_3 = \mathbf{0}}$$

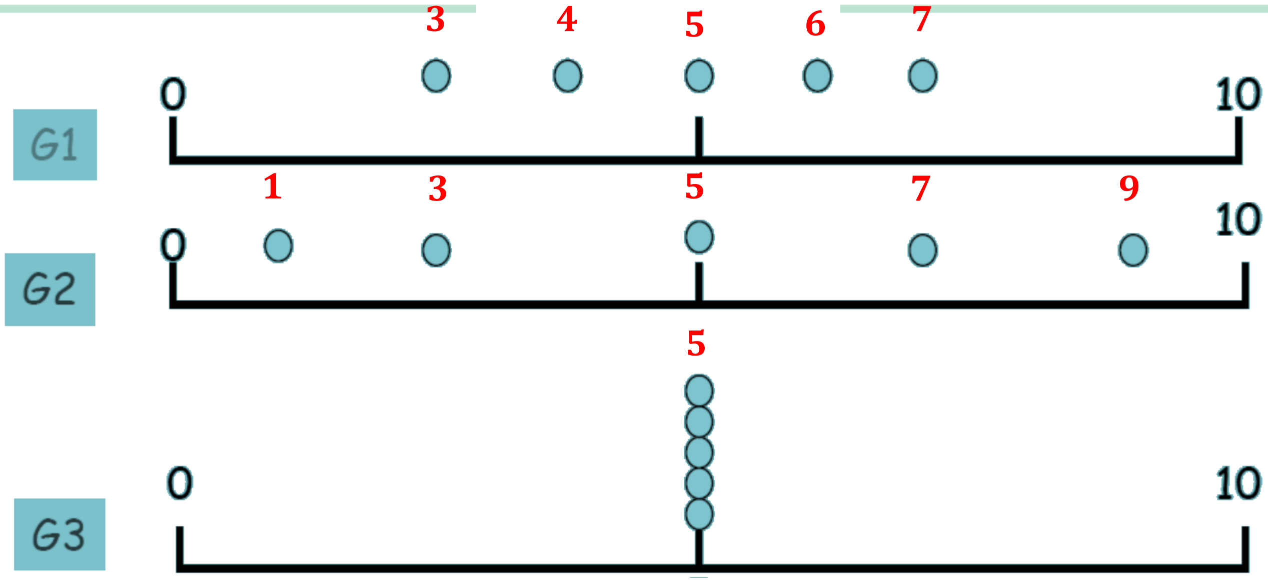
$$S^2(G1) = \frac{(3-5)^2+(4-5)^2+(5-5)^2+(6-5)^2+(7-5)^2}{5-1} = 2{,}5$$

$$S^2(G1) = \mathbf{2,5};\ S^2(G2) = \mathbf{10} \text{ e } S^2(G3) = \mathbf{0}$$

$$S(G1) = \mathbf{1,58};\ S(G2) = \mathbf{3,16} \text{ e } S(G3) = \mathbf{0}$$

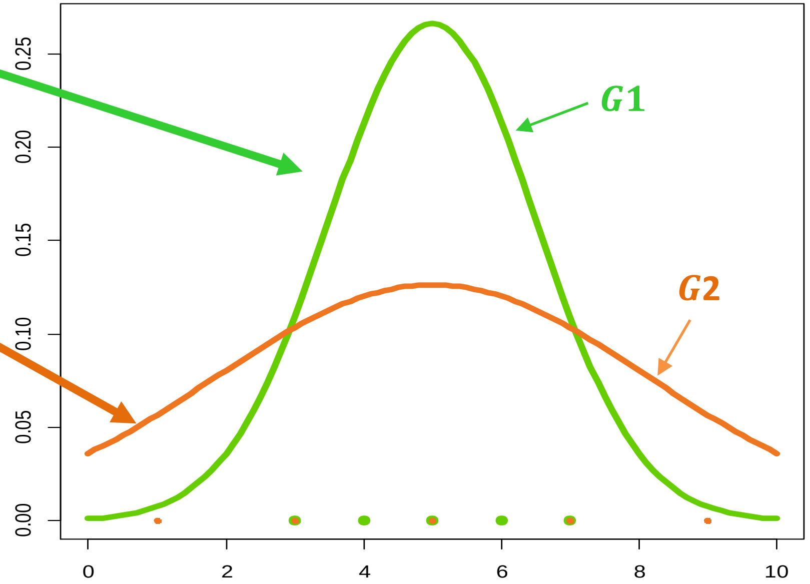
UNIFESP
Universidade Federal de São Paulo
1933

Exemplo



Desvio-padrão pequeno

Desvio-padrão grande



UNIFESP
Universidade Federal de São Paulo
1933

Para complementar

Variância para dados agrupados em classes. Mede a variação em relação a média dos pontos médios.

$$s^2 = \frac{\sum_{i=1}^{k} (PM_i - \overline{\overline{x}})^2 f_i}{n-1}$$

k : número de classes

PM_i : ponto médio da classe i

f_i : frequência da classe i

$$\overline{\overline{x}} = \frac{\sum_{i=1}^{k} f_i PM_i}{n} \quad \text{(média dos pontos médios)}$$

Desvio padrão: $s = \sqrt{s^2}$

UNIFESP

Universidade Federal de São Paulo

1933

COEFICIENTE DE VARIAÇÃO

Coeficiente de Variação: $CV = \frac{s}{\bar{x}} 100\%$

- Os valores do CV dependem do tipo de pesquisa e da variável em estudo para ser considerado aceitável.
- Na estatística experimental, ele indica a precisão do experimento, ou seja, a capacidade de o realizarmos novamente, sob as mesmas condições, e produzir resultados semelhantes.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

COEFICIENTE DE VARIAÇÃO

Coeficiente de Variação: $CV = \frac{s}{\bar{x}} 100\%$

- Os valores do CV dependem do tipo de pesquisa e da variável em estudo para ser considerado aceitável.
- Na estatística experimental, ele indica a precisão do experimento, ou seja, a capacidade de o realizarmos novamente, sob as mesmas condições, e produzir resultados semelhantes.

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

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Coeficiente de variação

Exemplo: Em um grupo de pacientes foram tomadas as pulsações (batimentos por minuto) e dosadas as taxas de ácido úrico (mg/100ml). As médias e os desvios padrão foram

Variável	$\bar{x}$	s
pulsação	68,7	8,7
ácido úrico	5,46	1,03

$$\boxed{CV_p = \frac{8{,}7}{68{,}7} 100\% \cong 12{,}7\%}$$

$$\boxed{CV_{au} = \frac{1{,}03}{5{,}46} 100\% \cong 19\%}$$

Conclusão: Com relação as médias, os pacientes são, aproximadamente, duas vezes mais dispersos quanto ao ácido úrico do que quanto a pulsação. Pulsação é mais estável.

UNIFESP
Universidade Federal de São Paulo
1933

Distância Interquartil

Distância Interquartil: $d_q = q(0{,}75) - q(0{,}25)$

Exemplo: $q(0{,}25)?\ q(0{,}75)?$

$$x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, x_{(5)} = 8, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

UNIFESP
Universidade Federal de São Paulo
1933

Distância Interquartil

Distância Interquartil: $d_q = q(0{,}75) - q(0{,}25)$

Exemplo: $q(0{,}25)$? $q(0{,}75)$?

$$x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7, \boxed{x_{(5)} = 8}, x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15$$

$$Md = 8$$

UNIFESP
Universidade Federal de São Paulo
1933

DISTÂNCIA INTERQUARTIL

Distância Interquartil: $d_q = q(0{,}75) - q(0{,}25)$

Exemplo: $q(0{,}25)?\ q(0{,}75)?$

$$x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$$

$$2 \leq 3 \leq 5 \leq 7 \leq 8 \leq 10 \leq 11 \leq 12 \leq 15$$

$$\underbrace{x_{(1)} = 2, x_{(2)} = 3, x_{(3)} = 5, x_{(4)} = 7}_{1^{\text{o}}\ Parte}, \boxed{x_{(5)} = 8}, \underbrace{x_{(6)} = 10, x_{(7)} = 11, x_{(8)} = 12, x_{(9)} = 15}_{2^{\text{o}}\ Parte}$$

1^{o} *Parte*

$$Md = \frac{X_{(2)} + X_{(3)}}{2} = 4$$

$$Md = 8$$

2^{o} *Parte*

$$Md = \frac{X_{(7)} + X_{(8)}}{2} = 11{,}5$$

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Distância Interquartil

Distância Interquartil: $d_q = q(0{,}75) - q(0{,}25)$

Exemplo: $q(0{,}25)?\ q(0{,}75)?$

$$x_1 = 15, x_2 = 5, x_3 = 3, x_4 = 8, x_5 = 10, x_6 = 2, x_7 = 7, x_8 = 11, x_9 = 12$$

$$2 \le 3 \le 5 \le 7 \le 8 \le 10 \le 11 \le 12 \le 15$$

$$x_{(1)} = 2, \boxed{x_{(2)} = 3, x_{(3)} = 5}, x_{(4)} = 7, \boxed{x_{(5)} = 8}, x_{(6)} = 10, \boxed{x_{(7)} = 11, x_{(8)} = 12}, x_{(9)} = 15$$

$1^{\underline{o}}$ *Parte* | $2^{\underline{o}}$ *Parte* | $3^{\underline{o}}$ *Parte* | $4^{\underline{o}}$ *Parte*

$q(0{,}25) = 4$

$Md = 8$

$q(0{,}75) = 11{,}5$

$$\boxed{d_q = 11{,}5 - 4 = \mathbf{7{,}5}}$$

UNIFESP

Universidade Federal de São Paulo

1933

RESUMO DA AULA

- Medidas de tendência central ou de posição;
- Quantis;
- Medidas de dispersão.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

PRÓXIMA AULA

- Construção do gráfico Box Plot;
- Discusão sobre outliers;
- Discusão sobre as formas da distribuição dos dados;
- Medidas de associação entre variáveis.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Referências

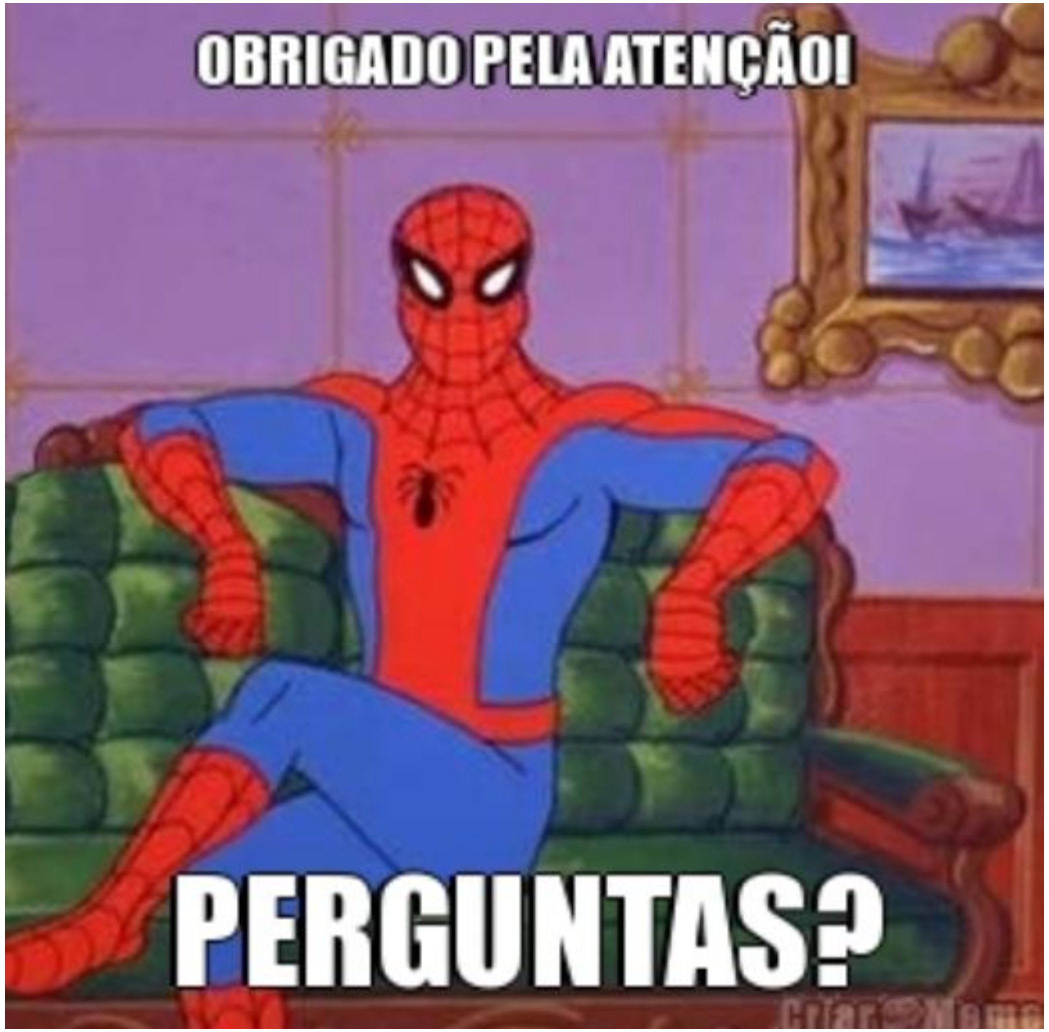
MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 6 ed. Rio de Janeiro: LTC, 2018. 628p.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 9 ed. São Paulo: Saraiva, 2017. 554p.

MOORE, D. S. NOTZ, W. I.; FLIGNER, M. A. A estatística básica e sua prática. 7 ed. Rio de Janeiro: LTC, 2017. 628p.

UNIFESP
UNIVERSIDADE FEDERAL DE SÃO PAULO
1933

Class finished



UNIFESP
Universidade Federal de São Paulo
1933