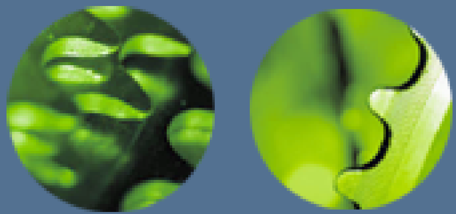
Medidas de Variação ou Dispersão

Estatística descritiva

- Recapitulando: As três principais características de um conjunto de dados são:
  - Um valor representativo do conjunto de dados: uma média (Medidas de Tendência Central)
  - Uma medida de dispersão ou variação.
  - A natureza ou forma da distribuição dos dados: sino, uniforme, assimétrica,... (Tabelas de frequência e histogramas)

Medidas de Variação

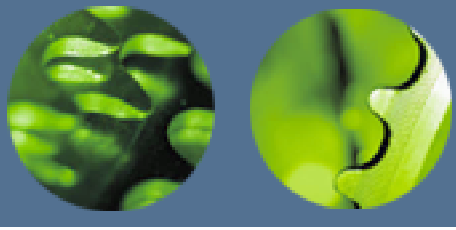
- Determina a característica de variação de um conjunto de dados
  - Amplitude
  - Desvio
  - Desvio médio ou desvio absoluto
  - Desvio padrão
  - Variância



Amplitude

- Diferença entre o maior e o menor valor
  - Subtraia o menor valor do maior
  - Amplitude = 1,88 – 1,60 = 0,28 m

Análise Estatística da Turma de Prob. e	
Eventos	x
Aluno 1	1,72
Aluno 2	1,60
Aluno 3	1,74
Aluno 4	1,88
Aluno 5	1,82
Aluno 6	1,75
Aluno 7	1,82
Aluno 8	1,75
Aluno 9	1,73
Aluno 10	1,75
Aluno 11	1,80
Aluno 12	1,75
Aluno 13	1,73
Aluno 14	1,84
Aluno 15	1,76
Aluno 16	1,78
Aluno 17	1,75
Aluno 18	1,69
Soma	31,66
Média	1,759
Amplitude	0,28



Desvio e desvio absoluto

- Desvio
  - diferença entre cada valor e a média

$$x - \bar{x}$$

- Desvio médio ou absoluto
  - Média dos desvios em termos absolutos

$$\frac{\sum |x - \bar{x}|}{n}$$

Análise Estatística da Turma de Prob. e Estatística			
Eventos	x	$x-\bar{x}$	$\lvert x-\bar{x}\rvert$
Aluno 1	1,72	-0,04	0,04
Aluno 2	1,60	-0,16	0,16
Aluno 3	1,74	-0,02	0,02
Aluno 4	1,88	0,12	0,12
Aluno 5	1,82	0,06	0,06
Aluno 6	1,75	-0,01	0,01
Aluno 7	1,82	0,06	0,06
Aluno 8	1,75	-0,01	0,01
Aluno 9	1,73	-0,03	0,03
Aluno 10	1,75	-0,01	0,01
Aluno 11	1,80	0,04	0,04
Aluno 12	1,75	-0,01	0,01
Aluno 13	1,73	-0,03	0,03
Aluno 14	1,84	0,08	0,08
Aluno 15	1,76	0,00	0,00
Aluno 16	1,78	0,02	0,02
Aluno 17	1,75	-0,01	0,01
Aluno 18	1,69	-0,07	0,07
	Média	Soma desvios	Desvio médio
	1,759	0,000	0,043

Desvio Padrão

- Desvio padrão: medida da variação dos valores em relação à média.
- Ex.: Calcular o desvio padrão do conjunto de dados ao lado.
  - Passo 1: Calcule a média;
  - Passo 2: Calcule o DESVIO de cada medida sobre a média

$$\text{Desvio} = \boxed{x - \bar{x}}$$

Análise Estatística da Turma de Prob. e Estatística		
Eventos	x	$x-\bar{x}$
Aluno 1	1,72	-0,04
Aluno 2	1,60	-0,16
Aluno 3	1,74	-0,02
Aluno 4	1,88	0,12
Aluno 5	1,82	0,06
Aluno 6	1,75	-0,01
Aluno 7	1,82	0,06
Aluno 8	1,75	-0,01
Aluno 9	1,73	-0,03
Aluno 10	1,75	-0,01
Aluno 11	1,80	0,04
Aluno 12	1,75	-0,01
Aluno 13	1,73	-0,03
Aluno 14	1,84	0,08
Aluno 15	1,76	0,00
Aluno 16	1,78	0,02
Aluno 17	1,75	-0,01
Aluno 18	1,69	-0,07
Soma	31,66	0,00
Média	1,759	-------

Desvio Padrão

- Calcule o desvio padrão do conjunto de dados ao lado.
  - Passo 3: Eleve ao quadrado cada uma das diferenças;
  - Passo 4: Some todos os quadrados obtidos

$$\boxed{\sum (x-\bar{x})^2}$$

Análise Estatística da Turma de Prob. e Estatística			
Eventos	x	$x-\bar{x}$	$(x-\bar{x})^2$
Aluno 1	1,72	-0,04	0,0015
Aluno 2	1,60	-0,16	0,0252
Aluno 3	1,74	-0,02	0,0004
Aluno 4	1,88	0,12	0,0147
Aluno 5	1,82	0,06	0,0037
Aluno 6	1,75	-0,01	0,0001
Aluno 7	1,82	0,06	0,0037
Aluno 8	1,75	-0,01	0,0001
Aluno 9	1,73	-0,03	0,0008
Aluno 10	1,75	-0,01	0,0001
Aluno 11	1,80	0,04	0,0017
Aluno 12	1,75	-0,01	0,0001
Aluno 13	1,73	-0,03	0,0008
Aluno 14	1,84	0,08	0,0066
Aluno 15	1,76	0,00	0,0000
Aluno 16	1,78	0,02	0,0004
Aluno 17	1,75	-0,01	0,0001
Aluno 18	1,69	-0,07	0,0047
Soma	31,66	0,00	0,065


Desvio Padrão

- Passo 5: Divida o total por (n-1), onde n é o número de dados coletados (amostra);
- Passo 6: Extraia a raiz quadrada do resultado anterior

$$\sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$$

Desvio Padrão

Análise Estatística da Turma de Prob. e Estatística			
Eventos	x	$x-\bar{x}$	$(x-\bar{x})^2$
Aluno 1	1,72	-0,04	0,0015
Aluno 2	1,60	-0,16	0,0252
Aluno 3	1,74	-0,02	0,0004
Aluno 4	1,88	0,12	0,0147
Aluno 5	1,82	0,06	0,0037
Aluno 6	1,75	-0,01	0,0001
Aluno 7	1,82	0,06	0,0037
Aluno 8	1,75	-0,01	0,0001
Aluno 9	1,73	-0,03	0,0008
Aluno 10	1,75	-0,01	0,0001
Aluno 11	1,80	0,04	0,0017
Aluno 12	1,75	-0,01	0,0001
Aluno 13	1,73	-0,03	0,0008
Aluno 14	1,84	0,08	0,0066
Aluno 15	1,76	0,00	0,0000
Aluno 16	1,78	0,02	0,0004
Aluno 17	1,75	-0,01	0,0001
Aluno 18	1,69	-0,07	0,0047
Soma	31,66	0,00	0,065
Média	1,759	-------	-------

$$\sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = 0{,}062$$

Desvio Padrão

- De uma amostra

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

- De uma população

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

Observação:

A unidade do desvio padrão é a mesma unidade dos valores originais, ou conjunto de dados.


Fórmula abreviada para o desvio padrão

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

Vantagens e desvantagens:

- Mais conveniente para uso com números extensos e com grandes conjuntos de valores
- Maior facilidade de uso com calculadoras e computadores (apenas três registros: n, Σx e Σx^2)
- Elimina erros de arredondamento
- Não evidencia o conceito de desvio médio da fórmula tradicional

Variância

- Desvio padrão ao quadrado
  - $s^2 \rightarrow$ variância amostral
  - $\sigma^2 \rightarrow$ variância populacional

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

$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{N}$$

Observação:

A unidade da variância é a mesma unidade do conjunto de dados, elevada ao quadrado.

Considerações finais

- Arredondamento:
  - Tomar uma casa decimal a mais em relação às que constam dos dados originais.
  - Arredondar apenas o resultado final e não os resultados intermediários.
  - Se necessitarmos arredondar os resultados intermediários, acrescente duas casas decimal a mais em relação às que constam dos dados originais

Para que serve o desvio padrão?

- Indica a dispersão dos dados; quanto mais dispersos, maior o desvio padrão
- Regra prática
  - Desvio padrão $\cong$ amplitude/4 *(só usar em casos muito extremos)
  - Portanto:
    - valor mínimo $\cong$ média – 2.(s)
    - Valor máximo $\cong$ média + 2.(s)
- Teorema de Tchebichev
  - A proporção de qualquer conjunto de dados a menos de K desvios-padrão a contar da média é sempre ao menos $1-1/k^2$, onde k é um número positivo maior do que 1. Para k=2 e k=3, temos:
    - Ao menos ¾ (75%) de todos os valores estão no intervalo de $\pm$ 2 desvios-padrão em torno da média
    - Ao menos 8/9 (89%) de todos os valores estão no intervalo de $\pm$ 3 desvios-padrão em torno da média

Medidas de dispersão

- O *Coeficiente de Variação* indica a magnitude relativa do desvio-padrão quando comparado com a média do conjunto de valores

$$CV = \frac{s}{\bar{x}} \quad \textbf{(amostra)} \qquad\qquad CV = \frac{\sigma}{\mu} \quad \textbf{(população)}$$

- O *Coeficiente de Variação* é útil para compararmos a variabilidade (dispersão) de dois conjuntos de dados de ordem de grandezas diferentes

Medidas de dispersão

- Seja o seguinte conjunto de preços de geladeiras em 7 lojas distintas

750,00	800,00	790,00	810,00	820,00	760,00	780,00

$$\bar{x} = 787{,}14 \qquad s = 25{,}63$$

- Seja o seguinte conjunto de preços de liquidificadores nas mesmas lojas acima

50,00	45,00	55,00	43,00	52,00	45,00	54,00

$$\bar{x} = 49{,}14 \qquad s = 4{,}81$$

- *Qual dos produtos têm uma maior variabilidade de preços?*


Medidas de dispersão

- Uma vez que, em geral, uma geladeira custa bem mais que um liquidificador, a tendência é que o desvio-padrão da geladeira seja também maior!

- O coeficiente de variação é uma medida adimensional que normaliza o desvio padrão em relação à média

$$CV_{geladeira} = \frac{25{,}63}{787{,}14} = 3{,}3\%$$

$$CV_{liquidificador} = \frac{4{,}81}{49{,}14} = 9{,}8\%$$

- *Com o CV podemos concluir que os preços da geladeira têm uma menor variabilidade que os do liquidificador*

Medida de Dispersão: Intervalo interquartil (amplitude interquartílica)

- Uma medida de dispersão alternativa que pode ser empregada é o chamado *intervalo interquartil ou amplitude interquartílica*
- É a diferença entre o terceiro e o primeiro quartis
- Só aproveita 50% dos dados
  - Pouco influenciada pelos valores extremos

$$D_j = Q_3 - Q_1 = P_{0,75} - P_{0,25}$$

Medidas de posição e dispersão

- Para o conjunto de valores abaixo:

05; 07; 08; 10; 12; 15; 18; 20; 28; 35; 40; 44

Q1 = 10 Q2 = Md = 16,5 Q3 = 28 Q4 = 44

Dj = 28 - 10 = 18

- Se alterarmos significativamente o último valor:

05; 07; 08; 10; 12; 15; 18; 20; 28; 35; 40; 200

Dj = 28 - 10 = 18 !!!

Teorema de Tchebichev

- A fração (porcentagem) de QUALQUER conjunto de dados, a menos de K desvios a contar da média, é SEMPRE *ao menos:*

$$1 - 1/K^2 \qquad \text{onde } K>1$$

- Para k = 2 e k = 3 isto significa, por exemplo:

$$[\bar{x}-2s, \bar{x}+2s] \rightarrow \text{75\% dos dados}$$

Ou seja, ao menos ¾ de todos os valores estão neste intervalo

$$[\bar{x}-3s, \bar{x}+3s] \rightarrow \text{89\% dos dados}$$

Teorema de Tchebichev

- Barbeadores elétricos sem fio da marca XYZ têm *vida média* de 8,0 anos, com desvio padrão de 3,0 anos.

- Faça uma estimativa:

  - da vida mais breve =>

  - da vida mais longa =>

- *Tchebichev também é útil para identificar valores "estranhos" em um conjunto de dados:* aqueles que ficam de fora do intervalo !

Identificando *“outliers”*

- “*Outliers*” são valores “*estranhos*” que se localizam muito distantes da média
- Por isso, as estatísticas descritivas são, usualmente, muito influenciadas (“*contaminadas*”) por eles
- Podem se originar em erros de coleta OU em desvios de processo
- Esses outliers devem ser muito bem analisados antes de um possível descarte!

Identificando *“outliers”*

- Tchebichev pode nos ajudar na identificação de *outliers*

- Valores fora do intervalo de +/- 2s devem ser analisados para um possível descarte

$[\bar{x}-2s, \bar{x}+2s]$ → fora deste intervalo, é *estranho*

Escore Padronizado

$$z = \frac{x - \bar{x}}{s}$$

$$z = \frac{x - \mu}{\sigma}$$

- Número de desvios-padrão pelo qual um valor dista da média (para mais ou para menos)

Exercício

- As alturas da população de homens adultos têm média μ=1,752m, desvio padrão σ=0,071m e distribuição gráfica em forma de sino (normal). O jogador de basquete Michael Jordan, que mede 1,98m, pode ser considerado excepcionalmente alto? Determine o escore padrão *z* para ele.

Resolução

- Calcula-se o escore *z* conforme segue:

$$z = \frac{x - \mu}{\sigma} = \frac{1{,}98 - 1{,}752}{0{,}071} \approx 3{,}211$$

- Este resultado indica que a altura de Michael Jordan está a 3,21 desvios-padrão acima da média da população. Considerando incomuns valores acima ou abaixo de 2 desvios da média, conclui-se que Michael Jordan é de fato excepcionalmente alto comparando com a população geral.