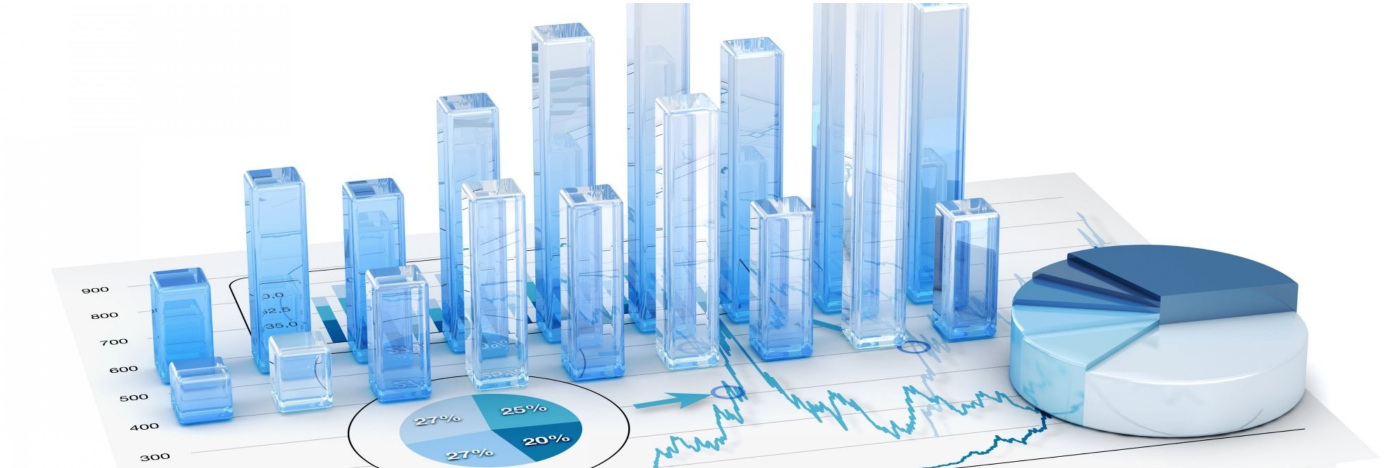
TTI109 – Estatística

Aula 05 – Estatística Descritiva 04



INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

Medidas de variação

¬ Uma *medida de variação* é um valor que representa a *dispersão* de um conjunto de dados.

Amplitude

Definição

A **amplitude** de um conjunto de dados é a diferença entre os valores máximo e mínimo. Para encontrar a amplitude, os dados devem ser quantitativos.

Amplitude = (valor máximo) – (valor mínimo)

(a) Salários iniciais para a empresa A (em milhares de dólares).

Salário	41	38	39	45	$\boxed{47}$	41	44	41	$\boxed{37}$	42

(b) Salários iniciais para a empresa B (em milhares de dólares).

Salário	40	$\boxed{23}$	41	50	49	32	41	29	52	$\boxed{58}$

Amplitudes

Empresa A:

Empresa B:

Medidas de variação

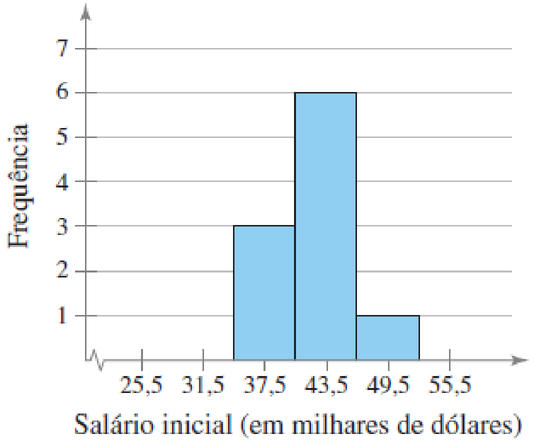
(a) Salários iniciais para a empresa A (em milhares de dólares).

Salário	41	38	39	45	47	41	44	41	37	42

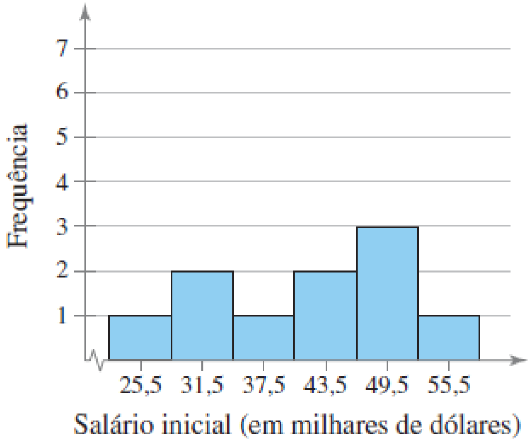
(b) Salários iniciais para a empresa B (em milhares de dólares).

Salário	40	23	41	50	49	32	41	29	52	58

(a) Empresa A.



(b) Empresa B.



- Ambos os conjuntos de dados têm média de 41,5, mediana e moda de 41.
- Ainda assim diferem significativamente, na medida em que os valores no segundo conjunto exibem uma variação maior.

Medidas de variação

Desvio

Definição

O **desvio** de um valor x em uma população é a diferença entre o valor e a média μ do conjunto de dados.

Desvio de $x = x - \mu$.

(a) Salários iniciais para a empresa A (em milhares de dólares).

Salário	41	38	39	45	47	41	44	41	37	42

$$Média = \mu = \frac{41+28+39+45+47+41+44+37+42}{10} = 41{,}5$$

Medidas de variação

Salário (em milhares de dólares) x	Desvio (em milhares de dólares) $x - \mu$
41	–0,5
38	–3,5
39	–2,5
45	3,5
47	5,5
41	–0,5
44	2,5
41	–0,5
37	–4,5
42	0,5
$\Sigma x = 415$	$\boxed{\Sigma (x - \mu) = 0}$

A soma dos desvios é 0

- *A soma dos desvios para qualquer conjunto de dados é zero*. Por isso, não faz sentido encontrar a média dos desvios.
- Sendo assim, calcula-se a soma dos quadrados dos desvios, .
- Em uma população, a média dos quadrados dos desvios é a *variância populacional*.

Medidas de variação

Definição

A **variância populacional** de um conjunto de dados com N elementos é:

$$\text{Variância populacional} = \sigma^2 = \frac{\Sigma (x-\mu)^2}{N}$$

O símbolo σ é a letra minúscula grega sigma.

Definição

O **desvio padrão populacional** de um conjunto de dados populacional de N elementos é a raiz quadrada da variância populacional.

$$\text{Desvio padrão populacional} = \sigma = \sqrt{\sigma^2} = \sqrt{\frac{\Sigma (x-\mu)^2}{N}}$$

¬ *Desvantagem*:
A unidade de medida da variância é diferente da unidade de medida do conjunto de dados (de fato, é o quadrado da unidade original).

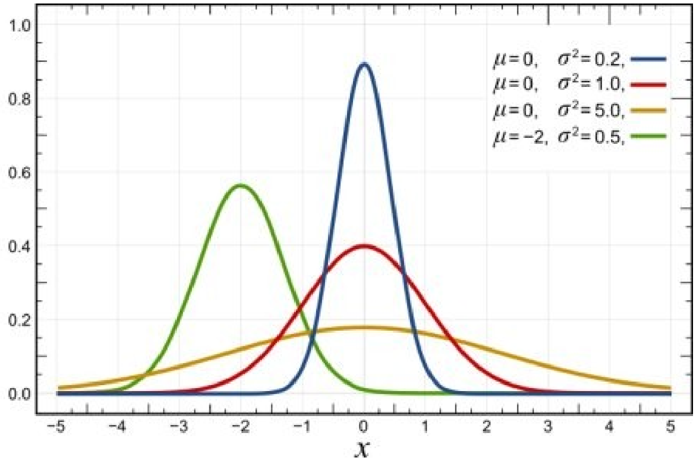
Medidas de variação

Observações sobre o desvio de padrão

- Mede a *variação dos dados com relação à média* e tem a mesma unidade de medida que o conjunto de dados.

- *É sempre maior ou igual a 0*. Quando , o conjunto de dados não apresenta variação (todos os elementos têm o mesmo valor).

- À medida que os valores se afastam da média (isto é, estão mais dispersos),
o valor de aumenta.



INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

(a) Salários iniciais para a empresa A (em milhares de dólares).

Salário	41	38	39	45	47	41	44	41	37	42

Salário x	Desvio $(x - \mu)$	Quadrados $(x - \mu)^2$
41	–0,5	0,25
38	–3,5	12,25
39	–2,5	6,25
45	3,5	12,25
47	5,5	30,25
41	–0,5	0,25
44	2,5	6,25
41	–0,5	0,25
37	–4,5	20,25
42	0,5	0,25
$\Sigma x = 415$		$SS_x = 88{,}5$

$$\boxed{\sigma^2 = \frac{88{,}5}{10} \approx 8{,}9} \Rightarrow \boxed{\sigma = \sqrt{\frac{88{,}5}{10}} \approx 3{,}0}$$

¬⅄ As expressões para a *variância amostral* e *o desvio padrão amostral* diferem ligeiramente daquelas válidas para uma população.

¬⅄ Para calcular , usa-se ao invés de . Além disso, é dividido por ao invés de .

Medidas de variação

Definição

As fórmulas para calcular a **variância amostral** e o **desvio padrão amostral** de um conjunto de dados amostral de *n* elementos estão listadas a seguir.

$$\text{Variância amostral} = s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}.$$

$$\text{Desvio padrão amostral} = s = \sqrt{s^2} = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}.$$

Por que usar ao invés de ?

As estatísticas e definidas ao lado são <u>*estimadores não viesados*</u> (ou seja, não viciados) dos parâmetros e .

Medidas de variação

Encontrando a variância e o desvio padrão amostral

$n = 12$

Em um estudo com jogadores de futebol americano do ensino médio que sofreram lesões, os pesquisadores colocaram os jogadores em dois grupos. Jogadores que se recuperaram das concussões em 14 dias ou menos foram colocados no grupo 1. Aqueles que levaram mais de 14 dias foram para o grupo 2. Os tempos de recuperação (em dias) para o grupo 1 estão listados a seguir. Encontre a variância e o desvio padrão amostrais dos tempos de recuperação.
(Adaptado de: The American Journal of Sports Medicine.)

4 7 6 7 9 5 8 10 9 8 7 10

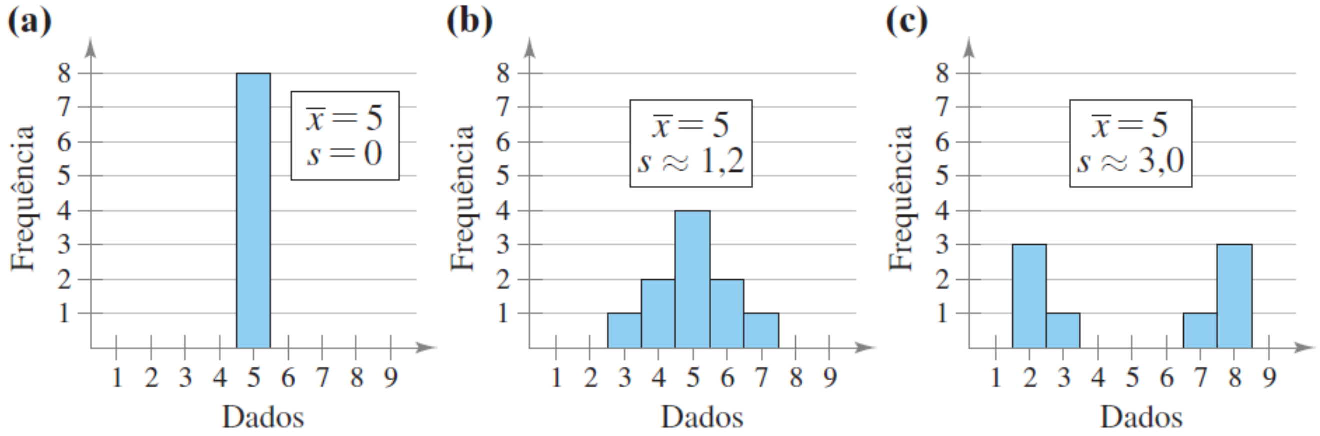
Tempo x	Desvio $(x - \mu)$	Quadrados $(x - \mu)^2$
4	–3,5	12,25
7	–0,5	0,25
6	–1,5	2,25
7	–0,5	0,25
9	1,5	2,25
5	–2,5	6,25
8	0,5	0,25
10	2,5	6,25
9	1,5	2,25
8	0,5	0,25
7	–0,5	0,25
10	2,5	6,25
$\Sigma x = 90$		$SS_x = 39$

$$s^2 = \frac{39}{11} \approx 3{,}5 \Rightarrow s = \sqrt{\frac{39}{11}} \approx 1{,}9$$

Medidas de variação

Interpretando o desvio de padrão

¬⅄ O desvio padrão de uma população / amostra é uma medida que indica o quanto, em média, os valores se desviam da média desse conjunto. *Quanto mais espalhados estiverem os valores, maior será o desvio padrão*.



Podemos usar o desvio padrão para comparar a variação em conjuntos de dados que têm a mesma unidade de medida e médias aproximadamente iguais.

INSTITUTO MAUÁ DE TECNOLOGIA

MAUÁ

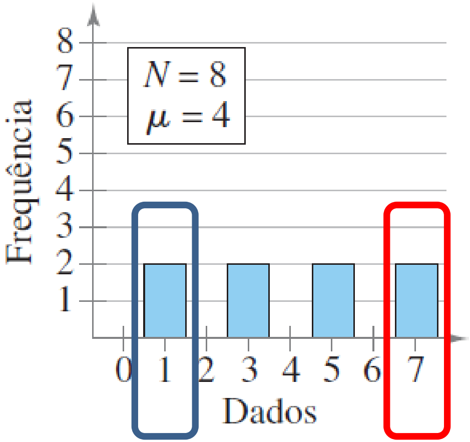
Medidas de variação

Interpretando o desvio de padrão

- Valores de dados que se encontram a mais ou menos dois desvios padrão da média são considerados incomuns, enquanto aqueles que se encontram a mais de três desvios padrão da média são muito incomuns.

- Valores incomuns e muito incomuns têm uma influência maior no desvio padrão do que aqueles que estão mais próximos da média.

python™



Valor x	Desvio $(x - \mu)$	Quadrados $(x - \mu)^2$
1	–3	9
3	–1	1
5	1	1
7	3	9

Medidas de variação

- Muitos conjuntos de dados da vida real têm distribuições que são aproximadamente simétricas e em forma de sino.

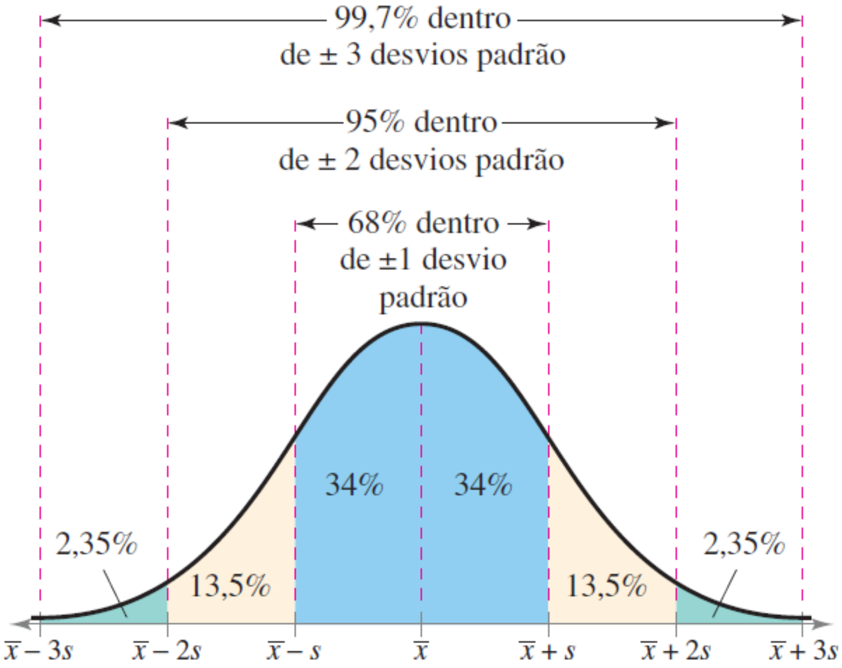
Empírico: Que se baseia somente na experiência ou observação, ou por elas se guia, sem levar em consideração teorias ou métodos científicos; experimental, prático.

Regra Empírica (ou Regra 68-95-99,7)

Para conjuntos de dados com distribuições que são aproximadamente simétricas e com forma de sino, o desvio padrão tem estas características:

1. Cerca de 68% dos dados encontram-se dentro do intervalo de ± 1 desvio padrão em relação à média.
2. Cerca de 95% dos dados encontram-se dentro do intervalo de ± 2 desvios padrão em relação à média.
3. Cerca de 99,7% dos dados encontram-se dentro do intervalo de ± 3 desvios padrão em relação à média.

Medidas de variação

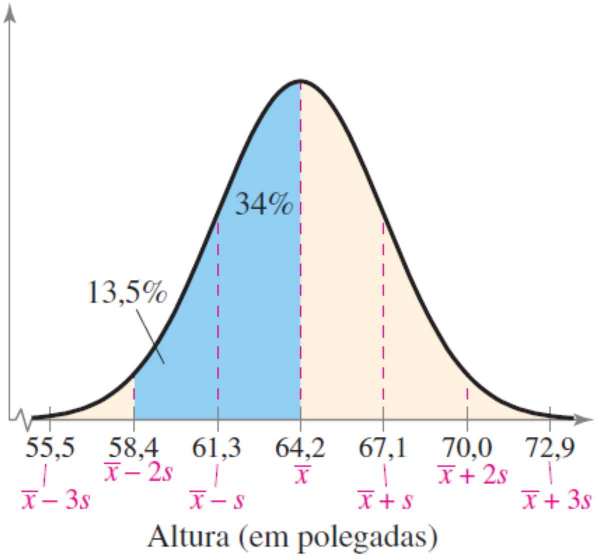


¬⅄ Em uma pesquisa conduzida pelo *National Center for Health Statistics*, a altura média amostral das mulheres nos Estados Unidos (com idades entre 20 e 29 anos) era de 64,2 polegadas, com um desvio padrão amostral de 2,9 polegadas.

Estime a percentagem de mulheres cujas alturas estão entre 58,4 e 64,2 polegadas.

Medidas de variação

Alturas das mulheres nos EUA com idade entre 20 e 29 anos.



- A altura média é 64,2, então, quando subtrairmos dois desvios padrão da altura média, teremos:

- Uma vez que 58,4 situa-se a dois desvios padrão abaixo da altura média, a percentagem das alturas entre 58,4 e 64,2 polegadas é cerca de:

Medidas de variação

Desvio padrão para dados agrupados

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

amostras

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2 f}{n-1}} = \sqrt{\frac{145{,}38}{49}} \approx 1{,}7$$

Número de crianças por domicílio — operações para o cálculo da média e desvio padrão.

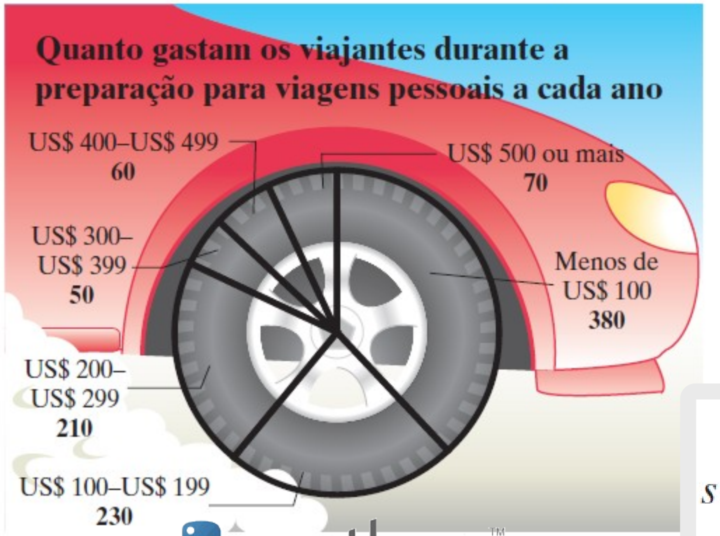
x	f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
0	10	0	–1,82	3,3124	33,1240
1	19	19	–0,82	0,6724	12,7756
2	7	14	0,18	0,0324	0,2268
3	7	21	1,18	1,3924	9,7468
4	2	8	2,18	4,7524	9,5048
5	1	5	3,18	10,1124	10,1124
6	4	24	4,18	17,4724	69,8896
	$\Sigma = 50$	$\Sigma = 91$			$\Sigma = 145{,}38$

$$\bar{x} = \frac{\Sigma xf}{n} = \frac{91}{50} = 1{,}82 \approx 1{,}8$$ média amostral

Medidas de variação

- Quando uma distribuição de frequência tem classes intervalores, podemos estimar a média e o desvio padrão amostral usando o ponto médio de cada classe.

Gastos antes de viajar.



Classe	x	f	xf	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^2 f$
0–99	49,5	380	18.810	–142,5	20.306,25	7.716.375,0
100–199	149,5	230	34.385	–42,5	1.806,25	415.437,5
200–299	249,5	210	52.395	57,5	3.306,25	694.312,5
300–399	349,5	50	17.475	157,5	24.806,25	1.240.312,5
400–499	449,5	60	26.970	257,5	66.306,25	3.978.375,0
500+	599,5	70	41.965	407,5	166.056,25	11.623.937,5

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2 f}{n-1}}=\sqrt{\frac{25.668.750}{999}}\approx 160{,}3 \quad \text{desvio padrão amostral}$$

python™

Medidas de variação

- Para comparar conjuntos de dados com unidades de medida diferentes ou médias diferentes, usa-se o *coeficiente de variação*, que mede a variação de um conjunto de dados com relação à média.

Definição

O **coeficiente de variação (CV)** de um conjunto de dados descreve o desvio padrão como uma percentagem da média.

População: $CV = \dfrac{\sigma}{\mu} \cdot 100\%$ Amostra: $CV = \dfrac{s}{\bar{x}} \cdot 100\%$

Alturas e pesos de um time de basquete.

Alturas	Pesos
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

Calcule o CV para os pesos e as alturas e compare os resultados

Medidas de variação

Alturas e pesos de um time de basquete.

Alturas	Pesos
72	180
74	168
68	225
76	201
74	189
69	192
72	197
79	162
70	174
69	171
77	185
73	210

Alturas: polegadas e polegadas

Pesos: libras e libras

$$CV_P = \frac{\sigma_P}{\mu_P} \cdot 100\% = \frac{17,7}{187,8} \cdot 100\% \approx 9,4\%$$

$$CV_A = \frac{\sigma_A}{\mu_A} \cdot 100\% = \frac{3,3}{72,8} \cdot 100\% \approx 4,5\%$$

Conclusão: A distribuição de pesos é mais dispersa do que a distribuição de alturas

TTI109 – Estatística

Aula 05 – Estatística Descritiva 04