IMPORTANTE

Vamos explorar um pouco mais esse resultado. Observe o Gráfico 8, abaixo:

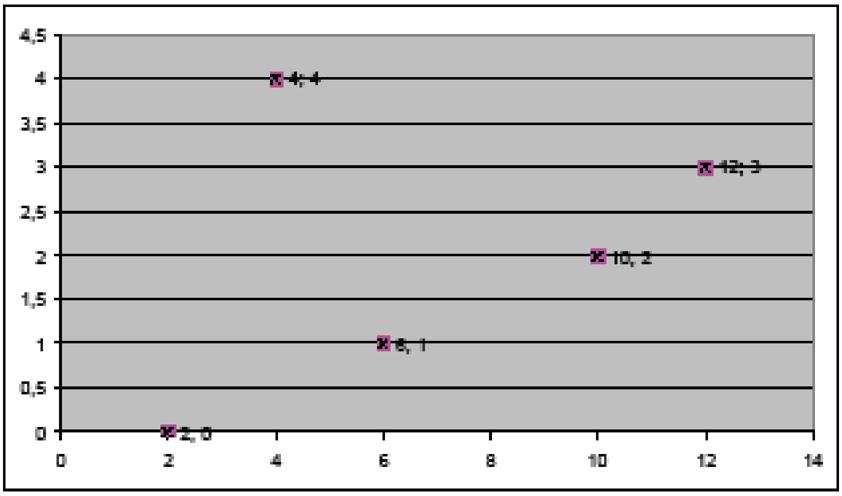

Gráfico 8: Mediana

O Gráfico 8 mostra que: duas famílias não possuem filhos meninos (2,0); 4 famílias possuem 4 meninos; seis famílias possuem 1 menino (6,1); 10 famílias possuem 2 meninos (10,2); 12 famílias possuem 3 meninos (12,3). Temos no nosso conjunto 78 meninos, por quê? Veja:

- 2 famílias não possuem meninos → 2 x 0 = 0;
- 4 famílias possuem 4 meninos → 4 x 4 = 16;
- 6 famílias possuem 1 menino → 6 x 1 = 6;
- 10 famílias possuem 2 meninos → 10 x 2 = 20;
- 12 famílias possuem 3 meninos → 12 x 3 = 36.

Logo, o total de meninos é 0 + 16 + 6 + 20 + 36 = 78 ($\sum = 78$).

A mediana encontrada foi 2, isso significa que **as famílias que possuem dois meninos** dividem nosso conjunto de 78 meninos ao meio: metade desses meninos estão nas famílias com nenhum filho, com um filho e com dois filhos; a outra metade é composta de famílias com dois meninos, com três meninos e famílias com quatro meninos. Agora ficou mais claro que a mediana divide nosso conjunto ao meio.

PRATIQUE

Vá à Secretaria de sua escola e pegue, aleatoriamente, dados sobre 10 famílias. Calcule a média e a mediana do número de filhas meninas.

Ainda não concluímos o estudo sobre mediana. É preciso, por último, calcular a mediana de **dados agrupados em intervalos de classe**. Mas isso, faremos mais à frente.

Moda

ATENÇÃO

Em um conjunto de números, chamamos de **moda** o valor que ocorre com maior freqüência, isto é, o valor mais comum. É assim que podemos dizer que "o salário modal dos empregados de uma indústria é o salário mais comum, isto é, o salário recebido pelo maior número de empregados dessa indústria". (CRESPO, 1995, p. 89). Por exemplo:[58]

a) O conjunto 2, 2, 5, 7, 9, 9, 9, 10, 10, 11, 12, 18 tem *moda* 9;

b) O conjunto 3, 5, 8, 10, 12, 15, 16 não tem moda;

c) O conjunto 2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9 tem duas modas, 4 e 7. Nesse caso é chamado *bimodal*.

Para **dados agrupados sem intervalos de classe**, é possível determinar imediatamente a moda, como nos exemplos acima. Mas, por exemplo, a Tabela 28, p. 77, indica que a moda é 3. Por quê? Porque o valor que mais se repete é aquele que possui maior freqüência simples, não é mesmo?

É ainda possível encontrar a moda para **dados agrupados com intervalos de classe**, mas deixaremos esse estudo para uma outra oportunidade.

Expressões gráficas da moda

ATENÇÃO

Em uma curva de freqüência, o maior valor de um conjunto é chamado **moda**. Na prática, a **moda** é o valor que corresponde, no eixo das abscissas, ao ponto de ordenada máxima, em outras palavras. Veja exemplos abaixo:

58 SPIEGEL (1975, p. 74).

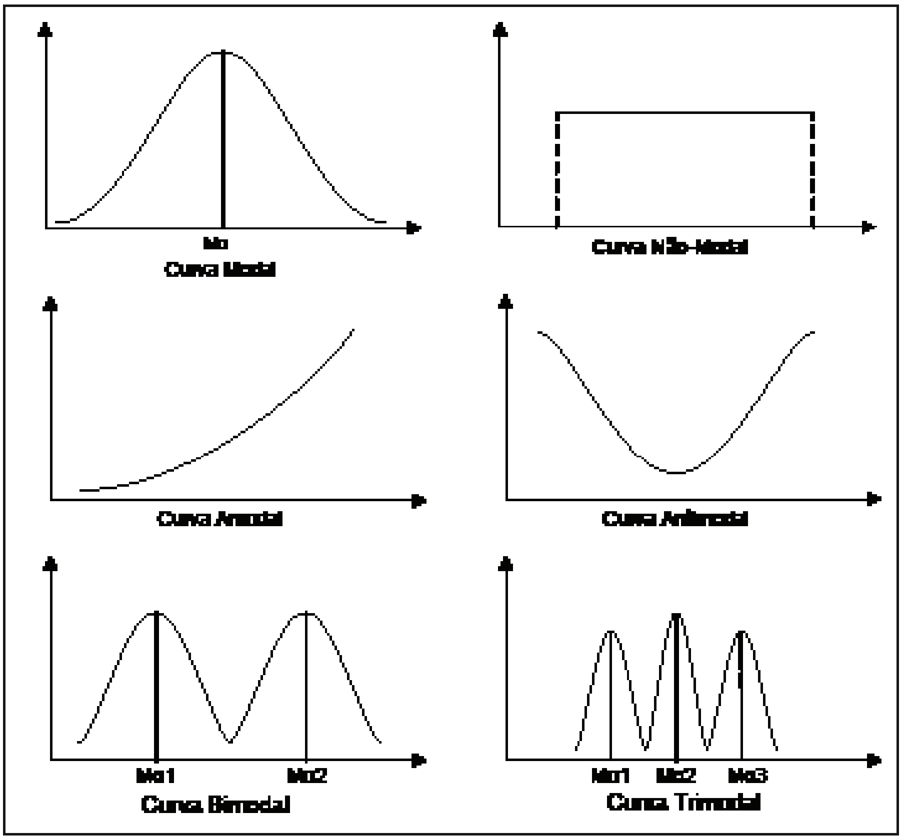

Figura 26: Curvas Modais

Vamos verificar a Curva Modal, acima (primeiro gráfico). Repare que ela possui um valor maior, mais alto no gráfico. O que isso indica? Indica que é o maior valor que o conjunto pode assumir, por isso, é a *moda* do conjunto.

Já no último gráfico – Curva Trimodal –, identificamos três valores de máximo, isto é, o conjunto possui três valores "maiores" que todos os demais, por isso, trimodal.

Relação entre Média, Mediana e Moda

Em curvas simétricas, unimodais, a média ($\bar{x}$), a mediana (Md) e a Moda (Mo) coincidem. Observe:

IMPORTANTE

SAIBA MAIS

Conjuntos com mais de três valores máximos são chamados de polimodais.

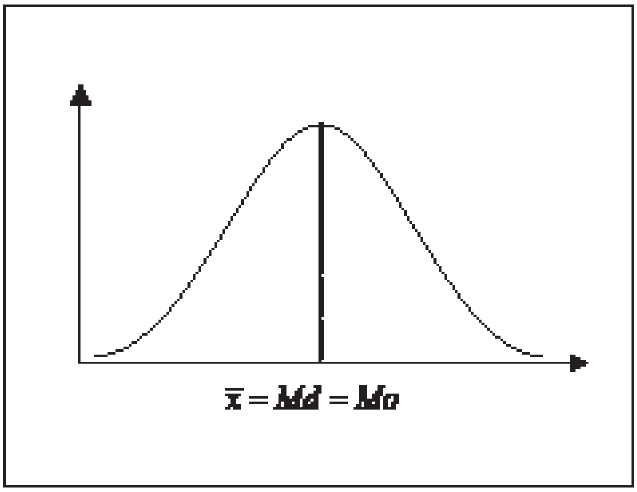

Figura 27: Média, Mediana, Moda: Curva Simétrica

Em curvas de freqüência desviadas para a direita e para a esquerda, as posições são diferentes. Veja:

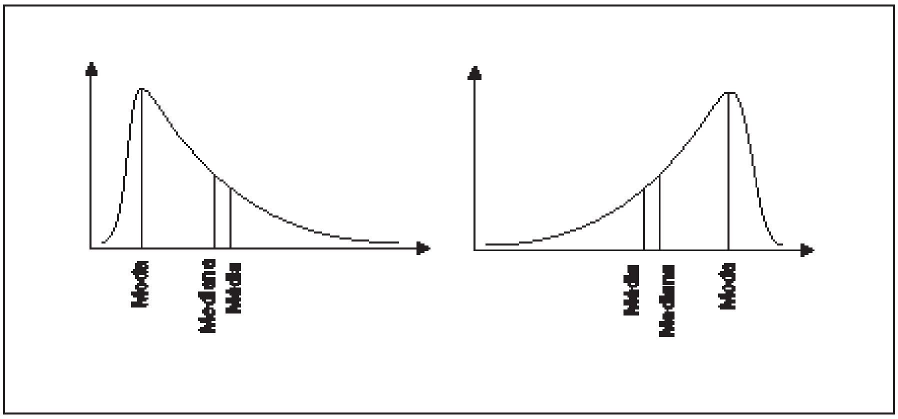

Figura 28: Média, Mediana, Moda: Curva Assimétrica

PRATIQUE

Determinar a média, a mediana e a moda dos conjuntos de números:[59]

A = 7, 4, 10, 9, 15, 12, 7, 9, 7

B = 8, 11, 4, 3, 2, 5, 10, 6, 4, 1, 10, 8, 12, 6, 5, 7

(Atenção: não se esqueça de colocar os conjuntos em rol).

59 SPIEGEL (1975, p. 105).

Seção 3: Medidas de Dispersão

Até aqui, vimos que **média**, **mediana** e **moda** são valores que podem servir de comparação, mas, fundamentalmente, fornecem a posição de qualquer elemento do conjunto. Mas para interpretar dados estatísticos, mesmo aqueles já convenientemente simplificados, é preciso conhecer a evolução desses dados.

Um exemplo clássico para a compreensão da importância das **medidas de dispersão** é o da comparação de temperaturas entre cidades[60]: saber que a temperatura média de duas cidades é de 24ºC não me diz muita coisa a respeito da variação dessas temperaturas.

Em uma cidade, o dia pode ter iniciado muito frio e terminado muito quente; aqui, ocorreu uma grande variação da temperatura.

Na outra cidade, o dia pode ter iniciado e terminado como 24º C; nesse caso, não haveria variação alguma de temperatura.

Viu? Embora as médias sejam importantes, elas não são suficientes para as *inferências estatísticas*, por isso, precisamos de outras medidas.

Vamos reforçar a importância das medidas de dispersão, por meio de um exercício. Consideraremos os três conjuntos abaixo, com seus respectivos valores:[61]

X: 70, 70, 70, 70, 70.

Y: 68, 69, 70, 71, 72.

Z: 5, 15, 50, 120, 160.

Vamos calcular a *média* das idades dos três conjuntos:

Solução:

Para calcular as médias, precisaremos da Fórmula 1, p. 71:

$$\bar{x} = \frac{\sum x_i}{n}, onde \begin{cases} \bar{x} = Média\ Aritmética \\ x_i = Valores\ da\ Variável \\ n = Número\ Total\ de\ Valores \end{cases}$$

60 CRESPO (1995, p. 108).
61 CRESPO (1995, p. 108).

IMPORTANTE

Então,

Para X: $\bar{x} = \dfrac{70+70+70+70+70}{5} = \dfrac{350}{5} = 70$

Para Y: $\bar{x} = \dfrac{68+69+70+71+72}{5} = \dfrac{350}{5} = 70$

Para Z: $\bar{x} = \dfrac{5+15+50+120+160}{5} = \dfrac{350}{5} = 70$

Como podemos observar, os três conjuntos possuem a mesma *média aritmética*: 70.

Mas também, podemos notar que o conjunto **X** é mais **homogêneo** do que os conjuntos **Y** e **Z**; o conjunto **Y**, por sua vez, é mais **homogêneo** que o conjunto **Z**; por fim, o conjunto **Z** é o mais **heterogêneo** de todos. Viu? Mesmo possuindo a mesma média, os conjuntos apresentam comportamentos muito diferentes. A isso chamamos de **dispersão**.

Dispersão e Variação

ATENÇÃO

***Dispersão** (ou **variabilidade**) de um conjunto refere-se à maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central[62] tomado como ponto de comparação.*

No nosso exercício acima, os conjuntos **X**, **Y** e **Z** apresentam como ponto de tendência central para fins de comparação a *média*. Essa *média* é a mesma para os três conjuntos: 70. Assim, o conjunto **X** apresenta *dispersão* nula, pois não há variação dos valores do conjunto em relação a essa média; o conjunto **Y** apresenta *dispersão* menor que o conjunto **Z**; isso porque os valores de **Y** estão mais próximos que os do conjunto **Z**.

Em resumo, a estatística recorre às **medidas de dispersão** (ou **de variabilidade**) quando deseja *qualificar* os valores de uma variável, ressaltando a maior ou menor dispersão entre

62 Ver Seção 2: Medidas de Tendência Central, p. 78.

esses valores e a sua medida de posição.[63] Dessas medidas de dispersão,[64] estudaremos apenas o *desvio padrão* e o *coeficiente de variação*.

IMPORTANTE

Desvio Padrão

*O **desvio padrão** é a medida da variação, da dispersão, de um conjunto.*

ATENÇÃO

Assim, quanto maior for o desvio padrão, maior será a heterogeneidade entre os valores que estão sendo analisados. Isso significa, portanto, que quanto maior for o desvio padrão, maior será a variação entre os valores. Vamos entender melhor isso.

SAIBA MAIS

Conjuntos mais homogêneos apresentam desvios-padrão menores.

De volta aos conjuntos **X**, **Y** e **Z** acima, vimos que a *média* de todos eles era 70. Notamos, também, que os conjuntos **X** e **Y** eram mais homogêneos que o conjunto **Z**. Agora vamos calcular essa medida matematicamente, utilizando mais uma fórmula:

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2}\text{, onde } \begin{cases} s \text{ é o desvio padrão} \\ n \text{ é a soma das freqüências} \end{cases}$$

Fórmula 4: Desvio Padrão: Dados Não Agrupados

ATENÇÃO

SAIBA MAIS

Muita atenção à diferença abaixo:

$$\frac{\sum x_i^2}{n} \neq \left(\frac{\sum x_i}{n}\right)^2.$$

Matematicamente, os parênteses alteram tudo. Acompanhe o exercício para detectar a diferença.

Os nossos conjuntos **X**, **Y** e **Z** são de **dados não agrupados**. Vamos representá-los em Tabelas, para melhor visualização.

63 Veremos, mais adiante, as *medidas de posição*. Por ora, podemos considerar, apenas, as medidas de tendência central.

64 A lista de medidas de dispersão é longa. Para Spiegel (1975), essas medidas são: a amplitude total; o desvio médio; a amplitude semi-interquartílica ou o desvio quartílico; o desvio-padrão; a variância; o coeficiente de variação.

Tabela 29: Desvio Padrão: Exercício

Tabela X		Tabela Y		Tabela Z	
x_i	x_i^2	x_i	x_i^2	x_i	x_i^2
70	490	68	4624	5	25
70	490	69	4761	15	225
70	490	70	4900	50	2500
70	490	71	5041	120	14400
70	490	72	5184	160	25600
$\sum = 350$	$\sum = 24500$	$\sum = 350$	$\sum = 24510$	$\sum = 350$	$\sum = 42750$

Note que cada valor do conjunto é representado por x_i e seu quadrado é x_i^2. Sabemos que *n* é igual a 5, para todos os conjuntos. Agora ficou fácil calcular o desvio padrão dos três conjuntos. Vejamos:

Solução:

Aplicando a Fórmula 4, temos que:

Para o conjunto X:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 2450 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24500}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4900 - 4900} = 0$$

Para o conjunto Y:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 24510 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{24510}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{4902 - 4900} = \sqrt{2} = 1{,}4$$

Para o conjunto Z:

$$\begin{cases} \sum x_i = 350 \\ \sum x_i^2 = 42750 \end{cases} \quad \text{Então,}$$

$$s = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} = \sqrt{\frac{42750}{5} - \left(\frac{350}{5}\right)^2} = \sqrt{8550 - 4900} = \sqrt{3650} = 60{,}4$$

Você reparou que colocando na tabela os elementos que iremos usar (x_i e x_i^2) fica mais fácil resolver o problema? Depois de todos esses cálculos, temos que:

- O desvio padrão do conjunto **X** é igual a 0. De fato, isso significa que não há variação alguma no conjunto, portanto, é um conjunto *homogêneo*;
- O desvio padrão do conjunto **Y** é igual a 1,4 e o do conjunto **Z** é igual a 60,4. Comparando-se os dois conjuntos, vemos que há uma pequena variação em **Y** (1,4) e uma alta variação em **Z** (60,4). Na prática, significa que os valores do conjunto **Y** estão mais próximos da *média*, ao passo que, em **Z**, os valores do conjunto estão muito distantes da *média*.

Graficamente, é ainda mais fácil identificar um conjunto *mais homogêneo*. Observe:

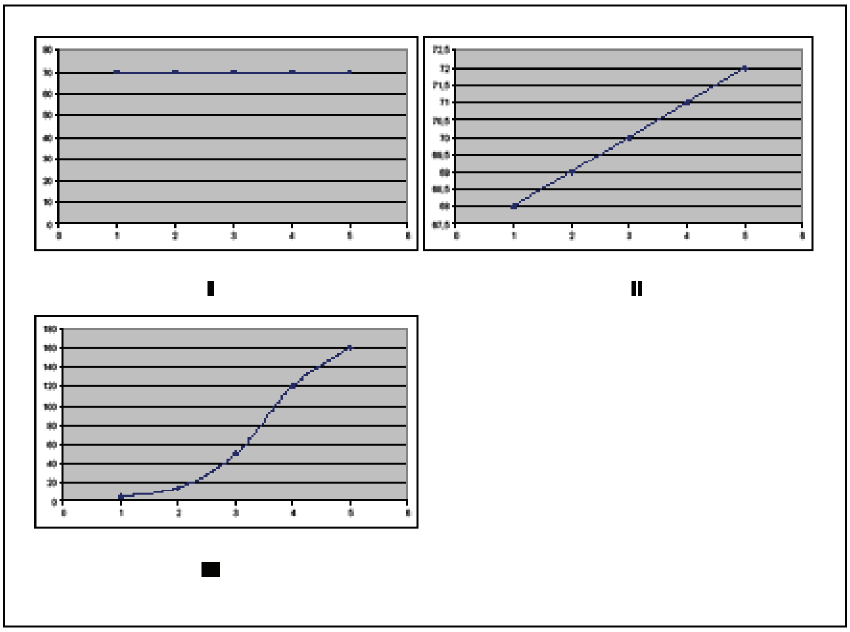

Figura 29: Desvio Padrão: Gráficos: Exercício

Você é capaz de dizer qual das três representações gráficas acima, é o conjunto **X**? E o conjunto **Y**? E o conjunto **Z**? Note que se o conjunto for homogêneo (**I**), o gráfico é uma linha reta paralela ao eixo *x*; observe também, que quanto menos homogêneo o conjunto, a reta tenderá a ser uma curva.

IMPORTANTE

PRATIQUE

Calcule o desvio padrão dos conjuntos abaixo:

A = 12, 6, 7, 3, 15, 10, 18, 5

B = 9, 3, 8, 8, 9, 8, 9, 18

Vamos fazer um exercício de cálculo do desvio padrão para conjuntos com **dados agrupados sem intervalos de classe**. Nesse caso, como temos freqüências (ou seja, como os valores se repetem), vamos fazer uma pequena alteração na Fórmula.

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}\text{, onde } \begin{cases} f_i x_i^2 \text{ é o produto do quadrado dos valores pela freqüencia} \\ f_i x_i \text{ é o produto dos valores pela freqüência} \end{cases}$$

Fórmula 5: Desvio Padrão: Dados Agrupados

Vamos encontrar o desvio padrão da Tabela 30, abaixo.

Tabela 30: Desvio Padrão: Dados Agrupados: Sem Intervalos de Classe: Exercício

x_i	f_i
0	2
1	6
2	12
3	7
4	3
	$\sum = 30$

Fonte: CRESPO (1995, p. 115).

Da mesma maneira que estamos resolvendo nossos exercícios, aqui, vamos acrescentar à Tabela três colunas que serão úteis.

Tabela 31: Desvio Padrão: Exercício: Continuação

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
0	2	0	0	0
1	6	6	1	6
2	12	24	4	48
3	7	21	9	63
4	3	12	16	48
	$\sum = 30$	$\sum = 63$		$\sum =$ 165

Com a Tabela assim, é fácil encontrar o desvio padrão. Veja:

Sabendo que: $\begin{cases} n = \sum f_i = 30 \\ \sum f_i x_i^2 = 165 \\ \sum f_i x_i = 63 \end{cases}$. Então,

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{165}{30} - \left(\frac{63}{30}\right)^2} = \sqrt{5,5 - 4,41} = \sqrt{1,09} = 1,044$$

Portanto, o desvio padrão é de 1,044.

Para encontrar o **desvio padrão** de um conjunto **com intervalos de classe**, utilizaremos o mesmo recurso de acrescentar à tabela os dados que iremos precisar na mesma Fórmula 5, acima. Como recurso didático, usaremos a mesma Fórmula para dados agrupados *sem intervalos de classe*.

Primeiro, vamos repetir a Fórmula 5:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}, \text{ onde } \begin{cases} f_i x_i^2 \text{ é o produto do quadrado dos valores pela freqüencia} \\ f_i x_i \text{ é o produto dos valores pela freqüência} \end{cases}$$

Suponha, agora, que queiramos encontrar o desvio padrão da Tabela 32, abaixo:

IMPORTANTE

SAIBA MAIS

***Relembrando:** Se n é quantidade de valores por que deu 30 se os valores são 0, 1, 2, 3 e 4? Ou seja, por que n não é 5?*
***Simples!** Porque, na verdade, a Tabela indica que temos os seguintes valores: 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 4, 4 e 4. Isso é que é a freqüência (f_i). Temos, portanto, 30 valores organizados por freqüências.*

Tabela 32: Desvio Padrão: Dados Agrupados: Com Intervalos de Classe: Exercício

Estaturas	f_i
150–154	4
154–158	9
158–162	11
162–166	8
166–170	5
170–174	3
	$\sum = 40$

Fonte: CRESPO (1995, p. 116)

SAIBA MAIS

Por exemplo, no intervalo 150–154, os valores podem assumir de 150 cm até 154 cm: esses são os valores de mínimo e de máximo.

O que essa tabela apresenta de diferente? Os dados são agrupados com intervalos de classe. Ou seja, os valores variam de *um valor mínimo para um máximo*. Portanto, temos um problema a resolver!

A Fórmula 5, acima, é para o cálculo do desvio padrão de um conjunto de *dados agrupados sem intervalos de classe*. Isso significa que nela temos x_i e não um *intervalo de classe*, como, por exemplo, 150–154. Mas se eu tivesse **um valor** ao invés de **um intervalo de valores** (como é o caso), a Fórmula 5 poderia ser a mesma, não é verdade?

SAIBA MAIS

O ponto médio é o ponto que está no meio do intervalo. Veja:
O que está no meio do intervalo que varia de 150 cm a 154 cm? 152 cm é o ponto médio.
Qual é o ponto médio do intervalo 154–158?
É 156 cm que está no meio.
E assim por diante.

Bem, vamos utilizar um recurso para manter a mesma Fórmula: vamos encontrar um ponto, que chamaremos **ponto médio**, para cada *intervalo de classe*. Dessa maneira, teremos x_i como no exercício anterior e, assim, poderemos utilizar a mesma Fórmula.

Os demais elementos ($f_i x_i$, x_i^2 e $f_i x_i^2$) já sabemos como encontrar. Agora, vamos à solução. Nossa Tabela, com os acréscimos necessários, ficará assim:

Tabela 33: Desvio Padrão: Exercício: Continuação

Estaturas	f_i	x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
150–154	4	152	608	23.104	92.416
154–158	9	156	1.404	24.336	219.024
158–162	11	160	1.760	25.600	281.600
162–166	8	164	1.312	26.896	215.168
166–170	5	168	840	28.224	141.120
170–174	3	172	516	29.584	88.752
	$\sum = 40$		$\sum = 6.440$		$\sum = 1.038.080$

Com a Tabela preenchida, vamos encontrar o desvio padrão.

Solução:

Sabendo que $\begin{cases} n = \sum f_i = 40 \\ \sum f_i x_i^2 = 1.038.080 \\ \sum f_i x_i = 6.440 \end{cases}$. Então,

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{1.038.080}{40} - \left(\frac{6.440}{40}\right)^2} = \sqrt{25.952 - 25.921} = \sqrt{31} = 5{,}567$$

Viu?! Acrescentando os dados que iremos necessitar para o cálculo à Tabela, tudo fica mais fácil. O desvio padrão é 5,57 cm.

Calcule o desvio padrão da distribuição abaixo:

PRATIQUE

CUSTO (R$)	450 – 550	550 – 650	650 – 750	750 – 850	850 – 950	950 – 1.050	1.050 – 1.150
f_i	8	10	11	16	13	5	1

Fonte: CRESPO (1995, p. 118).

IMPORTANTE

ATENÇÃO

Não se esqueça de montar a Tabela.

Estaturas	f_i	x_i	$f_i x_i$	x_i^2	$f_i x_i^2$
450–550	8				
550–650	10				
650–750	11				
750–850	16				
850–950	13				
950–1.050	5				
1.050–1.150	1				
	$\sum=$		$\sum=$		$\sum=$

Coeficiente de Variação

Até aqui, nossos esforços têm se voltado para caracterizar, com o maior rigor possível, a dispersão dos conjuntos. O **coeficiente de variação** é uma medida muito útil para essa intenção.

O **coeficiente de variação** (CV) está sempre relacionado ao valor médio de um conjunto porque, como já vimos, a dispersão é uma medida sempre relacionada a uma determinada média.

Sua fórmula é bastante simples:

$$CV = \frac{\text{desvio padrão}}{\text{média}} \times 100$$

De maneira mais simplificada:

$$CV = \frac{s}{\bar{x}} \times 100,\ \text{onde} \begin{cases} s \text{ é o desvio padrão} \\ \bar{x} \text{ é a média} \end{cases}$$

Fórmula 6: Coeficiente de Variação