## Capítulo 7 – Medidas de dispersão

### Introdução

Para a compreensão deste capítulo, é necessário que você tenha entendido os conceitos apresentados nos capítulos 4 (ponto médio, classes e frequência) e 6 (média). Assim será fácil para você calcular as medidas de dispersão (amplitude total, variância, desvio padrão e coeficiente de variação) para variáveis quantitativas e interpretar os valores dessas medidas em diferentes situações do cotidiano.

Estudamos, no capítulo anterior, mecanismos para encontrar valores (média, mediana e moda) que sintetizam o comportamento dos elementos de um conjunto de dados. Esses valores fornecem parâmetros significativos para uma análise dos dados, porém ainda é importante identificar como variam ou diferenciam as características dos elementos de um conjunto.

Imagine, por exemplo, que você precise analisar o valor da cotação do dólar no ano passado. Inicialmente, você precisará de uma média aritmética envolvendo os doze meses. Além da média, precisamos observar se houve ou não uma discrepância significativa do valor durante esses meses. Caso positivo, então podemos afirmar que ocorreu uma grande variabilidade ou dispersão dos dados em relação à média. Caso contrário, então o valor da cotação permaneceu bastante semelhante durante o ano.

Nesse capítulo, aprenderemos como medir o grau de concentração ou dispersão dos dados em torno da média. Por isso estudaremos as principais medidas de dispersão, que são: amplitude total, variância, desvio padrão e coeficiente de variação. A escolha de uma medida em relação à outra dependerá do objetivo que se pretende alcançar.

Para exemplificar os cálculos das medidas de dispersão, usaremos tanto dados não agrupados quanto dados agrupados com e sem intervalos de classe. Começaremos com a amplitude total.

### 7.1 Amplitude total

De acordo com Crespo (2002), a amplitude total é a diferença entre o maior e o menor valor analisado em uma variável. Vejamos, agora, como calcular amplitude total com dados agrupados e não agrupados.

### 7.1.1 Amplitude total – dados não agrupados

Para dados não agrupados, o cálculo da amplitude total pode ser expresso pela seguinte fórmula:

$$AT = x(máx) - x(mín)$$

Considerando os valores das variáveis A, B e C apresentados a seguir, vamos, então, calcular a amplitude total.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | 80 | 80 | 80 | 80 | 80 | 80 | 80 | 80 | 80 | 80 |
| **B** | 76 | 77 | 78 | 79 | 80 | 80 | 81 | 82 | 83 | 84 |
| **C** | 55 | 65 | 70 | 75 | 80 | 85 | 85 | 90 | 95 | 100 |

Assim, aplicando a fórmula anterior para esses dados, obteremos os seguintes resultados:

$$AT_A = x(máx) - x(mín) = 80 - 80 = 0$$

$$AT_B = x(máx) - x(mín) = 84 - 76 = 8$$

$$AT_C = x(máx) - x(mín) = 100 - 55 = 45$$

Nesse caso, podemos observar que a variável A obteve uma amplitude total igual a 0, ou seja, uma dispersão nula. Então significa que os valores não variam entre si. A variável B, por sua vez, obteve uma amplitude igual a 8. Já a variável C teve uma amplitude total igual a 45.

Embora o valor da amplitude total tenha sido diferente para as variáveis A, B e C, caso você calcule a média aritmética para essas variáveis, observará que o valor encontrado será igual a 80 para todas as situações. Independentemente da média, podemos verificar que a variável A possui elementos muito mais homogêneos do que as variáveis B e C. E, também, que os elementos da variável B são mais homogêneos do que os da variável C. Portanto é possível encontrar conjuntos de elementos que possuem o mesmo valor da média aritmética, mas eles podem ser compostos por elementos totalmente diferentes.

De um modo geral, quanto maior for o valor encontrado para a amplitude total, maior será a discrepância ou a variação entre os valores da variável.

### 7.1.2 Amplitude total – dados agrupados

Para dados agrupados sem intervalos de classe, a fórmula usada para a identificação da amplitude total é semelhante à utilizada para dados não agrupados. Porém o processo se diferencia porque identificaremos o maior e o menor valor em uma tabela de distribuição de frequência.

Por exemplo, considere os dados da variável $x$ apresentados no quadro 1.

Quadro 1: Valores e frequência da variável $x$

| $x_i$ | $f_i$ |
|---|---|
| 1 | 1 |
| 2 | 3 |
| 3 | 2 |
| 4 | 5 |
| 5 | 9 |

Nesse caso, como 5 e 1 são o maior e o menor valores, respectivamente, da variável $x$, aplicando a fórmula anterior, teríamos o seguinte resultado:

$$AT = x(máx) - x(mín) = 5 - 1 = 4$$

Cabe ressaltar que esses valores foram selecionados independentemente do valor da frequência associada a eles.

No entanto, se nos depararmos com dados agrupados com intervalos de classe, a amplitude total será encontrada pela diferença entre o limite superior da última classe e o limite inferior da primeira classe, conforme pode ser expresso na fórmula a seguir:

$$AT = L(máx) - l(mín)$$

Cabe ressaltar que no valor obtido para a amplitude total com dados agrupados com intervalos de classe é levado em consideração somente os dois valores extremos da série, descartando qualquer valor intermediário existente. Sendo assim, o valor obtido é somente uma indicação aproximada da dispersão dos dados.

Considere, agora, os dados do quadro 2 para o cálculo da amplitude total.

Quadro 2: Valores e frequência da variável salário

| SALÁRIO (EM REAIS) | $f_i$ |
|---|---|
| $1000 \vdash 2000$ | 07 |
| $2000 \vdash 3000$ | 13 |
| $3000 \vdash 4000$ | 21 |
| $4000 \vdash 5000$ | 10 |
| $5000 \vdash 6000$ | 04 |

Nesse caso, aplicando a fórmula apresentada anteriormente, obteremos o seguinte resultado:

$$AT = L(máx) - l(mín) = 6000 - 1000 = 5000$$

Logo $AT = 5000$ reais. Diante disso, podemos verificar que a amplitude total como medida de dispersão é um valor limitado e instável, visto que valores internos da série nunca são levados em consideração.

Veja agora a medida de dispersão variância e desvio padrão.

## 7.2 Variância e desvio padrão

A variância e o desvio padrão são medidas que levam em consideração a totalidade dos valores da variável em estudo, e não apenas os valores externos, como a amplitude total (CRESPO, 2002). Por isso, essas medidas são índices de variabilidade bastantes estáveis e, consequentemente, muito utilizados no cotidiano. Além disso, a variância e o desvio padrão complementam informações obtidas pelas medidas de tendência central.

A variância, denotada por $s^2$, é encontrada a partir dos desvios em torno da média aritmética, conforme pode ser observado na fórmula a seguir:

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

Aqui $x_i$ representa cada elemento do conjunto de dados, $\bar{x}$ é a média do conjunto e $n$ representa o número de elementos do conjunto.

O desvio padrão, denotado por $s$, é a raiz quadrada da variância. Assim teremos:

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

ou

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

O desvio padrão é a medida de dispersão mais utilizada porque aponta de forma mais precisa a dispersão dos valores em relação à média aritmética (NAZARETH, 2003). Para exemplificar o cálculo do desvio padrão para dados não agrupados, considere a série de valores a seguir:

18, 22, 15, 17, 19, 21, 16

Inicialmente, precisamos encontrar a média dessa série, que é 18,2. Depois, precisaremos encontrar, para cada elemento, a diferença do seu valor e a média. Para facilitar o processo, vamos criar uma tabela contendo duas colunas, uma para o valor de $x_i$, e outra para o valor de $x_i - \bar{x}$, conforme pode ser visto no quadro 3.

Quadro 3: Valores para $x_i$ e $x_i - \bar{x}$

| $x_i$ | $x_i - \bar{x}$ |
|---|---|
| 18 | -0,2 |
| 22 | 3,8 |
| 15 | -3,2 |
| 17 | -1,2 |
| 19 | 0,8 |
| 21 | 2,8 |
| 16 | -2,2 |

A partir dos valores da tabela, calculamos $\sum(x_i - \bar{x})^2$, que é igual a 0,36. Assim, aplicando a fórmula do desvio padrão, teríamos o seguinte:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{0,36}{7}} = \sqrt{0,05} = 0,02$$

O desvio padrão de uma série será sempre um valor positivo, e quanto maior esse valor, maior será a dispersão entre os elementos.

Quando nos deparamos com dados agrupados, o valor das frequências também precisa ser levado em consideração para o cálculo do desvio padrão. Assim sendo, a fórmula para o cálculo do desvio padrão para dados agrupados é a seguinte:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2}$$

Temos $f_i$, que representa a frequência de um determinado elemento.

Para exemplificar o cálculo do desvio padrão para dados agrupados sem intervalos de classe, vamos considerar os dados apresentados no quadro 1. Para auxiliar a aplicação da fórmula do desvio padrão, criaremos um quadro para encontrar os somatórios da frequência dos valores existentes (equivalente a *n*), de $f_i x_i$ e de $f_i x_i^2$, conforme pode ser visualizado no quadro 4.

Quadro 4: Valores para $x_i$, $f_i$, $f_i x_i$ e $f_i x_i^2$

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| 1 | 1 | 1 | 1 |
| 2 | 3 | 6 | 12 |
| 3 | 2 | 6 | 18 |
| 4 | 5 | 20 | 80 |
| 5 | 9 | 45 | 225 |
| | $\sum = 20$ | $\sum = 78$ | $\sum = 336$ |

Aplicando, agora, a regra do desvio padrão para dados agrupados, teríamos:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{336}{20} - \left(\frac{78}{20}\right)^2} = \sqrt{16{,}8 - 15{,}21} = \sqrt{1{,}59} = 1{,}26$$

O processo para encontrar o desvio padrão para dados agrupados com intervalos de classe é semelhante ao anterior, sendo apenas necessário encontrar o ponto médio ($x_i$) de cada uma das classes antes de calcular o produto de $f_i x_i$ e de $f_i x_i^2$. Assim realizaremos a multiplicação do ponto médio de cada classe com a sua respectiva frequência ao invés do valor exato da variável.

Para exemplificar o cálculo do desvio padrão, vamos considerar os dados da variável idade no quadro 5.

Quadro 5: Valores e frequência da variável idade

| IDADE | $f_i$ |
|---|---|
| $10 \vdash 20$ | 07 |
| $20 \vdash 30$ | 13 |

| | |
|---|---|
| $30 \vdash 40$ | 21 |
| $40 \vdash 50$ | 10 |
| $50 \vdash 60$ | 04 |

De modo similar, construiremos uma tabela para apresentar os valores de
$x_i$, $f_i x_i$ e $f_i x_i^2$ para cada uma das classes da variável idade, conforme pode ser observado no quadro 6.

Quadro 6: Valores para $x_i$, $f_i x_i$ e $f_i x_i^2$

| IDADE | $f_i$ | $x_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|---|
| $10 \vdash 20$ | 07 | 15 | 105 | 1.575 |
| $20 \vdash 30$ | 13 | 25 | 325 | 8.125 |
| $30 \vdash 40$ | 21 | 35 | 735 | 25.725 |
| $40 \vdash 50$ | 10 | 45 | 450 | 20.250 |
| $50 \vdash 60$ | 04 | 55 | 220 | 12.100 |
| | $\sum = 55$ | - | $\sum = 1.835$ | $\sum = 67.775$ |

Ao aplicar a fórmula com os dados obtidos na tabela, teríamos o seguinte:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{67.775}{55} - \left(\frac{1.835}{55}\right)^2}$$

$$= \sqrt{1.232{,}27 - 1.113{,}01} = \sqrt{119{,}25}$$

$$= 10{,}920$$

Logo $s = 10{,}92$.

Segundo Crespo (2002), o desvio padrão possui algumas propriedades que permitem introduzir, no cálculo do desvio padrão, simplificações úteis. Entre as propriedades existentes, destacam-se:

- somando-se (ou subtraindo-se) uma constante a (de) todos os valores de uma variável, o desvio padrão permanecerá o mesmo;
- multiplicando-se todos os valores de uma variável por uma constante (diferente de zero), o desvio padrão ficará multiplicado por essa constante.

Você já conheceu a amplitude total, a variância e o desvio padrão. Agora finalizaremos com o coeficiente de variação.

### 7.3 Coeficiente de variação

O coeficiente de variação é uma medida relativa de dispersão. Ela é útil quando se deseja comparar em termos relativos o grau de concentração em torno da média de séries distintas. É calculado por:

$$CV = \frac{S}{\bar{x}} * 100$$

Em que:

S = desvio-padrão

$\bar{x}$ = média

O coeficiente de variação é expresso em porcentagem.

Para exemplificar, suponha que o acesso médio de homens em um mês em um sítio *web* é de 3.500, com desvio padrão de 900; e das mulheres é em média 2.700, com desvio padrão de 1.100. Então:

Para os homens: $CV = \frac{S}{\bar{x}} = \frac{900}{3500} * 100 \cong 25{,}71\%$

Para as mulheres: $CV = \frac{S}{\bar{x}} = \frac{1100}{2700} * 100 \cong 40{,}74\%$

Assim concluímos que os acessos das mulheres apresentam maior dispersão relativa do que os dos homens.

Na prática, considera-se que um coeficiente de variância superior a 50% indica alto grau de dispersão e, por consequência, baixa representatividade da média. Por outro lado, quanto menor for o valor de seu coeficiente de variância, mais representativa é a média (MARTINS; DONAIRE, 2004, p. 164).

Chegamos ao fim do capítulo e da disciplina. Esperamos que você tenha conseguido vislumbrar as possibilidades de aplicações que a estatística fornece. Você aprendeu, neste capítulo, o porquê da necessidade de que um conjunto de dados seja complementado por uma medida de dispersão ou variabilidade, sendo capaz de calcular as medidas de dispersão para variáveis quantitativas e de interpretar os valores dessas medidas em diferentes situações do cotidiano.

**Resumo**

Algumas medidas utilizadas para identificar o grau de dispersão entre os elementos de um conjunto são amplitude total, variância, desvio padrão e coeficiente de variação, tanto para dados não agrupados quanto para dados agrupados com e sem intervalo de classes. A amplitude total é definida pela diferença entre o maior e o menor valor analisado em uma variável. A variância e o desvio padrão são medidas que levam em consideração a totalidade dos valores da variável em estudo, por isso são índices de variabilidade bastantes estáveis.

Especificamente, a variância é encontrada a partir dos desvios em torno da média aritmética. Já o desvio padrão é a medida de dispersão mais utilizada porque aponta de forma mais precisa a dispersão dos valores em relação à média aritmética. O coeficiente de variação é uma medida relativa de dispersão. É útil quando se deseja comparar em termos relativos o grau de concentração em torno da média de séries distintas.

**Atividades**

1. Em uma semana, a empresa X recebeu as seguintes quantidades de email:

| D | S | T | Q | Q | S | S |
|---|---|---|---|---|---|---|
| 1 | 10 | 14 | 12 | 7 | 6 | 3 |

Diante disso, determine a amplitude total e o desvio padrão.

2. Considere a distribuição a seguir que representa os valores de 20 lançamentos de um dado.

| $x_i$ | $f_i$ |
|---|---|
| 1 | 6 |
| 2 | 3 |
| 3 | 1 |
| 4 | 4 |
| 5 | 2 |
| 6 | 4 |

O valor do desvio padrão será de

a) 1,65.

b) 2,62.

c) 1,28.

d) 1,05.

e) 2,05.

3. Suponha que as notas finais de uma disciplina de programação foram:

3, 7, 4, 5, 1, 8, 4, 6, 5, 6, 2, 4, 6, 9, 8, 4, 5, 5, 6

Separe os dados em dois grupos: "aprovados", com nota maior ou igual que 5; "reprovados", com nota menor que cinco. Quais são os valores aproximados de seus respectivos coeficientes de variação?

a) Aprovados = 27,17%; Reprovados = 35,67%.

b) Aprovados = 17,18%; Reprovados = 11,2%.

c) Aprovados = 63,3%; Reprovados = 31,4%.

d) Aprovados = 22,35%; Reprovados = 28,18%.

e) Aprovados = 8,1%; Reprovados = 15,31%.

4. Considere que em uma rede existam quatro roteadores: A, B, C, D. Foram obtidas mostras do tempo de resposta de cada um dos roteadores, expressas na tabela a seguir. A partir dos tempos, foi possível determinar a média de tempo de resposta dos roteadores.

| ROTEADOR | TEMPOS (EM SEGUNDOS) | MÉDIA (EM SEGUNDOS) |
|---|---|---|
| A | 0.5, 0.25, 0.3, 0.45, 0.4, 0.5 | 0.40 |
| B | 0.1, 0.7, 0.25, 0.2, 0.3, 0.4, 0.35, 0.2 | 0.33 |
| C | 0.2, 0.4, 0.6, 0.7, 0.65, 0.5, 0.3 | 0.48 |
| D | 0.3, 0.15, 0.2, 0.1, 0.25, 0.1 | 0.18 |

Qual é a afirmação falsa?

a) A amplitude total do roteador 'A' é de 0.25.

b) A variância do roteador 'B' é de 0.031, aproximadamente.

c) O desvio padrão do roteador 'C' é de 0.173, aproximadamente.

d) O coeficiente de variação do roteador 'D' é de 40.656%, aproximadamente.

e) A maior variância e desvio padrão são os do roteador 'B'; o menor coeficiente de variação é o do roteador 'D'.

**Comentário das atividades**

Para responder à **atividade 1**, precisamos relembrar os conceitos de amplitude total e de desvio padrão para dados não agrupados. Assim, aplicando a fórmula da amplitude, obtivemos o seguinte resultado:

$$AT = x(máx) - x(mín) = 14 - 1 = 13$$

Para o desvio padrão, precisamos encontrar, inicialmente, a média que é 7,57. Em seguida, construímos a tabela a seguir para identificar a diferença entre o valor e a média.

| $x_i$ | $x_i - \bar{x}$ |
|---|---|
| **1** | **-6,57** |
| **10** | **2,43** |
| **14** | **6,43** |
| **8** | **0,43** |
| **7** | **-0,57** |
| **6** | **-1,57** |
| **3** | **-4,57** |

A partir dos valores da última coluna, calculamos $\sum(x_i - \bar{x})^2$, que é 16,40 e, depois, aplicamos a fórmula do desvio padrão, que pode ser observada a seguir:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{16{,}40}{7}} = \sqrt{2{,}34} = 1{,}53$$

Para a **atividade 2**, precisamos criar a tabela a seguir como auxílio na aplicação da fórmula do desvio padrão para dados agrupados.

| $x_i$ | $f_i$ | $f_i x_i$ | $f_i x_i^2$ |
|---|---|---|---|
| **1** | **6** | **6** | **6** |
| **2** | **3** | **6** | **12** |
| **3** | **1** | **3** | **9** |
| **4** | **4** | **16** | **64** |
| **5** | **2** | **10** | **50** |
| **6** | **4** | **24** | **144** |
| | $\sum = 20$ | $\sum = 71$ | $\sum = 285$ |

Depois, aplicando os somatórios encontrados na fórmula do desvio padrão, obtivemos o seguinte resultado:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \left(\frac{\sum f_i x_i}{n}\right)^2} = \sqrt{\frac{285}{20} - \left(\frac{71}{20}\right)^2} = \sqrt{14{,}25 - 12{,}60} = \sqrt{1{,}65} = 1{,}28$$

Logo a alternativa correta é a **(c)**, que apresenta o valor de 1,28.

Na **atividade 3**, a alternativa correta é a letra **(a)**. O primeiro passo foi separar os conjuntos dos “aprovados” {7, 5, 8, 6, 5, 6, 6, 9, 8, 5, 5, 6} e dos “reprovados” {3, 4, 1, 4, 2, 4, 4}. Na sequência, foi necessário calcular as médias: $\bar{x}_{\text{aprovados}}$=6,33 e $\bar{x}_{\text{reprovados}}$=3,14 e os desvios-padrão: $S_{\text{aprovados}}$=1,72 e $S_{\text{reprovados}}$=1,12. Por fim, calculamos os coeficientes de variância.

aprovados: $CV = \frac{S}{\bar{x}} = \frac{1{,}72}{6{,}33} * 100 = 27{,}17\%$

reprovados: $CV = \frac{S}{\bar{x}} = \frac{1{,}12}{3{,}14} * 100 = 35{,}67\%$

Na **atividade 4**, a resposta solicitada se refere à única alternativa falsa, correspondente à letra **(e)**. A partir dos dados fornecidos, foi possível montar a tabela a seguir.

| | ROTEADOR A | ROTEADOR B | ROTEADOR C | ROTEADOR D |
|---|---|---|---|---|
| **Amplitude total** | **0.25** | **0.009** | **0.096** | **23.936%** |
| **Variância** | **0.6** | **0.031** | **0.177** | **53.867%** |
| **Desvio padrão** | **0.5** | **0.030** | **0.173** | **36.131%** |
| **Coeficiente de variação** | **0.2** | **0.006** | **0.075** | **40.656%** |

De acordo com os dados expressos na tabela anterior, verificamos que a alternativa **(a)** é verdadeira, porque a amplitude total do roteador ‘A’ é 0.25. A alternativa **(b)** também é verdadeira, porque a variância do roteador ‘B’ é, em valores aproximados, 0.031. A alternativa **(c)** é verdadeira, pois o roteador ‘C’ tem como desvio padrão 0.173, aproximadamente. A alternativa **(d)** tem como coeficiente de variação cerca de 40.656%. A alternativa **(e)** é falsa porque, apesar de a maior variância e de o desvio padrão serem realmente do roteador ‘B’, o menor coeficiente de variação é do roteador ‘A’, e não do ‘D’.

**Referências**

CRESPO, Antônio Arnot. **Estatística fácil**. São Paulo: Saraiva, 2002.

MARTINS, Gilberto de A.; DONAIRE, Denis. **Princípios da estatística**: 900 exercícios resolvidos e propostos. São Paulo: Atlas 2004.

NAZARETH, Helenalda. **Curso básico de estatística**. São Paulo: Ática, 2003.