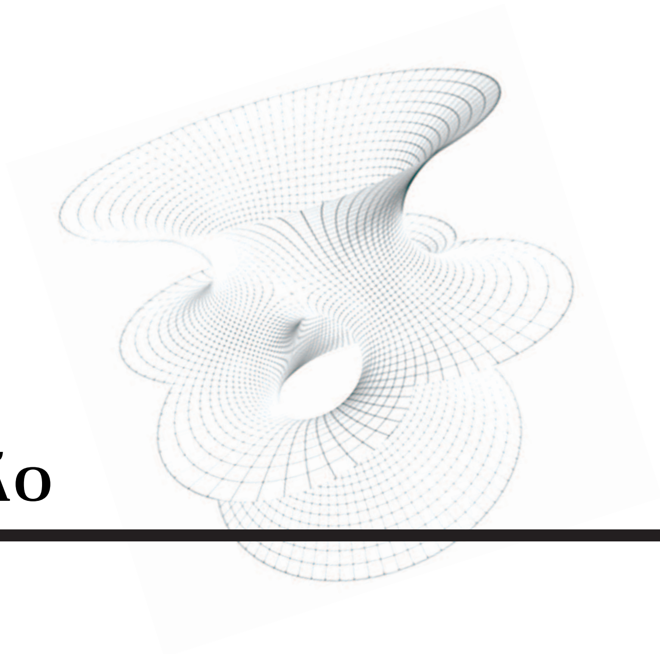
Aula 3



Medidas de Posição

Objetivo

Nesta aula, você estudará as medidas de posição de uma distribuição de dados e aprenderá os seguintes conceitos:

1. média;
2. mediana;
3. moda.

MEDIDAS DE POSIÇÃO OU TENDÊNCIA CENTRAL

A redução dos dados através de tabelas de frequências ou gráficos é um dos meios disponíveis para se ilustrar o comportamento de um conjunto de dados. No entanto, muitas vezes queremos resumir ainda mais esses dados, apresentando um único valor que seja "representativo" do conjunto original. As medidas de posição ou tendência central, como o próprio nome está indicando, são medidas que informam sobre a posição típica dos dados.

Na **Figura 3.1** podemos notar os seguintes fatos: em (a) e (b), as distribuições são idênticas, exceto pelo fato de que a segunda está deslocada à direita. Em (c), podemos ver que há duas classes com a frequência máxima e em (d), há uma grande concentração na cauda inferior e alguns poucos valores na cauda superior. As medidas de posição que apresentaremos a seguir irão captar essas diferenças.

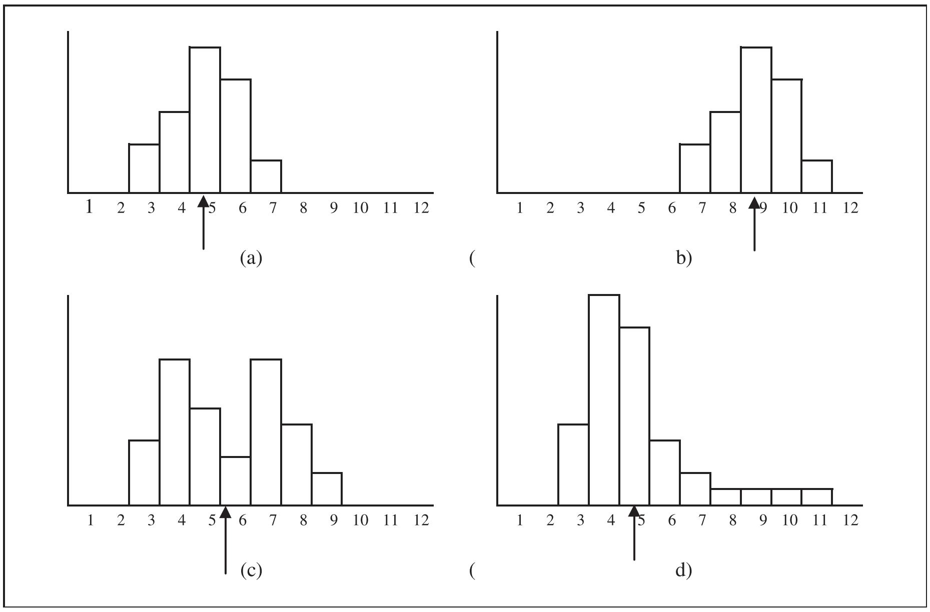


Figura 3.1: Exemplos ilustrativos do conceito de medidas de posição.

MÉDIA ARITMÉTICA SIMPLES

No nosso dia a dia, o conceito de média é bastante comum, quando nos referimos, por exemplo, à altura média dos brasileiros, à temperatura média dos últimos anos etc.

Definição 3.1.

Dado um conjunto de n observações $x_1, x_2, \ldots, x_n$, a **média aritmética simples** é definida como

$$\boxed{\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i}. \qquad (3.1)$$

A notação $\overline{x}$ (lê-se x barra), usada para indicar a média, é bastante comum; em geral, usa-se a mesma letra utilizada para indicar os dados com a barra em cima.

Na definição anterior, fazemos uso do símbolo de somatório, representado pela letra grega sigma maiúscula, Σ. Nesta aula, você ainda aprenderá mais sobre esse símbolo. Por enquanto, entenda como a média aritmética de um conjunto de dados é calculada. A primeira observação é que ela só pode ser calculada para dados quantitativos (não faz sentido somar masculino + feminino!). O seu cálculo é feito somando-se todos os valores e dividindo-se pelo número total de observações.

Consideremos as idades dos funcionários do Departamento de Recursos Humanos, analisadas na aula anterior e apresentadas no ramo e folhas da **Figura 3.2**.

Escala
1 | 0   10

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

Figura 3.2: Idade dos funcionários do Departamento de RH.

A idade média é

$$\begin{aligned}\overline{x} &= \frac{24+25+26+26+29+29+31+35+36+37+38+42+45+51+53}{15} \\ &= \frac{527}{15} = 35,13\end{aligned}$$

Como as idades estão em anos, a idade média também é dada nessa unidade, ou seja, a idade média é 35,13 anos. Em geral, *a média de um conjunto de dados tem a mesma unidade dos dados originais*.

A interpretação física da média aritmética é que ela representa o centro de gravidade da distribuição. Nos quatro histogramas da **Figura 3.1**, ela é o ponto de equilíbrio, indicado pela seta.

Note que o valor da média aritmética é um valor tal que, se substituíssemos todos os dados por ela, isto é, se todas as observações fossem iguais à média aritmética, a soma total seria igual à soma dos dados originais. Então, a média aritmética é uma forma de se distribuir o total observado pelos n elementos, de modo que todos tenham o mesmo valor.

Considere os seguintes dados fictícios referentes aos salários de cinco funcionários de uma firma: $136, 210, 350, 360, 2500$. O total da folha de pagamentos é 3236, havendo um salário bastante alto, discrepante dos demais. A média para esses dados é 647,20. Se todos os cinco funcionários ganhassem esse salário, a folha de pagamentos seria a mesma e todos teriam o mesmo salário.

MODA

No histograma (c) da **Figura 3.1**, duas classes apresentam a mesma frequência máxima. Esse é o conceito de *moda*.

Definição 3.2.

A **moda** de uma distribuição ou conjunto de dados, que representaremos por x^*, é o valor que mais se repete, ou seja, o valor mais frequente.

Podemos ter distribuições amodais (todos os valores ocorrem o mesmo número de vezes), unimodais (uma moda), bimodais (duas modas) etc. Para os dados da **Figura 3.2**, temos as seguintes modas: $x^* = 26$ e $x^* = 29$ anos e, portanto, essa é uma distribuição bimodal. Assim como a média, *a moda sempre tem a mesma unidade dos dados originais.*

MEDIANA

Vamos analisar novamente os seguintes dados referentes aos salários (em R$) de cinco funcionários de uma firma: 136, 210, 350, 360, 2500. Como visto, o salário médio é R$ 647,20. No entanto, esse valor não representa bem nem os salários mais baixos, nem o salário mais alto. Isso acontece porque o salário mais alto é muito diferente dos demais.

Esse exemplo ilustra um fato geral sobre a média aritmética: ela é muito influenciada por *valores discrepantes* (em inglês, *outliers*), isto é, valores muito grandes (ou muito pequenos) que sejam distintos da maior parte dos dados. Nesses casos, é necessário utilizar uma outra medida de posição para representar o conjunto; uma medida possível é a *mediana*.

Definição 3.3.

Seja $x_1, x_2, \ldots, x_n$ um conjunto de n observações e seja $x_{(i)}$, $i = 1, \ldots, n$ o conjunto das observações ordenadas, de modo que $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Então, a **mediana** Q_2 é definida como o valor tal que 50% das observações são menores que ela e 50% são maiores que ela. Para efeito de cálculo, valem as seguintes regras:

$$
\begin{array}{ll}
n \text{ ímpar}: & Q_2 = x_{\left(\frac{n+1}{2}\right)} \\
n \text{ par}: & Q_2 = \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}
\end{array}
\tag{3.2}
$$

Dessa definição, podemos ver que a mediana é o valor central dos dados e para calculá-la é necessário ordenar os dados. Para as idades na **Figura 3.2**, temos que o número total de observações é $n = 15$. Logo, a mediana é o valor central, que deixa sete observações abaixo e sete observações acima. Logo, a mediana é a oitava observação, uma vez que

$$\frac{n+1}{2} = \frac{15+1}{2} = 8.$$

Sendo assim, a idade mediana é $Q_2 = 35$ anos. *A unidade da mediana é a mesma dos dados.*

Exemplo 3.1.

Na aula anterior, analisamos os dados referentes ao número de dependentes dos funcionários do Departamento de Recursos Humanos, apresentados novamente na tabela abaixo.

Nome	Número de dependentes
João da Silva	3
Patrícia Silva	2
Pedro Fernandes	1
Regina Lima	2
Maria Freitas	0
Alfredo Souza	3
Paula Gonçalves	0
Margarete Cunha	0
Ana Freitas	1
Pedro Barbosa	2
Luiz Costa	3
Ricardo Alves	0
André Souza	4
Márcio Rezende	1
Ana Carolina Chaves	0

Vamos calcular as medidas de posição para esses dados. Ordenando-os, temos o seguinte:

$$0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 2 \quad 2 \quad 2 \quad 3 \quad 3 \quad 3 \quad 4$$

A média é

$$\overline{x} = \frac{5 \times 0 + 3 \times 1 + 3 \times 2 + 3 \times 3 + 1 \times 4}{15} = \frac{22}{15} = 1,47.$$

Então, em média temos 1,47 dependentes por funcionário do Departamento de RH. A moda é 0 dependente e a mediana é $(n = 15)$

$$Q_2 = x_{\left(\frac{15+1}{2}\right)} = x_{(8)} = 1 \text{ dependente}$$

Exercício 3.1.

No Exercício 2.1 (Aula 2), você analisou os dados sobre os salários dos funcionários do Departamento de Recursos Humanos, cujos valores (em R$) são os seguintes:

6300	5700	4500	3800	3200	7300	7100	5600
6400	7000	3700	6500	4000	5100	4500	

Calcule a média, a moda e a mediana para esses dados, especificando as respectivas unidades.

Exercício 3.2.

Calcule a nota média, a nota modal e a nota mediana para os dados da **Tabela 3.1**.

Tabela 3.1: Notas de 50 alunos.

2,9	3,7	3,8	4,7	4,9	5,2	5,6	5,8	6,0	6,2
6,3	6,3	6,3	6,5	6,5	6,6	6,8	6,8	6,9	6,9
7,0	7,0	7,1	7,3	7,3	7,4	7,4	7,5	7,5	7,6
7,6	7,7	7,7	7,9	8,1	8,1	8,2	8,2	8,3	8,3
8,4	8,5	8,7	8,7	8,8	8,9	9,0	9,1	9,4	9,7

SOMATÓRIO

A notação de somatório é bastante útil na apresentação de fórmulas, pois ele resume de forma bastante compacta a operação de soma de várias parcelas. Para compreender as propriedades do somatório, basta lembrar as propriedades da adição.

Para desenvolver um somatório, temos que substituir o valor do índice em cada uma das parcelas e em seguida realizar a soma dessas parcelas. Por exemplo:

$$\sum_{i=1}^{5} i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2$$

Em termos mais gerais, temos as seguintes propriedades:

$$\begin{aligned}\sum_{i=1}^{n}(x_i+y_i) &= (x_1+y_1)+(x_2+y_2)+\cdots+(x_n+y_n)= \\ &= (x_1+x_2+\cdots+x_n)+(y_1+y_2+\cdots+y_n)= \\ &= \sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}kx_i &= kx_1+kx_2+\cdots+kx_n= \\ &= k(x_1+x_2+\cdots+x_n)= \\ &= k\sum_{i=1}^{n}x_i\end{aligned}$$

$$\sum_{i=1}^{n}k=k+k+\cdots+k=nk$$

É importante salientar algumas diferenças:

$$\sum_{i=1}^{n}x_i^2 \neq \left(\sum_{i=1}^{n}x_i\right)^2$$

uma vez que

$$\sum_{i=1}^{n}x_i^2=x_1^2+x_2^2+\cdots+x_n^2$$

e

$$\left(\sum_{i=1}^{n}x_i\right)^2=(x_1+x_2+\cdots+x_n)^2$$

Temos também que

$$\sum_{i=1}^{n}x_iy_i \neq \left(\sum_{i=1}^{n}x_i\right)\left(\sum_{i=1}^{n}y_i\right)$$

uma vez que

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

e

$$\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right) = (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n)$$

à medida do necessário iremos apresentando mais propriedades do somatório.

Exercício 3.3.

Calcule as seguintes quantidades para os dados abaixo:

$$\sum_{i=1}^{6} x_i \qquad \sum_{i=1}^{6} f_i \qquad \sum_{i=1}^{6} f_i x_i \qquad \sum_{i=1}^{6} f_i x_i^2$$

i	1	2	3	4	5	6
f_i	3	5	9	10	2	1
x_i	10	11	15	19	21	26

MÉDIA ARITMÉTICA PONDERADA

Vimos que a média aritmética equivale a dividir o "todo" (soma dos valores) em partes iguais, ou seja, estamos supondo que os números que queremos sintetizar têm o mesmo grau de importância. Entretanto, há algumas situações em que não é razoável atribuir a mesma importância a todos os dados.

Por exemplo, o Índice Nacional de Preços ao Consumidor (INPC) é calculado com uma média dos Índices de Preço ao Consumidor (IPC) de diversas regiões metropolitanas do Brasil, mas a importância dessas regiões é diferente. Uma das variáveis que as diferencia é a população residente. Nesse tipo de situação, em vez de se usar a média aritmética simples, usa-se a *média aritmética ponderada*, que será representada por $\overline{x}_p$.

Definição 3.4.

A **média aritmética ponderada** de números $x_1, x_2, \dots, x_n$ com pesos $\rho_1, \rho_2, \dots, \rho_n$ é definida como

$$\bar{x}_p = \frac{\rho_1 x_1 + \rho_2 x_2 + \cdots + \rho_n x_n}{\rho_1 + \rho_2 + \ldots + \rho_n} = \frac{\sum_{i=1}^{n} \rho_i x_i}{\sum_{i=1}^{n} \rho_i}. \qquad (3.3)$$

Se definimos

$$\omega_i = \frac{\rho_i}{\sum_{j=1}^{n} \rho_j}, \qquad (3.4)$$

então, a média aritmética ponderada pode ser reescrita como

$$\bar{x}_p = \sum_{i=1}^{n} \omega_i x_i, \qquad (3.5)$$

em que $\sum_{i=1}^{n} \omega_i = 1$.

Note que a média aritmética simples é um caso particular da média aritmética ponderada, onde todas as observações têm o mesmo peso $\omega_i = \frac{1}{n}$.

Para a construção do Índice Nacional de Preços ao Consumidor – INPC, o peso de cada índice regional é definido pela população residente urbana, conforme dados da **Tabela 3.2**. Os pesos em porcentagem apresentados representam a participação da população residente urbana da região metropolitana no total da população residente urbana das 11 regiões metropolitanas pesquisadas. O índice geral é dado pela média ponderada é

$$\begin{aligned} \text{INPC}_{03/06} &= 0,0306 \times 0,75 + 0,0915 \times 0,64 + 0,0623 \times 0,55 + \\ &\quad 0,0919 \times 0,52 + 0,0749 \times 0,50 + 0,0425 \times 0,48 + \\ &\quad 0,0378 \times 0,48 + 0,0385 \times 0,44 + 0,3626 \times 0,37 + \\ &\quad 0,0334 \times 0,37 + 0,1340 \times 0,18 = 0,427137 \end{aligned}$$

Tabela 3.2: Estrutura básica de ponderação regional para cálculo do INPC - Março 2006

Área Geográfica	**Peso (%)**	**IPC - Mar/06**
Brasília	3,06	0,75
Belo Horizonte	9,15	0,64
Salvador	6,23	0,55
Porto Alegre	9,19	0,52
Curitiba	7,49	0,50
Recife	4,25	0,48
Goiânia	3,78	0,48
Belém	3,85	0,44
São Paulo	36,26	0,37
Fortaleza	3,34	0,37
Rio de Janeiro	13,40	0,18
INPC - Geral		0,42

Fonte: IBGE

Exercício 3.4.

Segundo o critério de avaliação adotado pelo Departamento de Estatística, cada aluno será submetido a duas provas, a primeira tendo peso 2 e a segunda tendo peso 3. Para ser aprovado sem precisar fazer prova final, a média nas duas provas tem que ser, no mínimo, 6.

Se um aluno tirar 5,5 na primeira prova, quanto deverá tirar na segunda prova para não precisar fazer prova final? E se as provas tivessem o mesmo peso?

PROPRIEDADES DAS MEDIDAS DE POSIÇÃO

Da interpretação física de média como centro de gravidade da distribuição, fica claro que a média é sempre um valor situado entre os valores mínimo e máximo dos dados. O mesmo resultado vale para a mediana e a moda, o que é imediato a partir das respectivas definições. Resumindo, temos:

Propriedade 1

$$\begin{aligned} x_{\min} &\leq \bar{x} \leq x_{\max} \\ x_{\min} &\leq Q_2 \leq x_{\max} \\ x_{\min} &\leq x^* \leq x_{\max} \end{aligned} \tag{3.6}$$

Vamos apresentar as outras duas propriedades através do seguinte exemplo:

Em uma turma de Estatística, os resultados de uma prova ficaram abaixo do que a professora esperava. Como todos os alunos vinham participando ativamente de todas as atividades, mostrando um interesse especial pela matéria, a professora resolveu dar 1 ponto na prova para todos os alunos. Além disso, ela deu os resultados com as notas variando de 0 a 10, mas a Secretaria da Faculdade exige que as notas sejam dadas em uma escala de 0 a 100. Sendo assim, a professora precisa multiplicar todas as notas por 10. O que acontece com a média, a moda e a mediana depois dessas alterações?

Vamos ver isso com um conjunto de cinco notas: $5,4,2,3,4$.

As notas ordenadas são $2,3,4,4,5$ e temos as seguintes medidas de posição:

$$\begin{aligned} \overline{x} &= \frac{5+4+2+3+4}{5} = \frac{18}{5} = 3,6 \\ Q_2 &= x^* = 4 \end{aligned}$$

Somando 1 ponto, as notas passam a ser $3,4,5,5,6$ com as seguintes medidas de posição:

$$\begin{aligned} \overline{y} &= \frac{3+4+5+5+6}{5} = \frac{23}{5} = 4,6 = 3,6+1 \\ Q_{2,y} &= y^* = 5 = 4+1 \end{aligned}$$

Ao somar 1 ponto em todas as notas, o conjunto de notas sofre uma translação, o que faz com que o seu centro também fique deslocado de 1 ponto. Sendo assim, todas as três medidas de posição ficam somadas de 1 ponto.

Multiplicando as novas notas por 10, obtemos 30, 40, 50, 50, 60 e

$$\begin{aligned} \overline{z} &= \frac{30+40+50+50+60}{5} = \frac{230}{5} = 46,0 = 4,6 \times 10 \\ Q_{2,z} &= z^* = 50 = 5 \times 10, \end{aligned}$$

ou seja, todas as medidas de posição ficam multiplicadas por 10.

Esse exemplo ilustra as seguintes propriedades:

Propriedade 2

Somando-se um mesmo valor a cada observação x_i, obtemos um novo conjunto de dados $y_i = x_i + k$ para o qual temos as seguintes medidas de posição:

$$y_i = x_i + k \Rightarrow \begin{cases} \overline{y} = \overline{x} + k \\ Q_{2,y} = Q_{2,x} + k \\ y^* = x^* + k \end{cases} \tag{3.7}$$

Propriedade 3

Multiplicando cada observação x_i por uma mesma constante não nula k, obtemos um novo conjunto de dados $y_i = kx_i$ para o qual temos as seguintes medidas de posição:

$$y_i = kx_i \Rightarrow \begin{cases} \overline{y} = k\overline{x} \\ Q_{2,y} = kQ_{2,x} \\ y^* = kx^* \end{cases} \tag{3.8}$$

Exercício 3.5.

A relação entre as escalas Celsius e Fahrenheit é a seguinte:

$$C = \frac{5}{9}(F - 32)$$

Se a temperatura média em determinada localidade é de $45°F$, qual é a temperatura média em graus Celsius?

Exercício 3.6.

Em uma certa pesquisa, foram levantados dados sobre o lucro líquido de uma amostra de grandes empresas, em reais, obtendo-se a média de R$ 1 035 420,00. Na divulgação dos resultados, os valores devem ser apresentados em milhares de reais. Qual é o valor a ser divulgado para o lucro médio?

MEDIDAS DE POSIÇÃO PARA DISTRIBUIÇÕES DE FREQUÊNCIAS AGRUPADAS

Considere a distribuição de frequências do salário dos funcionários do Departamento de Recursos Humanos reproduzida na **Tabela 3.3**.

Tabela 3.3: Distribuição da renda dos funcionários do Departamento de RH

Classe de renda	Ponto médio	Frequência simples		Frequência acumulada	
		absoluta	**relativa %**	**absoluta**	**relativa %**
[3200,4021)	3610,5	4	26,67	4	26,67
[4021,4842)	4431,5	2	13,33	6	40,00
[4842,5663)	5252,5	2	13,33	8	53,33
[5663,6484)	6073,5	3	20,00	11	73,33
[6484,7305)	6894,5	4	26,67	15	100,00
Total		15	100,00		

Essa tabela foi construída a partir dos dados da **Tabela 2.2**, analisada na aula anterior. Imagine, agora, que não dispuséssemos daqueles dados e só nos fosse fornecida a **Tabela 3.3**. Como poderíamos calcular a média, a moda e a mediana? Isso é o que você aprenderá nessa parte final da aula.

MÉDIA ARITMÉTICA SIMPLES

Quando agrupamos os dados em uma distribuição de frequências, estamos perdendo informação, uma vez que não apresentamos os valores individuais.

Informar apenas que existem quatro valores na classe $3200 \vdash 4021$ nos obriga a escolher um valor típico, representante de tal classe. Esse valor será sempre o *ponto médio* da classe.

Então, a informação anterior é interpretada como a existência de quatro valores iguais a 3610,5; que é o ponto médio dessa classe. Essa é a interpretação básica da tabela de frequências: *todos os valores de uma classe são considerados iguais ao ponto médio da classe*. O ponto médio da classe, por sua vez, é calculado como a média dos limites de classe. Veja a coluna criada com esses valores na **Tabela 3.3**.

A interpretação da tabela de frequências nos diz que há quatro observações iguais a 3610,5; duas observações iguais a 4431,5; duas iguais a 5252,5; três iguais a 6073,5 e quatro iguais a 6894,5. Então, esses dados podem ser vistos como o seguinte conjunto de observações:

$$\left.\begin{array}{r} 3610,5 \\ 3610,5 \\ 3610,5 \\ 3610,5 \end{array}\right\} \text{quatro ocorrências do 3610,5} \tag{3.9}$$

$$\left.\begin{array}{r} 4431,5 \\ 4431,5 \end{array}\right\} \text{duas ocorrências do 4431,5}$$

$$\left.\begin{array}{r} 5252,5 \\ 5252,5 \end{array}\right\} \text{duas ocorrências do 5252,5}$$

$$\left.\begin{array}{r} 6073,5 \\ 6073,5 \\ 6073,5 \end{array}\right\} \text{três ocorrências do 6073,5}$$

$$\left.\begin{array}{r} 6894,5 \\ 6894,5 \\ 6894,5 \\ 6894,5 \end{array}\right\} \text{quatro ocorrências do 6894,5}$$

Para calcular a média desse novo conjunto de dados, temos que fazer:

$$\begin{aligned}
\overline{x} &= \frac{4 \times 3610,5 + 2 \times 4431,5 + 2 \times 5252,5 + 3 \times 6073,5 + 4 \times 6894,5}{15} = \\
&= \frac{4}{15} \times 3610,5 + \frac{2}{15} \times 4431,5 + \frac{2}{15} \times 5252,5 + \frac{3}{15} \times 6073,5 + \frac{4}{15} \times 6894,5 = \\
&= 0,2667 \times 3610,5 + 0,1333 \times 4431,5 + 0,1333 \times 5252,5 + 0,20 \times 6073,5 + \\
&\qquad + 0,2667 \times 6894,5 = 5307,2333
\end{aligned}$$

Note, na penúltima linha da equação anterior, que os pontos médios de cada classe são multiplicados pela frequência relativa da classe. Então, a média dos dados agrupados em classes é uma *média ponderada dos pontos médios*, onde os pesos são definidos pelas frequências das classes.

Representando o ponto médio da classe por x_i e por f_i a frequência relativa (não multiplicada por 100), temos que

$$\overline{x} = \sum_{i=1}^{k} f_i x_i \tag{3.10}$$

Os pesos (frequências) aparecem exatamente para compensar o fato de que as classes têm números diferentes de observações.

MODA

Embora existam métodos geométricos para se calcular a moda de dados agrupados, tais métodos não são muito utilizados na prática. Sendo assim, estimaremos a moda de uma distribuição de frequências agrupadas pelo ponto médio da *classe modal*, que é a classe de maior frequência.

No exemplo anterior, temos uma distribuição bimodal com $x^* = 3610,5$ e $x^* = 6894,5$.

MEDIANA

Como já visto, a mediana é o valor que deixa 50% das observações acima e 50% abaixo dela. Estando os dados agrupados em classes, existe um método geométrico que produz uma estimativa da mediana. As ideias subjacentes a esse método são que a mediana divide ao meio o conjunto de dados (ou seja, a definição de mediana) e que, no histograma da distribuição, as áreas dos retângulos são proporcionais às frequências relativas.

Considere o histograma da **Figura 3.3**, referente aos salários dos funcionários do Departamento de Recursos Humanos. Nas duas primeiras classes, temos 40% das observações e, nas três primeiras classes, temos $53,33\%$. Logo, a mediana é algum ponto da *classe mediana* $4842 \vdash 5663$ e, abaixo desse ponto, temos que ter 50% da distribuição, ou seja, as áreas dos dois primeiros retângulos mais a área do retângulo hachurado representam 50% da frequência.

Então, para identificar a mediana, devemos notar que na classe mediana ficam faltando $50\% - 40\% = 10\%$ da distribuição para completar 50%. Então, a área A_1 do retângulo hachurado deve ser igual a 10%, enquanto que o retângulo da classe mediana tem área $A_m = 13,33\%$. Usando a fórmula que dá a área de um retângulo, obtém-se:

$$\begin{aligned} A_1 &= 0,10 = (Q_2 - 4842) \times h \\ A_m &= 0,1333 = (5663 - 4842) \times h \end{aligned}$$

em que h é a altura comum dos dois retângulos. Dividindo as duas igualdades, termo a termo, obtém-se a seguinte regra de proporcionalidade:

$$\frac{0,10}{0,1333} = \frac{Q_2 - 4842}{821} \Rightarrow Q_2 = 5457,904$$

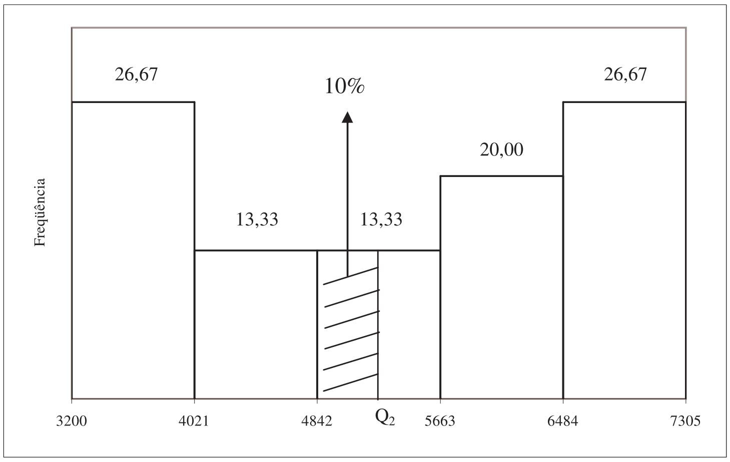


Figura 3.3: Cálculo da mediana dos salários dos funcionários de RH.

Exemplo 3.2.

Para fixar as ideias, vamos calcular a média e a mediana da seguinte distribuição:

Classes	**Frequência simples**		**Frequência acumulada**	
	Absoluta	**Relativa %**	**Absoluta**	**Relativa %**
$0 \vdash 5$	5	6,25	5	6,25
$5 \vdash 10$	15	18,75	20	25,00
$10 \vdash 15$	22	27,50	42	52,50
$15 \vdash 20$	18	22,50	60	75,00
$20 \vdash 25$	12	15,00	72	90,00
$25 \vdash 30$	8	10,00	80	100,00
Total	80	100,00		

Os pontos médios das classes são

$$\frac{0+5}{2} = 2,5 \qquad \frac{5+10}{2} = 7,5 \qquad \cdots \qquad \frac{25+30}{2} = 27,5$$

e a média é calculada como

$$\begin{aligned} \bar{x} &= 0,0625 \times 2,5 + 0,1875 \times 7,5 + 0,2750 \times 12,5 + 0,2250 \times 17,5 + \\ &\quad + 0,15 \times 22,5 + 0,10 \times 27,5 = 15,0625 \end{aligned}$$

Note que é preferível trabalhar com as frequências relativas em forma decimal, pois, se trabalhássemos com as frequências relativas em forma percentual, teríamos que dividir o resultado por 100. Lembre-se de que a média tem de que estar entre o valor mínimo 0 e o valor máximo 30.

Da coluna de frequências relativas acumuladas, vemos que a mediana está na terceira classe $10 \vdash 15$. Nas duas primeiras classes, temos 25% dos dados; assim, está faltando 25% para completar 50%. Veja a **Figura 3.4**.

A regra de três resultante é

$$\frac{Q_2 - 10}{25} = \frac{15 - 10}{27,5} \Rightarrow Q_2 = 14,545$$

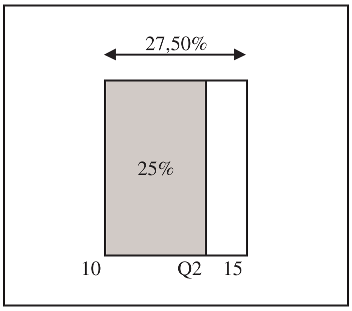


Figura 3.4: Cálculo da mediana do exemplo 3.2.

Exercício 3.7.

Calcule a média e a mediana da seguinte distribuição:

Classes	**Frequência**
$4 \vdash 6$	10
$6 \vdash 8$	12
$8 \vdash 10$	18
$10 \vdash 12$	6
$12 \vdash 14$	4
Total	50

Resumo

Nesta aula, você estudou as principais medidas de posição ou de tendência central, que ilustram a posição típica dos dados. Seja $x_1, x_2, \ldots, x_n$ o nosso conjunto de dados.

Média aritmética simples – é o valor dado por

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

cuja interpretação geométrica corresponde ao centro de gravidade da distribuição.

Moda – x^* é o valor que mais se repete.

Mediana – considerando os dados ordenados $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, a mediana Q_2 é o valor central, ou seja, a mediana é o valor tal que metade das observações é menor que ela:

$$Q_2 = x_{\left(\frac{n+1}{2}\right)} \qquad \text{se } n \text{ é ímpar}$$

$$Q_2 = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \qquad \text{se } n \text{ é par}$$

Média aritmética ponderada – se as observações têm pesos $\omega_1, \omega_2, \ldots, \omega_n$ tais que $\sum\limits_{i=1}^{n} \omega_i = 1$, a média ponderada é

$$\overline{x}_p = \omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n = \sum_{i=1}^{n} \omega_i x_i$$

Média de dados agrupados em classes – é a média ponderada dos pontos médios x_i das classes, em que os pesos são as frequências relativas f_i:

$$\overline{x} = \sum_i f_i x_i$$

Mediana de dados agrupados – é calculada pela proporcionalidade direta de áreas no histograma da distribuição.

Média, mediana e moda são medidas na mesma unidade dos dados e satisfazem as seguintes propriedades:

$$\begin{aligned} x_{\min} &\leq \overline{x} \leq x_{\max} \\ x_{\min} &\leq Q_2 \leq x_{\max} \\ x_{\min} &\leq x^* \leq x_{\max} \end{aligned}$$

$$y_i = k_1 x_i + k_2 \Rightarrow \begin{cases} \overline{y} = k_1 \overline{x} + k_2 \\ Q_{2,y} = k_1 Q_{2,x} + k_2 \\ y^* = k_1 x^* + k_2 \end{cases}$$

Exercício 3.8.

Quatro amigos trabalham em um supermercado em tempo parcial com os seguintes salários horários:

Pedro:	R$ 3,50	João:	R$ 2,60
Marcos:	R$ 3,80	Luiz:	R$ 2,20

Se Pedro trabalha 10 horas por semana, João 12 horas, Marcos 15 horas e Luiz 8 horas, qual é o salário horário médio desses quatro amigos?

Exercício 3.9.

Na UFF, o coeficiente de rendimento (CR) semestral dos alunos é calculado como uma média das notas finais nas disciplinas cursadas, levando em conta a carga horária (ou crédito) das disciplinas, de modo que disciplinas com maior carga horária têm maior peso no CR.

Suponha que um aluno tenha cursado cinco disciplinas em um semestre, obtendo médias finais de $7,5$; $6,1$; $8,3$; $6,5$; $7,5$. As três primeiras disciplinas tinham carga horária de 4 horas semanais, a quarta, carga horária de 6 horas e a última, duas horas semanais. Calcule o CR do aluno nesse semestre.

Exercício 3.10.

Em uma pesquisa sobre atividades de lazer realizada com uma amostra de 20 alunos de um campus universitário, perguntou-se o número de horas que os alunos gastaram "navegando" na internet na semana anterior. Os resultados obtidos foram os seguintes:

15	24	18	8	10	12	15	14	12	10
18	12	6	20	18	16	10	12	15	9

Calcule a média, a moda e a mediana desses dados, especificando as respectivas unidades.

Exercício 3.11.

No final do ano 2005, o dono de um pequeno escritório de administração deu a seus oito funcionários uma gratificação de 250 reais, paga junto com o salário de dezembro. Se em novembro o salário médio desses funcionários era de 920 reais, qual o salário médio em dezembro? Que propriedades você utilizou para chegar a esse resultado?

Exercício 3.12.

No mês de dissídio de determinada categoria trabalhista, os funcionários de uma empresa tiveram reajuste salarial de $8,9\%$. Se no mês anterior ao dissídio o salário médio desses funcionários era de 580 reais, qual o valor do salário médio depois do reajuste? Que propriedades você utilizou para chegar a esse resultado?

Exercício 3.13.

O número médio de empregados das empresas industriais do setor de fabricação de bebidas em determinado momento era de 117 empregados, enquanto o número mediano era de 27. Dê uma explicação para a diferença entre essas medidas de tendência central.

Exercício 3.14.

Na tabela a seguir, temos o número de empresas por faixa de pessoal ocupado (PO) do setor de fabricação de bebidas em determinado momento. Calcule a média e a mediana dessa distribuição, especificando as respectivas unidades.

Classe de PO	Número de empresas
$[10,30)$	489
$[30,100)$	269
$[100,500)$	117
$[500,1000)$	15
$[1000,2000)$	9
$[2000,4000)$	7

SOLUÇÃO DOS EXERCÍCIOS

Exercício 3.1.

Temos 15 funcionários. Os dados ordenados são os seguintes: 3200, 3780, 3800, 4000, 4500, 4500, 5100, 5600, 5700, 6300, 6400, 6500, 7000, 7100, 7300. A média é

$$\overline{x} = \frac{3200 + 3780 + \cdots + 7300}{15} = \frac{80700}{15} = 5380$$

A moda é

$$x^* = 4500$$

e a mediana é a observação de posição $\frac{15+1}{2} = 8$, ou seja,

$$Q_2 = x_{(8)} = 5600$$

Todas essas medidas estão em R$.

Exercício 3.2.

Note que os dados já estão ordenados; caso não estivessem, uma boa opção para ajudar na solução do exercício seria construir o diagrama de ramos e folhas. Temos 50 notas. Logo,

$$\overline{x} = \frac{2,9+3,7+\cdots+9,7}{50} = \frac{357,1}{50} = 7,142$$

A nota modal é $x^* = 6,3$, que aparece 3 vezes. Como o número de observações é par ($n = 50$), a mediana é a média das 2 observações centrais, cujas posições são

$\frac{50}{2}$ e $\frac{50}{2}+1$, ou seja, a mediana é a média da 25^a e da 26^a

Observações:

$$Q_2 = \frac{7,3+7,4}{2} = 7,35$$

Exercício 3.3.

Temos o seguinte:

$$\sum_{i=1}^{6} x_i = x_1+x_2+x_3+x_4+x_5+x_6 = 10+11+15+19+21+26 = 102$$

$$\sum_{i=1}^{6} f_i = f_1+f_2+f_3+f_4+f_5+f_6 = 3+5+9+10+2+1 = 30$$

$$\begin{aligned}\sum_{i=1}^{6} f_i x_i &= f_1x_1 + f_2x_2 + f_3x_3 + f_4x_4 + f_5x_5 + f_6x_6 = \\ &= 3\times 10 + 5\times 11 + 9\times 15 + 10\times 19 + 2\times 21 + 1\times 26 = 478\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{6} f_i x_i^2 &= f_1x_1^2 + f_2x_2^2 + f_3x_3^2 + f_4x_4^2 + f_5x_5^2 + f_6x_6^2 = \\ &= 3\times 10^2 + 5\times 11^2 + 9\times 15^2 + 10\times 19^2 + 2\times 21^2 + 1\times 26^2 = 8098\end{aligned}$$

Exercício 3.4.

Vamos denotar por x_1 e x_2 as notas na primeira e na segunda provas. Então, a média final é calculada como

$$\overline{x}_p = \frac{2x_1 + 3x_2}{2+3}$$

Para aprovação direta, sem prova final, temos que ter $\overline{x}_p \geq 6$. Logo,

$$\overline{x}_p \geq 6 \Leftrightarrow$$

$$\frac{2x_1 + 3x_2}{2+3} \geq 6 \Leftrightarrow$$

$$2\times 5,5 + 3x_2 \geq 30 \Leftrightarrow$$

$$3x_2 \geq 19 \Leftrightarrow$$

$$x_2 \geq \frac{19}{3} = 6,3\overline{3}$$

Se fosse média simples, teríamos que ter

$$\overline{x} \geq 6 \Leftrightarrow$$

$$\frac{x_1 + x_2}{2} \geq 6 \Leftrightarrow$$

$$5,5 + x_2 \geq 12 \Leftrightarrow$$

$$x_2 \geq 6,5$$

Exercício 3.5.

A mesma relação que se aplica às temperaturas individuais se aplica também à temperatura média, ou seja, a temperatura média em graus Celsius é

$$\overline{C} = \frac{5}{9}(\overline{F} - 32) = \frac{5}{9}(45 - 32) = 7,22^\circ C$$

Exercício 3.6.

Não é necessário recalcular a média em milhares de reais; basta dividir a média por 1000, ou seja, o lucro médio é de $1035,42$ milhares de reais.

Exercício 3.7.

A distribuição de frequências completa é a seguinte:

Classes	**Ponto**	**Freq. Simples**		**Freq.Acumulada**	
	Médio	**Absoluta**	**Relativa**	**Absoluta**	**Relativa**
$4 \vdash 6$	5	10	0,20	10	0,20
$6 \vdash 8$	7	12	0,24	22	0,44
$8 \vdash 10$	9	18	0,36	40	0,80
$10 \vdash 12$	11	6	0,12	46	0,92
$12 \vdash 14$	13	4	0,08	50	1,00
Total		50	1,00		

A média é

$$\overline{x} = 5 \times 0,20 + 7 \times 0,24 + 9 \times 0,36 + 11 \times 0,12 + 13 \times 0,08 = 8,28$$

A mediana está na classe $8 \vdash 10$. Abaixo desta classe, temos 44% das observações. Assim, para completar 50% ficam faltando 6%. Veja a **Figura 3.5**.

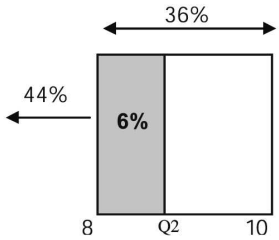


Figura 3.5: Cálculo da mediana do exercício 3.7.

A regra de proporcionalidade é

$$\frac{Q_2 - 8}{6} = \frac{10 - 8}{36} \Rightarrow Q_2 - 8 = \frac{12}{36} \Rightarrow Q_2 = 8,33$$

Exercício 3.8.

Para calcular o salário horário médio, temos que dividir o total dos vencimentos pelo total de horas trabalhadas pelos quatro amigos.

$$\begin{aligned}
\overline{x} &= \frac{10 \times 3,50 + 12 \times 2,6 + 15 \times 3,80 + 8 \times 2,20}{10 + 12 + 15 + 8} \\
&= \frac{10 \times 3,50 + 12 \times 2,6 + 15 \times 3,80 + 8 \times 2,20}{45} \\
&= \frac{10}{45} \times 3,50 + \frac{12}{45} \times 2,6 + \frac{15}{45} \times 3,80 + \frac{8}{45} \times 2,20 \\
&= \frac{140,8}{45} = 3,1289
\end{aligned}$$

Note que o salário médio é uma média ponderada dos salários individuais, com o peso sendo definido pelo número de horas de trabalho.

Exercício 3.9.

A carga horária semanal total é $4 + 4 + 4 + 6 + 2 = 20$. Logo, o CR do aluno é

$$\begin{aligned}
CR &= \frac{4}{20} \times 7,5 + \frac{4}{20} \times 6,1 + \frac{4}{20} \times 8,3 + \frac{6}{20} \times 6,5 + \frac{2}{20} \times 7,5 \\
&= \frac{141,6}{20} = 7,08
\end{aligned}$$

Exercício 3.10.

O diagrama de ramos e folhas é o seguinte:

0	6	8	9													
1	0	0	0	2	2	2	2	4	5	5	5	6	8	8	8
2	0	4													

A média é

$$\overline{x} = \frac{6+8+9+\cdots+20+24}{20} = \frac{274}{20} = 13,7$$

A moda é $x^* = 12$ e a mediana é a média dos valores centrais:

$$Q_2 = \frac{x_{(10)} + x_{(11)}}{3} = \frac{12+14}{2} = 13$$

Todos esses resultados estão medidos em horas por semana.

Exercício 3.11.

Todos os salários ficaram aumentados em 250 reais. Se chamamos de x_i o salário do funcionário i no mês de novembro e de y_i o salário desse mesmo funcionário em dezembro, então

$$y_i = x_i + 250.$$

De acordo com a Propriedade 2, temos que o salário médio em dezembro é

$$\overline{y} = \overline{x} + 250 = 920 + 250 = 1170 \text{ reais.}$$

Exercício 3.12.

Seja x_i o salário do funcionário i no mês anterior ao dissídio. Depois do aumento, seu salário passa a ser

$$y_i = x_i + 0,089x_i = 1,089x_i.$$

Logo, todos os salários ficam multiplicados por 1,089 e, pela Propriedade 3, a média também fica multiplicada por este valor, ou seja, depois do dissídio o salário médio passa a ser

$$\overline{y} = 1,089\overline{x} = 1,089 \times 580 = 631,62 \text{ reais.}$$

Exercício 3.13.

A diferença se deve à existência de grandes empresas no setor de bebidas, com muitos empregados. Como vimos, a média é bastante influenciada pelos valores discrepantes.

Exercício 3.14.

Completando a tabela, obtemos

Classe de PO	Ponto médio	Frequência Simples Absoluta	Frequência Simples Relativa (%)	Frequência Acumulada Absoluta	Frequência Acumulada Relativa (%)
$[10,30)$	20	489	53,9735	489	53,9735
$[30,100)$	65	269	29,6909	758	83,6645
$[100,500)$	300	117	12,9139	875	96,5784
$[500,1000)$	750	15	1,6556	890	98,2340
$[1000,2000)$	1500	9	0,9934	899	99,2274
$[2000,4000)$	3000	7	0,7726	906	100,0000
Total		906	100,0000		

Como as frequências relativas estão em forma percentual, temos que dividir o resultado por 100, ou seja:

$$\begin{aligned} \overline{x} &= (20\times 53,9735+65\times 29,6909+300\times 12,9139+\\ &\quad 750\times 1,6556+1500\times 0,9934+3000\times 0,7726)/100\\ &= 119,3322 \text{ empregados}\end{aligned}$$

A mediana está na classe $10 \vdash 30$. A frequência abaixo desta classe é nula. Logo, a regra de três é

$$\frac{Q_2-10}{50}=\frac{30-10}{53,9735}\Rightarrow Q_2-10=\frac{1000}{53,9735}\Rightarrow$$

$$Q_2 = 28,528 \text{ empregados}$$

Note a diferença da média para a mediana, resultado da presença de empresas com muitos empregados – muitas empresas têm poucos empregados, mas poucas empresas têm muitos empregados, o que “puxa” a média para cima.