

1. Contou-se o número de erros de impressão da primeira página de um jornal durante 50 dias, obtendo-se os resultados abaixo:

8 11 8 12 14 13 11 14 14 15
 6 10 14 19 6 12 7 5 8 8
 10 16 10 12 12 8 11 6 7 12
 7 10 14 5 12 7 9 12 11 9
 14 8 14 8 12 10 12 22 7 15

- (a) Agrupar os elementos em classes.
 (b) Construir o histograma e o polígono de frequência.
 (c) Com os dados agrupados em classes, calcular a média aritmética, a mediana e a moda.
 (d) Determine Q_1 ;
 (e) Determine D_2
2. Quer se estudar o número de erros de impressão de um livro. Para isso escolheu-se uma amostra de 50 páginas, encontrando-se o número de erros por página da tabela abaixo.

Erros	Frequência
0	25
1	20
2	3
3	1
4	1

- (a) Qual o número médio de erros por página?
 (b) E o número mediano?
 (c) Qual é o desvio padrão?
 (d) Construa o histograma e o polígono de frequência.
 (e) Se o livro tem 500 páginas, qual o número total de erros esperados no livro?
3. As taxas de juros recebidas por 10 ações durante um certo período foram (medidas em porcentagem) 2.59; 2.64,2.60,2.62,2.57,2.55,2.61,2.50,2.63,2.64. Calcule a média, a mediana e o desvio padrão.
4. Para facilitar um projeto de ampliação da rede de esgoto de uma certa região de uma cidade, as autoridades tomaram uma amostra de tamanho 50 dos 270 quarteirões que compõem a região, e foram encontrados os seguintes números de casas por quarteirão:

2 2 3 10 13 14 15 15 16 16
 18 18 20 21 22 22 23 24 25 25
 26 27 29 29 30 32 36 42 44 45
 45 46 48 52 58 59 61 61 61 65
 66 66 68 75 78 80 89 90 92 97

- (a) Use cinco intervalos e construa um histograma e o polígono de frequência.
 (b) Determine uma medida de posição central e uma medida de dispersão.
5. Numa pesquisa realizada com 100 famílias, levantaram-se as seguintes informações:

Número de filhos	0	1	2	3	4	5	mais que 5
Frequência de famílias	17	20	28	19	7	4	5

- (a) Qual a mediana do número de filhos?
 (b) E a moda?
 (c) Que problemas você encontraria para calcular a média? Faça alguma suposição e encontre-a.

6. Mostre que:

$$(a) \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$(b) \sum_{i=1}^k n_i(x_i - \bar{x})^2 = \sum_{i=1}^k n_i x_i^2 - n\bar{x}^2$$

7. Os dados abaixo representam as vendas semanais, em classes de salários mínimos, de vendedores de gêneros alimentícios:

Vendas semanais	Nº de vendedores
30 35	2
35 40	10
40 45	18
45 50	50
50 55	70
55 60	30
60 65	18
65 70	2

- (a) Faça o histograma e o polígono de frequência das observações.
 (b) Calcule a média da amostra.
 (c) Calcule o desvio padrão da amostra.
 (d) Qual a porcentagem das observações compreendidas entre $\bar{x} - 2s$ e $\bar{x} + 2s$?
 (e) Calcule a mediana.
8. O número de divórcios na cidade, de acordo com a duração do casamento, está representado na tabela abaixo.

Anos de casamento	Nº de divórcios
0 6	2.800
6 12	1.400
12 18	600
18 24	150
24 30	50

- (a) Qual a duração média dos casamentos? E a mediana?
 (b) Encontre a variância e o desvio padrão da duração dos casamentos.
 (c) Construa o histograma da distribuição e o polígono de frequência.
 (d) Encontre o 1º e o 9º decis.
 (e) Encontre o 2º e o 3º quartis.
9. O Departamento de Pessoal de uma certa firma fez um levantamento dos salários de 120 funcionários do setor administrativo, obtendo os resultados (em salários mínimos) da tabela abaixo:

Faixa salarial	Frequência relativa
0 2	0.25
2 4	0.40
4 6	0.20
6 10	0.15

- (a) Esboce o histograma correspondente.
- (b) Calcule a média, a variância e o desvio padrão.
- (c) Calcule o 1º quartil e a mediana.
- (d) Se for concedido um aumento de 100% para todos os funcionários, haverá alteração na média? E na variância? Justifique sua resposta.
- (e) Se for concedido abono de dois salários mínimos para todos os 120 funcionários, haverá alteração na média? E na variância? E na mediana? Justifique sua resposta.

10. O que acontece com a mediana, a média e o desvio padrão de uma série de dados quando:

- (a) cada observação é multiplicada por 2?
- (b) soma-se 10 a cada observação?
- (c) subtrai-se a média geral \bar{x} de cada observação?
- (d) de cada observação subtrai-se \bar{x} e divide-se pelo desvio padrão?

11. Na companhia A, a média dos salários é 10000 unidades e o 3º quartil é 5000.

- (a) Se você se apresentasse como candidato a funcionário nessa firma e se o seu salário fosse escolhido ao acaso entre todos os possíveis salários, o que seria mais provável: ganhar mais ou menos de 5000 unidades?
- (b) Suponha que na companhia B a média dos salários seja 7000 unidades, a variância praticamente zero e o salário também seja escolhido ao acaso. Em qual companhia você se apresentaria para procurar emprego?

12. Estudando-se o consumo diário de leite, verificou-se que, em certa região, 20% das famílias consomem até um litro, 50% consomem entre um e dois litros, 20% consomem entre dois e três litros e o restante consome entre três e cinco litros. Para a variável em estudo:

- (a) Escreva as informações acima na forma de uma tabela de frequências.
- (b) Construa o histograma e o polígono de frequência.
- (c) Calcule a média e a mediana.
- (d) Calcule a variância e o desvio padrão.
- (e) Qual o valor do 1º quartil?

13. A distribuição de frequências do salário anual dos moradores do bairro A que têm alguma forma de rendimento é apresentada na tabela abaixo:

Faixa salarial (x 10 salários mínimos)	Frequência
0 - 2	10000
2 - 4	3900
4 - 6	2000
6 - 8	1100
8 - 10	800
10 - 12	700
12 - 14	2000
Total	20500

- (a) Construa um histograma da distribuição.
- (b) Qual a média e o desvio padrão da variável salário?
- (c) O bairro B apresenta, para a mesma variável, uma média de 7,2 e um desvio padrão de 15,1. Em qual dos bairros a população é mais homogênea quanto à renda?

14. A idade média dos candidatos de um determinado curso de aperfeiçoamento sempre foi baixa, da ordem de 22 anos. Como esse curso foi planejado para atender a todas as idades, decidiu-se fazer uma campanha de divulgação. Para se verificar se a campanha foi ou não eficiente, fez-se um levantamento da idade dos candidatos à última promoção, e os resultados estão na tabela a seguir.

Idade	Frequência	Porcentagem
18 - 20	18	36
20 - 22	12	24
22 - 26	10	20
26 - 30	8	16
30 - 36	2	4
Total	50	100

- (a) Baseando-se nesses resultados, você diria que a campanha produziu algum efeito (isto é, aumentou a idade média)?
- (b) Um outro pesquisador decidiu usar a seguinte regra: se a diferença $\bar{x} - 22$ fosse maior que o valor do $2s/\sqrt{n}$, (s desvio padrão amostral) então a campanha tinha surtido efeito. Qual a conclusão dele, baseada nos dados?
- (c) Faça o histograma da distribuição.
15. Um órgão do governo do estado está interessado em determinar padrões sobre o investimento em educação, por habitante, realizado pelas prefeituras. De um levantamento de dez cidades, foram obtidos os valores (codificados) da tabela abaixo:

Cidade	A	B	C	D	E	F	G	H	I	J
Investimento	20	16	14	8	19	15	14	16	19	18

Nesse caso, será considerado como *investimento básico a média final* das observações, calculadas da seguinte maneira:

1. Obtém-se a média inicial;
2. Eliminam-se do conjunto aquelas observações que forem superiores à média inicial mais duas vezes o desvio padrão, ou inferiores à média inicial menos duas vezes o desvio padrão;
3. Calcula-se a média final com o novo conjunto de observações.

Qual investimento básico que você daria como resposta?

Observação: O procedimento do item 2 tem a finalidade de eliminar do conjunto a cidade cujo investimento é muito diferente dos demais.

16. Tanto a média quanto o desvio padrão podem não ser medidas adequadas para representar um conjunto de dados, pois:
- são afetados, de forma exagerada, por valores extremos;
 - apenas com estes dois valores não temos idéia da simetria ou assimetria da distribuição dos dados.

Para contornar esses fatos, outras medidas têm que ser consideradas, como a mediana, os quartis, decis e percentis.

Se quisermos comparar a variabilidade de dois conjuntos de dados podemos usar o coeficiente de variação, que é definido como a razão entre o desvio padrão, s , e a média amostral e usualmente expresso em porcentagem:

$$CV = \frac{s}{\bar{x}} 100\%.$$

Calcule o coeficiente de variação para as regiões A e B a seguir e comente o resultado.

Região	Média	Desvio Padrão
A	20	4
B	20	6

17. A tabela a seguir mostra os resultados de 25 funcionários em vários exames a que se submeteram. Sabe-se agora que os critérios adotados em cada exame não são comparáveis, por isso decidiu-se usar o *desempenho relativo* em cada exame. Essa medida será obtida do seguinte modo:

I Para cada exame serão calculados a média \bar{x} e o desvio padrão (s).

II A nota X de cada aluno será padronizada do seguinte modo:

$$Z = \frac{X - \bar{x}}{s}$$

Func.	Seção(*)	Administr.	Direito	Redação	Estatíst.	Política	Economia
1	P	8,0	9,0	8,6	9,0	9,0	8,5
2	P	8,0	9,0	7,0	9,0	6,5	8,0
3	P	8,0	9,0	8,0	8,0	9,0	8,5
4	P	6,0	9,0	8,6	8,0	6,0	8,5
5	P	8,0	9,0	8,0	9,0	6,5	9,0
6	P	8,0	9,0	8,5	10,0	6,5	9,5
7	P	8,0	9,0	8,2	8,0	9,0	7,0
8	T	10,0	9,0	7,5	8,0	6,0	8,5
9	T	8,0	9,0	9,4	9,0	10,0	8,0
10	T	10,0	9,0	7,9	8,0	9,0	7,5
11	T	8,0	9,0	8,6	10,0	10,0	8,5
12	T	8,0	9,0	8,3	7,0	6,5	8,0
13	T	6,0	9,0	7,0	7,0	6,0	8,5
14	T	10,0	9,0	8,6	9,0	10,0	7,5
15	V	8,0	9,0	8,6	9,0	10,0	7,0
16	V	8,0	9,0	9,5	7,0	9,0	7,5
17	V	8,0	9,0	6,3	8,0	10,0	7,5
18	V	6,0	9,0	7,6	9,0	6,0	8,5
19	V	6,0	9,0	6,8	4,0	6,0	9,5
20	V	6,0	9,0	7,5	7,0	6,0	8,5
21	V	8,0	9,0	7,7	7,0	6,5	8,0
22	V	6,0	9,0	8,7	8,0	6,0	9,0
23	V	8,0	9,0	7,3	10,0	9,0	7,0
24	V	8,0	9,0	8,5	9,0	6,5	9,0
25	V	8,0	9,0	7,0	9,0	9,0	8,5

(*) (P=departamento pessoal, T=seção técnica e V=seção de vendas)

- Interprete o significado de Z .
- Calcule as notas padronizadas dos funcionários para o exame de Estatística.
- Com os resultados obtido em (b), calcule \bar{z} e o desvio padrão de Z ($dp(Z)$).
- Se alguma das notas padronizadas estiver acima de $2dp(Z)$ ou abaixo de $-2dp(Z)$, esse funcionário deve ser considerado um caso atípico. Existe algum nessa situação?
- O funcionário 1 obteve 9,0 em Direito, em Estatística e em Política. Em que disciplina o seu desempenho relativo foi melhor?

Comparando a média, mediana e a moda

A média é a tendência central mais usada. Ela é fácil de calcular e interpretar. Infelizmente, a média é afetada por valores extremos ou atípicos e, diferente da mediana, pode ser drasticamente desviada por observações que ficam extremamente acima ou abaixo da maioria das observações.

Por exemplo, para os dados 4,5,6,6,7,8, a média e a mediana são 6 e representam uma medida excelente do ponto central do conjunto de dados. Se a última observação fosse 80 em vez de 6, a média seria 18, mas a mediana ainda seria 6. Como a mediana não é afetada pelo valor extremo ela representa melhor as seis observações.

A moda também é pouco afetada por algumas observações atípicas e continua sendo 6 mesmo se o último valor fosse 80. De qualquer maneira, se não há moda, ou se o conjunto de dados for bimodal, seu uso pode ser confuso.

Isso não significa que uma medida é melhor do que a outra. A medida escolhida depende da natureza dos dados ou da maneira com que os dados são usados. Por exemplo, uma loja de equipamentos de camping, fará pouco uso da informação de que o tamanho médio das botas de escaladas vendidas é de 41,3. Muito mais útil, para decisões nos negócios, seria conhecer a moda do tamanho, pois saberia que o tamanho mais vendido é o 39.

18. Sua empresa está introduzindo no mercado um novo computador que promete realizar cálculos estatísticos muito mais rápidos do que os concorrentes. Vinte cálculos diferentes são realizados nos tempos, em segundos, registrados na tabela abaixo. Você não quer deturpar as características de seu produto, mas deseja apresentar os resultados da melhor maneira para sua empresa. Calcule a média, a mediana e a moda. Comente a vantagem de cada uma.

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

19. Defina com suas palavras o quarto decil e o terceiro quartil.
20. Dadas as idades dos 100 maiores executivos das empresas citadas na *Fortune 500*, revelou-se que a média é de 56,2 anos de idade e o desvio padrão 12,7 anos. A renda média desses executivos é 89432 dólares, com $s = 16090$. Qual das variáveis têm a maior variação: idade ou renda?
21. Se a mediana da renda no exercício anterior é 87567:
- Os dados são assimétricos para a direita, para a esquerda ou são normalmente distribuídos?
 - Qual é o coeficiente de Pearson de assimetria?
22. Uma amostra das taxas de juros para empréstimos de 15 anos nas financiadoras em uma cidade do interior resultou em:

7, 1%, 7, 3%, 7%, 6, 9%, 6, 6%, 6, 9%, 6, 5%, 7, 3%, 6, 85

- Calcule a média, a mediana e a moda.
 - Esses dados estão distorcidos para a direita, distorcidos para a esquerda ou estão normalmente distribuídos? Calcule o coeficiente de Pearson como uma medida de assimetria.
 - Calcule e interprete a variância e o desvio padrão.
23. Calcule o coeficiente de Pearson de assimetria para as notas das provas de economia do Professor Pardal:

80 83 87 85 90 86 84 82 88