

1ª Lista de Exercícios - Estatística Descritiva

Exercício 1. 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
Total	$n = 50$

- Qual o número médio de erros por página?
- E o número mediano?
- Qual é o desvio padrão?
- Faça uma representação gráfica para a distribuição.
- Se o livro tem 500 páginas, qual o número total de erros esperado no livro?

Exercício 2. 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

- Determine uma medida de posição central e uma medida de dispersão.
- Use cinco intervalos e construa um histograma.
- Construa o boxplot para esses dados.
- Discuta sobre a simetria, caudas e presença de valores atípicos.

Exercício 3. 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

- Qual a mediana do número de filhos?
- E a moda?
- Que problemas você enfrentaria para calcular a média? Faça alguma suposição e encontre-a.

Exercício 4. 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	2800
6 – 12	1400
12 – 18	600
18 – 24	150
24 – 30	50

- Qual a duração média dos casamentos? E a mediana?
- Encontre a variância e o desvio padrão da duração dos casamentos.
- Construa o histograma da distribuição.
- Encontre o 1º e o 9º decis.
- Qual o intervalo interquantil?

Exercício 5. O que acontece com a mediana, a média e o desvio padrão de uma série de dados quando:

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

Exercício 6. Na companhia A, a média dos salários é 10.000 unidades e o 3º quartil é 5.000.

- 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 que 5.000 unidades?

- (b) Suponha que na companhia B a média dos salários seja 7.000 unidades, a variância praticamente zero e o salário também seja escolhido ao acaso. Em qual companhia você se apresentaria para procurar emprego?

Exercício 7. A média do nível de colesterol de um grupo de jovens recrutas é 205 mg/dl e o desvio padrão é 22 mg/dl. Para um grupo de oficiais, entretanto, a média obtida foi 244 mg/dl e o desvio padrão 45 mg/dl. Compare os dois grupos quanto à homogeneidade da variável nível de colesterol.

Exercício 8. Estudando-se a distribuição das idades dos funcionários de duas repartições públicas, obtiveram-se algumas medidas que estão no quadro abaixo. Esboce o histograma alisado (aquela curva de densidade aproximada do histograma) das duas distribuições, indicando nele as medidas descritas no quadro. Comente as principais diferenças entre os dois histogramas.

Repartição	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão
A	18	27	33	33	39	48	5
B	18	23	32	33	42	48	10

Exercício 9. Decidiu-se investigar a distribuição dos profissionais com nível universitário em duas regiões, A e B. As informações pertinentes foram obtidas e encontram-se no quadro abaixo, expressas em salários mínimos. Esboce a distribuição (histograma alisado) dos salários de cada região, indicando no gráfico as medidas apresentadas no quadro. Faça também uma descrição rápida das principais diferenças observadas nos gráficos.

Região	Média	DP	Mediana	Moda	q_1	q_3	$x_{(1)}$	$x_{(n)}$
A	20.00	4.00	20.32	20.15	17.32	22.68	8.00	32.00
B	20.00	6.00	18.00	17.00	16.00	24.00	14.00	42.00

Exercício 10. Para se estudar o desempenho de duas corretoras de ações, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada, computou-se a porcentagem de lucro apresentada durante um período fixado de tempo. Os dados estão a seguir.

Corretora A		
45	60	54
62	55	70
38	48	64
55	56	55
54	59	48
65	55	60

Corretora B		
57	55	58
50	52	59
59	55	56
61	52	53
57	57	50
55	58	54
59	51	56

- (a) Que tipo de informação revelam esses dados? (Análise esses dados para concluir)

- (b) Para verificar a homogeneidade das duas populações do problema anterior, um estatístico sugeriu que se usasse o quociente $F = \frac{Var(x_A)}{Var(x_B)}$, mas não disse qual decisão tomar baseado nesse valor. Que regra de decisão você adotaria para dizer se são homogêneas ou não? (As notações x_A e x_B representam o vetor de dados das Corretoras A e B, respectivamente)
- (c) Para decidir se o desempenho das duas corretoras são semelhantes ou não, adotou-se o seguinte teste: sejam

$$t = \frac{\bar{x}_A - \bar{x}_B}{S_*^2 \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}, \quad \text{onde} \quad S_*^2 = \frac{(n_A - 1)Var(x_A) + (n_B - 1)Var(x_B)}{n_A + n_B - 2},$$

onde n_A e n_B representam o número de dados dos vetores referentes às Corretoras A e B, respectivamente.

Caso $|t| < 2$, os desempenhos são semelhantes, caso contrário, são diferentes. Qual seria a sua conclusão?