

1ª Lista de PE

1. Dadas as populações abaixo, classifique as suas respectivas variáveis como qualitativas (ordinal ou nominal) ou quantitativas (contínuas ou discretas):
 - a) **População:** Alunos de uma escola.
Variável: Cor dos cabelos.
 - b) **População:** Casais residentes em uma cidade.
Variável: Números de filhos.
 - c) **População:** As jogadas sucessivas de um dado.
Variável: O ponto obtido em cada jogada.
 - d) **População:** Pacientes cancerígenos
Variável: Peso de cada paciente cancerígeno.
 - e) **População:** Estação meteorológica de uma cidade.
Variável: Precipitação pluviométrica, durante um ano.
 - f) **População:** Alunos de uma universidade.
Variável: Conceito obtido.
 - g) **População:** Nascidos vivos numa determinada data.
Variável: Altura.
 - h) **População:** Casais residentes em uma cidade.
Variável: Sexo dos filhos.

2. Um eleição irá ocorrer na próxima semana e, selecionando uma amostra aleatória da população de eleitores, tentaremos prever qual candidato (Republicano ou Democrata) prevalecerá. Qual dos métodos de seleção abaixo você acha que seria melhor para obtermos uma amostra representativa?
 - (a) Eleitores presentes em um jogo de basquete colegial.
 - (b) Eleitores presentes em um luxuoso restaurante metropolitano.
 - (c) Sortear 100 nomes da lista de eleitores com idade para votar.
 - (d) Utilizar as pesquisas de telespectadores de uma TV local.

3. Para estimar o número de fumantes em uma cidade, via pura observação, decidiu-se por escolher um dos locais abaixo para tomar uma amostra representativa:
 - (a) parque aquático;
 - (b) bar aberto;
 - (c) shopping;
 - (d) biblioteca.

4. Use os dados da tabela abaixo e construa a distribuição de frequências das variáveis:

- a) Estado civil.
- b) Região de procedência.
- c) Número de filhos dos empregados casados.
- d) Idade.

Número do funcionário	Estado civil	Grau de instrução	Número de filhos	Salário (x sal. mín.)	Idade (anos)	Região de procedência
1	solteiro	fundamental	—	4,00	26	interior
2	casado	fundamental	1	4,56	32	capital
3	casado	fundamental	2	5,25	36	capital
4	solteiro	médio	—	5,73	20	outra
5	solteiro	fundamental	—	6,26	40	outra
6	casado	fundamental	0	6,66	28	interior
7	solteiro	fundamental	—	6,86	41	interior
8	solteiro	fundamental	—	7,39	43	capital
9	casado	médio	1	7,59	34	capital
10	solteiro	médio	—	7,44	23	outra
11	casado	médio	2	8,12	33	interior
12	solteiro	fundamental	—	8,46	27	capital
13	solteiro	médio	—	8,74	37	outra
14	casado	fundamental	3	8,95	44	outra
15	casado	médio	0	9,13	30	interior
16	solteiro	médio	—	9,35	38	outra
17	casado	médio	1	9,77	31	capital
18	casado	fundamental	2	9,80	39	outra
19	solteiro	superior	—	10,53	25	interior
20	solteiro	médio	—	10,76	37	interior
21	casado	médio	1	11,06	30	outra
22	solteiro	médio	—	11,59	34	capital
23	solteiro	fundamental	—	12,00	41	outra
24	casado	superior	0	12,79	26	outra
25	casado	médio	2	13,23	32	interior
26	casado	médio	2	13,60	35	outra
27	solteiro	fundamental	—	13,85	46	outra
28	casado	médio	0	14,69	29	interior
29	casado	médio	5	14,71	40	interior
30	casado	médio	2	15,99	35	capital
31	solteiro	superior	—	16,22	31	outra
32	casado	médio	1	16,61	36	interior
33	casado	superior	3	17,26	43	capital
34	solteiro	superior	—	18,75	33	capital
35	casado	médio	2	19,40	48	capital
36	casado	superior	3	23,30	42	interior

5. A MB Indústria e Comércio, desejando melhorar o nível de seus funcionários em cargos de chefia, montou um curso experimental e indicou 25 funcionários para a primeira turma. Os dados referentes à seção a que pertencem, notas e graus obtidos no curso estão na tabela abaixo. Como havia dúvidas quanto à adoção de um único critério de avaliação, cada instrutor adotou seu próprio sistema de aferição. Usando dados da tabela referida, responda às questões:

- Após observar atentamente cada variável, e com o intuito de resumi-las, como você identificaria (qualitativa ordinal ou nominal e quantitativa discreta ou contínua) cada uma das 9 variáveis listadas?
- Apenas olhando a tabela, indique as diferenças existentes entre as distribuições das variáveis Direito, Estatística e Política.
- Construa o histograma com 8 classes para as notas da variável Redação.
- Construa a distribuição de frequências da variável Metodologia e faça um gráfico para indicar essa distribuição.
- Sorteado ao acaso um dos 25 funcionários, qual a probabilidade de que ele tenha obtido grau A em Metodologia?
- Se, em vez de um, sorteássemos dois, a probabilidade de que ambos tivessem tido A em Metodologia é maior ou menor do que a resposta dada no item e)?
- Olhando para a seção ao qual os funcionários pertencem, qual delas teve melhor aproveitamento na disciplina Estatística?

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

(*) - P = Departamento pessoal, T = Seção técnica e V = Seção de vendas.

6. As taxas médias geométricas de incremento anual (por 100 habitantes) dos 30 maiores municípios do Brasil estão listadas abaixo.

3,67 1,82 3,73 4,10 4,30
 1,28 8,14 2,43 4,17 5,36
 3,96 6,54 5,84 7,35 3,63
 2,93 2,82 8,45 5,28 5,41
 7,77 4,65 1,88 2,12 4,26
 2,78 5,54 0,90 5,09 4,07

- a) Construa a distribuição de frequências com classes de amplitude de tamanho 2 começando em zero e terminando em 10.
 b) Faça um histograma à partir do **item a)**.

7. Dispomos de uma relação de 200 aluguéis de imóveis urbanos e uma relação de 100 aluguéis de imóveis rurais.

Classes de aluguéis (codificados)	Zona urbana	Zona rural
2 † 3	10	30
3 † 5	40	50
5 † 7	80	15
7 † 10	50	5
10 † 15	20	0
Total	200	100

- a) Construa os histogramas das duas distribuições.
 b) Com base nos histogramas, discuta e compare as duas distribuições.

8. O número de divórcios em uma 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.
 d) Encontre $q(0, 25)$ e $q(0, 75)$.
 e) A distribuição parece simétrica?

9. 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 o total pelo desvio padrão $\mathcal{D}(x)$.

10. Na **Questão 5**, temos 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 $\mathcal{D}(x)$.
- (II) A nota X de cada aluno será padronizada do seguinte modo:

$$Z = \frac{X - \bar{x}}{\mathcal{D}(x)}.$$

- a) Interprete o significado de Z .
 - b) Calcule as notas padronizadas dos funcionários para o exame de Estatística.
 - c) Com os resultados obtidos no **item b)**, calcule \bar{z} e $\mathcal{D}(z)$.
 - d) Se alguma nota estiver fora do intervalo $(-2\mathcal{D}(z), 2\mathcal{D}(z))$ esse funcionário deve ser considerado um caso atípico. Existe alguém nessa situação?
 - e) O funcionário 1 obteve 9,0 em *Direito*, em *Estatística* e em *Política*. Em qual disciplina seu desempenho relativo foi melhor?
11. Supondo que podem existir valores repetidos no conjunto $X = \{x_1, \dots, x_n\}$, demonstre a validade da propriedade

$$\text{Var}(X) = \sum_{i=1}^k f_i x_i^2 - \bar{x}^2, \quad 1 \leq k \leq n.$$

12. A distribuição de frequências do salário anual (x 10 salários mínimos) dos moradores do bairro A que tem alguma forma de rendimento é apresentada na tabela abaixo:

Faixa salarial	Frequência
0 † 2	10.000
2 † 4	3.900
4 † 6	2.000
6 † 8	1.100
8 † 10	800
10 † 12	700
12 † 14	2.000
Total	20.500

- a) Construa um histograma da distribuição.
- b) Calcule a média e use o item anterior para calcular o desvio padrão.

- 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 em relação à renda?
- d) Com base na distribuição acumulada decida qual é a faixa salarial dos 10% mais ricos da população.
- e) Qual a “riqueza total” dos moradores do bairro?
13. Se $0 < \alpha < 1$, uma *média aparada* a $100\alpha\%$ é obtida eliminando-se $100\alpha\%$ dentre as maiores e menores observações e calculando-se a média aritmética das restantes. Por exemplo, se tivermos 10 observações ordenadas $x_{(1)} < x_{(2)} < \dots < x_{(10)}$, a média aparada a 10% é

$$\bar{x}(0, 10) = \frac{x_{(2)} + \dots + x_{(9)}}{8}.$$

- a) Calcule a média aparada a 10% e 25% para os dados da variável *Salários* referentes à tabela presente na **Questão 4**.
- b) Qual a intenção do autor ao criar a medida de posição *média aparada*?
14. Numa pesquisa sobre rotatividade de mão de obra, para uma amostra de 40 pessoas foram observadas duas variáveis: número de empregos nos últimos dois anos (X) e salário mais recente, em número de salários mínimos (Y). Os resultados foram:

Indivíduo	X	Y	Indivíduo	X	Y
1	1	6	21	2	4
2	3	2	22	3	2
3	2	4	23	4	1
4	3	1	24	1	5
5	2	4	25	2	4
6	2	1	26	3	2
7	3	3	27	4	1
8	1	5	28	1	5
9	2	2	29	4	4
10	3	2	30	3	3
11	2	5	31	2	2
12	3	2	32	1	1
13	1	6	33	4	1
14	2	6	34	2	6
15	3	2	35	4	2
16	4	2	36	3	1
17	1	5	37	1	4
18	2	5	38	3	2
19	2	1	39	2	3
20	2	1	40	2	5

- a) Usando a mediana, classifique os indivíduos em dois níveis, alto e baixo, para cada uma das variáveis, e construa a distribuição de frequências conjunta das duas classificações.
- b) Qual a porcentagem das pessoas com baixa rotatividade e ganhando pouco?

- c) Qual a porcentagem das pessoas que ganham pouco?
- d) Entre as pessoas com baixa rotatividade, qual a porcentagem das que ganham pouco?
- e) A informação adicional dada em d) mudou muito a porcentagem observada em c)? O que isso significa?
- f) Verifique se há associações entre as variáveis rotatividade e salário.
15. A companhia **A** de dedetização afirma que o processo por ela utilizado garante um efeito mais prolongado do que aquele obtido por seus concorrentes diretos. Uma amostra de vários ambientes dedetizados foi colhida e anotou-se a duração do efeito de dedetização. Os resultados estão na tabela abaixo. Você acha que existe alguma evidência a favor ou contra a afirmação feita pela companhia **A**?

Companhia	Duração do efeito de dedetização		
	Menos de 4 meses	De 4 a 8 meses	Mais de 8 meses
A	64	120	16
B	104	175	21
C	27	48	5

16. Prove que

$$\frac{1}{n} \sum_i \left(\frac{x_i - \bar{x}}{\mathcal{D}(x)} \right) \left(\frac{y_i - \bar{y}}{\mathcal{D}(y)} \right) = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{(\sum x_i^2 - n \bar{x}^2) (\sum y_i^2 - n \bar{y}^2)}}.$$

17. Abaixo estão os dados referentes à porcentagem da população economicamente ativa empregada no setor primário e o respectivo índice de analfabetismo para algumas regiões metropolitanas brasileiras.

Regiões metropolitanas	Setor primário	Índice de analfabetismo
São Paulo	2,0	17,5
Rio de Janeiro	2,5	18,5
Belém	2,9	19,5
Belo Horizonte	3,3	22,0
Salvador	4,1	26,5
Porto Alegre	4,3	16,6
Recife	7,0	36,6
Fortaleza	13,0	38,4

- a) Faça o diagrama de dispersão.
- b) Você acha que existe uma dependência linear entre as duas variáveis?
- c) Calcule o coeficiente de correlação.
- d) Existe alguma região com comportamento diferente das demais? Se existe, elimine o valor correspondente e recalcule o coeficiente de correlação.



Respostas

1.
 - a) Qualitativa nominal.
 - b) Quantitativa discreta.
 - c) Quantitativa discreta.
 - d) Quantitativa contínua.
 - e) Quantitativa contínua.
 - f) Qualitativa ordinal.
 - g) Quantitativa contínua.
 - h) Qualitativa nominal.
2. Item (c).
3. Item (b).
4. Construção.
5.
 - a) **Seção:** Variável qualitativa nominal.
Administração: Variável quantitativa contínua.
Direito: Variável quantitativa contínua.
Redação: Variável quantitativa contínua.
Estatística: Variável quantitativa contínua.
Política: Variável quantitativa contínua.
Economia: Variável quantitativa contínua.
Inglês: Variável qualitativa ordinal.
Metodologia: Variável qualitativa ordinal.
8.
 - a) $\bar{x} = 6,90$ e $\tilde{x} = 5,35$.
 - b) $\text{Var}(X) = 27,62$ e $\mathcal{D}(X) = 5,26$.
 - d) $q_1 = 2,68$ e $q_3 = 10,07$.
 - e) Assimétrica.
9.
 - a) $2\tilde{x}$, $2\bar{x}$ e $2\mathcal{D}(X)$.
 - b) $10 + \tilde{x}$, $10 + \bar{x}$ e $\mathcal{D}(X)$.
 - c) $\tilde{x} - \bar{x}$, 0 e $\mathcal{D}(X)$.
 - d) $\frac{\tilde{x} - \bar{x}}{\mathcal{D}(X)}$, 0 e 1 .
10.
 - a) Média seja zero e o desvio padrão 1.
 - b) 0,58; 1,35; 1,35; -0,95; -0,95; 0,58; -0,18; -0,95; -0,18; -0,18; -0,18; -0,18; -0,95; 0,58; 1,35; -0,18; 0,58; 0,58; -3,26; 0,58; 0,58; -0,18; 0,58; -0,95; 0,58
 - c) $\bar{z} = 0$ e $\mathcal{D}(Z) = 1$.
 - d) O funcionário 19 é um caso atípico.

- e) Política.
12. b) A média é 3,9172 e o o desvio padrão é 3,9631.
c) Bairro A.
d) A faixa salarial dos 10% mais ricos é [11,8592 ; 14,0000].
e) Aproximadamente 80.302,6.
13. a) $\bar{x}(0,10) = 10,84$ e $\bar{x}(0,25) = 10,52$.
b) Eliminar a contribuição exagerada que valores atípicos proporcionam à média aritmética.
14. a) Considere *baixa rotatividade* para valores menores que 2 e *baixo salário* para valores menores que 2,5
b) 2,5%.
c) 50%.
d) 12,5%.
e) Bastante modificada; observe que maioria das pessoas que ganham pouco tem alta rotatividade.
f) Parece haver uma associação entre *alta rotatividade* e *baixos salários*.
15. Não há diferença entre as 3 empresas.
16. Demonstração.
17. b) O gráfico indica uma dependência linear positiva entre as variáveis.
c) 0,87.
d) As regiões de Porto Alegre e Fortaleza parecem fugir à curva. O novo coeficiente de correlação é aproximadamente 0,99.