

Lista de exercícios propostos
Análise Exploratória
Estatística I

1. Mostre que:
 - a) $\sum_{i=1}^n (X_i - \bar{X}) = 0$
 - b) $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i)^2 - n\bar{X}^2 = \sum_{i=1}^n (X_i)^2 - \frac{(\sum_{i=1}^n X_i)^2}{n}$
2. O que acontece com a média, com a variância e com o desvio padrão de um conjunto 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 a média geral \bar{X} e divide-se pelo desvio padrão S_x
3. A tabela abaixo representa a distribuição do grau de instrução dos funcionários de uma empresa.

Grau de Instrução	%
<i>Fundamental</i>	40
<i>Médio Incompleto</i>	10
<i>Médio Completo</i>	25
<i>Superior</i>	17
<i>Pós – graduação</i>	8
Total	100

- (a) Construa um gráfico de barras (também conhecido como gráfico de colunas) e o gráfico de setores da distribuição do grau de instrução.
 - (b) Sabendo que a empresa tem 200 funcionários, quantos têm pós-graduação?
 - (c) No gráfico de setores, quantos graus haveria o setor que representa cada categoria de grau de instrução?
4. Uma indústria de componentes eletrônicos está interessada em determinar a vida útil de certo tipo de bateria. Uma amostra, em horas, segue abaixo:

123, 116, 122, 110, 175, 126, 125, 111, 118, 117.

- (a) Encontre a média e a mediana amostrais.
 - (b) Qual característica nessa amostra é responsável pela considerável diferença entre as duas?
5. (*Walpole et al. E.1.1 p.8 adaptado*). Foram registradas as seguintes medidas para o tempo de secagem, em horas, de certa marca de tinta látex:
 3,4 2,5 4,8 2,9 3,6 2,8 3,3 5,6 3,7 2,8 4,4 4,0 5,2 3,0 4,8
 Suponha que as medidas sejam uma amostra aleatória simples.
 - (a) Qual é o tamanho da amostra acima?
 - (b) Calcule a média, a mediana, a variância, o desvio padrão e o primeiro e o terceiro quartis amostrais para este conjunto de dados.
 - (c) Faça um gráfico de ramos-e-folhas destes dados. Interprete os dados.

6. (*Walpole et al. E.1.17 p. 17*). Um estudo dos efeitos do tabagismo nos padrões de sono é conduzido. A medida observada é o tempo, em minutos, que se leva para dormir. Os dados obtidos são:
 Fumantes: 69,3 56,0 22,1 47,6 53,2 48,1 52,7 34,4 60,2 43,8 23,2 13,8
 Não-fumantes: 28,6 25,1 26,4 34,9 29,8 28,4 38,5 30,2 30,6 31,8 41,6 21,1 36,0 37,9 13,9
 (a) Encontre a média amostral em cada grupo.
 (b) Encontre o desvio-padrão amostral em cada grupo.
 (c) Faça um gráfico de pontos ou histograma dos dois conjuntos de dados
 (d) Comente o tempo de impacto que o fumo aparenta ter no tempo que se leva para dormir.
7. A espessura (mm) de 9 parafusos foi mensurada e os dados obtidos foram: 2.8; 4.2; 5.5; 3.0; 4.1; 3.9; 2.7; 4.2 e 2.9. Calcule as medidas descritivas: mínimo, máximo, quartis, mediana, média, amplitude, desvio médio, variância, desvio padrão e coeficiente de variação.
8. (*Walpole et al. E.1.19 p. 17*). Os dados a seguir representam a duração da vida útil, em anos, medidos do décimo mais próximo, 30 bombas de combustível:
 2,0 3,0 0,3 3,3 1,3 0,4 0,2 6,0 5,5 6,5
 0,2 2,3 1,0 6,0 5,6 1,5 4,0 5,9 1,8 4,7
 0,7 4,5 0,3 1,5 0,5 2,5 5,0 6,0 1,2 0,2
 (a) Construa um diagrama de ramos-e-folha para a vida, em anos, das bombas de combustíveis, usando o dígito a esquerda da vírgula decimal como ramo para cada observação. Interprete os resultados.
 (b) Estabeleça a distribuição de frequência relativas.
 (c) Calcule a média, a amplitude e o desvio padrão amostrais.
9. Os dados a seguir correspondem aos recordes de atletas em 10 países nas Olimpíadas de Los Angeles em 1984 em algumas provas de atletismo.
 (a) Separadamente por gênero e modalidade, faça uma análise descritiva dos recordes.
 (b) Através de gráficos do tipo boxplot, faça uma comparação entre os sexos nas quatro modalidades. Em qual delas, ha maior diferença entre homens e mulheres?

Mulheres				
País	100m (seg)	400m(seg)	3000m (min)	Maratona(min)
<i>Argentina</i>	11,61	54,50	9,79	178,52
<i>Brasil</i>	11,31	52,80	9,77	168,75
<i>Chile</i>	12,00	54,90	9,37	171,38
<i>Colômbia</i>	11,6	53,26	9,46	165,42
<i>Alemanha</i>	11,01	48,16	8,75	148,53
<i>França</i>	11,15	51,73	8,98	155,27
<i>Portugal</i>	11,81	54,30	8,84	151,20
<i>Canadá</i>	11,00	50,06	8,81	149,50
<i>USA</i>	10,79	50,62	8,50	142,72
<i>Kenya</i>	11,73	52,70	9,20	181,05

Homens				
País	100m (seg)	400m(seg)	3000m (min)	Maratona(min)
<i>Argentina</i>	10,39	46,84	14,04	137,72
<i>Brasil</i>	10,22	45,21	13,62	133,13
<i>Chile</i>	10,24	46,20	13,61	134,03
<i>Colômbia</i>	10,43	46,10	13,49	131,35
<i>Alemanha</i>	10,16	44,50	13,21	132,23
<i>França</i>	10,11	45,28	13,34	132,30
<i>Portugal</i>	10,53	46,70	13,13	128,22
<i>Canadá</i>	10,17	45,68	13,55	131,15
<i>USA</i>	9,93	43,86	13,20	128,22
<i>Kenya</i>	10,46	44,92	13,10	129,75

10. Calcule a média, a mediana e a variância amostrais aproximadas para o conjunto de dados com a seguinte distribuição de frequência e construa seu histograma.

Intervalo de classe	Frequência
$-10 \leq x < 0$	3
$0 \leq x < 10$	8
$10 \leq x < 20$	12
$20 \leq x < 30$	16
$30 \leq x < 40$	9
$40 \leq x < 50$	4
$50 \leq x < 60$	2

11. Para se estudar o desempenho de 4 corretoras de ações, selecionaram-se de cada uma delas amostras de ações negociadas. Para cada ação negociada, computou-se a porcentagem de lucro durante um período fixado de tempo. Os dados estão a seguir:

Corretora A: 21 48 42 37 38 25 29 32 33 33

Corretora B: 12 42 32 28 26 26 16 18 35

Corretora C: 58 32 46 45 50 52 56 56 37

Corretora D: 42 44 22 32 24 37 37 32 31 28

Existe diferença entre as corretoras com relação à porcentagem de lucro durante o período estudado?

12. Mostre que o coeficiente de correlação de Pearson está entre -1 e 1 .

13. Um experimento foi delineado para determinar o efeito da taxa de fluxo de C_2F_6 sobre a uniformidade na queima de chapas de silicone usadas na integração de um circuito de manufatura. Três taxas de fluxo são utilizadas no experimento e os resultados da uniformidade (em %) de seis chapas avaliadas são:

C_2F_6	Uniformidade					
125	3.6	3.8	3.8	4.2	4.5	4.6
160	3.6	4.2	4.2	4.6	4.9	5
200	5.1	5.4	5.6	5.7	6.2	6.5

Você acredita que a uniformidade na queima de chapas de silicone é a mesma para as diferentes taxas de fluxo de C_2F_6 ? Justifique.

14. Na companhia A, a média dos salários é de R\$ 10.000,00 e o 3° quartil é R\$ 5.000,00.
 (a) Se você se apresentasse como candidato a essa firma e se o seu salário fosse escolhido ao acaso entre todos os possíveis salários, o que seria mais provável: ganhar mais ou menos do que R\$ 5.000,00?

(b) Suponha que na companhia B a média dos salário é de R\$ 7.000,00 e a variância é praticamente zero e que, lá, o seu salário também fosse escolhido ao acaso. Em qual companhia você se apresentaria para arrumar emprego?

15. Para estudar o efeito da condição do solo no crescimento de uma planta, amostras provenientes de 3 tipos de solo e seus crescimentos são classificados em 3 categorias. Calcule as freqüências relativas e compare a qualidade do crescimento para diferentes tipos de solo.

Tipo de solo			
Crescimento	Com pedregulho	Com areia	Argila
<i>Fraco</i>	16	8	14
<i>Médio</i>	31	16	21
<i>Bom</i>	18	36	25
<i>Total</i>	65	60	60

16. Uma indústria utiliza um determinado reagente químico (X) para obter uma dada substância (Y). Foram realizados 5 experimentos e as quantidades do reagente e da substância em cada um deles são as seguintes:

	1	2	3	4	5
<i>Reagente(mg)</i>	6,0	5,5	5,0	7,5	12,5
<i>Substância(mg)</i>	2,0	3,0	3,5	4,0	5,0

- a) Construa o gráfico de dispersão entre as variáveis X e Y.
 b) Determine o coeficiente de correlação de Pearson entre X e Y.
 c) A quantidade de reagente utilizada está relacionada com a quantidade de substância produzida? De que forma? Justifique.
 d) Se um outro reagente Z for utilizado para a obtenção da mesma substância (Y), e se o seu coeficiente de correlação de Pearson com a substância for $r_{yz} = 1$, como seria o gráfico de dispersão entre Y e Z?
17. Muitas vezes, a determinação da capacidade de produção instalada para certo tipo de indústria é um processo difícil e custoso. Como alternativa, pode-se estimar a capacidade de produção através da escolha de uma outra variável de medida mais fácil e que esteja linearmente relacionada com ela.
 Suponha que foram observados os valores para as variáveis: capacidade de produção instalada (X), potência instalada (Y) e área construída (Z). Com base num critério estatístico, qual das variáveis você escolheria para estimar a capacidade de produção instalada?

<i>X(ton.)</i> :	4	5	3	5	8	9	10	11	12	12
<i>Y(1000kW)</i> :	1	2	1	3	3	5	5	7	6	7
<i>Z(100m)</i> :	6	7	10	11	11	9	12	12	11	14