

3ª Lista de Exercícios - SME0803 Análise Exploratória de Dados

Data: 23/03/2012 Entregar dia 30/03/2012 às 19h os exercícios 6 e 7.

Exercício 1. Comente as principais vantagens e desvantagens do uso das medidas de posição média, mediana e moda. Dê exemplos de casos em que é preferível utilizar cada uma delas.

Exercício 2. Mostre as propriedades P1, P2, P3 e P4 da média.

Exercício 3. Obtenha a curva de sensibilidade da mediana quando $n - 1$ é ímpar.

Exercício 4. Mostre que, para x_1, x_2, \dots, x_n reais positivos, as médias aritmética, geométrica e harmônica são casos particulares da fórmula geral das médias, dada por

$$m(q) = \left[\sum_{i=1}^n \frac{x_i^q}{n} \right]^{\frac{1}{q}}$$

Obs: Para a média geométrica, use $\lim_{q \rightarrow 0} \left[\sum_{i=1}^n \frac{x_i^q}{n} \right]^{\frac{1}{q}} = \sqrt[n]{x_1 x_2 \dots x_n}$ para concluir que $\lim_{q \rightarrow 0} m(q) = m_g$.

Exercício 5. Em uma pesquisa realizada com 100 famílias foram coletados os dados apresentados na tabela abaixo. Apresente a mediana e a moda do número de filhos. Proponha uma aproximação para o número médio de filhos. Justifique. (Bussab e Morettin, Estatística Básica Saraiva 6ª Edição 2009).

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

Exercício 6. Dados sobre o tempo de falha (em h) de certo item foram coletados durante um período de 100h, fornecendo os valores

$$76, 63, 100^*, 36, 51, 45, 50, 40, 100^*, 35, 90,$$

sendo que “100*” indica que o item ainda não havia falhado ao término da coleta dos dados.

- (a) Apresente a(s) medida(s) de posição que você considera que pode calcular de forma exata.
- (b) O que você pode afirmar sobre a média?

Exercício 7. Um conjunto de dados com n observações foi dividido em dois grupos com n_1 e n_2 observações, $n_1 > 1$ e $n_2 > 1$. As médias dos dois grupos são dadas por \bar{x}_1 e \bar{x}_2 . Prove que $\bar{x} = (n_1 \bar{x}_1 + n_2 \bar{x}_2) / n$. Generalize o resultado para um número g de grupos, $g > 2$.

Exercício 8. O arquivo censo.txt disponível na WIKI contém dados de renda média (em reais) e número médio de anos de estudo (dos chefes de domicílios) nos estados brasileiros de acordo com o CENSO 2000. (Paula, G. A. Modelos de regressão com apoio computacional, IME USP, 2010).

- (a) Construa tabelas de frequências para as variáveis escolaridade e renda com 6 classes.
- (b) Construa o gráfico de ramos-e-folhas para a variável escolaridade e o histograma para a variável renda.
- (c) Descreva os dados utilizando medidas de posição que você considerar apropriadas.
- (d) Obtenha os quartis para as variáveis escolaridade e renda.

Exercício 9. O jornal “O Estado de São Paulo” publicou em uma reportagem em 2010 as seguintes afirmações:

“88% das cidades da região têm menos verde do que a média do estado”.

“78 cidades têm menos verde do que a média”.

“Só 11 dos 78 municípios têm cobertura vegetal acima do índice de 17% feito pelo Estado”.

O que podemos afirmar sobre a mediana? É maior do que ou menor do que a média estadual (= 17%)? Justifique.