

USP – ICMC – Análise Exploratória de Dados
1ª lista de exercícios – 1º/2010

1. Demonstre a propriedade P3 da média.
2. Demonstre a propriedade P4 da média.
3. Demonstre a propriedade P5 da média.
4. (a) Um conjunto de dados com n observações foi dividido em dois grupos com n_1 e n_2 observações, $n_1 \geq 1$ e $n_2 \geq 1$. As médias dos dois grupos são dadas por \bar{x}_1 e \bar{x}_2 . Prove que $\bar{x} = (n_1\bar{x}_1 + n_2\bar{x}_2)/n$.
(b) Generalize o resultado do item anterior para um número de grupos g qualquer, $g \geq 2$.
5. Obtenha a curva de sensibilidade da mediana na situação em que $n - 1$ é ímpar.
6. Apresente uma forma aproximada de obter os quantis utilizando as frequências acumuladas.
7. (Bussab e Morettin, 2009) 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. Encontre uma aproximação para o número médio de filhos.

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

8. O histograma dos dados da variável x encontra-se na Figura 1. Qual(is) medida(s) de posição você utilizaria?

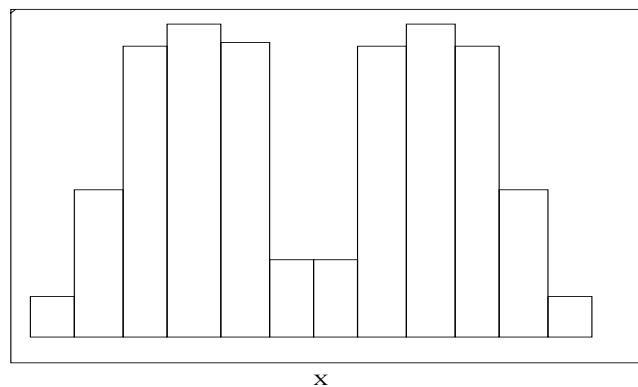


Figura 1: Histograma da variável x .

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

76, 63, 100⁺, 36, 51, 45, 50 e 90,

sendo que a observação “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?
10. O arquivo `mrpd2000.txt` contém os dados da percentagem total de mulheres responsáveis pelo domicílio (`mrpd`) nos municípios paulistas com base no Censo Demográfico 2000. Esta variável foi utilizada no estudo do Índice Paulista de Vulnerabilidade Social, que se encontra em http://www.al.sp.gov.br/web/ipvs/index_ipvs.htm.
- (a) O arquivo `mrpd2000.txt` está no tradicional formato CSV (*comma separated values*) utilizado, por exemplo, pelo Excel. Pode ser lido em R com o comando `read.csv("mrpd2000.txt", dec = ",", sep = ";", header = TRUE)`. Descreva o significado de cada um dos argumentos deste comando.
 - (b) Consultando a página indicada acima crie um arquivo com as variáveis município e `mrpd`. Leia o arquivo criado e compare com os dados lidos no item (10a).
 - (c) Apresente os dados de `mrpd` em uma tabela de frequências com 12 classes.
 - (d) Represente graficamente os dados de `mrpd`.
 - (e) No item (10d) algum município se destacou? Se sim, em que sentido? Se sim, qual(is) foi(ram) o(s) município(s)?
 - (f) Calcule as medidas de posição que você considerar apropriadas.
 - (g) Apresente os decis da variável `mrpd`.