USP – ICMC – Análise Exploratória de Dados
2ª lista de exercícios – 1º/2010

1. Obtenha a curva de sensibilidade da amplitude ($A = x_{(n)} - x_{(1)}$).

2. Obtenha as expressões da variância (s^2) nas situações descritas abaixo.

   (a) Os dados estão na forma $(x_1, f_1), \ldots, (x_k, f_k)$, sendo que f_j representa a frequencia absoluta de x_j, $j = 1, \ldots, k$.

   (b) Os dados estão na forma $(x_1, f_1^\star), \ldots, (x_k, f_k^\star)$, sendo que $f_j^\star$ representa a frequencia relativa de x_j, $j = 1, \ldots, k$.

   (c) Os dados estão na forma $(x_1^\star, f_1), \ldots, (x_k^\star, f_k)$, sendo que f_j representa a frequencia absoluta da classe com ponto médio $x_j^\star$, $j = 1, \ldots, k$. A expressão obtida corresponde ao verdadeiro valor de s^2?

3. Demonstre a propriedade P3 da variância.

4. Que transformação deve ser aplicada aos dados $x_1, \ldots, x_n$, $n > 1$, para que a variável resultante tenha média e desvio padrão iguais a 0 e 1, respectivamente?

5. Prove que $\sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - n\overline{x}^2$. Utililze este resultado para obter uma expressão alternativa da variância.

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

   $$76,\ 63,\ 100^+,\ 36,\ 51,\ 45, 50 \text{ e } 90,$$

   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 dispersão que você considera que pode calcular de forma exata.

   (b) O que você pode afirmar sobre a amplitude e o desvio padrão?

7. Considere um conjunto de dados $x_1, \ldots, x_n$, $n > 1$, tal que $x_i \in [\min, \max]$, $i = 1, \ldots, n$, com $\min < \max$, n par. Apresente a distribuição dos dados que maximiza o desvio padrão e o seu valor máximo. Se n for ímpar, qual a sua resposta?

8. Considere um conjunto de dados $x_1, \ldots, x_n$, $n > 1$, tal que $x_i \in [\min, \max]$, $i = 1, \ldots, n$, com $\min < \max$. Pode ser provado que $A_s \leq \sqrt{2(n-1)}$, sendo que A_s denota a amplitude studentizada. Conclua que este limite superior é atingido se tivermos um só valor igual a min, um só valor igual a max e todos os demais (se $n > 2$) iguais a $(\min + \max)/2$.

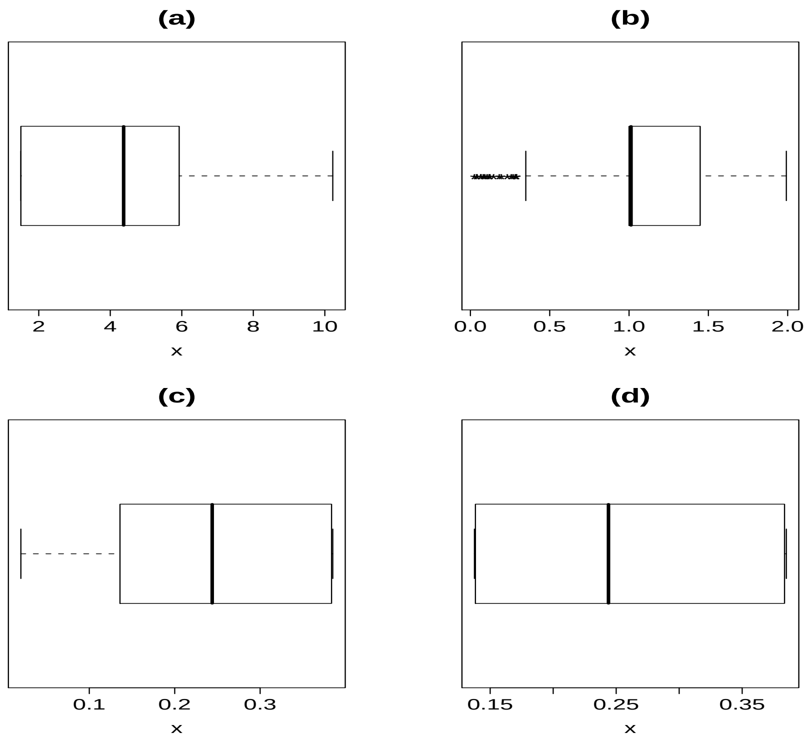

Figura 1: Gráficos de caixa.

9. No exercício 8 foi mencionada uma medida de posição chamada de centro (*mid-range* ou *midrange*), dada por $x_c = (\min + \max)/2$.

   (a) x_c é uma medida resistente?

   (b) Apresente a curva de sensibilidade de x_c.

10. Descreva conjuntos de dados correspondentes a cada um dos gráficos de caixa da Figura 1.

11. Os itens abaixo devem ser resolvidos utilizando os dados apresentados no exercício 10 da 1ª lista.

   (a) Apresente o gráfico de caixa para os dados de mrpd. Se existir(em) observação(ções) extrema(s), identifique o(s) município(s).

   (b) Calcule as medidas de dispersão que você considerar apropriadas.