ESTATÍSTICA

Professora: Patrícia Ferreira Paranaíba

Medidas de Posição: São valores que representam a tendência de concentração dos dados observados.

- Média Aritmética: A média de uma população ou amostra é a soma de todos os elementos da população (amostra) dividida pelo número de elementos. Esta medida apresenta a mesma unidade dos dados.
  - Para a população a média é representada por:

    $$\mu = \frac{\sum_{i=1}^{N} x_i}{N},$$ em que N é o tamanho da população.

  - Para a amostra a média é representada por:

    $$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n},$$ em que n é o tamanho da amostra.

- Exemplo: O tempo de vida útil (em horas) de uma amostra de 6 lâmpadas incadescentes é: 612, 983, 623, 883, 666 , 970. A média amostral do tempo de vida é dado por:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{612 + 983 + 623 + 883 + 666 + 970}{6} = 789,5h$$

Propriedades da média

- Adição ou Subtração por uma constante
  Seja $(X_1, X_2, X_3, ..., X_n)$ uma amostra aleatória de tamanho n, k uma constante e $\overline{X}$ a média da amostra. Se somarmos ou subtrairmos todos os valores de uma variável X pela constante k, o valor de $\overline{X}$ MÉDIA fica somada ou subtraída pela constante.
- Se no exemplo das lâmpadas somarmos a constante 2 a cada um dos valores da variável temos 614, 985, 625, 885, 667,972

$$\overline{X}^* = \frac{614 + 985 + 625 + 885 + 668 + 972}{6} = \frac{4749}{6} = 791,5h$$

Utilizando a propriedade,

$$\overline{X}^* = \overline{X} + k = 789,5 + 2 = 791,5h$$

Propriedades da média

- Multiplicação ou divisão por uma constante
  Seja $(X_1, X_2, X_3, ..., X_n)$ uma amostra aleatória de tamanho n, k uma constante e $\overline{X}$ a média da amostra. Se multiplicarmos ou dividirmos todos os valores de uma variável X pela constante k, o valor de $\overline{X}$ MÉDIA fica multiplicada ou dividida pela constante.
- Se no exemplo das lâmpadas multiplicarmos a constante 2 a cada um dos valores da variável temos 1224, 1966, 1246, 1766, 1332, 1940.

  $$\overline{X}^* = \frac{1224 + 1966 + 1246 + 1766 + 1332 + 1940}{6} = 1579h$$

  Utilizando a propriedade,

  $$\overline{X}^* = k\overline{X} = 2 \times 789,5 = 1579h$$

Propriedades da média

- Soma dos desvios
  Seja $(X_1, X_2, X_3, ..., X_n)$ uma amostra aleatória de tamanho n e $\overline{X}$ a média da amostra. Se subtrairmos cada valor da variável X pela média obtemos os desvios. A soma algébrica dos desvios é igual a zero.
- No exemplo da lampâda, temos:

Amostra	$\overline{X}$	Desvio
612	789,5	-177,5
983	789,5	193,5
623	789,5	-166,5
883	789,5	93,5
666	789,5	-123,5
970	789,5	180,5
	soma dos desvios	0

- Mediana: Num conjunto de dados ordenados, a mediana (M_d) é o valor que deixa metade da freqüência abaixo dele. A mediana, como a média, possui a mesma unidade de cada observação.
- A mediana pode ser obtida por meio da expressão:

$$M_d = \begin{cases} X_{\frac{n+1}{2}} & \text{se } n \text{ for ímpar} \\ \\ \frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} & \text{se } n \text{ for par} \end{cases}$$

- Exemplo 1: Considere o conjunto de dados: 5, 2, 6, 13, 9, 15, 10. Primeiro é necessário ordenar os dados: 2, 5, 6, 9, 10, 13, 15. Como se de uma conjunto com $n = 7$ (ímpar), então:

$$Md = X_{\frac{n+1}{2}} = X_{\frac{7+1}{2}} = X_4$$

Logo a mediana é igual ao elemento que está na quarta posição do conjunto de dados, assim

$$Md = 9$$

- Exemplo 2: Considere o conjunto de dados: 1, 3, 8, 6, 2, 4. Primeiro é necessário ordenar os dados: 1, 2, 3, 4, 6, 8. Como se de uma conjunto com $n = 6$ (par), então:

$$Md = \frac{X_{\frac{n}{2}} + X_{\frac{n+2}{2}}}{2} = \frac{X_{\frac{6}{2}} + X_{\frac{6+2}{2}}}{2} = \frac{X_3 + X_4}{2}$$

Logo para obter a mediana é necessário obter os elementos que estão na terceira e quarta posição do conjunto de dados, assim:

$$Md = \frac{3+4}{2} = 3,5$$

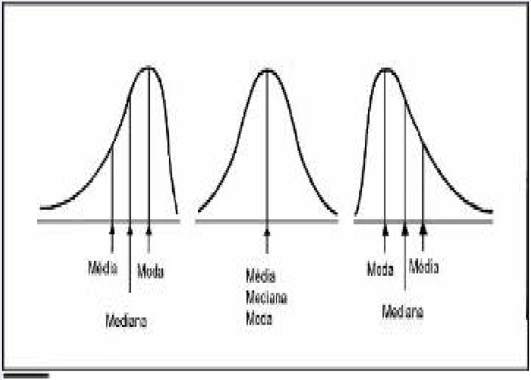
- Moda: A moda M_o de um conjunto de dados é o valor mais freqüente e também tem a mesma unidade dos dados. Para obter a moda basta observar qual o dado que mais se repete.
- Exemplo 1: No conjunto de dados 7 , 8 , 9 , 10 , 10 , 10 , 11 , 12 a moda é igual a 10, pois é único que se repete.
- Exemplo 2: No conjunto de dados 3 , 5 , 8 , 10 , 12 não apresenta moda. O conjunto é amodal
- Exemplo 3: No conjunto de dados 2 , 3 , 4 , 4 , 4 , 5 , 6 , 7 , 7 , 7 , 8 , 9 temos duas modas: 4 e 7. O conjunto é bimodal.

Simetria

A determinação das medidas de posição permite discutir sobre a simetria da distribuição dos dados.

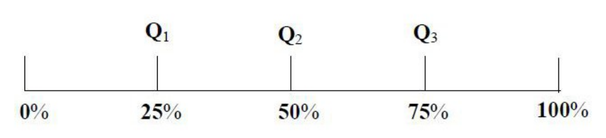
- Distribuição simétrica: $\overline{X} = M_d = M_o$
- Distribuição assimétrica: ocorrem diferenças entre os valores da média, mediana e moda. A assimetria pode ser:
  - à direita: $\overline{X} > M_d > M_o$
  - à esquerda: $\overline{X} < M_d < M_o$

Representação gráfica



Quantis

- Quartil: Denominamos quartis os valores de uma série que a dividem em quatro partes iguais.



Para determinar a ordem ou posição do quartil a ser calculado, usaremos a seguinte expressão:

$$EQ_i = \frac{i(n+1)}{4}$$

em que:

- i = número do quartil a ser calculado;
- n = número de observações.

Exemplo: Considere o conjunto de dados: 5, 2, 6, 9, 10, 13, 15

Primeiro é necessário ordenar os dados: 2, 5, 6, 9, 10, 13, 15.

$$EQ_1 = \frac{1}{4}(7+1) = 2$$

Logo o quartil 1 é o elemento da 2º posição. Logo $Q_1 = 5$.

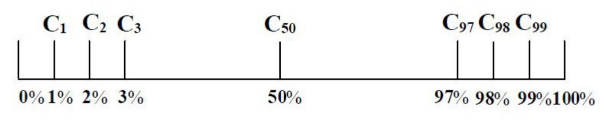
$$EQ_2 = \frac{2}{4}(7+1) = 4$$

Logo o quartil 2 é o elemento da 4º posição. Logo $Q_2 = 9$.

$$EQ_3 = \frac{3}{4}(7+1) = 6$$

Logo o quartil 3 é o elemento da 9º posição. Logo $Q_3 = 13$.

- Percentil: São as medidas que dividem a amostra em 100 partes iguais. Assim:



O elemento que definirá a ordem do percentil será encontrado pelo emprego da expressão:

$$E_{Pi} = \frac{i}{100}(n+1)$$

em que:

- i = número identificador do percentil;
- n = número total de observações.

Exemplo: Considere o conjunto de dados: 5, 2, 6, 9, 10, 13, 15

Primeiro é necessário ordenar os dados: 2, 5, 6, 9, 10, 13, 15.
Percentil 25

$$EP_{25} = \frac{25\times}{100}(7+1) = 2$$

Logo o percentil 25 é o elemento da 2º posição. Logo $P_{25} = 5$.
Percentil 50

$$EP_{50} = \frac{50\times}{100}(7+1) = 4$$

Logo o percentil 50 é o elemento da 4º posição. Logo $P_{50} = 9$.

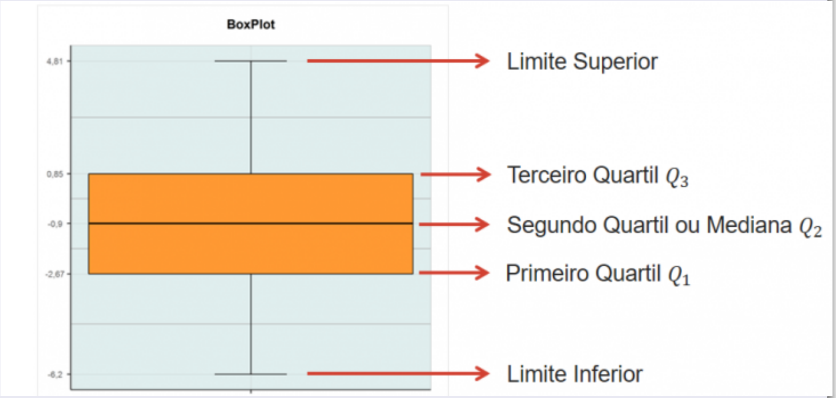
Boxplot

- O gráfico Boxplot (ou desenho esquemático) é uma análise gráfica que oferece a ideia da posição, dispersão, assimetria, caudas e dados discrepantes. Para construí-lo, desenhamos uma "caixa"com o nível superior dado pelo terceiro quartil (Q_3) e o nível inferior pelo primeiro quartil (Q_1). A mediana (Q_2) é representada por um traço no interior da caixa e segmentos de reta são colocados da caixa até dos limites inferior (LI) e superior (LS), dados por

$$\begin{aligned} LI &= Q_1 - 1,5dq \\ LS &= Q_3 + 1,5dq \end{aligned}$$

em que $dq = Q_3 - Q_1$ denominando diferença quartílica.

Representação

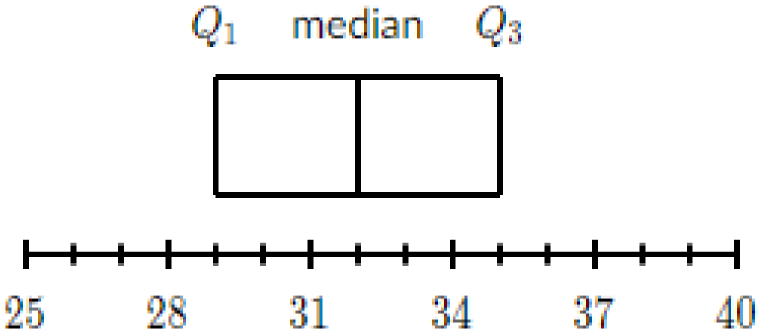


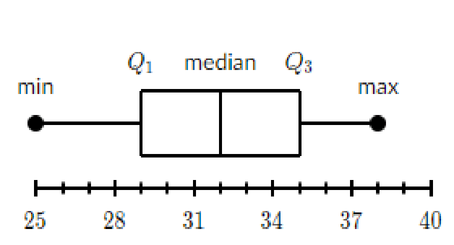
Para traçarmos o boxplot utilizamos as seguintes etapas:

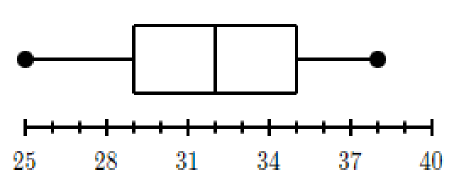
- Contruir um retângulo de tal maneira que os lados menores correspondem aos primeiro e terceiro quartis da distribuição.
- Cortar o retângulo por um segmento paralelo aos lados menores, na altura correspondente à mediana;
- Traçar um segmento paralelo ao eixo, partindo do ponto médio até o maior valor observado que NÃO supere LS;
- Traçar um segmento paralelo ao eixo, partindo do ponto médio até o menor valor que NÃO é menor LI;
- Caso tenha valores superiores a LS ou inferiores a LI, marcar os pontos, este valores são considerados observações discrepantes.

Exemplo: Considere um lote de 500 caixas de castanhas ensacadas que tem os seguintes pesos: 25g, 28g, 29g, 29g, 30g, 34g, 35g, 35g, 37g, 38g.

$$
\begin{aligned}
EQ_1 &= \frac{1(n+1)}{4} = \frac{(10+1)}{4} = 2,75 \approx 3 \Rightarrow Q_1 = 29 \\
EQ_2 &= \frac{2(10+1)}{4} = 5,5 \Rightarrow Q_2 = \frac{30+34}{2} = 32 \\
EQ_3 &= \frac{3(10+1)}{4} = 8,25 \approx 8 \Rightarrow Q_3 = 35 \\
LI &= 29 - 1,5.6 = 20 \\
LS &= 35 + 1,5.6 = 44
\end{aligned}
$$


Q_1
median
Q_3
25
28
31
34
37
40


Q_1
median
Q_3
min
max
25
28
31
34
37
40


25
28
31
34
37
40