ESTATÍSTICA

Professora: Patrícia Ferreira Paranaíba

Medidas de dispersão: As medidas de posição são importantes para caracterizar um conjunto de dados, mas não são suficientes para caracterizar completamente a distribuição dos dados. Para isso é necessário obter as medidas de dispersão, que medem a variabilidade dos dados.

- Por exemplo: Considere as amostras referentes a altura, em cm, de dois grupos de pessoas.

$$GrupoA : 185 \quad 185 \quad 185$$

$$GrupoB : 187 \quad 183 \quad 185$$

A média para os dois grupos é a mesma $\overline{X}_A = 185$ e $\overline{X}_B = 185$.
As medidas de variabilidade ou dispersão possibilitam que façamos distinção entre os conjuntos quanto à sua homogeneidade, isto é, o grau de concentração em torno de uma medida de tendência central.

Amplitude Total

- Amplitude Total (A) é a diferença entre o maior e o menor valor da amostra. Essa medida é bastante simples, e obtida pela expressão:

$$A = Max - Min$$

Para expressar variabilidade a amplitude total não é muito usada, pois baseia-se em apenas dois dados.

Variância e Desvio Padrão

- A variância é baseada pela quadrado dos desvios dos dados em relação à média. Esta medida é expressa na unidade dos dados ao quadrado.
    - Para a população a variância é representada por:

      $\sigma^2 = \dfrac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$, em que N é o tamanho da população.

    - Para a amostra a variância é representada por:

      $S^2 = \dfrac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n-1}$, em que n é o tamanho da população.

- O desvio padrão é a raíz quadrada positiva da variância. Esta medida é expressa na mesma unidade dos dados.
  - Para a população o desvio padrão é representada por:

$$\sigma = \sqrt{\sigma^2}$$

  - Para a amostra o desvio padrão é representada por:

$$S = \sqrt{S^2}$$

Coeficiente de Variação

- O coeficiente de variação (CV) é uma medida de dispersão que expressa o desvio padrão em termos da média de forma percentual:

$$CV = 100 \times \frac{S}{\overline{X}}$$

Se as amostras tiverem unidade diferentes ou médias diferentes o CV pode ser utilizado para comparar a variabilidade entre duas amostras.

Exemplo

Uma fábrica de ervilhas comercializa seu produto em embalagens de 300 gramas e em embalagens de um quilo ou 1000 gramas. Para efeitos de controle do processo de enchimento das embalagens, sorteia-se uma amostra de 10 embalagens de cada uma das máquinas e obtém-se os seguintes resultados:

$$300g \rightarrow \begin{cases} \overline{X} = & 296\text{g}; \\ S = & 5. \end{cases}$$

$$1000g \rightarrow \begin{cases} \overline{X} = & 996\text{g}; \\ S = & 5. \end{cases}$$

- Na primeira máquina, as embalagens deveriam fornecer peso de $300g$ mas devido a erros de ajuste da máquina de enchimento, o peso médio das 10 embalagens é de apenas $296g$. O desvio-padrão de $5g$ significa que, em média, os pesos das embalagens estão 5 gramas abaixo ou acima do peso médio das 10 latas. Uma interpretação análoga vale para a segunda máquina.
- Em qual das duas situações a variabilidade parece ser maior? Ou seja, em qual das duas máquinas parece haver um problema mais sério? Observe que, em ambos os casos, há uma dispersão de 5g em torno da média, mas $5g$ em $1000g$ é menos preocupante que $5g$ em $300g$.

- Os coeficientes de variação para as embalagens oriundas das duas máquinas são:

$$300g \rightarrow CV = 100 \times \frac{5}{296} = 1,69\%$$

$$1000g \rightarrow CV = 100 \times \frac{5}{996} = 0,5\%$$

Isso confirma a nossa observação anterior: a variabilidade na máquina de 300g é relativamente maior.