

MÉDIA, VARIÂNCIA E DESVIO PADRÃO

Para resumir dados quantitativos aproximadamente **simétricos**, é usual calcular a **média** aritmética como uma medida de locação.

Se x_1, x_2, \dots, x_n são os valores dos dados, então podemos escrever a média como $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$, onde $\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$ e frequentemente é simplificada para $\sum x_i$ ou até mesmo $\sum x$ que significa adicione todos os valores de x .

A **variância** é definida como o “desvio quadrático médio da média” e é calculada de uma amostra de dados como $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n (x_i^2) - n\bar{x}^2}{(n-1)}$.

A segunda versão é mais fácil de ser calculada, embora muitas calculadoras têm funções prontas para o cálculo de variâncias, e é raro ter que realizar todos os passos manualmente. Comumente as calculadoras fornecerão a raiz quadrada da variância, o **desvio padrão**, i.e. $s = \sqrt{\text{variância}} = \sqrt{s^2}$ a qual é medida nas mesmas unidades dos dados originais.

Uma informação útil é que para qualquer conjunto de dados, pelo menos 75% deles fica dentro de uma distância de 2 desvio padrão da média, i.e. entre $\bar{x} - 2s$ e $\bar{x} + 2s$.

Exemplo. Sete homens foram pesados, e os resultados em kg foram:

57,0; 62,9; 63,5; 64,1; 66,1; 67,1; 73,6.

A **média** é $\frac{454,3}{7} = 64,9 \text{ kg}$, a **variância** é $\frac{\left(\frac{29635,05 - 454,3^2}{7}\right)}{6} = 25,16 \text{ kg}$ e o **desvio padrão** é $\sqrt{25,16} = 5,02 \text{ kg}$.