Aula 12 - Média, desvio e variância

Matheus Roos de Oliveira

14 de junho de 2022

Preliminares

Motivação

- Legar códigos para o curso de física (Lab.);
- A tarefa maçante de produzir relatórios semanais;
- Software de produção de gráficos;
- Os dois problemas básicos em programação científica:
  1. Equações sem solução analítica;
  2. Tarefas maçantes e demoradas.

Erro como incerteza

- Física como ciência experimental;
- A definição de erro em ciência;
- A aplicação da ciência dependerá da medição, bem como é crucialmente importante a capacidade de avaliação das incertezas e de mantê-las as menores possíveis;
- Erros não podem ser eliminados!
- A incerteza é intrínseca ao ato de uma medição, se dividindo em dois tipos:
  1. Sistemático: atinge todas as medidas da mesma forma, como um equipamento descalibrado (um cronômetro adiantado);
  2. Aleatório: atinge as medidas de forma randômica (tempo de reação ao acionar um cronômetro), e poderá ser curado por métodos estatísticos.

Erro Sistemático vs Aleatório

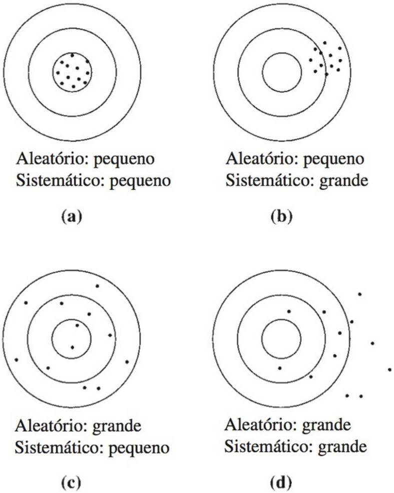

Figura: Taylor, Introdução à Análise de Erro

A melhor estimativa de uma grandeza

- Suponha que seja feita uma única medição de uma grandeza física como o tempo t;
- Partimos também do pressuposto que todos os erros aleatórios são conhecidos e que sejam insignificantes, portanto iremos desprezá-los;
- Suponha agora o processo de medição seja repetido por um número N de vezes;
- Dos valores registrados, qual corresponderá a melhor estimativa para a grandeza medida? O primeiro? O último? O que mais se repete?

A melhor estimativa de uma grandeza

- Suponha que seja feita uma única medição de uma grandeza física como o tempo t;
- Partimos também do pressuposto que todos os erros sistemáticos sejam conhecidos e que sejam insignificantes, portanto iremos desprezá-los;
- Suponha agora o processo de medição seja repetido por um número N de vezes;
- Dos valores registrados, qual corresponderá a melhor estimativa para a grandeza medida? O primeiro? O último? O que mais se repete?
- Podemos esperar que a melhor estimativa seja a média, tal que

$$t_{\text{melhor}} = \bar{t} = \frac{t_1 + t_2 + ... + t_N}{N},$$

- ou em notação totalmente equivalente

$$< t >= \frac{\sum_{i=1}^{N} t}{N}$$

Definição de variância

- Lembrando que nós devemos saber estimar nossas incertezas, uma boa estimativa para isso seria ver o quanto os valores desviam da média
- Definimos então o desvio (ou resíduo) d como

$$d = x_i - \overline{x}.$$

- Se $d \approx 0$ concluímos que nossas medidas são precisas.
- Mas se quisermos testar a confiabilidade de nossas medições, poderíamos tentar calcular a média dos desvios, mas

$$\overline{d} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right) = 0.$$

- Podemos contornar isso se quadrarmos o desvio e tomarmos a média,

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2.$$

- Onde definimos σ^2 como a variância.

Definindo o desvio padrão

- Mas nós gostaríamos de expressar a incerteza na medida de uma certa medida x como
$$x_{\text{melhor}} = \overline{x} \pm \delta x$$
- Então se extrairmos a raiz quadrada da variância,
$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$
- então $\delta x = \sigma$ é o desvio padrão, também chamado de *raiz média quadrática*, ou Root Mean Square (RMS). Mas ela também é conhecida como desvio padrão populacional.
- Há uma outra forma, um pouco mais precisa e com maior rigor teórico que é o desvio padrão amostral, abaixo:
$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}.$$

Refinando o valor da incerteza

- Mas se aumentarmos o número de medições, esperaríamos que a incerteza diminuiria lentamente, mas a equação

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

- não garante isso, então se escrevermos

$$\delta x = \sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}$$

- onde definimos $\sigma_{\overline{x}}$ como o *desvio padrão da média*, do inglês, Standard Deviation of the Mean (SDOM) teremos o efeito desejado.