Instituto de Física

FAP 2292

Experiência 0

Tratamento Estatístico de Dados Experimentais

1º semestre de 2010

26 de fevereiro de 2010

Os cursos FAP2292 e FAP2293 foram montados sobre o princípio de que a Física não se divide em partes *teórica* e *experimental* independentes. Ao contrário, os aspectos teóricos e experimentais de qualquer modelo físico formam um todo integrado. Isto se torna claro nos conceitos físicos necessários ao engenheiro elétrico moderno; não é possível desenvolver projetos práticos inovadores sem uma boa fundamentação teórica, como também não é possível entender e absorver todas as conseqüências de um modelo teórico sem verificação experimental. Por isso as experiência foram concebidas de forma a acompanhar o desenvolvimento teórico visto em sala de aula. No entanto, não são experiências de demonstração. No laboratório, o aluno é levado a se empenhar na montagem da experiência, na obtenção dos dados e análise dos resultados, utilizando praticamente conceitos aprendidos em aula. A unidade dos aprendizados teórico e experimental é realçada pela participação dos *mesmos* professores no ensino teórico e experimental, e pela avaliação dos alunos, para os quais as notas obtidas em provas teóricas, e nos relatórios e e prova de laboratório são integrados em um único conceito final.

Prof. Ricardo Magnus Osório Galvão
Coordenador do projeto

Muitos foram os autores e colaboradores desta apostila, entre os quais se destacam:

Delton O. Campos †
Aluisio N. Fagundes
Márcia C.A.Fantini
Ricardo M. O. Galvão
José Rafael Leon
Ernesto Lerche
Rene O. Medrano
Cássio Sanguini-Neto
Álvaro Vannucci

Equipe 2010:

Coraci Pereira Malta
Felix G. G. Hernandez
José Eduardo Padilha de Sousa
Luiz Carlos Nagamine
Manfredo H. Tabacniks
Paula S. Meirelles

---

0. Tratamento estatístico de dados experimentais

Objetivos

- Recordar conceitos associados ao tratamento de dados experimentais;
- Estabelecer fórmulas mais comuns;
- Procedimento gráfico: Ajuste de curvas a dados experimentais;

Esta apresentação destina-se *exclusivamente* a servir como recordação de conhecimentos já aprendidos. Para uma exposição extensiva do assunto, são indicados, entre outros:

- FUNDAMENTOS DA TEORIA DE ERROS, J. H. Vuolo, Ed Edgar Blücher Ltda, 1992.
- TRATAMENTO ESTATÍSTICO DE DADOS EM FÍSICA EXPERIMENTAL, O. A. M. Helene e Vito R. Vanin, Ed. Edgar Blücher Ltda, 1981.
- TÓPICOS AVANÇADOS EM TRATAMENTO ESTATÍSTICO DE DADOS EM FÍSICA EXPERIMENTAL Vito R. Vanin e P. Gouffon, LAL-IFUSP, 1996.
- EXPERIMENTAL MEASUREMENTS: PRECISION, ERROR AND TRUTH, N. C. Barford, Addison-Wesley Publishing Co., Inc. 1967.
- DATA REDUCTION AND ERROR ANALYSIS FOR THE PHYSICAL SCIENCES, P. R. Bevington, McGraw-Hill Book Co. 1969.

Introdução Toda medida experimental resulta afetada de erro. Duas medidas de uma mesma grandeza feitas nas mesmas condições, em geral, não apresentam o mesmo resultado. Considere o conjunto de medidas do diâmetro de um pino cilíndrico praticadas com um paquímetro (com incerteza nominal de $0{,}01$ mm), tabeladas a seguir.

d (mm)	0,88	0,95	0,86	0,89	0,93	0,90

Qual destas medidas *melhor* representa o diâmetro do cilindro? A resposta: *nenhuma!*

A melhor escolha para o diâmetro procurado é o *valor médio* do conjunto de medidas: [1]

$$\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i, \tag{1}$$

onde $\bar{x}$ representa o valor médio, n é o número de medidas e x_i é a i-ésima medida.

A dispersão da série de medidas pode ser aferida através do seu *desvio padrão* ou *desvio quadrático médio* σ:

$$\sigma^2 = \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{n-1}. \tag{2}$$

O *desvio padrão da média* σ_m é uma estimativa da dispersão que seria obtida em médias de diferentes conjuntos de medidas efetuadas nas mesmas condições:

$$\sigma_m^2 = \sum_{1}^{n} \frac{(x_i - \bar{x})^2}{n(n-1)} = \frac{\sigma^2}{n}. \tag{3}$$

Como o resultado da medida é representado pela *média*, a sua incerteza deve ser dada pelo *desvio padrão da média*. Entretanto, como a Eq. (3) indica, este desvio pode ser feito tão pequeno quanto se queira efetuando um número cada vez maior de medidas, o que não faz sentido. Na incerteza do resultado devemos levar em conta, também, a incerteza do instrumento de medida, σ_c. Tomamos, então, como incerteza do resultado:

$$\sigma_p = \sqrt{\sigma_m^2 + \sigma_c^2}. \tag{4}$$

Para o exemplo apontado de 6 medidas, tomando $\sigma_c = 0{,}01$ mm, obtemos:

$$\bar{d} = 0{,}901667\ldots \text{ mm}, \ \sigma = 0{,}033115\ldots \text{ mm}, \ \sigma_m = 0{,}013520\ldots \text{ mm} \ \text{ e } \sigma_p = 0{,}016816\ldots \text{ mm}.$$

Levando em conta a estimativa da incerteza da medida, vemos que o algarismo duvidoso em $\bar{d}$ é o segundo dígito depois da vírgula: o 0 à direita do 9. Assim todos os dígitos que o seguem não tem significado.

Para expressar a incerteza utilizamos, no máximo, dois algarismos significativos. Tanto a incerteza quanto o valor da medida devem ser arredondados convenientemente. Assim, para o resultado no exemplo escreveríamos:

$$\boxed{\bar{d} = (0{,}90 \pm 0{,}02) \text{ mm},} \tag{5}$$

ou

$$\boxed{\bar{d} = (0{,}902 \pm 0{,}017) \text{ mm.}} \tag{6}$$

Observe em (5) que **não** se pode omitir o zero à direita quando ele é um algarismo significativo.

Como regra, vamos expressar a incerteza com apenas um algarismo significativo. Entretanto, quando o arredondamento do valor da incerteza resultar em "1" ou "2" como único algarismo, devemos utilizar dois algarismos para minimizar o erro introduzido pelo arredondamento.

Segundo esta regra, a expressão (6) acima é a maneira correta de expressar o resultado final da medida do diâmetro do cilindro no exemplo. Se o resultado vai ser utilizado posteriormente para computar outra grandeza (como a densidade, por exemplo), é conveniente utilizar a forma

[1] Construamos uma função $D^2 = \sum_1^n (x_i - \bar{x})^2$. O *melhor* valor de $\bar{x}$ é aquele que torna mínima esta soma dos quadrados das diferenças:

$$\frac{\mathrm{d}D^2}{\mathrm{d}\bar{x}} = 0 \Rightarrow -2\sum_{1}^{n}(x_i - \bar{x}) = 0 \Rightarrow \sum_{1}^{n} x_i - \bar{x}\sum_{1}^{n} 1 = \sum_{1}^{n} x_i - n\bar{x} = 0 \Rightarrow \bar{x} = \frac{1}{n}\sum_{1}^{n} x_i.$$

Assim, a média é a melhor escolha para representar o conjunto de medidas x_i.

(6) para as grandezas intermediárias, mesmo que a regra anterior permita expressar a incerteza com um único algarismo significativo, para minimizar os erros de arredondamento.

O procedimento de computar média e desvio padrão de uma série de medidas pode ser tedioso, mas *é necessário*. A maioria das máquinas de calcular incluem programas de cálculo de médias e desvio-padrão. **Aprenda a usar sua máquina !**

Propagação de Incertezas Raras vezes a medida que fazemos constitui o resultado final do processo de investigação que fazemos. O número obtido deve ser transformado matematicamente — somado, subtraído, dividido, multiplicado... — para dar a conhecer a resposta. Outros números intervenientes irão, com boa probabilidade, também apresentar uma incerteza, o que nos leva à questão: como determinar a incerteza de uma grandeza z que resulta de operações matemáticas envolvendo resultados de medidas de outras grandezas, como por exemplo x_1, x_2, etc., com incertezas σ_1, σ_2, etc.?

Se z é dado por $z = f(x_1,x_2,\dots)$ e as incertezas σ_1, e σ_2, etc. são independentes, a incerteza em $\bar{z} = f(\bar{x}_1,\bar{x}_2,\dots)$ é dada por

$$\sigma_z^2 = \left(\frac{\partial f}{\partial x_1}\sigma_1\right)^2 + \left(\frac{\partial f}{\partial x_2}\sigma_2\right)^2 + \dots, \tag{7}$$

onde as derivadas parciais são computadas nos pontos $x_1 = \bar{x}_1$, $x_2 = \bar{x}_2$, etc..

Alguns casos particulares importantes desta forma geral:

- **Adição ou subtração**: Se $z = \pm x \pm y$, $|\partial z/\partial x| = 1$ e $|\partial z/\partial y| = 1$, resultando:

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2.$$

- **Potenciação**: Se $z = x^\alpha y^\beta$, $|\partial z/\partial x| = \left|\alpha x^{\alpha-1} y^\beta\right| = |\alpha z/x|$ e $|\partial z/\partial y| = |\beta z/y|$, resultando:

$$\left(\frac{\sigma_z}{z}\right)^2 = \left(\alpha\frac{\sigma_x}{x}\right)^2 + \left(\beta\frac{\sigma_y}{y}\right)^2.$$

A generalização destes resultados quando há mais de duas parcelas ou produtos é imediata. Resumindo:

> Quando somando ou subtraindo grandezas afetadas de erro, combine os quadrados das incertezas **absolutas.** Quando multiplicando ou dividindo envolvendo potências, devem ser somados os quadrados das incertezas **percentuais** multiplicadas pelos expoentes respectivos.

Ajuste de curvas Freqüentemente praticamos uma série de medidas as quais sabemos pertencer a uma relação matemática. O caso mais simples é o de uma *relação linear*, que se traduz numa *reta* quando os resultados são apresentados de forma gráfica. Seja, por exemplo, determinar o melhor valor da resistência elétrica de um resistor sobre o qual, dada a corrente, medimos a tensão. As medidas estão tabuladas na Tabela 1 e apresentadas graficamente na Figura 1. Tomamos a incerteza das medidas como constante — como é a situação prática mais encontrada.

Sabemos que o resistor obedece à lei de Ohm, e portanto $V(I) = RI$. Em nossas medidas, admitimos que a *variável independente* I não contem erros e as incertezas estão sobre a *variável dependente* V. Nosso problema é determinar *o melhor* valor para a resistência elétrica R. Neste exemplo, *todos* os pontos são utilizados para a determinação do *melhor* valor para a resistência elétrica do resistor. Que métodos podemos utilizar para isto?

O método mais simples é o olho. Por mais surpreendente que possa parecer, um ajuste visual dos dados normalmente fornece bons resultados – e deve ser a primeira estimativa do processo. Um outro método consiste em traçar *retas de inclinações extremas* sobre os pontos e praticar algum tipo de média sobre os resultados. Entretanto, se queremos um método que forneça os mesmos

i	I(A)	V(V) ($\pm$0,5 V)
1	0,2	1,0
2	0,4	2,2
3	0,6	3,0
4	0,7	3,5
5	0,8	4,0
6	1,2	6,0
7	1,4	7,0
8	1,6	7,8
9	1,8	9,2
10	2,0	11,0

i	x_i	y_i	x_i^2	x_iy_i
1	0,2	1,0	0,04	0,20
2	0,4	2,2	0,16	0,88
3	0,6	3,0	0,36	1,80
4	0,7	3,5	0,49	2,45
5	0,8	4,0	0,64	3,20
6	1,2	6,0	1,44	7,20
7	1,4	7,0	1,96	9,80
8	1,6	7,8	2,56	12,48
9	1,8	9,2	3,24	16,56
10	2,0	11,0	4,00	22,00
Σ	10,7	54,7	14,89	76,57

Tabela 1: Tensões medidas sobre um resistor. Admitimos que a *corrente*, variável independente, não apresenta erros. Apenas os valores de tensão apresentam erros, que supomos iguais a $\sigma_V = 0{,}5$ V na região medida. Na tabela à direita estão os dados necessários para ajuste de uma reta pelo método dos mínimos quadrados.

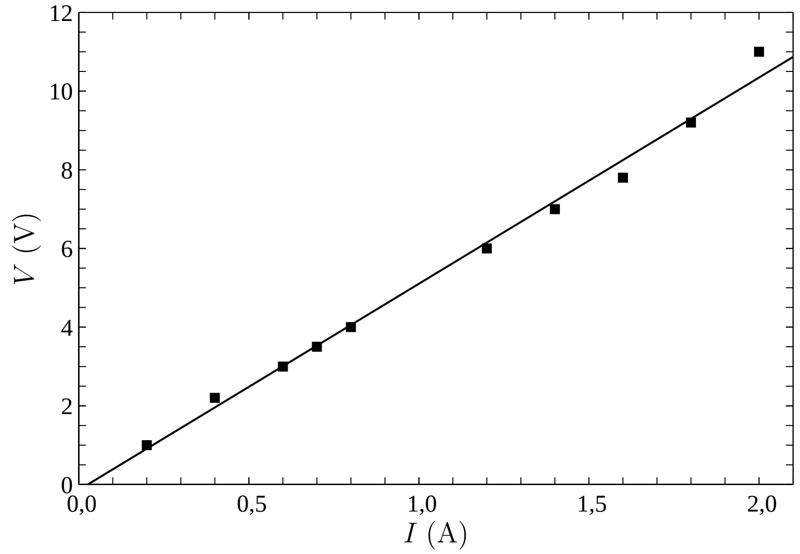

Figura 1: Dados pontos experimentais que sabemos obedecer a uma relação matemática conhecida, como determinar o melhor conjunto de parâmetros?

resultados, ainda que aplicado por pessoas diferentes, temos que recorrer a processos matemáticos mais elaborados.

O mais utilizado deles é denominado "Método dos mínimos quadrados" e consiste, em essência, em minimizar a diferença entre a curva a ser traçada e os pontos experimentais. Começamos construindo uma função matemática χ^2 (leia-se "qui quadrado") dos quadrados das *diferenças* entre os pontos experimentais y_i e os pontos teóricos correspondentes y_{ti}

$$\chi^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(y_i - y_{ti})^2 = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}(y_i - ax_i - b)^2$$

onde σ_i é a incerteza associada ao ponto medido y_i, n é o número de pontos experimentais, a e b são os coeficientes da reta que melhor ajusta os pontos — as nossas incógnitas.

Minimizamos χ^2 para determinar os melhores valores de a e b. Posto de outra forma, os *melhores* valores de a e b são exatamente aqueles que *tornam mínima a soma das diferenças entre os pontos* *experimentais e a reta procurada*. Então,

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= -2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(y_i - ax_i - b)\,x_i &= 0 \\ \frac{\partial \chi^2}{\partial b} &= -2\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(y_i - ax_i - b) &= 0 \end{aligned}$$

com solução

$$a = \frac{1}{\Delta}\left(\mathcal{S}_\sigma \mathcal{S}_{xy} - \mathcal{S}_x \mathcal{S}_y\right) \quad \text{e} \quad b = \frac{1}{\Delta}\left(\mathcal{S}_{x^2} \mathcal{S}_y - \mathcal{S}_x \mathcal{S}_{xy}\right)$$

onde

$$\Delta = \mathcal{S}_\sigma \mathcal{S}_{x^2} - \mathcal{S}_x^2$$

e

$$\mathcal{S}_\sigma = \sum_{i=1}^{n}\frac{1}{\sigma_i^2}, \quad \mathcal{S}_x = \sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}, \quad \mathcal{S}_y = \sum_{i=1}^{n}\frac{y_i}{\sigma_i^2}, \quad \mathcal{S}_{xy} = \sum_{i=1}^{n}\frac{x_i y_i}{\sigma_i^2}, \quad \mathcal{S}_{x^2} = \sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^2}.$$

Exercício. Faça a dedução detalhada das fórmulas.

As incertezas sobre a e b são dadas por

$$\sigma_a = \sqrt{\mathcal{S}_\sigma/\Delta}, \qquad \sigma_b = \sqrt{\mathcal{S}_{x^2}/\Delta}.$$

Nos casos como o do exemplo em que as incertezas de todos os pontos são iguais, $\sigma_i = \sigma$, os resultados ficam:

$$a = \frac{1}{D}\left(n\Sigma_{xy} - \Sigma_x \Sigma_y\right), \quad \sigma_a = \sigma\sqrt{n/D} \quad \text{e} \quad b = \frac{1}{n}\left(\Sigma_y - a\Sigma_x\right), \quad \sigma_b = \sigma\sqrt{\Sigma_{x^2}/D},$$

onde

$$\Sigma_x = \sum_{i=1}^{n} x_i, \quad \Sigma_y = \sum_{i=1}^{n} y_i, \quad \Sigma_{xy} = \sum_{i=1}^{n} x_i y_i, \quad \Sigma_{x^2} = \sum_{i=1}^{n} x_i^2 \quad \text{e} \quad D = n\Sigma_{x^2} - \left(\Sigma_x\right)^2.$$

A forma mais fácil de praticar ajustes de retas está em utilizar as máquinas de calcular ou programas de computador. As contas que se seguem constituem apenas um exemplo de aplicação. Vamos aplicar esta técnica aos dados experimentais da tabela 1. As somas necessárias para os cálculos estão indicadas na tabela à direita dos dados, com as quais obtemos:

$$\begin{aligned} a &= 5{,}24295\ \text{V/A}, & \sigma_a &= 0{,}26954\ \text{V/A} & &\Rightarrow a = (5{,}2 \pm 0{,}3)\ \text{V/A} \\ b &= -0{,}13995\ \text{V}, & \sigma_b &= 0{,}32891\ \text{V} & &\Rightarrow b = (-0{,}1 \pm 0{,}3)\ \text{V}. \end{aligned}$$

A melhor reta obtida está indicada na Figura 1.
A resposta, pois, para o valor da resistência é $R = (5{,}2 \pm 0{,}3)\ \Omega$.
