VI – O Método dos Mínimos Quadrados

Frequentemente, as leis físicas permitem prever o valor de uma grandeza y a partir de uma variável independente x. Um caso particular, mas bastante comum, é aquele em que y depende de x através de funções $f(x)$ e $g(x)$ na forma linear

$$y = a_0 f(x) + b_0 g(x) \qquad \text{(VI.1)}$$

onde a_0 e b_0 são constantes (isto é, independentes de x), cujos valores são relacionados ao sistema específico em estudo.

> A equação horária do movimento de um corpo lançado para cima em um plano inclinado por um ângulo θ em relação à horizontal, a partir da origem, é
>
> $$y = -\frac{g\mathrm{sen}\theta}{2}t^2 + v_0 t$$
>
> quando as forças de atrito puderem ser ignoradas. Neste caso, $a_0 = -\frac{g\mathrm{sen}\theta}{2}$, $b_0 = v_0$, $f(t) = t^2$ e $g(t) = t$.

Uma maneira de estimar a_0 e b_0 é variar o valor de x e medir os correspondentes valores de y. Se medimos apenas 2 pares, ou seja, (x_1, y_1, σ_1) e (x_2, y_2, σ_2), onde σ_1 e σ_2 são os desvios-padrões de y_1 e y_2, respectivamente, é possível resolver o sistema de duas equações a duas incógnitas que se obtém; a solução será um par de valores $\hat{a}$ e $\hat{b}$ que não são os parâmetros da equação (VI.1), mas sim suas estimativas, cujas incertezas decorrentes das incertezas nos valores y_1 e y_2 devem ser calculadas pelas fórmulas de propagação adequadas. Essa maneira, entretanto, é insuficiente quando as incertezas de medida são significativas e não podem ser diminuídas. Além disso, esse procedimento não permite de maneira nenhuma testar se a relação VI.1 é adequada para a descrição do fenômeno.

A prática em Física consiste em determinar N valores da grandeza y para diferentes valores de x, ou seja, determinar um conjunto de dados

$$\{(x_i, y_i, \sigma_i), i = 1..N\}$$

em que o índice i simplesmente identifica cada dado do conjunto de N pontos experimentais e incluímos os desvios-padrões dos dados, σ_i. O interesse em repetir a medição de y um número maior de vezes decorre da diminuição dos desvios-padrões de $\hat{a}$ e $\hat{b}$, por conta da redução da flutuação estatística quando são calculados a partir de mais medições, do mesmo jeito que o desvio-padrão da média diminui com o número de dados, $\sigma_m = \frac{\sigma}{\sqrt{N}}$, como já aprendemos na primeira experiência – na verdade, a média e o desvio-padrão da média constituem a estimativa de mínimos quadrados da medição direta de uma grandeza aleatória, caso em que a fórmula VI.1 fica $y = a_0$. A fim de testar se a relação (VI.1) é válida, procura-se escolher valores de x distribuídos por toda a região de interesse da variável independente.

No exemplo do corpo que é lançado da base da rampa para cima com velocidade v_0, deve-se observar a posição do corpo desde o início até o final do movimento e não apenas concentrar as medições nos primeiros ou nos últimos instantes do movimento.

Para explicar o método que vamos usar, relacionamos o dado experimental com a função (VI.1):

$$y_i = a_0 f(x_i) + b_0 g(x_i) + \epsilon_i \qquad \text{(VI.2)}$$

onde ϵ_i é o erro da medida experimental. O método que vamos apresentar baseia-se na impossibilidade de determinar esse erro experimental – se fosse possível conhecê-lo, subtrairíamos o erro de medida e nunca precisaríamos repetir uma medição, porque o resultado final, depois da subtração, seria exato. Embora não se possa conhecer o erro, supõe-se que se possa conhecer seu *valor médio,* bem como sua *média quadrática*, que são as duas hipóteses necessárias para aplicar o método dos mínimos quadrados:

i. $\langle \epsilon_i \rangle = 0$, que corresponde à hipótese que os instrumentos não dão leituras erradas sempre para o mesmo lado, ou seja, as medições são não-tendenciosas.
ii. $\langle {\epsilon_i}^2 \rangle = {\sigma_i}^2$, que é a medida de dispersão dos dados experimentais e pode ser estimado como vimos fazendo desde o primeiro experimento.

O método consiste em minimizar a soma dos quadrados dos resíduos ponderados pelos desvios-padrões dos dados, mais exatamente, minimizar a função $Q(a, b)$

$$Q(a,b) = \sum_{i=1}^{N} \frac{\left[y_i - \left(af(x_i) + bg(x_i)\right)\right]^2}{{\sigma_i}^2} \qquad \text{(VI.3)}$$

Note que, nessa expressão, a e b são variáveis, o que é estranho, mas inevitável, uma vez que a natureza não nos permite conhecer os seus valores verdadeiros, a_0 e b_0 – temos que admitir a possibilidade de que assumam quaisquer valores. Note também que a idéia intuitiva de minimizar a soma dos erros não funciona, por causa da propriedade (i) acima – o erro médio é zero. Assim, as estimativas dos parâmetros são os valores $\hat{a}$ e $\hat{b}$ que minimizam essa função Q.

Como $Q(a,b)$ é um polinômio do 2º grau em a e b, pode-se desenvolver o polinômio que corresponde a essa parábola bidimensional e encontrar seu mínimo usando técnicas de álgebra linear (veja, por exemplo, o livro de Barone [1]). A maneira que vamos indicar aqui é calcular as derivadas parciais de $Q(a,b)$ em relação a a e b, que devem se anular para um par de valores $\hat{a}$ e $\hat{b}$, que são as estimativas das grandezas físicas a_0 e b_0, ou seja,

$$\left.\frac{\partial Q}{\partial a}\right|_{\hat{a},\hat{b}} = 0$$

$$\left.\frac{\partial Q}{\partial b}\right|_{\hat{a},\hat{b}} = 0$$

As duas equações do sistema linear acima (a derivada de um polinômio do 2º grau é um polinômio do 1º grau) em função das duas incógnitas $\hat{a}$ e $\hat{b}$ são escritas de preferência em forma de matriz,

$$\begin{pmatrix} \sum \frac{y_i f(x_i)}{{\sigma_i}^2} \\ \sum \frac{y_i g(x_i)}{{\sigma_i}^2} \end{pmatrix} = \begin{pmatrix} \sum \frac{\left(f(x_i)\right)^2}{{\sigma_i}^2} & \sum \frac{f(x_i)g(x_i)}{{\sigma_i}^2} \\ \sum \frac{f(x_i)g(x_i)}{{\sigma_i}^2} & \sum \frac{\left(g(x_i)\right)^2}{{\sigma_i}^2} \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \tag{VI.4}$$

onde todas as somatórias se estendem desde $i = 1$ até $i = N$. Essa fórmula matricial pode ser escrita em forma compacta como

$$\vec{D} = \mathbf{M}\vec{A} \tag{VI.5}$$

A solução do sistema linear pode ser obtida por qualquer método. Aqui, definimos a matriz **V** que é igual à inversa da matriz **M**

$$\mathbf{V}{=}\mathbf{M}^{-1} \tag{VI.6}$$

de modo que simbolizamos a solução de (VI.4) por

$$\vec{A} = \mathbf{V}\vec{D} \tag{VI.7}$$

Os desvios-padrões das estimativas $\hat{a}$ e $\hat{b}$ são calculados a partir dos elementos da matriz **V** [3-5], que, por isso, é chamada matriz das variâncias:

$$\sigma_{\hat{a}} = \sqrt{v_{11}} \text{ e } \sigma_{\hat{b}} = \sqrt{v_{22}} \tag{VI.8}$$

O método dos mínimos quadrados é devido a Gauss e Legendre, no final do século XVIII ou início do XIX, que o aplicaram na redução de dados de observações astronômicas; a prioridade de descoberta do método é uma questão interessante, veja, por exemplo, Stiegler [2], que sugere que Gauss descobriu o método antes de Legendre, que, porém, o apresentou em uma publicação que despertou o interesse dos astrônomos da época, o que Gauss teria tentado vários anos antes, mas não conseguido.

A qualidade do ajuste pode ser avaliada pelo valor da função Q da equação (VI.3) calculada com as estimativas $\hat{a}$ e $\hat{b}$ dos parâmetros, $Q(\hat{a},\hat{b})$, cuja distribuição de probabilidade pode ser calculada quando os dados têm distribuição normal. Se conhecêssemos os valores exatos a_0 e b_0, poderíamos escrever

$$Q(a_0,b_0) = \sum_{i=1}^{N} \frac{\left[y_i - \left(a_o f(x_i) + b_0 g(x_i)\right)\right]^2}{{\sigma_i}^2} = \sum_{i=1}^{N} \frac{{\epsilon_i}^2}{{\sigma_i}^2}$$

onde usamos a relação VI.2 para identificar os erros ϵ_i. Calculando o valor médio dos dois membros da equação, obtemos

$$\langle Q(a_0,b_0)\rangle = \sum_{i=1}^{N} \frac{\langle {\epsilon_i}^2\rangle}{{\sigma_i}^2} = \sum_{i=1}^{N} \frac{{\sigma_i}^2}{{\sigma_i}^2} = N \qquad \text{(VI.9)}$$

Essa relação não é muito útil, uma vez que não conhecemos a_0 nem b_0. No entanto, se usamos as estimativas conhecidas $\hat{a}$ e $\hat{b}$ no lugar de a_0 e b_0, obtemos

$$Q(\hat{a},\hat{b}) = \sum_{i=1}^{N} \frac{\left[y_i - \left(\hat{a} f(x_i) + \hat{b} g(x_i)\right)\right]^2}{{\sigma_i}^2} = \sum_{i=1}^{N} \frac{{\epsilon'_i}^2}{{\sigma_i}^2} \qquad \text{VI.10}$$

onde definimos o resíduo do ajuste no ponto x_i, ϵ'_i, como

$$\epsilon'_i = y_i - \left(\hat{a} f(x_i) + \hat{b} g(x_i)\right) \qquad \text{(VI.11)}$$

É possível calcular o valor médio da expressão (VI.10) [5], obtendo-se uma expressão bastante parecida com a equação VI.9:

$$\langle Q(\hat{a},\hat{b})\rangle = N - 2 \qquad \text{(VI.12)}$$

Como Q é uma função definida positiva ou nula, o fato de seu valor médio ser nulo para $N = 2$ significa que ele vale zero sempre. Isso porque, nesse caso de $N = 2$, as estimativas $\hat{a}$ e $\hat{b}$ são a solução do sistema linear, de modo que os resíduos ϵ'_1 e ϵ'_2 para os únicos dois dados são nulos e, portanto, sua soma quadrática é nula.

Quando fazemos *um* ajuste, porém, temos *um* único valor $Q(\hat{a},\hat{b})$, de modo que, se $N > 2$, ele pode assumir qualquer valor no intervalo $[0, \infty[$. Quando os dados têm uma distribuição gaussiana, a função de probabilidade dos diferentes valores pode ser calculada e demonstra-se que $Q(\hat{a},\hat{b})$ tem a função de probabilidade da variável χ^2_{N-2} (diz-se qui-quadrado com N–2 graus de liberdade), de forma que se pode realizar um teste de hipótese rigoroso, dentro do quadro teórico da teoria estatística. Aqui, vamos nos limitar a uma interpretação qualitativa dessa grandeza.

Valores de $Q(\hat{a},\hat{b})$ muito menores que N–2 são devidos a desvios-padrões superestimados. Já valores muito maiores que N–2 sugerem que o modelo seja inadequado, quando devemos buscar outra função para relacionar y com x no lugar da (VI.1), ou, então, os desvios-padrões estão muito subestimados. Se desenvolvermos as expressões algébricas da equação (VI.8), que determinam os desvios padrões de $\hat{a}$ e $\hat{b}$, veremos que, quando se subestimam (superestimam) os desvios padrões dos dados, os desvios dos resultados também estarão subestimados (superestimados).

Na equação (VI.12), o número 2 do membro direito, N–2, está associado aos 2 parâmetros a_0 e b_0 da relação entre y e x. Caso a relação que substitua (VI.1) tenha m parâmetros, esse número 2 será substituído por m. Por exemplo, se forem três os parâmetros: a_0, b_0 e c_0, então a relação (VI.10) fica

$$\langle Q(\hat{a},\hat{b},\hat{c})\rangle = N - m = N - 3 \qquad \text{(VI.13)}$$

Uma introdução ao método dos mínimos quadrados do ponto de vista do tratamento estatístico dos dados em física experimental pode ser encontrada no livro de Helene e Vanin [3] ou em Vuolo [4] e uma apresentação mais profunda em Helene [5].

Referências

[1] *Álgebra Linear*. Mário Barone Jr., 3ª edição, São Paulo, *IME*-USP (2002).

[2] *Gauss and the Invention of Least Squares*. Stephen M. Stigler. Annals of Statistics, **9** (1981) 465- 474 - doi:10.1214/aos/1176345451

[3] *Tratamento Estatístico de Dados em Física Experimental*. O. Helene e V.R. Vanin. Ed. Edgard Blucher, SP, 2ª edição (1991)

[4] *Fundamentos da teoria de erros.* José Henrique Vuolo. Ed. Edgard Blücher, São Paulo, SP, 2ª edição (1992)

[5] *Método Dos Mínimos Quadrados Com Formalismo Matricial*. Otaviano Helene. Ed. Livraria da Física (2006).

Referências e fontes bibliográficas

- R.P. Feynman., R.B. Leighton and M. Sands, *Lectures on Physics*, Vol 1. 1971.
- A.Hudson, R. Nelson, *University Physics*, 2nd Ed. Saunders College Publ. 1990.
- Diretório Central dos Estudantes. *Normatização de trabalhos acadêmicos & referências bibliográficas.* 2a. Ed. Pontifícia Universidade Católica de Campinas. - (1998). 52p.
- Fernandes, Normando C. O laboratório de projetos: inúmeras variações sobre o mesmo tema. *Preprint IFUSP*/ P-564. (1986).
- Frota, Maurício Nogueira, Ohayon, Pierre. eds. *Padrões e Unidades de Medida - Referências Metrológicas da França e do Brasil.* INMETRO - Rio de Janeiro: Qualitymark Ed. 1999. 120p.
- Helene, Otaviano A.M. e Vanin, Vito R. *Tratamento estatístico de dados em física experimental.* Ed. Edgard Blücher, São Paulo, SP. 1981.
- INMETRO, SBM. *Guia para expressão da incerteza de medição*. ABNT, Rio de Janeiro. (1998). 120p.
- *Referências Bibliográficas de Multimeios e Documentos Eletrônicos.* Pontifícia Universidade Católica de Campinas. Projeto Disque-Biblio, (1998) 19p.
- Saad, Fuad Daher, Yamamura, Paulo; Watanabe, Kazuo . *Introdução à interpretação gráfica de dados, gráficos e equações.* 25p. IFUSP (sem data).
- Vuolo, José Henrique. Fundamentos da teoria de erros. Ed. Edgard Blücher, São Paulo, SP. 2a Ed. 1992.
- Yamamura, Paulo e Watanabe, Kazuo *Instrumentos de Medição in Manuais Didáticos de Física.* 18p