

# Mínimos Quadrados

Em 1809, Carl Friedrich Gauss (1777-1855) publicou um artigo no [Werke, 4, 1-93](#), demonstrando que a melhor maneira de **determinar um parâmetro desconhecido de uma equação de condições é minimizando a soma dos quadrados dos resíduos**, mais tarde chamado de Mínimos Quadrados por Adrien-Marie Legendre (1752-1833). Em abril de 1810, Pierre-Simon Laplace (1749-1827) apresenta no memoir da Academia de Paris, ["Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et sur leur application aux probabilités (suite)". Mémoires l'Institut 1809 (1810), 353-415, 559-565. Oeuvres 12 p.301-345, p.349-353] a generalização a problemas com vários parâmetros desconhecidos.

Um programa de mínimos quadrados sempre começa com a minimização da soma:

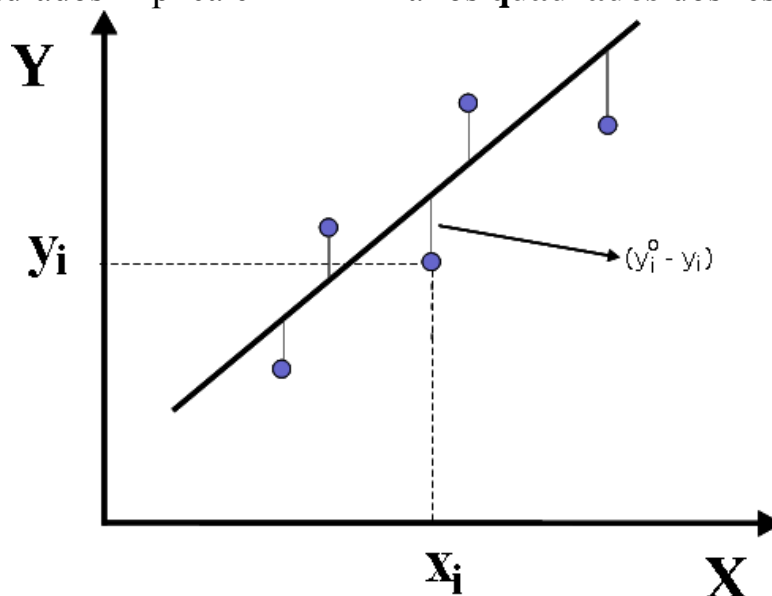
$$S \equiv \sum_{i=1}^N (y_i^o - y_i)^2 \quad (1)$$

onde chamamos de

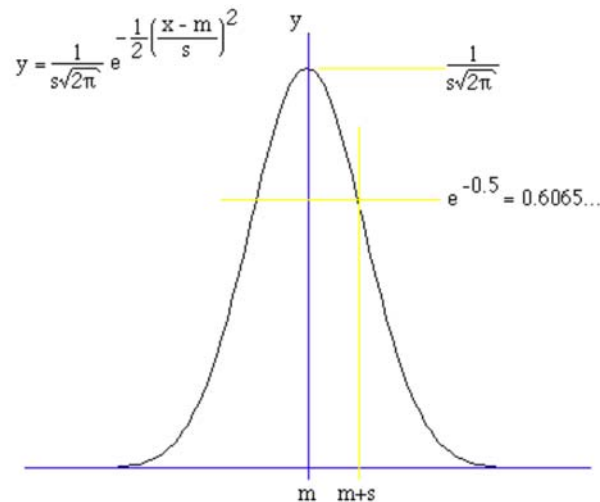
$y_i^o$  = valores observados de y

$y_i$  = valores calculados de y

ou seja, mínimos quadrados implica em minimizar os **quadrados** dos resíduos.



Por que este critério é considerado um bom critério e não simplesmente minimizar os resíduos ou o cubo dos resíduos? A resposta formal é que os mínimos quadrados são corretos se os resíduos tiverem uma distribuição gaussiana (normal).



Distribuição gaussiana (normal) para uma variável  $x$ , com média  $m$  e desvio padrão  $s$ . É simples notar que se minimizarmos os resíduos diretamente, um grande resíduo negativo pode ser anulado por um grande resíduo positivo, enquanto que com o quadrado minimizamos os módulos das diferenças.

Suponhamos que temos um conjunto de dados  $y$  com uma distribuição normal:

$$P(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y - \bar{y})^2 \right]$$

onde

$P(y)$  = probabilidade de obter o valor  $y$

$y$  = quantidade a ser observada

$\bar{y}$  = valor médio de  $y$

$\sigma$  = desvio padrão de  $y$

Por exemplo,

a probabilidade de se encontrar uma medida entre  $-\sigma$  e  $+\sigma$  é de 68,3%,

a probabilidade de se encontrar uma medida entre  $-1,5\sigma$  e  $+1,5\sigma$  é de 86,6%,

a probabilidade de se encontrar uma medida entre  $-2\sigma$  e  $+2\sigma$  é de 95,4%,

a probabilidade de se encontrar uma medida entre  $-2,5\sigma$  e  $+2,5\sigma$  é de 98,76%,

a probabilidade de se encontrar uma medida entre  $-3\sigma$  e  $+3\sigma$  é de 99,74%.

Suponhamos que medimos o valor de  $y$  várias vezes, obtendo uma série de valores  $\{y_i\}$ . A probabilidade de observar este conjunto é dada por

$$\begin{aligned}
 P(y_1, y_2, y_3, \dots, y_N) &= \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_1 - \bar{y})^2 \right] \right\} \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_2 - \bar{y})^2 \right] \right\} \dots \\
 &\dots \left\{ \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} (y_N - \bar{y})^2 \right] \right\} \\
 &= \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2 \right]
 \end{aligned}$$

Queremos agora saber qual é o melhor valor de  $\bar{y}$ , o valor médio. O melhor valor será aquele que **maximiza a probabilidade de obter os valores observados**  $\{y_i\}$ , ou seja, o melhor valor de  $\bar{y}$  é obtido colocando a derivada da probabilidade como nula:

$$\frac{d}{d\bar{y}} [P(y_1, y_2, y_3, \dots, y_N)] = 0$$

Ou seja

$$\begin{aligned}
 \frac{d}{d\bar{y}} \left\{ \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2 \right] \right\} &= 0 \\
 \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^N \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2 \right] \frac{d}{d\bar{y}} \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \bar{y})^2 \right] &= 0
 \end{aligned}$$

Como o termo  $\exp[\dots]$  não pode ser nulo, obtemos

$$\frac{d}{d\bar{y}} \left[ \sum_{i=1}^N (y_i - \bar{y})^2 \right] = 0$$

ou na nossa notação

$$S = \sum_{i=1}^N (y_i - \bar{y})^2$$

$$\boxed{\frac{dS}{d\bar{y}} = 0}$$

Continuando com a derivação, obtemos:

$$0 = \frac{d}{d\bar{y}} \left[ \sum_{i=1}^N (y_i - \bar{y})^2 \right]$$

$$= \sum_{i=1}^N -2(y_i - \bar{y})$$

$$= \sum_{i=1}^N y_i - \sum_{i=1}^N \bar{y}$$

$$N\bar{y} = \sum_{i=1}^N y_i$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

que é a média simples, como queríamos.

O próximo passo é simplesmente reconhecer que  $y$  pode ser uma função, [por exemplo](#):

$$y_i = a + b x_i$$

de modo que

$$S = \sum_{i=1}^N (y_i^0 - a - b x_i^0)^2$$

Minimizando  $S$  em relação a  $a$  e  $b$ , obtemos:

$$\frac{dS}{da} = \sum_{i=1}^N -2(y_i^0 - a - b x_i^0) = 0$$

$$\frac{dS}{db} = \sum_{i=1}^N -2x_i^0 (y_i^0 - a - b x_i^0) = 0$$

ou as duas condições:

$$\sum_{i=1}^N y_i^0 - Na - b \sum_{i=1}^N x_i^0 = 0$$

$$\sum_{i=1}^N x_i^0 y_i^0 - a \sum_{i=1}^N x_i^0 - b \sum_{i=1}^N (x_i^0)^2 = 0$$

Em notação matricial podemos escrever as duas condições como:

$$\begin{pmatrix} N & \sum_{i=1}^N x_i^0 \\ \sum_{i=1}^N x_i^0 & \sum_{i=1}^N (x_i^0)^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^N y_i^0 \\ \sum_{i=1}^N x_i^0 y_i^0 \end{pmatrix}$$

Para simplificar a notação, vamos definir os colchetes:

$$\sum_{i=1}^N x_i \equiv [x]$$

$$\sum_{i=1}^N 1 = N \equiv [1]$$

desta forma, nossa equação matricial se escreve como:

$$\begin{pmatrix} [1] & [x] \\ [x] & [x^2] \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} [y] \\ [xy] \end{pmatrix}$$

Estas equações são chamadas de equações normais. Elas podem ser resolvidas com a matriz inversa:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} [1] & [x] \\ [x] & [x^2] \end{pmatrix}^{-1} \begin{pmatrix} [y] \\ [xy] \end{pmatrix}$$

Para sabermos se, quando passamos de uma fitagem de uma reta para uma fitagem de uma parábola, a redução no  $S \equiv \chi^2 \equiv \sigma^2$  é suficiente para que o termo quadrático seja

significativo, podemos definir um parâmetro

$$\lambda = \frac{\sigma_{\text{reta}}^2 - \sigma_{\text{parab}}^2}{\sigma_{\text{parab}}^2} (N - 3)$$

e determinar o nível de confiabilidade que podemos descartar a hipótese do termo quadrático ser nulo por

$$\lambda = F_p(1, n - 3)$$

onde a distribuição F é dada por

$$F(a, b) = \frac{\chi^2(a)/a}{\chi^2(b)/b}$$

[Brownlee, K.A. 1960, Statistical theory & methodology in science and engineering, Wiley]

## Mínimos quadrados lineares

Os mínimos quadrados são lineares quando podemos resolver as equações normais usando álgebra linear. O exemplo dado acima é um mínimo quadrado linear, pois podemos resolver as equações para  $a$  e  $b$  exatamente.

## Mínimos quadrados não lineares

Muitos problemas interessantes não podem ser resolvidos linearmente. Por exemplo, qual é o melhor valor de  $\alpha$  em

$$y = \exp(-\alpha x)$$

Pela nossa definição de  $S$ :

$$S = \sum_{i=1}^N [y_i^0 - \exp(-\alpha x_i)]^2$$

e quando minimizamos  $S$ :

$$0 = \frac{dS}{d\alpha} = \sum_{i=1}^N 2[y_i^0 - \exp(-\alpha x_i)] [-\exp(-\alpha x_i)] (-x_i)$$

ou seja

$$0 = \sum_{i=1}^N y_i^0 x_i \exp(-\alpha x_i) - \sum_{i=1}^N x_i \exp(-2\alpha x_i)$$

Que podemos escrever, na notação de colchetes, como:

$$0 = [xy \exp(-\alpha x)] - [x \exp(-2\alpha x)]$$

Esta equação não pode ser resolvida usando-se álgebra linear.

Precisamos utilizar técnicas diferentes. A técnica mais empregada e, de fato, a técnica que se utiliza quando chamamos de **mínimos quadrados não lineares**, é a linearização do problema.

A idéia, aplicada ao problema acima, é:

$$y = \exp(-\alpha x)$$

Escolhemos um valor inicial de  $\alpha$ , chamado  $\alpha_0$ , e definimos:

$$\alpha = \alpha_0 + \Delta \alpha$$

Definindo

$$y_0 = \exp(-\alpha_0 x)$$

Em primeira ordem, isto é, linear:

$$y = y_0 + \left. \frac{dy}{d\alpha} \right|_{\alpha_0} \Delta \alpha$$

$$y = \exp(-\alpha_0 x) - x \exp(-\alpha_0 x) \Delta \alpha$$

Agora  $y$  é linear em  $\Delta \alpha$  e podemos usar os mínimos quadrados lineares para encontrar a correção  $\Delta \alpha$ :

$$S \equiv \sum_{i=1}^N (y_i^0 - y_i)^2$$

que se torna

$$S \equiv \sum_{i=1}^N \left( y_i^0 - y_{0,i} - \left. \frac{dy}{d\alpha} \right|_{\alpha_0} \Delta \alpha \right)^2$$

que minimizando

$$0 = \frac{dS}{d(\Delta \alpha)} = \sum_{i=1}^N -2 \left. \frac{dy}{d\alpha} \right|_{\alpha_0} \left( y_i^0 - y_{0,i} - \left. \frac{dy}{d\alpha} \right|_{\alpha_0} \Delta \alpha \right)$$

$$0 = \sum_{i=1}^N \left[ y_i^0 \left. \frac{dy}{d\alpha} \right|_{\alpha_0} - y_{0,i} \left. \frac{dy}{d\alpha} \right|_{\alpha_0} - \left( \left. \frac{dy}{d\alpha} \right|_{\alpha_0} \right)^2 \Delta \alpha \right]$$

ou, na notação dos colchetes:

$$0 = \left[ y^0 \frac{dy}{d\alpha} \right] - \left[ y_0 \frac{dy}{d\alpha} \right] - \left[ \left( \frac{dy}{d\alpha} \right)^2 \Delta \alpha \right]$$

Que podemos resolver para  $\Delta \alpha$ :

$$\Delta \alpha = \frac{\left[ y^0 \frac{dy}{d\alpha} \right] - \left[ y_0 \frac{dy}{d\alpha} \right]}{\left[ \left( \frac{dy}{d\alpha} \right)^2 \right]}$$

e finalmente obter o valor revisado de  $\alpha$ :

$$\alpha = \alpha_0 + \Delta \alpha$$

Note entretanto que o valor revisado de  $\alpha$  não é o melhor valor de  $\alpha$ , mas somente uma melhor aproximação do que  $\alpha_0$ . Isto ocorre porque  $\Delta \alpha$  é a solução do problema linearizado e **não** do problema real. Portanto, **precisamos iterar**, isto é, utilizar este valor revisado de  $\alpha$  como um novo  $\alpha_0$  e obter um novo valor revisado.

## Formulação Geral

Se a função  $y$  for uma função de  $k$  parâmetros que queremos ajustar:

$$y_i = y(x_i, a_1, a_2, \dots, a_k)$$

colocamos

$$\begin{aligned} a_1 &= a_{1,0} + \Delta a_1 \\ a_2 &= a_{2,0} + \Delta a_2 \\ &\vdots \\ a_k &= a_{k,0} + \Delta a_k \end{aligned}$$

com a hipótese de que  $\Delta a_i \ll a_i, 0$ .

Então podemos linearizar

$$\begin{aligned} y_i = & y(x_i, a_1, 0, a_2, 0, \dots, a_k, 0) + \frac{dy}{da_1} \Big|_{a_n=a_{n,0}} \Delta a_1 + \frac{dy}{da_2} \Big|_{a_n=a_{n,0}} \Delta a_2 + \dots \\ & + \dots + \frac{dy}{da_k} \Big|_{a_n=a_{n,0}} \Delta a_k \end{aligned}$$

notando que as derivadas são calculadas para todos  $a_n = a_{n,0}$ .



Chamando

$$y_{i,0} = y(x_i, a_{1,0}, a_{2,0}, \dots, a_{k,0})$$

e

$$\left. \frac{dy_i}{da_j} \right|_{a_0} \Delta a_j = \left. \frac{dy_i}{da_j} \right|_{a_n=a_{n,0}} \Delta a_j$$

podemos escrever

$$y_i = y_{i,0} + \sum_{j=1}^k \left. \frac{dy_i}{da_j} \right|_{a_0} \Delta a_j$$

onde o subscrito  $i$  significa calculado no ponto  $x_i$ .

Podemos agora calcular  $S$ :

$$S \equiv \sum_{i=1}^N (y_i^0 - y_i)^2$$

$$S \equiv \sum_{i=1}^N \left( y_i^0 - y_{i,0} - \sum_{j=1}^k \left. \frac{dy_i}{da_j} \right|_{a_0} \Delta a_j \right)^2$$

que minimizando com respeito a  $\Delta a_m$ :

$$0 = \frac{dS}{d(\Delta a_m)} = \sum_{i=1}^N 2 \left( y_i^0 - y_{i,0} - \sum_{j=1}^k \left. \frac{dy_i}{da_j} \right|_{a_0} \Delta a_j \right) \left( - \left. \frac{dy_i}{da_m} \right|_{a_0} \right)$$

$$0 = \sum_{i=1}^N (y_i^0 - y_{i,0}) \left( \left. \frac{dy_i}{da_m} \right|_{a_0} \right) - \sum_{j=1}^k \left. \frac{dy_i}{da_j} \right|_{a_0} \Delta a_j \left( \sum_{i=1}^N \left. \frac{dy_i}{da_m} \right|_{a_0} \right)$$

que na notação dos colchetes pode ser escrita como:

$$0 = \left[ (y^0 - y_0) \frac{dy}{da_m} \right] - \sum_{j=1}^k \left[ \frac{dy}{da_m} \frac{dy}{da_j} \right] \Delta a_j$$

Ou, em notação matricial

$$\begin{pmatrix} \left[ \frac{dy}{da_1} \frac{dy}{da_1} \right] & \left[ \frac{dy}{da_1} \frac{dy}{da_2} \right] & \dots & \left[ \frac{dy}{da_1} \frac{dy}{da_k} \right] \\ \left[ \frac{dy}{da_2} \frac{dy}{da_1} \right] & \left[ \frac{dy}{da_2} \frac{dy}{da_2} \right] & \dots & \left[ \frac{dy}{da_2} \frac{dy}{da_k} \right] \\ \vdots & \vdots & & \vdots \\ \left[ \frac{dy}{da_k} \frac{dy}{da_1} \right] & \left[ \frac{dy}{da_k} \frac{dy}{da_2} \right] & \dots & \left[ \frac{dy}{da_k} \frac{dy}{da_k} \right] \end{pmatrix} \begin{pmatrix} \Delta a_1 \\ \Delta a_2 \\ \vdots \\ \Delta a_k \end{pmatrix} = \begin{pmatrix} \left[ (y^0 - y_0) \frac{dy}{da_1} \right] \\ \left[ (y^0 - y_0) \frac{dy}{da_2} \right] \\ \vdots \\ \left[ (y^0 - y_0) \frac{dy}{da_k} \right] \end{pmatrix}$$

Esta equação matricial pode agora ser revolvida por álgebra matricial para encontrar as

correções  $\Delta a_m$ .

No caso linear

$$\begin{pmatrix} \begin{bmatrix} \frac{dy}{da_1} & \frac{dy}{da_1} \\ \frac{dy}{da_2} & \frac{dy}{da_1} \\ \vdots & \vdots \\ \frac{dy}{da_k} & \frac{dy}{da_1} \end{bmatrix} & \begin{bmatrix} \frac{dy}{da_1} & \frac{dy}{da_2} \\ \frac{dy}{da_2} & \frac{dy}{da_2} \\ \vdots & \vdots \\ \frac{dy}{da_k} & \frac{dy}{da_2} \end{bmatrix} & \cdots & \begin{bmatrix} \frac{dy}{da_1} & \frac{dy}{da_k} \\ \frac{dy}{da_2} & \frac{dy}{da_k} \\ \vdots & \vdots \\ \frac{dy}{da_k} & \frac{dy}{da_k} \end{bmatrix} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} (y^o - y_0) \frac{dy}{da_1} \\ (y^o - y_0) \frac{dy}{da_2} \\ \vdots \\ (y^o - y_0) \frac{dy}{da_k} \end{pmatrix}$$

## Determinação das incertezas

A maneira correta de determinar as incertezas nos parâmetros é calculando a variância de um parâmetro qualquer  $z$ :

$$\sigma_z^2 \equiv \langle (\Delta z)^2 \rangle$$

onde o quadrado é necessário pelas mesmas considerações do valor de  $S$ ,  $\langle a \rangle$  significa a média e onde

$$\Delta z = z_{\text{calculado}} - z_{\text{observado}}$$

Agora suponhamos que  $z$  seja uma função de duas variáveis  $x$  e  $y$ ,  $z = z(x, y)$ . Podemos expandir por série de Taylor [Brook Taylor (1685-1731), *Methodus incrementorum directa et inversa* (1715)]:

$$\Delta z = \frac{dz}{dx} \Delta x + \frac{dz}{dy} \Delta y$$

de onde obtemos:

$$\begin{aligned} \sigma_z^2 \equiv \langle (\Delta z)^2 \rangle &= \left\langle \left( \frac{dz}{dx} \Delta x + \frac{dz}{dy} \Delta y \right)^2 \right\rangle \\ &= \left\langle \left( \frac{dz}{dx} \right)^2 (\Delta x)^2 + 2 \frac{dz}{dx} \frac{dz}{dy} \Delta x \Delta y + \left( \frac{dz}{dy} \right)^2 (\Delta y)^2 \right\rangle \end{aligned}$$

Se as variáveis  $x$  e  $y$  são separáveis, podemos reduzir a equação acima a

$$\sigma_z^2 = \left( \frac{dz}{dx} \right)^2 \langle (\Delta x)^2 \rangle + 2 \frac{dz}{dx} \frac{dz}{dy} \langle \Delta x \Delta y \rangle + \left( \frac{dz}{dy} \right)^2 \langle (\Delta y)^2 \rangle$$

e, por definição:

$$\sigma_x^2 \equiv \langle (\Delta x)^2 \rangle$$

$$\sigma_y^2 \equiv \langle (\Delta y)^2 \rangle$$

$$\sigma_{xy}^2 \equiv \langle \Delta x \Delta y \rangle$$

de modo que

$$\sigma_z^2 = \left( \frac{dz}{dx} \right)^2 \sigma_x^2 + 2 \frac{dz}{dx} \frac{dz}{dy} \sigma_{xy}^2 + \left( \frac{dz}{dy} \right)^2 \sigma_y^2$$

E como obtemos  $\sigma_x$ ,  $\sigma_y$  e  $\sigma_{xy}$ ? Definindo a matriz covariância.

## Matrix Covariância

Definimos a matriz covariância como

$$COV = \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y^2 \end{pmatrix}$$

De modo que, se a [equação normal matricial](#) que usamos para achar os coeficientes é escrita como

$$\mathbf{MA} = \mathbf{Y}$$

de modo que a solução é dada por

$$\mathbf{A} = \mathbf{M}^{-1}\mathbf{Y}$$

onde  $\mathbf{M}^{-1}$  é a matriz inversa da matriz  $\mathbf{M}$ , vemos que a matriz covariância é dada por

$$COV(\mathbf{A}) = \sigma^2 \mathbf{M}^{-1}$$

onde

$$\sigma^2 = \frac{S}{N - k} = \frac{\sum_{i=1}^N (y_i^o - y_i)^2}{N - k}$$

Desta maneira é muito fácil calcular as incertezas nos parâmetros, pois somente precisamos multiplicar a matriz inversa que usamos para obter os valores dos parâmetros, por  $\sigma^2$ , obtendo a matriz covariância.

$\chi^2$

Mas o que fazemos **se a função a ser fitada não é analítica**, como por exemplo, um espectro resultante de um modelo de atmosferas? Er-Ho Zhang, Edward L. Robinson e R. Edward

Nather (1986, *Astrophysical Journal*, 305, 740) demonstraram que no caso **medirmos vários parâmetros simultaneamente, onde a mudança de alguns pode gerar mudanças significativas, enquanto que a mudança de outros parâmetros gera pequenas diferenças** e, ainda, que os parâmetros não são independentes, isto é, estão correlacionados, precisamos maximizar a probabilidade (*likelihood*) de que os parâmetros  $x_1, \dots, x_k$  sejam medidos pelos valores  $x_1^0, \dots, x_k^0$ , definindo incertezas  $\epsilon_i = x_i^0 - x_i$ ,

$$f(\epsilon_1, \dots, \epsilon_k) = \frac{1}{(2\pi)^k/2 \sigma_1 \dots \sigma_k} \exp \left[ -\frac{1}{2} \left( \frac{\epsilon_1^2}{\sigma_1^2} + \dots + \frac{\epsilon_k^2}{\sigma_k^2} \right) \right]$$

e, portanto, **maximizarmos a probabilidade é equivalente a minimizarmos**

$$\chi^2 \equiv \tilde{S} \equiv \frac{\epsilon_1^2}{\sigma_1^2} + \dots + \frac{\epsilon_k^2}{\sigma_k^2}$$

Note que o valor de  $\tilde{S}$  tem normalização diferente da de  $S$  definido na equação (1).

Se medimos os valores  $I_i^0$ , por exemplo, o espectro, e calculamos  $I_i$  com os parâmetros  $x_1, \dots, x_k$ , se assumirmos que a distribuição de erros é normal e que todas as variâncias são iguais ( $\sigma_1 = \sigma_2 = \dots = \sigma_k = W$ ), obtemos

$$\tilde{S} = \frac{\sum_i^N (I_i^0 - I_i)^2}{W^2}$$

de modo que, se  $\tilde{S}_0$  é o valor mínimo de  $\tilde{S}$ , e se mudarmos o valor do parâmetro  $i$  por um delta  $d$ , mantendo todos os outros parâmetros constantes, obteremos um novo valor de  $\tilde{S}$

$$\tilde{S} = \tilde{S}_0 + d^2 / \sigma_{ii}$$

e, portanto,

$$\sigma_i^2 = \sigma_{ii} = \frac{d^2}{\tilde{S} - \tilde{S}_0} \quad (2)$$

Portanto **podemos obter a incerteza em cada parâmetro** minimizando  $\tilde{S}$  com todos os parâmetros livres, para obter  $\tilde{S}_0$ , fixando o valor do parâmetro  $x_i$  para o valor que minimizou

$\tilde{S}$  mais um delta, digamos 5%, e minimizando novamente  $\tilde{S}$  com este valor de  $x_i$  fixo, obtendo um novo  $\tilde{S}$ . Com a equação (2) podemos então estimar o valor de sua incerteza, e repetir o processo para os outros parâmetros.

Podemos encontrar o termo de correlação usando uma transformação de variáveis

$$d_{ij}' = (d_i + d_j) / \sqrt{2}$$

$$d_{ij}' = (d_i - d_j)/\sqrt{2}$$

com  $d_i = d_j = d = d_{ij}/\sqrt{2}$ . Pela propagação de erros,

$$\sigma_{ij}'^2 = \frac{1}{2}(\sigma_i^2 + 2\sigma_{ij} + \sigma_j^2)$$

de modo que

$$\sigma_{ij} = \sigma_{ij}'^2 - \frac{1}{2}(\sigma_i^2 + \sigma_j^2) = \frac{d^2}{\tilde{S} - \tilde{S}_0} - \frac{1}{2}(\sigma_i^2 + \sigma_j^2)$$

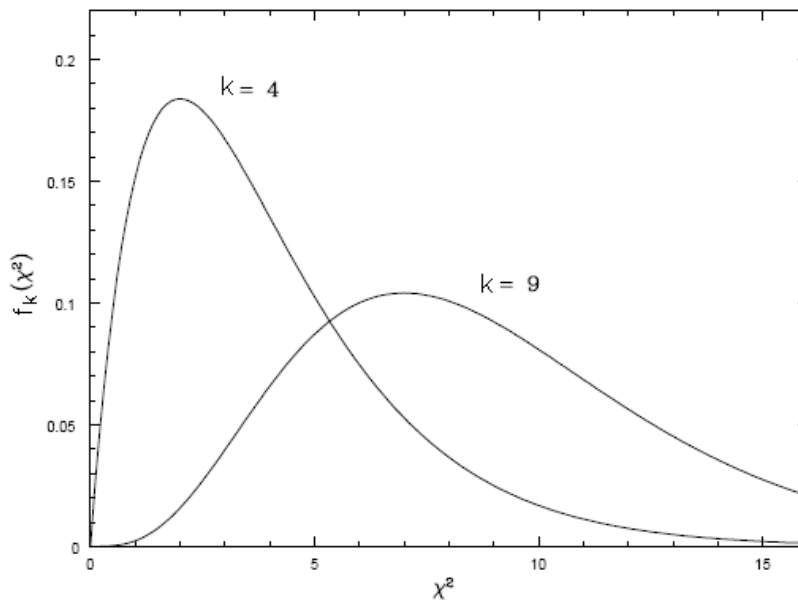
Se normalização não for unitária, devemos usar:

$$\sigma_i^2 = \sigma_{ii} = \frac{\tilde{S}}{n - k + 1} \frac{d^2}{\tilde{S} - \tilde{S}_0} \quad (1)$$

onde  $n$  é o número de pontos,  $k$  o número de parâmetros fitados, e o  $+1$  é porque um dos parâmetros foi mantido fixo, reduzindo o número de graus de liberdade.

A distribuição de probabilidades  $\chi^2$  para  $k$  graus de liberdade é dada por

$$f_k(\chi^2) = \frac{(\chi^2)^{(k-2)/2}}{2^{k/2} (k/2 - 1)!} \exp\left[-\frac{1}{2}\chi^2\right]$$



Note que o valor médio de  $\chi^2$  para  $k$  graus de liberdade é  $k$ . Algumas vezes se usa um  $\chi^2$  reduzido de modo que sua média seja 1:

$$\chi_{\text{red}}^2 \equiv \frac{1}{k}\chi^2 = \frac{1}{k} \sum_i^k \frac{\epsilon_i^2}{\sigma_i^2}$$

Sabemos que o teorema do limite central garante que a distribuição  $\chi^2$  para  $k$  graus de

liberdade se aproxima da distribuição gaussiana para grandes valores de  $k$ . Esta aproximação é boa para  $k > 30$ .

## Estimativa Robusta

O grande problema do método de mínimos quadrados é a grande influência de pontos com resíduo muito grande, os *outliers*. Para se reduzir a influência destes resíduos muito grandes, pode-se minimizar funções que não crescem tão rapidamente quanto  $S^2$ , como por exemplo, minimizando a função discrepância, definida como:

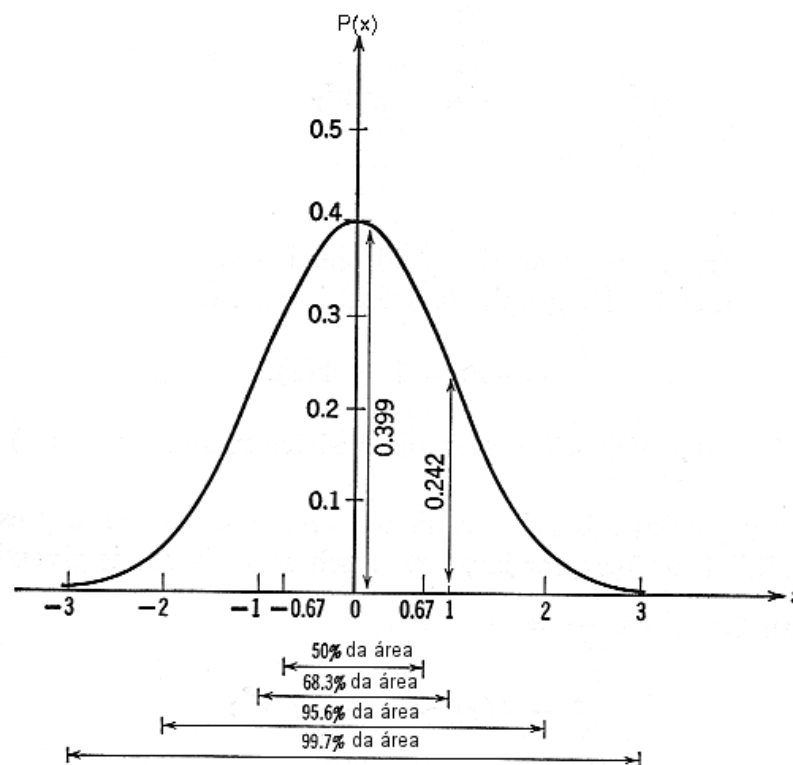
$$f(S) = \begin{cases} S^2 & \text{se } |S| \leq c \\ c^2 & \text{se } |S| > c \end{cases}$$

onde  $c$  é uma constante.

## Probabilidade

A função normal, centrada em  $x=0$  e com largura em  $1/e$  em  $x=1$  e  $x=-1$ , é definida como

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$



e sua integral é a distribuição normal

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}y^2} dy$$

Com esta definição

$$\int_{-\infty}^{+\infty} P(x) dx = 1$$

Se tivermos uma distribuição centrada em  $x_o$ , e de largura  $\sigma^2$ ,

$$P(x - x_o) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_o)^2}{2\sigma^2}}$$

O valor da integral de x até infinito, centrada em 0 e com desvio padrão 1, é

x=0	P(>x)=0,5
x=2	P(>x)=0,02275
x=3	P(>x)=0,00135
x=4	P(>x)=0,00003167
x=5	P(>x)=2,867 × 10 <sup>-7</sup>
x=6	P(>x)=9,866 × 10 <sup>-10</sup>
x=7	P(>x)=1,280 × 10 <sup>-12</sup>
x=8	P(>x)=6,221 × 10 <sup>-16</sup>
x=9	P(>x)=1,129 × 10 <sup>-19</sup>

Se a distribuição é discreta, com N valores:

$$\sum_{i=0}^N P(x_i) \equiv 1$$

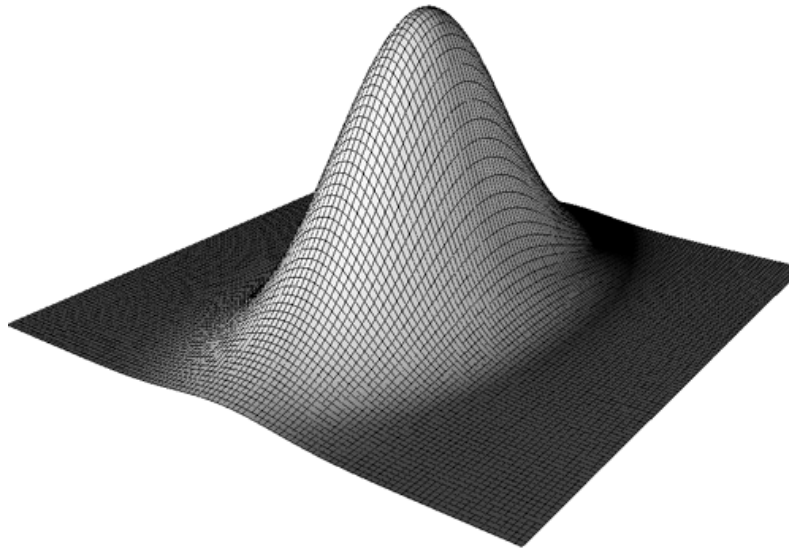
de modo que precisamos normalizar nossas probabilidades

$$P(x - x_o) = \frac{1}{\sum_{i=0}^N P(x_i)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-x_o)^2}{2\sigma^2}}$$

## Distribuição Binominal

Seja um problema que só admite duas soluções, como ao jogarmos uma moeda. Chamemos de  $b(k;n,p)$  a probabilidade que n tentativas com probabilidade intrínseca de sucesso p, e  $q=1-p$  a probabilidade intrínseca de insucesso, resulte em k sucessos e n-k insucessos. Esta probabilidade é dada pela distribuição binominal

$$b(k; n, p) = \binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{npq}} P\left(\frac{k - np}{\sqrt{npq}}\right)$$



Distribuição gaussiana (normal) para duas variáveis não correlacionadas.  
 Para duas variáveis correlacionadas,  $x_1$  medida a partir de  $\mu_1$  e  $x_2$  medida a partir de  $\mu_2$ :

$$P(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \exp\left[-\frac{z}{2(1-\rho^2)}\right]$$

onde

$$z = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}$$

$$\rho = \text{cor}(x_1, x_2) = \frac{\sigma_{1,2}}{\sigma_1\sigma_2}$$

[Variança da média](#)

[Pesos nas medidas](#)

[Cálculo de Distribuição Normal](#)

[Outras Distribuições](#)

Data analysis recipes: Fitting a model to data. Hogg et al. 2010. [arXiv:1008.4686](#)

Tutorial sobre diferentes métodos de ajuste linear a dados, incluindo exercícios. Fornece uma visão crítica sobre os métodos comumente adotados. É interessante por exemplo a crítica dos autores à metodologia usada para derivar a relação de Tully-Fisher. E a crítica ao uso de PCA.

Error estimation in astronomy: A guide. Andrae. 2010. [arXiv:1009.2755](#) Tutorial sobre estimativa das incertezas em parâmetros baseados em modelos (ou não).

[Dos and don'ts of reduced chi-squared](#)

No Excel, mínimos quadrados lineares são fitados em Tools-Data Analysis-Regression. Se



este pacote não estiver instalado, ele pode ser adicionado com Tools-Add Ins-Analysis Tool Pack e Solver. [O OpenOffice Calc](#) também pode ser usado. Com uma [subrotina em Fortran de inversão de matrizes](#), se pode facilmente implementar um algoritmo de mínimos quadrados.

---

**Volta:** [Astronomia e Astrofísica](#)

© [Kepler de Souza Oliveira Filho](#)

*Modificada em 13 dez 2010*