

ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 4

Inferência estatística e ajuste de parâmetros

...quando não está ao nosso alcance discernir as opiniões mais verdadeiras, nós devemos seguir as mais prováveis... (R. Descartes, Discurso do Método, 1637)

Não se pode seleccionar os métodos de análise dos dados verificando se fornecem estimativas das grandezas que sejam próximas aos seus *valores verdadeiros*, porque estes são desconhecidos. Assim, o procedimento que adotaremos será propor formas de construir estimadores das grandezas e verificar as propriedades *estatísticas* das estimativas obtidas. Discutiremos neste capítulo os métodos da *Máxima Verossimilhança* e dos *Mínimos Quadrados*, que são muito usados nas ciências experimentais.

Nos restringiremos aqui ao estudo das estimativas de máxima verossimilhança da média e da variância de dados distribuídos de acordo com a função de probabilidade normal e ao ajuste, pelo método dos mínimos quadrados, de parâmetros lineares de funções modelo a dados experimentais com variâncias conhecidas. Nos capítulos 8 e 9 voltaremos a discutir esses métodos com mais detalhes.

4.1 O Método da Máxima Verossimilhança

Como o nome indica, o método se caracteriza por estimar os parâmetros como aqueles que mais provavelmente originariam os dados observados — ou seja, os mais *verossímeis*.

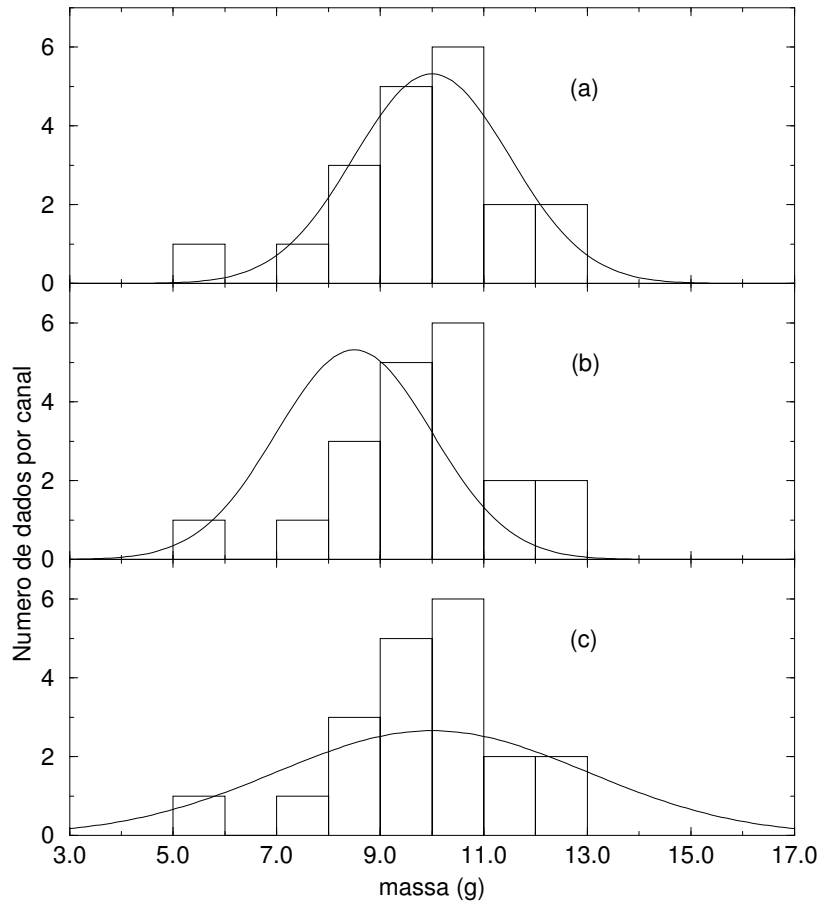


Figura 4.1: Três cópias do histograma de uma medida $\{x_i, i = 1, 2, \dots, 20\}$, às quais foram superpostas as funções: **(a)** $20 \cdot N(x; x_0 = 10, \sigma_0 = 1, 5)$. **(b)** $20 \cdot N(x; x_0 = 8, 5, \sigma_0 = 1, 5)$; **(c)** $20 \cdot N(x; x_0 = 10, \sigma_0 = 3)$.

Ha figura 4.1 estão representadas três gaussianas diferentes, convenientemente normalizadas, que poderiam originar a medida histogramada, que é a mesma em todas elas; dessas três, a da figura **4.1(a)** é mais verossímil que as demais.

Q4.1 *Explicite, o mais objetivamente possível, porque a f.d.p. da figura 4.1(a) é a mais verossímil das três.*

A probabilidade de observarmos o particular conjunto de dados $\{x_i\} \equiv \{x_i, i = 1, \dots, N\}$ quando a variável aleatória x é contínua é, estritamente falando, nula. No entanto, podemos calcular a probabilidade de obtermos um primeiro dado no intervalo $[x_1, x_1 + \Delta x]$, um segundo dado no intervalo $[x_2, x_2 + \Delta x]$, etc, como o produto das probabilidades de cada um deles,

$$\mathcal{P}(\{x_i\} | \mathbf{p}_0) = f(x_1 | \mathbf{p}_0)\Delta x \cdot \dots \cdot f(x_i | \mathbf{p}_0)\Delta x \cdot \dots \cdot f(x_N | \mathbf{p}_0)\Delta x \quad ,$$

onde $f(x_i | \mathbf{p}_0)$ é o valor, em $x = x_i$, da f.d.p. de x , cujos parâmetros verdadeiros estão representados pelo vetor \mathbf{p}_0 . Se a f.d.p. é normal, $\mathbf{p}_0 = (x_0, \sigma_0)$, sendo x_0 e σ_0 o valor esperado e o desvio padrão, respectivamente. Na comparação de diversas f.d.p.s f , mantém-se Δx constante, o que permite ignorar o fator $(\Delta x)^N$. Assim, a grandeza

$$\mathcal{P}(\{x_i\} | \mathbf{p}_0) \propto \prod_{i=1}^N f(x_i | \mathbf{p}_0) \quad (4.1)$$

é proporcional à probabilidade de obtenção do conjunto de dados. Este resultado, embora lógico e de interpretação indiscutível, não nos permite resolver o problema de estimar as grandezas simbolizadas por \mathbf{p}_0 , o que requer o procedimento a seguir.

Define-se a *Função Verossimilhança* a partir da fórmula (4.1), mas substituindo \mathbf{p}_0 pela variável \mathbf{p} , ou seja,

$$\mathcal{L}(\mathbf{p} | \{x_i\}) = \prod_{i=1}^N f(x_i | \mathbf{p}) \quad , \quad (4.2)$$

que é uma função de \mathbf{p} . O *Método da Máxima Verossimilhança* consiste em estimar \mathbf{p}_0 pelo vetor $\hat{\mathbf{p}}$ que maximiza $\mathcal{L}(\mathbf{p} | \{x_i\})$. Esse procedimento tem a interpretação qualitativa induzida pela discussão acima, relacionada à figura 4.1, mas não tem nenhuma outra justificativa *a priori*. Veremos no capítulo 9,

como justificativa *a posteriori* do método, que as estimativas baseadas nele têm certas propriedades convenientes, em particular, à proporção que o número de dados da medida aumenta, as f.d.p.s das estimativas obtidas tendem a ser normais e centradas nos valores verdadeiros e as estimativas tendem a ser as de menores incertezas que podem ser alcançadas em termos absolutos. No jargão da estatística, diz-se que o *estimador de Máxima Verossimilhança é assintoticamente normal, não tendencioso e eficiente*. Note que o termo “assintoticamente” aplica-se às três qualidades, ou seja, esse estimador é possivelmente tendencioso e não tem eficiência máxima quando o número de dados não for muito grande.

4.2 Estimativas da média e do desvio padrão

Na medição de uma grandeza, a média e o desvio-padrão informam sobre a posição e a localização da função densidade de probabilidade dos valores observados, sempre que ela for unimodal. Nesta seção, trataremos do caso particular, mas muito importante, em que se obtêm N dados, $\{x_i, i = 1, 2, \dots, N\} \equiv \{x_i\}$, todos estatisticamente independentes e provenientes de observações diretas de uma grandeza x , cujo valor verdadeiro é x_0 , com uma f.d.p. normal centrada nesse valor e com desvio-padrão σ_0 . Nessas condições, a função verossimilhança é

$$\begin{aligned} \mathcal{L}(x, \sigma | \{x_i\}) &= \prod_{i=1}^N f(x_i | x, \sigma) = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^N \exp \left\{ - \sum_{i=1}^N \frac{(x_i - x)^2}{2\sigma^2} \right\} . \end{aligned}$$

Como \mathcal{L} é definida positiva, maximizar $\ln(\mathcal{L})$ ao invés de \mathcal{L} é equivalente, porque ambas as funções têm máximo no mesmo ponto, e simplifica consideravelmente o cálculo.

Q4.2 *Demonstre que $\ln(\mathcal{L})$ e \mathcal{L} tem máximo no mesmo ponto \hat{x} , $\hat{\sigma}$. Mostre que esse resultado é geral, valendo mesmo quando a f.d.p. f não é gaussiana e para qualquer número de parâmetros.*

Q4.3 *Mostre que*

$$\hat{x} = \frac{1}{N} \sum_{i=1}^N x_i = \bar{x} \quad e \quad (4.3)$$

4.3. CORREÇÃO DA TENDENCIOSIDADE DA ESTIMATIVA DA VARIÂNCIA 103

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (\text{diferente da estimativa usual}) \quad , \quad (4.4)$$

resolvendo o sistema de equações em \hat{x} , $\hat{\sigma}$:

$$\left. \frac{\partial \ln(\mathcal{L})}{\partial x} \right|_{\hat{x}, \hat{\sigma}} = 0 \quad \text{e}$$

$$\left. \frac{\partial \ln(\mathcal{L})}{\partial \sigma} \right|_{\hat{x}, \hat{\sigma}} = 0 \quad .$$

4.3 Correção da tendenciosidade da estimativa da variância.

O valor esperado da estimativa \bar{x} de x_0 pode ser calculado a partir da expressão algébrica da fórmula (4.3) e das propriedades da f.d.p. dos dados, f . Quando o valor esperado de uma variável α se escreve $\langle \alpha \rangle$, obtemos

$$\langle \bar{x} \rangle = \frac{1}{N} \sum_{i=1}^N \langle x_i \rangle = \frac{1}{N} \sum_{i=1}^N x_0 = x_0 \quad ,$$

onde usamos o fato que $\langle x_i \rangle = x_0$, para qualquer i . Assim, embora não saibamos se nosso resultado \bar{x} de uma medida está próximo ou não de x_0 , sabemos que *em média* ele está correto. Estimativas que têm esta propriedade são chamadas de *não tendenciosas*. Em certas situações, o estimador de máxima verossimilhança tem esta propriedade, mas nem sempre, como no caso da estimativa da variância, $\hat{\sigma}^2$, obtida na seção anterior. A partir da fórmula (4.4), calcula-se

$$\langle \hat{\sigma}^2 \rangle = \frac{1}{N} \left\langle \sum_{i=1}^N (x_i - \bar{x})^2 \right\rangle \quad .$$

Quando ϵ é a diferença entre a média e o valor verdadeiro,

$$\bar{x} = x_0 + \epsilon \quad , \quad (4.5)$$

podemos reescrever cada termo da somatória como

$$(x_i - \bar{x})^2 = (x_i - x_0)^2 - 2\epsilon(x_i - x_0) + \epsilon^2 \quad .$$

Substituindo este resultado na somatória acima, obtemos

$$\begin{aligned} \langle \hat{\sigma}^2 \rangle &= \frac{1}{N} \left\langle \sum_{i=1}^N (x_i - x_0)^2 - 2\epsilon \sum_{i=1}^N (x_i - x_0) + \sum_{i=1}^N \epsilon^2 \right\rangle \\ \langle \hat{\sigma}^2 \rangle &= \frac{1}{N} \left\langle \sum_{i=1}^N (x_i - x_0)^2 - 2\epsilon(N\bar{x} - Nx_0) + N\epsilon^2 \right\rangle \\ \langle \hat{\sigma}^2 \rangle &= \frac{1}{N} \left\langle \sum_{i=1}^N (x_i - x_0)^2 - N\epsilon^2 \right\rangle \\ \langle \hat{\sigma}^2 \rangle &= \frac{1}{N} \left\langle \sum_{i=1}^N (x_i - x_0)^2 \right\rangle - \langle \epsilon^2 \rangle . \end{aligned}$$

Nesta expressão, identifica-se a variância $\langle (x_i - x_0)^2 \rangle = \sigma_0^2$ e, considerando a fórmula (4.5) acima, verifica-se que o último termo corresponde ao valor quadrático médio da diferença entre a média e o valor verdadeiro,

$$\langle \epsilon^2 \rangle = \langle (\bar{x} - x_0)^2 \rangle = \sigma_m^2 = \frac{\sigma_0^2}{N} ,$$

que é a variância da média. Finalizando o cálculo, obtém-se

$$\langle \hat{\sigma}^2 \rangle = \frac{1}{N} (N\sigma_0^2) - \frac{\sigma_0^2}{N} = \frac{N-1}{N} \sigma_0^2 . \quad (4.6)$$

A estimativa obtida através da Máxima Verossimilhança para a variância é, portanto, *tendenciosa*, desde que em *média* essa estimativa *subestima* o valor verdadeiro. Note, porém, que para $N \gg 1$ a estimativa não é tendenciosa, uma vez que $(N-1)/N \rightarrow 1$ quando $N \rightarrow \infty$. Dizemos então que a estimativa de Máxima Verossimilhança é *assintoticamente* não tendenciosa. Toda a estimativa obtida pelo método da máxima verossimilhança tem essa propriedade, mas não é verdade que a estimativa seja não tendenciosa para qualquer N . Neste caso particular da variância, a tendenciosidade pode ser corrigida analiticamente, mas sempre é possível determinar essa correção numericamente.

Q4.4 Mostre que a estimativa da variância que usamos habitualmente,

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 , \quad (1.10)$$

não é tendenciosa qualquer que seja N . Assim, poderíamos chamá-la de “estimativa pelo método da Máxima Verossimilhança corrigida da tendenciosidade”.

Q4.5 Considere duas medidas da mesma grandeza y_0 , com desvios-padrões das médias verdadeiros σ e σ' diferentes, em que se obteve as médias y e y' .

a) Mostre que

$$\hat{y} = \frac{\frac{y}{\sigma^2} + \frac{y'}{\sigma'^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma'^2}}$$

é a estimativa de máxima verossimilhança de y_0 .

b) Mostre que \hat{y} é uma estimativa não-tendenciosa do valor verdadeiro y_0 , se y e y' são estimativas não-tendenciosas de y_0 .

c) Use a fórmula de cálculo da variância de uma função de variáveis aleatórias (seção 3.1) e mostre que a variância de \hat{y} é dada por $\sigma_{\hat{y}}^2$ da expressão

$$\frac{1}{\sigma_{\hat{y}}^2} = \frac{1}{\sigma^2} + \frac{1}{\sigma'^2}$$

Finalmente, note que a estimativa da variância da fórmula (1.10) será não tendenciosa independentemente da f.d.p. dos dados. O que mudará com a f.d.p. é a probabilidade associada ao intervalo de confiança com largura de $\pm 1\sigma$ em torno do valor da média, que é 68,3% somente para a normal (e com muitos dados, caso contrário, é preciso usar t de Student).

4.4 Exemplos de aplicação do Método da Máxima Verossimilhança

Na seção anterior, usamos o método da máxima verossimilhança para estimar x_0 e σ_0^2 a partir de dados $\{x_i\}$ que obedecem a f.d.p.s gaussianas e estatisticamente independentes. Nesta seção, vamos discutir algumas outras aplicações.

4.4.1 Dados normais correlacionados

Na seção 2.4, apresentamos a f.d.p. multinormal. Neste exemplo, vamos aplicar o método da máxima verossimilhança para calcular a média de dois dados

correlacionados, x_1 e x_2 , com matriz de covariância

$$\mathbf{V} = \sigma_0^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} .$$

A f.d.p. de x_1 e x_2 é (veja a seção 2.4)

$$f(x_1, x_2; x_0, \sigma_0) = \frac{1}{2\pi\sigma_0^2\sqrt{1-\rho^2}} \exp\left(-\frac{(x_1-x_0)^2 - 2\rho(x_1-x_0)(x_2-x_0) + (x_2-x_0)^2}{2\sigma_0^2(1-\rho^2)}\right) .$$

Supondo que σ_0 seja o valor verdadeiro e conhecido do desvio padrão e que apenas x_0 deva ser estimado, a função verossimilhança é

$$\mathcal{L}(x \mid x_1, x_2) = f(x_1, x_2; x) ,$$

em que x é a variável que substitui x_0 e o parâmetro σ_0 não aparece mais, por que é suposto conhecido. Derivando $\ln \mathcal{L}$ em relação a x e determinando o valor \hat{x} em que a derivada é nula, obtém-se

$$\hat{x} = \frac{x_1 + x_2}{2} ,$$

que é a média simples dos dois dados. A demonstração que essa estimativa é não tendenciosa vem de $\langle x_1 \rangle = \langle x_2 \rangle = x_0$.

A variância de \hat{x} pode ser calculada usando os resultados das seções 3.1 e 3.3 e vale

$$\sigma_{\hat{x}}^2 = \sigma^2 \frac{1+\rho}{2} .$$

Deixamos o caso em que as variâncias dos dados são diferentes para os exercícios no final do capítulo, bem como o caso de infinitos dados covariantes, que também será visto de outra forma na seção 8.15.

4.4.2 Grandezas distribuídas como Poisson

Considere que a variável n obedece à f.p. de Poisson $P_a(n)$. Quando se realiza uma única medição com o resultado n , a função verossimilhança é

$$\mathcal{L}(a \mid n) = \frac{e^{-a} a^n}{n!} .$$

4.4. EXEMPLOS DE APLICAÇÃO DO MÉTODO DA MÁXIMA VEROSSIMILHANÇA 107

Obtém-se a estimativa de máxima verossimilhança ao maximizar $\mathcal{L}(a | n)$ (ou seu logaritmo) e determinar o valor \hat{a} correspondente. Neste caso, obtém-se

$$\hat{a} = n \quad .$$

Essa estimativa é não tendenciosa (porque $\langle \hat{a} \rangle = \langle n \rangle = a$) e sua variância é $\sigma_{\hat{a}}^2 = a$. Note que, como não se conhece a , tudo que se pode fazer é aproximar $\sigma_{\hat{a}}^2 \cong n$, aproximação essa também não tendenciosa.

Como um segundo exemplo de medida de uma grandeza com distribuição de Poisson de média a , suponha que uma primeira medição apresente n_1 eventos e outra, n_2 eventos. Neste caso, a função verossimilhança é

$$\mathcal{L}(a | n_1, n_2) = \frac{e^{-a} a^{n_1}}{n_1!} \frac{e^{-a} a^{n_2}}{n_2!} \quad .$$

A estimativa de máxima verossimilhança é

$$\hat{a} = \frac{n_1 + n_2}{2}$$

com variância $\sigma_{\hat{a}}^2 = a/2$, que é aproximada por $(n_1 + n_2)/4$, pela mesma razão apresentada no exemplo anterior.

Como um terceiro exemplo, considere um experimento para determinar a intensidade s de uma fonte na presença de um ruído de fundo com intensidade f , ambas desconhecidas. Em uma primeira medição, observou-se n eventos provenientes da fonte e do ruído de fundo, que juntos têm intensidade $s + f$. Em uma segunda medição, a fonte foi removida, observando-se m eventos, portanto todos provenientes do ruído de fundo. A função verossimilhança é

$$\mathcal{L}(s, f | n, m) = \frac{e^{-(s+f)} (s+f)^n}{n!} \frac{e^{-f} f^m}{m!} \quad .$$

O primeiro fator do membro da direita corresponde à probabilidade de se observar n eventos provenientes de uma distribuição de Poisson com parâmetro $s + f$; o segundo fator, a uma distribuição com parâmetro f . Derivando $\ln \mathcal{L}$ em relação a s e a f e impondo a condição que essas derivadas sejam nulas em \hat{s} e \hat{f} , obtém-se

$$\begin{aligned} \hat{f} &= m \\ \hat{s} &= n - m \quad , \end{aligned}$$

que é o resultado esperado. As expressões acima permitem estimar a matriz de covariâncias de \hat{f} e \hat{s} a partir das variâncias de n e m .

Q4.6 Mostre que as estimativas \hat{f} e \hat{s} não são tendenciosas.

4.5 O Método dos Mínimos Quadrados

Este método consiste em estimar as grandezas a partir da minimização do somatório dos quadrados das diferenças entre os valores calculados e os experimentais. Ele é devido a Gauss e Legendre, no final do século XVIII e início do XIX, que o aplicaram na redução de dados de observações astronômicas. A prioridade de descoberta do método é uma questão interessante. Stigler [Stigler] sugere que Gauss descobriu o método antes de Legendre, que, porém, o apresentou em uma publicação, datada de 1805, que despertou o interesse dos astrônomos da época, algo que Gauss teria tentado vários anos antes, sem sucesso. Já Grosser [Grosser] relata que o método teria chamado a atenção quando da descoberta do asteroide Ceres, que ocorreu da seguinte maneira, de acordo com esse livro e também com a referência [Forbes].

Após a descoberta de Urano, acreditava-se que deveria haver um planeta entre Marte e Júpiter. No início de 1801, um possível planeta nessa região foi avistado antes de esconder-se sob o Sol. Nenhuma das órbitas determinadas para esse planeta, com base nessas observações, permitiram reencontrá-lo no céu. Gauss interessou-se pelo problema e, em 2 meses, desenvolveu o Método dos Mínimos Quadrados, com que previu a posição do objeto, que foi encontrado imediatamente, no final do ano. Esse objeto era Ceres, o maior dos planetas menores internos a Netuno, com uma órbita bastante excêntrica e que pertence ao cinturão de asteroides. Embora seja difícil ter certeza da razão pela qual os astrônomos erravam a previsão, é de se imaginar que escolhessem das observações aquelas que davam órbitas de acordo com suas expectativas. Esse episódio ilustra, então, a objetividade dos métodos estatísticos.

É mais fácil explicar o método no caso em que se ajustam apenas dois parâmetros lineares. Assim, adotaremos como hipótese que a relação entre a variável dependente y e as grandezas físicas procuradas, a e b , é linear, com a forma particular

$$y = a f(x) + b g(x) \quad (4.7)$$

e que se pode observar experimentalmente valores y_i para certos x_i ; normalmente, essa equação é a representação matemática de um modelo, por isso é frequentemente chamada de *função modelo*. A partir dos dados $\{(x_i, y_i, \sigma_i), i = 1, \dots, N\}$, pretende-se estimar os valores das grandezas a e b . A figura 4.2 ilustra os significados de x_i, y_i e σ_i .

As funções $f(x)$ e $g(x)$ são funções quaisquer, com a única restrição de não envolverem os parâmetros a e b . Podem ser, por exemplo, $\sin(x)$ e $\cos(x)$, ou

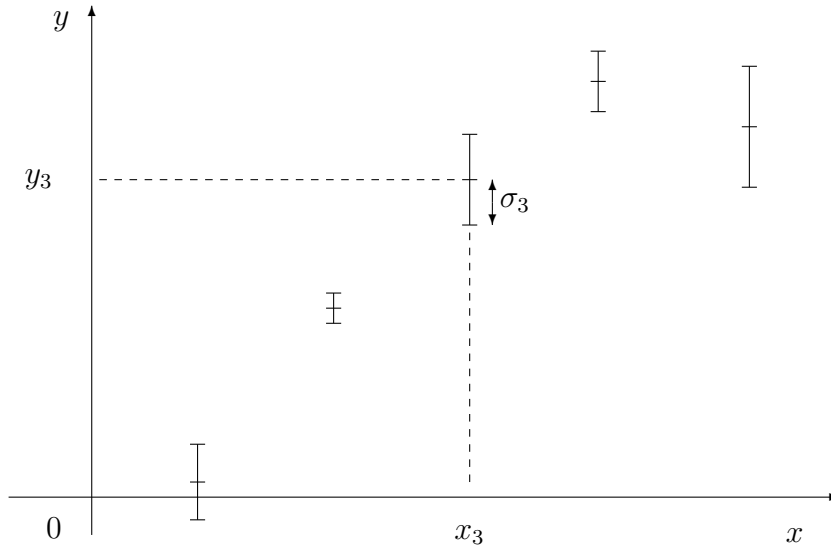


Figura 4.2: Representação de uma medida $\{(x_i, y_i, \sigma_i), i = 1, 2, \dots, 5\}$. As barras de incerteza representadas no gráfico são iguais a σ_i , de acordo com o procedimento convencional.

mesmo $\sin(kx)$ e $\cos(kx)$, com k uma constante conhecida e independente de a e b . De nenhuma maneira se exige linearidade de f e g em x . Veremos logo mais, na seção 4.8, que f e g não precisariam sequer serem funções de uma mesma e única variável x , podendo ser substituídas por grandezas numéricas conhecidas sem erro e diferentes para cada dado.

Vamos adotar que os parâmetros a e b têm valores verdadeiros iguais a a_0 e b_0 , respectivamente, e que as abscissas x_i das observações são conhecidas exatamente. Assim, os valores y_i observados relacionam-se com a variável independente x_i pela expressão

$$y_i = a_0 f(x_i) + b_0 g(x_i) + \epsilon_i \quad , \quad (4.8)$$

em que ϵ_i é o erro experimental, desconhecido. Além disso, consideramos a inexistência de erros sistemáticos, o que significa, como vimos no capítulo 1,

$$\langle \epsilon_i \rangle = 0 \quad . \quad (1.7)$$

De acordo com a definição de variância, relação (1.5), temos

$$\sigma_i^2 = \langle (\epsilon_i - \langle \epsilon_i \rangle)^2 \rangle \quad ,$$

que, pela ausência de erros sistemáticos, implica em

$$\sigma_i^2 = \langle \epsilon_i^2 \rangle . \quad (4.9)$$

É importante para o método que essas variâncias σ_i^2 sejam conhecidas ou que sejam todas iguais, o que deixaremos para explicar no capítulo 8. Vamos supor adicionalmente, por enquanto, que os dados não estejam correlacionados, ou seja,

$$\langle \epsilon_i \epsilon_j \rangle = 0 \quad , \quad i \neq j \quad .$$

Essa hipótese não é necessária para o método, mas torna os cálculos menos complicados; voltaremos ao assunto também no capítulo 8, onde mostraremos como proceder quando as covariâncias não são nulas, o que é relativamente frequente. Definimos a função $Q(a, b)$ neste exemplo particular como

$$Q(a, b) = \sum_{i=1}^N \frac{(y_i - af(x_i) - bg(x_i))^2}{\sigma_i^2} . \quad (4.10)$$

A estimativa (\hat{a}, \hat{b}) de (a_0, b_0) é obtida da condição de máximo de $Q(a, b)$, ou seja, calculam-se as derivadas parciais, que devem se anular simultaneamente no ponto de mínimo,

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

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

Q4.7 Mostre que a solução do sistema de equações pode ser obtida como

$$\hat{\mathbf{p}} = \mathbf{M}^{-1} \mathbf{D} \quad , \quad (4.11)$$

onde $\hat{\mathbf{p}}$ e \mathbf{D} são os vetores

$$\hat{\mathbf{p}} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} \sum_{i=1}^N \frac{y_i f(x_i)}{\sigma_i^2} \\ \sum_{i=1}^N \frac{y_i g(x_i)}{\sigma_i^2} \end{pmatrix} \quad (4.12)$$

e \mathbf{M} é a matriz

$$\mathbf{M} = \begin{pmatrix} \sum_{i=1}^N \frac{f(x_i)^2}{\sigma_i^2} & \sum_{i=1}^N \frac{f(x_i)g(x_i)}{\sigma_i^2} \\ \sum_{i=1}^N \frac{f(x_i)g(x_i)}{\sigma_i^2} & \sum_{i=1}^N \frac{g(x_i)^2}{\sigma_i^2} \end{pmatrix} . \quad (4.13)$$

Note que não foi adotada nenhuma hipótese sobre a forma da f.d.p. dos dados, apenas que suas variâncias são conhecidas. Em outras palavras, o MMQ pode ser usado qualquer que seja a f.d.p. dos dados.

Q4.8 Mostre que as matrizes necessárias para a determinação dos parâmetros de uma reta

$$y = a + bx \quad (4.14)$$

pelo MMQ são

$$\hat{\mathbf{p}} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \quad \mathbf{D} = \begin{pmatrix} \sum_{i=1}^N \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^N \frac{y_i x_i}{\sigma_i^2} \end{pmatrix} \quad (4.15)$$

$$\mathbf{M} = \begin{pmatrix} \sum_{i=1}^N \frac{1}{\sigma_i^2} & \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \\ \sum_{i=1}^N \frac{x_i}{\sigma_i^2} & \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} \end{pmatrix}. \quad (4.16)$$

É muito comum apresentar o MMQ com este exemplo, uma vez que as fórmulas são mais simples e tem inúmeras aplicações, mas pode levar quem estuda a pensar que somente coeficientes de polinômios podem ser ajustados pelo MMQ. A linearidade de y com os parâmetros a e b é fundamental para o MMQ, mas a forma polinomial é totalmente irrelevante para o método.

A forma particular da função de mérito Q do MMQ, fórmula (4.10), carece de justificativa fundamental tanto quanto a escolha de \mathcal{L} da fórmula (4.2) para o método da Máxima Verossimilhança, mas também tem uma interpretação. Estimam-se os parâmetros pelos valores que, quando substituídos na função modelo (4.7), dão a curva que passa à menor distância dos pontos experimentais em seu conjunto ou, mais exatamente, à menor distância quadrática da curva toda em relação ao *conjunto* dos pontos experimentais, escolhida como a soma ponderada do quadrado da distância a cada ponto na *direção* Oy , com peso igual ao inverso da variância. Com essa escolha, $Q(a, b)$ é uma função quadrática em a e b , portanto apresenta um único mínimo. Assim, o MMQ define um critério objetivo e único para obter estimativas \hat{a} e \hat{b} dos parâmetros

que, substituídos na função modelo (4.7), aproximam os valores experimentais. Essa justificativa, embora racional e razoável, é também bastante arbitrária.

No capítulo 8, estudaremos detalhadamente as propriedades desse estimador e concluiremos que, no caso de funções **lineares** nos parâmetros, o MMQ é o estimador linear dos parâmetros de **menor variância, mesmo que as f.d.p.s dos dados não sejam normais**. Veremos adiante que, para funções não lineares nos parâmetros, o método da Máxima Verossimilhança é o mais adequado, embora seja equivalente ao MMQ quando as f.d.p.s dos dados forem gaussianas, como mostraremos na seção seguinte.

4.6 Dados normais: equivalência entre Máxima Verossimilhança e Mínimos Quadrados

Considere a medida $\{(x_i, y_i, \sigma_i), i = 1, 2, \dots, N\} \equiv \{(x_i, y_i, \sigma_i)\}$, onde cada y_i é obtido com erro ϵ_i , de modo que

$$y_i = h(x_i; \mathbf{p}_0) + \epsilon_i \quad , \quad (4.17)$$

em que os parâmetros \mathbf{p}_0 são as grandezas a determinar; essa relação generaliza a fórmula (4.8) acima.

Vamos manter as hipóteses do exemplo da seção 4.5: erro médio nulo, relação (1.7), variância medida igual ao erro quadrático médio, (4.9), e dados não correlacionados, $\langle \epsilon_i \epsilon_j \rangle = \text{cov}(y_i, y_j) = 0, i \neq j$. Nessas condições, o MMQ consiste em estimar os parâmetros pela minimização da função

$$Q(\mathbf{p}; \{(x_i, y_i, \sigma_i)\}) = \sum_{i=1}^N \frac{(y_i - h(x_i; \mathbf{p}))^2}{\sigma_i^2} \quad , \quad (4.18)$$

o que generaliza a fórmula (4.10) acima. Nessa última expressão, $h(x_i; \mathbf{p})$ representa o valor da função h no ponto x_i com parâmetros \mathbf{p} . A função $Q(\mathbf{p}; \{(x_i, y_i, \sigma_i)\})$ é, portanto, a soma dos quadrados dos resíduos (*resíduo* é a diferença entre o valor experimental em um certo ponto e o valor ajustado, nesse mesmo ponto) ponderados pelos inversos das variâncias.

O procedimento do MMQ para estimar os parâmetros consiste em determinar o valor $\hat{\mathbf{p}}$ dos parâmetros que minimiza $Q(\mathbf{p}; \{(x_i, y_i, \sigma_i)\})$.

A fim de aplicar o Método da Máxima Verossimilhança, é preciso construir

4.7. PROPRIEDADES DAS ESTIMATIVAS DOS PARÂMETROS AJUSTADOS NO EXEMPLO

a função verossimilhança da fórmula (4.2),

$$\mathcal{L}(\mathbf{p} \mid \{(x_i, y_i, \sigma_i)\}) = \prod_{i=1}^N f(x_i, y_i, \sigma_i \mid \mathbf{p}) \quad .$$

Juntando a hipótese da f.d.p. do erro ser *gaussiana*, obtemos

$$\begin{aligned} \mathcal{L}(\mathbf{p} \mid \{(x_i, y_i, \sigma_i)\}) &= \left[\prod_{j=1}^N \left(\frac{1}{\sqrt{2\pi}\sigma_j} \right) \right] \exp \left\{ - \sum_{i=1}^N \frac{[y_i - h(\mathbf{p}; x_i)]^2}{2\sigma_i^2} \right\} \\ \mathcal{L}(\mathbf{p} \mid \{(x_i, y_i, \sigma_i)\}) &= \left[\prod_{j=1}^N \left(\frac{1}{\sqrt{2\pi}\sigma_j} \right) \right] \exp \left\{ - \frac{Q(\mathbf{p}; \{(x_i, y_i, \sigma_i)\})}{2} \right\} \end{aligned} \quad (4.19)$$

onde $Q = Q(\mathbf{p}; \{(x_i, y_i, \sigma_i)\})$ da fórmula acima é *exatamente* o mesmo da fórmula (4.18).

Como o máximo da função verossimilhança desta última expressão corresponde ao mínimo de Q , dada a dependência exponencial de \mathcal{L} em Q , com o sinal negativo multiplicando o expoente, as estimativas obtidas pelos dois métodos, no caso muito comum de dados com f.d.p. gaussiana, são idênticas, qualquer que seja a forma da função h , tanto linear quanto não-linear nos parâmetros.

4.7 Exemplo – propriedades das estimativas dos parâmetros de uma reta

As estimativas pelo MMQ dos coeficientes linear e angular de uma reta em função dos dados podem ser obtidas com a equação (4.11), usando as matrizes das fórmulas desenvolvidas na questão 4.5 acima:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{1}{\sum_i^N \frac{1}{\sigma_i^2} \sum_i^N \frac{x_i^2}{\sigma_i^2} - \left[\sum_i^N \frac{x_i}{\sigma_i^2} \right]^2} \begin{pmatrix} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} & - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} \\ - \sum_{i=1}^N \frac{x_i}{\sigma_i^2} & \sum_{i=1}^N \frac{1}{\sigma_i^2} \end{pmatrix} \cdot \begin{pmatrix} \sum_{i=1}^N \frac{y_i}{\sigma_i^2} \\ \sum_{i=1}^N \frac{y_i x_i}{\sigma_i^2} \end{pmatrix} \quad . \quad (4.20)$$

em que a inversa da matriz \mathbf{M} foi determinada analiticamente. Nosso objetivo é mostrar que essas estimativas não são tendenciosas, ou seja, $\langle \hat{a} \rangle = a_0$ e $\langle \hat{b} \rangle = b_0$. Para isso, podemos escrever os valores médios de \hat{a} e \hat{b} em função

de a_0 e b_0 , substituindo na expressão acima os dados y_i como definidos na fórmula (4.8), que adaptamos para o caso da reta,

$$y_i = a_0 + b_0 x_i + \epsilon_i \quad . \quad (4.21)$$

A fim de facilitar o cálculo, vamos adotar uma notação compacta para os somatórios:

$$S = \sum_i^N \frac{1}{\sigma_i^2} \quad , \quad S_x = \sum_i^N \frac{x_i}{\sigma_i^2} \quad \text{e} \quad S_{xx} = \sum_i^N \frac{x_i^2}{\sigma_i^2} \quad .$$

Após substituição da relação (4.21) acima, os somatórios que envolvem os dados y_i ficam

$$\begin{aligned} \sum_{i=1}^N \frac{y_i}{\sigma_i^2} &= \sum_{i=1}^N \frac{a_0 + b_0 x_i + \epsilon_i}{\sigma_i^2} = a_0 S + b_0 S_x + \sum_i^N \frac{\epsilon_i}{\sigma_i^2} \quad \text{e} \\ \sum_{i=1}^N \frac{y_i x_i}{\sigma_i^2} &= a_0 S_x + b_0 S_{xx} + \sum_{i=1}^N \frac{\epsilon_i x_i}{\sigma_i^2} \quad . \end{aligned}$$

Isolando o valor de \hat{a} da expressão matricial, obtemos, após a substituição dos somatórios que acabamos de calcular,

$$\hat{a} = \frac{S_{xx} \left(a_0 S + b_0 S_x + \sum_i^N \frac{\epsilon_i}{\sigma_i^2} \right) - S_x \left(a_0 S_x + b_0 S_{xx} + \sum_i^N \frac{\epsilon_i x_i}{\sigma_i^2} \right)}{S \cdot S_{xx} - S_x^2} \quad .$$

Ao fatorar a_0 e b_0 , verificamos que o coeficiente de a_0 é 1 e o de b_0 é nulo, de modo que

$$\hat{a} = a_0 + \frac{S_{xx} \sum_i^N \frac{\epsilon_i}{\sigma_i^2} - S_x \sum_i^N \frac{\epsilon_i x_i}{\sigma_i^2}}{S \cdot S_{xx} - S_x^2} \quad .$$

Calculando o valor esperado, obtemos

$$\langle \hat{a} \rangle = a_0 + \frac{S_{xx} \sum_i^N \frac{\langle \epsilon_i \rangle}{\sigma_i^2} - S_x \sum_i^N \frac{\langle \epsilon_i \rangle x_i}{\sigma_i^2}}{S \cdot S_{xx} - S_x^2} \quad .$$

A hipótese de inexistência de erros sistemáticos corresponde a $\langle \epsilon_i \rangle = 0$, de maneira que se deduz

$$\langle \hat{a} \rangle = a_0 \quad ,$$

o que mostra que a estimativa é não tendenciosa. Um resultado análogo vale para o parâmetro b .

4.8. FUNÇÕES LINEARES NOS PARÂMETROS E DADOS ESTATISTICAMENTE INDEPENDENTES

Q4.9 Mostre que $\langle \hat{b} \rangle = b_0$.

Mostramos que, se não há erros sistemáticos nos dados, as estimativas dos parâmetros não são tendenciosas, isto é, distribuem-se em torno dos valores verdadeiros de maneira que suas médias igualam os valores verdadeiros. Note que esse resultado não depende das f.d.p.s dos dados serem normais, nem das variâncias serem as verdadeiras; ele decorre da linearidade do estimador em relação ao conjunto dos dados $\{y_i, i = 1, 2, \dots, N\}$ (e da ausência de erros sistemáticos). No capítulo 7, abordaremos de novo essa questão.

Note que as estimativas \hat{a} e \hat{b} são funções lineares dos dados y_i . Pode-se mostrar que uma combinação linear de variáveis aleatórias gaussianas tem uma f.d.p. gaussiana também, numa demonstração muito parecida com aquela da f.d.p. da média na seção 2.6. Acabamos de verificar que os valores esperados de \hat{a} e \hat{b} são os valores verdadeiros, a_0 e b_0 , de maneira que as f.d.p.s de \hat{a} e \hat{b} são

$$\begin{aligned}\phi(\hat{a}) &= \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left\{-\frac{(\hat{a} - a_0)^2}{2\sigma_a^2}\right\} & \text{e} \\ \gamma(\hat{b}) &= \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{-\frac{(\hat{b} - b_0)^2}{2\sigma_b^2}\right\} & .\end{aligned}\tag{4.22}$$

Em palavras, dizemos que ambas as estimativas são *normais* e *não tendenciosas*. Esse resultado vale qualquer que seja o número de dados. Na seção seguinte, calcularemos as variâncias das estimativas que, aliadas ao conhecimento da forma da f.d.p., permitem calcular a probabilidade de um certo intervalo $[a, a']$ conter o valor verdadeiro a_0 e de um certo intervalo $[b, b']$ conter b_0 , o que resolve o problema que nos colocamos.

Veremos adiante, na seção 4.10, que

$$\text{f.d.p.}(\hat{a}, \hat{b}) \neq \phi(\hat{a}) \times \gamma(\hat{b})$$

ou seja, as estimativas \hat{a} e \hat{b} são correlacionadas.

4.8 Funções lineares nos parâmetros e dados estatisticamente independentes, em geral

Nesta seção, mostraremos que o MMQ pode ser aplicado aos casos em que a relação entre os dados e os parâmetros não é expressa por uma função de uma ou poucas variáveis.

Considere um vetor \mathbf{p} , que representa um conjunto de μ parâmetros

$$\mathbf{p} = (p_1, p_2, \dots, p_\nu, \dots, p_\mu) \quad ,$$

Uma função linear em \mathbf{p} pode ser escrita como

$$y_i(\mathbf{p}) = \sum_{\nu=1}^{\mu} p_\nu g_{\nu i} \quad , \quad (4.23)$$

em que os $g_{\nu i}$ são valores conhecidos exatamente. Por exemplo, no caso do ajuste dos parâmetros da reta $y = a + bx$, y_i representa um dos dados experimentais, $g_{1i} = 1$ e $g_{2i} = x_i$.

Procedendo como nos outros casos, deduz-se que a estimativa de mínimos quadrados dos parâmetros p_ν a partir dos dados sa medida $\{(x_i, y_i, \sigma_i), i = 1, 2, \dots, N\}$ e do conjunto de coeficientes $\{g_{\nu i}, \nu = 1, 2, \dots, \mu, i = 1, 2, \dots, N\}$ é dada por

$$\hat{\mathbf{p}} = \mathbf{M}^{-1}\mathbf{D} \quad , \quad (4.24)$$

em que as componentes do vetor \mathbf{D} e os elementos da matriz \mathbf{M} são dados por

$$D_\nu = \sum_{i=1}^N \frac{y_i g_{\nu i}}{\sigma_i^2} \quad \text{e} \quad M_{\kappa\nu} = \sum_{i=1}^N \frac{g_{\kappa i} g_{\nu i}}{\sigma_i^2} \quad . \quad (4.25)$$

Note que a matriz \mathbf{M} é simétrica, seus elementos diagonais são definidos positivos, e pode-se mostrar que $\det(\mathbf{M}) \geq 0$, ou seja, \mathbf{M} é uma matriz *simétrica semi-definida positiva*. Os parâmetros p_κ são funções lineares dos valores experimentais y_i , como no caso da equação (4.20), do exemplo de ajuste dos coeficientes de uma reta.

Note que, nas expressões (4.26) e (4.25), usamos sub-índices gregos sempre que nos referimos às componentes do vetor de parâmetros e esses sub-índices variam de 1 a μ . Já para identificar os dados experimentais $\{y_i, \sigma_i\}$, usamos sub-índices do nosso alfabeto, variando de 1 a N . Para ser possível ajustar μ parâmetros a N pontos experimentais, é preciso¹ $N \geq \mu$.

¹O caso $N = \mu$ não se trata propriamente de um ajuste, pois os dados experimentais serão idênticos aos valores calculados com os parâmetros, mas as fórmulas deduzidas nesta seção podem ser usadas sem restrições .

Exemplo 4.1

Uma pessoa deseja determinar a massa de uma amostra, que deve ser acondicionada dentro de um recipiente. A tabela abaixo lista uma série de medições, todas com mesmo desvio padrão $\sigma = 1$ g e estatisticamente independentes.

i	y_i (g)	O que está na balança	g_{1i}	g_{2i}
1	10,5	Recipiente vazio	0	1
2	9,9	Recipiente vazio	0	1
3	9,1	Recipiente vazio	0	1
4	14,4	Recipiente com amostra	1	1
5	15,8	Recipiente com amostra	1	1

A segunda coluna dessa tabela traz os valores de massa obtidos na pesagem do material descrito na 3ª coluna. As colunas 4 e 5 trazem os valores das variáveis explicativas associadas aos parâmetros “massa do recipiente”, p_1 , e “massa da amostra”, p_2 , respectivamente. A primeira coluna ordena os dados, mas não interfere nos resultados, que serão os mesmos se os dados forem ordenados de outra forma.

Substituindo os valores da tabela nas fórmulas (4.25), obtém-se

$$D_1 = \sum_{i=1}^N \frac{y_i g_{1i}}{\sigma_i^2} = 30,2 \quad \text{e} \quad D_2 = \sum_{i=1}^N \frac{y_i g_{2i}}{\sigma_i^2} = 59,7$$

em g e

$$M_{11} = \sum_{i=1}^N \frac{g_{1i} g_{1i}}{\sigma_i^2} = 2, \quad M_{12} = M_{21} = \sum_{i=1}^N \frac{g_{1i} g_{2i}}{\sigma_i^2} = 2 \quad \text{e} \quad M_{22} = \sum_{i=1}^N \frac{g_{2i} g_{2i}}{\sigma_i^2} = 5$$

adimensionais.

Pela equação (4.24), obtém-se os valores ajustados para as massas do recipiente e da amostra, respectivamente $\hat{p}_1 = 5,3$ g e $\hat{p}_2 = 9,8$ g.

Uma forma comum de função modelo linear nos parâmetros é

$$h(\mathbf{x}; \mathbf{p}) = \sum_{\nu=1}^{\mu} p_{\nu} g_{\nu}(\mathbf{x}) \quad , \quad (4.26)$$

em que $g_\nu(\mathbf{x})$ é uma função qualquer da variável independente \mathbf{x} , mas que não depende de qualquer um dos parâmetros p_ζ . Em geral, \mathbf{x} é um *vetor* de r componentes,

$$\mathbf{x} = (x_1, x_2, \dots, x_q, \dots, x_r) \quad ,$$

sendo que r não precisa guardar qualquer relação com μ ou N . É o conjunto de variáveis independentes

$$(x_{1,i}, x_{2,i}, \dots, x_{q,i}, \dots, x_{r,i})$$

que define \mathbf{x}_i e, como essa é a variável *independente*, o número de coordenadas que a define não tem qualquer restrição.

4.9 Desvios padrão das estimativas

Podemos calcular o desvio-padrão de p_ν a partir da fórmula desenvolvida na seção 3.1. Como a dependência da função modelo com os parâmetros é linear, o resultado será exato². Calculando a covariância entre os parâmetros \hat{p}_ν e \hat{p}_κ usando a fórmula (3.6),

$$\text{cov}(p_\nu, p_\kappa) = \sum_{i=1}^N \frac{\partial p_\nu}{\partial y_i} \frac{\partial p_\kappa}{\partial y_i} \sigma_i^2 \quad , \quad (4.27)$$

em que se considerou que os dados y_i são estatisticamente independentes, isto é, $\text{cov}(y_i, y_j) = 0$ para $i \neq j$; essa é a mesma simplificação adotada ao definir a função Q das fórmulas (4.10) e (4.18).

Calculamos as derivadas a partir da forma explícita dos parâmetros, obtida da expressão (4.24),

$$p_\nu = \sum_{\tau=1}^{\mu} (\mathbf{M}^{-1})_{\nu\tau} D_\tau \quad .$$

Define-se uma matriz \mathbf{V} por

$$\mathbf{V} = \mathbf{M}^{-1} \quad , \quad (4.28)$$

que, no momento, permite simplificar a notação, mas, ao fim deste cálculo, veremos que essa matriz é muito importante. Assim, reescreve-se

$$p_\nu = \sum_{\tau=1}^{\mu} V_{\nu\tau} D_\tau \quad ,$$

²Na medida em que as variâncias dos dados sejam conhecidas exatamente.

donde se deduz

$$\frac{\partial p_\nu}{\partial y_i} = \sum_{\tau=1}^{\mu} V_{\nu\tau} \frac{\partial D_\tau}{\partial y_i} = \sum_{\tau=1}^{\mu} V_{\nu\tau} \frac{g_{\tau i}}{\sigma_i^2} \quad , \quad (4.29)$$

onde substituí-se

$$\frac{\partial D_\tau}{\partial y_i} = \frac{\partial}{\partial y_i} \sum_{j=1}^N \frac{y_j g_{\tau j}}{\sigma_j^2} = \sum_{j=1}^N \frac{\delta_{ji} g_{\tau j}}{\sigma_j^2} = \frac{g_{\tau i}}{\sigma_i^2} \quad .$$

Usando (4.29) em (4.27) para calcular as derivadas parciais da fórmula da covariância entre os parâmetros, obtém-se

$$\text{cov}(p_\nu, p_\kappa) = \sum_{i=1}^N \sum_{\tau=1}^{\mu} V_{\nu\tau} \frac{g_{\tau i}}{\sigma_i^2} \sum_{\eta=1}^{\mu} V_{\kappa\eta} \frac{g_{\eta i}}{\sigma_i^2} \sigma_i^2 \quad ,$$

que, depois de rearranjar a ordem das somatórias, fica

$$\text{cov}(p_\nu, p_\kappa) = \sum_{\tau=1}^{\mu} V_{\nu\tau} \sum_{\eta=1}^{\mu} V_{\kappa\eta} \sum_{i=1}^N \frac{g_{\tau i} g_{\eta i}}{\sigma_i^2} \quad ,$$

onde é possível identificar a última somatória como $M_{\tau\eta} = M_{\eta\tau}$. Assim,

$$\text{cov}(p_\nu, p_\kappa) = \sum_{\tau=1}^{\mu} V_{\nu\tau} \sum_{\eta=1}^{\mu} V_{\kappa\eta} M_{\eta\tau} = \sum_{\tau=1}^{\mu} V_{\nu\tau} \delta_{\kappa\tau} \quad ,$$

que, usando a definição (4.28), $\mathbf{V} = \mathbf{M}^{-1}$, equivalente a $\mathbf{VM} = \mathbf{I}$, leva ao resultado procurado,

$$\text{cov}(p_\nu, p_\kappa) = V_{\nu\kappa} \quad . \quad (4.30)$$

Esse resultado explica o símbolo usado para a matriz inversa da matriz \mathbf{M} — ela é a matriz das Variâncias e covariâncias entre os parâmetros: sua diagonal contém as variâncias dos parâmetros e os termos fora da diagonal, as covariâncias. Note que \mathbf{V} não depende dos valores das grandezas y_i , apenas das variáveis $g_{\tau i}$ e das variâncias, de modo que é possível antecipar a precisão dos resultados a partir da precisão dos dados e do procedimento de medida usado, que, juntamente com o modelo adotado, é quem determina $g_{\tau i}$.

Os diferentes parâmetros do ajuste não são, portanto, independentes, a menos que o elemento correspondente na matriz \mathbf{V} seja nulo. Para que todos os μ parâmetros sejam independentes, \mathbf{V} precisa ser diagonal, o que implica em \mathbf{M} diagonal — algo que só ocorre quando *seleccionamos* de maneira particular as

variáveis $g_{\tau i}$ ou, ainda, por um acaso. Em geral, portanto, devemos considerar que as estimativas de parâmetros provenientes de um ajuste sobre um conjunto de dados *são covariantes*.

É interessante perceber, também, que a matriz de variâncias, calculada como $\mathbf{V} = \mathbf{M}^{-1}$, é exata neste caso de ajuste de funções lineares *mesmo que os dados não obedecem à f.d.p. normal*, na medida em que as variâncias σ_i^2 dos dados y_i sejam *exatas* também. De qualquer maneira, \mathbf{V} não depende da f.d.p. dos dados, apenas é necessário que *exista* uma f.d.p., qualquer que seja, com variância conhecida e igual a σ_i^2 .

4.10 Interpretação estatística dos parâmetros ajustados

O que discutimos na seção 4.7 pode ser estendido à situação onde ajusta-se um número qualquer de parâmetros, porque as estimativas $\hat{\mathbf{p}}$ obtidas por meio das equações (4.24 e 4.25) correspondem a funções lineares dos valores experimentais y_i .

4.10.1 Geral

O MMQ aplicado nos casos em que a relação entre os dados e os parâmetros é linear fornece parâmetros ajustados que dependem linearmente dos dados, de modo que, se os dados não são tendenciosos, então as estimativas dos parâmetros não são tendenciosas e suas variâncias são os elementos da matriz das variâncias, \mathbf{V} . Assim, se a f.d.p. dos dados for gaussiana, a função densidade de probabilidade das estimativas p_ν é a multinormal,³

$$\phi(\hat{\mathbf{p}}, \mathbf{V}) = \frac{1}{(2\pi)^{\mu/2} \det(\mathbf{V})^{1/2}} \exp \left[-\frac{(\hat{\mathbf{p}} - \mathbf{p}_0)^t \cdot \mathbf{V}^{-1} \cdot (\hat{\mathbf{p}} - \mathbf{p}_0)}{2} \right] . \quad (4.31)$$

Os parâmetros ajustados são, portanto, estatisticamente dependentes e a f.d.p. conjunta dos parâmetros não pode ser fatorada em funções densidade de probabilidade independentes, ou seja, cada uma dependendo de um único parâmetro, como havíamos antecipado na seção 4.7.

³A f.d.p. das estimativas pelo MMQ no caso linear é a multinormal sempre que o número de dados for grande, desde que as variâncias dos dados sejam finitas, independentemente das suas f.d.p.s ou f.p.s, como veremos mais à frente.

No entanto, como vimos na seção 2.4, a f.d.p. marginal de um parâmetro em particular, ou seja, a densidade de probabilidade deste parâmetro quando se ignoram totalmente os demais, é uma gaussiana com média igual ao valor verdadeiro e variância igual ao termo correspondente da diagonal da matriz de variâncias, como já havíamos antecipado na equação (4.22). Como a f.d.p. das estimativas tem forma definida, podemos calcular as probabilidades dos diferentes intervalos construídos em torno delas conterem os valores verdadeiros. Em particular, um determinado parâmetro $p_{\nu,0}$ tem 68% de probabilidade de se encontrar no intervalo $[\hat{p}_{\nu} - \sqrt{V_{\nu\nu}}, \hat{p}_{\nu} + \sqrt{V_{\nu\nu}}]$, com resultados análogos para cada um dos μ parâmetros, sempre lembrando que cada um desses intervalos supõe desconhecimento dos demais parâmetros. Se for necessário determinar um intervalo de confiança para a observação simultânea de dois parâmetros em uma certa região do espaço bidimensional coberta por esse par, é necessário levar em conta a covariância entre os parâmetros envolvidos.

4.10.2 Exemplo da reta

No caso do ajuste dos coeficientes da reta da seção 4.7, pode-se obter formas analíticas para os desvios-padrões dos coeficientes, que, em (4.31), definem a forma da f.d.p. dos parâmetros.

As fórmulas são:

$$\sigma_a^2 = V_{11} = \frac{\sum_i^N \frac{x_i^2}{\sigma_i^2}}{\sum_i^N \frac{1}{\sigma_i^2} \sum_i^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_i^N \frac{x_i}{\sigma_i^2}\right)^2}, \quad (4.32)$$

$$\sigma_b^2 = V_{22} = \frac{\sum_i^N \frac{1}{\sigma_i^2}}{\sum_i^N \frac{1}{\sigma_i^2} \sum_i^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_i^N \frac{x_i}{\sigma_i^2}\right)^2}, \quad (4.33)$$

que, nas f.d.p.s de \hat{a} e \hat{b} das expressões (4.22), permitem determinar os intervalos de confiança desejados. Por outro lado, calculamos

$$\text{cov}(a, b) = V_{12} = \frac{-\sum_i^N \frac{x_i}{\sigma_i^2}}{\sum_i^N \frac{1}{\sigma_i^2} \sum_i^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_i^N \frac{x_i}{\sigma_i^2}\right)^2}, \quad (4.34)$$

deixando claro que, na maioria dos casos, a inclinação e o coeficiente linear são estatisticamente dependentes. Na próxima seção, essas fórmulas serão simplificadas.

4.11 Análise de previsão

Na seção 4.8, a fórmula (4.25) da matriz \mathbf{M} mostra que ela *independe* dos valores experimentais y_i . Na seção 4.9, vimos que os desvios padrão dos parâmetros ajustados dependem apenas de \mathbf{M} . Juntando os dois resultados, conclui-se que as incertezas dos parâmetros ajustados são independentes dos resultados das observações. Como a matriz \mathbf{M} depende tanto do conjunto dos valores da variável independente (\mathbf{x}_i ou os $g_{\nu i}$) quanto das variâncias, mesmo que não haja possibilidade de reduzir as incertezas experimentais para cada observação y_i , muitas vezes é possível escolher os valores de \mathbf{x}_i (ou o conjunto dos $g_{\nu i}$) para os quais observaremos a grandeza y , e uma escolha adequada desse conjunto de valores pode melhorar muito a precisão final dos parâmetros ajustados.

Análise de previsão tem, aqui, o sentido de previsão dos desvios padrão dos parâmetros ajustados, o que é possível, conforme a discussão acima. Na prática, a maior limitação consiste na frequente dependência entre y_i e σ_i , o que obriga a desenvolver um modelo físico para y , a partir do qual se estimam os desvios padrão. A inclusão do modelo físico torna a análise complexa e dificulta a previsão dos desvios padrão independentemente da antecipação (aproximada) do resultado, o que não é um problema de estatística.

Há situações como a da seção seguinte, onde retomamos o exemplo do ajuste dos parâmetros de uma reta em uma situação particular, para as quais é possível uma solução analítica do problema da previsão. Normalmente, porém, o roteiro de um experimento de medida de uma grandeza, quando se pretende realizar a análise de previsão, é o seguinte: efetua-se uma medida rápida e preliminar, a partir da qual se estimam os parâmetros, mesmo que com baixa precisão; constrói-se um modelo para os desvios padrão dos dados, com o conhecimento adquirido na medida preliminar; define-se a precisão desejada na medida final dos parâmetros e seleciona-se o conjunto de valores da variável independente necessário para se atingir essa precisão; mede-se “para valer” e, se tudo foi bem, o resultado obtido terá a precisão pretendida. Note que é boa prática realizar uma medida preliminar, a fim de localizar os aspectos críticos do experimento. Assim, usar os resultados desse protótipo para basear a análise de previsão aumenta pouco a carga de trabalho e contribuirá para que a medida definitiva resulte como desejado.

4.12 Análise de previsão—parâmetros da reta

A fim de realizar uma discussão completa, vamos nos limitar ao ajuste dos parâmetros de uma reta, dada pela fórmula (4.14), a um conjunto de dados $\{(x_i, y_i, \sigma_i), i = 1, 2, \dots, N\}$, em que σ_i é o desvio padrão de y_i , igual para todos os N dados, ou seja,

$$y = a + bx$$

em que a e b são os parâmetros a ajustar e

$$\sigma_i = \sigma, \quad \forall i \quad . \quad (4.35)$$

Nesse caso, aplicam-se as fórmulas (4.32), (4.33) e (4.34), as quais têm um mesmo denominador, que, usando (4.35), se reduz a

$$\sum_i^N \frac{1}{\sigma_i^2} \sum_i^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_i^N \frac{x_i}{\sigma_i^2} \right)^2 = \frac{N \sum x_i^2 - N^2 \bar{x}^2}{\sigma^4} = \frac{N \sum (x_i - \bar{x})^2}{\sigma^4} \quad .$$

Na fórmula acima, usamos

$$\bar{x} = \frac{1}{N} \sum_i^N x_i \quad , \quad (4.36)$$

que é a média das coordenadas independentes x_i e, portanto, não é uma variável aleatória — os valores de x_i não costumam ser aleatórios, mas sim escolhidos a priori. A raiz quadrada do valor médio do quadrado da diferença das coordenadas x_i em relação ao seu valor médio pode, num sentido razoável, ser chamada de dispersão em x dos dados, mesmo que não seja uma grandeza estatística nos casos em que ela é *completamente determinada* pelo conjunto $\{x_i\}$ escolhido pelo experimentador. A dispersão dos valores desse conjunto, δ_x , pode ser definida a partir de

$$\delta_x^2 = \frac{1}{N} \sum_i^N (x_i - \bar{x})^2 \quad . \quad (4.37)$$

Com as definições (4.36) e (4.37), as fórmulas das variâncias dos parâmetros ajustados e da covariância entre eles reduzem-se a

$$\sigma_a^2 = \sigma^2 \frac{\sum_i^N x_i^2}{N^2 \delta_x^2} = \frac{\sigma^2}{N} \frac{\delta_x^2 + \bar{x}^2}{\delta_x^2} = \frac{\sigma^2}{N} \left(1 + \frac{\bar{x}^2}{\delta_x^2} \right) \quad , \quad (4.38)$$

$$\sigma_b^2 = \sigma^2 \frac{1}{N\delta_x^2} \quad \text{e} \quad (4.39)$$

$$\text{cov}(a, b) = -\sigma^2 \frac{\bar{x}}{N\delta_x^2} \quad , \quad (4.40)$$

neste caso particular em que todos os desvios-padrão são iguais, conforme relação (4.35). Assim, concluímos que a condição necessária e suficiente para uma boa definição do coeficiente angular b é dispersar os valores de x para os quais efetuam-se as observações. Já para definir com precisão o termo constante (coeficiente linear da reta) é preciso observar-se os pontos experimentais em torno de $x = 0$, para minimizar \bar{x} (ou $\sum x_i^2$, que dá no mesmo). Finalmente, a covariância entre os parâmetros só é nula quando $\bar{x} = 0$, ou seja, é preciso efetuar observações tanto para valores negativos quanto para valores positivos de x a fim de obter-se parâmetros *estatisticamente* independentes.

EXERCÍCIOS

- 4.1. Considere dois dados experimentais x_1 e x_2 correspondentes a medidas de uma mesma grandeza x_0 cuja matriz de covariância é

$$\mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} .$$

com distribuição multinormal.

- (a) Escreva a função verossimilhança $\mathcal{L}(x | x_1, x_2)$.
 (b) Mostre que a estimativa de x_0 pelo método da máxima verossimilhança é

$$\tilde{x} = \frac{x_1(\sigma_2^2 - \rho\sigma_1\sigma_2) + x_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} .$$

- (c) Determine a variância de \tilde{x} .

- 4.2. Considere uma medida de uma certa grandeza com N dados experimentais $\{x_i, i = 1, \dots, N\}$, todos com o mesmo desvio padrão σ e igualmente correlacionados entre si ($\text{cov}(x_i, x_j) = \rho\sigma^2, \forall i \neq j$).

- (a) Mostre que a estimativa da grandeza pelo método dos mínimos quadrados é a média aritmética dos dados.
- (b) Mostre que a variância da média é

$$\sigma_m^2 = \frac{\sigma^2}{N} + \frac{N-1}{N} \sigma^2 \rho \quad .$$

Note que este último resultado já havia sido obtido em um exercício do capítulo anterior.

- 4.3. Em N ensaios de Bernoulli observaram-se n eventos de um determinado tipo.

- (a) Sabendo que esse número de eventos obedece à distribuição binomial

$$P_{Np}(n) = \frac{N!}{(N-n)!n!} p^n (1-p)^{N-n} \quad ,$$

com p desconhecido, estime p pelo Método da Máxima Verossimilhança.

- (b) Verifique que essa estimativa, \hat{p} , é não tendenciosa.
- (c) Estime o desvio padrão de \hat{p} .

- 4.4. Na medida de uma grandeza, foram obtidos três dados, $\{2, 11, 3\}$, que seguem a f.p. de Poisson de média a ,

$$P_a(n) = \frac{a^n e^{-a}}{n!} \quad .$$

- (a) Determine a estimativa da grandeza, \hat{a} , pelo Método da Máxima Verossimilhança.
- (b) Estime a incerteza de \hat{a} .

- 4.5. Realizaram-se 5 observações da posição de um objeto que se move com velocidade constante. Os dados obtidos são ternos $(t_i; x_i; \sigma_i)$, onde x_i representa a posição do objeto observada no instante t_i , com desvio padrão σ_i . A medida consiste no seguinte conjunto de dados:

t	x	σ
1	1,06	0,25
2	2,19	0,15
3	3,12	0,15
4	4,12	0,15
5	5,32	0,15

onde t é dado em unidades de segundo e tanto x como σ são dados em centímetro. A incerteza em t deve ser ignorada.

- Calcule a posição inicial (em $t = 0$ s) e a velocidade, bem como a matriz de covariância desses parâmetros.
- Determine a posição no instante $t = 3,5$ s, bem como seu desvio padrão.
- Faça um gráfico do desvio padrão dos valores interpolados $x(t)$ em função do tempo t .
- Determine para que valor de t o desvio padrão do valor interpolado é mínimo e explique porque, apesar dos dados terem sido obtidos a intervalos de tempo regulares, ele não é o instante do meio do intervalo, $t = 3$ s, ou seja, diga qual característica particular do conjunto de dados é responsável por esse fato.

4.6. Para medir a defasagem de um movimento harmônico de frequência angular ω conhecida, o experimentador ajusta os parâmetros da função

$$y = a_1 \sin(x) + a_2 \cos(x) \quad , \quad (4.41)$$

onde a incerteza em $x = \omega \cdot \text{tempo}$ pode ser ignorada. A tabela abaixo apresenta os dados experimentais obtidos. A primeira coluna dá os valores da variável independente, a segunda, da variável dependente e a terceira, o desvio padrão de y_i .

Os itens a seguir encaminham o ajuste de a_1 e a_2 da função modelo (4.41) aos dados da tabela e exploram o resultado obtido. Os dados foram escolhidos de forma a facilitar o cálculo a mão.

- Mostre que a matriz \mathbf{M} , da fórmula (4.25) do texto, é diagonal neste caso e para esses dados.

x (rad)	y (m)	σ (m)
$-\pi/2$	-1,07	0,10
$-\pi/4$	-0,33	0,10
0	0,51	0,10
$\pi/4$	1,00	0,10
$\pi/2$	1,11	0,10

- (b) Determine as estimativas de a_1 e a_2 , bem como seus respectivos desvios padrão. Dê o intervalo que tem 68% de probabilidade de conter o valor verdadeiro de a_1 . Dê o intervalo que tem 68% de probabilidade de conter o valor verdadeiro de a_2 .
- (c) A função da fórmula (4.41) pode ser reescrita como

$$y = \alpha \operatorname{sen}(x + \phi) \quad .$$

Calcule α e ϕ a partir das estimativas a_1 e a_2 obtidas acima. Determine seus desvios padrão, bem como sua covariância, a partir dos desvios padrão em a_1 e a_2 obtidos acima.

4.7. A relação entre as variáveis x e y tem a forma

$$y = a_0 + a_1x \quad .$$

Realizam-se três observações experimentais (x_i, y_i, σ_i) , onde σ_i é o desvio padrão de y_i ,

$$\{(-1; -9,9; 1,0), (0; -0,4; 1,0), (1; 11,3; 1,0)\} \quad .$$

Ignore o erro em x_i e use o Método dos Mínimos Quadrados.

Determine:

- (a) as estimativas de a_0 e a_1 .
- (b) os desvios padrão destas estimativas.
- (c) se as estimativas de a_0 e a_1 são estatisticamente independentes; explique sua conclusão.
- (d) o valor estimado e o desvio padrão de y para $x = 0,5$.

4.8. Uma superfície plana, inclinada em relação à vertical, é descrita por

$$z = ax + by$$

em um sistema de eixos triortogonal, definido de maneira que Oz seja o eixo vertical e corte o plano yOx em $(0, 0, 0)$. Para determinar a e b , realizou-se um conjunto de 4 observações de pontos do plano, obtendo-se quadras $(x_i, y_i, z_i, \sigma_i)$, onde σ_i é o desvio padrão de z_i , que tem f.d.p. normal, e os erros nas variáveis x_i e y_i podem ser ignorados frente a σ_i . Os dados obtidos constam da tabela abaixo.

x_i	y_i	z_i
1	1	3,03(10)
1	2	5,17(15)
2	1	4,01(15)
2	2	5,93(20)

- Determine a e b com seus respectivos desvios padrão.
- Determine z do plano para o ponto $x = 1,5$ e $y = 1,5$ e determine um intervalo com 68% de probabilidade de conter o valor verdadeiro desse z .
- Seria possível, observando-se apenas quatro pontos do plano, determinar a e b com maior precisão? Discuta esta questão e dê uma diretriz genérica sobre como selecionar os pontos (x, y) onde deveria ser observado z para aumentar as precisões de a e b .

4.9. Deseja-se ajustar os parâmetros a e b de uma reta

$$y = a + bx$$

aos pontos experimentais (x_i, y_i, σ_i) onde os desvios padrão σ_i são todos iguais. Pode-se observar no máximo 6 pontos entre $-10 \leq x \leq 10$ e nenhum fora deste intervalo, e a separação mínima entre os pontos é $\Delta x = 2$. Determine:

- A maneira com que as observações devem ser distribuídas para que as estimativas \hat{a} e \hat{b} , dos parâmetros a e b respectivamente, sejam independentes.

- (b) O conjunto de pontos que torna o desvio padrão de \hat{a} mínimo, dado que \hat{a} e \hat{b} são independentes.
- (c) O conjunto de pontos que torna o desvio padrão de \hat{b} mínimo, dado que \hat{a} e \hat{b} são independentes.
- 4.10. O governo deseja saber a eficiência energética do motor de um determinado tipo de veículo antes de decidir se autoriza ou não a sua comercialização. Para isso, dispõe dos dados da tabela abaixo, em que L é a abreviação a unidade litro.

consumo (L)	distância percorrida em km	
	estrada	cidade
22,7	100	200
15,9	300	100
24,7	200	200
30,8	0	400

Os desvios-padrão dos volumes consumidos são todos iguais a 1,0 L e não há correlação entre os dados de consumo; as incertezas nas distâncias percorridas podem ser ignoradas.

Determine, usando o MMQ, a eficiência desse veículo, em litros por quilômetro, na cidade e na estrada.

- 4.11. Uma pessoa tem duas crianças gêmeas idênticas, inclusive quanto às suas massas. Em um passeio, essa pessoa resolveu pesar a si mesma e as crianças em cada balança de farmácia que encontrasse pelo caminho. A tabela abaixo detalha as medições efetuadas e os resultados que ele obteve. As medições foram realizadas em balanças diferentes.

massa (kg)	quem subiu na balança
65,9	apenas a pessoa adulta
70,2	a pessoa adulta com uma criança
9,8	as duas crianças
74,3	a pessoa adulta com as duas crianças
64,9	apenas a pessoa adulta

130 *CAPÍTULO 4. INFERÊNCIA ESTATÍSTICA E AJUSTE DE PARÂMETROS*

Adote 0,5 kg para o desvio-padrão das massas e que as medições são independentes umas das outras.

Determine as estimativas dos valores das massas de cada uma das pessoas pelo MMQ.

Bibliografia

- [Arfken] Mathematical Methods for Physicists, G.Arfken & H.Weber, Academic Press, 4ª edição (1995)
- [Bard] Nonlinear Parameter Estimation, Yonathan Bard, Academic Press (1974)
- [Benzécri] Histoire et Préhistoire de l'Analyse des Données, J.P.Benzécri, Ed. Bordas, Paris 1982
- [Bevington] Data Reduction and Error Analysis for the Physical Sciences, P.Bevington, McGraw-Hill, 1969
- [Birge] The calculation of errors by the method of least squares, Raymond T. Birge, Phys Rev 40 (1932) 207-227
- [Conover] Practical Nonparametric Statistics, W.J.Conover, John Wiley & Sons Inc. 1971
- [CRC] Handbook of Tables for Probability and Statistics, CRC
- [Eadie] Statistical Methods for Physicists, W.T.Eadie et al., North Holland Pub.Co. 1971
- [Escoubes] Experimental Signs Pointing to a Bayesian Instead of a Classical Approach for Experiments with Small Number of Events, B.Escoubes, S.De Unamuno e O. Helene, Nuclear Instruments and Methods A257(1987)346
- [Feller] Feller, An Introduction to Probability Theory and its Applications, John Wiley, 2ª Ed. (1957)
- [Feynman] Lectures on Physics Vol.I, Chap.6, Feynman Leighton & Sands

- [Firestone] Analysis of α , β , and γ ray emission probabilities, R.B. Firestone, Nuclear Instruments and Methods A286(1990)584
- [Forbes] Forbes, Eric G., *Gauss and the Discovery of Ceres*. Journal for the History of Astronomy. 2 (1971) 195-199.
- [Frieden] Fisher's Information as the basis for the Schrödinger wave equation, B. Roy Frieden, Am. J.Phys. 57(1989)11
- [Geraldo] L.P. Geraldo e D.L Smith, Nuclear Instruments and Methods A290(1990)499
- [Grosser] Morton Grosser, The Discovery of Neptune, Harvard University Press, Cambridge, Massachusetts (1962)
- [Gray] C.G.Gray, Am. J.Phys. 59(1991)282
- [Guimarães-Filho] Z.O. Guimarães-Filho e O. Helene, One Step Self-Calibration Procedure in Gamma-Ray Energy Measurements. Brazilian Journal of Physics, v. 33, n.2, (2003) 280-281.
- [Lyons] How to combine correlated estimates of a single physical quantity, L.Lyons, D.Gibaut e P. Clifford, Nuclear Instruments and Methods A270(1988)110
- [Helene] Tratamento Estatístico de dados em Física Experimental, O.Helene, V. R. Vanin, Ed. Edgard Blücher, 2ª Ed., 1991
- [Helene 83] Upper Limit of Peak Area, O.Helene, Nuclear Instruments and Methods 212(1983)319
- [Helene 84] Errors in Experiments with Small Number of Events, O.Helene, Nuclear Instruments and Methods 228(1984)120
- [Helene 91b] Determination of the Upper Limit of Peak Area, O.Helene, Nuclear Instruments and Methods A300(1991)132
- [Helene 91] O que é uma medida?, O. Helene, Shan.P.Tsai, R.P.Teixeira, preprint IFUSP/P-854 (1990) e Revista de Ensino de Física, Vol.13 p.12, SBF (1991).

- [Helene 93] O.Helene and V.R.Vanin, Nuclear Instruments and Methods A335(1993)227
- [Helene 2013] O. Helene, Método dos Mínimos Quadrados com formalismo matricial, Editora Livraria da Física, São Paulo, 2ª edição (2013).
- [James] A review of pseudorandom number generators, F.James, Computer Physics Communications 60(1990)329-344
- [Kendall] The Advanced Theory of Statistics, M.Kendall, A.Stuart & J.K.Ord, Charles Griffin & Company Limited, London
- [Magalhães] Noções de Probabilidade e Estatística, Marcos N. Magalhães e Antonio Carlos P. Lima, Editora da Universidade de São Paulo - EDUSP, 2011
- [Mannhart] A Small Guide to Generating Covariances of Experimental Data, Report PTB-FMRD 84, Berlin, 1981. ISSN 0341-6666
- [Marquardt] An Algorithm for Least-Squares Estimation of Nonlinear Parameters, D. Marquardt, SIAM J. Appl. Math. 11, 431-441, 1963
- [Merzbacher] Quantum Mechanics, E.Merzbacher, John Wiley & sons, New York 1961
- [Mises] Probability, Statistics and Truth, R.von Mises, Dover, 1955
- [Moralles] M.Morales, P.R.Pascholati, V.R.Vanin and O.Helene, Applied Radiation and Isotopes 46-2(1995)133
- [Mucciolo] E.R.Mucciolo and O.Helene, Nuclear Instruments and Methods A256(1987)153
- [Noether] Introdução à Estatística – Uma abordagem não paramétrica, G.E.Noether, Guanabara Dois, 1983
- [Smith] D.L. Smith, Nuclear Instruments and Methods A257(1987)361
- [Stigler] *Gauss and the Invention of Least Squares*. Stephen M. Stigler, Annals of Statistics, 9 (1981) 465-474 - doi:10.1214/aos/1176345451
- [Vanin 1989] V.R.Vanin e M.Aiche, Nuclear Instruments and Methods A284(1989)452

- [Vanin 1997] V.R.Vanin, G.Kenchian, M.Morales, O.Helene e P.R. Pascholati, Nuclear Instruments and Methods A391(1997)338
- [Vuolo] Fundamentos da Teoria de Erros, J.H.Vuolo, Ed. Edgard Blücher, 1992
- [Youden] Statistical Methods for Chemists, W.J.Youden, John Wiley 1951
- [Zar] J.H. Zar, Appl. Statist. 27(1978)n.3, 280-290