

ANÁLISE ESTATÍSTICA DE MEDIDAS EM CIÊNCIAS EXATAS

Vito R. Vanin, Philippe Gouffon, Otaviano Helene

Março 2023

Capítulo 8

O Método dos Mínimos Quadrados

Apresentaremos aqui o método dos mínimos quadrados de forma completamente geral, que se aplica tanto aos casos mais comuns quanto àqueles em que os dados não são estatisticamente independentes, além de permitir a imposição de vínculos lineares entre os parâmetros.

Discutiremos em que situação as propriedades deste estimador são ótimas. Verificaremos que este é o método que se deve usar sempre que os dados experimentais se relacionarem linearmente com os parâmetros a serem ajustados e conhecermos as variâncias (e covariâncias) dos dados, mesmo que os erros não tenham f.d.p.s normais. O método também se aplica quando as variâncias são desconhecidas, desde que elas sejam iguais ou suas razões sejam conhecidas, para que essa variância comum possa ser estimada, como mostraremos na seção 8.8.

O fato mencionado no parágrafo precedente, que este método dá bons resultados independentemente da forma da f.d.p. dos dados, não significa que a interpretação dos resultados não venha a depender dela, especialmente quando os parâmetros forem ajustados a poucos dados.

O método dos mínimos quadrados também é o melhor a usar quando a relação entre os parâmetros do ajuste e as grandezas medidas *não* for linear, desde que a f.d.p. dos dados seja gaussiana. Nesse caso, as boas propriedades dos mínimos quadrados decorrem da relação entre a função Verossimilhança e a soma dos quadrados dos resíduos Q ,

$$\ln \mathcal{L}(\mathbf{a} \mid \{x_i, y_i, \sigma_i\}) = -Q/2 + \text{constante} , \quad (4.19)$$

da seção 4.6, e serão discutidas no próximo capítulo. Neste caso, porém, as propriedades de mínima variância (seção 8.6) e não tendenciosidade (seção 8.3) não são garantidas.

8.1 O modelo linear

Vamos retomar o caso em que a função que expressa o modelo, na forma de uma relação entre os dados experimentais e os parâmetros de interesse, é linear nos parâmetros, como no capítulo 4 (seção 4.8). Vamos escrevê-la de forma geral, mas com o vetor de parâmetros chamado \mathbf{a} , com

$$\mathbf{a} = (a_1, \dots, a_\nu, \dots, a_\mu) \quad ,$$

em que seguimos usando uma letra grega como sub-índice para representar uma componente genérica e μ , para o número de parâmetros. Usaremos um zero adicional subscrito para denotar o valor verdadeiro do parâmetro, por exemplo, $a_{\nu,0}$. Denominamos y a variável dependente, N o número de dados, e um dado genérico tem uma letra do nosso alfabeto no sub-índice.

Neste modelo, a relação entre cada dado experimental e os parâmetros, quando o erro na medida é isolado como na relação (4.21), tem a forma da relação (4.26),

$$y_i = \sum_{\nu=1}^{\mu} a_{\nu,0} x_{i,\nu} + \epsilon_i \quad \text{para} \quad i = 1, \dots, N, \quad (8.1)$$

onde ϵ_i é o erro, com as propriedades

$$E(\epsilon_i) = 0 \quad \text{e} \quad (8.2)$$

$$E(\epsilon_i^2) = \sigma_i^2 \quad . \quad (8.3)$$

Os elementos $x_{i,\nu}$ definem uma matriz \mathbf{X} retangular de N linhas e μ colunas, e a fórmula (8.1) pode ser reescrita como

$$\mathbf{y} = \mathbf{X}\mathbf{a}_0 + \vec{\epsilon} \quad . \quad (8.4)$$

A matriz \mathbf{X} tem o nome de *matriz de planejamento* ou *de projeto*. Ela relaciona os dados experimentais com os parâmetros de ajuste, ou seja, é a matriz que contém as informações significativas do *arranjo* experimental adotado. Nessa

fórmula, os dados e os erros respectivos constituem dois vetores coluna, portanto, cada um tem N linhas, e os parâmetros formam outro vetor coluna, com μ linhas. Note que os $x_{i,\nu}$ são valores exatos, conhecidos sem erro, e essa característica está relacionada à não tendenciosidade e eficiência do método. No entanto, veremos ao final do capítulo que poderemos utilizar o método dos mínimos quadrados mesmo quando os $x_{i,\nu}$ forem variáveis aleatórias, sob certas condições que especificaremos e que *precisam* ser verificadas em cada caso particular.

Na notação da fórmula (4.26), os elementos da matriz de planejamento são tais que

$$x_{i,\nu} = g_\nu(\mathbf{x}_i) \quad .$$

No entanto, não há necessidade dos diferentes elementos da matriz \mathbf{X} se relacionarem entre si de forma específica, de modo que a relação (8.4) é a mais geral possível.

Exemplo 8.1

Determinação da matriz de planejamento

As relações matemáticas entre os parâmetros de ajuste e os dados experimentais dependem da maneira que se realiza o experimento, e a matriz de planejamento, \mathbf{X} , resume todas as equações correspondentes no conjunto dos seus elementos, como se ilustra nos exemplos a seguir.

i) Ajuste dos coeficientes de um polinômio

Considere o ajuste dos parâmetros a, b e c da expressão polinomial $y = a + bx + cx^2$ a dados experimentais (x_i, y_i) , em que \mathbf{V} é a matriz de covariância dos dados experimentais y_i . Na forma da eq. (8.4), esse problema se escreve como

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad ,$$

onde a matriz retangular que contém as potências das coordenadas x_i é a matriz de planejamento.¹

ii) Média de dados

Considere n dados (y_i) , correspondentes a medições de uma mesma grandeza y_0 , cujo valor é desconhecido. A eq. (8.4), nesse caso, toma a forma

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \cdot (y_0) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} .$$

A matriz de planejamento é o vetor coluna cujos elementos são todos iguais a 1, e a matriz de parâmetros é a matriz 1 por 1 cujo único elemento é o valor desconhecido y_0 a ser ajustado.

iii) Equações diferentes para dados distintos

Considere um experimento que busca medir o peso p_0 de uma amostra, que precisa ser colocada dentro de um recipiente, cujo peso p_r também não se conhece. Numa primeira medição, pesou-se a amostra junto com seu recipiente e se obteve o resultado y_1 . Posteriormente, pesou-se apenas o recipiente, que deu o valor y_2 . Uma terceira medição, com uma balança especial, permitiu pesar apenas a amostra, sem qualquer recipiente, com o resultado y_3 . Finalmente, uma quarta medição da amostra dentro de um outro recipiente, cujo peso é conhecido e igual a p_{rc} , resultou no valor y_4 .

As equações desse processo, dentro do esquema da eq. (8.4), ficam na forma

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 - p_{rc} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} p_0 \\ p_r \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix} .$$

Note que foi necessário alterar o quarto dado experimental para que o conjunto de relações ficasse na forma da eq. (8.4). Caso o valor

¹Uma matriz que tem os elementos da mesma linha em progressão geométrica, como a desse exemplo, chama-se Matriz de Vandermonde.

p_{rc} tenha uma incerteza, a variância desse dado alterado $y'_4 = y_4 - p_{rc}$ precisará ser calculada por propagação de incerteza.

Da relação (8.4), obtém-se

$$E(\mathbf{y}) = \mathbf{X}\mathbf{a}_0$$

como consequência da inexistência de erros sistemáticos, propriedade explicitada pela fórmula (8.2). Também note que as variâncias σ_i^2 são *exatamente* conhecidas por hipótese, relação (8.3), o que traz consequências que detalharemos adiante.

A fim de obter a solução de mínimos quadrados do modelo linear em forma compacta, reescrevemos a expressão da soma dos quadrados dos resíduos, equação (4.18), na forma de um produto de matrizes, que inclui a matriz de variâncias,

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{bmatrix} .$$

que é diagonal, quadrada e de dimensão N . Adiante, verificaremos que não há qualquer necessidade desta matriz ser diagonal e, na maioria das deduções, usaremos apenas o fato de \mathbf{V} ser simétrica e possuir uma inversa. O método dos mínimos quadrados estima os parâmetros com o vetor \mathbf{a} que minimiza a função

$$Q = (\mathbf{y} - \mathbf{X}\mathbf{a})^t \cdot \mathbf{V}^{-1} \cdot (\mathbf{y} - \mathbf{X}\mathbf{a}) \quad . \quad (8.5)$$

Q8.1 Verifique que, quando \mathbf{V} é diagonal, a expressão acima é idêntica à (4.18).

Se $\hat{\mathbf{a}}$ corresponde ao vetor de parâmetros que minimiza Q , então

$$\left. \frac{\partial Q}{\partial a_\nu} \right|_{\hat{\mathbf{a}}} = 0 \quad \text{para } \nu = 1, \dots, \mu \quad . \quad (8.6)$$

Escrevendo os produtos das matrizes da relação (8.5) como somatórios,

$$Q = \sum_{i=1}^N \left([y_i - (\mathbf{X}\mathbf{a})_i] \sum_{j=1}^N (\mathbf{V}^{-1})_{ij} [y_j - (\mathbf{X}\mathbf{a})_j] \right) \quad ,$$

fica mais fácil calcular as derivadas,

$$\begin{aligned} \frac{\partial Q}{\partial a_\nu} = & - \sum_{i=1}^N \left(\frac{\partial(\mathbf{X}\mathbf{a})_i}{\partial a_\nu} \sum_{j=1}^N (\mathbf{V}^{-1})_{ij} [y_j - (\mathbf{X}\mathbf{a})_j] \right) \\ & - \sum_{i=1}^N \left([y_i - (\mathbf{X}\mathbf{a})_i] \sum_{j=1}^N (\mathbf{V}^{-1})_{ij} \frac{\partial(\mathbf{X}\mathbf{a})_j}{\partial a_\nu} \right) . \end{aligned}$$

Depois de permutar i e j no segundo somatório — afinal, são índices mudos — e usar o fato que \mathbf{V}^{-1} é uma matriz simétrica, percebe-se que os termos do lado direito são idênticos. Após efetuar a derivada de $(\mathbf{X}\mathbf{a})_i$ requerida, chega-se a

$$\begin{aligned} \frac{\partial Q}{\partial a_\nu} = & -2 \sum_{i=1}^N \left((\mathbf{X}^t)_{\nu i} \sum_{j=1}^N (\mathbf{V}^{-1})_{ij} [y_j - (\mathbf{X}\mathbf{a})_j] \right) \\ \frac{\partial Q}{\partial a_\nu} = & -2 \left(\mathbf{X}^t \mathbf{V}^{-1} [\mathbf{y} - \mathbf{X}\mathbf{a}] \right)_\nu . \end{aligned}$$

O conjunto das μ equações (8.6) pode, então, ser representado pela equação matricial

$$\mathbf{X}^t \mathbf{V}^{-1} [\mathbf{y} - \mathbf{X}\hat{\mathbf{a}}] = \mathbf{0} .$$

Deduz-se imediatamente que

$$\mathbf{X}^t \mathbf{V}^{-1} \mathbf{y} = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{X} \hat{\mathbf{a}} . \quad (8.7)$$

Quando $\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}$ não é singular, pode-se determinar a estimativa de mínimos quadrados pela relação

$$\hat{\mathbf{a}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y} . \quad (8.8)$$

Podemos identificar na equação acima a expressão $\hat{\mathbf{a}} = \mathbf{M}^{-1} \mathbf{D}$ da equação (4.24) do capítulo 4, com $\mathbf{D} = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}$ e $\mathbf{M} = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{X}$.

Note que

$$\mathbf{M}^t = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^t = \mathbf{X}^t (\mathbf{V}^{-1})^t (\mathbf{X}^t)^t = \mathbf{X}^t \mathbf{V}^{-1} \mathbf{X} = \mathbf{M} ,$$

ou seja, a matriz \mathbf{M} é simétrica, como já sabíamos. Essa dedução vale mesmo que a matriz \mathbf{V} não seja diagonal, que é o caso geral de interesse. Fizemos isso a fim de deixar claro que as equações (8.7,8.8) valem tanto para dados estatisticamente independentes quanto para dados correlacionados.

8.2 Exemplos de ajustes de parâmetros

Discutiremos aqui três exemplos simples de estimação pelo método dos mínimos quadrados a fim de ilustrar a teoria na forma em que a estamos desenvolvendo. Nesta seção, obteremos $\hat{\mathbf{a}}$. Retomaremos estes exemplos adiante para calcular as variâncias nos parâmetros e, mais no final do capítulo, para estimar as variâncias dos dados quando elas forem desconhecidas.

8.2.1 Exemplo A. Determinação do volume específico v a partir da medição de volume e massa de fragmentos

Quando se observa N fragmentos de um mesmo material, se V_i e M_i representam o volume e a massa do fragmento i com σ_i igual ao desvio padrão no volume V_i , a equação (8.1) pode ser escrita

$$V_i = M_i v + \epsilon_i \quad ,$$

com ϵ_i o erro (desconhecido) em V_i . Os elementos da fórmula matricial (8.4) são

$$\mathbf{y} = \begin{bmatrix} V_1 \\ \vdots \\ V_N \end{bmatrix} \quad , \quad \mathbf{X} = \begin{bmatrix} M_1 \\ \vdots \\ M_N \end{bmatrix} \quad , \quad \mathbf{V} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{bmatrix} \quad \text{e} \quad \hat{\mathbf{a}} = [\hat{v}] \quad .$$

A fórmula (8.8) fica

$$\hat{v} = \left\{ \begin{array}{l} [M_1 \dots M_N] \begin{bmatrix} \sigma_1^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^{-2} \end{bmatrix} \begin{bmatrix} M_1 \\ \vdots \\ M_N \end{bmatrix} \\ [M_1 \dots M_N] \begin{bmatrix} \sigma_1^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^{-2} \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_N \end{bmatrix} \end{array} \right\}^{-1} .$$

que dá

$$\hat{v} = \frac{\sum_{i=1}^N \frac{M_i V_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{M_i^2}{\sigma_i^2}} \quad . \quad (8.9)$$

A estimativa de mínimos quadrados é, portanto, a média de V_i/M_i ponderada por M_i^2/σ_i^2 . Note que σ_i/M_i é o desvio padrão associado a V_i/M_i , portanto esse resultado é a familiar média ponderada pelos inversos das variâncias (veja a questão Q4.5 do capítulo 4). Aqui, porém, mostramos que a média ponderada é o estimador da grandeza *mesmo que a f.d.p. dos dados não seja gaussiana*.

8.2.2 Exemplo B. Inclinações de um plano $y_i = a_1z_i + a_2x_i$

Neste exemplo, a variável dependente é y_i , com desvio padrão σ_i . As variáveis independentes são z_i e x_i e, de novo, temos N dados. Nesse caso,

$$\mathbf{X} = \begin{bmatrix} z_1 & x_1 \\ z_2 & x_2 \\ \vdots & \vdots \\ z_N & x_N \end{bmatrix} = [[\mathbf{z}] \ [\mathbf{x}]] = [\mathbf{z} \ \mathbf{x}] \quad .$$

A matriz de covariância, \mathbf{V} , é a mesma do Exemplo A. A fórmula (8.8) fica

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \left\{ \begin{bmatrix} \mathbf{z}^t \\ \mathbf{x}^t \end{bmatrix} \begin{bmatrix} \sigma_1^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{z} & \mathbf{x} \end{bmatrix} \right\}^{-1} \times \\ \begin{bmatrix} \mathbf{z}^t \\ \mathbf{x}^t \end{bmatrix} \begin{bmatrix} \sigma_1^{-2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{y} \end{bmatrix}$$

em que, após substituir os dados, obtém-se

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N \frac{z_i^2}{\sigma_i^2} & \sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} \\ \sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} & \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N \frac{z_i y_i}{\sigma_i^2} \\ \sum_{i=1}^N \frac{x_i y_i}{\sigma_i^2} \end{bmatrix} \quad . \quad (8.10)$$

8.2.3 Exemplo C. Coeficientes da reta $y_i = a_1 + a_2x_i$

Este caso pode ser analisado a partir do exemplo B acima, quando se faz

$$z_i = 1 \quad , \quad \forall i \quad .$$

Vamos detalhar o caso em que todas as observações têm mesma variância,

$$\sigma_i^2 = \sigma^2, \quad \text{igual para todo } i,$$

ficando para a questão 8.2 lidar com o caso de variâncias diferentes. Com essa simplificação,

$$\begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \frac{\sigma^2}{N \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum_{i=1}^N x_i^2 & -\sum_{i=1}^N x_i \\ -\sum_{i=1}^N x_i & N \end{bmatrix} \frac{1}{\sigma^2} \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

Assim, as estimativas dos coeficientes linear e angular são

$$\hat{a}_1 = \left(Y \sum x_i^2 - X \sum x_i y_i \right) \left(N \sum x_i^2 - X^2 \right)^{-1} \quad \text{e}$$

$$\hat{a}_2 = \left(-XY + N \sum x_i y_i \right) \left(N \sum x_i^2 - X^2 \right)^{-1},$$

respectivamente, onde escrevemos

$$X = \sum x_i \quad \text{e} \quad Y = \sum y_i \quad .$$

É interessante observar que

$$Y - \hat{a}_2 X = N \hat{a}_1 \quad , \quad (8.11)$$

o que pode ser reescrito como

$$\frac{Y}{N} = \hat{a}_1 + \hat{a}_2 \frac{X}{N} \quad . \quad (8.12)$$

Vemos, então, que o ponto $(X/N, Y/N)$ *pertence à reta ajustada*. Esse ponto pode, em um sentido razoável, ser chamado de *baricentro dos dados* e o simbolizaremos por (x_C, y_C) . Podemos também definir uma *dispersão da variável independente* como δ_x , por meio da relação

$$\delta_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - x_C)^2 = \frac{1}{N^2} \left(N \sum x_i^2 - X^2 \right) \quad . \quad (8.13)$$

As estimativas dos parâmetros podem ser reescritas como

$$\hat{a}_1 = \frac{1}{N \delta_x^2} \left(y_C \sum x_i^2 - x_C \sum x_i y_i \right) \quad \text{e}$$

$$\hat{a}_2 = \frac{1}{\delta_x^2} \left(-x_C y_C + \frac{1}{N} \sum x_i y_i \right) = \frac{1}{\delta_x^2} \frac{1}{N} \sum_{i=1}^N (x_i - x_C) (y_i - y_C) \quad . \quad (8.14)$$

Nesta última forma, fica claro que \hat{a}_2 *independe* da origem do sistema de coordenadas. Já \hat{a}_1 é dependente da origem dos eixos, o que talvez fique ainda mais claro reescrevendo a expressão (8.12),

$$\hat{a}_1 = y_C - \hat{a}_2 x_C \quad . \quad (8.12')$$

Esse resultado — \hat{a}_1 dependente da origem do sistema de coordenadas e \hat{a}_2 , não — pode ser interpretado se imaginamos que os coeficientes linear e angular representem, respectivamente, a posição inicial e a velocidade de um corpo em movimento retilíneo e uniforme, quando a velocidade e sua incerteza devem independer da origem do sistema, mas não a posição inicial.

Q8.2 *Escreva uma equação para x_C e y_C quando as variâncias dos diferentes dados experimentais são diferentes. O baricentro pertence à reta ajustada?*

8.3 O estimador de mínimos quadrados não é tendencioso no modelo linear

Substituindo (8.4) em (8.8), obtém-se

$$\hat{\mathbf{a}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} (\mathbf{X} \mathbf{a}_0 + \vec{\epsilon}) \quad ,$$

que pode ser rearranjado como

$$\hat{\mathbf{a}} = \mathbf{a}_0 + (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \vec{\epsilon} \quad . \quad (8.15)$$

A expressão acima explicita a relação entre os erros dos dados e os erros nas estimativas dos parâmetros. Como sempre, poderemos estimar médias e variâncias, ficando eternamente não resolvida a questão de descobrir os valores verdadeiros.

O valor esperado do vetor de estimativas dos parâmetros é

$$E(\hat{\mathbf{a}}) = E(\mathbf{a}_0) + (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} E(\vec{\epsilon}) \quad ,$$

porque nem \mathbf{V} nem \mathbf{X} são variáveis aleatórias. O último termo é nulo, por causa da hipótese de não-tendenciosidade dos dados, relação (8.2), de modo que

$$E(\hat{\mathbf{a}}) = \mathbf{a}_0 \quad , \quad (8.16)$$

o que demonstra que o estimador de mínimos quadrados não é tendencioso. Note que não se supôs que \mathbf{V} fosse diagonal, de modo que a propriedade de não tendenciosidade independe da existência de covariâncias entre os dados. Na realidade, a demonstração indica que essa propriedade não depende da escolha de \mathbf{V} . Além disso, frequentemente as estimativas $\hat{\mathbf{a}}$ são pouco sensíveis às estimativas dos desvios padrões e das covariâncias.

Essa propriedade não depende da forma da f.d.p. dos dados, de modo que o MMQ não é tendencioso tanto para dados gaussianos quanto não gaussianos. Já a dependência linear de y em relação aos parâmetros foi essencial na demonstração.

8.4 As variâncias das estimativas

O resultado (8.15) permite calcular os desvios padrões dos parâmetros. De fato, eles já foram calculados na seção 4.9, mas aqui a dedução emprega o formalismo matricial.

A matriz de covariâncias dos parâmetros é $\mathbf{V}(\hat{\mathbf{a}})$, definida por

$$\mathbf{V}(\hat{\mathbf{a}}) = E\left((\hat{\mathbf{a}} - \mathbf{a}_0)(\hat{\mathbf{a}} - \mathbf{a}_0)^t\right) \quad (8.17)$$

que, usando a eq. (8.15), fica

$$\mathbf{V}(\hat{\mathbf{a}}) = E\left((\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\vec{\epsilon}\vec{\epsilon}^t\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\right),$$

onde se levou em conta que as matrizes \mathbf{V} e $\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}$ são simétricas, bem como suas inversas. Como só $\vec{\epsilon}$ é variável aleatória e

$$E(\vec{\epsilon}\vec{\epsilon}^t) = \mathbf{V}$$

pela hipótese expressa por (8.3), deduz-se

$$\begin{aligned} \mathbf{V}(\hat{\mathbf{a}}) &= (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1} \\ \mathbf{V}(\hat{\mathbf{a}}) &= (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}, \end{aligned} \quad (8.18)$$

que pode ser identificada com \mathbf{M}^{-1} definida na seção 4.8.

O resultado (8.18) é independente da forma da f.d.p. dos dados, ou seja, vale para dados não normais também. No entanto, as probabilidades associadas a intervalos de confiança dependem da f.d.p. dos dados. Por exemplo, a

probabilidade do valor verdadeiro de um parâmetro estar contido no intervalo definido pela estimativa mais ou menos um desvio padrão só é garantidamente 68% caso os dados sejam gaussianos, embora a linearidade nos dados da fórmula (8.8) de cálculo das estimativas permite deduzir que sua f.d.p. tenderá à multinormal com o aumento do número de dados, pelo Teorema Central do Limite.

Veja que, na dedução da fórmula (8.18), consideramos que os elementos da matriz \mathbf{X} são exatos. No entanto, em muitas situações experimentais, as grandezas que determinam esses elementos são variáveis aleatórias. Nesse caso, quando a flutuação dessas grandezas causa um impacto pequeno em $\hat{\mathbf{a}}$, uma aproximação melhor é adicionar a incerteza nessas grandezas propagadas de acordo com a relação (3.9),

$$\mathbf{V}(\hat{\mathbf{a}}) = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} + \mathbf{C} \mathbf{V}_x \mathbf{C}^t \quad , \quad (8.19)$$

em que \mathbf{V}_x é a matriz de covariâncias dessas grandezas x_κ que influem na matriz de planejamento e

$$C_{\nu, \kappa} = \frac{\partial \hat{a}_\nu}{\partial x_\kappa}$$

em que a derivada deve ser calculada nos valores estimados de todas as grandezas envolvidas, o que normalmente é feito de forma numérica.

Além das covariâncias entre os parâmetros ajustados entre si, a relação entre os parâmetros ajustados e os dados experimentais conduz também a covariâncias entre eles, dados experimentais e parâmetros. Esse assunto é discutido na referência [Guimarães-Filho].

8.5 Exemplos: variâncias dos parâmetros

Retomamos os exemplos **A**, **B** e **C** da seção 8.2, calculando as variâncias dos parâmetros.

A) A partir da fórmula (8.18), obtém-se

$$\sigma_v^2 = \left(\sum_{i=1}^N \frac{M_i^2}{\sigma_i^2} \right)^{-1} \quad . \quad (8.20)$$

Note que a fórmula usual de propagação de incerteza, relação (3.5), exata no caso de um estimador linear como o da expressão (8.9), fornece o mesmo

resultado, e que ambos os procedimentos — usando a fórmula (8.18) ou a (3.5), da variância de uma função de variáveis aleatórias — não exigem que os dados sejam normais, sendo, de fato, válidos mesmo para dados não gaussianos.

B) Com a relação (8.18), calcula-se

$$V(\hat{a}) = \frac{1}{\sum_{i=1}^N \frac{z_i^2}{\sigma_i^2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} \right)^2} \begin{bmatrix} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} & - \sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} \\ - \sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} & \sum_{i=1}^N \frac{z_i^2}{\sigma_i^2} \end{bmatrix} .$$

Ao isolar a variância de \hat{a}_1 , chega-se a

$$\sigma_{a_1}^2 = \frac{\sum_{i=1}^N \frac{x_i^2}{\sigma_i^2}}{\sum_{i=1}^N \frac{z_i^2}{\sigma_i^2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} \right)^2} . \quad (8.21)$$

A variância de \hat{a}_2 pode ser obtida a partir da expressão acima por permutação de z_i com x_i , o que modifica apenas o numerador. A covariância entre os parâmetros é

$$\text{cov}(\hat{a}_1, \hat{a}_2) = \frac{- \sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{z_i^2}{\sigma_i^2} \sum_{i=1}^N \frac{x_i^2}{\sigma_i^2} - \left(\sum_{i=1}^N \frac{z_i x_i}{\sigma_i^2} \right)^2} ,$$

que será nula somente se o somatório do numerador for nulo.

C) As variâncias dos coeficientes linear, \hat{a}_1 , e angular, \hat{a}_2 , e a covariância entre eles são, respectivamente,

$$\sigma_{a_1}^2 = \frac{\sigma^2 \sum_{i=1}^N x_i^2}{N \sum x_i^2 - (\sum x_i)^2} = \frac{\sigma^2 \sum x_i^2}{N N \delta_x^2} , \quad (8.22)$$

$$\sigma_{a_2}^2 = \frac{N \sigma^2}{N \sum x_i^2 - (\sum x_i)^2} = \frac{\sigma^2}{N \delta_x^2} \quad \text{e} \quad (8.23)$$

$$\text{cov}(\hat{a}_1, \hat{a}_2) = - \frac{\sigma^2 x_C}{N \delta_x^2} , \quad (8.24)$$

onde x_C é a abscissa do baricentro dos dados e a dispersão da variável independente, δ_x , já foi definida anteriormente,

$$\delta_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - x_C)^2 = \frac{\sum x_i^2}{N} - \left(\frac{\sum x_i}{N} \right)^2 . \quad (8.13)$$

Note que $\sigma_{a_2}^2$ é independente da escolha de origem do sistema de coordenadas conforme a relação (8.23), em acordo com a afirmação do final da discussão deste exemplo na seção 8.2, e sua dependência nos x_i restringe-se ao inverso do quadrado da dispersão dos dados. Já $\sigma_{a_1}^2$ e $\text{cov}(a_1, a_2)$ dependem da origem, sendo que a primeira é tanto maior quanto mais afastado da origem estiver o conjunto dos pontos. A covariância é nula para $x_C = 0$, ou seja, se o baricentro dos pontos experimentais estiver sobre o eixo Oy . Desses resultados, o mais difícil de entender é a situação em que a covariância é nula, cuja interpretação pode ser feita usando o resultado (8.12'). Note a dependência da variância com o inverso do número de dados, já familiar.

As calculadoras com regressão linear, salvo raríssimas exceções, fazem um ajuste com incerteza igual a 1 para todos os pontos e não devolvem as incertezas dos parâmetros. No entanto, pode-se recuperar as somatórias em x_i , x_i^2 , y_i e $x_i y_i$ e calcular as incertezas como em (8.22) e (8.23), sem esquecer de multiplicar por σ . Isso, porém, não funciona quando os pontos experimentais têm incertezas diferentes ou há covariâncias entre os dados experimentais.

8.6 Variância mínima no modelo linear

Vamos mostrar que o estimador da fórmula (8.8) é o de Variância Mínima não só para os parâmetros a_ν , mas também para qualquer combinação linear $A_\ell(\mathbf{a})$ dos parâmetros,

$$A_\ell(\mathbf{a}) = \sum_{\nu} \phi_{\ell\nu} a_\nu \quad , \quad (8.25)$$

um resultado bastante abrangente, que pode ser especializado para um único parâmetro a_η com a escolha $\phi_{\ell\nu} = \delta_{\nu\eta}$. Essa maneira mais geral foi escolhida porque, assim, não ficam dúvidas acerca da validade da propriedade de mínima variância em situações particulares interessantes, como por exemplo a interpolação por meio da função ajustada, uma vez que $\hat{y} = y(x, \hat{\mathbf{a}})$ é função linear dos \hat{a}_ν . Como diversas combinações lineares nos a_ν , $A_\ell(\mathbf{a})$, serão estimadas ao mesmo tempo (os vários parâmetros ou diversos pontos interpolados, como exemplos), o objetivo será determinar toda a matriz de covariâncias. Atribui-se a descoberta dessa propriedade de variância mínima a Gauss.

O conjunto das relações (8.25) pode ser escrita em forma matricial,

$$\mathbf{A} = \mathbf{F}\mathbf{a} \quad , \quad (8.26)$$

em que o vetor \mathbf{A} tem como componentes as combinações $A_\ell(\mathbf{a})$ dos parâmetros e \mathbf{F} é a matriz formada pelos coeficientes $\phi_{\ell\nu}$, com μ colunas e tantas linhas quantas grandezas se pretende estimar. Assim, a relação entre os valores verdadeiros dos A_ℓ e os dos parâmetros é dada por

$$\mathbf{A}_0 = \mathbf{F}\mathbf{a}_0 \quad . \quad (8.27)$$

A estimativa de mínimos quadrados de \mathbf{A} da expressão (8.26) é

$$\hat{\mathbf{A}} = \mathbf{F}\hat{\mathbf{a}} = \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\mathbf{y} = \mathbf{T}\mathbf{y} \quad , \quad (8.28)$$

onde $\hat{\mathbf{a}}$ foi substituído pela expressão (8.8) e \mathbf{T} é a matriz

$$\mathbf{T} = \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1} \quad , \quad (8.29)$$

que independe de \mathbf{y} , de modo que a equação (8.28) põe em evidência a relação linear entre A_ℓ e os dados experimentais.

Suponha a existência de outro estimador de \mathbf{A} que também seja linear em \mathbf{y} , de modo que

$$\mathbf{A}' = \mathbf{T}'\mathbf{y} \quad . \quad (8.30)$$

Agora, impomos a condição de \mathbf{T}' estimar \mathbf{A} sem tendenciosidade, porque essa é uma propriedade mais relevante que a de variância mínima. Isso corresponde a exigir que o valor esperado dessa outra estimativa seja idêntico a \mathbf{A}_0 ,

$$E(\mathbf{A}') = \mathbf{A}_0 \quad .$$

Substituindo \mathbf{A}' pela expressão (8.30), obtém-se

$$E(\mathbf{A}') = E(\mathbf{T}'\mathbf{y}) = E[\mathbf{T}'(\mathbf{X}\mathbf{a}_0 + \vec{\epsilon})] = \mathbf{T}'\mathbf{X}\mathbf{a}_0 \quad ,$$

em que se usou y da expressão (8.4) e a condição de não-tendenciosidade dos dados, relação (8.2). Assim,

$$\mathbf{T}'\mathbf{X}\mathbf{a}_0 = \mathbf{A}_0 \quad .$$

Agora, usando (8.27), chega-se à relação

$$\mathbf{T}'\mathbf{X} = \mathbf{F} \quad . \quad (8.31)$$

Essa é a condição necessária e suficiente para \mathbf{T}' estimar \mathbf{A} sem tendenciosidade, o que deixa claro que existe uma infinidade de estimadores lineares não

tendenciosos para \mathbf{A} . É imediato verificar que $\mathbf{TX} = \mathbf{F}$ a partir da definição de \mathbf{T} , fórmula (8.29), ou seja, o estimador de mínimos quadrados também não é tendencioso.

A partir de uma expressão análoga à (8.17), a matriz de covariância de \mathbf{A}' é

$$\begin{aligned} V(\mathbf{A}') &= E((\mathbf{A}' - \mathbf{F}\mathbf{a}_0)(\mathbf{A}' - \mathbf{F}\mathbf{a}_0)^t) \\ V(\mathbf{A}') &= E(\mathbf{T}'\epsilon\epsilon^t\mathbf{T}^t) = \mathbf{T}'\mathbf{V}\mathbf{T}^t \quad , \end{aligned} \quad (8.32)$$

com uma expressão semelhante para $\mathbf{V}(\hat{\mathbf{A}})$.

Desejamos demonstrar que os elementos diagonais da matriz de covariância de $\hat{\mathbf{A}}$ são menores que os elementos correspondentes da matriz de covariância de \mathbf{A}' , em outras palavras, a tese é

$$(\mathbf{T}\mathbf{V}\mathbf{T}^t)_{ii} \leq (\mathbf{T}'\mathbf{V}\mathbf{T}^t)_{ii} \quad \text{para todo } i.$$

A fim de estabelecer se essa desigualdade é satisfeita, recalcula-se $\mathbf{V}(\mathbf{A}')$ com um truque:

$$\begin{aligned} V(\mathbf{A}') &= E\left([\mathbf{A}' - \hat{\mathbf{A}}] + [\hat{\mathbf{A}} - \mathbf{A}_o][\mathbf{A}' - \hat{\mathbf{A}}] + [\hat{\mathbf{A}} - \mathbf{A}_o]^t\right) = \\ &= E\left([\mathbf{A}' - \hat{\mathbf{A}}][\mathbf{A}' - \hat{\mathbf{A}}]^t\right) + E\left([\hat{\mathbf{A}} - \mathbf{A}_o][\hat{\mathbf{A}} - \mathbf{A}_o]^t\right) + 2E\left([\mathbf{A}' - \hat{\mathbf{A}}][\hat{\mathbf{A}} - \mathbf{A}_o]^t\right) . \end{aligned}$$

Escrevendo $E\left([\mathbf{A}' - \hat{\mathbf{A}}][\mathbf{A}' - \hat{\mathbf{A}}]^t\right) = V(\mathbf{A}' - \hat{\mathbf{A}})$, cujos termos diagonais são positivos porque são os valores esperados dos quadrados das diferenças (não muito grandes, porque $E(\mathbf{A}' - \hat{\mathbf{A}}) = 0$, uma vez que ambos os estimadores não são tendenciosos), chega-se a

$$V(\mathbf{A}') = V(\mathbf{A}' - \hat{\mathbf{A}}) + V(\hat{\mathbf{A}}) + 2E\left([\mathbf{A}' - \hat{\mathbf{A}}][\hat{\mathbf{A}} - \mathbf{A}_o]^t\right) \quad . \quad (8.33)$$

Assim, para comparar $V(\mathbf{A}')$ e $V(\hat{\mathbf{A}})$, falta entender o último termo, que é, a menos do fator 2,

$$\begin{aligned} &E\left([\mathbf{A}' - \hat{\mathbf{A}}][\hat{\mathbf{A}} - \mathbf{A}_o]^t\right) = \\ &= E\left([\mathbf{T}'\bar{\epsilon} - \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\bar{\epsilon}][\mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}\bar{\epsilon}]^t\right) = \\ &= E\left([\mathbf{T}' - \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}]\bar{\epsilon}\bar{\epsilon}^t\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}^t\right) = \\ &= [\mathbf{T}' - \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}]E(\bar{\epsilon}\bar{\epsilon}^t)\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}^t = \\ &= [\mathbf{T}' - \mathbf{F}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1}]\mathbf{X}(\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}^t = \end{aligned}$$

$$\begin{aligned} & [\mathbf{T}'\mathbf{X} - \mathbf{F}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}](\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}' = \\ & [\mathbf{T}'\mathbf{X} - \mathbf{F}](\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}' = 0 \quad , \end{aligned}$$

onde foram usadas diversas propriedades de simetria de matrizes, a definição de \mathbf{V} e, na última passagem, a condição (8.31). Assim, de (8.33), obtém-se a relação

$$V(\mathbf{A}') = V(\mathbf{A}' - \hat{\mathbf{A}}) + V(\hat{\mathbf{A}}) \quad .$$

Como a *diagonal* da matriz $\mathbf{V}(\mathbf{A}' - \hat{\mathbf{A}})$ é definida positiva ou nula, decorre

$$V(\mathbf{A}')_{ii} \geq V(\hat{\mathbf{A}})_{ii} \quad ,$$

o que mostra que qualquer estimador linear não tendencioso diferente do estimador de mínimos quadrados tem variância maior ou igual ao de mínimos quadrados.

Aproveitamos para deduzir mais uma vez a fórmula da matriz de variâncias de uma função de variáveis aleatórias, substituindo \mathbf{T} da expressão (8.28) na fórmula (8.32), em que trocamos \mathbf{A}' por $\hat{\mathbf{A}}$ e \mathbf{T}' por \mathbf{T} , o que dá

$$\begin{aligned} V(\hat{\mathbf{A}}) &= \mathbf{F}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}' \\ V(\hat{\mathbf{A}}) &= \mathbf{F}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{F}' \quad , \end{aligned} \quad (8.34)$$

que é a mesma que a expressão (3.9), sem o \cong por causa da linearidade da transformação.

8.7 A média como estimativa linear de variância mínima

Vamos ilustrar o resultado da seção anterior com o caso em que todas as medições $\{y_i\}$ de uma grandeza têm o mesmo desvio-padrão e são independentes, explorando o aspecto geométrico do problema.

Nas fórmulas que seguem, o valor esperado da grandeza é θ e o desvio-padrão dos dados é σ . Esta dedução é válida independentemente da f.d.p. dos dados, assim como o resultado obtido na seção precedente.

Neste caso, a relação entre as observações e θ é

$$y_j = \theta + \epsilon_j \quad j = 1, \dots, N \quad ,$$

onde ϵ_j é o erro associado ao j -ésimo dado.

Os estimadores de θ lineares nas observações y_i têm a forma geral

$$t = \sum_{j=1}^N c_j y_j \quad (8.35)$$

e, para que não sejam tendenciosos, impõe-se a condição

$$\sum_{j=1}^N c_j = 1 \quad . \quad (8.36)$$

Q8.3 *Mostre que a condição (8.36) é necessária e suficiente para estimar θ sem tendenciosidade por uma relação linear, fórmula (8.35).*

A partir da fórmula (3.5), da variância de uma função de variáveis aleatórias, que é exata neste caso, ou usando que a variância é um cumulante (seção 6.15), a variância de t é

$$\text{var}(t) = \sigma^2 \sum_{j=1}^N c_j^2 \quad . \quad (8.37)$$

No espaço N -dimensional dos c_j , a equação (8.36) representa um hiper-plano que corta todos os eixos Oc_k em $c_k = 1, \forall k$. A equação (8.37) dá, a menos do fator σ^2 , a distância entre um ponto definido pelo vetor \mathbf{c} , correspondente a um certo t , e a origem do sistema coordenado. A perpendicular ao (hiper)plano que passa pela origem define, então, o ponto de distância mínima (ponto onde a perpendicular intercepta o plano), correspondente ao vetor \mathbf{c} que, por simetria, terá todas as componentes iguais. A condição (8.36) impõe que este vetor \mathbf{c} tem componentes iguais a $1/N$. Assim,

$$t \text{ de variância mínima} = \sum_{j=1}^N \frac{1}{N} y_j = \bar{y} \quad \text{QED.}$$

A figura 8.1 mostra, no caso de duas variáveis, a função de vínculo entre os parâmetros e a condição de mínimo.

A expressão (8.37) fornece a variância

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{N} \quad ,$$

resultado que já nos é familiar.

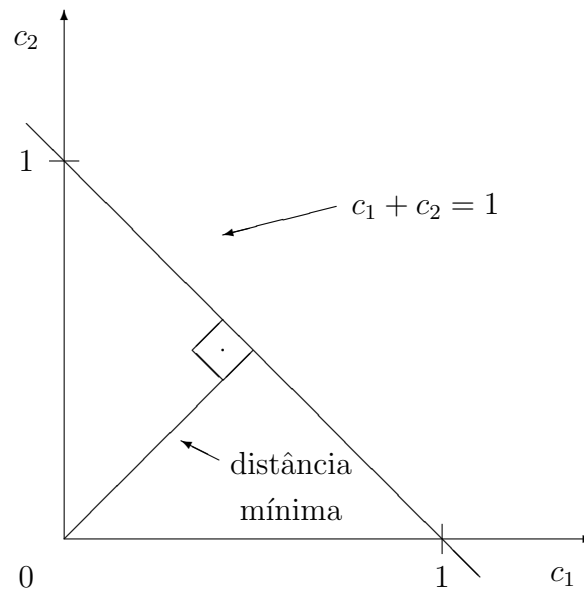


Figura 8.1: Ilustração geométrica do vínculo entre os coeficientes c_j para que o estimador seja não-tendencioso, fórmula (8.36), e da condição de variância mínima, fórmula (8.37).

A variância dada por essa expressão é exata e independe da f.d.p. dos dados, mas, se ela não for gaussiana, então a f.d.p. de \bar{y} também não será, embora o Teorema Central do Limite indique que ela tende à gaussiana com o aumento de N desde que cada dado tenha variância finita.

8.8 Estimativas das variâncias

De acordo com o desenvolvimento do capítulo 4, a estimativa não tendenciosa da variância de um conjunto de dados experimentais $\{x_i, i = 1, \dots, N\}$ que obedecem a uma mesma f.d.p. é

$$\sigma^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2 \quad . \quad (1.10)$$

Um exame dessa expressão mostra que a variância é estimada a partir dos resíduos observados, $x_i - \bar{x}$. Um resultado análogo pode ser obtido quando o conjunto de dados $\{y_i\}$ é descrito por um modelo linear com parâmetros \mathbf{a}_0 , ou seja, quando $\mathbf{Y} = \mathbf{X}\mathbf{a}_0 + \epsilon$, com \mathbf{X} a matriz de planejamento e ϵ um

vetor de erros aleatórios. Como há muitos resíduos estimados $(\mathbf{y} - \mathbf{X}\hat{\mathbf{a}})_i$, pode-se calcular uma certa média dos seus quadrados e usá-la como estimativa da variância; com o procedimento adequado, o resultado obtido não é tendencioso. Nesta seção, deduziremos detalhadamente como estimar essa variância comum, começando por encontrar a relação entre os resíduos — observáveis — e os erros — impossíveis de determinar.

Começaremos com a hipótese das variâncias dos dados serem todas iguais a σ^2 e as covariâncias, nulas; essa é uma hipótese muito restritiva, que será suavizada adiante, mas vamos adotá-la por enquanto. Assim, a matriz de covariâncias é

$$\mathbf{V} = \sigma^2 \mathbf{I}_N \quad , \quad (8.38)$$

com \mathbf{I}_N a matriz identidade de ordem N . Quando a variância independe do dado y_i , ela pode ser fatorada nos somatórios, o que simplifica muitas das fórmulas. Assim, a função Q da expressão (8.5) fica, simplesmente,

$$Q = (\mathbf{y} - \mathbf{X}\mathbf{a})^t (\mathbf{y} - \mathbf{X}\mathbf{a}) / \sigma^2 \quad .$$

A estimativa de mínimos quadrados, fórmula (8.8), fica

$$\hat{\mathbf{a}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} \quad ,$$

independente de σ^2 , o que permite calcular as variâncias dos parâmetros ajustados.

Os valores da grandeza y calculados com os parâmetros ajustados nos valores observados da variável independente são dados por

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{a}} = \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} = \mathbf{H}\mathbf{y} \quad , \quad (8.39)$$

onde se define a matriz de projeção ortogonal \mathbf{H} ,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \quad , \quad (8.40)$$

também chamada *matriz chapéu* (no caso geral, em que as covariâncias não são nulas e/ou as variâncias não são idênticas, ela deve ser calculada como $\mathbf{H} = \mathbf{X}(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1}$). Os resíduos são, então,

$$\mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{H}\mathbf{y} = [\mathbf{I}_N - \mathbf{H}]\mathbf{y} \quad , \quad (8.41)$$

que podem ser reescritos em função dos erros quando se substitui a relação (8.4) entre \mathbf{y} e $\tilde{\boldsymbol{\epsilon}}$, com o resultado

$$\mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I}_N - \mathbf{H}](\mathbf{X}\mathbf{a}_0 + \tilde{\boldsymbol{\epsilon}}) = [\mathbf{I}_N - \mathbf{H}]\mathbf{X}\mathbf{a}_0 + [\mathbf{I}_N - \mathbf{H}]\tilde{\boldsymbol{\epsilon}} \quad .$$

O primeiro termo do lado direito desta última equação, substituindo \mathbf{H} da equação (8.40), fica

$$[\mathbf{I}_N - \mathbf{H}]\mathbf{X}\mathbf{a}_0 = \mathbf{I}_N\mathbf{X}\mathbf{a}_0 - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X}\mathbf{a}_0 = \mathbf{X}\mathbf{a}_0 - \mathbf{X}\mathbf{a}_0 = \mathbf{0} \quad ,$$

porque $(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{X} = \mathbf{I}_\mu$. Assim, os resíduos podem ser calculados pela expressão

$$\mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I}_N - \mathbf{H}]\vec{\epsilon} = [\mathbf{I}_N - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]\vec{\epsilon} \quad . \quad (8.42)$$

Observando (8.42) e (8.41), verifica-se que os resíduos observados, $\mathbf{y} - \hat{\mathbf{y}}$, dependem dos dados, \mathbf{y} , da mesma forma com que dependem de $\vec{\epsilon}$, o vetor de erros (verdadeiros). Assim, com os resíduos observados, pode-se estimar o valor médio *quadrático do erro*.²

Q8.4 Mostre que a matriz \mathbf{A} ,

$$\mathbf{A} = [\mathbf{I}_N - \mathbf{H}] = \mathbf{I}_N - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \quad , \quad (8.43)$$

é simétrica ($\mathbf{A} = \mathbf{A}^t$) e idempotente ($\mathbf{A} \cdot \mathbf{A} = \mathbf{A}$, de modo que $\mathbf{A}^n = \mathbf{A}$, para qualquer n inteiro positivo).

A soma dos quadrados dos resíduos é

$$S = (\mathbf{y} - \hat{\mathbf{y}})^t(\mathbf{y} - \hat{\mathbf{y}}) \quad ,$$

donde, usando (8.42), obtém-se

$$S = ([\mathbf{I}_N - \mathbf{H}]\vec{\epsilon})^t [\mathbf{I}_N - \mathbf{H}]\vec{\epsilon} = \vec{\epsilon}^t [\mathbf{I}_N - \mathbf{H}]\vec{\epsilon} = \vec{\epsilon}^t \mathbf{A} \vec{\epsilon} \quad .$$

Para chegar ao resultado acima, considerou-se que $\mathbf{A} = [\mathbf{I}_N - \mathbf{H}]$ é simétrica e idempotente, conforme questão Q8.4. O valor esperado de S é

$$\langle S \rangle = \left\langle \sum_{i,j} \epsilon_i A_{ij} \epsilon_j \right\rangle = \sum_{i,j} A_{ij} \langle \epsilon_i \epsilon_j \rangle = \sum_i A_{ii} \sigma^2 \quad ,$$

onde usou-se que $\langle \epsilon_i \epsilon_j \rangle = \sigma^2 \delta_{ij}$, que é a hipótese da relação (8.38). Note que os resíduos observados são correlacionados, isto é, $\langle [y_i - \hat{y}_i][y_j - \hat{y}_j] \rangle \neq 0$

²O fato de $\vec{\epsilon}$ ter N componentes e existirem N resíduos dá a impressão de que se possa achar $\vec{\epsilon}$ a partir do sistema de N equações a N incógnitas. Para entender que isso não é possível, note que \mathbf{y} e $\vec{\epsilon}$ são ambos soluções desse sistema de N equações, conforme (8.41) e (8.42), respectivamente, de modo que a matriz dos coeficientes do sistema linear, $[\mathbf{I}_N - \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]$, não é inversível.

em geral,³ como detalharemos na seção 8.10 abaixo, mas na expressão acima aparecem os erros verdadeiros, que *não são correlacionados*, devido à hipótese das covariâncias entre os dados serem nulas. Pode-se concluir que

$$\langle S \rangle = \sigma^2 \sum_i^N A_{ii} = \sigma^2 \text{tr}(\mathbf{A})$$

onde $\text{tr}(\mathbf{A})$ lê-se *traço de A*, que é a soma dos elementos diagonais da matriz \mathbf{A} . Assim,

$$\langle S \rangle = \sigma^2 \text{tr}[\mathbf{I}_N - \mathbf{H}] = \sigma^2 \text{tr}(\mathbf{I}_N - \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t) \quad . \quad (8.44)$$

Com as propriedades do traço de uma matriz,

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B}) \quad \text{e}$$

$$\text{tr}(\mathbf{X}\mathbf{Y}) = \text{tr}(\mathbf{Y}\mathbf{X}) \quad ,$$

calcula-se⁴ que

$$\begin{aligned} \langle S \rangle &= \sigma^2 \{ \text{tr}(\mathbf{I}_N) - \text{tr}[\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] \} \\ \langle S \rangle &= \sigma^2 \{ N - \text{tr}[\mathbf{X}^t \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1}] \} \\ \langle S \rangle &= \sigma^2 \{ N - \text{tr}(\mathbf{I}_\mu) \} = \sigma^2 (N - \mu) \quad . \end{aligned} \quad (8.45)$$

Finalmente,

$$\hat{\sigma}^2 = \frac{S}{N - \mu} = \frac{(\mathbf{y} - \hat{\mathbf{y}})^t (\mathbf{y} - \hat{\mathbf{y}})}{N - \mu} \quad (8.46)$$

é um estimador não tendencioso da variância dos dados experimentais, quando todos têm mesma variância.

A estimativa tradicional da variância da medida de uma grandeza com N dados (fórmula (1.10), reproduzida no início desta seção) é um caso particular da fórmula (8.46).

³A curva ajustada é “atraída” pelos pontos experimentais, de forma que ela se situa entre eles. Assim, uma curva cujos parâmetros foram ajustados a partir de um conjunto de pontos experimentais nunca pode passar acima de todos eles (nem abaixo), embora a totalidade dos pontos experimentais possa estar abaixo (ou acima) da curva verdadeira. Esta é uma evidência qualitativa da correlação entre os resíduos observados.

⁴Se $\mathbf{X}\mathbf{Y}$ é quadrada, então $\mathbf{Y}\mathbf{X}$ também será, embora possa ser de ordem diferente. Assim, se um dos traços existe, o outro também existe — lembre que só é possível calcular traço de uma matriz quadrada.

8.9 Exemplos: estimativa da variância dos dados

Vamos retomar os exemplos **A**, **B** e **C** da seção 8.2. Em todos eles, suporemos $\sigma_i^2 = \sigma^2$, igual para todos os N dados, mas desconhecido, e usaremos o método da seção 8.8 para estimar essa variância comum. Uma vez obtida essa estimativa da variância do dado, ela pode ser usada para calcular as variâncias dos parâmetros, com as fórmulas apresentadas na seção 8.5.

A) Se $\sigma_i^2 = \sigma^2, \forall i$,

$$\hat{v} = \frac{\sum_i M_i V_i}{\sum_i M_i^2} .$$

Com essa estimativa de v , calcula-se

$$S = \sum_i^N (V_i - M_i \hat{v})^2$$

e, com a fórmula (8.46) acima, obtém-se a estimativa da variância dos dados,

$$\hat{\sigma}^2 = \frac{S}{N-1} .$$

A semelhança da expressão acima com a estatística não-tendenciosa da variância habitual (fórmula 1.10) não é acidental, uma vez que essa última fórmula aplica-se também ao caso da medida de um único parâmetro com um conjunto de dados $\{x_i\}$. Esse resultado mostra que a variância deve ser estimada sempre pela fórmula (1.10), sejam os dados x_i normais ou não, mas no primeiro caso a f.d.p. desse $\hat{\sigma}^2$ é análoga à de χ^2 , caso contrário é necessário, em princípio, determiná-la. Porém, note que, se o número de dados é grande, essa f.d.p. será normal nas situações em que o Teorema Central do Limite se aplica, porque S é uma soma de muitos termos.

B) Com $\sigma_i^2 = \sigma^2$, as estimativas dos parâmetros são

$$\hat{a}_1 = \frac{(\sum x_i^2 \sum z_i y_i - \sum z_i x_i \sum x_i y_i)}{\sum z_i^2 \sum x_i^2 - (\sum z_i x_i)^2} \quad \text{e}$$

$$\hat{a}_2 = \frac{(\sum z_i^2 \sum x_i y_i - \sum z_i x_i \sum z_i y_i)}{\sum z_i^2 \sum x_i^2 - (\sum z_i x_i)^2} .$$

A estimativa da variância dos y_i é, simplesmente,

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^N (y_i - \hat{a}_1 z_i - \hat{a}_2 x_i)^2 \quad .$$

C) Neste caso, supusemos desde o início que todas as variâncias fossem iguais. As estimativas \hat{a}_1 e \hat{a}_2 são, portanto, dadas pelas expressões (8.14).

A soma dos quadrados dos resíduos é

$$S = \sum_{i=1}^N (y_i - \hat{a}_1 - \hat{a}_2 x_i)^2 \quad , \quad (8.47)$$

e a fórmula (8.46) permite estimar a variância,

$$\hat{\sigma}^2 = \frac{S}{N-2} \quad . \quad (8.48)$$

Substituindo σ^2 das fórmulas (8.22, 8.23 e 8.24) por esta estimativa, determinam-se as variâncias de \hat{a}_1 e \hat{a}_2 e a covariância entre eles, o que completa o exemplo. Este caso particular é muito comum e permite uma análise detalhada, que está apresentada na subseção abaixo.

8.9.1 Análise de variância — duas variáveis

Quando se ajustam os parâmetros a e b da reta $y = a+bx$ a dados experimentais $\{(x_i, y_i, \sigma_i)\}$, onde $\sigma_i^2 = \text{var}(y_i)$, não é incomum que se deseje saber quanto da variação da grandeza y se deve ao efeito da variação em x e quanto, à flutuação estatística.

A primeira etapa é relacionar a soma dos quadrados dos resíduos com as dispersões dos dados. Inicialmente, reduz-se a expressão de S da relação (8.47) acima pela eliminação de \hat{a}_1 com a relação (8.12') entre \hat{a}_1 e \hat{a}_2 , o que fornece

$$S = \sum_{i=1}^N ((y_i - y_C) - \hat{a}_2(x_i - x_C))^2 \quad .$$

Essa forma deixa evidente que S é independente da origem do sistema de coordenadas. Em seguida, expande-se o quadrado e substitui-se \hat{a}_2 , ficando com uma expressão que depende de δ_x , que é a dispersão em x dos dados conforme a fórmula (8.13), com o resultado

$$S = \sum_i (y_i - y_C)^2 - \frac{1}{N\delta_x^2} \left(\sum_i (x_i - x_C)(y_i - y_C) \right)^2 \quad .$$

Agora se definem δ_y^2 e δ_{xy} com fórmulas análogas à (8.13) para δ_x ,

$$\delta_y^2 = \frac{1}{N} \sum_i^N (y_i - y_C)^2 \quad \text{e} \quad (8.49)$$

$$\delta_{xy} = \frac{1}{N} \sum_i^N (x_i - x_C)(y_i - y_C) \quad , \quad (8.50)$$

e repetimos neste texto a fórmula de δ_x para juntar todas definições,

$$\delta_x^2 = \frac{1}{N} \sum_i^N (x_i - x_C)^2 \quad . \quad (8.13)$$

Assim, pode-se escrever a somatória como a expressão compacta

$$S = \frac{N(\delta_y^2 \delta_x^2 - \delta_{xy}^2)}{\delta_x^2} \quad . \quad (8.51)$$

Define-se o coeficiente de correlação r entre x e y à grandeza adimensional

$$r = \frac{\delta_{xy}}{\delta_x \delta_y} \quad , \quad (8.52)$$

que, substituído na fórmula (8.46) da variância dos y , dá

$$\hat{\sigma}^2 = \frac{N}{N-2} \delta_y^2 (1 - r^2) \quad , \quad (8.53)$$

em que se usou S da fórmula (8.51). A partir da expressão acima, verifica-se que, para um valor fixo de δ_y^2 , o desvio padrão dos dados y_i aumenta quando a correlação r entre x e y diminui. Em particular, se $|r| = 1$, a incerteza em y é nula, caso em que todos os pontos se alinham exatamente em uma reta comum, com o valor de cada y_i *exatamente* determinado pelo valor de x_i . Por outro lado, quando $r = 0$, a equação (8.53) simplifica, de modo que, ao substituir nela a relação (8.49), a variância dos y_i fica

$$\hat{\sigma}^2 = \frac{N}{N-2} \frac{1}{N} \sum_i^N (y_i - y_C)^2 = \frac{1}{N-2} \sum_i^N (y_i - y_C)^2 \quad ,$$

que é praticamente a mesma estimativa da variância de $\{y_i\}$ que seria feita se y fosse uma grandeza constante, independente de x ,

$$\sigma^2 = \frac{1}{N-1} \sum_i^N (y_i - y_C)^2 \quad .$$

Quando $|r| = 0$, então, o conjunto dos $\{y_i\}$ comporta-se como uma variável aleatória na qual a variável independente x_i não influi *nada*. Valores intermediários de $|r|$ denotam que fração da dispersão δ_y é explicada pela dependência com x através da relação *linear* $y = a_1 + a_2x$.

Apenas como ilustração, a variância de \hat{a}_2 , utilizando σ^2 da expressão (8.53) na relação (8.23), fica

$$\sigma_{a_2}^2 = \frac{1}{N-2} \frac{\delta_y^2}{\delta_x^2} (1-r^2) \quad .$$

Note que r é totalmente distinto da correlação ρ entre os parâmetros \hat{a}_1 e \hat{a}_2 ajustados, que pode ser calculada a partir das expressões (8.22), (8.23) e (8.24),

$$\rho = \frac{\text{cov}(\hat{a}_1, \hat{a}_2)}{\sigma_{a_1} \sigma_{a_2}} = -\frac{x_c}{\sqrt{\frac{\sum x_i^2}{N}}} \quad . \quad (8.54)$$

Em particular, r depende tanto dos x_i quanto dos y_i , veja (8.52), enquanto ρ depende **apenas** dos x_i . Para ilustrar esta questão, a figura 8.2 apresenta dois conjuntos de pontos (x_i, y_i) que têm coeficientes de correlação entre x e y muito diferentes, mas que, ao ajustar-se uma reta $y = a_1 + a_2x$ com a suposição de variâncias iguais, têm o **mesmo** coeficiente de correlação entre \hat{a}_1 e \hat{a}_2 .

8.10 Variâncias e covariâncias dos resíduos

A partir da relação entre os resíduos e os erros, obtida na seção 8.8,

$$\mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I}_N - \mathbf{H}] \vec{\epsilon} \quad , \quad (8.42)$$

é possível determinar a matriz das covariâncias dos resíduos calculando, diretamente, a matriz quadrada de ordem N

$$\mathbf{V}_{\text{residuo}} = \langle (\mathbf{y} - \hat{\mathbf{y}}) (\mathbf{y} - \hat{\mathbf{y}})^t \rangle = \langle [\mathbf{I}_N - \mathbf{H}] \vec{\epsilon} ([\mathbf{I}_N - \mathbf{H}] \vec{\epsilon})^t \rangle =$$

$$[\mathbf{I}_N - \mathbf{H}] \langle \vec{\epsilon} \vec{\epsilon}^t \rangle [\mathbf{I}_N - \mathbf{H}]^t = [\mathbf{I}_N - \mathbf{H}] \mathbf{V}_y [\mathbf{I}_N - \mathbf{H}]^t$$

em que se identificou $\mathbf{V}_y = \langle \vec{\epsilon} \vec{\epsilon}^t \rangle$, que é uma matriz diagonal, com todos os elementos diagonais iguais, $\mathbf{V}_y = \sigma^2 \mathbf{I}_N$, de acordo com a hipótese da seção 8.8.

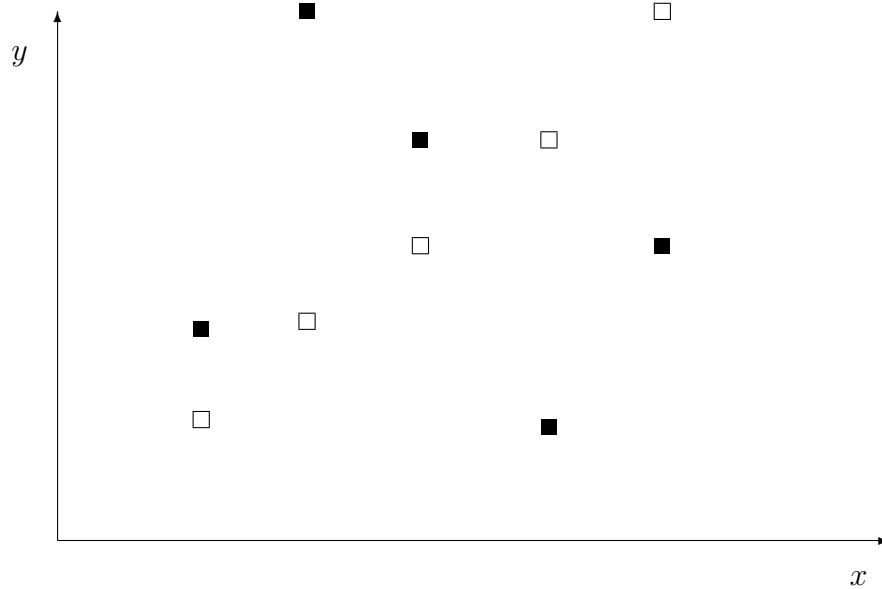


Figura 8.2: Dois conjuntos de pontos experimentais com dispersões iguais, um representado pelos quadrados cheios e outro, pelos quadrados vazios. Ambos têm mesmas dispersões em x e y , mas coeficientes de correlação r , entre x e y , diferentes. O conjunto de quadrados vazios representa uma situação onde $r \cong 1$, ou seja, a variação de x *explica* a variação de y . Já o conjunto de quadrados cheios tem $r \cong 0$, ou seja y parece comportar-se de maneira *independente* de x .

Assim, pode-se inverter a ordem do último produto de matrizes da equação acima e obter o resultado final,

$$\mathbf{V}_{\text{residuo}} = [\mathbf{I}_N - \mathbf{H}] \mathbf{V}_y \quad , \quad (8.55)$$

onde se usou que $[\mathbf{I}_N - \mathbf{H}]$ é idempotente, conforme questão Q8.4.

Embora deduzido no caso em que a matriz de variâncias é diagonal, com todos os elementos diagonais iguais, este resultado vale qualquer que seja a matriz de variâncias, basta usar

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \quad .$$

8.11 Generalização do modelo para dados covariantes

Muitos dos resultados deste capítulo foram obtidos com a suposição de covariâncias nulas e, na seção 8.6, também com a hipótese de igualdade das variâncias de todos os dados. Essas suposições ajudaram a desenvolver a álgebra, mas elas não são necessárias, e tomamos o cuidado de obter resultados de validade geral. Assim, se as covariâncias entre os dados não são nulas, mas se conhece a matriz de covariância exatamente, tudo que foi deduzido nas seções anteriores vale, ou seja, no modelo linear, o método dos mínimos quadrados fornece a estimativa linear não-tendenciosa de variância mínima, independentemente da forma da f.d.p. dos dados. As estatísticas que fornecem os parâmetros ajustados e sua matriz de covariância, dadas pelas expressões (8.8) e (8.18), já estão adequadas para incluir essa situação. Também a extensão para o cálculo da matriz de covariância entre funções lineares dos parâmetros, expressão (8.34), e a matriz de covariância dos resíduos, (8.55), já estão adequadas ao caso geral.

Esse fato, dessas demonstrações efetuadas com a matriz de variância diagonal valerem quando ela não é diagonal, decorre da linearidade do estimador de mínimos quadrados. Quando a matriz de covariância não é nula, pode-se fazer uma transformação linear dos dados, $\mathbf{y}' = \mathbf{T}\mathbf{y}$, de modo a diagonalizar a matriz de covariâncias de \mathbf{y}' e, por uma mudança de escala adequada para cada y'_i , igualar todas as variâncias. Todos os resultados anteriores valem para esse conjunto transformado e, agora, os resultados podem ser devolvidos ao sistema original, efetuando transformações inversas. Aqui, vamos apenas enunciar, sem demonstrar, que toda essa manipulação algébrica corresponde apenas a substituir a matriz de covariância completa nas expressões obtidas.

Uma extensão da discussão da seção 8.8, sobre a estimativa da variância dos dados quando ela é desconhecida, permite trabalhar com a situação em que a matriz de covariâncias e variâncias é conhecida a menos de um fator multiplicativo. A estimativa desse fator pelo método descrito na seção 8.8 permite estimar as variâncias dos dados sem tendenciosidade.

Em todos os casos, porém, quando \mathbf{V} corresponde a uma estimativa, de fato não vale a propriedade de variância mínima, que caracteriza o método dos mínimos quadrados no modelo linear. Qualitativamente, porém, pode-se avaliar que os resultados serão tão bons quanto for boa a estimativa da matriz \mathbf{V} de covariâncias.

8.12 Interpretação dos parâmetros ajustados

8.12.1 f.d.p. dos dados normal

Se a f.d.p. dos dados for gaussiana, a f.d.p. dos parâmetros será gaussiana, uma vez que a estimativa de mínimos quadrados é uma combinação *linear* dos dados. Nesse caso, a f.d.p. de $\hat{\mathbf{a}}$ é a *multinormal*, centrada em \mathbf{a}_0 , porque o estimador não é tendencioso, e com matriz de covariâncias dada pela fórmula (8.18). A expressão analítica da multinormal foi discutida anteriormente nos capítulos 2 e 4, onde também se chamou a atenção para o fato das f.d.p.s marginais também serem gaussianas. Assim, pode-se construir intervalos de confiança para cada um dos parâmetros independentemente dos outros, do tipo⁵

$$(a_i - \sigma_{a_i} \leq a_{i,0} \leq a_i + \sigma_{a_i}) \text{ e } \forall a_{j \neq i} \quad ,$$

com nível de confiança de 68%. Note que esse intervalo explicita um e apenas um dos parâmetros, sem especificar valores ou intervalos para os demais. Em outras palavras, à região definida por

$$(a_i - \sigma_{a_i} \leq a_{i,0} \leq a_i + \sigma_{a_i}), (a_k - \sigma_{a_k} \leq a_{k,0} \leq a_k + \sigma_{a_k}) \text{ e } \forall a_j, j \neq i \text{ e } j \neq k,$$

o nível de confiança não é 68%, mas bem menor. Intervalos de confiança multidimensionais devem levar em conta as correlações entre os parâmetros, veja [Eadie, seção 9.1.2].

8.12.2 f.d.p. dos dados não normal

Se os dados não são gaussianos, há duas possibilidades de obter respostas completas para as estimativas dos parâmetros:

i) Se a f.d.p. dos dados é conhecida, é possível em certas situações determinar a f.d.p. dos parâmetros, o que pode não ser tão difícil uma vez que as estimativas dos parâmetros são lineares nos dados.

ii) Se a f.d.p. dos dados é desconhecida ou se ela é conhecida, mas não é possível ou prático determinar a f.d.p. das estimativas, o Teorema Central do Limite (seção 6.16) estabelece que, quando o número de dados for suficientemente grande, a f.d.p. das estimativas, $\hat{\mathbf{a}}$, será gaussiana. A questão do quanto grande o número de dados precisa ser não tem resposta geral, mas

⁵O conjunto de símbolos $\forall a_{j \neq i}$ pode ser lido como *quaisquer* que sejam os valores de $a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_\mu$.

normalmente o número de graus de liberdade do ajuste deve medir-se em *dezenas*. Não costuma haver necessidade de centenas de dados, mas também não convém lidar com apenas 2 ou 3 graus de liberdade nessa situação. Se de todo for impossível observar um número razoável de pontos, é conveniente ao menos simular o experimento, introduzindo a forma conhecida da f.d.p., para avaliar a forma da f.d.p. das estimativas e a adequação da matriz de covariância dos parâmetros ajustados.

8.12.3 f.d.p. dos dados não normal e desconhecida e poucos dados

Este é, talvez, o pior caso encontrado com alguma frequência em ciências experimentais. Não é possível avaliar o comportamento estatístico das estimativas através de simulação, desde que para isso é necessário conhecer a f.d.p. dos dados. No entanto, a estimativa é não tendenciosa, tem variância mínima e Q tem valor esperado igual a $N - \mu$, o que frequentemente basta. Discutiremos um pouco mais acerca disto na seção seguinte.

8.13 Teste de χ^2

Se a f.d.p. dos dados for normal, pode-se conduzir um teste de hipótese baseado na f.d.p. de χ^2 para avaliar a qualidade do ajuste. A grandeza

$$S = (\mathbf{y} - \hat{\mathbf{y}})^t \mathbf{V}^{-1}(\mathbf{y} - \hat{\mathbf{y}}) \quad , \quad (8.56)$$

onde $\hat{\mathbf{y}}$ são os valores calculados com os parâmetros ajustados, $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{a}}$, é distribuída como χ^2 com um número de graus de liberdade igual a $N - \mu$. Veja que o valor esperado de S é *exatamente* $N - \mu$, confira na fórmula (8.45). No capítulo 5, discutimos a questão do *teste de hipótese* e os critérios de escolha dos valores críticos de χ^2 e não nos estenderemos aqui sobre estes assuntos.

Mesmo que os dados não sejam gaussianos, o valor esperado de S é $N - \mu$, uma vez que a relação (8.45) independe da distribuição estatística dos dados. A f.d.p. de S , porém, não será a de χ^2 , embora o Teorema Central do Limite garanta que S tenda, assintoticamente, à gaussiana (não à f.d.p. de χ^2 !), cuja variância possivelmente *não* será $2N$. De qualquer maneira, o valor médio de S vale $N - \mu$ e podemos ter uma confiança razoável no ajuste quando S for aproximadamente igual a esse valor.

Um procedimento comum quando o valor de S é muito maior que $N - \mu$ é não rejeitar a função ajustada e, portanto, adotar os parâmetros ajustados, $\hat{\mathbf{a}}$, mas multiplicar os desvios-padrões nos \hat{a}_k por $\sqrt{S/(N - \mu)}$. A justificativa desse procedimento reside em supor que a matriz de covariância dos dados usada no ajuste, \mathbf{V} , esteja errada apenas por um fator multiplicativo. Assim, utiliza-se uma generalização do resultado da seção 8.8 para estimar σ^2 . Como multiplicar \mathbf{V} por um fator não modifica $\hat{\mathbf{a}}$, a única consequência desse procedimento reside em alterar o cálculo da matriz de covariância dos parâmetros, que passa a ser

$$\text{Cov}(\hat{\mathbf{a}}) = \frac{S}{N - \mu} (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1} . \quad (8.57)$$

Este resultado deve ser aplicado com discrição e, de preferência, apenas quando se tem absoluta certeza de estar usando o modelo correto e for viável a hipótese das incertezas *relativas* dos dados estarem corretas. Note que se perde a possibilidade de usar o teste de χ^2 quando (8.57) é utilizada. Também não compare este procedimento com a situação em que *se sabe* que as variâncias são iguais entre si, com um valor desconhecido, caso em que o procedimento da seção 8.8 e o uso da fórmula (8.46) constituem, muito provavelmente, a melhor maneira de resolver o problema.

8.14 A inversão de $\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}$ na prática

Como os elementos das diferentes colunas da matriz de planejamento \mathbf{X} podem exprimir grandezas de naturezas totalmente diferentes, é comum se obter uma matriz

$$\mathbf{M} = \mathbf{X}^t\mathbf{V}^{-1}\mathbf{X}$$

com elementos de ordens de grandeza numéricas muito diversas, sendo frequente encontrar diferenças de 10 ordens de grandeza entre o maior e o menor elemento. A maneira prática de invertê-la consiste em construir uma matriz \mathbf{M}' cujos elementos sejam adimensionais por meio da transformação bilinear

$$\mathbf{M}' = \mathbf{TMT} \quad , \quad (8.58)$$

onde \mathbf{T} é a matriz diagonal definida por

$$T_{ij} = \frac{1}{\sqrt{M_{ii}}}\delta_{ij} \quad . \quad (8.59)$$

O resultado é que \mathbf{M}' tem todos os elementos diagonais idênticos a 1 e os não-diagonais são adimensionais. Também pode-se verificar que todos os elementos não-diagonais de \mathbf{M}' tem módulo menor que 1 e, na prática, exceto os elementos que são nulos em razão de alguma particularidade do modelo, os demais não são muito menores que 1 em módulo. A matriz \mathbf{M}' transformada dessa maneira pode, quase sempre, ser invertida com grande precisão numérica por qualquer dos algoritmos usuais, mesmo para matrizes de ordem elevada.

A matriz \mathbf{M}^{-1} é obtida refazendo a transformação (8.58),

$$\mathbf{M}^{-1} = \mathbf{T}(\mathbf{M}')^{-1}\mathbf{T} \quad . \quad (8.60)$$

Q8.5 *Demonstre a fórmula (8.60) a partir da (8.58).*

Embora a transformação (8.58) seja desnecessária na maior parte das situações quando se usam muitos dígitos nos cálculos, ocasionalmente sua aplicação resolve problemas numéricos, especialmente quando os parâmetros são expressos por valores de ordens de grandeza muito distintas e/ou o número de parâmetros envolvido é muito grande.

8.15 Exemplo: dados correlacionados

Nas seções 8.2, 8.5 e 8.9, analisamos três medidas com dados independentes. Aqui, retomaremos apenas o exemplo **C** com a inclusão de covariâncias entre os dados; os exemplos **A** e **B** podem ser tratados de maneira análoga. Na próxima seção, discutiremos outros exemplos de análise de dados correlacionados, explorando as possibilidades do método dos mínimos quadrados tal como o desenvolvemos neste capítulo.

Vamos supor que as variâncias de todos os dados sejam iguais a σ^2 e que as correlações entre todos os pares de dados sejam iguais a ρ . A matriz de covariância fica

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & & \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & & \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & & \\ & & & \ddots & \\ & & & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & & \\ \rho & 1 & \rho & & \\ \rho & \rho & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix} \quad . \quad (8.61)$$

A matriz inversa, assim como \mathbf{V} , tem todos os elementos da diagonal iguais, com os demais elementos iguais entre si e diferentes dos diagonais, de maneira que esses dois valores podem ser obtidos resolvendo um sistema de duas equações a duas incógnitas. Os elementos da matriz inversa são

$$\begin{aligned} (\mathbf{V}^{-1})_{ii} &= \frac{1}{\sigma^2} \left\{ 1 - \frac{(N-1)\rho^2}{\rho^2(N-1) - \rho(N-2) - 1} \right\} \quad \text{e} \\ (\mathbf{V}^{-1})_{ij} &= \frac{1}{\sigma^2} \left\{ \frac{\rho}{\rho^2(N-1) - \rho(N-2) - 1} \right\} \quad \text{para } i \neq j \end{aligned} \quad (8.62)$$

onde o polinômio do denominador, $P(\rho)$, pode ser fatorado,

$$P(\rho) = \rho^2(N-1) - \rho(N-2) - 1 = [(N-1)\rho + 1] \cdot [\rho - 1] \quad . \quad (8.63)$$

C) A matriz de planejamento, \mathbf{X} , independe da matriz de covariância ser diagonal ou não, por isso é a mesma da seção 8.2. Obtém-se, agora, depois de muita álgebra

$$(\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} = \frac{\sigma^2}{N^2 \delta_x^2} \begin{pmatrix} [\rho(N-1) + 1] \cdot \sum x_i^2 - \rho X^2 & X(\rho - 1) \\ X(\rho - 1) & N(1 - \rho) \end{pmatrix} \quad (8.64)$$

na mesma notação da seção 8.2. Definindo

$$S_{xy} = \sum x_i y_i \quad ,$$

calcula-se

$$\mathbf{X}^t \mathbf{V}^{-1} \mathbf{y} = \frac{1}{\sigma^2 P(\rho)} \begin{pmatrix} Y(\rho - 1) \\ XY\rho - S_{xy}[\rho(N-1) + 1] \end{pmatrix} \quad .$$

As estimativas dos parâmetros são dadas por $\hat{\mathbf{a}} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}$, fórmula (8.8), o que resulta em

$$\begin{aligned} \hat{a}_1 &= \frac{1}{N\delta_x^2} \left(y_C \sum x_i^2 - x_C S_{xy} \right) \quad \text{e} \\ \hat{a}_2 &= \frac{1}{\delta_x^2} \frac{1}{N} \sum_{i=1}^N (x_i - x_C) \cdot (y_i - y_C) \quad , \end{aligned} \quad (8.14)$$

em que o número (8.14) atribuído à fórmula acima enfatiza a identidade deste resultado com aquele obtido no caso de dados independentes, veja seção 8.2.

Esta coincidência decorre das correlações serem todas idênticas, mas, em geral, ignorar as correlações modifica os parâmetros ajustados, embora o mais comum seja conduzir a diferenças pequenas.

As variâncias e covariâncias dos parâmetros ajustados são dadas pela matriz da fórmula (8.18). Neste caso, ela está apresentada na fórmula (8.64), da qual isolamos os resultados

$$\text{var}(\hat{a}_1) = \sigma_{a_1}^2 = \frac{\sigma^2 [\rho(N-1) + 1] \cdot \sum x_i^2 - \rho X^2}{N \delta_x^2} \quad \text{de onde}$$

$$\sigma_{a_1}^2 = \rho\sigma^2 + \frac{\sigma^2 \sum x_i^2}{N \delta_x^2} (1 - \rho) \quad ,$$

$$\text{var}(\hat{a}_2) = \sigma_{a_2}^2 = \frac{\sigma^2}{N \delta_x^2} (1 - \rho) \quad \text{e}$$

$$\text{cov}(\hat{a}_1, \hat{a}_2) = -\frac{\sigma^2 x_C}{N \delta_x^2} (1 - \rho) \quad .$$

A variância de \hat{a}_2 e o módulo da covariância entre \hat{a}_1 e \hat{a}_2 reduzem-se muito se a correlação for próxima de 1.⁶ Já quando $\rho \approx 0$, pode-se praticamente ignorar a correlação entre os dados, do ponto de vista das incertezas nos parâmetros ajustados. A variância de \hat{a}_1 , porém, muda até mesmo qualitativamente, porque o primeiro termo da expressão, $\rho\sigma^2$, não depende do número de dados, representando um limite inferior que não existe quando os dados não são correlacionados.

As mudanças introduzidas na matriz de variância dos parâmetros pelas correlações dos dados refletem-se nas variâncias dos pontos interpolados pela função ajustada, que podem ser muito diferentes do caso em que os dados são independentes. Após mais álgebra, obtém-se

$$\text{var}(\hat{y}(x)) = \sigma_{y(x)}^2 = \rho\sigma^2 + \frac{\sigma^2}{N} (1 - \rho) \frac{\sum (x_i - x)^2}{N \delta_x^2} \quad , \quad (8.65)$$

⁶À primeira vista, pode-se estranhar o fato de $\rho = 1$ ser tão especial, mas não $\rho = -1$. Afinal, em ambos os casos o comportamento dos dados é determinístico. A diferença vem do fato que é impossível ter todos os dados com $\rho = -1$. Por exemplo, em uma medida com 3 dados, se a correlação entre todos pudesse ser -1 , se o primeiro estivesse superestimado, o segundo e o terceiro estariam subestimados — mas se o segundo estiver subestimado, então o terceiro deveria estar superestimado, uma contradição. Em uma matriz de covariância como a (8.61), o valor de ρ precisa pertencer ao intervalo $\left] -\frac{1}{N-1}, 1 \right[$.

com a propriedade

$$\text{var}(\hat{y}(x)) \xrightarrow{N \rightarrow \infty} \rho\sigma^2 \quad .$$

A expressão acima mostra que o desvio padrão mínimo de um ponto interpolado por meio de uma reta ajustada a um número muito grande de pontos experimentais é $\rho\sigma^2$, a parcela comum das variâncias dos dados. Caso houvéssemos, equivocadamente, efetuado os cálculos ignorando a correlação, o valor mínimo do desvio padrão do ponto interpolado tenderia a zero quando o número de dados tendesse a infinito.⁷ Esta talvez seja a mudança mais marcante que decorre de considerar as covariâncias no ajuste de parâmetros e que, após alguma reflexão, parecerá bastante natural.

Finalizamos com o cálculo de χ^2 , que envolve mais álgebra por conta da matriz \mathbf{V}^{-1} . Obtemos, na notação da seção 8.9,

$$\chi^2 = \frac{N}{\sigma^2(1-\rho)} \delta_y^2 (1-r^2) \quad (8.66)$$

Assim, ignorar a correlação implica, na maioria das vezes, em subestimar χ^2 , uma vez que na maior parte dos experimentos $\frac{1}{1-\rho} > 1$ porque $\rho > 0$ ocorre muito mais frequentemente que o oposto.

8.16 Média de dois dados correlacionados

Este exemplo pode ser associado ao exemplo **A** discutido nas seções 8.2, 8.5 e 8.9, restrito ao caso em que há apenas *dois* dados experimentais, d_1 e d_2 , de medições de uma mesma grandeza física, cujo valor deve ser estimado. A matriz de covariâncias desses dois dados é conhecida e vale

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

de maneira que ρ representa o coeficiente de correlação entre os dados d_1 e d_2 . Busca-se \hat{d} , a estimativa da grandeza física, e sua variância, $\text{var}(\hat{d})$.

O modelo estatístico na forma da equação fica (8.4),

$$\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} d_0 + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \quad ,$$

⁷Note que, quando $N \rightarrow \infty$, $\rho \geq 0$ (veja nota de rodapé 6), compatível com o fato que $\text{var}(\hat{y}(x)) \geq 0$, desde que uma variância nunca pode ser negativa.

em que d_0 é o valor (verdadeiro) da grandeza e o último vetor tem como componentes os erros das medidas. Vê-se que, neste caso, a *matriz de planejamento* apropriada é

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \quad (8.67)$$

A estimativa é dada pela equação (8.8), $\hat{d} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1} \mathbf{y}$, em que o vetor das variáveis dependentes, \mathbf{y} , é

$$\mathbf{y} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} .$$

Efetuando os cálculos, obtém-se

$$\hat{d} = \frac{d_1(\sigma_2^2 - \rho\sigma_1\sigma_2) + d_2(\sigma_1^2 - \rho\sigma_1\sigma_2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} . \quad (8.68)$$

Um exame desta expressão permite verificar que \hat{d} é uma média ponderada dos dois dados experimentais, com pesos que envolvem a covariância entre eles, além das variâncias. A variância desta estimativa pode ser calculada pela expressão (8.18), $\text{var}(\hat{d}) = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1}$, obtendo-se o resultado

$$\text{var}(\hat{d}) = \sigma_d^2 = \frac{\sigma_1^2 \sigma_2^2 (1 - \rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} . \quad (8.69)$$

Os resultados (8.68) e (8.69) merecem alguns comentários.

i) Caso aconteça que $\sigma_2 < \rho\sigma_1$ (ou $\sigma_1 < \rho\sigma_2$), d_1 (ou d_2) terá um peso *negativo* na média calculada pela fórmula (8.68), o que resulta em \hat{d} fora do intervalo definido pelos dois dados, d_1 e d_2 . Apesar de surpreendente, este resultado é apenas uma consequência do uso **correto** das covariâncias. Tal situação, entretanto, é muito rara, uma vez que corresponde a uma covariância superior à variância de um dos dados.

ii) Caso $\sigma_2 = \rho\sigma_1$ (ou $\sigma_1 = \rho\sigma_2$), d_1 (ou d_2) terá peso nulo, de modo que o dado experimental correspondente não contribui em nada para o valor adotado, \hat{d} . Como no caso *i*, este resultado é apenas consequência do uso **correto** das covariâncias, embora também muito raro.

iii) Caso $\sigma_1 = \sigma_2 = \sigma$, \hat{d} será simplesmente a média entre d_1 e d_2 . A variância de \hat{d} ainda depende, porém, da correlação dos dados,

$$\sigma_d^2 = \sigma^2 \frac{(1 + \rho)}{2} \quad (\text{quando } \sigma_1 = \sigma_2 = \sigma) . \quad (8.70)$$

8.17. A MEDIDA DE UMA GRANDEZA PODE ALTERAR AS ESTIMATIVAS DE OUTRAS GR

Se $\rho = 1$, esta equação se reduz a $\sigma_d^2 = \sigma^2$, o que corresponde a d_1 e d_2 serem o mesmo dado.

Q8.6 Demonstre, no caso particular $\sigma_1 = \sigma_2 = \sigma$ discutido acima, que $\hat{d} = (d_1 + d_2)/2$ e obtenha a fórmula (8.70).

iv) Se $\sigma_1 = \sigma_2 = \sigma$, como no item *ii)* acima, com $\rho = -1$, obtém-se $\sigma_d^2 = 0$ e, portanto, d_0 será conhecido exatamente, $\hat{d} = d_0$. Isto deve-se ao fato de, nesta situação, se d_1 é uma superestimação (subestimação) de d_0 , então d_2 é uma subestimação (superestimação) exatamente do mesmo valor.

Q8.7 Procure demonstrar analiticamente o resultado interpretado no item *(iv)* acima, utilizando as relações $d_1 = d_0 + \epsilon_1$ e $d_2 = d_0 + \epsilon_2$ e as definições de variância e covariância.

Concluimos este exemplo com o cálculo da soma dos quadrados dos resíduos ponderados da equação (8.56),

$$S = \begin{pmatrix} d_1 - \hat{d} \\ d_2 - \hat{d} \end{pmatrix}^t \mathbf{V}^{-1} \begin{pmatrix} d_1 - \hat{d} \\ d_2 - \hat{d} \end{pmatrix} = \frac{(d_1 - d_2)^2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad . \quad (8.71)$$

Se d_1 e d_2 têm f.d.p. gaussiana, então S tem a mesma f.d.p. que χ^2 para 1 grau de liberdade, $F_1(S)$ da fórmula (2.45'). Serve, então, de base a um teste de hipótese, a partir do qual é possível, por exemplo, avaliar a compatibilidade entre os dois dados.

8.17 A medida de uma grandeza pode alterar as estimativas de outras grandezas com as quais é covariante

Considere que x e y são os resultados de medidas de grandezas diferentes, cujos valores verdadeiros são x_0 e y_0 , com matriz de variâncias

$$\mathbf{V}_{xy} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \quad .$$

Suponha, agora, que se meça x_0 de novo, com o resultado x' com variância $\sigma_x'^2$, não correlacionado com a estimativa x nem com y . Neste caso, o modelo

linear toma a forma

$$\begin{pmatrix} x \\ y \\ x' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{pmatrix},$$

a partir da qual identificamos o vetor \mathbf{y} e a matriz \mathbf{X} da equação (8.4). A estimativa do par x_0 e y_0 que leve em conta todos os dados dependerá também da matriz de covariância dos 3 dados,

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{xy} & \vdots & \mathbf{0} \\ \dots & \vdots & \dots \\ \mathbf{0} & \vdots & \sigma_x'^2 \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y & 0 \\ \rho\sigma_x\sigma_y & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_x'^2 \end{pmatrix}.$$

A equação (8.8) fornece, então,

$$\hat{x} = \frac{x/\sigma_x^2 + x'/\sigma_x'^2}{1/\sigma_x^2 + 1/\sigma_x'^2} \quad \text{e} \quad (8.72)$$

$$\hat{y} = y + \frac{(x' - x)\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_x'^2}, \quad (8.73)$$

com matriz de covariância dada pela equação (8.18),

$$\mathbf{V} = \frac{1}{\sigma_x^2 + \sigma_x'^2} \begin{pmatrix} \sigma_x^2\sigma_x'^2 & \rho\sigma_x\sigma_y\sigma_x'^2 \\ \rho\sigma_x\sigma_y\sigma_x'^2 & \sigma_y^2[\sigma_x^2(1 - \rho^2) + \sigma_x'^2] \end{pmatrix}. \quad (8.74)$$

A estimativa \hat{y} **não** dependerá do novo resultado x' **somente** se a correlação entre os valores x e y **anteriores** à nova medida for nula, $\rho = 0$. Para avaliar quanto é significativa a dependência de \hat{y} com o novo dado x' , examinemos a situação onde $\sigma_x = \sigma_y = \sigma_x'$ e $\rho = 0,5$. Neste caso, as estimativas dadas pelas fórmulas (8.72) e (8.73) são

$$\hat{x} = x + \frac{x' - x}{2} \quad \text{e}$$

$$\hat{y} = y + \frac{x' - x}{4}.$$

Verifica-se, então, que as modificações das estimativas de ambos os parâmetros são da mesma ordem de grandeza.

8.18 Exemplo: vínculos entre parâmetros

Apresentaremos aqui um caso particular de redução de dados onde há um *vínculo linear* entre os parâmetros, com um procedimento que serve em qualquer situação de ajuste de parâmetros em funções lineares.

Antes de entrar nos detalhes, uma palavra deve ser dita a respeito do número de graus de liberdade do ajuste quando se impõe um vínculo. Se o modelo tem μ parâmetros e ν relações de vínculo, o número de parâmetros efetivamente livres é $\mu - \nu$. É esse número de parâmetros livres que deve entrar no cálculo do número de graus de liberdade:

$$\ell = N - \mu + \nu \quad ,$$

onde N é o número de pontos. Veja o exemplo numérico abaixo, no teste de hipótese com base em χ^2 .

Imagine um experimento em que se medem os ângulos internos de um triângulo plano e se obtém os valores θ_1 , θ_2 e θ_3 . Para simplificar a discussão e interpretação dos resultados, considere que estes dados são estatisticamente independentes e têm mesmo desvio-padrão, de modo que a matriz de variâncias é diagonal e vale

$$\mathbf{V} = \sigma^2 \mathbf{I}_3 \quad .$$

Há, entretanto, um vínculo entre as três grandezas observadas: a soma dos ângulos internos é 180° . Pode-se, então, ajustar os valores de qualquer par desses 3 ângulos e determinar o terceiro pela relação de vínculo. Escolhendo os ângulos θ_1 e θ_2 como parâmetros, a relação de vínculo que determina o terceiro ângulo é

$$\theta_{3,0} = 180^\circ - \theta_{1,0} - \theta_{2,0} \quad ,$$

em que os valores verdadeiros dos ângulos estão simbolizados por $\theta_{i,0}$. O modelo estatístico fica

$$\theta_1 = \theta_{1,0} + \epsilon_1$$

$$\theta_2 = \theta_{2,0} + \epsilon_2$$

$$\theta_3 = \theta_{3,0} + \epsilon_3 = 180^\circ - \theta_{1,0} - \theta_{2,0} + \epsilon_3 \quad \text{ou}$$

$$180^\circ - \theta_3 = \theta_{1,0} + \theta_{2,0} - \epsilon_3 \quad .$$

O conjunto dessas equações pode ser reescrito nos moldes da equação (8.4),

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ 180^\circ - \theta_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_{1,0} \\ \theta_{2,0} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ -\epsilon_3 \end{pmatrix} \quad .$$

Pode-se identificar na relação acima a matriz de planejamento,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix},$$

e o vetor de variáveis dependentes,

$$\mathbf{y} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ 180^\circ - \theta_3 \end{pmatrix},$$

lembrando que o vetor de parâmetros a ajustar é

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix}.$$

A equação (8.8) resulta em

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2\theta_1 & + & (180^\circ - \theta_2 - \theta_3) \\ 2\theta_2 & + & (180^\circ - \theta_1 - \theta_3) \end{pmatrix} \quad (8.75)$$

e a equação (8.18) dá a matriz de variâncias,

$$\mathbf{V}_{\theta_1, \theta_2} = \frac{\sigma^2}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \quad (8.76)$$

Os resultados para $\hat{\theta}_3$ (estimativa de ponto, variância e covariâncias com $\hat{\theta}_1$ e $\hat{\theta}_2$) podem ser calculados diretamente a partir de (8.75) e (8.76) com a relação de vínculo e a fórmula da variância de uma função de variáveis aleatórias.

Q8.8 Calcule $\hat{\theta}_3$, $\text{var}(\hat{\theta}_3)$, $\text{cov}(\hat{\theta}_3, \hat{\theta}_1)$ e $\text{cov}(\hat{\theta}_3, \hat{\theta}_2)$. Nesse caso em que as variâncias dos dados são iguais, os resultados são mais simples de interpretar. A equação (8.75) nos mostra que a estimativa $\hat{\theta}_1$ é a média entre o dado θ_1 e o valor deduzido da relação de vínculo, usando os dados θ_2 e θ_3 . O mesmo ocorre em relação a $\hat{\theta}_2$. A equação (8.76) mostra que a imposição do vínculo reduz as variâncias das estimativas $\hat{\theta}_1$ e $\hat{\theta}_2$ a $2/3$ das variâncias de θ_1 e θ_2 , mas tem como contrapartida a perda da independência, uma vez que existe uma covariância não nula entre $\hat{\theta}_1$ e $\hat{\theta}_2$, apesar da independência estatística entre os dados θ_1 e θ_2 . Um pouco mais de interpretação está no exemplo numérico abaixo.

Q8.9 Calcule as estimativas dos três ângulos, impondo a condição adicional do triângulo ser isósceles. Escolha os ângulos θ_2 e θ_3 como os ângulos idênticos.

Exemplo 8.2

EXEMPLO NUMÉRICO

A medição dos três ângulos resultou em $\theta_1 = 40,1(5)$, $\theta_2 = 69,8(5)$ e $\theta_3 = 72,0(5)$, em graus. Todos os dados têm mesma variância, o que permite aplicar as equações (8.75) e (8.76) acima. Substituindo esses dados naquelas equações, obtém-se

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} 39,5 \\ 69,2 \end{pmatrix} \quad \text{e}$$

$$\mathbf{V}_{\theta_1, \theta_2} = \frac{0,5^2}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad ,$$

em graus para as estimativas e em graus ao quadrado para as variâncias e a covariância; estas unidades serão adotadas ao longo deste exemplo, e não serão mais repetidas.

Pode-se deduzir a estimativa do terceiro ângulo a partir da relação de vínculo, com o resultado

$$\hat{\theta}_3 = 180^\circ - \hat{\theta}_1 - \hat{\theta}_2 = 71,4 \quad .$$

Note que todas as estimativas são menores que os dados. Como a soma dos três dados é 181,9, há um excesso de 1,9 que deve ser retirado dos dados. Como eles são independentes e têm mesma variância, o corte em cada dado deve ser o mesmo e, portanto, igual a $1,9/3 \cong 0,6$. Verifique!

O cálculo da soma dos quadrados dos resíduos ponderados pela matriz de variâncias dos dados é dada pela fórmula (8.56),

$$Q_m = (\vec{\theta} - \mathbf{X}\hat{\vec{\theta}})^t \mathbf{V}^{-1} (\vec{\theta} - \mathbf{X}\hat{\vec{\theta}}) \quad ,$$

onde

$$(\vec{\theta} - \mathbf{X}\hat{\vec{\theta}}) = \begin{pmatrix} \theta_1 \\ \theta_2 \\ 180^\circ - \theta_3 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \begin{pmatrix} 0,6 \\ 0,6 \\ 0,6 \end{pmatrix} \quad ,$$

sendo que os cálculos intermediários foram efetuados com um dígito a mais que os exibidos acima. Assim,

$$Q_m = (0,6 \quad 0,6 \quad 0,6) \begin{pmatrix} 0,5^{-2} & 0 & 0 \\ 0 & 0,5^{-2} & 0 \\ 0 & 0 & 0,5^{-2} \end{pmatrix} \begin{pmatrix} 0,6 \\ 0,6 \\ 0,6 \end{pmatrix} = 4,3 \quad ,$$

um resultado que, nesta situação em que o número de graus de liberdade é $\nu = 1$, pode não indicar a rejeição do ajuste. A hipótese só poderá ser testada quantitativamente se a f.d.p. dos dados for conhecida. Caso os dados sejam gaussianos, então Q_m tem a f.d.p. de χ^2 e calcula-se

$$P(\chi_{\ell=1}^2 > 4,3) \cong 4\% \quad ,$$

o que corresponde a um ajuste no limite da rejeição.

Encerramos o exemplo com a determinação da matriz de covariâncias entre $\hat{\theta}_2$ e $\hat{\theta}_3$, o que você efetuou ao resolver a questão Q8.8. Calcular a matriz de variâncias de $\hat{\theta}_2$ e $\hat{\theta}_3$ a partir da matriz de variâncias de $\hat{\theta}_1$ e $\hat{\theta}_2$ requer a relação entre estes dois pares de variáveis aleatórias, que é dada por

$$\begin{aligned} \hat{\theta}_2 &= \hat{\theta}_2 \quad \text{e} \\ \hat{\theta}_3 &= 180^\circ - \hat{\theta}_1 - \hat{\theta}_2 \quad . \end{aligned}$$

A regra de transformação pode ser escrita como

$$\begin{pmatrix} \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 180^\circ \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \quad .$$

A dependência de $\hat{\theta}_2$ e $\hat{\theta}_3$ nas variáveis aleatórias está na matriz \mathbf{F} ,

$$\mathbf{F} = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \quad ,$$

que, na fórmula (3.9), dá

$$\begin{aligned} \mathbf{V}_{\theta_2\theta_3} &= \mathbf{F}\mathbf{V}_{\theta_1\theta_2}\mathbf{F}^t \\ \mathbf{V}_{\theta_2\theta_3} &= \frac{\sigma^2}{3} \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} \quad , \end{aligned}$$

onde foi substituída a matriz de variância da fórmula (8.76). O resultado é

$$\mathbf{V}_{\theta_2\theta_3} = \frac{\sigma^2}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad .$$

Neste caso particular, a matriz acima é idêntica à da fórmula (8.76) porque o problema é completamente simétrico pela permutação dos índices 1, 2 e 3 — todas as variâncias e covariâncias dos dados são iguais e os ângulos entram na equação de vínculo da mesma maneira.

Exemplo 8.3

A MATRIZ DE COVARIÂNCIA DE $\hat{\theta}_1$, $\hat{\theta}_2$ e $\hat{\theta}_3$

Analogamente ao desenvolvimento dos cálculos usado no exemplo acima, pode-se escrever

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 180^\circ \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} .$$

Usando novamente a eq. (3.9), tem-se

$$\begin{aligned} V_{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix} \frac{\sigma^2}{3} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \frac{\sigma^2}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} . \end{aligned}$$

Esta última matriz é singular, o que se deve à relação de vínculo entre as três variáveis θ_1 , θ_2 e θ_3 : especificadas duas delas, a terceira passa a ser determinada. Consequentemente, a f.d.p. multidimensional de θ_1 , θ_2 e θ_3 é nula a menos que a relação $\theta_1 + \theta_2 + \theta_3 = 180^\circ$ seja verdadeira e, portanto, a f.d.p. conjunta está concentrada no subespaço em que o vínculo imposto é satisfeito.

Exemplo 8.4

VERIFICANDO SE O TRIÂNGULO PODE SER ISÓSCELES

O resultado do teste dessa condição depende da f.d.p. dos dados. Aqui, vamos supor os dados gaussianos.

Há duas maneiras de testar essa hipótese. Uma, corresponde a impor o vínculo $\theta_{2,0} = \theta_{3,0}$ e realizar um teste de χ^2 . Este caminho, iniciado na questão Q8.9, fica para o exercício 8.8. Outro, corresponde a testar a hipótese diretamente sobre os resultados para $\hat{\theta}_2$ e $\hat{\theta}_3$ obtidos acima, uma vez que a f.d.p. das estimativas é conhecida — gaussiana porque os dados são gaussianos.

Na abordagem adotada aqui, formula-se a hipótese

$$H : \theta_{2,0} = \theta_{3,0} \quad ,$$

que pode ser reescrita como

$$H : \theta_{2,0} - \theta_{3,0} = 0 \quad .$$

Chamando a diferença dos ângulos de d ,

$$d = \theta_{2,0} - \theta_{3,0} \quad ,$$

a estimativa de d será

$$\hat{d} = \hat{\theta}_2 - \hat{\theta}_3$$

com variância

$$\text{var}(\hat{d}) = \sigma_d^2 = \text{var}(\hat{\theta}_2) + \text{var}(\hat{\theta}_3) - 2 \text{cov}(\hat{\theta}_2, \hat{\theta}_3) \quad .$$

Substituindo os valores apropriados, tem-se

$$\hat{d} = 2,2 \quad \text{e} \quad \sigma_d = 0,71 \quad .$$

O maior nível de significância com que não rejeitaríamos a hipótese, $\alpha_{\text{máximo}}$, é (veja capítulo 5)

$$\alpha_{\text{máximo}} = 2 \int_{\frac{2,2}{0,7}}^{\infty} N(x; \text{média } 0 \text{ e desvio padrão } 1) dx = 0,2\% \quad ,$$

onde $N(x)$ está representando a f.d.p. gaussiana e $x = (\hat{d} - d_0)/\sigma_d$ é a variável normalizada. Esse resultado leva a rejeitar a hipótese H de maneira bastante segura, porque qualquer nível de significância superior a 0,2% que seja adotado (uma probabilidade diminuta nesta situação em que se efetua um único teste) conduz à sua rejeição.

Note que a estimativa da diferença entre os ângulos está a mais de 3 desvios padrão do valor da hipótese, o que sugere a rejeição mesmo que a f.d.p. dos dados não seja gaussiana. Lembre que as variâncias calculadas pelo método dos mínimos quadrados quando a função é linear nos parâmetros são corretas mesmo que a f.d.p. dos dados não seja gaussiana e a probabilidade do dado estar afastado três desvios padrão do valor médio é pequena para as f.d.p.s mais comuns.

Outras situações em que a imposição dos vínculos é interessante podem ser encontradas no livro *Método dos Mínimos Quadrados com formalismo matricial* [Helene 2013].

8.19 Vínculo a priori ou a posteriori

Suponha que se deseja ajustar os parâmetros da função $y = a + ax$ aos dados $\{(x_i, y_i), i = 1, \dots, N\}$, onde \mathbf{V}_y é a matriz de covariância entre os y_i . Neste caso, dentro do esquema da equação (8.4), tem-se

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 + x_1 \\ 1 + x_2 \\ \vdots \\ 1 + x_N \end{pmatrix} (a_0) + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix} .$$

Identificando na expressão acima a matriz de planejamento,

$$\mathbf{X}_a = \begin{pmatrix} 1 + x_1 \\ 1 + x_2 \\ \vdots \\ 1 + x_N \end{pmatrix} , \quad (8.77)$$

calcula-se \hat{a} ,

$$\hat{a} = (\mathbf{X}_a^t \mathbf{V}_y^{-1} \mathbf{X}_a)^{-1} \mathbf{X}_a^t \mathbf{V}_y^{-1} \mathbf{y} . \quad (8.78)$$

Esse resultado é a *estimativa com imposição do vínculo a priori*, que será comparado com uma outra maneira de estimar essa grandeza.

Este outro procedimento terá dois passos. Inicialmente, ajusta-se α_1 e α_2 de uma função $y = \alpha_1 + \alpha_2 x$ e, posteriormente, ajusta-se um único valor, \hat{a} , a $\hat{\alpha}_1$ e $\hat{\alpha}_2$. Este resultado é a *estimativa com imposição do vínculo a posteriori*.

Para o ajuste dos parâmetros da reta $y = \alpha_1 + \alpha_2 x$, usa-se a matriz

$$\mathbf{X}_\alpha = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}, \quad (8.79)$$

e calculam-se as estimativas e sua matriz de covariância,

$$\hat{\alpha} = (\mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{y} \quad \text{e} \quad (8.80)$$

$$\mathbf{V}_\alpha = (\mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha)^{-1}. \quad (8.81)$$

O segundo passo corresponde a um cálculo como o do exemplo da seção 8.16. Reescrevendo a equação (8.4) com os símbolos adequados a este problema,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (\hat{a}) + \begin{pmatrix} \epsilon'_1 \\ \epsilon'_2 \end{pmatrix}.$$

A matriz de planejamento é a mesma do exemplo 8.16,

$$\mathbf{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (8.67)$$

Ao invés de efetuar explicitamente o cálculo de \hat{a} , o resultado será escrito em função das matrizes \mathbf{X} e \mathbf{V}_α e do vetor $\hat{\alpha}$ para facilitar a demonstração pretendida. Assim,

$$\hat{a}' = (\mathbf{X}^t \mathbf{V}_\alpha^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}_\alpha^{-1} \hat{\alpha}. \quad (8.82)$$

Substituindo nesta última expressão os resultados (8.80) e (8.81), da etapa anterior, obtém-se

$$\hat{a}' = (\mathbf{X}^t \mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha \mathbf{X})^{-1} \mathbf{X}^t \mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha (\mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{y}.$$

Observando que $\mathbf{X}_\alpha \mathbf{X} = \mathbf{X}_\alpha$, tem-se

$$\hat{a}' = (\mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{X}_\alpha)^{-1} \mathbf{X}_\alpha^t \mathbf{V}_y^{-1} \mathbf{y},$$

exatamente a expressão de \hat{a} obtida pelo ajuste direto com a função $y = a + ax$, fórmula (8.78).

Esse resultado – as estimativas de mínimos quadrados obtidas pela imposição do vínculo *a priori* e *a posteriori* serem idênticas – é geral no ajuste de funções lineares nos parâmetros com vínculos também lineares, não limitando-se a esse exemplo particular. Essa característica decorre das propriedades ótimas do método dos mínimos quadrados no ajuste de funções lineares. Assim, não importa o número de etapas na obtenção de uma estimativa, desde que toda a informação estatística acerca dos dados seja passada de uma etapa a outra. Também a f.d.p. dos dados não tem qualquer importância nessa discussão.

EXERCÍCIOS

- 8.1. Mostre que o valor a ser adotado para uma determinada grandeza quando se tem N dados $(x_1, \sigma_1), (x_2, \sigma_2), \dots, (x_N, \sigma_N)$ não covariantes é

$$\hat{x} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad \text{com} \quad \text{var}(\hat{x}) = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \quad ,$$

independentemente da f.d.p. dos x_i (sequer as f.d.p.s dos diferentes x_i precisam ser iguais para que esta estimativa seja a estimativa linear ótima).

- 8.2. Considere dois dados experimentais com seus respectivos desvios padrões, $x_1(\sigma_1)$ e $x_2(\sigma_2)$, não correlacionados e não tendenciosos, $\langle x_i \rangle = x_0$. Mostre que:

- (a) O valor adotado $\tilde{x} = \alpha \cdot x_1 + (1 - \alpha) \cdot x_2$ não é tendencioso, qualquer que seja o valor de α .
- (b) O valor de α que minimiza o desvio padrão de \tilde{x} coincide com o resultado obtido no exercício 8.1.

- 8.3. Determine, usando o método dos mínimos quadrados, o valor a ser adotado para uma grandeza x_0 quando os dados experimentais são x_1, x_2 e x_3 com matriz de covariância \mathbf{V} .

- 8.4. Considere a matriz de covariância das variáveis aleatórias a e b , \mathbf{V}_{ab} , que provêm do ajuste dos parâmetros a e b da função $y = a + bx$ a dados experimentais $\{(x_i, y_i)\}$, com sua respectiva matriz de covariância.

- (a) A partir de \mathbf{V}_{ab} , determine a matriz de covariância dos valores interpolados em x_1 e x_2 , $y_1 = a + bx_1$ e $y_2 = a + bx_2$, com x_1 e x_2 conhecidos exatamente.
- (b) Mostre que o coeficiente de correlação entre y_1 e y_2 tende a 1 quando x_1 tende a x_2 .

8.5. As energias de três transições gama: $E_{1,0}$, $E_{2,0}$ e $E_{3,0}$, foram medidas independentemente, obtendo-se os resultados E_1 , E_2 e E_3 , com matriz de covariâncias $\sigma^2 \mathbf{I}_3$. A figura 8.3 mostra a relação física entre essas energias. Determine:

- (a) As estimativas \hat{E}_1 e \hat{E}_2 , bem como sua matriz de covariâncias, quando o vínculo $E_{1,0} + E_{2,0} = E_{3,0}$ é imposto.
- (b) A estimativa \hat{E}_3 e a matriz de covariância de \hat{E}_1 , \hat{E}_2 e \hat{E}_3 .

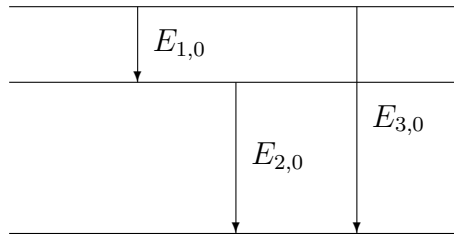


Figura 8.3: Parte do diagrama de níveis de energia de um núcleo, onde estão assinaladas as transições cujas energias são vinculadas.

8.6. Na medida de uma grandeza y que depende de x de acordo com a relação $y = \alpha + \beta x$, foram obtidos os dados da tabela 8.1. Todos os dados y_i são estatisticamente independentes ($\text{cov}(y_i, y_j) = 0$ para $i \neq j$) e têm a mesma variância, enquanto que x foi medido sem erro. Estime:

Tabela 8.1: Dados para o exercício 8.6.

x	-3	-2	-1	0	1	2	3
y	-10,9	-5,4	-5,0	-0,3	1,3	5,9	7,1

- (a) α , β e a correspondente matriz de covariância.
- (b) a variância dos dados y_i .

8.7. Uma partícula se desintegra em vôo, originando dois fragmentos. Estime o módulo da quantidade de movimento da partícula, sabendo que antes da partícula desintegrar-se ele tinha sido medido como P com desvio padrão σ e que os fragmentos provenientes da desintegração tinham quantidades de movimento de módulos $P_1(\sigma_1)$ e $P_2(\sigma_2)$ e formavam com a direção da quantidade de movimento da partícula os ângulos indicados na figura abaixo. Ignore os erros de medida dos ângulos e suponha que os dados P , P_1 e P_2 sejam estatisticamente independentes.

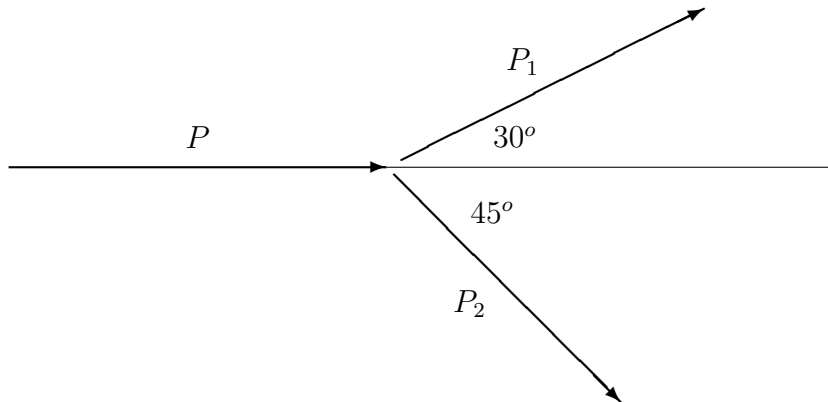


Figura 8.4: Esquema para o exercício 8.7.

- 8.8. Considere o exemplo numérico da seção 8.18. Calcule as estimativas $\hat{\theta}'_1$, $\hat{\theta}'_2$ e $\hat{\theta}'_3$ impondo os vínculos $\theta_{1,0} + \theta_{2,0} + \theta_{3,0} = 180^\circ$ e $\theta_{2,0} = \theta_{3,0}$. Calcule a soma dos quadrados dos resíduos ponderados pela inversa da matriz das variâncias. Suponha os dados gaussianos e que os desvios padrão dos dados sejam corretos e aplique o teste de χ^2 para verificar a hipótese do triângulo ser isósceles. Você poderia chegar à mesma conclusão (rejeitar, ou não, a hipótese) se os dados não fossem gaussianos?
- 8.9. Considere o modelo linear (eq. 8.4), $\mathbf{y} = \mathbf{X}\mathbf{a}_0 + \vec{\epsilon}$, onde os parâmetros ajustados são dados pela equação 8.8,

$$\tilde{\mathbf{a}} = \mathbf{M}\mathbf{y} \quad , \quad \text{onde} \quad \mathbf{M} = (\mathbf{X}^t\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{V}^{-1} \quad .$$

Considerando a identidade $\mathbf{y} = \mathbf{I} \cdot \mathbf{y}$, onde \mathbf{I} é a matriz identidade de ordem igual ao número de dados experimentais no vetor \mathbf{y} , mostre que a matriz de covariância entre $\tilde{\mathbf{a}}$ e os dados experimentais é dada por

$$\mathbf{V}_{\tilde{\mathbf{a}}\mathbf{y}} = \begin{pmatrix} (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} & (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \cdot \mathbf{X}^t \\ \mathbf{X} \cdot (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} & \mathbf{V} \end{pmatrix} .$$

Sugestão: considere que $\tilde{\mathbf{a}}$ e \mathbf{y} dependem dos dados segundo a equação

$$\begin{pmatrix} \tilde{\mathbf{a}} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{M} \\ \mathbf{I} \end{pmatrix} \cdot \mathbf{y} .$$

- 8.10. Uma determinada grandeza foi medida por dois equipamentos, simultaneamente. Os resultados fornecidos pelos equipamentos são $y_1 = y_0 + e_1 + e_2$ e $y_2 = y_0 + 2e_1 + e_3$, onde os valores esperados dos erros e_i são nulos, $\langle e_i \rangle = 0$, com desvios padrões dados por $\sigma_i = 1$ e não correlacionados, $\langle e_i \cdot e_j \rangle = 0$ para $i \neq j$. Uma das fontes de erro, e_1 , afeta ambos os resultados, mas de forma diferente; isso pode ter como origem, por exemplo, um ruído da rede elétrica que tem efeito duas vezes maior no segundo dado do que no primeiro.

- (a) Escreva a matriz de covariância dos três erros.
 (b) Mostre que a matriz de covariância de y_1 e y_2 é dada por

$$\mathbf{V} = \begin{pmatrix} 2 & 2 \\ 2 & 5 \end{pmatrix} .$$

- (c) Considerando essa matriz de covariância, verifique que o valor adotado para a grandeza medida, dada pelo MMQ, é uma média entre os dois dados experimentais em que o segundo deles tem peso negativo.

- 8.11. Considere os dados experimentais da tabela 8.2, em que os valores de x_i são exatos e as incertezas de y_i são todas iguais a 0,2, com covariâncias nulas entre eles. Considere o ajuste dos parâmetros de uma reta, $y = a + b \cdot x$, cujos valores ajustados são calculados como

$$\hat{\mathbf{a}} = \mathbf{M}\mathbf{y} \quad , \quad \text{em que} \quad \mathbf{M} = (\mathbf{X}^t \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{V}^{-1}$$

é uma matriz que indica quanto cada um dos dados experimentais contribui para cada um dos parâmetros.

Tabela 8.2: Dados para o exercício 8.11.

x	1	2	3	4	5
y	3,3	4,9	6,4	8,0	9,5

- (a) Escreva explicitamente a matriz \mathbf{M} e verifique que o terceiro dado, y_3 , em nada contribui para a definição do parâmetro b .
- (b) Obtenha os parâmetros ajustados bem como a correspondente matriz de covariância.
- (c) Faça novamente o ajuste descartando o terceiro dado e compare os valores obtidos, tanto dos parâmetros como da matriz de covariância, confirmando o resultado obtido no item (a).

8.12. A massa de uma amostra líquida foi medida dentro de um recipiente e, em seguida, usando-se outra balança, mediu-se a massa do recipiente vazio. Esse procedimento foi repetido várias vezes, sempre com balanças diferentes. Os resultados estão na tabela 8.3. Considere que todos os dados têm o mesmo desvio padrão, desconhecido, não havendo covariância entre eles.

Tabela 8.3: Dados para o exercício 8.12. A primeira linha contém as massas do recipiente com amostra e a segunda, sem amostra. Valores em gramas.

com	10,5	12,9	12,1	9,8	10,3	12,3	11,1	11,4	9,3	11,2
sem	9,8	10,3	9,4	9,0	9,6	11,1	12,6	10,2	9,7	10,7

- (a) Usando os dados da tabela, ajuste os valores a serem adotados para as massas do recipiente e da amostra.
- (b) Com o resultado do item (a), estime o valor do desvio padrão dos dados e, a partir dele, a matriz de covariância das massas estimadas.
- (c) Em dois casos, a massa do recipiente com a amostra líquido resultou em um valor menor do que a massa do recipiente sem amostra; argumente no sentido de justificar porque esses dados “não físicos” não devem ser descartados.

- 8.13. Após o ajuste dos parâmetros a e b da função $y = a + bx$ por pontos experimentais, estes foram descartados, guardando-se apenas os valores ajustados, \tilde{a} e \tilde{b} , e a correspondente matriz de covariância, \mathbf{V} . Depois disso, um novo dado (x, y) foi obtido, sendo que y tem desvio padrão σ e não é covariante com os dados inicialmente usados no ajuste dos parâmetros.
- (a) Escreva a expressão do modelo linear do MMQ, relação (8.4), de modo que permita incluir esse novo dado no ajuste.
 - (b) Obtenha a equação que leva aos novos valores ajustados dos parâmetros bem como à nova matriz de covariância.

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