Matriz de Variância e Covariância e o Teorema de Gauss-Markov

Reginaldo J. Santos
Departamento de Matemática-ICEx
Universidade Federal de Minas Gerais
http://www.mat.ufmg.br/~regi

26 de setembro de 2001

Seja $X = [x_1 \ \dots \ x_n]^t$ um vetor de variáveis aleatórias $x_1, \dots, x_n$. A operação de tomar a **esperança** será denotada por E. Definimos $E(X)$ como sendo o vetor dos valores esperados de cada elemento de X, ou seja,

$$E(X) = E \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} E(x_1) \\ E(x_2) \\ \vdots \\ E(x_n) \end{bmatrix}.$$

1 Matriz de Variância e Covariância

Sejam $x_1, \dots, x_n$ variáveis aleatórias com variâncias $\sigma_1^2, \dots, \sigma_n^2$ e covariâncias $\sigma_{12}, \sigma_{13}, \dots, \sigma_{(k-1)k}$. Ou seja,

$$\sigma_i^2 = E[(x_i - E(x_i))^2], \quad \sigma_{ij} = E[(x_i - E(x_i))(x_j - E(x_j))], \text{ para } i \neq j$$

Reunindo as variâncias e covariâncias em uma matriz, ficamos com

$$\mathrm{Var}(X) = V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{bmatrix}$$

que é chamada de **matriz de variância e covariância** ou **matriz de dispersão** das variáveis aleatórias $x_1, \ldots, x_n$. Ela é simétrica ($V^t = V$) e o elemento de posição i, i é a variância da variável x_i e o elemento de posição i, j, para $i \neq j$, é a covariância, entre as variáveis x_i e x_j. Assim, podemos expressar V como

$$V = \mathrm{Var}(X) = E[(X - E(X))(X - E(X))^t]. \tag{1}$$

Seja A uma matriz $m \times n$. Então,

$$E(AX) = AE(X), \tag{2}$$

pois

$$\begin{aligned} E(AX) &= E \begin{bmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n \\ \vdots & & & & \ldots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} a_{11}E(x_1) & + & a_{12}E(x_2) & + & \ldots & + & a_{1n}E(x_n) \\ a_{21}E(x_1) & + & a_{22}E(x_2) & + & \ldots & + & a_{2n}E(x_n) \\ \vdots & & & & \ldots & & \vdots \\ a_{m1}E(x_1) & + & a_{m2}E(x_2) & + & \ldots & + & a_{mn}E(x_n) \end{bmatrix} = AE(X) \end{aligned}$$

De forma análoga podemos mostrar que se B é uma matriz $n \times m$, então

$$E(XB) = E(X)B \quad \text{e} \quad E(AXB) = AE(X)B. \tag{3}$$

De (2) e (3) segue que

$$\begin{aligned} \mathrm{Var}(AX) &= E[(AX - E(AX))(AX - E(AX))^t] = E[(AX - AE(X))(AX - AE(X))^t] \\ &= E[A(X - E(X))(X - E(X))^t A^t] = AE[(X - E(X))(X - E(X))^t]A^t \\ &= A\mathrm{Var}(X)A^t \end{aligned} \tag{4}$$

2 Introdução à Análise de Regressão

Vamos supor que um vetor de variáveis aleatórias $Y = [y_1, \ldots, y_m]^t$ seja tal que a sua esperança seja uma combinação linear de outros vetores, ou seja, que

$$\begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_m) \end{bmatrix} = E(Y) = b_1X_1 + \ldots + b_nX_n = b_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \end{bmatrix} + \ldots + b_n \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{mn} \end{bmatrix} \tag{5}$$

A equação (5) pode ainda ser escrita de duas outras formas:

$$E(y_i) = b_1 x_{i1} + \ldots + b_n x_{in}, \quad \text{para } i = 1, \ldots, m$$

ou simplesmente

$$E(Y) = XB, \tag{6}$$

onde

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & & \ldots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{bmatrix}, \quad \text{e} \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

O problema aqui é determinar os parâmetros b_i a partir de observações y_i, para $i = 1, \ldots, m$. Para cada i, a diferença $y_i - E(y_i)$ é o desvio do valor observado y_i em relação ao valor esperado $E(y_i)$ e é escrito como

$$\varepsilon_i = y_i - E(y_i), \quad \text{para } i = 1, \ldots, m \tag{7}$$

Assim, em termos das observações e dos erros, o nosso modelo pode ser escrito como

$$y_i = b_1 x_{i1} + \ldots + b_n x_{in} + \varepsilon_i, \quad \text{para } i = 1, \ldots, m$$

ou de forma mais compacta, simplesmente

$$Y = XB + \varepsilon, \tag{8}$$

onde

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \quad X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \vdots & & \ldots & \vdots \\ x_{m1} & x_{m2} & \ldots & x_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}.$$

A equação (8) é chamada de **equação do modelo**. Ela é a base para estimar B a partir dos dados obtidos armazenados em X e Y.

Os erros ε_i por definição, têm média zero, pois de (7) temos que

$$E(\varepsilon) = E(Y - E(Y)) = E(Y) - E(Y) = \bar{0}.$$

Vamos assumir também que os erros ε_i têm a mesma variância σ^2 e que cada um deles é não correlacionado (tem covariância zero) com os outros.

Assim a matriz de variância-covariância dos ε_i's é $\sigma^2 I_n$. Portanto,

$$\sigma^2 I_n = \mathrm{Var}(\varepsilon) = E[(\varepsilon - E(\varepsilon))(\varepsilon - E(\varepsilon))^t] = E(\varepsilon\varepsilon^t). \tag{9}$$

É usual se tomar como estimador do vetor de parâmetros B, a solução de quadrados mínimos, ou seja, a solução de

$$\min \sum_{i=1}^{n} \varepsilon_i^2 = \min \ ||XB - Y||^2.$$

Este problema é equivalente a resolver as equações normais

$$X^t X B = X^t Y.$$

A solução de quadrados mínimos foi usada por Gauss em 1801 para predizer a trajetória do asteróide Ceres. Dias após o asteróide ter sido descoberto, o seu rastreamento foi perdido. Vários astrônomos publicaram artigos fazendo a previsão da trajetória do asteróide. Entretanto, quando o asteróide foi novamente localizado, a sua posição era muito próxima daquela prevista por Gauss e diferia muito das previsões feitas pelos outros.

Na maioria dos casos a matriz X tem posto máximo, ou equivalentemente $X^t X$ é não singular. Nestes casos temos uma fórmula para o estimador $\hat{B}$

$$\hat{B} = (X^t X)^{-1} X^t Y \tag{10}$$

O estimador de quadrados mínimos é não viesado, ou seja, $E(\hat{B}) = B$, pois de (10) e (6), temos

$$E(\hat{B}) = E[(X^t X)^{-1} X^t Y] = (X^t X)^{-1} X^t E(Y) = (X^t X)^{-1} X^t X B = B$$

Este estimador, que vamos chamar de $\hat{B}$ é o "melhor" estimador linear não viciado, como mostra o próximo teorema.

Teorema 2.1 (Gauss-Markov). *Considere o modelo linear*

$$Y = XB + \varepsilon,$$

com as hipóteses

$$E(\varepsilon) = \bar{0} \quad e \quad \mathrm{Var}(\varepsilon) = \sigma^2 I_n.$$

Seja $\hat{B} = (X^tX)^{-1}X^tY$ *o estimador de quadrados mínimos. Se* $\tilde{B}$ *é um outro estimador de* B *tal que* $\tilde{B} = CY$*, onde* C *é uma matriz* $n \times n$*, e* $E(\tilde{B}) = B$ *(não viciado), então:*

$$Z^t\mathrm{Var}(\tilde{B})Z \geq Z^t\mathrm{Var}(\hat{B})Z, \quad \textit{para todo } Z \in \mathbb{R}^n$$

Demonstração. Vamos demonstrar para o caso em que a matriz X tem posto máximo, ou equivalentemente X^tX é não singular. Por (4), temos que

$$\mathrm{Var}(\hat{B}) = (X^tX)^{-1}X^t\mathrm{Var}(Y)X(X^tX)^{-1}. \tag{11}$$

Mas, por (1), (6) e (9) a variância de Y é dada por

$$\mathrm{Var}(Y) = E[(Y - E(Y))(Y - E(Y))^t] = E(\varepsilon\varepsilon^t) = \sigma^2 I_n. \tag{12}$$

Substituindo (12) em (11) obtemos

$$\mathrm{Var}(\hat{B}) = \sigma^2(X^tX)^{-1}. \tag{13}$$

Por outro lado, por (4) e usando (12), obtemos

$$\mathrm{Var}(\tilde{B}) = C\mathrm{Var}(Y)C^t = \sigma^2(CC^t). \tag{14}$$

Agora, como por hipótese $E(\tilde{B}) = B$, segue que

$$B = E(\tilde{B}) = E(CY) = CE(Y) = CXB, \quad \text{para todo } B \in \mathbb{R}^n.$$

O que implica que

$$I_n = CX. \tag{15}$$

Assim, usando (13), (14) e (15) segue que

$$\mathrm{Var}(\tilde{B}) - \mathrm{Var}(\hat{B}) = \sigma^2[CC^t - CX(X^tX)^{-1}X^tC^t] = \sigma^2C[I_n - X(X^tX)^{-1}X^t]C^t. \tag{16}$$

A matriz $M = I_n - X(X^tX)^{-1}X^t$ é simétrica e tal que $M^2 = M$ (**idempotente**). Assim,

$$\begin{aligned} Z^t\mathrm{Var}(\tilde{B})Z - Z^t\mathrm{Var}(\hat{B})Z &= Z^t[\mathrm{Var}(\tilde{B}) - \mathrm{Var}(\hat{B})]Z \\ &= \sigma^2 Z^t(CMC^t)Z = \sigma^2(Z^tCM)(M^tC^tZ) \\ &= \sigma^2||Z^tCM||^2 \geq 0, \end{aligned}$$

o que prova o resultado. $\square$

Referências

[1] David C. Lay. *Álgebra Linear e suas Aplicações*. Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 2a. edition, 1999.

[2] Steven J. Leon. *Álgebra Linear com Aplicações*. Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 5a. edition, 1998.

[3] Reginaldo J. Santos. *Um Curso de Geometria Analítica e Álgebra Linear*. Imprensa Universitária da UFMG, Belo Horizonte, 2010.

[4] Shayle R. Searle. *Matrix Algebra Useful for Statistics*. John Wiley and Sons, New York, 1982.

[5] S. D. Silvey. *Statistical Inference*. Penguin, Harmondsworth, 1970.