Quadrados Mínimos Generalizados

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

16 de abril de 2012

Proposição 1. *Seja A uma matriz simétrica. Se L e M são matrizes triangulares inferiores com 1's na diagonal e D uma matriz diagonal invertível tais que $A = LDM^t$, então $M = L$.*

Demonstração. Vamos mostrar que $M^{-1}L = I$. Seja $B = M^{-1}AM^{-t}$. Então $B^t = B$ e $B = M^{-1}LD$. Como M^{-1} e L são triangulares inferiores, então B é diagonal. Como D é invertível, então $M^{-1}L$ também é diagonal e como L e M são matrizes triangulares inferiores com 1's na diagonal, então $M^{-1}L = I$. $\square$

A decomposição $A = LDL^t$, em que L é uma matriz triangular inferior com 1's na diagonal e D uma matriz diagonal, é chamada decomposição de Cholesky de A.

Uma matriz A simétrica é chamada positiva definida se $X^tAX > 0$, para todo $X \neq \bar{0}$.

Proposição 2. *Se $A = LDL^t$ é a decomposição de Cholesky de uma matriz A positiva definida, então os elementos da diagonal de D são maiores que zero.*

Demonstração. Se A é positiva definida, então $X^tAX > 0$, para todo $X \neq \bar{0}$. Substituindo-se $A = LDL^t$ em X^tAX obtemos que $X^tLDL^tX > 0$, para todo $X \neq \bar{0}$. Seja $Y = L^tX$. Então $Y^tDY > 0$, para todo $Y \neq \bar{0}$. Em particular para $Y = E_i = [0\ \cdots\ 0\ 1\ 0\ \cdots\ 0]^t$ temos que $d_{ii} = E_i^tDE_i > 0$. $\square$

Proposição 3. *Se Ω é simétrica e e positiva definida e L é tal que $L\Omega = DU$, com U triangular superior com 1's na diagonal e D uma matriz diagonal, então $\Omega = L^{-1}DL^{-t}$ é a sua decomposição de Cholesky e $P = D^{-1/2}L$ é triangular inferior, invertível e tal que*

$$\Omega^{-1} = P^t P.$$

Demonstração. Multiplicando-se à esquerda $L\Omega = DU$ por L^{-1} obtemos

$$\Omega = L^{-1}DU.$$

Pela proposição anterior $U = L^{-t}$ e portanto $\Omega = L^{-1}DL^{-t}$. Logo

$$P^t P = L^t D^{-1/2} D^{-1/2} L = L^t D^{-1} L = \Omega^{-1}.$$

□

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

$$\begin{bmatrix} E(y_1) \\ E(y_2) \\ \vdots \\ E(y_m) \end{bmatrix} = E(Y) = b_1 X_1 + \ldots + b_n X_n = b_1 \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{m1} \end{bmatrix} + \ldots + b_n \begin{bmatrix} x_{1n} \\ x_{2n} \\ \vdots \\ x_{mn} \end{bmatrix} \tag{1}$$

A equação (1) 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

$$E(Y) = XB, \tag{2}$$

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

$$\varepsilon_i = y_i - E(y_i), \quad \text{para } i = 1, \ldots, m \tag{3}$$

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

$$Y = XB + \varepsilon, \tag{4}$$

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 (4) 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 (3) temos que

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

Vamos assumir também que os erros ε_i têm matriz variância e covariância Ω. Portanto,

$$\Omega = \mathrm{Var}(\varepsilon) = E[(\varepsilon - E(\varepsilon))(\varepsilon - E(\varepsilon))^t] = E(\varepsilon\varepsilon^t). \tag{5}$$

Como Ω é simétrica e invertível pela proposição anterior existe uma matriz P triangular inferior e invertível tal que

$$\Omega^{-1} = P^t P.$$

Multiplicando-se a equação (4) à esquerda por P obtemos

$$PY = PXB + P\varepsilon \tag{6}$$

Vamos fazer as mudanças de variáveis $X^* = PX$, $Y^* = PY$ e $\varepsilon^* = P\varepsilon$. Então a equação (6) se transforma em

$$Y^* = X^* B + \varepsilon^*. \tag{7}$$

Assim

$$E(\varepsilon^*) = PE(\varepsilon) = 0$$

e

$$\operatorname{Var}(\varepsilon^*) = P\operatorname{Var}(\varepsilon)P^t = P\Omega P^t = P(P^tP)^{-1}P^t = PP^{-1}P^{-t}P^t = I.$$

Logo o problema dado pela equação (7) pode ser resolvido usando o Teorema de Gauss-Markov por

$$\hat{B} = (X^{*t}X^*)^{-1}X^{*t}Y^* = (X^tP^tPX)^{-1}X^tP^tPY = (X^t\Omega^{-1}X)^{-1}X^t\Omega^{-1}Y.$$

Proposição 4. *Se*

$$\Omega = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \cdots & \rho & 1 \end{bmatrix}$$

então

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \cdots & 0 \\ -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\rho & 1 \end{bmatrix}$$

é tal que $\Omega^{-1} = P^tP$.

Demonstração. As matrizes elementares

$$E_{i1} = \begin{bmatrix} 1 & 0 & & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & & 0 \\ \vdots & & \ddots & & & \vdots \\ -\rho^{i-1} & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & \cdots & & 0 & 1 \end{bmatrix}$$

são tais que

$$E_{21}\cdots E_{n1}=\begin{bmatrix} 1 & 0 & & \cdots & & 0\\ -\rho & 1 & 0 & \cdots & & 0\\ \vdots & & \ddots & & & \vdots\\ -\rho^{i-1} & 0 & \cdots & 1 & \cdots & 0\\ \vdots & & & & \ddots & \vdots\\ -\rho^{n-1} & 0 & \cdots & & 0 & 1 \end{bmatrix}$$

e

$$E_{21}\cdots E_{n1}\Omega=\begin{bmatrix} \frac{1}{1-\rho^2} & \frac{\rho}{1-\rho^2} & \frac{\rho^2}{1-\rho^2} & \cdots & & \frac{\rho^{n-1}}{1-\rho^2}\\ 0 & 1 & \rho & \cdots & & \rho^{n-2}\\ 0 & \rho & 1+\rho^2 & \cdots & & \rho^{n-3}(1+\rho^2)\\ \vdots & \vdots & \vdots & & & \vdots\\ 0 & \rho^{n-2} & \cdots & & \rho & 1+\rho^2+\cdots+\rho^{2(n-2)} \end{bmatrix}.$$

As matrizes elementares

$$E_{i2}=\begin{bmatrix} 1 & 0 & & \cdots & & 0\\ 0 & 1 & 0 & \cdots & & 0\\ \vdots & & \ddots & & & \vdots\\ 0 & -\rho^{i-2} & \cdots & 1 & \cdots & 0\\ \vdots & & & & \ddots & \vdots\\ 0 & 0 & \cdots & & 0 & 1 \end{bmatrix}$$

são tais que

$$E_{32}\cdots E_{n2}E_{21}\cdots E_{n1}=\begin{bmatrix} 1 & 0 & & \cdots & & 0\\ -\rho & 1 & 0 & \cdots & & 0\\ \vdots & & \ddots & & & \vdots\\ 0 & -\rho^{i-2} & \cdots & 1 & \cdots & 0\\ \vdots & & & & \ddots & \vdots\\ 0 & -\rho^{n-2} & \cdots & & 0 & 1 \end{bmatrix}$$

e

$$E_{32}\cdots E_{n2}E_{21}\cdots E_{n1}\Omega=\begin{bmatrix} \frac{1}{1-\rho^2} & \frac{\rho}{1-\rho^2} & \frac{\rho^2}{1-\rho^2} & \cdots & \frac{\rho^{n-1}}{1-\rho^2}\\ 0 & 1 & \rho & \cdots & \rho^{n-2}\\ 0 & 0 & 1 & \cdots & \rho^{n-3}\\ \vdots & \vdots & \vdots & & \vdots\\ 0 & 0 & \rho^{n-3} & \cdots & 1+\rho^2+\cdots+\rho^{2(n-3)} \end{bmatrix}.$$

Continuando com as colunas $3, 4, \ldots, n$ obtemos que

$$L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\rho & 1 \end{bmatrix}.$$

$$L\Omega = \begin{bmatrix} \frac{1}{1-\rho^2} & \frac{\rho}{1-\rho^2} & \frac{\rho^2}{1-\rho^2} & \cdots & \frac{\rho^{n-1}}{1-\rho^2} \\ 0 & 1 & \rho & \cdots & \rho^{n-2} \\ 0 & 0 & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\rho^2} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ 0 & 1 & \rho & \cdots & \rho^{n-2} \\ 0 & 0 & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Então, pela Proposição anterior

$$P = \begin{bmatrix} \frac{1}{1-\rho^2} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}^{-\frac{1}{2}} \cdot L = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \cdots & 0 \\ -\rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & -\rho & 1 \end{bmatrix}$$

é tal que $\Omega^{-1} = P^t P$. $\square$