Linear models with applications in R

PUBHLTH 744: Handout 6(Introduction to Linear Models)

Instructor: Andrea S. Foulkes

Division of Biostatistics and Epidemiology
UMass School of Public Health and Health Sciences

Fall 2007

Solutions to systems of linear equations

Consider the equation $Y = X\beta$ where $Y_{n\times 1}$, $X_{n\times p}$ and $\beta_{p\times 1}$. For a given X and Y (observed data), does there exist a solution β to this equation?

- If $p = n$ (i.e. X square) and X is nonsingular, then *yes* and the unique solution is $\beta = X^{-1}Y$. Note that in this case, the number of parameters is equal to the number of subjects, and we could not make inference.

Solutions to systems of linear equations

- Suppose $p \leq n$ and $Y \in C(X)$, then *yes* though the solution is not necessarily unique. In this case, $\beta = X^{-}Y$ is a solution since $X\beta = XX^{-}Y = Y$ for all $Y \in C(X)$ by definition of generalized inverse. Consider following 2 cases:
  - If $r(X) = p$, (X full rank) then the columns of X form a basis for $C(X)$ and the coordinates of Y relative to that basis are unique (recall notes section 2.2) and therefore the solution β is unique.
  - Suppose $r(X) < p$. If β^* is a solution to $Y = X\beta$ then $\beta^* + w$, $w \in N(X)$ is also a solution. So we have the set of all solutions to the equation equal to $\{\beta^* : \beta^* = X^{-}Y + (I - X'(X'X)^{-1}X)z,\ z \in \mathcal{R}^p\}$. Note that $X'(X'X)^{-1}X$ is the orthogonal projection operator onto $C(X')$ and so $(I - X'(X'X)^{-1}X)$ is the orthogonal projection operator onto $C(X')^{\perp} = N(X)$.

Solutions to systems of linear equations

- In general, $Y \neq C(X)$ and no solution exists. In this case, we look for a vector in $C(X)$ that is "closest" to Y and solve the equation with this vector in place of Y. This is given by MY where $M = X(X'X)^{-1}X'$ is the orthogonal projection operator onto X. Now solve:

$$MY = X\beta$$

Solutions to systems of linear equations

- The general solution (for $r(X) \leq p$) is given by $X^{-}MY + (I - X'(X'X)^{-}X)z$ and again there are infinite solutions. Let the SVD of X be given by $X = V_1 \Delta U_1'$. We know the MP generalized inverse of X is $X^{+} = U_1 \Delta^{-1} V_1'$. Therefore,

$$
\begin{aligned}
\beta^* &= X^{+}MY \\
&= X^{+}X(X'X)^{-1}X'Y \\
&= (U_1\Delta^{-1}V_1')(V_1\Delta U_1)(U_1\Delta^2 U_1)'(U_1\Delta V_1')Y \\
&= (U_1\Delta^{-1}V_1')Y \\
&= X^{+}Y
\end{aligned}
$$

So the general solution is given by $X^{+}Y + (I - X'(X'X)^{+}X)z$

Solutions to systems of linear equations

- Now assume $r(X) = p$. In this case, we have $X^{+} = (X'X)^{-1}X'$ and so

$$\begin{aligned}\beta^{*} &= X^{+}MY \\ &= (X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'Y\end{aligned}$$

Random vectors and matrices

Definition: Let $Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}$ be a random vector with $E(Y_i) = \mu_i$, $Var(Y_i) = \sigma_{ii}$ and $Cov(Y_i, Y_j) = \sigma_{ij}$. The expectation of Y is given by

$$E(Y) = \begin{pmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix} = \mu$$

Similarly, the expectation of a matrix is the matrix of expectations of the elements of that matrix.

Random vectors and matrices

Definition: Suppose Y is an $n \times 1$ vector of random variables. The covariance of Y is given by the matrix:

$$\Sigma = Cov(Y) = E\left[(Y-\mu)(Y-\mu)'\right] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & & & \\ \vdots & & & \\ \sigma_{n1} & & & \sigma_{nn} \end{pmatrix}$$

where $\sigma_{ij} = E[(Y_i - \mu_i)(Y_j - \mu_j)'] =$
$E[Y_i Y_j - Y_j \mu_i - Y_i \mu_j + \mu_i \mu_j] = E(Y_i Y_j) - E(Y_i)E(Y_j)$.

Random vectors and matrices

Theorem: Suppose Y is a random $n \times 1$ vector with mean $E(Y) = \mu$ and covariance $Cov(Y) = \Sigma$. Further suppose the elements of $A_{r\times n}$ and $b_{r\times 1}$ are scalar constants. Then,

$$E(AY + b) = AE(Y) + b = A\mu + b$$

and

$$Cov(AY + b = ACov(Y)A' = A\Sigma A'$$

Random vectors and matrices

Definition: Let $Y_{s\times 1}$ and $W_{r\times 1}$ be random vectors with $E(Y) = \mu$ and $E(W) = \gamma$. The covariance between Y and W is given by

$$Cov(W, Y) = E[(W - \gamma)(Y - \mu)']$$

We call this a matrix of covariances (not necessarily square) which is distinct from a covarince matrix.

Random vectors and matrices

Theorem: Let $Y_{s\times 1}$ and $W_{r\times 1}$ be random vectors with $Cov(Y) = \Sigma_Y$, $Cov(W) = \Sigma_W$, $Cov(W,Y) = \Sigma_{WY}$ and $Cov(Y,W) = \Sigma_{YW}$. Further suppose $A_{n\times r}$ and $B_{n\times s}$ are matrices of constant scalars. Then

$$Cov(AW + BY) = A\Sigma_W A' + B\Sigma_Y B' + A\Sigma_{WY} B' + B\Sigma_{YW} A'$$

Random vectors and matrices

Theorem: Covariance matrices are always positive semi-definite.
Proof: Let $Y_{n\times 1}$ be a random vector and

$\Sigma = Cov(Y) = E[(Y-\mu)(Y-\mu)']$ where $\mu = E(Y)$. We need to show that for any $x \in \mathcal{R}^n$, $x'\Sigma x \geq 0$. Let $Z = (Y-\mu)$, then we have:

$$\begin{aligned} x'\Sigma x = x'E(ZZ')x &= E(xZZ'x) \qquad \text{(since } x \text{ is a vector of scalars)} \\ &= E(w'w) \qquad \text{(where } w = Z'x\text{)} \\ &= E\left(\sum_{i=1}^{n} w_i^2\right) \\ &= \sum_{i=1}^{n} E\left(w_i^2\right) \geq 0 \end{aligned}$$

Random vectors and matrices

Since the expectation of a non-negative random variable will always be non-negative. Note that if $w_i = 0$ for all i, then we have $Z'x = z_1x_1 + z_2x_2 + \ldots + z_nx_n = 0$ where z_i is the ith column of Z'. This implies dependency among the columns and singularity of the covariance matrix.