多元线性回归的参数估计方法

吴仕勋，赵东方，金秀云
华中师范大学数学与统计学学院，湖北武汉（430079）
E-mail: wushixun333@163.com

摘 要：本文依据高斯—马尔可夫定理，通过对最小二乘估计方法得出的参数估计值的分析，从另外两个角度出发得出了参数估计的值与最小二乘估计的值是相同的。

关键词：最小二乘法；参数估计；线性
中图分类号：f224.0

1． 前言

多元线性回归的参数估计方法有很多种，对于线性的来说当然是最小二乘法得出的参数估计是最好的，它是最优线性无偏估计量(BLUE)。事实上最小二乘法是根据残差平方和最小化准则获得的。

2． 高斯－马尔可夫定理

假设样本数据以如下形式给出，

$$Y_{n\times 1}=\begin{pmatrix} y1 \\ \vdots \\ y_n \end{pmatrix} \quad X_{n\times k}=\begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}$$

假设有如下的一个简单的线性关系，变量 y_i 是 k 个变量 $(x_{i1},x_{i2},\cdots,x_{ik})$ 和不可观察随机扰动项 u_i 的线性函数，$y_i=\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik}+u_i, i=1,2,\cdots,n$

用矩阵形式 $Y=X\beta+u$ （1），其中 $\beta=(\beta_1,\cdots,\beta_k)'$ 是 $k\times 1$ 维常量

回归的目的是利用观察到的 y 和 X 对（1）中的未知参数进行推断，包括回归系数 β。为了使回归得以进行，必须对数据生成过程进行若干假定。除了对模型（1）的线性假设之外，还有如下假定：

1.1 $\mathrm{E}(u\,|\,X)=0$

1.2 $\mathrm{E}(uu'\,|\,X)=\sigma^2\mathrm{I}_{n\times n}$，其中 $\mathrm{I}_{n\times n}$ 为 $n\times n$ 单位矩阵。

1.3 $rank(X)=k$

在 1.1 和 1.2 假定下，基本线性回归模型（1）有 $k+1$ 个未知参数，即 k 个系数 β 以及方差 σ^2。回归分析就是通过观察到的 y 和 X 推断出这些参数。事实上有多种方法估计参数 β 和 σ^2。给定 β 值，在 1.1 假定下，y_i 在给定 $k\times 1$ 向量 x_i 下的条件均值 $E(y_i)=X_i'\beta$，相应的与真实值的离差 $u_i=y_i-E(y_i)$。最小二乘法[1]就是选择 β 的值使得 y_i 的真实值与预测的条件均值 $E(y_i)$ 的离差平方和最小。更正规地，最小二乘法就是用沿 y 轴估计地残差平方和最小值 S 作为 β 的估计值，S 定义为：

$$S=\sum_{i=1}^{n}(y_i-x_i^{'}\beta)^2=\sum_{i=1}^{n}u_i^2=(Y-X\beta)^{'}(Y-X\beta) \qquad (2)$$
$$=Y^{'}Y-2\beta^{'}X^{'}Y+\beta^{'}X^{'}X\beta$$

为得到使 S 最小化的 β 值，将(2)中的 S 对 β 求偏导得:

$$\frac{\partial S}{\partial \beta}=-2X^{'}Y+2X^{'}X\beta$$

令 $\frac{\partial S}{\partial \beta}=0$，并用 $\hat{\beta}$ 表示 β 的估计值，得到最小二乘估计量 $\hat{\beta}$ 的方程为 $X^{'}X\hat{\beta}=X^{'}Y$

如果 $X^{'}X$ 满秩，β 的最小二乘估计量可由下面 k 个线性方程求得

$$\hat{\beta}=(X^{'}X)^{-1}X^{'}Y \qquad (3)$$

3. 矩阵的分解

我们可以通过(3)得到最小二乘估计量 $\hat{\beta}$，假定 $\hat{\beta}$ 可以分拆为 $\hat{\beta}=(\hat{\beta}_1,\hat{\beta}_2)$，其中 $\hat{\beta}_1,\hat{\beta}_2$ 分别为 $k_1\times 1$ 和 $k_2\times 2$ 向量，且 $k_1+k_2=k$。相应地，将 X 分拆为 $X=(X_1,X_2)$,那么方程 $X^{'}X\hat{\beta}=X^{'}Y$ 可以表示为 $\begin{pmatrix}X_1^{'}\\X_2^{'}\end{pmatrix}(X_1,X_2)\begin{pmatrix}\hat{\beta}_1\\\hat{\beta}_2\end{pmatrix}=\begin{pmatrix}X_1^{'}\\X_2^{'}\end{pmatrix}Y$ 化简得:

$$X_1^{'}X_1\hat{\beta}_1+X_1^{'}X_2\hat{\beta}_2=X_1^{'}Y \quad (4) \qquad X_2^{'}X_1\hat{\beta}_1+X_2^{'}X_2\hat{\beta}_2=X_2^{'}Y \quad (5)$$

对(4)两边左乘 $(X_1^{'}X_1)^{-1}$ 得 $\hat{\beta}_1=(X_1^{'}X_1)^{-1}X_1^{'}Y-(X_1^{'}X_1)^{-1}X_1^{'}X_2\hat{\beta}_2$ （6）

将(6)式代入(5)得 $\hat{\beta}_2=(X_2^{'}(I-X_1(X_1^{'}X_1)^{-1}X_1^{'})X_2)^{-1}X_2^{'}(I-X_1(X_1^{'}X_1)^{-1}X_1^{'})Y$ （7）

类似的可以对(5) 两边左乘 $(X_2^{'}X_2)^{-1}$ 得

$$\hat{\beta}_1=(X_1^{'}(I-X_2(X_2^{'}X_2)^{-1}X_2^{'})X_1)^{-1}X_1^{'}(I-X_2(X_2^{'}X_2)^{-1}X_2^{'})Y \qquad (8)$$

由于多元线性回归常数项的存在，对于多元线性回归的设计矩阵 $X_{n\times k}$ 来说，有一列都为 1。令 $X_{n\times k}$ 的最后一列(第 k 列)为 e，$e=\begin{pmatrix}1\\\vdots\\1\end{pmatrix}_{n\times 1}$ 即 $X_2=e$ 代入(5)得

$$(n\bar{x}_1,\cdots,n\bar{x}_{k-1})\begin{pmatrix}\hat{\beta}_1\\\vdots\\\hat{\beta}_{k-1}\end{pmatrix}+n\hat{\beta}_k=n\bar{Y} \text{ 即 } \hat{\beta}_k=\bar{Y}-\bar{x}_1\hat{\beta}_1-\cdots-\bar{x}_{k-1}\hat{\beta}_{k-1} \qquad (9)$$

其中 $\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{i,j},j=1,\cdots,k-1,\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$

在将 $X_2 = e$ 代入（8）式，再令 $M = I - X_2(X_2{}'X_2)^{-1}X_2{}'$，那么 $M = I - e(e'e)^{-1}e$.

我们知道 $M = M'$, $MM' = (I - e(e'e)^{-1}e)(I - e(e'e)^{-1}e) = I - \frac{1}{n}e_{n\times n} = M$ 那么(8)式可以改写为 $\hat{\beta}_1 = (X_1{}'MM'X_1)^{-1}X_1{}'MM'Y = (X_1{}'M(X_1{}'M)')^{-1}X_1{}'MM'Y$ （10）

其中 $X_1 = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1,k-1} \\ x_{21} & x_{22} & \cdots & x_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n,k-1} \end{pmatrix}$ 代入(10)式得

$$X_1{}'M = \begin{pmatrix} x_{11}-\bar{x}_1 & x_{21}-\bar{x}_1 & \cdots & x_{n1}-\bar{x}_1 \\ x_{12}-\bar{x}_2 & x_{22}-\bar{x}_2 & \cdots & x_{n2}-\bar{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,k-1}-\bar{x}_{k-1} & x_{2,k-1}-\bar{x}_{k-1} & \cdots & x_{n,k-1}-\bar{x}_{k-1} \end{pmatrix}_{k-1\times n} \qquad M'Y = \begin{pmatrix} y_1-\bar{Y} \\ y_2-\bar{Y} \\ \vdots \\ y_n-\bar{Y} \end{pmatrix}_{n\times 1}$$

$$\hat{\beta}_1 = \begin{pmatrix} \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2 & \cdots & \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i,k-1}-\bar{x}_{k-1}) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i,k-1}-\bar{x}_{k-1}) & \cdots & \sum_{i=1}^{n}(x_{i,k-1}-\bar{x}_{k-1})^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{n}(x_{i,1}-\bar{x}_1)(y_i-\bar{Y}) \\ \vdots \\ \sum_{i=1}^{n}(x_{i,k-1}-\bar{x}_{k-1})(y_i-\bar{Y}) \end{pmatrix}$$

从 $\hat{\beta}_1$ 的形式我们可以得出另外一种求最小二乘估计量的方法。首先，将每个变量变化为与各自均值的离差形式，即 $\dot{y}_i = y_i - \bar{y}, \dot{x}_{ij} = x_{i,j} - \bar{x}_j, j = 1\cdots k-1$。

然后作 $\dot{y}_i$ 对 $\dot{x}_{i1},\cdots,\dot{x}_{i,k-1}$ 的回归求得最小二乘估计量 $\hat{\beta}_1,\cdots,\hat{\beta}_{k-1}$，最后将它们代入(9)式中得到 $\hat{\beta}_k$。

4. 估计量的方差最小化

最小二乘估计是从残差的平方和最小的角度出发而得出的估计量。那么我们能不能从估计量的方差最小这个角度出发而得出估计量呢？我们的回答是肯定的。

首先我们来考虑简单线性回归情形，从(3)式我们知道参数的估计量是因变量的线性组合。那么不妨设 $\hat{\beta}_1 = \sum c_i y_i = \sum(\beta_1 x_i + \beta_2 + u_i)$

我们要满足的条件是 $E(\hat{\beta}_1) = \beta_1\sum c_i x_i + \beta_2\sum c_i = \beta_1$ 即 $\sum c_i x_i = 1, \sum c_i = 0$

在这两个约束条件下，求方差最小化

$$\operatorname{var}(\hat{\beta}_1) = \sum c_i^2 \operatorname{var}(y_i) = \sigma^2\sum c_i^2$$

这是一个条件极值问题，因此我们可以利用构造拉格朗日函数如下：

$L(c_i,\lambda_1,\lambda_2) = \sigma^2\sum c_i^2 + \lambda_1(\sum c_i x_i - 1) + \lambda_2(\sum c_i)$

分别对 c_i,λ_1,λ_2 求偏导得

$$\frac{\partial L}{\partial c_i}=2\sigma^2 c_i+\lambda_1 x_i+\lambda_2,\frac{\partial L}{\partial \lambda_1}=\sum c_i x_i-1,\frac{\partial L}{\partial \lambda_2}=\sum c_i$$

从 $\frac{\partial L}{\partial c_i}=0$ 得出 $c_i=-\frac{1}{2\sigma^2}(\lambda_1 x_i+\lambda_2)$ （11）

从 $\frac{\partial L}{\partial \lambda_1}=0$，$\frac{\partial L}{\partial \lambda_2}=0$ 得出 $\sum c_i x_i=1,\sum c_i=0$

对(11)式两边同时加上 $\sum$ 得 $\lambda_1\sum x_i+n\lambda_2=0$ （12）

对(11)式两边同时乘以 x_i，在加上 $\sum$ 得 $\lambda_1\sum x_i^2+\lambda_2\sum x_i=-2\sigma^2$ （13）

联立(12),(13)式可得 $\lambda_1=\frac{2n\sigma^2}{(\sum x_i)^2-n\sum x_i^2}$ $\lambda_2=-\frac{2\sigma^2\sum x_i}{(\sum x_i)^2-n\sum x_i^2}$

再将 λ_1，λ_2 代入(11)得出 $c_i=\frac{\sum x_i-nx_i}{(\sum x_i)^2-n\sum x_i^2}$

那么可以得出 $\hat{\beta}_1$ 的参数估计值为 $\hat{\beta}_1=\frac{(\sum x_i)(\sum y_i)-n\sum x_i y_i}{(\sum x_i)^2-n\sum x_i^2}$

同样我们也可以通过此方法得出 $\hat{\beta}_2=\overline{y}-\hat{\beta}_1\overline{x}$，这与最小二乘估计得出的估计量是一样的。现在我们要做的是把一元情况推广到多元，考虑估计量为 $\hat{\beta}_{k\times1}=c_{k\times n}y_{n\times1}$

满足 $E(\hat{\beta}_{k\times1})=c_{k\times n}E(y_{n\times1})=cE(X\beta+u)=\beta$ 得出 $cX=I_{n\times n}$ 我们很容易的看出

$c=(X'X)^{-1}X'$ 是满足无偏性的约束条件的。因此我们考虑估计量 $\hat{\beta}_{k\times1}$ 的方差是否达到最小。不妨设最小二乘估计量为 β，那么 $\beta=(X'X)^{-1}X'y$

$$\hat{\beta}_{k\times1}=\beta-(X'X)^{-1}X'y+cy=\beta+(c-(X'X)^{-1}X')u$$

$$Cov(\hat{\beta}_{k\times1})=\sigma^2(X'X)^{-1}+\sigma^2(c-(X'X)^{-1}X')(c-(X'X)^{-1}X')' \quad （14）$$

从(14)中的第二项是非负半正定矩阵可以知道最小二乘估计量的方差是所有的线性无偏估计量中是最小的。

5. 结束语

本文考虑的估计量都是无偏的，事实上有些有偏的估计量比无偏估计量的效果要好。评价一个估计量的好坏还是要看该估计量是否是有效估计量。

参考文献

[1] [美] 因特里格特 博德金 萧政 著 李双杰 张涛 主译[M].2004 年 4 月.

The parameter estimation method for Multi-dimensional linear regression

Wu Shixun, Zhao Dongfang, Jin Xiuyun
Central China Normal university mathematics and statistics institute, wuhan, hubei (430079)

Abstract
According to the Gauss-Markov theorem, through the analysis of parameter estimation from the least square estimation method, this paper gains the conclusion that the value of parameter estimation from other two angles is the same to the least square estimation method.
Keywords: Least squares method; parameter estimation; linear