纯粹数学与应用数学专著 第14号

线性模型参数的估计理论

陈希孺 陈桂景
吴启光 赵林城 著

科学出版社

纯粹数学与应用数学专著　第 14 号

线性模型参数的估计理论

陈希孺　陈桂景
吴启光　赵林城　著

科　学　出　版　社

内 容 简 介

本书为作者近几年在数理统计线性模型参数估计理论方面所做的研究工作的总结。

全书共分四章．第一章是预备知识；第二章讨论线性模型回归系数的最小二乘估计及一般线性估计的相合性问题；第三章介绍误差方差估计的大样本理论；第四章讨论小样本理论，即回归系数的线性估计与误差方差的二次型估计的容许性问题．

本书读者对象为大学数学系高年级学生、研究生、教师和数理统计科学研究工作者。

纯粹数学与应用数学专著　第 14 号

线性模型参数的估计理论

陈希孺　陈桂景　吴启光　赵林城　著

责任编辑　刘嘉善

科学出版社 出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行　各地新华书店经售

*

1985 年 4 月第 一 版　　开本：850×1168　1/32

1985 年 4 月第一次印刷　　印张：8 1/8

印数：精 1—3,700　　插页：精 2

　　　平 1—9,000　　字数：212,000

统一书号：13031・2869

本社书号：3968・13—1

定价：布面精装 3.30 元

　　　平 装 2.30 元

科技新书目：精 21 平 22

序　言

本书的目的是介绍线性模型理论的若干新发展．对数理统计知识有过一点接触的人，都了解线性模型的重要地位．一些富有实用意义的统计分支，诸如回归分析、方差分析和多元分析等，都以这种模型理论为基础，或与之有密切联系．因此，有关这种统计模型的一些较为古典的内容，在一般数理统计教科书中都有不同程度的介绍．近几十年来，特别是六十年代以来，线性模型理论无论在广度和深度上都有不少新发展，像大样本理论、可容许的线性与二次型估计、非参数和 Robust 估计、序贯和 Bayes 方法以及自变量也带随机误差的所谓"Error In Variables"模型等等．这些发展大都有实用上的意义：有的改进了传统的估计方法而提供了较好的估计，有的扩大了模型的应用范围，有的在误差的正态性不成立的情况下提供了可用的大样本检验和区间估计等．另一些发展的主要意义则在于纯理论方面，它加深了我们对这个重要模型的性质的认识．

本书作者近年来在这个领域里做了一点研究工作，对其现状作了一些了解，写作这本专著的念头就是由此而起．但由于篇幅所限而且由于不少新的发展目前还远未达到比较成熟和定型，所以要写一本详尽的，包括到目前为止的所有主要成果的专著是不现实的．我们希望本书内容以我们自己的工作为基础．这样，对所涉及的课题能作较深入的论述．因此，我们挑选了线性模型参数的线性和二次型估计的大样本理论和容许性这些题材，并把书名定为《线性模型参数的估计理论》．

本书共分四章．第一章是预备知识．在这一章开头列举了为阅读本书所需的各种预备知识．由于本书使用的方法没有超出古典分析、矩阵及一般分析概率论的范围，即使只读过少数基本课程

的概率统计专业学生，也不难看懂本书的绝大部分内容．第二章讨论线性模型的回归系数的最小二乘估计及一般线性估计相合性问题，介绍了在各种意义下的相合性条件．从概率论角度，可以把本章内容看作是古典大数定律的某种推广．因为它的主题无非是关于线性型 $\sum_{i=1}^{n} a_{ni} e_i$（这里 a_{ni} 是常数，e_i 为满足一定性质的随机变量）在各种意义下收敛到 0 的问题．只是 $\{a_{ni}\}$ 是由试验点列 $\{x_i\}$ 所决定，而条件必须加在 $\{x_i\}$ 上，而不能直接涉及 $\{a_{ni}\}$，因而增加了复杂性．第三章讨论线性模型的另一重要参数——误差方差估计的大样本理论，主要是讨论基于残差平方和的二次型估计．在相合性，渐近于正态分布的一致和非一致性速度等问题上，都得到了较理想的结果．在一个特殊情况下，这方面的工作早在四十年代已由许宝騄教授开其端．本章的工作可以看作是他的工作的继续和发展．第四章讨论回归系数的线性估计与误差方差的二次型估计的容许性问题．前两节主要是介绍 Rao, Cohen, Stein, James 和 Brown 等人的工作，其中包括了近代参数估计理论中的若干重大成果，过去在中文文献中还很少介绍．以后几节包括在矩阵损失下线性估计的容许性及误差方差二次型估计的容许性，则主要是本书作者及其合作者的工作．

作者的一个希望是使一些初进入研究工作的青年读者相信，使用初等工具也可以解决数理统计学中比较困难的理论问题，并达到比较深入的结果．总观近四十年来数理统计学发展的状况，给人的印象是：虽然新的结果大量涌现，但这个数学分支仍保持了这样一个特点：其多数重大结果依赖于熟练和深入地使用古典方法的技巧，而不是更新的数学工具．

本书写作分工情况如下：第一到四章的初稿分别由陈希孺、陈桂景、赵林城和吴启光执笔，最后由陈希孺写成定稿．由于作者水平所限，书中不妥和错误之处肯定不少．希望同行专家和广大读者不吝赐教．

作者

1931 年 11 月 24 日

目　　录

第一章　预备知识

阅读本书所需的预备知识有以下三方面：一是相当于大学二年级程度的数学分析与线性代数．二是数理统计．除了初等教本中包含的一般性内容外，还需要一点线性模型的估计理论知识、判决函数的基本概念、Bayes 方法初步和估计的容许性理论中若干较不常见的结果．三是概率论．需要测度论、强弱极限理论及鞅论的初步知识，程度大体上相当于 Loève 的专著[1]，个别地方还需用到 Petrov 专著[2]和 Stout 的专著[3]中的材料．

在本章中，我们打算对本书中常用的一部分知识(如线性模型的最小二乘估计理论和矩阵的广义逆等)以及某些在文献中不易查阅的事实给以较仔细的叙述．对其他内容，主要是概率论中极限理论方面，因涉及面太广，自无法在此详细叙述．但准备将一些常用结果不加证明地汇集一下，以便于查阅．了解这些结果的确切意义而不必涉及其证明细节，就可以读懂本书的有关部分．当然，如果要进一步作这方面的研究工作，则必须去钻研上面提到的有关著作，或与之相当的著作．

§1.1　矩阵与线性模型

先提出本书中常用的一些记号． m 行 n 列的矩阵 A 常称为 $m\times n$ 矩阵 A，或 $A: m\times n$．当 $m=n$ 时称为 n 阶方阵．方阵 A 的行列式记为 $|A|$．矩阵 A 的转置记为 A'．一列矩阵称为列向量，而一行矩阵称为行向量．我们总是以不加“′”的向量表列向量．例如，a 为一列向量，而 a' 则为行向量．向量 $a=(a_1,\cdots,a_n)'$ 的长为 $\left(\sum_1^n a_i^2\right)^{1/2}$，记为 $\|a\|$．矩阵 A 的秩记为 $\mathrm{rk}(A)$．若 A

为方阵，则其迹(trace)，即主对角线元之和，记为 $\mathrm{tr}(A)$. 本书中涉及的矩阵都是实的.

若 $\mathscr{A}$ 为一线性子空间，则其正交补空间记为 $\mathscr{A}^{\perp}$. 设 $A=(a_1 \vdots \cdots \vdots a_n)$，则由 A 的列向量 $a_1, \cdots, a_n$ 张成（或生成）的线性子空间记为 $\mathscr{M}(A)$. 其正交补空间记为 $\mathscr{M}^{\perp}(A)$.

正定方阵 A 记为 $A>0$. 半正定（又称非负定）方阵 A 记为 $A \geqslant 0$. 在本书中，正定与半正定方阵必为对称. 若 $A-B \geqslant 0$ $(A-B>0)$，则记为 $A \geqslant B(A>B)$. n 阶单位阵记为 I_n，或简记为 I.

设 $A: m \times n$ 的 (i,j) 元为 a_{ij}，则记为 $A=(a_{ij})_{i=1,\cdots,m,\, j=1,\cdots,n}$. 设 $A_n=(a_{ij}^{(n)})_{i=1,\cdots,u,\, j=1,\cdots,v}$, $n=1,2,\cdots$. 若

$$\lim_{n\to\infty} a_{ij}^{(n)}=0,\quad i=1,\cdots,u,\quad j=1,\cdots,v,$$

则称 $\lim\limits_{n\to\infty} A_n=O$ 或 $A_n \to O$. 若 $\{b_n\}$ 为一列正数，且当 $n\to\infty$ 时，$a_{ij}^{(n)}=O(b_n)\,(o(b_n))$ 对 $i=1,\cdots,u$, $j=1,\cdots,v$，则称 $A_n=O(b_n)\,(o(b_n))$.

（一）广义逆 设 A 为任一矩阵. 若矩阵 B 满足关系

$$ABA=A, \tag{1.1}$$

则称 B 为 A 的一个“减号广义逆”，记为 $B=A^{-}$. 在本书中，我们只用这种方式定义的广义逆，因此在以后，“减号”两字常省去.

若 A 为一满秩（非异）方阵，则 A^{-} 唯一且等于 A 的逆矩阵 A^{-1}. 以下将看到，这事实之逆亦真. 由这个性质可知，A^{-} 是 A^{-1} 的某种推广. 这在以下的性质 1 中看得更为明显.

广义逆的性质：（以下只涉及在后面有用的）

1. $B=A^{-}$ 的充要条件为，若 $Ax=c$ 有解，则 $x=Bc$ 为其一解.

证. 设 $B=A^{-}$，则(1.1)成立. 记 $A=(a_1 \vdots \cdots \vdots a_n)$，有 $ABa_i=a_i$, $i=1,\cdots,n$. 因为 $Ax=c$ 有解，任取其一解 $x^*=(x_1^*,\cdots,x_m^*)'$. 有 $c=\sum\limits_{1}^{m} x_i^* a_i$，因而

$$ABc = A\sum_1^m x_i^* \ Ba_i = \sum_1^m x_i^* \ ABa_i = \sum_1^m x_i^* a_i = c,$$

即 Bc 为一解. 反过来，若所设条件成立，则因 $Ax = a_i$ 有解 ($(0,\cdots,0,1,0,\cdots,0)'$即为一解)，有 $ABa_i = a_i$, $i = 1,\cdots,n$, 因而(1.1)成立，故 $B = A^-$.

2. 对任何 A, A^- 必存在. 更进一步：若通过初等变换将 A 变为

$$A = P\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}Q$$

的形状，则 $B = A^-$ 的充要条件为：B 有形式

$$B = Q^{-1}\begin{pmatrix} I_r & D \\ E & F \end{pmatrix}P^{-1},$$

此处 D. E. F 任意(当然，若 A 为 $m \times n$，则 $\begin{pmatrix} I_r & D \\ E & F \end{pmatrix}$ 必须为 $n \times m$).

证. 设 A 为 $m \times n$. 任取一个 $n \times m$ 的矩阵 B，表之为 $B = Q^{-1}\begin{pmatrix} G & D \\ E & F \end{pmatrix}P^{-1}$. 则

$$\begin{aligned}
ABA = A &\Leftrightarrow P\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}\begin{pmatrix} G & D \\ E & F \end{pmatrix}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}Q = P\begin{pmatrix} I_r & O \\ O & O \end{pmatrix}Q \\
&\Leftrightarrow \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}\begin{pmatrix} G & D \\ E & F \end{pmatrix}\begin{pmatrix} I_r & O \\ O & O \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \\
&\Leftrightarrow \begin{pmatrix} G & O \\ O & O \end{pmatrix} = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix} \\
&\Leftrightarrow G = I_r,
\end{aligned}$$

明所欲证. 由这个结果推出以下几点：

a. A^- 必存在.

b. $\mathrm{rk}(A^-)\geqslant \mathrm{rk}(A)$. 更进一步：若 $A:m\times n$，则对任何 t，$\mathrm{rk}(A)\leqslant t\leqslant \min(m,n)$，存在 A^-，致 $\mathrm{rk}(A^-)=t$.

c. 当且仅当 A 为满秩方阵时，A^- 才唯一.

3. 对任何矩阵 A，有

$$AA'(AA')^-A=A, \tag{1.2}$$

$$A'(AA')^-AA'=A'. \tag{1.3}$$

证. 因为任一实矩阵 $B=O$ 的充要条件是 $BB'=O$，记 $B=AA'(AA')^-A-A$. 有 $B'=(CA)'=A'C'$，其中 $C=AA'(AA')^--I$，故由(1.1)（改其中的 A 为 AA'，B 为 $(AA')^-$）

$$BB'=[AA'(AA')^-AA'-AA']C'=O,$$

因而 $B=O$. 这证明了(1.2). (1.3)类似证明.

4. 设 A 为对称方阵，将 A 表为

$$A=P\mathrm{diag}(\lambda_1,\cdots,\lambda_n)P',$$

此处 P 为正交阵，$\lambda_1,\cdots,\lambda_n$ 为 A 的全部特征根. 又 $\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 记一对角阵，其主对角线元依次为 $\lambda_1,\cdots,\lambda_n$. 则 A 之一广义逆为

$$A^-=P\mathrm{diag}(\lambda_1^{-1},\cdots,\lambda_n^{-1})P',$$

此处约定 $0^{-1}=0$. 由此可知，若 A 对称(半正定)，则必存在对称(半正定)的 A^-.

(二) 投影矩阵与幂等矩阵 设 $\mathscr{A}$ 为 n 维实向量空间中之一线性子空间. 如所周知，任一向量 x 可分解为 $x=a_x+b_x$，其中 $a_x\in\mathscr{A}$，而 $b_x\in\mathscr{A}^\perp$，且这种分解是唯一的. 变换 $x\to a_x$ 显然是一线性变换，称为"向 $\mathscr{A}$ 的投影变换". 将此变换用矩阵表为 $a_x=Bx$，则 B 称为"向 $\mathscr{A}$ 的投影(变换)阵". 一般地，若 B 为向某一线性子空间的投影阵，则称 $\mathscr{B}$ 为投影阵.

有如下的重要结果.

定理 1.1 设 A 为任一矩阵，则向 $\mathscr{M}(A)$ 的投影阵为 $P_A=A(A'A)^-A'$（由投影阵的唯一性，这结果也表示 $A(A'A)^-A'$ 与 $(A'A)^-$ 的取法无关).

证. 只需验证以下三条：a. $P_Ax\in\mathscr{M}(A)$，对任何 x. b. P_Ax

$=x$，当 $x\in\mathscr{M}(A)$. c. $x-P_Ax\perp\mathscr{M}(A)$，对任何 x.

a 是显然的. 对 b，注意若 $x\in\mathscr{M}(A)$，则存在向量 c 使 $x=Ac$. 于是由(1.2)式

$$P_Ax=P_AAc=A(A'A)^-A'Ac=Ac=x.$$

为证 c，任取 $y=Ad\in\mathscr{M}(A)$，则由(1.3)式

$$\begin{aligned}y'(x-P_Ax)&=d'A'(I-A(A'A)^-A')x\\&=d'[A'-A'A(A'A)^-A']x=O.\end{aligned}$$

明所欲证.

这个结果给投影阵以一明确的表达式，是广义逆的一重要应用. 易见，向 $\mathscr{M}^{\perp}(A)$ 的投影阵为 $I-P_A$. 又 P_A 之秩等于 A 之秩. 事实上，由 $P_A=A(A'A)^-A'$ 知 $\mathrm{rk}(P_A)\leqslant\mathrm{rk}(A)$. 又在(1.2) 中改 A 为 A'，得 $A'P_A=A'$，因而 $\mathrm{rk}(P_A)\geqslant\mathrm{rk}(A')=\mathrm{rk}(A)$.

若一方阵 A 满足 $A^2=A$，则称 A 为幂等阵. 以后我们只考虑对称的幂等阵. 易见这种方阵的特征根只能为 1 或 0，故得

1. A 为对称幂等阵的充要条件为：存在正交阵 P，致

$$A=P'\begin{pmatrix}I_r & O\\ O & O\end{pmatrix}P,\qquad r=\mathrm{rk}(A).\tag{1.4}$$

设 $P=(p_{ij})_{i,j=1,\cdots,n}$，则由(1.4)得

$$x'Ax=\sum_{i=1}^{r}\left(\sum_{j=1}^{n}p_{ij}x_j\right)^2,\qquad x=(x_1,\cdots,x_n)'.\tag{1.5}$$

又利用关系式 $\mathrm{tr}(AB)=\mathrm{tr}(BA)$，由(1.4)立得

2. 对称幂等阵的迹等于其秩(此性质对一般幂等阵也成立).

幂等阵之一重要性质为其与投影阵的联系：

3. A 为投影阵的充要条件为：A 对称幂等.

证. 设 A 为向线性子空间 $\mathscr{A}$ 的投影阵. 在 $\mathscr{A}$ 中找向量 $b_1,\cdots,b_n$，使 $\mathscr{A}=\mathscr{M}(B)$，$B=(b_1\vdots\cdots\vdots b_n)$. 则 $A=P_B=B(B'B)^-B'$. 因 $B'B$ 对称，故由广义逆的性质 4，知存在对称的 $(B'B)^-$. 取此作为 P_B 中的 $(B'B)^-$(前已指出，由 P_B 的唯一性，

P_B 与 $(B'B)^-$ 的取法无关),知 $A = P_B$ 为对称阵. 又由(1.2)(改 A 为 B')有

$$A^2 = B(B'B)^-B'B(B'B)^-B' = B(B'B)^-B' = A,$$

知 A 为幂等的. 反过来，若 A 为对称幂等，则易见，A 即为向 $\mathscr{M}(A)$ 的投影阵. 因 $Ax \in \mathscr{M}(A)$，又

$$\begin{aligned} x \in \mathscr{M}(A) &\Rightarrow x = Ac \Rightarrow Ax = AAc = Ac = x; \\ y \in \mathscr{M}(A) &\Rightarrow y = Ac \Rightarrow y'(x - Ax) = c'A'(I - A)x \\ &\qquad\qquad\qquad\qquad\qquad = c'A(I - A)x \\ &\qquad\qquad\qquad\qquad\qquad = c'(A - A^2)x = 0. \end{aligned}$$

明所欲证.

(三)线性模型与最小二乘估计 设 $\beta = (\beta_1, \cdots, \beta_p)'$ 为未知的 p 维向量，$x_i = (x_{1i}, \cdots, x_{pi})'$，$i = 1, \cdots, n$ 为已知的 p 维向量. $e_1, \cdots, e_n$ 为随机变量，则称

$$Y_i = x_i'\beta + e_i, \quad i = 1, \cdots, n \tag{1.6}$$

为一线性（回归）模型. 实际上，这表示一个包含 p 个自变量 $X_1, \cdots, X_p$ 和一个因变量 Y 的结构. 在第 i 次试验或观察时，自变量 $X_1, \cdots, X_p$ 分别取值 $x_{1i}, \cdots, x_{pi}$，而因变量则取值 Y_i. 在一些问题中，$x_{1i}, \cdots, x_{pi}$ 的值可事先指定. 这时称 $x_1, \cdots, x_n$ 为"试验点列"而 $X = (x_1 \vdots \cdots \vdots x_n)'$ 为"设计矩阵". 即使在 x_i 之值不能由试验者自由选择的场合，我们为方便计也沿用以上术语. $\beta_1, \cdots, \beta_p$ 称为回归系数. $x_i'\beta$ 可视为因变量 Y_i 的值中，依赖于自变量的部分，而 e_i 则被视为第 i 次试验的随机性误差. 因此，常假定

$$Ee_i = 0, \quad i = 1, \cdots, n. \tag{1.7}$$

引进 $Y = (Y_1, \cdots, Y_n)'$，$e = (e_1, \cdots, e_n)'$，可将线性模型(1.6)写为矩阵形式:

$$Y = X\beta + e. \tag{1.8}$$

在大样本理论中，(1.6)中的 n 是不固定的，这时写成 $Y_i = x_i'\beta + e_i$，$i = 1, \cdots, n, \cdots$. 相应于(1.8)的形式则上式为 $Y_{(n)} = X_{(n)}\beta + e_{(n)}$，$n = 1, 2, \cdots$.

线性模型理论中首要问题之一，就是利用 $x_i, Y_i, i=1,\cdots,n$，以对 β 及其线性函数作出估计，及检验关于它的假设．在此我们只讨论估计问题．称 $\hat{\beta}$ 为 β 的最小二乘（Least Squares，简记为 LS）估计，若 $\hat{\beta}$ 满足条件

$$\|Y - X\hat{\beta}\|^2 = \min\{\|Y - X\beta\|^2 : \beta \in R^p\}. \tag{1.9}$$

用微分法，得出必要条件为：$\hat{\beta}$ 是方程

$$S\beta = X'Y \quad (S = X'X) \tag{1.10}$$

之解．方程(1.10)称为正则方程．

定理 1.2 1° 方程(1.10)必有解．2°(1.10)之任一解必满足(1.9)．3° 反之，若 β^* 满足(1.9)，则 β^* 为(1.10)之一解．

证．1° 显然．因由(1.2)易知：

$$\hat{\beta} = S^- X'Y \tag{1.11}$$

就是(1.10)之一解．若 $\hat{\beta}$ 为(1.10)之任一解，则对任何 $\beta \in R^p$ 有

$$\begin{aligned}\|Y - X\beta\|^2 &= \|Y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2 \\ &\quad + 2(\hat{\beta} - \beta)'X'(Y - X\hat{\beta}) \\ &= \|Y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \beta)\|^2 \geqslant \|Y - X\hat{\beta}\|^2.\end{aligned} \tag{1.12}$$

这证明了 2°．若 β^* 满足(1.9)，则以 $\beta = \beta^*$ 代入(1.12)，应有 $\|X(\hat{\beta} - \beta^*)\|^2 = 0$，因而 $X\hat{\beta} = X\beta^*$，故

$$S\beta^* = X'X\beta^* = X'X\hat{\beta} = S\hat{\beta} = X'Y,$$

即 β^* 确为方程(1.10)之一解．证毕．

若 $\mathrm{rk}(X) = p$，则 $|S| \neq 0$，这时称为“满秩情况”．在满秩情况下，(1.10)的解，即 β 的 LS 估计，是唯一的：

$$\hat{\beta} = S^{-1}X'Y. \tag{1.13}$$

在这种情况下称 β 为可估的．一般，只有在 β 可估时，才谈到其 LS 估计．但我们以后不坚持这一点，而称可表为(1.11)的 $\hat{\beta}$ 为 β 的 LS 估计．

容易证明，若 β 可估而条件(1.7)满足，则 β 的 LS 估计 $\hat{\beta}$ 为无偏的．又若 $e = (e_1, \cdots, e_n)'$ 的协差阵为 Σ，则 $\hat{\beta}$ 的协差阵为

$$\mathrm{COV}(\hat{\beta}) = S^{-1}X'\Sigma XS^{-1}. \tag{1.14}$$

事实上，有 $\hat{\beta}=S^{-1}X'Y=S^{-1}X'(X\beta+e)=\beta+S^{-1}X'e$. 故由 $Ee=0$ 立知 $\hat{\beta}$ 为无偏的. 再注意到 S 对称，即得

$$\mathrm{COV}(\hat{\beta})=S^{-1}X'\Sigma(S^{-1}X')'=S^{-1}X'\Sigma XS^{-1}.$$

一个重要的特例是

$$\Sigma=\sigma^2I_n,\quad 0<\sigma^2<\infty, \tag{1.15}$$

这时由(1.14)得

$$\mathrm{COV}(\hat{\beta})=\sigma^2S^{-1}. \tag{1.16}$$

如果线性模型 (1.8) 满足条件 (1.7) 和 (1.15)，则称它为 Gauss-Markov 模型，简称为 GM 模型(在此并未要求 X 满秩). 条件(1.7)和(1.15)也称为 GM 条件或 GM 假定. 在讨论线性模型时，一般常假定(1.7)，故 GM 条件一般特指(1.15). 在(1.15)中，σ^2 为随机误差的(公共)方差. 它一般假定为未知，是线性模型的一重要参数.

在证明定理 1.3 时曾得到表达式

$$\hat{\beta}-\beta=S^{-1}X'e, \tag{1.17}$$

它把 $\hat{\beta}-\beta$ 表为 $e_1,\cdots,e_n$ 的线性组合，是一很有用的关系式.

(四) 可估函数与 Gauss-Markov (GM) 定理，BLUE

设有线性模型(1.8)，并设条件(1.7)成立. 设 $c'\beta$ 为 β 之一线性函数. 若存在 $c'\beta$ 之一无偏估计量 $\varphi(Y)$，则称 $c'\beta$ 为可估的 ($\varphi(Y)$ 自然还可依赖于 $x_1,\cdots,x_n$. 因 $x_1,\cdots,x_n$ 已知且无随机性，这个依赖关系不必指出). 若 C 为 $r\times p$ 矩阵，其各行向量为 $c_1',\cdots,c_r'$，而 $c_i'\beta$ 皆可估($i=1,\cdots,r$)，则称 $C\beta$ 为可估的.

定理 1.3 设 $c'\beta$ 可估，则必存在 $c'\beta$ 的形如 $a'Y$ 的线性无偏估计. 又 $c'\beta$ 可估的充要条件为 $c\in\mathscr{M}(X')$.

证. 首先，由于 $c'\beta=E_\beta\varphi(Y)$，而 Y 的分布只通过 $X\beta$ 而依赖于 β[1]，有

$$X\beta=X\beta^*\Rightarrow c'\beta=c'\beta^*.$$

---

1) 此处隐含了这样的假定：在模型(1.8)中，e 的分布与 β 无关. 若给线性模型以更一般的定义：$EY=X\beta$，$\mathrm{COV}(Y)=\sigma^2I$，则定理 1.3 的前一结论不必成立.

换句话说，若 $\beta-\beta^*$（它可取 R^p 中任何向量为值）与 $\mathscr{M}(X')$ 正交，则必与 c 正交．因而 $c\in\mathscr{M}(X')$．反过来，若 $c\in\mathscr{M}(X')$，则存在 a，使 $c=X'a$．这时 $E_\beta(a'Y)=a'X\beta=c'\beta$．因而 $c'\beta$ 可估．这一举证明了定理中的两个结论．

以 $\hat{\beta}$ 记 β 的任一个 LS 估计．当 $c'\beta$ 可估时，称 $c'\hat{\beta}$ 为 $c'\beta$ 的 LS 估计．易见此时 $c'\hat{\beta}$ 与 $\hat{\beta}$ 的取法无关．事实上，若 $\hat{\beta}_{(1)}$ 和 $\hat{\beta}_{(2)}$ 为两个 LS 估计，则 $X'Y=S\hat{\beta}_{(1)}=S\hat{\beta}_{(2)}$．因 $c'\beta$ 可估，有 $c\in\mathscr{M}(X')=\mathscr{M}(X'X)=\mathscr{M}(S)=\mathscr{M}(S')$，故存在 d，致 $c=S'd$，而

$$c'\hat{\beta}_{(1)}=d'S\hat{\beta}_{(1)}=d'S\hat{\beta}_{(2)}=c'\hat{\beta}_{(2)}.$$

明所欲证．

定理 1.4（GM 定理） 设 GM 条件满足．若 $c'\beta$ 可估，则在 $c'\beta$ 的一切线性无偏估计类中，其 LS 估计 $c'\hat{\beta}$ 是唯一的方差一致最小的估计．

证．先明确一点：从字面上看，$c'\beta$ 的线性估计可以包括形如 $a'Y+a_0$ 的非齐次估计．但要这估计为无偏，必须 $c'\beta=E_\beta(a'Y+a_0)=a'X\beta+a_0$，对一切 $\beta\in R^p$．特别，取 $\beta=0$ 得 $a_0=0$．因此，可限于考虑齐次线性估计．

现证 $c'\hat{\beta}$ 无偏．因 $c'\beta$ 可估，由定理 1.3 及其证明，知存在 a，致 $c=Sa$．故 $c'\hat{\beta}=a'S\hat{\beta}$．据 (1.10)，有 $c'\hat{\beta}=a'X'Y$，故 $E_\beta(c'\hat{\beta})=a'X'X\beta=a'S\beta=c'\beta$．因而 $c'\hat{\beta}$ 为 $c'\beta$ 之一无偏估计．

现任取 $c'\beta$ 之一无偏估计 $a'Y$．由无偏性有 $c'\beta=E_\beta(a'Y)=a'X\beta$ 对一切 β，故 $a'X=c'$．而 $c'\hat{\beta}=a'X\hat{\beta}$．依(1.15)，有

$$\begin{aligned}\mathrm{var}_{\beta,\sigma}(c'\hat{\beta})&=a'XS^-X'a\sigma^2=a'P_Xa\sigma^2\\&=a'P_XP_Xa\cdot\sigma^2=\sigma^2\|P_Xa\|^2.\end{aligned}$$

另一方面，有 $\mathrm{var}_{\beta,\sigma}(a'Y)=\sigma^2\|a\|^2$．因为对任何 a 有 $\|a\|\geqslant\|P_Xa\|$，故 $c'\hat{\beta}$ 的方差总不会超过 $a'Y$ 的方差．而要 $\|a\|=\|P_Xa\|$，充要条件为 $a\in\mathscr{M}(X)$，即存在 d，使 $a=Xd$．这时有

$$c'\hat{\beta}=a'X\hat{\beta}=d'X'X\hat{\beta}=d'S\hat{\beta}=d'X'Y=a'Y.$$

这证明了 $c'\hat{\beta}$ 为使方差最小的唯一的线性无偏估计．定理证毕．

这个重要定理奠定了在 GM 模型下，LS 估计的重要地位．如果有线性模型 (1.8) 且仍假定条件 (1.7) 成立，但代替 GM 条件 (1.15)，假定

$$\mathrm{COV}(e)=\sigma^2V,\ V>0 \text{ 已知}, 0<\sigma^2<\infty, \sigma^2 \text{ 未知}, \tag{1.18}$$

则关于线性函数 $c'\beta$ 的可估性定义，LS 估计等，并无改变．因这些只依赖于 Y 的均值向量而不依赖于其协差阵．但是，在 $V\neq I$ 的场合，可估函数 $c'\beta$ 的 LS 估计 $c'\hat{\beta}$ 一般已不再具有定理 1.4 中所描述的性质．事实上，上述模型不难转化到 GM 模型：作变换 $Z=V^{-1/2}Y$，则有

$$Z=\tilde{X}\beta+\tilde{e},$$

其中 $\tilde{X}=V^{-1/2}X$，$\tilde{e}=V^{-1/2}e$．而

$$\mathrm{COV}(\tilde{e})=V^{-1/2}\sigma^2VV^{-1/2}=\sigma^2I,$$

故对 Z 而言，GM 条件满足．若 $c'\beta$ 可估，在 Z 模型求 β 之任一 LS 估计，为

$$\tilde{\beta}=(X'V^{-1}X)^-X'V^{-1/2}Z=(X'V^{-1}X)^-X'V^{-1}Y. \tag{1.19}$$

因而得到(据 GM 定理)在一切线性无偏估计类中，唯一的一个无偏方差最小估计，为

$$c'\tilde{\beta}=c'(X'V^{-1}X)^-X'V^{-1}Y. \tag{1.20}$$

这个估计在文献中常称为"最佳线性无偏估计"(Best Linear Unbiased Estimate，简记为 BLUE)．它与 $c'\beta$ 的在 Y 模型下的 LS 估计 $c'(X'X)^-X'Y$ 一般当然不同．有时，也称(1.20)为 GM 估计．于是，在 GM 条件下，GM 估计重合于 LS 估计．

（五）σ^2 的估计 假定线性模型(1.8)满足 GM 条件 (1.7) 与 (1.15)．于是有估计 σ^2 的问题．任取 β 之一 LS 估计 $\hat{\beta}$ (在估计 σ^2 时，不必要求 β 可估)．称

$$\delta_i=Y_i-x_i'\hat{\beta},\quad i=1,\cdots,n$$

为残差．易见它们不依赖于 $\hat{\beta}$ 的选择 (因为 $x_i'\beta$ 为可估，而 $x_i'\hat{\beta}$ 为其 LS 估计．前已指出，它与 $\hat{\beta}$ 的选择无关)．通常，以 $\delta_1,\cdots,\delta_n$ 的平方和，除以适当常数(使之成为无偏的)，作为 σ^2 的估计，称之