科學前沿丛书

线性模型中的最小二乘法

Least Squares in Linear Models

陈希孺　王松桂　著

上海科学技术出版社

科學前沿丛书

线性模型中的最小二乘法

陈希孺　王松桂　著

上海科学技术出版社

图书在版编目（CIP）数据

线性模型中的最小二乘法／陈希孺，王松桂著．—上海：上海科学技术出版社，2003.3
（科学前沿丛书）
ISBN 7-5323-6602-2

Ⅰ．线…　Ⅱ．①陈…②王…　Ⅲ．线性模型—最小二乘法　Ⅳ．O241.5

中国版本图书馆 CIP 数据核字（2002）第 087916 号

责任编辑　静晓英　田廷彦

上海科学技术出版社出版发行
（上海瑞金二路 450 号　邮政编码 200020）
上海新华印刷厂印刷　新华书店上海发行所经销
2003 年 3 月第 1 版　2003 年 3 月第 1 次印刷
开本 787×1092　小 1/16　印张 18.75　插页 4　字数 279 千
印数 1—1 200　定价：42.00 元

内 容 提 要

本书所论述的主要内容是，作者及其合作者在线性统计模型与最小二乘法这一领域多年来的研究成果，以及相关的最新发展。全书可分两大部分。第一部分(1—4 章)讨论的是样本量固定时的问题，分别讨论了线性模型中各种最小二乘估计以及它们之间的关系；最小二乘估计的相对效率和各种意义下的优良性；当模型误差协方差阵含有未知参数时，两步估计的一些重要性质。第二部分(5、6 章)讨论的是当样本量趋向无穷时，模型中的误差方差有界和误差只有 r 阶矩 ($1 \leqslant r < 2$) 时估计量的情况。

本书主要读者对象是从事数理统计理论与方法研究的专家、科研人员及实际工作者。也可以供数理统计专业的大学教师、研究生作教材或参考书。

《科學前沿丛书》序

人类文明发展的长河正浩浩荡荡地流向又一个千年，在世界格局的综合国力竞争中，基础研究的发展水平已经成为一个民族的智慧、能力和国家科学技术进步的基本标志之一。

基础研究是人类对未知世界的探求，它在各门学科的前沿上展开，以认识客观世界的物质结构、各种基本运动形态和运动规律为己任，它的重大发现常常带来社会生产的革命性变化。

基础研究在科学前沿向未知领域迈进的每一步，都有赖于创新，创新是基础研究的灵魂，而创新需要很高水平的理论思维。正如 19 世纪的一位伟人所说，一个民族想要站在科学的最高峰，就一刻也不能没有理论思维。

自然科学的理论来自关于自然现象和探索实践认识的总结。这种总结通过去粗取精、去伪存真、由此及彼、由表及里的过程，实现关于自然规律认识的飞跃，在人类认识自然的知识体系上编织出新的结点。这样的结点往往又是在新的高度编织下一个结点的支撑点。一个民族想要攀登到科学的最高峰，进行高水平的理论思维，既需要一批批科学家不懈地在科学前沿上探索，也需要他们不断地进行这种实现认识上飞跃的总结。

著书立说，对一个专题或一个领域的研究成果，进行去粗取精、去伪存真、由此及彼、由表及里的总结，使之系统化、理论化，是提高理论思维水平和持续创新能力的必须。在攀登科学高峰的历程中，一部好的基础科学学术著作常常能为众多继续向上攀登的人们提供一块坚实的平台。因此，出版好基础性研究领域的学术著作，是一件十分有意义的工作。

《科學》杂志的编者和出版者，自 1915 年《科學》创刊以来，始终以传播科学为己任，在办好刊物的同时，积极地参与出版科学著作这件有意义的工作。在 20 世纪的最后五年，《科學》的出版者——上海科学技术出版社推出了一套《科學专著丛书》，出版了 14 部专著，受到了科学界和出版

界的欢迎和好评。

我高兴地看到，在迎来21世纪之时，作为上述努力的继续，该社又推出这套《科学前沿丛书》，着重于从基础性研究的前沿交叉领域选题，出版学术著作。我期望，这套丛书的编者、作者和出版者能通力合作，通过自己的辛勤劳动，以一部部精心选题、精心著述、精心编辑、精心出版的著作，参与铺筑通向中国科学再度辉煌的大道！

周光召

（《科学》杂志编委会主编）

2001年元旦

本 书 序

线性统计模型与最小二乘法是数理统计学中两个重要而有密切联系的主题，虽然最小二乘法在统计中的应用不限于线性模型，但无可怀疑，线性模型是最小二乘法(及其变种)在统计中应用最成功的领域，其成果具有理论上的深刻性和应用上的广泛性.

从历史上考察，最小二乘法起源于求解线性矛盾方程组的问题. 用现在的术语说，后者则是源于线性模型参数的估值问题. 至于线性模型本身，则是起源于天文和测地中的误差分析问题. 其一般模式是这样的：设在一个问题中涉及量 x_0, x_1, $\cdots$, x_p 和 β_1, $\cdots$, β_p，前者是易于或可以直接观测的量，而后者是难于或不能直接观测的量. 根据某种理论，这些量之间在一定的近似程度内有关系 $x_0 + x_1\beta_1 + \cdots + x_p\beta_p = 0$. 问题是要依据($x_0$, $\cdots$, x_p)的一些观测值(x_{0i}, $\cdots$, x_{pi}), $1 \leqslant i \leqslant n$, 去估计 β_1, $\cdots$, β_p. 按上述理论关系应有 $x_{0i} + x_{1i}\beta_1 + \cdots + x_{pi}\beta_p = 0$, $1 \leqslant i \leqslant n$. 由于一般总是 $n > p$, 故得到矛盾方程组. 包括 Euler, Laplace 在内的一些大学者都曾致力于研究这个问题，但未能提出合适的解法. 最后是法国学者 Legendre 在 1805 年提出用最小二乘法来解这个问题，得到学界的公认，并在应用上得到推广. 从这段历史可以看出线性模型与最小二乘法这两个主题的密切联系.

接着在 1809 年 Gauss 提出以正态分布作为测量误差的分布，并在这个基础上发展了最小二乘估计的统计理论. 其中包括现在人们所周知的，作为这个方法的基础理论成果——Gauss-Markov 定理. 有了这一发展，才使最小二乘法与统计联系起来. 以上这些发展在统计史上享有极高的评价，被认为是 19 世纪数理统计学的“中心主题”.

但在 19 世纪大部分时间，以上的发展被认为纯属误差论的范围，与统计分析不相干. 当时流行的观点认为误差论处理的是对一个对象多次测量的数据，而统计学处理的是对一个群体中一些不同个体的测量数据，

两者有本质的不同. 但当时也有些学者,主要是 Quetelet,从经验中注意到统计数据也往往很拟合于 Gauss 的正态分布,而主张 Legendre-Gauss 这一套方法也可用于统计数据. 到 19 世纪后期,Galton, Edgeworth, Pearson 和 Yule 等人提出和发展了相关回归分析,引进多元正态分布作为描述多维数据的统计模型,特别是发现了在多维正态中一个变元的条件期望是其他变元的线性函数,从而把相关回归分析与线性模型联系起来,也为最小二乘法在此领域中的应用敞开了大门. 这些发展确立了线性模型-最小二乘法在统计分析尤其是回归分析中的中心地位,意义十分重大,被视为在 Legendre-Gauss 以后线性模型-最小二乘法这个体系发展过程中的又一块里程碑.

下一个重大发展是在 20 世纪初的 20 余年,Student, Fisher 等导出了正态样本的一些重要统计量的精确抽样分布,Fisher 在 20 世纪 20 年代提出方差分析法——这可视为线性模型的离散化,为线性模型-最小二乘法这个体系找到了一种全新的且十分重要的应用,是这个体系发展史上第三块有重大意义的里程碑. 经过这些发展,线性模型-最小二乘法这个体系最终确立了其在统计方法中的中心地位. 虽然在以后,许多更复杂的模型和方法先后被引进到统计学的研究领域中,但从应用的角度看,这个体系的中心地位并未基本动摇.

自这以后,这个体系还没有像上述那样的基本意义的成果,但 70 余年来,在深度和广度上,还是有不少重要的发展,其中一些很有实用意义. 本书的目的就是介绍有关的成果. 由于论题涉及面太广且为作者的知识面和书的篇幅所限,只能着重在作者及其合作者曾涉足因而相对来说比较了解的那些领域进行阐述.

全书 6 章,可分为两大部分:第一部分(前四章)讨论的是样本量 n 固定时的问题;第二部分(后两章)则讨论当 $n \to \infty$ 时估计量的渐近性质问题.

第 1 章讨论线性模型中各种最小二乘估计以及它们之间的关系. 第 2 章和第 3 章分别讨论最小二乘估计的相对效率和各种意义下的优良性. 当模型误差协方差阵含有未知参数时,两步估计是由广义最小二乘估计衍生出的一种常用估计,第 4 章讨论了这种估计的一些重要性质.

后两章讨论的是最小二乘估计的相合性问题,即在何种条件下,当样

本量 $n \to \infty$ 时,最小二乘估计在一定意义下收敛于被估计的参数值.附带也研究了这个问题:在何种条件下线性相合估计或一般相合估计存在或不存在.在不少情况下结论是:若最小二乘估计不为相合,则别无其他的相合估计存在.因此可以说,对这个问题的研究有助于揭示最小二乘估计的深层性质.

第 5 章讨论的是模型中的误差方差有界的情况.在作者之一参与写作的完成于 1985 年的著作[18]中,对这个情况有所讨论.本书收录了其中部分内容,并在本章中主要介绍了 20 世纪 90 年代以来的工作.第 6 章讨论误差只有 r 阶矩 ($1 \leqslant r < 2$) 的情况,是近几年来的成果.

本书的写作,得到上海科学技术出版社和赵序明同志的支持和关心,朱力行教授也给了不少协助.作者愿借这个机会,对上述机构和人士表示衷心的感谢.

作 者

目　录

第 1 章　最小二乘估计

若无特殊声明,本书所讨论的各种量,包括数、向量、矩阵和函数都是实的. 我们用 a 表示列向量,a'表示 a 的转置. 若 $a' = (a_1, \cdots, a_n)$,则 $\| a \| = (a'a)^{1/2} = \left(\sum_{i=1}^{n} a_i^2\right)^{1/2}$, 表示向量 a 的欧氏长度. 用 A, B, C 等表示矩阵,A^- 和 A^+ 分别表示 A 的任一广义逆和 Moore-Penrose 广义逆. $|A|$表示方阵 A 的行列式,$\mu(A)$表示 A 的列向量张成的子空间. $A > 0$ 和 $A \geqslant 0$ 分别表示 A 为对称正定阵和对称半正定阵. 当 A 与 B 为同阶方阵时, $A \geqslant B$ 或$B \leqslant A$ 表示$A \geqslant 0$, $B \geqslant 0$ 且 $A - B \geqslant 0$, 称在半正定意义下 B 小于或等于A.

本章我们讨论线性模型中未知参数的各种最小二乘估计.

§1.1　最小二乘估计

线性模型是描述变量之间线性相关关系的一类统计模型的总称. 采用矩阵符号,可以把它写成如下简洁形式:

$$y = X\beta + e,\ \mathrm{E}(e) = 0,\ \mathrm{COV}(e) = \sigma^2 I_n, \tag{1.1}$$

这里 $y = (y_1, y_2, \cdots, y_n)'$ 是因变量 Y 的 n 次观测值,X 为 $n \times p$ 的设计矩阵,一般它的第 1 列对应于模型的常数项,因而所有元素皆为 1,其余各列分别是自变量的 n 次试验中的取值. $\beta = (\beta_0, \beta_1, \cdots, \beta_{p-1})'$ 为未知的回归参数向量,也常常称为回归系数. $e = (e_1, e_2, \cdots, e_n)'$ 表示 n 次试验的随机误差,其均值为零,协方差阵为 $\sigma^2 I_n$. 假设条件 $\mathrm{COV}(e) = \sigma^2 I_n$, 表明 n 次试验的误差是等方差的,且互不相关,常常把这个条件称为 Gauss-Markov 假设.

在模型(1.1)式中,设计矩阵 X 的秩可以等于它的列数 p,也可以小

于 p,分别称为 X 是列满秩和列降秩的. 线性回归模型多属于前一种情况,而方差分析模型和协方差分析模型都属于后一种情况. 设计矩阵 X 是否为列满秩,对模型(1.1)式的参数估计问题将会产生很大影响.

在一些实际应用中,我们有充分的理由认为误差是彼此相关的或者误差是不等方差的,这时 Gauss-Markov 假设不再成立,一个合理的线性模型应该是:

$$y = X\beta + e,\ \mathrm{E}(e) = 0,\ \mathrm{COV}(e) = \sigma^2 V, \tag{1.2}$$

这里 n 阶方阵 V 可以是正定的,也可以是半正定的,可以是已知的,也可以包含若干未知参数. 当 V 是半正定方阵时,模型(1.2)式称为奇异线性模型. 近 20 余年来,关于这个模型的参数估计理论研究取得了相当深入的结果.

本段讨论模型(1.1)式的参数估计问题. 估计回归系数 β 的基本出发点是最小二乘原理. 这个原理认为,回归系数的真值应该使模型误差 $e = y - X\beta$ 达到最小,这就导致了用最小化 $Q(\beta) = \| e \|^2 = \| y - X\beta \|^2$ 来求 β 的估计,即最小二乘法.

对 $Q(\beta)$关于 β 求导数并令其等于零,得到关于 β 的线性方程组

$$X'X\beta = X'y, \tag{1.3}$$

称为正则方程,或称估计方程. 很明显,这个方程组是相容的. 当设计阵 X 的秩 $\mathrm{rank}(X) = p$ 时,这个方程组有唯一解

$$\hat{\beta} = (X'X)^{-1}X'y, \tag{1.4}$$

它就是回归系数 β 的最小二乘估计(least squares estimate,以下缩写为 LSE).

若 $\mathrm{rank}(X) = r < p$, 此时方程组(1.3)式有无穷多个解,这些解可记为

$$\hat{\beta} = (X'X)^{-}X'y, \tag{1.5}$$

设 A 为任意矩阵,则 A^- 表示矩阵 A 的任一广义逆,即 A^- 为满足 $AA^-A = A$ 的矩阵. 对任一 $c \in \mu(X')$, 称 $c'\beta$ 为可估函数,称 $c'\hat{\beta}$ 为 $c'\beta$ 的 LSE, 它与$\hat{\beta}$表达式中所含的广义逆$(X'X)^-$的选择无关,于是具有唯一性. 容

易证明，$c'\hat{\beta}$ 是 $c'\beta$ 的无偏估计.

设 l 为任一满足条件 $l'X=0$ 的 n 维非零向量，只要 $\operatorname{rank}(X)<n$，这样的 l 总存在而且有无穷多个. 对任一可估函数 $c'\beta$，$c'\hat{\beta}+l'y$ 也是 $c'\beta$ 的无偏估计. 可见 $c'\beta$ 的无偏估计有无穷多个. 在这无穷多个无偏估计中，方差最小者称为最佳线性无偏估计(best linear unbiased estimate，简记为 BLUE). 下面的定理表明，LSE $c'\hat{\beta}$ 就是 BLUE，它奠定了 LSE 在线性模型估计理论中的中心地位.

定理 1.1 (Gauss-Markov)对于线性模型(1.1)式和任一可估函数 $c'\beta$，$c'\hat{\beta}$ 为 $c'\beta$ 的唯一 BLUE.

假设 W 为 $m\times p$ 矩阵，$\mu(W')\subset\mu(X')$，$\operatorname{rank}(W)=m$，于是 $W\beta$ 为 m 个线性无关的可估函数. 如果 Ay 为 $W\beta$ 的任一无偏估计，那么从 Gauss-Markov 定理可以证明

$$\operatorname{COV}(W\hat{\beta})\leqslant\operatorname{COV}(Ay),$$

这就是说，在 $W\beta$ 的所有无偏估计中，LSE $W\hat{\beta}$ 具有最小的协方差阵，这里“最小”是在半正定意义下而言的，以后我们也称 $W\hat{\beta}$ 为 $W\beta$ 的 BLUE. 特别取 $W=X$，$\mu=X\beta$ 就是观测向量 y 的均值向量，$\hat{\mu}=X\hat{\beta}$ 就是 μ 的 BLUE. 如果 $\operatorname{rank}(X)=p$，我们可以取 $W=I$，这时 LSE $\hat{\beta}$ 就是 β 的 BLUE.

记
$$\hat{e}=y-X\hat{\beta}=(I-H)y, \tag{1.6}$$

称 $\hat{e}$ 为残差向量，这里 $H=X(X'X)^{-}X'$. 残差向量 $\hat{e}$ 作为 e 的一个“估计”，在回归诊断中起着重要作用. 基于 $\hat{e}$，我们可以构造模型误差方差 σ^2 的如下估计：

$$\hat{\sigma}^2=\frac{\|\hat{e}\|^2}{n-r}=\frac{\|y-X\hat{\beta}\|^2}{n-r}, \tag{1.7}$$

这里 $r=\operatorname{rank}(X)$. 为简单计，常常也称 $\hat{\sigma}^2$ 为 σ^2 的 LSE. 它是 σ^2 的无偏估计.

如果在模型(1.1)式中，误差服从正态分布，那么 LSE 还有许多优良的统计性质.

定理 1.2 对模型(1.1)式，设 $e\sim N(0,\sigma^2 I_n)$，$c'\beta$ 为任一可估函数.

则

(1) $c'\hat{\beta}$ 为 $c'\beta$ 的极大似然估计，也是唯一的最小方差无偏估计，且 $c'\hat{\beta}\sim N(c'\beta,\ \sigma^2 c'(X'X)^- c)$.

(2) $\dfrac{(n-r)\,\hat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-r}$，且 $\hat{\sigma}^2$ 为 σ^2 的唯一最小方差无偏估计.

(3) $c'\hat{\beta}$ 与 $\hat{\sigma}^2$ 相互独立.

以下讨论线性模型(1.2)式的参数估计问题. 假设 $V>0$ 且已知. 用 $V^{-1/2}$ 左乘(1.2)式得到一个新线性模型，它的误差协方差阵为 $\sigma^2 I_n$. 对这个模型应用最小二乘法，就归结为通过最小化

$$Q(\beta)=(y-X\beta)'V^{-1}(y-X\beta) \tag{1.8}$$

来求 β 的估计. 这就导致了下面的正则方程，也称估计方程：

$$X'V^{-1}X\beta=X'V^{-1}y. \tag{1.9}$$

当 $\mathrm{rank}(X)=p$ 时，上述方程的唯一解为

$$\beta^*=(X'V^{-1}X)^{-1}X'V^{-1}y, \tag{1.10}$$

称为 β 的广义最小二乘估计(generalized least squares estimate，简记为 GLSE). 若 $\mathrm{rank}(X)=r<p$ 时，(1.9)式的解可表为

$$\beta^*=(X'V^{-1}X)^- X'V^{-1}y. \tag{1.11}$$

对于任意可估函数 $c'\beta$，即 $c\in\mu(X')$，称 $c'\beta^*$ 为 $c'\beta$ 的 GLSE，并且可以证明，它是 $c'\beta$ 的唯一 BLUE.

设 $W\beta$ 为 m 个线性无关的可估函数，Ay 为它的任一无偏估计，则有

$$\mathrm{COV}(W\beta^*)\leqslant\mathrm{COV}(Ay),$$

这就是说，在 $W\beta$ 的所有无偏估计中，GLSE $W\beta^*$ 具有最小的协方差阵，因而它是 $W\beta$ 的 BLUE. 进一步，对 y 的均值向量 $\mu=X\beta$，$\mu^*=X\beta^*$ 就是 μ 的 BLUE. 特别地，当 $\mathrm{rank}(X)=p$ 时，β 是可估的，对 β 的任一无偏估计 Ay，这时 A 满足 $AX=I_n$，总有

$$\mathrm{COV}(\beta^*)\leqslant\mathrm{COV}(Ay). \tag{1.12}$$

需要特别注意的是，对于线性模型(1.2)式，如果我们忽略掉 $\mathrm{COV}(e)=$

$\sigma^2 V \neq \sigma^2 I_n$ 这个事实，而认为 $\mathrm{COV}(e) = \sigma^2 I_n$，应用最小二乘法就得到 LSE $\hat{\beta} = (X'X)^{-1}X'y$. 在线性模型(1.2)式中，$\hat{\beta}$ 也是 β 的一个无偏估计，于是应用(1.12)式我们得到

$$\mathrm{COV}(\beta^*) \leqslant \mathrm{COV}(\hat{\beta}), \tag{1.13}$$

这就是说，对于线性模型(1.2)式，当 $\mathrm{rank}(X) = p$ 时，虽然 LSE $\hat{\beta}$ 和 GLSE β^* 都是 β 的无偏估计，但后者优于前者.

类似于模型(1.1)式，称 $e^* = y - X\beta^*$ 为残差向量，基于 e^*，可构造 σ^2 的如下无偏估计：

$$\sigma^{*2} = \frac{\| e^* \|^2}{n-r} = \frac{\| y - X\beta^* \|^2}{n-r}. \tag{1.14}$$

对于 $c'\beta^*$ 和 σ^{*2}，类似于定理 1.1 和定理 1.2 的结论都成立.

§1.2 最小二乘统一理论

在线性模型

$$y = X\beta + e,\ \mathrm{E}(e) = 0,\ \mathrm{COV}(e) = \sigma^2 V \tag{1.15}$$

中，如果 $|V| = 0$，则称该模型为奇异线性模型. 对这样的模型，因为 V^{-1} 不存在，所以我们不能够通过最小化(1.8)式所定义的 $Q(\beta)$ 来求 β 的最小二乘估计. 20 世纪 60 年代以来，许多统计学家研究了这种模型的参数估计，提出了几种估计方法. 在这些估计方法中，著名统计学家 Rao 应用推广的最小二乘法所导出的估计以其形式简单便于理论研究而得到普遍采用. 本节的目的是讨论这个方法.

对于奇异线性模型，因为 V^{-1} 不存在，于是(1.8)式的 $Q(\beta)$ 无定义. 如果用任一广义逆 V^- 代替 V^{-1}，把 $Q(\beta)$ 定义为 $Q(\beta) = (y - X\beta)'V^-(y - X\beta)$，因为这样的 $Q(\beta)$ 与所含的广义逆 V^- 有关，取不同的广义逆得到不同的 $Q(\beta)$，因而(1.8)式失去意义. 于是对于奇异线性模型，一个核心的问题是寻找一个新矩阵 T，它能够充当(1.8)式中 V^{-1} 所担负的作用. Rao[70] 成功地解决了这个问题. 他定义

$$T = V + XUX',\text{其中 } U \geqslant 0,\ \mathrm{rank}(T) = \mathrm{rank}(V \vdots X), \tag{1.16}$$

然后定义

$$Q(\beta) = (y - X\beta)'T^{-}(y - X\beta). \tag{1.17}$$

用最小化 $Q(\beta)$ 求出最小值点

$$\beta^{*} = (X'T^{-}X)^{-}X'T^{-}y, \tag{1.18}$$

后面我们将证明,对任一可估函数 $c'\beta$, $c'\beta^{*}$ 为其 BLUE. 这个结论既适用于设计阵 X 列满秩或列降秩的情形,又适用于 V 奇异或非奇异的情形. 正是由于这个原因,通常把这个结果称为最小二乘统一理论,见文献[70, 79].

在 T 的定义中,包含一个可以选择的半正定阵 U. 事实上满足条件的方阵 U 是很多的. 例如,一个最简单的选择是 $U = I_n$, 这是因为等式

$$\text{rank}(V + XX') = \text{rank}(V \vdots X)$$

对一切 $V \geqslant 0$ 和 X 都成立. 另外,当 $V > 0$ 时,可取 $U = 0$,此时 $T = V$, (1.18)式就变成了(1.11)式.

为了证明 $c'\beta^{*}$ 为 $c'\beta$ 的 BLUE,先证明几个预备事实.

引理 1.1 对于线性模型(1.15)式,不管 $V > 0$ 或 $V \geqslant 0$, $y \in \mu(V \vdots X)$ 总是成立.

证明 将 V 分解为 $V = LL'$, 这里 L 为 $n \times t$ 矩阵, $t = \text{rank}(V) = \text{rank}(L)$. 记 $e = L\varepsilon$, $\text{E}(\varepsilon) = 0$, $\text{COV}(\varepsilon) = \sigma^2 I_n$, 则 y 可表为如下新线性模型的形式:

$$y = X\beta + L\varepsilon,\ \text{E}(\varepsilon) = 0,\ \text{COV}(\varepsilon) = \sigma^2 I_n,$$

于是 $y \in \mu(X \vdots L)$. 再利用 $\mu(L) = \mu(LL') = \mu(V)$, 结论得证. ▌

引理 1.2 对(1.16)式所定义的 T,总有

(1) $\mu(T) = \mu(V \vdots X)$.

(2) $X'T^{-}X$, $X'T^{-}y$ 和 $(y - X\beta)'T^{-}(y - X\beta)$ 都与广义逆 T^{-} 的选择无关.

证明 (1)是(1.16)式的直接推论. 因为 $y \in \mu(T)$, $\mu(X) \subset \mu(T)$, $y - X\beta \in \mu(T)$, 再利用事实:若 $\mu(A) \subset \mu(B)$, 则 $A'B^{-}A$ 与 B^{-} 的选择无关,便可证得(2). 引理证毕. ▌