XIANXING MOXING DE YUCE LILUN JIQI YINGYONG

线性模型的预测理论及其应用

徐礼文 著

中国水利水电出版社
www.waterpub.com.cn

线性模型的预测理论及其应用

徐礼文　著

中国水利水电出版社
www.waterpub.com.cn

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预测这一领域近些年的研究成果，以及相关的最新进展．全书共分 6 章：第 1 章通过实例引进线性模型预测问题和预备知识；从第 2 章起，系统讨论线性模型预测的基本理论、方法及其应用，包括未来观察值的最优线性无偏预测，稳健性预测，可容许性预测，极小、极大预测和混合效应模型中的预测以及它们之间的关系．

本书主要读者对象是数理统计理论与方法研究的专家，数学、工程、经济、金融等领域的科研人员和实际工作者．本书也可用作高等院校数理统计专业本科生、研究生的教材或参考书．

图书在版编目（CIP）数据

线性模型的预测理论及其应用 / 徐礼文著．-- 北京
：中国水利水电出版社，2011.6
ISBN 978-7-5084-8695-6

Ⅰ．①线… Ⅱ．①徐… Ⅲ．①线性模型—数学预测
Ⅳ．①O212

中国版本图书馆CIP数据核字(2011)第110231号

书　　名	**线性模型的预测理论及其应用**
作　　者	徐礼文　著
出版发行	中国水利水电出版社
	(北京市海淀区玉渊潭南路 1 号 D 座　100038)
	网址：www.waterpub.com.cn
	E-mail：sales@waterpub.com.cn
	电话：(010) 68367658（营销中心）
经　　售	北京科水图书销售中心（零售）
	电话：(010) 88383994、63202643
	全国各地新华书店和相关出版物销售网点
排　　版	中国水利水电出版社微机排版中心
印　　刷	北京瑞斯通印务发展有限公司
规　　格	184mm×260mm　16 开本　9.5 印张　225 千字
版　　次	2011 年 6 月第 1 版　2011 年 6 月第 1 次印刷
印　　数	0001—2000 册
定　　价	**26.00** 元

前　言

线性模型是在现代统计学理论方法中得到广泛应用的一个重要分支，也是其他现代统计学分支的重要基础和新思想的源泉．线性模型中的统计推断理论和方法，在生物医学、经济管理、金融工程、工业农业和工程技术等众多领域的应用获得了长足发展．线性模型预测理论作为线性模型统计推断理论的重要分支，以线性模型估计和检验等理论为基础，具有重要的实用价值，一直受到统计学家的重视．

从历史上来看，预测理论一直伴随着统计学的发展．通过一定建模方法，得到经验模型后，就存在使用估计的模型进行预测和控制的问题．预测理论的发展，有相当一部分借助于估计理论中的概念与方法，但也有自成一体的结构体系和问题驱动．经过长期的发展，在深度和广度上已有了很多重要进步，其中包括了很多有实用意义的理论、方法和技巧．在已有的结论中，多数是在模型的设计矩阵和协方差矩阵满秩要求下得出的．本书的目的是介绍在任意秩有限总体情形下各种预测理论和方法的一般结论，这也是本书的突出特色．因为论题涉及面较广，且为作者的知识和本书的篇幅所限，只能着重在作者曾涉足且相对比较了解的领域来论述．

全书共有 6 章，分为两部分：第一部分（前 5 章）主要是在超总体观点下，讨论了有限总体中几种优良性准则下未来观察值的预测问题；第二部分（第 6 章）则讨论线性混合效应模型中固定效应和随机效应线性组合的预测问题．

第 1 章通过实例引进和比较各种线性模型，介绍模型的设计背景和本书常用的矩阵论方面的知识．第 2 章分别讨论了一元和多元有限总体中线性可预测量的最优线性无偏预测．对于一元有限总体，考虑两种不同预测风险函数：二次损失和矩阵损失函数；另外，还考虑了线性预测充分性问题．对多元有限总体，考虑了矩阵迹等意义下的最优预测问题．在真实模型的协方差阵扰动情形下，线性可预测量的最优线性无偏预测能否保持其优良性呢？针对这一问题，第 3 章讨论了简单投影预测的最优性及最优线性无偏预测的稳健性．第 4 章系统地研究了在一般 Gauss - Markov 模型、随机系数模型、带有不等

式约束模型和多元线性模型中线性预测的可容许性．第5章则考虑线性预测的Minimax性．第6章，转入线性混合效应模型中固定和随机效应的线性组合的预测问题．

借本书出版之际，我要向我的三位导师——王松桂教授、陆璇教授和喻胜华教授表示衷心的感谢．三位导师在我攻读硕士、博士学位和博士后研究工作期间，一直都给予我亲切的关怀、悉心的指导和热情的帮助．导师们渊博的学识和杰出的工作引导我进入了统计学研究领域．特别是喻胜华教授，在我硕士学习阶段之初，他就指导我从事线性模型预测的研究．在本书写作期间，爱妻袁伶俐给予了我全力的支持和积极的鼓励，她在完成自己工作的同时，承担了生活的全部责任，在此衷心的表达我对她的感激与谢意．

本书的顺利出版还要感谢国家自然科学基金（多元复杂数据的混合效应模型理论及其应用，No. 1112008)、中国博士后科学基金（No. 20070410544）和北京市自然科学基金（No. 1112008）的资助和中国水利水电出版社的良好合作与支持，作者愿借此机会向上述机构和人士表示诚挚的谢意．

徐礼文

2011年1月于北京

北方工业大学

符 号 表

$\triangleq$	“定义为”或“记为”
$A\geqslant 0$	A 为对称半正定方阵
$A>0$	A 为对称正定方阵
$A\geqslant B$	$A\geqslant 0$，$B\geqslant 0$ 且 $A-B\geqslant 0$
A^-	矩阵 A 的任一广义逆
A^+	矩阵 A 的 Moore - Penrose 广义逆
$A^\perp$	满足 $A'A^\perp=0$ 且具有最大秩的矩阵
$\mathrm{rk}(A)$	矩阵 A 的秩
$\mid A \mid$	矩阵 A 的行列式
$\Vert A \Vert$	矩阵 A 的范数
$\mathrm{tr}(A)$	方阵 A 的迹
$\mathscr{R}(A)$	矩阵 A 的列向量张成的子空间
$\mathscr{N}(A)$	矩阵 A 的零空间
P_A	向 $\mathscr{R}(A)$ 的正交投影变换阵
N_A	$N_A=I-P_A$
$\mathbf{1}=(1, \cdots, 1)'$	分量皆为 1 的列向量
$\mathrm{Vec}(A)$	将 A 的列向量依次排成的列向量
$A\otimes B$	A 与 B 的 Kronecker 乘积
$E(X)$	随机变量或向量 X 的均值
$\mathrm{Var}(X)$	随机变量 X 的方差
$\mathrm{Cov}(X, Y)$	随机变量或向量 X，Y 的协方差
$u\sim(\mu, \Sigma)$	均值为 μ，协方差阵为 Σ 的随机向量
$u\sim N_p(\mu, \Sigma)$	均值为 μ，协方差阵为 Σ 的 p 维正态向量

目　　录

第1章 模型及预备知识

线性模型是一类统计模型的总称，它包括了线性回归模型，方差分析模型，协方差分析模型，线性混合效应模型，纵向数据模型，生长曲线模型和半相依回归模型等. 许多生物、医学、经济、管理、地质、气象、农业、工业、工程技术等领域的现象都可以用线性模型来近似描述. 而非线性模型中有一部分可以通过构造合适的函数化为线性模型. 另一方面，非参数和半参数统计模型中许多推断的思想方法是在局部线性化，再利用线性模型的基础理论后得出结果. 因此，线性模型成为现代统计学中应用最为广泛的模型之一. 本书在介绍线性模型预测的基本理论与方法的基础上，论述作者及其合作者线性模型预测这一领域近些年的研究工作，以及相关的最新发展.

本章将先通过不同背景引入线性模型预测的问题，使读者对问题的丰富背景有足够的了解，有助于对后面引进的观点、统计概念和理论方法的理解. 另外，在本章还将提供以后常用到的有关矩阵论的预备知识.

1.1 模型简介

线性模型的一般形式为

$$y=X\beta+\varepsilon,\quad E(\varepsilon)=0,\quad \mathrm{Cov}(\varepsilon)=\Sigma \tag{1.1}$$

式中：y 为 $n\times1$ 的观测向量；X 为 $n\times p$ 的已知设计阵；β 为 $p\times1$ 的未知参数向量；ε 为随机误差向量.

式（1.1）中，$\mathrm{Cov}(y)=\Sigma>0$ 或 $\Sigma\geqslant0$（根据具体情形而定）. 在经济和社会科学等领域中的许多模型可以表示成式（1.1）的形式，可以参见 Searle（1971），王松桂（1987），Christensen（1987），Wang 和 Chow（1994），Rao 和 Toutenburg（1995）.

若参数向量 β 中既包含固定效应也包含随机效应，则式（1.1）表示的模型称为线性混合效应模型，其一般形式为

$$y=X\beta+Z\xi+\varepsilon \tag{1.2}$$

式中：y 为 $n\times1$ 的观测向量；β 为 $p\times1$ 的非随机参数向量，称为固定效应；X 为对应于固定效应的设计阵；ξ 为 $q\times1$ 的随机向量，称为随机效应；Z 为对应于随机效应的设计阵；ε 为随机误差向量.

一般我们假设 $E(\xi)=0$，$E(\varepsilon)=0$，ξ 和 ε 互不相关且

$$\mathrm{Cov}(\xi)=G,\quad \mathrm{Cov}(\varepsilon)=R$$

于是 $\mathrm{Cov}(y)=\Sigma=ZGZ'+R$，这里 G 和 R 分别为已知或未知的非负定和正定阵，在它们未知时，它们可以依赖于一个未知参数向量 θ. 即此时式（1.2）表示的模型的随机部分 $Z\xi+\varepsilon$ 可以分解为 $Z\xi+\varepsilon=U_1\xi_1+U_2\xi_2+\cdots+U_k\xi_k$，则得到一般的方差分量模型为

$$y = X\beta + U_1\xi_1 + U_2\xi_2 + \cdots + U_k\xi_k \tag{1.3}$$

式中：ξ_i 为 $q_i \times 1$ 的随机效应向量；U_i 为 $n \times q_i$ 的已知设计阵.

假设

$$E(\xi_i) = 0, \quad \mathrm{Cov}(\xi_i) = \sigma_i^2 I_{q_i}, \quad \mathrm{Cov}(\xi_i, \xi_j) = 0, \quad i \neq j$$

于是

$$E(y) = X\beta, \quad \mathrm{Cov}(y) = \sum_{i=1}^{k} \sigma_i^2 U_i U_i'$$

其中，σ_i^2 称为方差分量，最后一个随机效应向量 ξ_k 是通常的随机误差向量 ε，而 $U_k = I_n$. 式（1.3）表示的模型包含多类具有广泛应用背景的线性模型，如 Panel 数据模型、单向（两向，多向）分类混合模型、套分类混合模型，参见 Rao（1967，1973），Searle（1987），Rao 和 Kleffe（1988），Searle，Casella 和 McCulloch（1992），Diggle，Liang 和 Zeger（1994），Khuri 和 Mathew（1998），Verbeke 和 Molenberghs（2000）以及 Baitagi（2005）等.

当式（1.3）表示的模型中固定效应只有常数项，即 $X\beta = 1_n u$，这里 1_n 表示元素皆为 1 的 n 维列向量，从上下文知道其维数时，下标 n 可略去，u 是未知的总体均值，此时，式（1.3）表示的模型又常常被称为随机模型.

另一方面，若式（1.1）中的未知参数向量是随机回归系数，即 β 的各分量看作是来自于相应总体的随机抽样，可得如下模型

$$y = X\beta + e, \quad E(\beta) = A\alpha, \quad \mathrm{Cov}(\beta) = \sigma^2 V, \quad E(\beta e') = 0 \tag{1.4}$$

式中：X、A 分别为已知的 $n \times p$ 和 $p \times k$ 的矩阵；V 为已知的 $p \times p$ 的对称非负定矩阵.

当回归系数随时间、个体、单元、区域等变化而发生变化时，随机回归系数模型是很好的选择（见 Burnett，1970；Swamy，1971；Rosenberg，1972；Sarris，1973）.

当问题的因变量不是一个而是多个时，这就导致多因变量的线性模型，如多元线性回归模型、生长曲线模型等. 其中生长曲线模型是一种在生物医学和社会科学等领域被广泛用来处理重复测量数据的多元模型. 它的一般形式为

$$Y = X_1 B X_2 + \varepsilon, \quad E(\varepsilon) = 0, \quad \mathrm{Cov}[\mathrm{Vec}(\varepsilon)] = V_2 \otimes V_1 \tag{1.5}$$

式中：Y 为 $n \times q$ 的随机观测矩阵；X_1、X_2 分别为 $n \times p$ 和 $q \times k$ 的已知矩阵；B 为未知的参数矩阵；ε 为随机误差矩阵；$V_2 \otimes V_1$ 为 Vec（ε）的协方差矩阵，其中 $V_2 \otimes V_1$ 表示 V_2 和 V_1 的 Kronecker 乘积，Vec（ε）表示将 ε 的列向量依次排成的列向量.

【例 1.1】 Panel 数据模型

这类模型被广泛用于计量经济学中（见 Baltagi，2005）. 假设对 N 个个体（如个人、家庭、公司、城市或国家）进行了 T 个时刻的观测，观测数据可表示为

$$y_{it} = x_{it}\beta + \xi_i + \varepsilon_{it}, i = 1, \cdots, N, t = 1, \cdots, T \tag{1.6}$$

式中：y_{it} 为第 i 个个体第 t 个时刻的某项经济指标；x_{it} 为 $p \times 1$ 的已知向量，它刻画了第 i 个个体在第 t 个时刻的一些自身特征；ξ_i 为 i 个个体的个体效应；ε_{it} 为随机误差项.

若要研究整个市场的运行规律，而不是关心这特定的 N 个个体，这 N 个个体只不过是从总体中抽取的随机样本，这时个体效应就是随机的. 假定所有 ξ_i 和 ε_{it} 互相独立，且 $\xi_i \sim N(0, \sigma_\xi{}^2)$，$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. 记

$$y=(y_{11},\cdots,y_{1T},y_{21},\cdots,y_{NT})',\quad X=(x_{11},\cdots,x_{1T},x_{21},\cdots,x_{NT})'$$
$$\xi=(\xi_1,\cdots,\xi_N)',\varepsilon=(\varepsilon_{11},\cdots,\varepsilon_{1T},\varepsilon_{21},\cdots,\varepsilon_{NT})',U_1=I_N\otimes 1_T$$

则式（1.6）表示的模型可以表示成

$$y=X\beta+U_1\xi+\varepsilon$$

由上面的假设可知

$$\mathrm{Cov}(y)=\sigma_\xi^2U_1U_1'+\sigma_\varepsilon^2I_{NT}$$

这里 σ_ξ^2 和 σ_ε^2 即为方差分量．式（1.6）表示的模型有时也称为纵向数据（longitudinal data）模型，常用于生物医药统计的研究领域中．

在上述问题中，若把时间效应也考虑进来，则式（1.6）表示的模型可以改写为

$$y_{it}=x_{it}'\beta+\xi_i+\eta_t+\varepsilon_{it},\quad i=1,\cdots,N,t=1,\cdots,T$$

其中，η_t 表示与第 t 个时刻相联系的时间效应，若把其看成随机的，并且假设 $\mathrm{Var}(\eta_t)=\sigma_\eta^2$，$\eta_t$ 与所有的 ξ_i 和 ε_{it} 相互独立，记 $U_2=1_N\otimes I_T$，$\eta=(\eta_1,\ \cdots,\ \eta_T)'$，则可得如下模型

$$y=X\beta+U_1\xi+U_2\eta+\varepsilon$$

此时，观测向量的协方差阵为

$$\mathrm{Cov}(y)=\sigma_\xi^2U_1U_1'+\sigma_\eta^2U_2U_2'+\sigma_\varepsilon^2I_{NT}$$

其中，σ_ξ^2，σ_η^2 和 σ_ε^2 为方差分量．

【例 1.2】 生长曲线（Growth - curve）模型

生物学家欲研究白鼠的某个特征随时间变化情况，随机选用 n 只小白鼠做试验．在时刻 t_1，…，t_p 对每只小白鼠观察该特征的值．设第 i 只小白鼠的 p 次观测值为 y_{i1}，…，y_{ip}，$i=1$，…，n．假定不同白鼠的观测值是不相关的，而同一只白鼠的观测却是相关的，且其协方差阵为 V．从理论上分析认为，这些观测值与观测时间 t 的关系为 $k-1$ 阶的多项式

$$Y=f(t)=\beta_0+\beta_1t+\cdots+\beta_{k-1}t^{k-1}\tag{1.7}$$

这就是所谓的理论生长曲线．生物学家的目的是估计 β_0，β_1，…，β_{k-1}，以得到经验生长曲线．若以 ε_{ij} 记为 y_{ij} 所含的误差，则对观测数据 y_{ij}，有模型

$$\begin{pmatrix}y_{11}&y_{12}&\cdots&y_{1p}\\y_{21}&y_{22}&\cdots&y_{2p}\\\vdots&\vdots&\vdots&\vdots\\y_{n1}&y_{n2}&\cdots&y_{np}\end{pmatrix}=\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}(\beta_0,\beta_1,\cdots,\beta_{k-1})\begin{pmatrix}1&1&\cdots&1\\t_1&t_2&\cdots&t_p\\\vdots&\vdots&\vdots&\vdots\\t_1^{k-1}&t_2^{k-1}&\cdots&t_p^{k-1}\end{pmatrix}+\varepsilon_{ij}$$

它具有式（1.5）的形式，且 $\varepsilon=(\varepsilon_{ij})$ 也满足所做的假设．

1.2　预测的有关问题和研究进展

关于线性模型的预测问题，我们主要感兴趣的有：有限总体中的预测，利用历史数据对未来观察值的预测，固定与随机效应线性组合的预测．关于以上问题，已经有了较为丰富的研究内容，下面将对此作简要的介绍．

1.2.1　有限总体中的预测

抽样调查中的统计推断通常在两种观点下进行：一种是基于抽样方案（设计）的观

点；另一种是基于超总体模型的观点．二者主要的区别在于是否假定有限总体是来自于另一个超总体的样本．本书中，除特别说明外，均指在超总体观点下进行．

假设 $\mathscr{P}$ 是一个含有 n 个个体的有限总体．在经典的有限总体抽样理论中，考虑的是利用与这 n 个个体相联系的 n 个变量的值对总体的某个函数做出统计推断．根据某个给定的随机程序（抽样设计），此有限总体容量为 s 的一个样本记为 $\boldsymbol{s}$．那么，上述统计推断的基础就是这个样本 $\boldsymbol{s}$．抽样结果 $\boldsymbol{s}$ 出现的概率 p_s 由抽样设计而定．例如，在无放回的简单随机抽样情形下，$p_s=1/C_n^s$．

与之相比，超总体观点把感兴趣的有限总体看成来自另外一个无限总体的容量为 n 的样本，而把产生容量为 s 的样本的随机程序看成由以下两步组成．

第一步：从一个无穷超总体中抽取容量为 n 的“大的样本”．

第二步：从由第一步得到的“大的样本”中抽取容量为 $s(s<n)$ 的样本．

事实上，第一步是一个假想的步骤，往往假设得到的样本元素是独立同分布．第二步是由调查者根据其具体的设计 p_s 执行的“真实的抽样调查”．超总体理论就是考虑包含第一步和第二步的两步随机程序的重复实施．因此，在超总体观点下，有限总体抽样则可以看作是以给定第一步产生的具体结果的情况下的条件分布为基础的．

设 $\mathscr{P}=\{1，2，\cdots，n\}$ 表示容量为 n 的有限总体，$\mathscr{P}$ 的第 k 个元对应着 $p+1$ 个标量：$y_{(k)}$，x_{k1}，x_{k2}，$\cdots$，x_{kp}，这里除 $y_{(k)}$ 外都是已知的，$k=1，2，\cdots，n$．记 $y=(y_{(1)}，y_{(2)}，\cdots，y_{(n)})'$，$X=(X_{(1)}，X_{(2)}，\cdots，X_{(n)})'$，其中 $X_{(k)}=(x_{k1}，x_{k2}，\cdots，x_{kp})'$，$k=1，2，\cdots，n$．考虑这些变量满足如下线性模型

$$y = X\beta+\varepsilon,\quad E(\varepsilon)=0,\quad \mathrm{Cov}(\varepsilon)=\sigma^2 V \tag{1.8}$$

式中：ε 为 n 维随机误差向量；X 为 $n\times p$ 的已知矩阵；β 为 p 维未知参数向量；V 为已知的对称非负定矩阵，记为 $V\geqslant 0$；σ^2 是未知参数，$\sigma^2>0$．

形式上，这个模型与一般的线性模型并无两样，但两者有本质上的不同．这里的总体是有限的，且向量 y 是部分被观测的，因而在抽样调查的文献中称其为超总体（Super-population）模型（见 Cassel，1977）．

根据某个确定的抽样方案，选取容量为 s 的样本 $\boldsymbol{s}$．记 $\boldsymbol{r}=\mathscr{P}-\boldsymbol{s}$ 表示容量为 r 的未观测部分，这里，$s+r=n$．样本 $\boldsymbol{s}$ 被选取后，可以重新安排 y 的元素，使得 y，X 和 V 有一致的分块，即

$$y=\begin{pmatrix} y_s \\ y_r \end{pmatrix},\quad X=\begin{pmatrix} X_s \\ X_r \end{pmatrix},\quad V=\begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix}$$

在抽样调查理论和应用研究中，我们最感兴趣的问题是对 y 的函数 $\theta(y)$ 进行有效的预测．这样的函数有总体的线性函数 Qy 和总体的二次函数 $y'Hy$ 等，这里 Q 和 H 分别是 $h\times n$ 和 $n\times n$ 的已知矩阵．具体而言，比如要预测总体总量 $T=\sum\limits_{i=1}^{n} y_{(i)}$，有限总体回归系数 $\beta_n=(X'V^{-1}X)^{-1}X'V^{-1}y$ 和总体方差 $S_n^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(y_{(i)}-\bar{y}_n)^2=y'\left(I-\dfrac{1}{n}11'\right)y$ 等，这里假设 X 和 V 为列满秩的，且 $\bar{y}_n=\dfrac{1}{n}T$ 为总体均值．

根据上面的论述，似乎还看不出有限总体抽样调查理论与超总体模型的联系．事实上，可以对前面提到的产生样本的两步作进一步分析．显然，第一步的基本问题就是寻找最优的预测方法；第二步的问题是基于得到的最优预测，在某些优良性准则下实施最优抽样方案的选择．下面通过 Royall 和 Herson（1973）中的一个例子来说明这一点．

考虑超总体模型

$$y_i = x_i\beta + e_i x_i^{1/2}, i = 1,\cdots,N \tag{1.9}$$

其中，β 为一标量，$e=(e_1, \cdots, e_N)'$，$E(e)=0$，$\text{Cov}(e)=\sigma^2 \boldsymbol{I}$.

为预测总体总量 $T=\sum_{i=1}^{N} y_i$，选择容量为 s 的一个样本 $\boldsymbol{s}$．设 $\boldsymbol{r}=\mathscr{P}-\boldsymbol{s}$ 表示总体 $\mathscr{P}$ 的容量为 r 的未观测部分，其中 $s+r=N$．对于给定的容量为 s 的一个样本 $\boldsymbol{s}$ 和式（1.9）表示的模型，标准的比率预测是最优的，即

$$\hat{T} = \sum_s y_i + \left(\sum_s y_i \Big/ \sum_s x_i\right)\sum_r x_i = \left(\sum_{i=1}^{N} x_i \Big/ \sum_s x_i\right)\sum_s y_i$$

预测的均方误差为

$$E(\hat{T}-T)^2 = \sigma^2\left(\sum_r x_i \Big/ \sum_s x_i\right)\sum_{i=1}^{N} x_i \tag{1.10}$$

由式（1.10）易见，使得 $\sum_s x_i$ 达到最大值的抽样为最优抽样，其为 s 个抽样单元对应的 x_i 值是最大的抽样．显然，上述的有限总体抽样问题依赖于最优比率预测和抽样方案 $\boldsymbol{s}$.

1.2.2 未来观察值的预测

所谓未来观察值的预测，就是在一般线性模型中，对指定的自变量的值，预测对应的因变量所可能取的值．在线性模型中，自变量往往代表一组试验条件或生产条件或社会经济条件，由于试验和生产等方面的耗费或经济周期长的原因，在根据以往记录的数据得到经验模型后，希望对感兴趣的试验和生产条件不真正去做试验，而利用经验模型关于对应的因变量的取值作出合理的估计和分析．可见，未来观察值的预测是实际当中普遍存在且很有意义的问题．

假设历史数据满足如下线性模型

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, \quad \text{Cov}(\varepsilon) = \sigma^2 V \tag{1.11}$$

式中：y 为 n 维观测向量；ε 为 n 维随机误差向量；X 为 $n\times p$ 的已知矩阵；V 为已知的对称非负定矩阵．

例如，想要预测 m 个点 $x_{0i}=(x_{0i1}, \cdots, x_{0ip})$ 所对应的因变量 y_{0i} 的值，$i=1, \cdots, m$，且已知 y_{0i} 和历史数据服从同一个线性模型，即

$$y_{0i} = x'_{0i}\beta + \varepsilon_{0i}, i = 1,\cdots,m$$

采用矩阵形式，这个模型变为

$$y_0 = X_0\beta + \varepsilon_0, \quad E(\varepsilon_0) = 0, \quad \text{Cov}(\varepsilon_0) = \sigma^2 V_0 \tag{1.12}$$

其中，$y_0=(y_{01}, \cdots, y_{0n})'$，$X_0=(x_{01}, \cdots, x_{0m})'$，$\varepsilon_0=(\varepsilon_{01}, \cdots, \varepsilon_{0m})'$．$y$ 和 y_0 之间的相关性可以用 $\text{Cov}(\varepsilon, \varepsilon_0)=\sigma^2 W$ 来度量．此时有

$$\text{Cov}\begin{pmatrix} y \\ y_0 \end{pmatrix} = \sigma^2\begin{pmatrix} V & W \\ W' & V_0 \end{pmatrix}$$

从以上可以看出，未来观察值的预测问题可以化为与在超总体观点下有限总体中的预测相同的框架下来讨论. 因此，以后的章节就只讨论超总体观点下有限总体的预测问题.

1.2.3　固定和随机效应组合的预测

在线性混合效应模型中，只对固定效应部分估计的研究文献相对较少. 而在许多的实际应用中，经常要考虑固定效应和随机效应线性组合的预测问题. 例如，质量指标的估计、纵向数据研究、植物育种试验和小域（small area）估计问题（见 Robinson，1991）. 记 $\mu=l'\beta+m'\xi$，这里 l 和 m 是已知的常数向量. 在式（1.3）表示的模型中，Henderson（1975）得到了 μ 的最佳线性无偏预测（BLUP），即

$$\begin{aligned} t(\sigma^2) &= t(\sigma^2,\ y) \\ &= l'\tilde{\beta}+m'GU'\Sigma^{-1}(y-X\tilde{\beta}) \end{aligned} \tag{1.13}$$

其中，$\tilde{\beta}=(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ 是 β 的最佳线性无偏估计（BLUE）；$\sigma^2=(\sigma_1^2,\ \sigma_2^2,\ \cdots,\ \sigma_k^2)'$ 为方差分量向量. $t(\sigma^2)$ 称为 μ 的无偏预测，如果对一切参数，有 $E[t(\sigma^2,\ y)-\mu]=0$，μ 称为可预测变量. 如果 $l'\beta$ 是一个线性可估函数，$t(\sigma^2)$ 的均方误差（Mean squared error，MSE）可以表示成

$$\mathrm{MSE}[t(\sigma^2)]=E[t(\sigma^2)-\mu]^2=g_1(\sigma^2)+g_2(\sigma^2)$$

其中

$$g_1(\sigma^2)=m'(G-GU'\Sigma^{-1}UG)m$$

$$g_2(\sigma^2)=(l-X'\Sigma^{-1}UGm)'(X'\Sigma^{-1}X)^{-1}(l-X'\Sigma^{-1}UGm)$$

参见 Henderson（1975）. 当 $\Sigma\geqslant 0$，Harville（1976）研究了 μ 的最佳线性无偏估计（预测）问题，但没有给出一个具体的表达式.

由于 μ 的最佳线性无偏预测 $t(\sigma^2)$ 含有未知参数向量 σ^2，因此在实际应用时，必须用 σ^2 的一个估计 $\hat{\sigma}^2$ 代入 $t(\sigma^2)$ 得到 μ 的两步预测（经验预测）$t(\hat{\sigma}^2)$. 估计 σ^2 的方法有很多，如极大似然估计（The maximum likelihood estimate，MLE）和限制极大似然估计（The restricted maximum likelihood estimate，REMLE）、方差分析估计（The analysis of variance estimate，ANOVAE）和最小范数二次无偏估计（The minimum norm quadratic unbiased estimate，MINQUE）（参见陈希孺和王松桂，2003）. 这些估计都是观测向量的偶函数和变换不变的，也就是说，对一切 y，有 $\hat{\sigma}^2(y)=\hat{\sigma}^2(-y)$，且对一切的 y 和 β，有 $\hat{\sigma}^2(y+X\beta)=\hat{\sigma}^2(y)$.

Kackar 和 Harville（1981）证明了如果 $\hat{\sigma}^2$ 是偶函数且是变换不变的，$Et(\hat{\sigma}^2)$ 是有限的且 ξ 和 ε 具有对称分布，那么经验预测 $t(\hat{\sigma}^2)$ 是无偏的. Kackar 和 Harville（1984）在正态假设下研究了经验预测 $t(\hat{\sigma}^2)$ 的均方误差的逼近问题. 他们证明了如果 $\mathrm{MSE}[t(\hat{\sigma}^2)]$ 是有限的，则当估计 $t(\hat{\sigma}^2)$ 是观测向量的偶函数和变换不变的时，有

$$\mathrm{MSE}[t(\hat{\sigma}^2)]=\mathrm{MSE}[t(\sigma^2)]+E[t(\hat{\sigma}^2)-t(\sigma^2)]^2$$

从上式可以看出，当参数真值 σ^2 用它的估计代替时，均方误差将会增大. Kackar 和 Harville（1984），Prasad 和 Rao（1990），Harville 和 Jeske（1992），Datta 和 Lahiri（2000）以及 Das、Jiang 和 Rao（2004）分别研究了均方误差增大的程度和 $\mathrm{MSE}[t(\hat{\sigma}^2)]$ 的逼近.

另外，对于式（1.4）表示的模型，Rao（1965），Harville（1976）和 Pfeffermann（1984）在回归系数随机时推广了 Gauss-Markov 定理，给出了随机回归系数和参数的最

佳线性无偏估计．在式（1.4）表示的模型中，设随机回归系数和参数的线性组合 $S\beta+Q\alpha$ 为可估函数，这里 S 和 Q 为已知的常数矩阵．吴启光（1988）和董莉明等（1988）对随机回归系数和参数的线性估计的可容许性问题作了深入的研究．他们应用统计判决理论，分别在二次损失和矩阵损失下得到了一个线性估计在齐次和非齐次线性估计类中是可容许估计的充要条件．但对于另一个重要的方面，估计的 Minimax 性的研究还较少在文献中见到．

1.3 矩阵论的预备知识

在 1.1 节和 1.2 节，为了描述线性模型，矩阵起了显著的作用．因而，在线性模型的预测理论中，矩阵的有关理论扮演了一个重要的角色．为了适应后续章节的需要，本节将讨论有关矩阵论的一些预备知识．无特别说明，本书所涉及的矩阵均指实矩阵．

1.3.1 线性空间

线性空间是几何学研究中的空间概念的一般化．为适应本书需要，使用线性空间的矩阵表示，叙述线性空间的某些基本结果．1.3.1 部分限于讨论 $n\times 1$ 的实向量组成的向量空间．

我们称实向量的集合 S 是实数域 $\mathbb{R}$ 上的线性空间，若 S 对向量加法和数乘两种运算封闭，即 S 中任意两个向量之和仍在 S 中，S 中任一实数和任一向量的乘积也仍在 S 中；且这两种运算满足加法结合律和交换律，数乘结合律和分配律等八条基本性质．记全体 $n\times 1$ 的实向量组成的集合为 $\mathbb{R}^n$，它是一个线性空间．考虑 $\mathbb{R}^n$ 中向量组 a_1，a_2，…，a_k 的一切可能的线性组合构成的集合为

$$S_1=\left\{x=\sum_{i=1}^{k}t_ia_i,t_1,t_2,\cdots,t_k\ \text{均为实数}\right\}$$

易知，S_1 也是一个线性空间，称为 $\mathbb{R}^n$ 的子空间．

若将 a_1，a_2，…，a_k 排成 $n\times k$ 的矩阵 $A=(a_1,\ a_2,\ \cdots,\ a_k)$，则 S_1 可以表示为

$$S_1=\{x=At,\quad t\in\mathbb{R}^k\}$$

其为 A 的列向量张成的子空间，记为 $S_1=\mathscr{R}(A)$．易证，$\mathbb{R}^n$ 的任意子空间都是某一矩阵的列向量张成的子空间．

设 a_1，a_2，…，a_k 为 $\mathbb{R}^n$ 中的一组向量，若存在不全为 0 的实数 t_1，t_2，…，t_k，使得 $t_1a_1+t_2a_2+\cdots+t_ka_k=0$，则称向量组 a_1，a_2，…，a_k 是线性线性相关的；否则称它们是线性无关的．如果子空间 S_1 由一组线性无关的向量 a_1，a_2，…，a_k 张成，则称 a_1，a_2，…，a_k 为 S_1 的一组基，k 称为子空间 S_1 的维数，记作 $k=\dim(S_1)$．

设矩阵 $A=(a_1,\ a_2,\ \cdots,\ a_k)$，$B=(b_1,\ b_2,\ \cdots,\ b_m)$，则容易证明：

(1) $\dim\mathscr{R}(A)=rk(A)$.

(2) $\mathscr{R}(A)\subset\mathscr{R}(A\,\vdots\,B)$.

对于 $\mathbb{R}^n$ 中的任意两个向量 $a_1=(a_{11},\ a_{12},\ \cdots,\ a_{1n})'$，$a_2=(a_{21},\ a_{22},\ \cdots,\ a_{2n})'$，定义它们的内积为 $(a_1,a_2)=a'_1a_2=\sum_{i=1}^{n}a_{1i}a_{2i}$．若 $(a_1,\ a_2)=0$，则称 a_1 与 a_2 正交，记为 $a_1\perp a_2$．若 a 与一个子空间 S 的每一个向量都正交，则称 a 和 S 正交，记为 $a\perp S$．称

$(a'_1 a_1)^{1/2} = \left(\sum_{i=1}^{n} a_{1i}^2\right)^{1/2}$ 为向量 a_1 的长度，记为 $\|a_1\|$.

设 S 为一子空间，令 $S^{\perp}=\{x: x\perp S\}$，则 $S^{\perp}$ 也是一个子空间，称为 S 的正交补空间. 设 A 为 $n\times k$ 矩阵，记 $A^{\perp}$ 表示满足 $A'A^{\perp}=0$ 的矩阵中具有最大秩的矩阵，则 $\mathscr{R}(A^{\perp})=\mathscr{R}(A)^{\perp}$.

下面的事实，在以后的讨论会常用到.

定理 1.1 对任意矩阵 A，有 $\mathscr{R}(A)=\mathscr{R}(AA')$.

证明 显然 $\mathscr{R}(AA')\subset\mathscr{R}(A)$，故只需证 $\mathscr{R}(A)\subset\mathscr{R}(AA')$. 事实上，对于任意给定的 $x\perp\mathscr{R}(AA')$，有 $x'AA'=0$. 两边右乘 x，得 $x'AA'x=\|A'x\|^2=0$，故 $A'x=0$. 于是 $x\perp\mathscr{R}(A)$. 明所欲证.

1.3.2 矩阵的广义逆

矩阵的广义逆的研究源于 Moore（1935）的著名论文. 对任意一个矩阵 A，Moore 用如下四个矩阵方程

$$AXA = A,\quad XAX = X,\quad (AX)' = AX,\quad (XA)' = XA$$

定义了 A 的广义逆 X. 但当时并未引起人们的注意. 直到 1955 年，Penrose 证明了满足上述方程的广义逆具有唯一性后，广义逆的研究才真正为人们所重视. 基于此，人们把这样定义的矩阵广义逆称为 Moore - Penrose 广义逆. 另外，Penrose 首先注意到了广义逆和线性方程组之间的关系. 假设

$$Ax = b \tag{1.14}$$

是相容的线性方程组，这里 A 是 $m\times n$ 矩阵，其秩 $rk(A)=r\leqslant\min(m, n)$. 显然，当 $r=m=n$ 时，方程组（1.14）有唯一解 $x=A^{-1}b$. 但是，当 A 不可逆或不是方阵时，方程组（1.14）有无穷多个解，如何用 A 和 b 通过简单形式表征方程组（1.14）的全体解，之前是很困难的. Penrose 指出，在研究方程组（1.14）的解时，所用的广义逆只需满足上述四个矩阵方程组的第一个矩阵方程. 自此以后的几十年，出现了大量关于这种广义逆的研究，并利用这种广义逆彻底解决了相容线性方程组解的表征问题. 1.3.2 部分讨论这种广义逆及其重要的性质和应用.

定义 1.1 对任意矩阵 $A_{m\times n}$，一切满足方程组

$$AXA = A$$

的矩阵 X，称为矩阵 A 的广义逆，记为 A^{-}.

下面的定理 1.2～1.6 和推论 1.1 引自 Wang 和 Chow（1994），pp. 16 - 30.

首先，由如下定理给出 A^{-} 的存在性和构造性问题的解.

定理 1.2 设 A 是 $m\times n$ 的矩阵，秩 $\mathrm{rk}(A)=s$. 若

$$A = P\begin{pmatrix} I_s & 0 \\ 0 & 0 \end{pmatrix}Q$$

其中，P 和 Q 分别是 $m\times m$ 和 $n\times n$ 的可逆阵，则

$$A^{-} = Q^{-1}\begin{pmatrix} I_s & B \\ C & D \end{pmatrix}P^{-1}$$

其中，B、C 和 D 为任意适当阶数的矩阵.

推论 1.1 对任意矩阵 A，有：

(1) A^- 总是存在的.

(2) A^- 是唯一的，当且仅当 A 是非奇异矩阵时，且 $A^-=A^{-1}$.

(3) 假设 $F\neq 0$，$G\neq 0$，那么 FA^-G 关于任意广义逆 A^- 是不变的，当且仅当 $\mathscr{R}(F')\subset\mathscr{R}(A')$ 且 $\mathscr{R}(G)\subset\mathscr{R}(A)$.

(4) $P_A=A(A'A)^-A'$ 和广义逆 $(A'A)^-$ 的选择无关，且 $P_AA=A$.

定理 1.3 设 $A_y=b$ 是一个相容的线性方程组. 则有：

(1) 对任意广义逆 A^-，$y=A^-b$ 必为 $Ay=b$ 的一个解.

(2) 齐次线性方程组 $Ay=0$ 的通解为

$$y=(I-A^-A)z$$

其中，A^- 是 A 的任一给定的广义逆，z 为任意向量.

(3) $Ay=b$ 的通解为

$$y=A^-b+(I-A^-A)z$$

其中，A^- 为 A 的任一给定的广义逆；z 为任意向量.

定理 1.4 对于一个给定的分块矩阵

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}$$

若 $\mathscr{R}(A_{12})\subset\mathscr{R}(A_{11})$，$\mathscr{R}(A'_{21})\subset\mathscr{R}(A'_{11})$，则

$$A^-=\begin{pmatrix}A_{11}^-+A_{11}^-A_{12}A_{22\cdot 1}^-A_{21}A_{11}^- & -A_{11}^-A_{12}A_{22\cdot 1}^-\\ -A_{22\cdot 1}^-A_{21}A_{11}^- & A_{22\cdot 1}^-\end{pmatrix} \tag{1.15}$$

若 $\mathscr{R}(A_{21})\subset\mathscr{R}(A_{22})$，$\mathscr{R}(A'_{12})\subset\mathscr{R}(A'_{22})$，则

$$A^-=\begin{pmatrix}A_{11\cdot 2}^- & -A_{11\cdot 2}^-A_{12}A_{22}^-\\ -A_{22}^-A_{21}A_{11\cdot 2}^- & A_{22}^-+A_{22}^-A_{21}A_{11\cdot 2}^-A_{12}A_{22}^-\end{pmatrix} \tag{1.16}$$

其中，$A_{22\cdot 1}=A_{22}-A_{21}A_{11}^-A_{12}$，$A_{11\cdot 2}=A_{11}-A_{12}A_{22}^-A_{21}$.

注 式 (1.13) 和式 (1.14) 给出的是 A 的全体广义逆构成的集合的一个子集.

定理 1.5 (镶边矩阵广义逆) 设 $S\geqslant 0$ 是 $n\times n$ 的矩阵，L 是 $m\times n$ 的矩阵，那么

$$\begin{pmatrix}S & L'\\ L & 0\end{pmatrix}^-=\begin{pmatrix}T^--T^-L'U^-LT^- & T^-L'U^-\\ U^-LT^- & U^-U-U^-\end{pmatrix}$$

其中，$T=S+L'L$，$U=LT^-L'$.

定义 1.2 对任意一个矩阵 A，若 X 满足如下四个矩阵方程

$$AXA=A,\quad XAX=X,\quad (AX)'=AX,\quad (XA)'=XA$$

则称矩阵 X 为 A 的 Moore - Penrose 广义逆，记为 A^+.

定理 1.6 设 A 是 $m\times n$ 的矩阵，秩为 s，则：

(1) 存在两个正交矩阵 P 和 Q，使得

$$A=P\begin{pmatrix}\Delta_s & 0\\ 0 & 0\end{pmatrix}Q'$$

其中，$\Delta_s=\text{diag}(\zeta_1,\ \cdots,\ \zeta_s)$，$\zeta_i>0$，$i=1,\ 2,\ \cdots,\ s$，而 ζ_1^2，$\cdots$，ζ_s^2 为 $A'A$ 的非零特征值.