

责任编辑	高尚华
封面设计	张楠
责任绘图	吴文信
版式设计	马静如
责任校对	胡晓琪
责任印制	杨明

# 前 言

线性统计模型是一类很重要的统计模型,它包括了线性回归模型、方差分析模型等应用十分广泛的许多模型,同时线性模型的理论和方法也是学习和研究其它统计方法的基础.正是由于这些原因,线性统计模型不仅已成为统计专业本科生和研究生的必修课,而且也是生物、医学、经济、管理、商业、金融、工程技术以及社会科学等学科本科生和研究生统计课程的重要内容.为了适应各方面的需要,我们在教育部高教司的关心和支持下编写了这本教材.

本书的目的是为需要学习线性统计模型基础知识的各专业的学生提供一本教材,因此,阅读本书所要求的数学方面的预备知识并不多.我们认为,读者如果掌握了工科大学的微积分、线性代数和初等概率统计知识,就可以顺利阅读本书的绝大部分内容.

全书共分七章.前两章是预备性知识.第一章通过例子引进了我们所要研究的模型,第二章介绍了随机向量特别是正态向量的基础知识.接下来的三章,即第三、四、五章对线性回归模型的估计和检验做了系统讨论.第六章介绍了方差分析模型.第七章概要介绍了几类在经济、生物、医学等领域颇为有用的线性回归模型.为了帮助读者理解基本内容,掌握其中的方法,书中给出了一些应用实例.书末有三个附录.附录1给出了矩阵论的一些预备知识.附录2提供了书中部分例子计算机软件如SAS软件、SPSS软件的计算机输出结果,以帮助读者了解如何使用计算机求解线性模型.除第七章外,各章末配有一定数量的习题.

本书第一至第四章由王松桂执笔,第五章和第七章由陈敏执笔,第六章由陈立萍执笔,最后由王松桂统一修改定稿.陈立萍选配了大部分章末的习题.陈立萍和程维虎曾经在北京工业大学应用数学系讲授过本书;王松桂曾多次在北京工业大学工科博士生“数学模型”课上讲授过本书的部分内容.

本书的写作得到国家自然科学基金、北京市自然科学基金和北京市教委科技发展计划支持.另外,1997年12月在苏州大学召开了本书的审稿会.汪仁官、茹诗松、王静龙、陈庆云、傅珏生和高尚华等专家对本书进行了认真评审,提出了许多宝贵意见.这对提高本书质量起了重要作用.趁本书出版之际,我们谨向以上单位和专家表示衷心感谢.

## 2 前 言

---

由于作者水平所限,书中的缺点和错误在所难免,恳请同行和广大读者批评指正.

编 者

1999年4月30日

# 目 录

<b>第一章 引论</b> .....	1
§ 1.1 线性回归模型 .....	1
§ 1.2 方差分析模型 .....	6
§ 1.3 应用概述 .....	8
习题一 .....	10
<b>第二章 随机向量</b> .....	12
§ 2.1 均值向量与协方差阵 .....	12
§ 2.2 随机向量的二次型 .....	15
§ 2.3 正态随机向量 .....	15
§ 2.4 $\chi^2$ 分布 .....	21
习题二 .....	25
<b>第三章 回归参数的估计</b> .....	28
§ 3.1 最小二乘估计 .....	28
§ 3.2 最小二乘估计的性质 .....	34
§ 3.3 约束最小二乘估计 .....	40
§ 3.4 回归诊断 .....	43
§ 3.5 Box-Cox 变换 .....	53
§ 3.6 广义最小二乘估计 .....	55
§ 3.7 复共线性 .....	58
§ 3.8 岭估计 .....	64
§ 3.9 主成分估计 .....	71
习题三 .....	75
<b>第四章 假设检验与预测</b> .....	81
§ 4.1 一般线性假设 .....	81
§ 4.2 回归方程的显著性检验 .....	87
§ 4.3 回归系数的显著性检验 .....	90
§ 4.4 异常点检验 .....	92
§ 4.5 因变量的预测 .....	96
习题四 .....	99
<b>第五章 回归方程的选择</b> .....	102
§ 5.1 评价回归方程的标准 .....	102

## 2 目 录

---

§ 5.2 计算所有可能的回归·····	112
§ 5.3 计算最优子集回归·····	117
§ 5.4 逐步回归·····	126
习题五·····	134
<b>第六章 方差分析模型</b> ·····	<b>138</b>
§ 6.1 单因素方差分析·····	138
§ 6.2 两因素方差分析·····	147
§ 6.3 正交试验设计与方差分析·····	155
习题六·····	161
<b>*第七章 其它线性回归模型</b> ·····	<b>163</b>
§ 7.1 引言·····	163
§ 7.2 具有异方差误差的线性回归模型·····	164
§ 7.3 具有自回归误差的线性回归模型·····	178
§ 7.4 具有一阶自回归误差的线性回归模型·····	186
§ 7.5 对一阶自回归误差的假设检验·····	202
§ 7.6 半相依线性回归模型·····	207
<b>附录 1 关于矩阵的若干基础知识</b> ·····	<b>221</b>
<b>附录 2 本书部分例题常用统计软件包计算机输出结果</b> ·····	<b>226</b>
<b>附录 3 Durbin-Watson 统计量的上、下界值表</b> ·····	<b>233</b>
<b>参考文献</b> ·····	<b>236</b>

# 第一章 引 论

线性统计模型是现代统计学中应用最为广泛的模型之一,而且也是其它统计模型研究或应用的基础.之所以如此,其原因主要是

1. 在现实世界中,许多量之间具有线性或近似的线性依赖关系.
2. 在现实世界中,虽然许多量之间的关系是非线性的,但是经过适当的变换,变换过后的新变量之间具有近似的线性关系.
3. 线性关系是数学中最基本的关系,因而比较容易处理.在数学中已经积累了处理线性关系的丰富的理论与方法,为实际应用提供了坚实的理论依据和有效算法.

本章我们通过一些实例引进线性统计模型,使读者对这种模型丰富的实际背景有一定了解,这对后面要引进的一些统计概念和方法的理解将是大有裨益的.

## § 1.1 线性回归模型

在现实世界中,存在着大量这样的情况:两个变量例如  $X$  和  $Y$  有一些依赖关系.由  $X$  可以部分地决定  $Y$  的值,但这种决定往往不很确切.常常用来说明这种依赖关系的最简单、直观的例子是体重与身高.若用  $X$  表示某人的身高,用  $Y$  表示他的体重.众所周知,一般说来,当  $X$  大时,  $Y$  也倾向于大,但由  $X$  不能严格地决定  $Y$ .又如,城市生活用电量  $Y$  与气温  $X$  有很大的关系.在夏天气温很高或冬天气温很低时,由于室内空调、冰箱等家用电器的使用,可能用电量就高.相反,在春秋季节气温不高也不低,用电量就可能少.但我们不能由气温  $X$  准确地决定用电量  $Y$ .类似的例子还很多.变量之间的这种关系称为“相关关系”,回归模型就是研究相关关系的一个有力工具.

在以上诸例中,  $Y$  通常称为因变量或响应变量,  $X$  称为自变量或预报变量.我们可以设想,  $Y$  的值由两部分组成:一部分是由  $X$  能够决定的部分,它是  $X$  的函数,记为  $f(X)$ .而另一部分则由其它众多未加考虑的因素(包括随机因素)所产生的影响,它被看作随机误差,记为  $e$ .于是我们得到如下模型:

$$Y = f(X) + e. \quad (1.1.1)$$

这里  $e$  作为随机误差,我们有理由要求它的均值  $E(e) = 0$ ,其中  $E(\cdot)$  表示随机

变量的均值.

特别,当  $f(X)$  是线性函数  $f(X) = \beta_0 + \beta_1 X$  时,我们得到

$$Y = \beta_0 + \beta_1 X + e. \quad (1.1.2)$$

在这个模型中,若忽略掉  $e$ ,它就是一个通常的直线方程.因此,我们称(1.1.2)为线性回归模型或线性回归方程.关于“回归”一词的由来,我们留在后面作解释.常数项  $\beta_0$  是直线的截距, $\beta_1$  是直线的斜率,也称为回归系数.在实际应用中, $\beta_0$  和  $\beta_1$  皆是未知的,需要通过观测数据来估计.

假设自变量  $X$  分别取值为  $x_1, x_2, \dots, x_n$  时,因变量  $Y$  对应的观测值分别为  $y_1, y_2, \dots, y_n$ .于是我们有  $n$  组观测值  $(x_i, y_i), i = 1, \dots, n$ .如果  $Y$  与  $X$  有回归关系(1.1.2),则这些  $(x_i, y_i)$  应该满足

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n, \quad (1.1.3)$$

这里  $e_i$  为对应的随机误差.基于(1.1.3),应用适当的统计方法(这将在第三章讨论)可以得到  $\beta_0$  和  $\beta_1$  的估计值  $\hat{\beta}_0, \hat{\beta}_1$ ,将它们代入(1.1.2),再略去误差项  $e_i$  得到

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X, \quad (1.1.4)$$

称之为经验回归直线,也称为经验回归方程.这里“经验”两字表示这个回归直线是基于前面的  $n$  次观测数据  $(x_i, y_i), i = 1, \dots, n$  而获得的,在许多文献中,人们往往把“经验”两字省略掉.经过适当的统计检验后,我们可以认为(1.1.4)描述了因变量  $Y$  与自变量  $X$  之间的相关关系.

**例 1.1.1** 肥胖是现代社会人们关注的一个重要问题,那么体重多少才算是肥胖呢?这当然跟每个人的身高有关,于是许多学者应用直线回归方法研究人的体重与身高的关系.假设  $X$  表示身高(cm), $Y$  表示体重(kg).我们假设  $Y$  与  $X$  之间具有回归关系(1.1.2).在这里误差  $e$  表示除了身高  $X$  之外,所有影响体重  $Y$  的其它因素,例如遗传因素、饮食习惯、体育锻炼多少等.为了估计其中的参数  $\beta_0$  和  $\beta_1$ ,研究者测量了很多人的身高  $x_i$  和他们的体重  $y_i, i = 1, \dots, n$ ,得到关系(1.1.3).从而应用统计方法可以估计出  $\beta_0$  和  $\beta_1$ .一种研究结果是,若用  $X - 150$  作自变量,则得  $\hat{\beta}_0 = 50, \hat{\beta}_1 = 0.6$ ,也就是说我们有经验回归直线.

$$Y = 50 + (X - 150) \times 0.6.$$

我们可以把它改写成如下形式:

$$Y = -40 + 0.6X, \quad (1.1.5)$$

这个经验回归方程在一定程度上描述了体重与身高的相关关系.给定  $X$  的一个具体值  $x_0$ ,我们可以算出对应的  $Y$  值  $y_0 = -40 + 0.6x_0$ .例如某甲身高  $x_0 = 160$  (cm),代入(1.1.5)可以算出对应  $y_0 = 56$  (kg).我们称 56kg 为身高是 160cm 的

人的体重的预测.这就是说,对于一个身高 160cm 的人,我们预测他的体重大致为 56kg,但实际上,他的体重不可能恰为 56kg.可能比 56kg 多,也可能比 56kg 少.

**例 1.1.2** 我们知道,一个公司的商品销售量与其广告费有密切关系,一般说来在其它因素(如产品质量等)保持不变的情况下,它用在广告上的费用愈高,它的商品销售量也就会愈多.但这也只是一种相关关系.某公司为了进一步研究这种关系.用  $X$  表示在某地区的年度广告费, $Y$  表示年度商品销售量.根据过去一段时间的销售记录  $(x_i, y_i), i = 1, \dots, n$ , 采用线性回归模型(1.1.3), 假定计算出  $\hat{\beta}_0 = 1\ 608.5, \hat{\beta}_1 = 20.1$ , 于是得到经验回归直线

$$Y = 1\ 608.5 + 20.1X. \quad (1.1.6)$$

这个经验回归直线告诉我们,广告费  $X$  每增加一个单位,该公司销售收入就增加 20.1 个单位.如果某地区人口增加得很快,那么很可能人口总数也是影响销售量的一个重要因素.若记  $X_1$  为年度广告费, $X_2$  为某地区人口总数.我们可以考虑如下含两个自变量的线性回归模型:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e. \quad (1.1.7)$$

同样,根据记录的历史数据,应用适当统计方法可以估计出  $\beta_i, i = 0, 1, 2$ . 假定估计出的

$$\hat{\beta}_0 = 320.3, \quad \hat{\beta}_1 = 18.4, \quad \hat{\beta}_2 = 0.2,$$

则我们得到经验回归方程

$$Y = 320.3 + 18.4X_1 + 0.2X_2. \quad (1.1.8)$$

从这个经验回归方程我们可以看出,当广告费  $X_1$  增加或人口总数  $X_2$  增加时,商品销售量都增加,且当人口总数保持不变时,广告费每增加 1 个单位,销售量增加 18.4 个单位.而当广告费保持不变,而该地区人口总数每增加一个单位,该公司销售量增加 0.2 个单位.当然,在实际应用中,并不是每个经验回归方程都能描述变量之间的客观存在的真正的关系.关于这一点,将在第四章详细讨论.

在实际问题中,影响因变量的主要因素往往很多,这就需要考虑含多个自变量的回归问题.假设因变量  $Y$  和  $p-1$  个自变量  $X_1, \dots, X_{p-1}$  之间有如下关系:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e, \quad (1.1.9)$$

这是多元线性回归模型,其中  $\beta_0$  为常数项, $\beta_1, \dots, \beta_{p-1}$  为回归系数, $e$  为随机误差.

假设我们对  $Y, X_1, \dots, X_{p-1}$  进行了  $n$  次观测,得到  $n$  组观测值

$$x_{i1}, \dots, x_{i,p-1}, y_i, \quad i = 1, \dots, n,$$

它们满足关系式



$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1} + e_i, \quad i = 1, \cdots, n, \quad (1.1.10)$$

这里  $e_i$  为对应的随机误差. 引进矩阵记号

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

(1.1.10)就写为如下简洁形式:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (1.1.11)$$

这里  $\mathbf{y}$  为  $n \times 1$  的观测向量.  $\mathbf{X}$  为  $n \times p$  已知矩阵, 通常称为设计矩阵. 对于线性回归模型, 术语“设计矩阵”中的“设计”两字并不蕴含任何真正设计的含义, 只是习惯用法而已. 近年来, 有一些学者建议改用“模型矩阵”. 但就目前来讲, 沿用“设计矩阵”者居多.  $\boldsymbol{\beta}$  为未知参数向量. 其中  $\beta_0$  称为常数项, 而  $\beta_1, \cdots, \beta_{p-1}$  为回归系数. 而  $\mathbf{e}$  为  $n \times 1$  随机误差向量, 关于  $\mathbf{e}$  最常用的假设是:

(a) 误差项均值为零, 即  $E(e_i) = 0, i = 1, \cdots, n$ .

(b) 误差项具有等方差, 即

$$\text{Var}(e_i) = \sigma^2, \quad i = 1, \cdots, n. \quad (1.1.12)$$

(c) 误差是彼此不相关的, 即

$$\text{Cov}(e_i, e_j) = 0, \quad i \neq j, \quad i, j = 1, \cdots, n.$$

通常称以上三条为 Gauss - Markov 假设. 模型(1.1.11)和假设(1.1.12)构成了我们以后要讨论的最基本的线性回归模型.

在 Gauss - Markov 假设中, 第一条表明误差项不包含任何系统的趋势, 因而观测值  $y_i$  的均值

$$E(y_i) = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1}, \quad i = 1, \cdots, n.$$

这就是说, 观测值  $y_i$  大于或小于其均值  $E(y_i)$  的波动完全是一种随机性的, 这种随机性来自误差项  $e_i$ . 我们知道, 一个随机变量的方差刻画了该随机变量取值散布程度的大小, 因此假设(b)要求  $e_i$  等方差, 也就是要求不同次的观测  $y_i$  在其均值附近波动程度是一样的. 这个要求有时显得严厉一些. 在一些情况下, 我们不得不放松为  $\text{Var}(e_i) = \sigma_i^2, i = 1, \cdots, n$ , 这种情况将在 § 7.2 讨论. 第三条假设等价于要求不同次的观测是不相关的. 在实际应用中这个假设比较容易满足. 但是在一些实际问题中, 误差往往是相关的. 这时估计问题比较复杂, 本书中不少地方要讨论这种情形.

对于模型(1.1.10), 假设  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_{p-1})'$  为  $\boldsymbol{\beta}$  的一种估计, 将它们代入(1.1.9), 并略去其中的误差项  $e$ , 得到经验回归方程

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_{p-1} X_{p-1}. \quad (1.1.13)$$

和单个回归自变量的情形一样,这个经验回归方程是否真正描述了因变量  $Y$  与自变量  $X_1, \dots, X_{p-1}$  之间的关系,还需要适当的统计检验.

上面我们讨论的都是线性回归模型.有一些模型虽然是非线性的,但经过适当变换,可以化为线性模型.

**例 1.1.3** 在经济学中,著名的 Cobb-Douglas 生产函数为

$$Q_t = aL_t^b K_t^c, \quad (1.1.14)$$

这里  $Q_t, L_t$  和  $K_t$  分别为  $t$  年的产值、劳力投入量和资金投入量,  $a, b$  和  $c$  为参数.在上式两边取自然对数,得到

$$\ln(Q_t) = \ln a + b \ln(L_t) + c \ln(K_t).$$

若令  $y_t = \ln(Q_t), x_{t1} = \ln(L_t), x_{t2} = \ln(K_t),$

$$\beta_0 = \ln a, \beta_1 = b, \beta_2 = c,$$

则再加上误差项,便得到线性关系

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + e_t, \quad t = 1, \dots, T. \quad (1.1.15)$$

因此我们把非线性模型(1.1.14)化成了线性模型.

**例 1.1.4** 多项式回归模型

假设因变量  $Y$  和自变量  $X$  之间具有关系

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + e.$$

这是二次多项式回归模型.若令  $X_1 = X, X_2 = X^2, X_3 = X^3,$  则有

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e.$$

这就是一个线性模型.从这里我们看到,线性模型(1.1.9)的基本特征是:它关于未知参数  $\beta_1, \beta_2, \dots, \beta_{p-1}$  是线性的.

在结束这一节之前,我们解释一下“回归”一词的由来.“回归”英文为“regression”,是由英国著名生物学家兼统计学家高尔顿(Galton)在研究人类遗传问题时提出的.为了研究父代与子代身高的关系,高尔顿收集了 1 078 对父亲及其一子的身高数据.用  $X$  表示父亲身高,  $Y$  表示儿子身高.单位为英寸(1 英寸为 2.54cm).将这 1 078 对  $(x_i, y_i)$  标在直角坐标纸上,他发现散点图大致呈直线状.也就是说,总的趋势是父亲的身高  $X$  增加时,儿子的身高  $Y$  也倾向于增加,这与我们的常识是一致的.但是,高尔顿对数据的深入分析,发现了一个很有趣的现象——回归效应.

因为这 1 078 个  $x_i$  值的算术平均值  $\bar{x} = 68$  英寸,而 1 078 个  $y_i$  值的平均值为  $\bar{y} = 69$  英寸,这就是说,子代身高平均增加了 1 英寸.人们自然会这样推想,若父亲身高为  $x$ ,他儿子的平均身高大致应为  $x + 1$ ,但高尔顿的仔细研究所得

结论与此大相径庭. 他发现, 当父亲身高为 72 英寸时(请注意, 比平均身高  $\bar{x} = 68$  英寸要高), 他们的儿子平均身高仅为 71 英寸. 不但达不到预期的  $72 + 1 = 73$  英寸, 反而比父亲身高低了 1 英寸. 反过来, 若父亲身高为 64 英寸(请注意, 比平均身高  $\bar{x} = 68$  英寸要矮), 他们的儿子平均身高为 67 英寸, 竟比预期的  $64 + 1 = 65$  英寸高出了 2 英寸. 这个现象不是个别的, 它反映了一个一般规律: 即身高超过平均值  $\bar{x} = 68$  英寸的父亲, 他们的儿子的平均身高将低于父亲的平均身高. 反之, 身高低于平均身高  $\bar{x} = 68$  英寸的父亲, 他们的儿子的平均身高将高于父亲的平均身高. 高尔顿对这个一般结论的解释是: 大自然具有一种约束力, 使人类身高的分布在一定时期内相对稳定而不产生两极分化, 这就是所谓的回归效应. 通过这个例子, 高尔顿引进了“回归”一词. 用他的数据, 可以计算出儿子身高  $Y$  与父亲身高  $X$  的经验关系

$$Y = 35 + \frac{1}{2}X,$$

它代表一条直线, 人们也就把这条直线称为回归直线. 当然, 这个经验回归直线只反映了变量相关关系中具有回归效应的一种特殊情况, 对更多的相关关系, 并非都是如此. 特别是在涉及多个自变量的情况中, 回归效应便不复存在. 因此将 (1.1.9) 或 (1.1.11) 或 (1.1.13) 分别称为线性回归模型和经验回归方程, 并把对应的统计分析称为回归分析, 不一定恰当. 但“回归”这个名词沿用已久, 实无改变之必要与可能.

## § 1.2 方差分析模型

在上节引进的线性回归模型中, 所涉及的自变量——一般来说都可以是连续变量, 研究的基本目的则是寻求因变量与自变量之间客观存在的依赖关系. 而本节所要引进的模型则不同, 它的自变量是示性变量, 这种变量往往表示某种效应的存在与否, 因而只能取 0, 1 两个值. 这种模型是比较两个或多个因素效应大小的一种有力工具. 因为比较因素效应的统计分析在统计学上叫做方差分析, 所以对应地, 人们将这种模型称为方差分析模型. 在一些文献中, 也把这种模型称为试验设计模型, 这是因为它所分析的数据往往跟一个预先安排的试验相联系.

**例 1.2.1** 某农业科学研究机构欲比较三种小麦品种的优劣, 安排了一种比较试验. 为保证试验结果的客观性, 他们选择了六块面积相等, 土质肥沃程度一样的田地, 每一种小麦播种在其中的两块田内, 并给予几乎完全相同的田间管理.

这是一个简单的试验, 用  $y_{ij}$  表示种第  $i$  种小麦的第  $j$  块田的产量, 那么我们可以对  $y_{ij}$  作如下分解:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i = 1, 2, 3; j = 1, 2. \quad (1.2.1)$$

这里  $\mu$  表示总平均值,  $\alpha_i$  表示第  $i$  种小麦品种的效应,  $e_{ij}$  是随机误差, 它表示所有其它未加控制因素以及各种误差的总效应. 若用矩阵记号, 则(1.2.1)变为

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} + \begin{pmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{pmatrix}.$$

记

$$\begin{aligned} y &= (y_{11}, y_{12}, y_{21}, y_{22}, y_{31}, y_{32})', \\ e &= (e_{11}, e_{12}, e_{21}, e_{22}, e_{31}, e_{32})', \end{aligned}$$

它们分别是  $6 \times 1$  的观测向量和随机误差向量.

$\beta = (\mu, \alpha_1, \alpha_2, \alpha_3)'$  为  $4 \times 1$  未知参数向量,

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

这是  $6 \times 4$  矩阵, 称为设计矩阵. 引进这些矩阵之后, 上面的模型就具有形式

$$y = \mathbf{X}\beta + e. \quad (1.2.2)$$

它和上节引进的线性回归模型(1.1.11)形式上完全一样. 不同的是在(1.1.11)中, 除第 1 列之外,  $\mathbf{X}$  的元素原则上可以取任意连续值, 而对现在的模型, 设计矩阵  $\mathbf{X}$  的所有元素只能取 0 和 1 两个值.  $\mathbf{X}$  的第 2, 3 和 4 列分别对应于小麦品种 1, 2 和 3. 例如, 第 2 列对应于第一种小麦品种. 在这一列中, 1 表示在对应的那块田中, 播种的是第一种小麦, 即效应  $\alpha_1$  存在. 0 表示在对应的那块田中, 播种的不是第一种小麦, 其余类推. 我们进行统计分析的目的是比较这三种小麦品种, 即比较它们的效应  $\alpha_1, \alpha_2$  和  $\alpha_3$  的大小.

这个例子所引进的模型是方差分析模型中最简单的一种, 称为单因素方差分析模型, 这是因为它只涉及“小麦品种”这一个因素.

**例 1.2.2** (续上例) 假定小麦产量与所施化学肥料品种有很大关系. 设在六块田中施用了两种化肥. 这时观测值可表示为  $y_{ijk}$ , 它为种第  $i$  种小麦并施了第  $j$  种化肥的第  $k$  块田的产量, 且可分解为

$$y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk},$$

$$i = 1, 2, 3; j = 1, 2; k = 1, 2. \quad (1.2.3)$$

其中  $\mu$  为总平均,  $\alpha_i$  表示第  $i$  种小麦的效应,  $\beta_j$  表示第  $j$  种化肥的效应. 这个问题包含了两个因素: 一是小麦品种, 二是化肥品种. 因此, 这是两个因素的方差分析模型. 假定我们分析的目的仍是比较小麦品种, 而不是比较这两种化肥的效力. 在试验设计中, 把要比较的这个因素通称为“处理”, 而把为消除试验条件差异而引进的“化肥”这个因素通称为“区组”. 这个例子就是所谓的区组设计的一种特殊情况. 记

$$y = (y_{111}, y_{112}, y_{121}, y_{122}, y_{211}, y_{212}, y_{221}, y_{222}, y_{311}, y_{312}, y_{321}, y_{322})',$$

$$e = (e_{111}, e_{112}, e_{121}, e_{122}, e_{211}, e_{212}, e_{221}, e_{222}, e_{311}, e_{312}, e_{321}, e_{322})',$$

它们分别是  $12 \times 1$  的观测向量和随机误差向量. 记  $\beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)'$  它是  $6 \times 1$  的未知参数向量. 设计矩阵  $X$  为  $12 \times 6$  矩阵, 具有形式

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

引进这些矩阵记号之后, 我们把两因素方差分析模型也表成了(1.2.2)的形式.

从前面的讨论我们看到, 应用矩阵记号, 线性回归模型和方差分析模型都具有相同形式. 因此, 可以对它们用统一的方法来进行统计分析. 但是它们的用处却有较大不同, 线性回归模型本质上用于描述变量之间的依赖关系, 方差分析模型主要用于比较效应的大小.

还存在着其它几类线性模型, 如协方差分析模型、方差分量模型等. 这些模型的讨论超出了本教材的范围. 感兴趣的读者可以参阅王松桂(1987).

### § 1.3 应用概述

对回归模型所进行的统计分析, 通常称为回归分析. 本节将介绍回归分析的

实际应用. 归纳起来主要有以下四个方面.

### 1. 描述变量之间的关系

从 § 1.1 我们已经看到, 有了一组因变量和自变量的数据, 我们通过回归分析方法, 可以建立一个线性经验回归方程. 通过一定的检验, 我们认为这个方程描述了因变量和自变量之间的相依关系. 在 § 1.1 的讨论中, 我们已经举了这方面的两个例子. 这里我们要强调的是, 我们所建立的刻画变量之间关系的“函数”之所以称为经验回归方程, 是因为这个方程建立在现有的一组数据的基础上, 它是对现有数据中所包含的变量关系信息的一种归纳. 这种归纳当然与所获得的数据有关, 因此它不十分准确. 当我们观察的数据增加时, 经验回归方程随之会有一些改变. 一般说来, 当观察数据愈多, 经验回归方程所归纳的信息也就愈多, 因而准确程度会随之提高.

另一方面, 当我们通过一组数据, 获得了一个经验回归方程后, 我们还要考察这个经验回归方程是否真正刻画了因变量与自变量之间客观存在的依赖关系. 这是因为, 当我们应用线性回归模型对数据进行分析时, 面临着模型选择, 自变量选择, 误差假设适用性等问题. 诸多问题有一个处理得不当, 都会导致一定的偏差. 因此, 在实际应用中, 建立经验回归方程的过程是一个迭代过程. 先选择一个初始模型, 基于数据得到经验回归方程后, 经过一些统计检验后结合问题专业知识的解释, 若认为初始模型不够合理, 则对其进行适当修正, 或改变估计方法, 然后再重新建立经验回归方程. 重复上述过程, 直到所得到的经验回归方程从诸多角度考察都比较满意为止.

### 2. 分析变量之间的相互关系

当我们建立了一个比较满意的经验回归方程之后, 就可以利用它分析变量之间的关系. 用  $Y$  表示因变量,  $X_1, \dots, X_p$  表示自变量. 设经验回归方程为

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p. \quad (1.3.1)$$

在适当消除了自变量  $X_1, \dots, X_p$  所选用计量单位的影响之后, 回归系数  $\beta_i$  的估计值  $\hat{\beta}_i$  的大小在一定程度上, 反映了变量  $X_i$  对因变量  $Y$  的影响大小. 当其余自变量保持不变时,  $X_i$  每增加一个单位, 因变量  $Y$  平均“增加”  $\hat{\beta}_i$  个单位. 因此, 当  $\hat{\beta}_i > 0$  时,  $Y$  与  $X_i$  是正的相关关系, 即  $Y$  随着  $X_i$  的增加而增加. 相反, 当  $\hat{\beta}_i < 0$  时,  $Y$  与  $X_i$  是负的相关关系, 即  $Y$  随着  $X_i$  的增加而减少. 同时回归系数  $\hat{\beta}_i$  也刻画了它们相关的程度. 这样我们可以根据  $\hat{\beta}_i$  的符号以及绝对值  $|\hat{\beta}_i|$  的大小将自变量进行分类: 正相关变量、负相关变量或重要自变量和次要自变量.

另一方面, 应用一些统计分析方法, 我们还可以分析自变量之间存在的相关关系. 随着电子计算机的发展与普及, 人们愈来愈多地处理含较多自变量的大型

回归问题.例如在一些经济问题中,回归自变量可以多到数百个,甚至上千个.这么多的自变量,它们之间往往有一些依赖关系,回归分析中的主成分估计法就可以帮助我们寻找回归自变量之间的相依关系(详细讨论见第三章).这种关系的发现往往对我们认识问题的内在规律具有相当大的帮助.

### 3. 预测

预测就是通过经验回归方程,对自变量  $X_1, \dots, X_p$  的一组特定值去预测因变量  $Y$  将可能取的值.这个问题相对说来比较简单,但它却构成了回归分析的一个极为重要的应用.例如,若在(1.3.1)中,  $Y$  表示一个公司的商品销售值,  $X_1, \dots, X_p$  表示影响销售的自变量,那么对这些自变量的一组特定值  $X_1 = x_1^{(0)}, \dots, X_p = x_p^{(0)}$ ,从(1.3.1)就可以预测对应的销售值

$$y_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^{(0)} + \dots + \hat{\beta}_p x_p^{(0)}.$$

当然,这里会有一些误差.其误差的大小与所选模型、参数估计方法都有密切关系.

这里我们谈的是点预测.在应用上,我们往往需要用去一个区间去预测因变量,这称为区间预测.这种预测相对要复杂一些,留在 § 4.5 讨论.

线性统计模型是一类最简单、应用最广泛的数学模型.它的应用遍及工业、农业、经济、保险、生物、医学、工程技术和社会科学等领域,是分析数据,寻求变量关系的一种有力工具.

## 习题一

1.1 假设一物体长度为  $\mu$ ,但  $\mu$  是未知的,我们欲估计它,于是将其测量了  $n$  次,得到测量值为  $y_1, \dots, y_n$ .如果测量过程没有系统误差,我们可以认为  $y_i, i=1, \dots, n$  为来自正态总体  $N(\mu, \sigma^2)$  的一组随机样本.试将这些观测数据表成线性模型的形式.

1.2 某公司采用一项新技术试验以求提高产品质量.设在试验前,随机抽取的  $n_1$  件产品的质量指标值为  $y_1, \dots, y_{n_1}$ .它们可以看成来自正态总体  $N(\mu_1, \sigma^2)$  的一组随机样本.而在试验之后,随机抽取的  $n_2$  件产品的质量指标值为  $z_1, \dots, z_{n_2}$ .它们可以看成来自正态总体  $N(\mu_2, \sigma^2)$  的一组随机样本.为了考察这项新技术的效果,需要比较  $\mu_1$  和  $\mu_2$ .因此需要先估计它们.

(1) 试将这些数据表成线性模型的形式.

(2) 在实际问题中,如果  $z_1, \dots, z_{n_2}$  的值相比  $y_1, \dots, y_{n_1}$ ,有很大不同,往往它们的变异程度也就不同.于是我们不能再假定这两个正态总体有公共的方差.这时认为它们分别来自正态总体  $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$  比较适宜,试问这时(1)中所表示的线性模型应该有怎样的修正?

1.3 用两台仪器测量同一批材料的各 3 件样品的某种成分的含量.记测量值分别为  $y_{11}, y_{12}, y_{13}$  和  $y_{21}, y_{22}, y_{23}$ .由于两台仪器可能存在着性能上的差异,在表示这些数据时需要

考虑仪器的效应,记之为  $\alpha_1$  和  $\alpha_2$ ,试将这些测量数据表成某成分含量  $\mu$  和  $\alpha_1, \alpha_2$  的线性模型.

1.4 下面的模型是否表示一般线性模型?如果不是,能否通过适当的变换使之成为线性模型?

$$(1) y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \beta_3 \ln x_{i2} + e_i;$$

$$(2) y_i = e_i \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2);$$

$$(3) y_i = [1 + \exp(\beta_0 + \beta_1 x_{i1} + e_i)]^{-\frac{1}{2}};$$

$$(4) y_i = \beta_0 + \beta_1(x_{i1} + x_{i2}) + \beta_2 e^{x_{i1} + x_{i2}} + \beta_3 \ln(x_{i1}^2) + e_i.$$

1.5 考虑如下两因素设计模型:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad i = 1, \dots, I; j = 1, \dots, J,$$

其中  $\mu, \alpha_i, \beta_j$  为未知参数.试将其表示为矩阵形式的线性模型  $y = X\beta + e$ ,并写出其设计矩阵  $X$ .



## 第二章 随机向量

在第一章我们已经看到,当用矩阵形式来表示一个线性模型时,观测向量和误差向量都是随机向量.因此,在对线性模型做进一步讨论之前,有必要对随机向量,特别是正态随机向量的一些基本性质,作一些介绍.

### § 2.1 均值向量与协方差阵

设  $\mathbf{X} = (X_1, \dots, X_n)'$  为  $n \times 1$  随机向量.称

$$E(\mathbf{X}) = (EX_1, \dots, EX_n)'$$

为  $\mathbf{X}$  的均值向量.

**定理 2.1.1** 设  $\mathbf{A}$  为  $m \times n$  非随机矩阵,  $\mathbf{X}$  和  $\mathbf{b}$  分别为  $n \times 1$  和  $m \times 1$  随机向量,记  $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ , 则

$$E(\mathbf{Y}) = \mathbf{A}E(\mathbf{X}) + E(\mathbf{b}).$$

**证明** 设  $\mathbf{A} = (a_{ij})$ ,  $\mathbf{b} = (b_1, \dots, b_m)'$ ,  $\mathbf{Y} = (Y_1, \dots, Y_m)'$ , 于是

$$Y_i = \sum_{j=1}^n a_{ij}X_j + b_i, \quad i = 1, \dots, m.$$

求均值,得

$$E(Y_i) = E\left(\sum_{j=1}^n a_{ij}X_j + b_i\right) = \sum_{j=1}^n a_{ij}E(X_j) + E(b_i), \quad i = 1, \dots, m.$$

定理得证.

$n$  维随机向量  $\mathbf{X}$  的协方差阵定义为

$$\text{Cov}(\mathbf{X}) = E[(\mathbf{X} - E\mathbf{X})(\mathbf{X} - E\mathbf{X})'].$$

这是一个  $n \times n$  对称阵,它的  $(i, j)$  元为  $\text{Cov}(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)]$ ,特别当  $i = j$  时,就是  $X_i$  的方差  $\text{Var}(X_i)$ .所以  $\mathbf{X}$  的协方差阵的对角元为  $\mathbf{X}$  的分量的方差,而非对角元为相应分量的协方差.若对某个  $i$  和  $j$ ,  $\text{Cov}(X_i, X_j) = 0$ ,则称  $X_i$  与  $X_j$  是不相关的.

**推论 2.1.1**  $\text{tr}(\text{Cov}(\mathbf{X})) = \sum_{i=1}^n \text{Var}(X_i)$ , 这里  $\text{tr}\mathbf{A}$  表示方阵  $\mathbf{A}$  的迹,即对角元之和.

**定理 2.1.2** 设  $\mathbf{X}$  为任意  $n \times 1$  随机向量,则它的协方差阵是半正定的对称

阵.

**证明** 对称性是显然的. 下面证明它是半正定的. 事实上, 对任意  $n \times 1$  非随机向量  $c$ , 考虑随机变量  $Y = c'X$  的方差. 根据定义, 我们有

$$\begin{aligned}\text{Var}(Y) &= \text{Var}(c'X) = E[(c'X - Ec'X)^2], \\ &= E[(c'X - Ec'X)(c'X - Ec'X)] \\ &= c'E[(X - EX)(X - EX)']c \\ &= c'\text{Cov}(X)c.\end{aligned}$$

因为左端总是非负的, 于是对一切  $c$ , 右端也是非负的. 根据定义, 这说明矩阵  $\text{Cov}(X)$  是半正定的. 定理证毕.

**定理 2.1.3** 设  $A$  为  $m \times n$  阵,  $X$  为  $n \times 1$  随机向量,  $Y = AX$ , 则

$$\text{Cov}(Y) = A\text{Cov}(X)A'.$$

**证明** 依协方差阵的定义, 有

$$\begin{aligned}\text{Cov}(Y) &= E[(Y - EY)(Y - EY)'] \\ &= E[(AX - EAX)(AX - EAX)'] \\ &= AE[(X - EX)(X - EX)']A' \\ &= A\text{Cov}(X)A'.\end{aligned}$$

定理证毕.

设  $X$  和  $Y$  分别为  $n \times 1, m \times 1$  随机向量, 我们定义

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)'].$$

**定理 2.1.4** 设  $X$  和  $Y$  分别为  $n \times 1$  和  $m \times 1$  随机向量,  $A$  和  $B$  分别为  $p \times n$  和  $q \times m$  非随机矩阵, 则

$$\text{Cov}(AX, BY) = A\text{Cov}(X, Y)B'.$$

**证明** 
$$\begin{aligned}\text{Cov}(AX, BY) &= E[(AX - EAX)(BY - EBY)'] \\ &= AE[(X - EX)(Y - EY)']B' \\ &= A\text{Cov}(X, Y)B'.\end{aligned}$$

定理证毕.

## § 2.2 随机向量的二次型

假设  $X = (X_1, \dots, X_n)'$  为  $n \times 1$  随机向量,  $A$  为  $n \times n$  对称阵, 则随机变量

$$X'AX = \sum_{i=1}^n \sum_{j=1}^n a_{ij}X_iX_j$$

称为  $X$  的二次型. 本节只要求  $\text{Cov}(X)$  存在, 在对  $X$  的分布不做进一步假定的情况下, 给出它的均值计算公式, 这就是下面的定理 2.2.1. 如果  $X$  服从多元正态分布, 那末,  $X'AX$  还有进一步的性质, 这将在下一节讨论.

定理 2.2.1 设  $E(\mathbf{X}) = \boldsymbol{\mu}$ ,  $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ , 则

$$E(\mathbf{X}'\mathbf{A}\mathbf{X}) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \text{tr}(\mathbf{A}\boldsymbol{\Sigma}). \quad (2.2.1)$$

证明 因为

$$\begin{aligned} \mathbf{X}'\mathbf{A}\mathbf{X} &= (\mathbf{X} - \boldsymbol{\mu} + \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu} + \boldsymbol{\mu}) \\ &= (\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) + \boldsymbol{\mu}'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu}) \\ &\quad + (\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}\boldsymbol{\mu} + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}, \end{aligned} \quad (2.2.2)$$

利用定理 2.1.1, 有

$$E[\boldsymbol{\mu}'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] = E(\boldsymbol{\mu}'\mathbf{A}\mathbf{X}) - \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = \boldsymbol{\mu}'\mathbf{A}E(\mathbf{X}) - \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} = 0,$$

于是(2.2.2)式中第二、三两项的均值都等于零. 为了证明(2.2.1), 只需证明

$$E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] = \text{tr}(\mathbf{A}\boldsymbol{\Sigma}). \quad (2.2.3)$$

注意到

$$E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] = E[\text{tr}(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})],$$

利用矩阵迹的性质:  $\text{tr}\mathbf{A}\mathbf{B} = \text{tr}\mathbf{B}\mathbf{A}$ , 并交换求均值和求迹的次序, 上式变为

$$\begin{aligned} E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})] &= E[\text{tr}\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ &= \text{tr}E[\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ &= \text{tr}\mathbf{A}E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] \\ &= \text{tr}(\mathbf{A}\boldsymbol{\Sigma}). \end{aligned}$$

定理证毕.

注 在定理证明中, 我们应用了一个很重要的技巧. 这就是, 首先注意到二次型  $(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$  的迹就是它本身, 然后利用迹的可交换性  $\text{tr}\mathbf{A}\mathbf{B} = \text{tr}\mathbf{B}\mathbf{A}$ , 交换  $\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$  与  $(\mathbf{X} - \boldsymbol{\mu})'$  的位置, 最后再交换求  $E(\cdot)$  和  $\text{tr}(\cdot)$  的次序. 这样一来, 把求  $E[(\mathbf{X} - \boldsymbol{\mu})'\mathbf{A}(\mathbf{X} - \boldsymbol{\mu})]$  的问题归结为求协方差阵  $E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = \boldsymbol{\Sigma}$ . 这个技巧在后面的讨论中会多次用到.

推论 2.2.1 在定理假设条件下,

(1) 若  $\boldsymbol{\mu} = \mathbf{0}$ , 则

$$E(\mathbf{X}'\mathbf{A}\mathbf{X}) = \text{tr}(\mathbf{A}\boldsymbol{\Sigma});$$

(2) 若  $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ , 则

$$E(\mathbf{X}'\mathbf{A}\mathbf{X}) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \sigma^2\left(\sum_{i=1}^n a_{ii}\right) = \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} + \sigma^2\text{tr}\mathbf{A};$$

(3) 若  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\boldsymbol{\Sigma} = \mathbf{I}$ , 则

$$E(\mathbf{X}'\mathbf{A}\mathbf{X}) = \text{tr}\mathbf{A}.$$

例 2.2.1 假设一维总体的均值为  $\mu$ , 方差为  $\sigma^2$ .  $X_1, \dots, X_n$  为从此总体中抽取的随机样本, 试求样本方差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

的均值, 这里  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

解 记  $Q = (n-1)S^2$ ,  $\mathbf{X} = (X_1, \dots, X_n)'$ . 我们首先把  $Q$  表为  $\mathbf{X}$  的一个二次型. 用  $\mathbf{1}_n$  (在不会引起误解时也常用  $\mathbf{1}$ ) 表示所有元素为 1 的  $n$  维向量, 则  $E(\mathbf{X}) = \mu \mathbf{1}_n$ ,  $\text{Cov}(\mathbf{X}) = \sigma^2 \mathbf{I}_n$ . 另外

$$\bar{X} = \frac{1}{n} \mathbf{1}' \mathbf{X},$$

$$\mathbf{X} - \mathbf{X}\mathbf{1} = \mathbf{X} - \frac{1}{n} \mathbf{1}\mathbf{1}' \mathbf{X} = (\mathbf{I}_n - \frac{1}{n} \mathbf{1}\mathbf{1}') \mathbf{X} = \mathbf{C}\mathbf{X},$$

这里  $\mathbf{C} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}\mathbf{1}'$ , 这是一个幂等对称阵, 即  $\mathbf{C}^2 = \mathbf{C}$ ,  $\mathbf{C}' = \mathbf{C}$ . 于是

$$\begin{aligned} Q &= \sum_{i=1}^n (X_i - \bar{X})^2 = (\mathbf{X} - \bar{X}\mathbf{1})'(\mathbf{X} - \bar{X}\mathbf{1}) \\ &= (\mathbf{C}\mathbf{X})' \mathbf{C}\mathbf{X} = \mathbf{X}' \mathbf{C}\mathbf{X}. \end{aligned} \quad (2.2.4)$$

应用定理 2.2.1, 得

$$\begin{aligned} E(Q) &= (E(\mathbf{X}))' \mathbf{C} (E(\mathbf{X})) + \sigma^2 \text{tr} \mathbf{C} \\ &= \mu^2 \mathbf{1}' \mathbf{C} \mathbf{1} + \sigma^2 \text{tr} \mathbf{C}. \end{aligned}$$

容易验证

$$\mathbf{C}\mathbf{1} = \mathbf{0}, \quad \text{tr} \mathbf{C} = n - 1,$$

故有

$$E(Q) = \sigma^2(n - 1).$$

因而

$$E(S^2) = \sigma^2.$$

这就得到了所要的结论.

这个例子表明了初等数理统计中的一个重要事实: 不管总体的具体分布形式如何, 样本方差总是总体方差的一个无偏估计.

## § 2.3 正态随机向量

若随机变量  $X$  具有密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < +\infty,$$

则称  $X$  为具有均值  $\mu$ , 方差  $\sigma^2$  的正态随机变量, 记为  $N(\mu, \sigma^2)$ . 推广到多元情形, 我们可以作如下定义:

**定义 2.3.1** 设  $n$  维随机向量  $\mathbf{X} = (X_1, \dots, X_n)'$  具有密度函数

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} (\det \boldsymbol{\Sigma})^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \quad (2.3.1)$$

其中  $\mathbf{x} = (x_1, \dots, x_n)'$ ,  $-\infty < x_i < +\infty$ ,  $i = 1, \dots, n$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ ,  $\boldsymbol{\Sigma}$  是正定矩阵, 则称  $\mathbf{X}$  为  $n$  维正态随机向量, 记为  $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . 在不致引起混淆的情况下, 也简记为  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , 这里  $\boldsymbol{\mu}$  和  $\boldsymbol{\Sigma}$  分别为分布参数.

我们首先证明, 其中的参数  $\boldsymbol{\mu}$  为  $\mathbf{X}$  的均值向量,  $\boldsymbol{\Sigma}$  为  $\mathbf{X}$  的协方差阵. 在 (2.3.1) 中, 用到了  $\boldsymbol{\Sigma}^{-1}$ , 因此我们假定  $\boldsymbol{\Sigma}$  是正定阵, 记为  $\boldsymbol{\Sigma} > 0$ . 用  $\boldsymbol{\Sigma}^{\frac{1}{2}}$  记  $\boldsymbol{\Sigma}$  的平方根阵 (关于平方根阵的定义, 见附录), 记  $\boldsymbol{\Sigma}^{-\frac{1}{2}}$  为  $\boldsymbol{\Sigma}^{\frac{1}{2}}$  的逆矩阵, 即  $\boldsymbol{\Sigma}^{-\frac{1}{2}} = (\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}$ . 定义

$$\mathbf{Y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}),$$

故  $\mathbf{X} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Y} + \boldsymbol{\mu}$ . 于是  $\mathbf{Y}$  的密度函数为

$$g(\mathbf{y}) = f(\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu}) |J|,$$

这里  $J$  为变换的 Jacobi 行列式,

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \det(\boldsymbol{\Sigma}^{\frac{1}{2}}) = (\det \boldsymbol{\Sigma})^{\frac{1}{2}}.$$

从 (2.3.1) 得到  $\mathbf{Y}$  的密度函数

$$\begin{aligned} g(\mathbf{y}) &= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathbf{y}'\mathbf{y}} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}} = \prod_{i=1}^n f(y_i), \end{aligned}$$

这里

$$f(y_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_i^2}{2}}$$

是标准正态分布的密度函数. 这表明,  $\mathbf{Y}$  的  $n$  个分量的联合密度等于每个分量的密度函数的乘积. 于是,  $\mathbf{Y}$  的  $n$  个分量相互独立, 且  $Y_i \sim N(0, 1)$ ,  $i = 1, \dots, n$ . 因而有

$$E(\mathbf{Y}) = \mathbf{0}, \quad \text{Cov}(\mathbf{Y}) = \mathbf{I}.$$

利用关系  $\mathbf{X} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Y} + \boldsymbol{\mu}$  及定理 2.1.1 和定理 2.1.3, 得

$$E(\mathbf{X}) = \boldsymbol{\mu}, \quad \text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}.$$

这就完成了我们所要证明的。

从定义 2.3.1 可以看出,多元正态分布完全由它的均值向量  $\boldsymbol{\mu}$  和协方差阵  $\boldsymbol{\Sigma}$  所决定.特别,若  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\boldsymbol{\Sigma} = \mathbf{I}$ , 此时称其为多元标准正态分布.

设  $\mathbf{X}$  的协方差阵具有如下分块对角矩阵形式:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad (2.3.2)$$

这里  $\boldsymbol{\Sigma}_{11}$  为  $m \times m$  矩阵.相应地,将  $\mathbf{X}$  和  $\boldsymbol{\mu}$  也分块为

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad (2.3.3)$$

其中  $\mathbf{X}_1$  和  $\boldsymbol{\mu}_1$  皆为  $m \times 1$  矩阵.则(2.3.1)可分解为

$$f(\mathbf{x}) = f_1(x_1)f_2(x_2), \quad (2.3.4)$$

其中

$$f_1(x_1) = \frac{1}{(2\pi)^{\frac{m}{2}} \det \boldsymbol{\Sigma}_{11}} e^{-\frac{1}{2}(x_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_{11}^{-1} (x_1 - \boldsymbol{\mu}_1)},$$

$$f_2(x_2) = \frac{1}{(2\pi)^{\frac{n-m}{2}} \det \boldsymbol{\Sigma}_{22}} e^{-\frac{1}{2}(x_2 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_{22}^{-1} (x_2 - \boldsymbol{\mu}_2)}.$$

这表明  $\mathbf{X}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$ ,  $i=1, 2$  且相互独立.于是我们有如下定理.

**定理 2.3.1** (a) 设  $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , 且  $\mathbf{X}$  和  $\boldsymbol{\mu}$  分别有分块形式(2.3.3), 而  $\boldsymbol{\Sigma}$  具有形式(2.3.2), 则  $\mathbf{X}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$ ,  $i=1, 2$  且相互独立.

(b) 若  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ , 且记  $\mathbf{X} = (X_1, \dots, X_n)'$ ,  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ , 则  $\mathbf{X}_i \sim N(\mu_i, \sigma^2)$ ,  $i=1, \dots, n$ , 且相互独立.

这个定理的结论(a)刻画了正态分布的一个重要性质, 因为从(2.3.2)知  $\text{Cov}(\mathbf{X}_1, \mathbf{X}_2) = \boldsymbol{\Sigma}_{12} = \mathbf{0}$ , 所以  $X_1$  与  $X_2$  不相关. 结论(a)说明: 从  $\mathbf{X}_1$  和  $\mathbf{X}_2$  不相关, 可以推出  $\mathbf{X}_1$  和  $\mathbf{X}_2$  相互独立. 反过来, 我们知道, 随机变量相互独立, 一定是不相关的. 所以对正态向量而言, 相互独立与不相关是等价的.

**例 2.3.1** 二元正态分布.

从初等概率统计教科书我们已经知道, 二元正态分布密度为

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\}, \quad (2.3.5)$$

若写成(2.3.1)的形式, 则其中的  $\boldsymbol{\mu}$  和  $\boldsymbol{\Sigma}$  分别为

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

它们分别是二元正态向量的均值向量和协方差阵,  $\rho$  表示相关系数. 因为  $\det \boldsymbol{\Sigma} = (1 - \rho^2)\sigma_1^2\sigma_2^2$ , 所以为了保证  $\boldsymbol{\Sigma}$  可逆, 我们要求  $|\rho| < 1$ .

当  $\rho = 0$  时,  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2)$ , 依定理 2.3.1 知, 此时  $X_1$  与  $X_2$  相互独立, 且  $X_i \sim N(\mu_i, \sigma_i^2)$ . 关于这个事实, 我们也可以从密度函数得到证明. 当  $\rho = 0$  时, 它们的联合密度可分解为

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}} \\ &= f_1(x_1)f_2(x_2), \end{aligned}$$

这里

$$f_i(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}}, \quad i = 1, 2.$$

可见  $X_i \sim N(\mu_i, \sigma_i^2)$ , 且相互独立.

**定理 2.3.2** 设  $n$  维随机向量  $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{A}$  为  $n \times n$  非随机可逆阵,  $\boldsymbol{b}$  为  $n \times 1$  向量, 记  $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ , 则

$$\boldsymbol{Y} \sim N(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}').$$

**证明** 因为  $\boldsymbol{A}$  可逆, 于是  $\boldsymbol{X} = \boldsymbol{A}^{-1}(\boldsymbol{Y} - \boldsymbol{b})$ , 利用(2.3.1),  $\boldsymbol{Y}$  的密度函数为

$$g(\boldsymbol{y}) = f(\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b})) |J|,$$

这里  $J$  为变换的 Jacobi 行列式

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \det(\boldsymbol{A}^{-1}).$$

注意到

$$\begin{aligned} (\det \boldsymbol{\Sigma})^{\frac{1}{2}} |J|^{-1} &= (\det \boldsymbol{\Sigma} (\det \boldsymbol{A})^2)^{\frac{1}{2}} = (\det(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'))^{\frac{1}{2}}, \\ (\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{\mu}) \\ &= (\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}))' (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')^{-1} (\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b})), \end{aligned}$$

于是

$$\begin{aligned} g(\boldsymbol{y}) &= \frac{1}{(2\pi)^{n/2} (\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2} (\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{b}) - \boldsymbol{\mu})} |J| \\ &= \frac{1}{(2\pi)^{n/2} (\det \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')^{1/2}} e^{-\frac{1}{2} (\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}))' (\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')^{-1} (\boldsymbol{y} - (\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}))}. \end{aligned}$$

这正是  $N(A\mu + b, A\Sigma A')$  的分布密度, 定理证毕.

这个定理表明, 多元正态向量经过可逆线性变换仍为多元正态向量, 这个事实还可以推广到一般线性变换的情形(见下面的定理 2.3.4).

在定理中, 若取  $A = \Sigma^{-\frac{1}{2}}$ , 注意到  $\Sigma^{-1}$  也是对称阵, 因此

$$\Sigma^{-\frac{1}{2}}\Sigma(\Sigma^{-\frac{1}{2}})' = \Sigma^{-\frac{1}{2}}\Sigma\Sigma^{-\frac{1}{2}} = I,$$

于是我们有如下推论:

**推论 2.3.1** 设  $X \sim N(\mu, \Sigma)$ , 则  $Y = \Sigma^{-\frac{1}{2}}X \sim N(\Sigma^{-\frac{1}{2}}\mu, I)$ .

注意, 这里  $X$  的诸分量可以是彼此相关且方差互不相等, 但变换过的  $Y$  的诸分量相互独立, 且方差皆为 1. 这个推论表明, 我们可以用一个线性变换把诸分量相关且方差不等的多元正态向量变换为多元标准正态向量.

**推论 2.3.2** 设  $X \sim N_n(\mu, \sigma^2 I)$ ,  $Q$  为  $n \times n$  正交阵, 则  $QX \sim N_n(Q\mu, \sigma^2 I)$ .

这个推论的证明是容易的, 留给读者作练习.

本推论表明, 诸分量相互独立且具有等方差的正态向量, 经过正交变换后, 变为诸分量仍然相互独立且具有等方差的正态向量.

以上两个推论所给出的正态向量的性质很重要, 在本书后面的讨论中多次会用到.

**定理 2.3.3** 设  $X \sim N_n(\mu, \Sigma)$ , 将  $X, \mu, \Sigma$  分块为

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

其中  $X_1$  和  $\mu_1$  为  $m \times 1$  向量, 而  $\Sigma_{11}$  为  $m \times m$  矩阵, 则  $X_1 \sim N_m(\mu_1, \Sigma_{11})$ .

**证明** 在定理 2.3.2 中, 取

$$A = \begin{bmatrix} I_m & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{n-m} \end{bmatrix}, b = 0,$$

则  $Y = AX \sim N_n(A\mu, A\Sigma A')$ .

因为

$$A\Sigma A' = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{bmatrix},$$

其中  $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ , 于是

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 - \Sigma_{21}\Sigma_{11}^{-1}X_1 \end{bmatrix} \sim N \left[ \begin{bmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{bmatrix} \right],$$

根据定理 2.3.1(a) 知  $X_1 \sim N_m(\mu_1, \Sigma_{11})$ , 定理得证.

类似地, 我们也可以证明  $X_2 \sim N(\mu_2, \Sigma_{22})$ . 这说明, 若  $X$  服从多元正态分



布,将其砍成两段后,各段仍服从正态分布.这称为正态分布的“再生性”.犹如一条蚯蚓,砍成两段后,仍各成一条蚯蚓,这就是所谓的蚯蚓的再生性.事实上,我们不难证明,更一般的结论也成立,即正态向量  $X$  的任意维数的子向量也是正态向量.设  $i_1 < i_2 < \cdots < i_k$ , 则  $(X_{i_1}, \cdots, X_{i_k})' \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ , 这里

$$\boldsymbol{\mu}_0 = (\mu_{i_1}, \cdots, \mu_{i_k})',$$

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_{i_1 i_1} & \cdots & \sigma_{i_1 i_k} \\ \vdots & & \vdots \\ \sigma_{i_k i_1} & \cdots & \sigma_{i_k i_k} \end{pmatrix}.$$

**定理 2.3.4** 设  $X \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $A$  为  $m \times n$  矩阵,且秩为  $m (< n)$ , 则  $Y = AX \sim N_m(A\boldsymbol{\mu}, A\boldsymbol{\Sigma}A')$ .

**证明** 因为  $A$  的秩为  $m$ , 故可将其扩充为  $n \times n$  可逆阵

$$C = \begin{pmatrix} A \\ B \end{pmatrix}.$$

应用定理 2.3.2,

$$Z = CX = \begin{pmatrix} A \\ B \end{pmatrix} X \sim N\left(\begin{pmatrix} A\boldsymbol{\mu} \\ B\boldsymbol{\mu} \end{pmatrix}, \begin{pmatrix} A\boldsymbol{\Sigma}A' & A\boldsymbol{\Sigma}B' \\ B\boldsymbol{\Sigma}A' & B\boldsymbol{\Sigma}B' \end{pmatrix}\right),$$

再应用定理 2.3.3 知

$$Y = AX \sim N_m(A\boldsymbol{\mu}, A\boldsymbol{\Sigma}A').$$

定理证毕.

在定理 2.3.4 中,若  $A$  为一个非零行向量,就得到如下推论:

**推论 2.3.3** 设  $X \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $c$  为  $n \times 1$  非零向量, 则

$$c'X \sim N(c'\boldsymbol{\mu}, c'\boldsymbol{\Sigma}c).$$

于是我们证明了,多元正态向量的任一线性组合是一元正态变量.

**例 2.3.2** 设  $X_1, \cdots, X_n$  为从正态总体  $N(\mu, \sigma^2)$  抽取的简单随机样本. 则

$$\text{样本均值 } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

事实上,若记  $X = (X_1, X_2, \cdots, X_n)'$ ,  $c = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right)'$ , 则  $\bar{X} = c'X$ . 依推论 2.3.3 知  $\bar{X}$  服从正态分布. 其余结论的证明是容易的,留给读者作练习.

在推论 2.3.3 中,取  $c' = (0, \cdots, 0, 1, 0, \cdots, 0)$ , 则  $c'X = X_i$ ,  $c'\boldsymbol{\mu} = \mu_i$ ,  $c'\boldsymbol{\Sigma}c = \sigma_{ii}$ . 于是我们有如下推论.

**推论 2.3.4** 设  $X \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)'$ ,  $\boldsymbol{\Sigma} = (\sigma_{ij})$ , 则  $X_i \sim N(\mu_i, \sigma_{ii})$ ,  $i = 1, \cdots, n$ .

这个推论表明,若  $\mathbf{X} = (X_1, \dots, X_n)'$  为  $n$  维正态向量,则它的任一分量也是正态变量.但反过来的结论未必成立.即  $X_1, \dots, X_n$  均为正态变量, $\mathbf{X} = (X_1, \dots, X_n)'$  未必为正态向量.我们可以举出很多这样的例子.下面就是其中的一个.

**例 2.3.3** 设  $(X, Y)$  的联合密度为

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \left[ 1 - \frac{xy}{(x^2+1)(y^2+1)} \right],$$

$$-\infty < x < +\infty, -\infty < y < +\infty,$$

显然这不是二元正态分布的密度.而  $X$  和  $Y$  的边缘分布均为  $N(0, 1)$ .

事实上

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, y) dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dy - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{xy}{(x^2+1)(y^2+1)} e^{-\frac{1}{2}(x^2+y^2)} dy. \end{aligned}$$

上式第二项被积函数对固定的  $x$ , 是  $y$  的奇函数,因此第二项积分等于零.于是

$$\begin{aligned} f_1(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \end{aligned}$$

这里利用了 
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

这就证明了  $X \sim N(0, 1)$ , 在  $f(x, y)$  表达式中,  $x, y$  的地位完全一样,故  $Y \sim N(0, 1)$  也成立.

## § 2.4 $\chi^2$ 分布

在初等概率统计中我们已经知道,若  $X_1, \dots, X_n$  相互独立,都服从正态分布  $N(0, 1)$ , 则  $X = X_1^2 + \dots + X_n^2$  服从自由度为  $n$  的卡方分布,记为  $X \sim \chi_n^2$ .

可以证明,自由度为  $n$  的  $\chi^2$  分布密度函数为

$$g(x) = \begin{cases} \frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} e^{-\frac{x}{2}} x^{\frac{n}{2}-1}, & \text{若 } x > 0, \\ 0, & \text{若 } x \leq 0. \end{cases}$$

这个事实的证明在一般概率统计教材中都可以找到.

**定理 2.4.1** 设  $\mathbf{X} \sim N_n(\mathbf{0}, \Sigma)$ ,  $\Sigma$  为正定阵, 则  $\mathbf{X}'\Sigma^{-1}\mathbf{X} \sim \chi_n^2$ .

**证明** 记  $\mathbf{Y} = \Sigma^{-\frac{1}{2}}\mathbf{X}$ . 依推论 2.3.1 知,  $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{I})$ . 但

$$\mathbf{X}'\Sigma^{-1}\mathbf{X} = (\Sigma^{-\frac{1}{2}}\mathbf{X})'\Sigma^{-\frac{1}{2}}\mathbf{X} = \mathbf{Y}'\mathbf{Y},$$

依定义知  $\mathbf{X}'\Sigma^{-1}\mathbf{X} \sim \chi_n^2$ . 定理证毕.

**定理 2.4.2** 设  $X \sim \chi_n^2$ , 则  $E(X) = n$ ,  $\text{Var}(X) = 2n$ .

**证明** 设  $X_1, \dots, X_n$  相互独立, 且  $X_i \sim N(0, 1)$ ,  $i = 1, \dots, n$ . 令

$$X = X_1^2 + \dots + X_n^2,$$

于是

$$E(X) = \sum_{i=1}^n E(X_i^2), \quad (2.4.1)$$

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i^2). \quad (2.4.2)$$

问题归结为求  $E(X_i^2)$  和  $\text{Var}(X_i^2)$ . 因为  $E(X_i) = 0$  所以

$$E(X_i^2) = E(X_i^2) - (EX_i)^2 = \text{Var}(X_i) = 1.$$

按定义直接计算可证明  $E(X_i^4) = 3$ , 于是

$$\text{Var}(X_i^2) = E(X_i^4) - (EX_i^2)^2 = 3 - 1 = 2.$$

代入(2.4.1)和(2.4.2), 命题得证.

下面的定理刻画了  $\chi^2$  分布的所谓“可加性”.

**定理 2.4.3** 设  $X_1$  和  $X_2$  相互独立, 且  $X_1 \sim \chi_n^2$ ,  $X_2 \sim \chi_m^2$ . 则  $X_1 + X_2 \sim \chi_{n+m}^2$ .

**证明** 依  $\chi^2$  分布的定义,  $X_1$  和  $X_2$  可分别写为

$$X_1 = Y_1^2 + \dots + Y_n^2,$$

$$X_2 = Y_{n+1}^2 + \dots + Y_{n+m}^2,$$

这里  $Y_1, \dots, Y_{n+m}$  相互独立, 且有分布  $N(0, 1)$ . 因而

$$X_1 + X_2 = Y_1^2 + \dots + Y_{n+m}^2,$$

于是  $X_1 + X_2$  表成了  $n + m$  个相互独立的标准正态变量的平方和, 于是  $X_1 + X_2 \sim \chi_{n+m}^2$ . 定理证毕.

假设  $\mathbf{X} = (X_1, \dots, X_n)'$ , 若  $X_i \sim N(0, 1)$ , 且相互独立, 则  $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$ . 根据定义知二次型  $\mathbf{X}'\mathbf{X} \sim \chi_n^2$ . 但是在后面的讨论中, 我们常常碰到一般二次型  $\mathbf{X}'\mathbf{A}\mathbf{X}$ , 这里  $\mathbf{A}$  是一个对称阵. 下面的定理给出了这样的二次型服从  $\chi^2$  分布的

充分条件.

**定理 2.4.4** 设  $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ ,  $\mathbf{A}$  为  $n \times n$  对称阵, 秩为  $r$ , 则当  $\mathbf{A}$  满足:  $\mathbf{A}^2 = \mathbf{A}$ , 即  $\mathbf{A}$  为幂等阵时, 二次型

$$\mathbf{X}'\mathbf{A}\mathbf{X} \sim \chi_r^2. \quad (2.4.3)$$

**证明** 因为  $\mathbf{A}$  为幂等阵, 所以  $\mathbf{A}$  的特征值非零即 1. 因而  $\mathbf{A}$  的特征值中有  $r$  个为 1,  $n-r$  个为 0 (见附录). 再由  $\mathbf{A}$  的对称性, 知存在一个正交阵  $\mathbf{Q}$ , 使得

$$\mathbf{A} = \mathbf{Q}' \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}, \quad (2.4.4)$$

命  $\mathbf{Y} = \mathbf{Q}\mathbf{X}$ , 则  $\mathbf{Y} \sim N(\mathbf{0}, \mathbf{I}_n)$ . 于是  $\mathbf{Y} = (Y_1, \dots, Y_n)'$  的分量  $Y_i \sim N(0, 1)$ , 且相互独立, 而

$$\begin{aligned} \mathbf{X}'\mathbf{A}\mathbf{X} &= \mathbf{X}'\mathbf{Q}' \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}\mathbf{X} \\ &= Y_1^2 + \dots + Y_r^2, \end{aligned}$$

这就证明了  $\mathbf{X}'\mathbf{A}\mathbf{X} \sim \chi_r^2$ . 定理证毕.

**例 2.4.1** 设  $X_1, \dots, X_n$  为取自  $N(\mu, \sigma^2)$  的随机样本, 根据例 2.2.1, 我们知道样本方差

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

为  $\sigma^2$  的一个无偏估计. 这里  $\bar{X}$  为样本均值. 现在我们证明  $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ . 记  $\mathbf{X} = (X_1, \dots, X_n)$ , 则  $\mathbf{X} \sim N(\mu\mathbf{1}, \sigma^2\mathbf{I})$ , 其中  $\mathbf{1} = (1, \dots, 1)'$ , 即  $n$  个分量皆为 1 的向量. 则

$$\mathbf{Y} = \frac{(\mathbf{X} - \mu\mathbf{1})}{\sigma} \sim N(\mathbf{0}, \mathbf{I}_n).$$

记

$$\mathbf{C} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}',$$

这是秩为  $n-1$  的幂等阵, 利用(2.2.3), 我们有

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\mathbf{X}'\mathbf{C}\mathbf{X}}{\sigma^2} = \mathbf{Y}'\mathbf{C}\mathbf{Y},$$

再由定理 2.4.4 知,  $\mathbf{Y}'\mathbf{C}\mathbf{Y} \sim \chi_{n-1}^2$ . 这就证明了所要的结论.

**定理 2.4.5** 设  $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I}_n)$ ,  $\mathbf{A}$  为  $n \times n$  对称阵,  $\mathbf{B}$  为  $m \times n$  阵. 若  $\mathbf{B}\mathbf{A} = \mathbf{0}$ , 则  $\mathbf{B}\mathbf{X}$  与  $\mathbf{X}'\mathbf{A}\mathbf{X}$  相互独立.

**证明** 因为  $\mathbf{A}$  为对称阵, 故存在正交阵  $\mathbf{Q}$ , 使得

$$\mathbf{A} = \mathbf{Q} \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Q}',$$

其中  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r, \lambda_{r+1}, \dots, \lambda_n)$ ,  $\lambda_i, i=1, \dots, r$ , 为  $\Lambda$  的非零特征根,  $r$  为  $\Lambda$  的秩. 将  $Q$  分块为  $Q = (Q_1 : Q_2)$ , 其中  $Q_1$  为  $n \times r$  阵, 做变换

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = Q'X,$$

于是  $Y_i = Q_i'X, i=1, 2$ . 依推论 2.3.2 有  $Y \sim N(\mathbf{0}, I)$ . 所以,  $Y_1 \sim N_r(\mathbf{0}, I_r), Y_2 \sim N_{n-r}(\mathbf{0}, I_{n-r})$ , 且  $Y_1$  与  $Y_2$  相互独立. 注意到

$$X'AX = X'Q_1\Lambda Q_1'X = Y_1'\Lambda Y_1, \quad (2.4.5)$$

$$BX = BQY = DY, \quad (2.4.6)$$

这里  $D = BQ$ . 因为  $BA = \mathbf{0}$ , 于是

$$BQQ'AQ = DQ'AQ = D \begin{pmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{0}.$$

将  $D$  分块为  $D = (D_1 : D_2)$ , 其中  $D_1$  为  $m \times r$  矩阵. 由上式知  $D_1 = \mathbf{0}$ . 代入 (2.4.6) 式, 知

$$BX = D_2 Y_2. \quad (2.4.7)$$

由  $Y_1$  和  $Y_2$  的独立性, 从 (2.4.5) 和 (2.4.7) 便可推出  $X'AX$  与  $BX$  独立. 定理证毕.

**例 2.4.2** 设  $X_1, \dots, X_n$  为取自  $N(0, 1)$  的随机样本, 则样本均值  $\bar{X}$  与样本方差  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  相互独立.

事实上, 若记  $\mathbf{X} = (X_1, \dots, X_n), \mathbf{1} = (1, \dots, 1)'$ , 即  $\mathbf{1}$  为所有分量皆为 1 的  $n \times 1$  向量, 则

$$\bar{X} = \frac{1}{n} \mathbf{1}'\mathbf{X}.$$

由例 2.4.1 知

$$(n-1)S^2 = \mathbf{X}'\mathbf{C}\mathbf{X},$$

这里

$$\mathbf{C} = I_n - \frac{1}{n} \mathbf{1}\mathbf{1}'.$$

容易验证  $\mathbf{1}'\mathbf{C} = \mathbf{0}$ , 由定理 2.4.5 知  $\bar{X}$  与  $S^2$  独立.

**定理 2.4.6** 设  $\mathbf{X} \sim N_n(\mathbf{0}, I)$ ,  $\mathbf{A}$  和  $\mathbf{B}$  皆为  $n$  阶对称阵, 且  $\mathbf{AB} = \mathbf{0}$ , 则二次型  $\mathbf{X}'\mathbf{A}\mathbf{X}$  与  $\mathbf{X}'\mathbf{B}\mathbf{X}$  相互独立.

**证明** 因为  $\mathbf{AB} = \mathbf{0}$ , 结合  $\mathbf{A}$  与  $\mathbf{B}$  的对称性, 容易推出  $\mathbf{BA} = \mathbf{0}$ , 于是  $\mathbf{AB} = \mathbf{BA}$ , 即  $\mathbf{A}, \mathbf{B}$  可交换, 因而可用一个正交阵将这两个矩阵同时对角化 (见王松桂等 (1994), p. 14). 即存在正交阵  $Q$ , 使得

$$A = QA_1Q', B = QA_2Q',$$

其中  $A_i = \text{diag}(\lambda_1^{(i)}, \dots, \lambda_n^{(i)}), i = 1, 2.$

由  $AB = \mathbf{0}$ , 可推出  $A_1A_2 = \mathbf{0}$ , 于是

$$\lambda_j^{(1)}\lambda_j^{(2)} = 0, j = 1, \dots, n. \quad (2.4.8)$$

命  $Y = Q'X$ , 则  $Y \sim N(\mathbf{0}, I)$ . 于是  $Y$  的所有分量都相互独立. 另一方面

$$X'AX = Y'A_1Y,$$

$$X'BX = Y'A_2Y,$$

由(2.4.8)知,  $X'AX$  与  $X'BX$  依赖于  $Y$  的不同分量, 因而相互独立, 定理证毕.

多元正态分布是许多统计推断的基础, 因此在数理统计的许多分支都有重要应用. 关于多元正态分布的更详细的讨论, 读者可以参阅王松桂(1987), Wang等(1994)或 Tong(1990).

## 习题二

2.1 设  $X$  为  $n \times p$  随机矩阵,  $A$  为  $p \times n$  常数矩阵, 证明

$$E\text{tr}(AX) = \text{tr}E(AX) = \text{tr}[AE(X)].$$

2.2 设  $X_1, \dots, X_n$  为随机变量,  $Y_1 = X_1, Y_i = X_i - X_{i-1}, i = 2, 3, \dots, n$ . 记  $X = (X_1, X_2, \dots, X_n)'$ ,  $Y = (Y_1, Y_2, \dots, Y_n)'$ .

(1) 若  $\text{Cov}(X) = I$ , 其中  $I$  是  $n$  阶单位阵, 求  $\text{Cov}(Y)$ ;

(2) 若  $\text{Cov}(Y) = I$ , 求  $\text{Cov}(X)$ .

2.3 (1) 设  $X$  和  $Y$  为具有相同方差的任意两个随机变量. 证明  $\text{Cov}(X + Y, X - Y) = 0$ .

(2) 设  $X_{n \times 1}, Y_{m \times 1}$  均为随机向量,  $\text{Cov}(X) > 0$  (即  $\text{Cov}(X)$  是正定矩阵), 求常数矩阵  $A_{n \times m}$ , 使得

$$\text{Cov}(X, Y - AX) = \mathbf{0}.$$

(3) 利用(1)和(2), 试构造例子说明不相关的随机变量或向量不一定相互独立.

2.4 设随机变量  $X_1, \dots, X_n$  相互独立, 具有公共均值  $\theta$  和方差  $\sigma^2$ .

(1) 定义  $Y_i = X_i - X_{i-1}, i = 1, \dots, n-1$ . 证明  $Y_i$  均值为  $\theta$ , 方差为  $2\sigma^2$ ;

(2) 定义  $Q = (X_1 - X_2)^2 + \dots + (X_{n-1} - X_n)^2$ , 求  $E(Q)$ .

2.5 设随机向量  $X$  的密度函数为(2.3.1), 且  $\Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_{nn})$ . 证明  $x_1, \dots, x_n$  相互独立, 且  $X_i \sim N(\mu_i, \sigma_{ii}^2), i = 1, 2, \dots, n$ .

2.6 设二维随机向量  $X = (X_1, X_2)'$  具有密度(2.3.5), 证明  $X_1$  与  $X_2$  的相关系数为  $\rho$ .

2.7 设  $X_i \sim N_p(\mu_i, \Sigma_i), i = 1, 2$ , 相互独立.

(1) 证明

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{2p}(\mu, \Sigma),$$

其中  $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ ,  $\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix}$ ;

(2) 设  $a, b$  为常数, 证明

$$aX_1 + bX_2 \sim N_p(a\mu_1 + b\mu_2, a^2\Sigma_1 + b^2\Sigma_2).$$

2.8 设  $X_1, \dots, X_n$  相互独立, 且均服从  $N(0, \sigma^2)$ . 证明

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ 与 } \sum_{i=1}^{n-1} (X_i - X_{i+1})^2$$

相互独立.

2.9 设  $X \sim N_n(\mathbf{0}, I)$ . 令  $U = AX, V = BX, W = CX$ , 这里  $A, B, C$  皆为  $r \times n$  矩阵, 且秩为  $r$ , 若  $\text{Cov}(U, V) = \text{Cov}(U, W) = \mathbf{0}$ . 证明  $U$  与  $V + W$  独立.

2.10 设线性模型  $Y = X\beta + e, E(e) = \mathbf{0}, \text{Cov}(e) = \sigma^2 V$ . 若要  $Y'AY$  为  $\sigma^2$  的无偏估计 ( $A$  是非随机矩阵),  $A$  应满足什么条件?

2.11 若  $X$  和  $Y$  皆为  $n \times 1$  随机向量,  $A$  和  $B$  为  $m \times n$  非随机矩阵. 记  $Z = AX + BY$ . 证明

$$\text{Cov}(Z) = A\text{Cov}(X)A' + B\text{Cov}(Y)B' + B\text{Cov}(Y, X)A' + A\text{Cov}(X, Y)B'.$$

2.12 设  $X = (X_1, X_2, X_3)' \sim N_3(\mu, \Sigma)$ , 其中

$$\mu = (2, 1, 2)', \quad \Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

求  $Y_1 = X_1 + X_2 + X_3$  与  $Y_2 = X_1 - X_2$  的联合分布.

2.13 设  $X = (X_1, X_2, X_3)' \sim N_3(\mu, \Sigma)$ , 其中

$$\Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}.$$

试问:  $\rho$  取什么值时,  $X_1 + X_2 + X_3$  与  $X_1 - X_2 - X_3$  独立?

2.14 已知  $X \sim N_n(\mu, I)$ .

(1) 求  $Y_1 = \alpha'X$  和  $Y_2 = \beta'X$  的联合分布, 其中  $\alpha, \beta$  皆为  $n \times 1$  非随机向量 ( $\alpha \neq k\beta$ );

(2) 若  $\alpha'\beta = 0$ , 证明  $Y_1$  与  $Y_2$  独立.

2.15 设  $X \sim N_n(\mu, \Sigma)$ ,  $A$  是  $n \times n$  对称矩阵, 证明当  $A\Sigma A = A$  时,  $(X - \mu)'A(X - \mu) \sim \chi_r^2$ , 其中  $r$  为矩阵  $A$  的秩.

2.16 若  $X \sim N_n(\mu, \Sigma)$ , 试给出二次型  $(X - \mu)'A(X - \mu)$  与  $(X - \mu)'B(X - \mu)$  独立的条件.

2.17 设  $X$  是二元正态分布, 其密度函数

$$f(x_1, x_2) = k^{-1} \exp \left\{ -\frac{1}{2} (x_1^2 + 2x_2^2 - x_1x_2 - 3x_1 - 2x_2 + 4) \right\},$$

求  $E(X)$  及  $\text{Cov}(X)$ .

2.18 设  $X \sim N_p(\mu, \Sigma)$ . 证明  $X$  的特征函数为

$$E(e^{t'X}) = e^{t'\mu + \frac{1}{2}t'\Sigma t},$$

其中  $t' = (t_1, \dots, t_p)$ ,  $i = \sqrt{-1}$ .

**2.19** 利用特征函数证明:  $p$  维随机向量  $\mathbf{X}$  为  $p$  元正态分布, 当且仅当对一切  $p$  维实向量  $\alpha$  ( $\alpha \neq \mathbf{0}$ ),  $\alpha'X$  是一元正态分布.

**2.20** 利用二次型的均值及二元正态分布, 计算积分

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x^2 + xy + 3y^2) \exp[-(x^2 + 2xy + 2y^2)] dx dy.$$



## 第三章 回归参数的估计

估计回归参数的最基本方法是最小二乘法. 这个方法不仅仅在统计学中, 就是在数学的其它分支, 例如运筹学、计算数学、逼近论和控制论等, 都是很重要的求解方法. 本章前三节我们首先讨论如何应用最小二乘法求回归参数的最小二乘估计, 并研究这种估计的基本性质. 在 § 3.4 和 § 3.5 两节, 我们讨论当有了一批实际数据之后, 如何考察导出最小二乘估计的基本假设的适用性以及当这些假设不适用时, 对数据应该做的变换, 即 Box-Cox 变换. 在 § 3.6 我们讨论了广义最小二乘估计. 在一些情况下, 最小二乘估计并不是一个好的估计, 于是统计学家们提出了一些新的估计. 在本章最后两节, 我们将讨论其中最重要的两种估计——岭估计和主成分估计.

### § 3.1 最小二乘估计

假设  $Y$  为因变量,  $X_1, \dots, X_{p-1}$  为对  $Y$  有影响的  $p-1$  个自变量, 并且它们之间具有线性关系

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e, \quad (3.1.1)$$

其中  $e$  为误差项, 它表示除了  $X_1, \dots, X_{p-1}$  之外其它因素对  $Y$  的影响以及试验或测量误差.  $\beta_0, \beta_1, \dots, \beta_{p-1}$  是待估计的未知参数. 假定我们有了因变量  $Y$  和自变量  $X_1, \dots, X_{p-1}$  的  $n$  组观测值

$$(x_{i1}, \dots, x_{i,p-1}, y_i), \quad i = 1, \dots, n,$$

它们满足

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{i,p-1}\beta_{p-1} + e_i, \quad i = 1, \dots, n \quad (3.1.2)$$

误差项  $e_i, i = 1, \dots, n$  满足如下假设:

- (a)  $E(e_i) = 0,$
- (b)  $\text{Var}(e_i) = \sigma^2,$  (等方差) (3.1.3)
- (c)  $\text{Cov}(e_i, e_j) = 0, \quad i \neq j.$  (不相关)

这就是所谓的 Gauss-Markov 假设. 在 § 1.1 我们已经解释了这些假设的实际意义. 若用矩阵形式, (3.1.2) 变形为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,p-1} \\ 1 & x_{21} & \cdots & x_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix},$$

等价地

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (3.1.4)$$

其中所有矩阵和向量的定义是不言自明的,  $\mathbf{y}$  是  $n \times 1$  的变量观测向量,  $\mathbf{X}$  为  $n \times p$  的已知设计矩阵,  $\boldsymbol{\beta}$  为  $p \times 1$  未知参数向量,  $\mathbf{e}$  为随机误差向量. 用矩阵形式可将 Gauss-Markov 假设(3.1.3)写成

$$E(\mathbf{e}) = \mathbf{0}, \quad \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}_n. \quad (3.1.5)$$

将(3.1.4)和(3.1.5)合并在一起, 就得到最基本、最重要的线性回归模型

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad E(\mathbf{e}) = \mathbf{0}, \quad \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}_n. \quad (3.1.6)$$

获得参数向量  $\boldsymbol{\beta}$  的估计的一个最重要方法是最小二乘法. 这个方法是找  $\boldsymbol{\beta}$  的估计, 使得偏差向量  $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$  的长度之平方  $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$  达到最小. 记

$$Q(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

将此式展开,

$$Q(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}.$$

对  $\boldsymbol{\beta}$  求偏导数, 并命其为零, 可以得到方程组

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}. \quad (3.1.7)$$

它称为正则方程. 这个线性方程组有唯一解的充要条件是  $\mathbf{X}'\mathbf{X}$  的秩为  $p$ . 等价地,  $\mathbf{X}$  的秩为  $p$ . 以后, 在线性回归模型的讨论中, 我们总假定这个条件满足. 于是我们得到(3.1.7)的唯一解

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (3.1.8)$$

根据微积分的极值理论,  $\hat{\boldsymbol{\beta}}$  只是函数  $Q(\boldsymbol{\beta})$  的一个驻点. 我们还需要证明,  $\hat{\boldsymbol{\beta}}$  确实使  $Q(\boldsymbol{\beta})$  达到最小. 事实上, 对任意一个  $\boldsymbol{\beta}$ , 有

$$\begin{aligned} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 &= \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|^2 \\ &= \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &\quad + 2(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \end{aligned} \quad (3.1.9)$$

因为  $\hat{\boldsymbol{\beta}}$  满足正则方程(3.1.7), 于是  $\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}$ , 因而上式第三项等于零. 这就证明了对任意的  $\boldsymbol{\beta}$ , 有

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \quad (3.1.10)$$

又因为  $\mathbf{X}'\mathbf{X}$  是一个正定阵, 故上式第二项总是非负的, 于是

$$Q(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \geq \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 = Q(\hat{\boldsymbol{\beta}}),$$

且等号成立当且仅当

$$(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) = 0.$$

注 关于最小二乘法的发明权,在数学史的研究中尚存在一些争议.但目前比较普遍接受的想法是,两位数学家 Gauss(1777—1855)和 Legendre(1752—1833)独立地提出了这种方法. Legendre 是 1805 年第一次公开发表关于最小二乘法的材料,这时 Gauss 指出,他早在 1795 年之前就使用了这个方法,但数学史研究者只找到了 Gauss 约在 1803 年之前使用这个方法的证据.对此问题感兴趣的读者可参阅 Stigler, S. M. “Gauss and the invention of least squares”, *Annals of Statistics*, 9(1981) 465—474 以及所引文献.

记  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{p-1})'$  并将其代入(3.1.1). 去掉误差项,得到

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_{p-1} X_{p-1}, \quad (3.1.11)$$

称为经验线性回归方程. 这个方程是不是描述了  $Y$  与自变量  $X_1, \dots, X_{p-1}$  的真实关系,还需要做进一步的统计分析. 这将在后续章节讨论.

### 例 3.1.1 一元线性回归.

假设影响因变量  $Y$  的因素只有一个,记为  $X$ . 现在我们对  $Y$  和  $X$  获得了  $n$  次观测  $(y_i, x_i), i = 1, \dots, n$ , 于是我们有

$$y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, n.$$

这时,正则方程(3.1.7)变为

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}, \quad (3.1.12)$$

当  $x_i, i = 1, \dots, n$  不全相等时,  $\sum_i (x_i - \bar{x})^2 \neq 0$ . 这里  $\bar{x} = \sum x_i / n$ . 于是正则方程(3.1.12)左端的系数行列式  $= n \sum (x_i - \bar{x})^2 \neq 0$ . 经过初等计算可以解得  $\alpha$  和  $\beta$  的最小二乘估计分别为

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x},$$

$$\hat{\beta} = \frac{\sum x_i y_i - \sum x_i \sum y_i}{\sum x_i^2 - n \bar{x}^2},$$

其中  $\bar{y} = \sum y_i / n$ .

在回归分析的应用中,我们常常要把原始观测数据进行中心化和标准化,这对于我们的统计分析将是有益的. 记

$$x_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = 1, \dots, p-1.$$

$\bar{x}_j$  就是第  $j$  个回归自变量  $n$  次取值的平均值. 将(3.1.2)改写为

$$y_i = \alpha + (x_{i1} - \bar{x}_1)\beta_1 + \cdots + (x_{i,p-1} - \bar{x}_{p-1})\beta_{p-1} + e_i, i = 1, \cdots, n. \quad (3.1.13)$$

这里  $\alpha$  与(3.2)中的  $\beta_0$  有如下关系:

$$\alpha = \beta_0 + \bar{x}_1\beta_1 + \cdots + \bar{x}_{p-1}\beta_{p-1}.$$

在(3.1.13)中,我们把每个回归自变量减去了它们的平均值,这称为中心化.若记

$$\mathbf{X}_c = \begin{bmatrix} x_{11} - \bar{x}_1 & x_{12} - \bar{x}_2 & \cdots & x_{1,p-1} - \bar{x}_{p-1} \\ x_{21} - \bar{x}_1 & x_{22} - \bar{x}_2 & \cdots & x_{2,p-1} - \bar{x}_{p-1} \\ \vdots & \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & x_{n2} - \bar{x}_2 & \cdots & x_{n,p-1} - \bar{x}_{p-1} \end{bmatrix} \quad (3.1.14)$$

则(3.1.13)可改写为

$$\mathbf{y} = \alpha \mathbf{1}_n + \mathbf{X}_c \boldsymbol{\beta} + \mathbf{e}, \quad (3.1.15)$$

这里  $\mathbf{1}_n$  表示所有分量皆为1的  $n \times 1$  向量.  $\boldsymbol{\beta}' = (\beta_1, \cdots, \beta_{p-1})$  称为回归系数. 这个模型跟前面模型不同之处是:我们把回归常数项  $\alpha$  与回归系数  $\boldsymbol{\beta}$  分离开来了且设计矩阵  $\mathbf{X}_c$  已经中心化,它满足

$$\mathbf{1}'\mathbf{X}_c = \mathbf{0}, \quad (3.1.16)$$

也就是  $\mathbf{X}_c$  的每个列向量的  $n$  个元素之和等于零. 由于这个性质,正则方程(3.1.7)变形为

$$\begin{bmatrix} n & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_c'\mathbf{X}_c \end{bmatrix} \begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \mathbf{1}'\mathbf{y} \\ \mathbf{X}_c'\mathbf{y} \end{pmatrix}, \quad (3.1.17)$$

等价地,

$$\begin{cases} n\alpha = \mathbf{1}'\mathbf{y}, \\ \mathbf{X}_c'\mathbf{X}_c \boldsymbol{\beta} = \mathbf{X}_c'\mathbf{y}. \end{cases} \quad (3.1.18)$$

于是回归参数的最小二乘估计为

$$\begin{cases} \hat{\alpha} = \bar{y}, \\ \hat{\boldsymbol{\beta}} = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}. \end{cases} \quad (3.1.19)$$

这样,对经过中心化程序的线性回归模型(3.1.13),回归常数项的最小二乘估计总是等于因变量的观测平均值,而  $\boldsymbol{\beta}$  的最小二乘估计  $\hat{\boldsymbol{\beta}} = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}$  就相当于从线性回归模型  $\mathbf{y} = \mathbf{X}_c \boldsymbol{\beta} + \mathbf{e}$ ,按原来的最小二乘估计公式(3.1.8)计算得到的.于是我们就把常数项和回归系数的估计也分离开了,因为在实际应用中,我们总是特别关心回归系数,所以中心化是很必要的.

### 例 3.1.2 一元线性回归(续).

将例 3.1.1 所研究的一元线性回归模型进行中心化,得到

$$y_i = \alpha + (x_i - \bar{x})\beta + e_i, \quad i = 1, \cdots, n. \quad (3.1.20)$$

根据(3.1.18),我们可以得到  $\alpha$  和  $\beta$  的最小二乘估计

$$\hat{\alpha} = \bar{y},$$

$$\hat{\beta} = \frac{\sum_i y_i (x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}.$$

除了中心化,对自变量经常做的另一种处理叫做标准化.记

$$s_j^2 = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2, \quad j = 1, \dots, p-1,$$

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}. \quad (3.1.21)$$

我们刚才讨论过,将  $x_{ij}$  减去  $\bar{x}_j$  称为中心化,现在再除以  $s_j$ ,这称为标准化.命  $Z = (z_{ij})$ ,则  $Z$  就是将原来的设计矩阵  $X$  经过中心化和标准化得到的新设计矩阵,这个矩阵具有如下性质:

- (a)  $\mathbf{1}'Z = \mathbf{0}$ ,
- (b)  $\mathbf{R} \stackrel{\text{d}}{=} Z'Z = (r_{ij})$ ,

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{s_i s_j}, \quad i, j = 1, \dots, p-1. \quad (3.1.22)$$

性质(a)是中心化的作用,它使设计阵每列之和都为零.性质(b)是中心化后再施以标准化后的结果.如果把回归自变量都看成随机变量, $X$  的第  $j$  列为第  $j$  个自变量的  $n$  个随机样本.那末  $\mathbf{R} = Z'Z$  的第  $(i, j)$  元正是回归自变量  $X_i$  与  $X_j$  的样本相关系数,因此  $\mathbf{R}$  是回归自变量的相关阵,于是  $r_{ii} = 1$ ,对一切  $i$  成立.这样做的好处有两条.其一是,用  $\mathbf{R}$  可以分析回归自变量之间的相关关系;其次是,在一些问题中,诸回归自变量所用的单位可能不相同,取值范围大小也不同,经过标准化,消去了单位和取值范围的差异,这便于对回归系数估计值的统计分析.

需要注意的是,如果把模型(3.1.2)既经过中心化,又经过标准化,则变形为

$$y_i = \alpha + \left( \frac{x_{i1} - \bar{x}_1}{s_1} \right) \beta_1 + \dots + \left( \frac{x_{i,p-1} - \bar{x}_{p-1}}{s_{p-1}} \right) \beta_{p-1} + e_i. \quad (3.1.23)$$

这里  $\alpha$  和前面的  $\alpha$  当然不同,但我们没有必要写出它和  $\beta_0$  的关系,用矩阵形式,模型(3.1.23)就是

$$\mathbf{y} = \alpha \mathbf{1} + Z\boldsymbol{\beta} + \mathbf{e}.$$

当我们对这个模型求到了参数的最小二乘估计:

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta}_1, \dots, \hat{\beta}_{p-1}$$

之后,所对应的经验回归方程应为

$$\begin{aligned} \hat{Y} &= \hat{\alpha} + \left( \frac{X_1 - \bar{x}_1}{s_1} \right) \hat{\beta}_1 + \cdots + \left( \frac{X_{p-1} - \bar{x}_{p-1}}{s_{p-1}} \right) \hat{\beta}_{p-1} \\ &= \left( \hat{\alpha} - \sum_{i=1}^{p-1} \frac{\bar{x}_i}{s_i} \hat{\beta}_i \right) + \sum_{i=1}^{p-1} \left( \frac{\hat{\beta}_i}{s_i} \right) X_i. \end{aligned} \quad (3.1.24)$$

**例 3.1.3** 一个试验容器靠蒸汽供应热量,使其保持恒温,表 3.1.1 中,自变量  $X$  表示容器周围空气单位时间的平均温度( $^{\circ}\text{C}$ ),  $Y$  表示单位时间内消耗的蒸汽量(L),共观测了 25 个单位时间.图 3.1.1 是这些数据的散点图,对这组数据,应用中心化线性回归模型(3.1.20),我们得到

$$\begin{aligned} \bar{y} &= 9.424, \\ \bar{x} &= 52.60. \end{aligned}$$

常数项  $\alpha$  和回归系数  $\beta$  的最小二乘估计分别为

$$\begin{aligned} \hat{\alpha} &= \bar{y} = 9.424, \\ \hat{\beta} &= -0.0798. \end{aligned}$$

于是,经验回归方程为

$$\hat{Y} = 9.424 - 0.0798(X - 52.60),$$

即

$$\hat{Y} = 13.623 - 0.0798X.$$

从这个经验回归方程,我们看到,周围环境的温度每上升  $1^{\circ}\text{C}$  为保持容器恒温在单位时间内所需要的蒸汽量减少  $0.0798(\text{L})$ .一般说来,一个经验回归方程是不是真正描述了因变量和自变量之间的关系,最重要的是应该根据实践来检验,从问题本身专业知识的角度来分析.当然从数理统计的角度也有一些辅助方法,这将留在下一章讨论.

表 3.1.1 蒸汽数据

序号	Y(L)	X( $^{\circ}\text{C}$ )	序号	Y(L)	X( $^{\circ}\text{C}$ )
1	10.98	35.3	14	9.57	39.1
2	11.13	29.7	15	10.94	46.8
3	12.51	30.8	16	9.58	48.5
4	8.40	58.8	17	10.09	59.3
5	9.27	61.4	18	8.11	70.0
6	8.73	71.3	19	6.83	70.0
7	6.36	74.4	20	8.88	74.5
8	8.50	76.7	21	7.68	72.1
9	7.82	70.7	22	8.47	58.1
10	9.14	57.5	23	8.86	44.6
11	8.24	46.4	24	10.36	33.4
12	12.19	28.9	25	11.08	28.6
13	11.88	28.1			

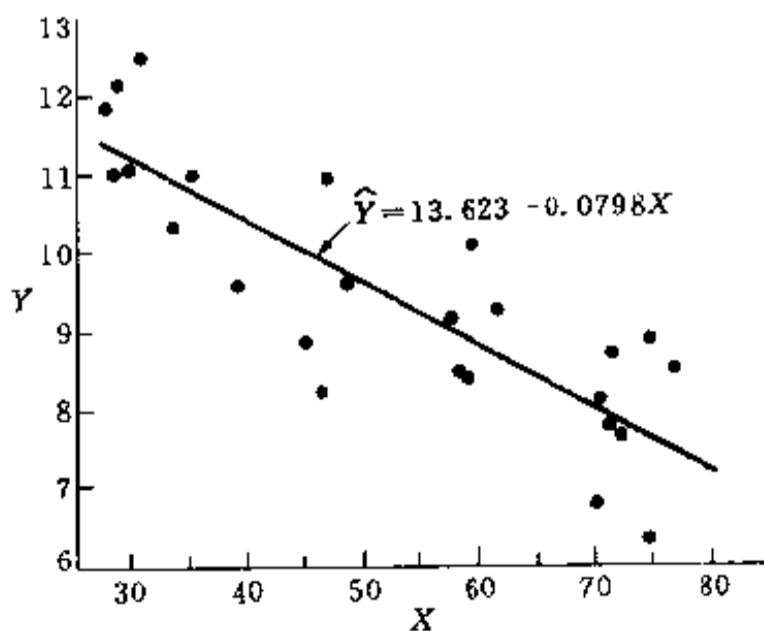


图 3.1.1 蒸汽与温度关系

## § 3.2 最小二乘估计的性质

无论从理论上还是从应用上讲,最小二乘估计都是最重要的估计,其原因是这种估计具有许多优良性质.

**定理 3.2.1** 对于线性回归模型(3.1.6),最小二乘估计  $\hat{\beta} = (X'X)^{-1}X'y$  具有下列性质:

$$(a) E(\hat{\beta}) = \beta;$$

$$(b) \text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

**证明** (a) 因为  $Ey = X\beta$ , 于是

$$E(\hat{\beta}) = (X'X)^{-1}X'E(y) = (X'X)^{-1}X'X\beta = \beta.$$

(b) 因为  $\text{Cov}(y) = \text{Cov}(e) = \sigma^2I$ , 利用定理 2.1.3 有

$$\begin{aligned} \text{Cov}(\hat{\beta}) &= \text{Cov}[(X'X)^{-1}X'y] \\ &= (X'X)^{-1}X'\text{Cov}(y)X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2IX(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}. \end{aligned}$$

定理证毕.

这个定理的第一条结论表明,最小二乘估计  $\hat{\beta}$  是  $\beta$  的无偏估计,这就是说,我们的估计没有系统性偏差.在实际问题中当我们用  $\hat{\beta}$  去估计  $\beta$  时,有时可能会偏高,有时则可能会偏低,但这些正负偏差在概率上平均起来等于零.

设  $c$  为  $p \times 1$  常数向量,对于线性函数  $c'\beta$ ,我们称  $c'\hat{\beta}$  为  $c'\beta$  的最小二乘估计.从定理 3.2.1 我们容易推出

$$\text{推论 3.2.1} \quad (a) E(c'\hat{\beta}) = c'\beta;$$

$$(b) \text{Cov}(c'\hat{\beta}) = \sigma^2c'(X'X)^{-1}c.$$

这就是说,对任意线性函数  $c'\beta$ ,最小二乘估计  $c'\hat{\beta}$  是  $c'\beta$  的无偏估计,且方差为  $\sigma^2c'(X'X)^{-1}c$ . 我们知道  $c'\hat{\beta} = c'(X'X)^{-1}X'y$  是观测值  $y_1, \dots, y_n$  的线性函数,于是  $c'\hat{\beta}$  是  $c'\beta$  的一个具有无偏性的线性估计.当然我们还可以构造出  $c'\beta$  的许多其它的线性无偏估计,构成  $c'\beta$  的线性无偏估计类.我们的问题是,最小二乘估计  $c'\hat{\beta}$  在这个类中有什么特别的优良性呢? 下面的著名定理回答了这个问题.

**定理 3.2.2** (Gauss - Markov) 对于线性回归模型(3.1.6),在  $c'\beta$  的所有

线性无偏估计中,最小二乘估计  $c'\hat{\beta}$  是唯一具有最小方差的估计.

证明 设  $a'y$  为  $c'\beta$  的任意一个线性无偏估计,于是

$$c'\beta = E(a'y) = a'X\beta,$$

此式对一切  $p \times 1$  向量  $\beta$  都成立,因而必然有

$$a'X = c'. \quad (3.2.1)$$

因为  $\text{Var}(a'y) = \sigma^2 a'a = \sigma^2 \|a\|^2$ , 对  $\|a\|^2$  作分解:

$$\begin{aligned} \|a\|^2 &= \|a - X(X'X)^{-1}c + X(X'X)^{-1}c\|^2 \\ &= \|a - X(X'X)^{-1}c\|^2 + \|X(X'X)^{-1}c\|^2 + 2c'(X'X)^{-1}X'(a - X(X'X)^{-1}c). \end{aligned} \quad (3.2.2)$$

记(3.2.2)中第二项为  $\Delta_1$ , 利用推论 3.2.1(b) 得

$$\begin{aligned} \sigma^2 \Delta_1 &= \sigma^2 c'(X'X)^{-1}X'X(X'X)^{-1}c \\ &= \sigma^2 c'(X'X)^{-1}c = \text{Var}(c'\hat{\beta}). \end{aligned}$$

再记(3.2.2)中第三项为  $\Delta_2$ , 利用(3.2.1)有

$$\begin{aligned} \Delta_2 &= 2a'X(X'X)^{-1}X'(a - X(X'X)^{-1}X'a) \\ &= 2a'(X(X'X)^{-1}X'a - X(X'X)^{-1}X'X(X'X)^{-1}X'a) \\ &= 2a'(X(X'X)^{-1}X'a - X(X'X)^{-1}X'a) = 0. \end{aligned}$$

于是从(3.2.2), 我们证明了, 对  $c'\beta$  的任意一个无偏估计  $a'y$ ,

$$\begin{aligned} \text{Var}(a'y) &= \text{Var}(c'\hat{\beta}) + \|a - X(X'X)^{-1}c\|^2 \\ &\geq \text{Var}(c'\hat{\beta}), \end{aligned} \quad (3.2.3)$$

且等号成立当且仅当

$$\|a - X(X'X)^{-1}c\|^2 = 0,$$

等价地

$$a = X(X'X)^{-1}c.$$

于是(3.2.3)等号成立当且仅当  $a'y = c'\hat{\beta}$ , 这就证明了  $c'\hat{\beta}$  是  $c'\beta$  的唯一方差最小的线性无偏估计. 定理证毕.

我们知道, 对于无偏估计, 方差愈小愈好, 因此 Gauss-Markov 定理说明: 最小二乘估计  $c'\hat{\beta}$  在  $c'\beta$  的一切线性无偏估计中是最优的, 所以我们也称  $c'\hat{\beta}$  是  $c'\beta$  的最佳线性无偏估计. 这个事实奠定了最小二乘估计在线性回归估计中的地位.

在线性回归模型(3.1.6)中还有一个重要参数  $\sigma^2$ , 它是模型误差项的方差, 因而有时简称为误差方差.  $\sigma^2$  反映了模型误差以及观察误差的大小, 在回归分析中起着重要作用. 现在我们讨论  $\sigma^2$  的估计问题.

误差向量  $e = y - X\beta$ , 它是一个不可观测的随机向量. 用最小二乘估计  $\hat{\beta}$  代替其中的  $\beta$ , 得到



$$\hat{e} = y - X\hat{\beta}, \quad (3.2.4)$$

称为残差向量. 若用  $x_i'$  表示设计矩阵  $X$  的第  $i$  行, 则上式用分量形式写出来就是

$$\hat{e}_i = y_i - x_i'\hat{\beta}, \quad i = 1, \dots, n \quad (3.2.5)$$

称为第  $i$  次试验或观测的残差.

我们将  $\hat{e}$  看成误差向量  $e$  的一个估计, 很自然我们用

$$RSS = \hat{e}'\hat{e} = \sum_{i=1}^n \hat{e}_i^2 \quad (3.2.6)$$

来衡量  $\sigma^2$  的大小, 这里  $RSS$  是残差平方和 (Residual Sum of Squares, 简记为  $RSS$ ), 它的大小反映了实际数据与理论模型 (3.1.6) 的偏离程度或者说拟合程度.  $RSS$  愈小, 数据与模型拟合得愈好. 下面的定理给出了  $RSS$  的一个有用表达式以及利用  $RSS$  构造的  $\sigma^2$  的无偏估计.

**定理 3.2.3** (a)  $RSS = y'(I - X(X'X)^{-1}X')y$ ;

$$(b) \hat{\sigma}^2 = \frac{RSS}{n-p} \quad (3.2.7)$$

是  $\sigma^2$  的无偏估计.

$$\begin{aligned} \text{证明} \quad (a) \quad RSS &= \hat{e}'\hat{e} = (y - X\hat{\beta})'(y - X\hat{\beta}) \\ &= [(I - X(X'X)^{-1}X')y]'[(I - X(X'X)^{-1}X')y] \\ &= y'(I - X(X'X)^{-1}X')y. \end{aligned}$$

(b) 因为  $E(y) = X\beta$ ,  $\text{Cov}(y) = \sigma^2 I$ , 由定理 2.2.1,

$$\begin{aligned} E(RSS) &= E[y'(I - X(X'X)^{-1}X')y] \\ &= \beta'X'(I - X(X'X)^{-1}X')X\beta \\ &\quad + \sigma^2 \text{tr}(I - X(X'X)^{-1}X') \\ &= \sigma^2 [n - \text{tr}(X(X'X)^{-1}X')]. \end{aligned}$$

利用  $\text{tr}AB = \text{tr}BA$ , 可得

$$\text{tr}X(X'X)^{-1}X' = \text{tr}(X'X)^{-1}X'X = \text{tr}I_p = p.$$

于是

$$E(RSS) = \sigma^2(n-p).$$

定理证毕.

本定理的结论 (a) 给出的  $RSS$  的表达式在理论推导上有其方便之处. 但在实际计算时, 利用下面的表达式:

$$RSS = y'y - \hat{\beta}'X'y, \quad (3.2.8)$$

其中  $y'y$  称为总平方和, 而第二项称为回归平方和. 于是残差平方和等于总平方和减去回归平方和.

前面我们是在 Gauss - Markov 假设下研究了回归系数的最小二乘估计  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  和误差方差  $\sigma^2$  的估计  $\hat{\sigma}^2$  的统计性质. 如果假设误差向量  $\mathbf{e}$  服从正态分布, 即  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ , 那末我们可以得到  $\hat{\boldsymbol{\beta}}$  和  $\hat{\sigma}^2$  更多的重要性质, 它们在后续讨论中具有重要应用.

**定理 3.2.4** 对于线性回归模型 (3.1.6), 若进一步假设误差向量  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ . 则

$$(a) \hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1});$$

$$(b) \frac{RSS}{\sigma^2} \sim \chi_{n-p}^2;$$

(c)  $\hat{\boldsymbol{\beta}}$  与  $RSS$  相互独立.

**证明** (a) 在定理的假设下,  $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ . 注意到  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  是  $\mathbf{y}$  的线性变换, 利用定理 2.3.4, 便证明了 (a).

(b) 根据定义

$$RSS = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \triangleq \mathbf{y}'\mathbf{N}\mathbf{y},$$

这里  $\mathbf{N} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . 注意到  $\mathbf{N}\mathbf{X} = \mathbf{0}$ , 于是

$$\begin{aligned} RSS &= (\mathbf{X}\boldsymbol{\beta} + \mathbf{e})'\mathbf{N}(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) \\ &= \mathbf{e}'\mathbf{N}\mathbf{e}. \end{aligned} \quad (3.2.9)$$

又因为  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ ,  $\mathbf{N}^2 = \mathbf{N}$ , 即  $\mathbf{N}$  是幂等阵, 根据定理 2.4.4, 我们只需证明  $\mathbf{N}$  的秩为  $n - p$ . 因为  $\mathbf{N}$  是幂等阵, 它的秩等于它的迹 (这个事实的证明见附录), 于是

$$\begin{aligned} \text{rank}(\mathbf{N}) &= \text{tr}(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= n - \text{tr}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= n - \text{tr}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = n - p. \end{aligned}$$

这就完成了 (b) 的证明.

(c) 因为  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$ , 而  $RSS = \mathbf{e}'\mathbf{N}\mathbf{e}$ , 注意到  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{N} = \mathbf{0}$ , 由定理 2.4.5 便有  $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}$  与  $RSS$  相互独立. 因而  $\hat{\boldsymbol{\beta}}$  与  $RSS$  相互独立. 定理证毕.

注意到, 在线性回归模型 (3.1.2) 中,  $\boldsymbol{\beta}$  的第一个分量为常数项  $\beta_0$ , 于是若在前面的讨论中, 取  $\mathbf{c} = (0, \dots, 0, 1, 0, \dots, 0)'$ , 这里 1 是  $\mathbf{c}$  的第  $i+1$  个元素. 则  $\mathbf{c}'\boldsymbol{\beta} = \beta_i$ , 即第  $i$  个回归自变量  $X_i$  的回归系数. 记  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{p-1})'$ , 于是  $\mathbf{c}'\hat{\boldsymbol{\beta}} = \hat{\beta}_i$ . 再用  $(A)_{ii}$  表示矩阵  $A$  的第  $(i, i)$  元. 从前面的几个定理和推论, 容易得到如下推论.

**推论 3.2.2** 对于线性回归模型(3.1.6),若  $e \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , 则

$$(a) \hat{\beta}_i \sim N(\beta_i, \sigma^2 [(\mathbf{X}'\mathbf{X})^{-1}]_{i+1, i+1});$$

(b) 在  $\beta_i, i = 0, 1, \dots, p-1$  的一切线性无偏估计中,  $\hat{\beta}_i, i = 0, 1, \dots, p-1$  是唯一方差最小者.

**证明** 结论(a)和(b)可以直接从定理 3.2.4 和定理 3.2.2 得出. 证毕.

下面的推论是将定理 3.2.1 和定理 3.2.4 应用于中心化回归模型(3.1.13)的直接结果.

**推论 3.2.3** 对于中心化的回归模型(3.1.15),

$$(a) E(\hat{\alpha}) = \alpha, \quad E(\hat{\beta}) = \beta.$$

这里  $\hat{\alpha} = \bar{y}, \hat{\beta} = (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{y}$ .

$$(b) \text{Cov} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & (\mathbf{X}'_c \mathbf{X}_c)^{-1} \end{pmatrix}.$$

(c) 若进一步假设  $e \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , 则

$$\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2}{n}\right),$$

$$\hat{\beta} \sim N(\beta, \sigma^2 (\mathbf{X}'_c \mathbf{X}_c)^{-1}),$$

且  $\hat{\alpha}$  与  $\hat{\beta}$  相互独立.

这个推论的证明,留给读者作练习. 从这里我们进一步看到中心化的作用, 它不仅仅是把回归常数项  $\alpha$  和回归系数  $\beta$  的最小二乘估计有分离的简洁表达式(3.1.19), 而且当误差分布为正态时, 两者的估计也是相互独立的.

在结束这一节之前, 我们介绍一个量, 它的定义为

$$R^2 = \frac{SS_{\text{回}}}{SS_{\text{总}}}, \quad (3.2.10)$$

其中  $SS_{\text{回}} = \hat{\beta}' \mathbf{X}'_c \mathbf{y} = \mathbf{y}' \mathbf{X}_c (\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_c \mathbf{y}$  称为回归平方和,

$$SS_{\text{总}} = (\mathbf{y} - \hat{\alpha} \mathbf{1})' (\mathbf{y} - \hat{\alpha} \mathbf{1}) = (\mathbf{y} - \bar{y} \mathbf{1})' (\mathbf{y} - \bar{y} \mathbf{1})$$

称为修正的总平方和.  $R^2$  称为判定系数或测定系数, 它度量了回归自变量  $X_1, \dots, X_{p-1}$  对因变量  $Y$  的拟合程度的好坏. 显然  $0 \leq R^2 \leq 1$ .  $R^2$  是因变量  $Y$  与诸回归自变量的复相关系数之平方. 因此, 它的值愈大, 表明  $Y$  与诸  $X$  有较大的相依关系.

为了进一步理解判定系数  $R^2$  的统计意义, 下面我们来看一看例 3.1.1 和例 3.1.2 所讨论的一元线性回归模型, 这时

$$\begin{aligned}
 SS_{\text{总}} &= \sum_{i=1}^n (y_i - \bar{y})^2, \\
 SS_{\text{回}} &= \hat{\beta} \cdot \sum_{i=1}^n (x_i - \bar{x}) y_i = \frac{\left[ \sum_{i=1}^n y_i (x_i - \bar{x}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
 &= \frac{\left[ \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.
 \end{aligned}$$

于是判定系数

$$R^2 = \frac{\left[ \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \right]^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}.$$

它是回归因变量  $Y$  与自变量  $X$  之间的样本相关系数的平方. 因为相关系数是两个量之间线性相关程度的一种度量, 因此  $R^2$  愈大, 就表示回归方程与数据拟合得愈好.

**例 3.2.1** 根据经验知, 在人的身高相等的条件下, 其血压的收缩压  $y$  与体重  $x_1$ 、年龄  $x_2$  有关. 现收集了 13 名男子的测量数据, 见表(3.2.1). 试建立  $y$  关于  $x_1, x_2$  的线性回归方程.

表 3.2.1 血压数据

序号	$x_{i1}$	$x_{i2}$	$y_i$	序号	$x_{i1}$	$x_{i2}$	$y_i$
1	152	50	120	8	158	50	125
2	183	20	141	9	170	40	132
3	171	20	124	10	153	55	123
4	165	30	126	11	164	40	132
5	158	30	117	12	190	40	155
6	161	50	125	13	185	20	147
7	149	60	123				

利用中心化模型

$$y_i = a + (x_{i1} - \bar{x}_1)\beta_1 + (x_{i2} - \bar{x}_2)\beta_2 + e_i, i = 1, \dots, 13.$$

由表 3.2.1 中数据计算可得

$$\bar{x}_1 = \frac{1}{13} \sum_{i=1}^{13} x_{i1} = 166.8$$

$$\bar{x}_2 = \frac{1}{13} \sum_{i=1}^{13} x_{i2} = 38.85$$

$$y = \frac{1}{13} \sum_{i=1}^{13} y_i = 130$$

中心化设计矩阵

$$\mathbf{X}_c = \begin{pmatrix} -14.08 & 11.15 \\ 16.92 & -18.85 \\ 4.92 & -18.85 \\ 1.08 & -8.85 \\ -8.08 & -8.85 \\ 5.08 & 11.15 \\ -17.08 & 21.15 \\ -8.08 & 11.15 \\ 3.92 & 1.15 \\ 13.08 & 16.15 \\ -2.08 & 1.15 \\ 23.92 & 1.15 \\ 18.92 & -18.85 \end{pmatrix}$$

中心化正则方程组  $\mathbf{X}'_c \mathbf{X}_c \boldsymbol{\beta} = \mathbf{X}'_c \mathbf{y}$  为

$$\begin{cases} 2\,078.92 \beta_1 - 1\,533.85 \beta_2 = 1\,607.00 \\ -1\,533.85 \beta_1 + 2\,307.69 \beta_2 = -715.00 \end{cases}$$

解得

$$\hat{\beta}_1 = 1.068, \quad \hat{\beta}_2 = 0.400.$$

$$\text{又} \quad \hat{\alpha} = \bar{y} = 130,$$

所以得到经验回归方程为

$$\begin{aligned} \hat{Y} &= \hat{\alpha} + (x_1 - \bar{x}_1) \hat{\beta}_1 + (x_2 - \bar{x}_2) \hat{\beta}_2 \\ &= 130 + 1.068(x_1 - 166.8) + 0.400(x_2 - 38.85) \\ &= -62.963 + 1.068x_1 + 0.400x_2. \end{aligned}$$

我们还可以计算出判定系数: 由于  $SS_{\text{回}} = \hat{\boldsymbol{\beta}}' \mathbf{X}'_c \mathbf{y} = 1\,430.276$ ,  $SS_{\text{总}} = \sum_{i=1}^{13} (y_i - \bar{y}_i)^2 = 1\,512$ , 所以  $R^2 = 1\,430.276/1\,512 = 0.945\,9$ . 可见上面得到的二元线性回归方程拟合程度良好.

### § 3.3 约束最小二乘估计

对于线性回归模型(3.1.6), 在对参数向量  $\boldsymbol{\beta}$  没有附加任何约束条件的情况下, 在前面两节我们求出了最小二乘估计并讨论了它的基本性质. 但是, 在-

些检验问题的讨论中或其它一些场合,我们需要求带一定线性约束的最小二乘估计.

假设

$$A\boldsymbol{\beta} = \boldsymbol{b} \quad (3.3.1)$$

是一个相容线性方程组,其中  $A$  为  $k \times p$  的已知矩阵,且秩为  $k$ ,  $\boldsymbol{b}$  为  $k \times 1$  已知向量.我们用 Lagrange 乘子法求模型(3.1.6)满足线性约束(3.3.1)的最小二乘估计.记

$$A = \begin{bmatrix} a_1' \\ \vdots \\ a_k' \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}, \quad (3.3.2)$$

则线性约束(3.3.1)可以改写为

$$a_i' \boldsymbol{\beta} = b_i, \quad i = 1, \dots, k. \quad (3.3.3)$$

我们的问题是在(3.3.3)的  $k$  个条件下求  $\boldsymbol{\beta}$  使  $Q(\boldsymbol{\beta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2$  达到最小值.为了应用 Lagrange 乘子法,构造辅助函数

$$\begin{aligned} F(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + 2 \sum_{i=1}^k \lambda_i (a_i' \boldsymbol{\beta} - b_i) \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + 2\boldsymbol{\lambda}'(\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{b}) \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}'(\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{b}), \end{aligned}$$

其中  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$  为 Lagrange 乘子.对函数  $F(\boldsymbol{\beta}, \boldsymbol{\lambda})$  求对  $\beta_0, \beta_1, \dots, \beta_{p-1}$  的偏导数,整理并令它们等于零,得到

$$-\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{A}'\boldsymbol{\lambda} = \mathbf{0}. \quad (3.3.4)$$

然后解(3.3.4)和线性约束(3.3.1)组成的联立方程组.

为方便计,我们用  $\hat{\boldsymbol{\beta}}_c$  和  $\hat{\boldsymbol{\lambda}}_c$  表示(3.3.4)和(3.3.1)的解.用  $(\boldsymbol{X}'\boldsymbol{X})^{-1}$  左乘(3.3.4),整理后得到

$$\begin{aligned} \hat{\boldsymbol{\beta}}_c &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\hat{\boldsymbol{\lambda}}_c \\ &= \hat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\hat{\boldsymbol{\lambda}}_c. \end{aligned} \quad (3.3.5)$$

代入(3.3.1)得

$$\boldsymbol{b} = \boldsymbol{A}\hat{\boldsymbol{\beta}}_c = \boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\hat{\boldsymbol{\lambda}}_c,$$

等价地

$$\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'\hat{\boldsymbol{\lambda}}_c = (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}). \quad (3.3.6)$$

这是一个关于  $\hat{\boldsymbol{\lambda}}_c$  的线性方程组.因为  $A$  的秩为  $k$ ,于是  $\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}'$  是  $k \times k$  的可逆矩阵,故(3.3.6)有唯一解

$$\hat{\boldsymbol{\lambda}}_c = (\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}')^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{b}),$$

将  $\hat{\lambda}_c$  代入(3.3.5)得到

$$\hat{\beta}_c = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{b}). \quad (3.3.7)$$

现在我们证明  $\hat{\beta}_c$  确实是线性约束  $\mathbf{A}\beta = \mathbf{b}$  下  $\beta$  的最小二乘估计. 为此我们只需证明如下两点:

(a)  $\mathbf{A}\hat{\beta}_c = \mathbf{b}$ ;

(b) 对一切满足  $\mathbf{A}\beta = \mathbf{b}$  的  $\beta$ , 都有

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 \geq \|\mathbf{y} - \mathbf{X}\hat{\beta}_c\|^2.$$

根据(3.3.7)结论(a)是很容易验证的. 为了证明(b), 我们将平方和  $\|\mathbf{y} - \mathbf{X}\beta\|^2$  作分解. 利用(3.1.10), 得

$$\begin{aligned} \|\mathbf{y} - \mathbf{X}\beta\|^2 &= \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + (\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \beta) \\ &= \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + (\hat{\beta} - \hat{\beta}_c + \hat{\beta}_c - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \hat{\beta}_c + \hat{\beta}_c - \beta) \\ &= \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + (\hat{\beta} - \hat{\beta}_c)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \hat{\beta}_c) + (\hat{\beta}_c - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta}_c - \beta) \\ &= \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + \|\mathbf{X}(\hat{\beta} - \hat{\beta}_c)\|^2 + \|\mathbf{X}(\hat{\beta}_c - \beta)\|^2, \end{aligned} \quad (3.3.8)$$

这里我们利用了应用(3.3.5)得到的下述关系:

$(\hat{\beta} - \hat{\beta}_c)' \mathbf{X}'\mathbf{X}(\hat{\beta}_c - \beta) = \hat{\lambda}_c' \mathbf{A}(\hat{\beta}_c - \beta) = \hat{\lambda}_c' (\mathbf{A}\hat{\beta}_c - \mathbf{A}\beta) = \hat{\lambda}_c' (\mathbf{b} - \mathbf{b}) = 0$ , 这个等式对一切满足  $\mathbf{A}\beta = \mathbf{b}$  的  $\beta$  成立.

(3.3.8)式表明, 对一切满足  $\mathbf{A}\beta = \mathbf{b}$  的  $\beta$ , 总有

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 \geq \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + \|\mathbf{X}(\hat{\beta} - \hat{\beta}_c)\|^2, \quad (3.3.9)$$

且等号成立当且仅当(3.3.8)式的第三项等于零, 也就是

$$\mathbf{X}(\hat{\beta}_c - \beta) = \mathbf{0}.$$

因为  $\mathbf{X}$  的秩为  $p$ , 故上式等价于  $\beta = \hat{\beta}_c$ . 于是在(3.3.9)中用  $\hat{\beta}_c$  代替  $\beta$ , 等式成立, 即

$$\|\mathbf{y} - \mathbf{X}\hat{\beta}_c\|^2 = \|\mathbf{y} - \mathbf{X}\hat{\beta}\|^2 + \|\mathbf{X}(\hat{\beta} - \hat{\beta}_c)\|^2. \quad (3.3.10)$$

综合(3.3.9)和(3.3.10), 便证明了结论(b).

我们把估计  $\hat{\beta}_c$  称为  $\beta$  的约束最小二乘估计, 于是我们证明了下面的定理.

**定理 3.3.1** 对于线性回归模型(3.1.6), 满足(3.3.1)的约束最小二乘估计为

$$\hat{\beta}_c = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{b}),$$

其中  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  是无约束条件下的最小二乘估计.

**例 3.3.1** 在天文测量中,对天空中三个星位点构成的三角形  $ABC$  的三个内角  $\theta_1, \theta_2, \theta_3$ , 进行测量,得到的测量值分别为  $y_1, y_2, y_3$ , 由于存在测量误差, 所以需对  $\theta_1, \theta_2, \theta_3$  进行估计, 我们利用线性模型表示有关的量:

$$\begin{cases} y_1 = \theta_1 + e_1 \\ y_2 = \theta_2 + e_2 \\ y_3 = \theta_3 + e_3 \\ \theta_1 + \theta_2 + \theta_3 = \pi \end{cases} \quad (3.3.11)$$

其中  $e_i, i=1, 2, 3$  表示测量误差. (3.3.11) 就是一个带有约束条件的线性模型. 可将它写成矩阵形式

$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \\ \mathbf{A}\boldsymbol{\beta} = \mathbf{b}, \end{cases}$$

其中  $\mathbf{y} = (y_1, y_2, y_3)'$ ,  $\boldsymbol{\beta} = (\theta_1, \theta_2, \theta_3)'$ ,  $\mathbf{X} = \mathbf{I}_3$ ,  $\mathbf{I}_3$  表示 3 阶单位阵,  $\mathbf{A} = (1, 1, 1)$ ,  $\mathbf{b} = \pi$ . 利用定理 3.3.1 可得到  $\boldsymbol{\beta}$  的约束最小二乘估计为

$$\hat{\boldsymbol{\beta}}_c = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{b}),$$

其中  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y}$  是  $\boldsymbol{\beta}$  的无约束最小二乘估计, 经计算可得

$$\hat{\boldsymbol{\beta}}_c = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} - \frac{1}{3} \left( \sum_{i=1}^3 y_i - \pi \right) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

即  $\hat{\theta}_i = y_i - \frac{1}{3}(y_1 + y_2 + y_3 - \pi)$ ,  $i=1, 2, 3$  为  $\theta_i$  的约束最小二乘估计.

## § 3.4 回归诊断

在前面几节, 我们讨论了线性回归模型的最小二乘估计, 并证明了它的一些优良性质. 当进行上述讨论时, 我们对模型做了一些假设, 其中最主要的是 Gauss-Markov 假设, 即假定模型误差  $e_i$  满足下列条件:

- (a)  $E(e_i) = 0$ ;
- (b)  $\text{Var}(e_i) = \sigma^2$  (等方差);
- (c)  $\text{Cov}(e_i, e_j) = 0, i \neq j$  (不相关).

对于涉及到估计量统计分布的一些性质, 我们还假设了  $e_i$  服从正态分布, 即  $e_i \sim N(0, \sigma^2)$ . 这里有一个重要问题, 就是在一个具体场合, 当有了一批实际数据之后, 怎样考察我们的数据基本上满足这些假设, 这就是回归诊断中要研究的第



一个问题.因为这些假设都是关于误差项的,所以很自然我们要从分析它们的“估计量”——残差的角度来解决.正是这个原因,这部分内容在文献中也称为残差分析.

回归诊断所要研究的另一个重要问题,是探查对参数估计或预测有异常大的影响的数据.这样的数据称为强影响数据,因为一组数据对应空间中的一个点,所以也称为强影响点.在回归分析中,因变量  $Y$  的取值  $y_i$  具有随机性,而自变量  $X_1, \dots, X_{p-1}$  的取值  $\mathbf{x}' = (x_{i1}, \dots, x_{i,p-1})$ ,  $i = 1, \dots, n$  也只是许多可能取到的值中的  $n$  组.我们希望每组数据  $(\mathbf{x}'_i, y_i)$  对未知参数的估计有一定的影响,但这种影响不能过大.这样,我们得到的经验回归方程就具有一定的稳定性.不然的话,如果个别一两组数据对估计有异常大的影响,当我们剔除这些数据之后,就能得到与原来差异很大的经验回归方程,这样我们就有理由怀疑所建立的经验回归方程是否真正描述了因变量与诸自变量之间的客观存在的相依关系.正是这个原因,我们在做回归分析时,有必要考察每组数据对参数估计的影响大小.这部分内容在回归诊断中,统称为影响分析.影响分析只是研究探查强影响数据的统计方法,至于对已确认的强影响数据如何处理,这需要具体问题具体分析.往往先要仔细核查数据获得的全过程,如果强影响数据是由于试验条件失控或记录失误或其它一些过失所致,那么这些数据应该剔除.不然的话,应该考虑收集更多的数据(从几何上讲,这些数据应该跟强影响数据比较接近)或采用一些稳健估计方法以缩小强影响数据对估计的影响,从而获得较稳定的经验回归方程.

回归诊断是一个很复杂的问题,实施起来很有点像医生给病人诊病,有时一个症状往往是多种不同疾病的征兆,必须从多方面做检查分析,才能断言毛病出在什么地方.在这方面,理论虽然起一定指导作用,但“临床”经验是十分重要的,这一点从后面的讨论中可以看到.我们先讨论第一个问题——残差分析.

考虑线性回归模型

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}, \quad (3.4.1)$$

如果用  $\mathbf{x}'_1, \dots, \mathbf{x}'_n$  表示  $\mathbf{X}$  的  $n$  个行向量,在 §3.2 我们已经定义

$$\hat{e}_i = y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}, \quad i = 1, \dots, n \quad (3.4.2)$$

为第  $i$  次试验或观测的残差.我们把残差  $\hat{e}_i$  看作误差  $e_i$  的一次观测值,如果模型(3.4.1)正确的话,它应该具有  $e_i$  的一些性状,因此我们应该首先研究残差的性质.

记  $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ , 称  $\hat{\mathbf{y}}$  为拟合值向量,称其第  $i$  个分量  $\hat{y}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$  为第  $i$  个拟合值,则

$$\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \stackrel{d}{=} \mathbf{H}\mathbf{y}. \quad (3.4.3)$$

这里  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ . 从上式看出,  $\mathbf{H}$  作用在观测向量  $\mathbf{y}$  上是给  $\mathbf{y}$  戴上了帽子

“ $\wedge$ ”,故文献中通常称  $H$  为帽子矩阵.容易验证,帽子矩阵  $H$  满足

$$H^1 = H, \quad H^2 = H, \quad (3.4.4)$$

即帽子矩阵是一个对称幂等阵.利用帽子矩阵  $H$ ,残差向量  $\hat{e}$  可表为

$$\hat{e} = y - \hat{y} = (I - H)y = (I - H)e. \quad (3.4.5)$$

从这个表达式我们很容易证明残差向量的下列重要性质.

**定理 3.4.1** (a)  $E(\hat{e}) = \mathbf{0}$ ,  $\text{Cov}(\hat{e}) = \sigma^2(I - H)$ ;

(b) 若进一步假设误差服从正态分布,即  $e \sim N(\mathbf{0}, \sigma^2 I)$ , 则

$$\hat{e} \sim N(\mathbf{0}, \sigma^2(I - H)).$$

**证明** (a) 利用(3.4.5),容易证明  $E(\hat{e}) = \mathbf{0}$ .再利用定理 2.1.3.

$$\begin{aligned} \text{Cov}(\hat{e}) &= (I - H)\text{Cov}(e)(I - H)' \\ &= \sigma^2(I - H). \end{aligned}$$

(b) 因为  $e \sim N(\mathbf{0}, \sigma^2 I)$ ,对(3.4.5)应用定理 2.3.2,便可推出  $\hat{e} \sim N(\mathbf{0}, \sigma^2(I - H))$ .定理证毕.

我们看到  $\text{Var}(\hat{e}_i) = \sigma^2(1 - h_{ii})$ ,这里  $h_{ii}$  为  $H$  的第  $i$  个对角元.可见一般情况下残差  $\hat{e}_i$  的方差不相等,这有碍于  $\hat{e}_i$  的实际应用.将其标准化为  $\hat{e}_i / (\sigma \sqrt{1 - h_{ii}})$ ,再用  $\hat{\sigma}$  代替  $\sigma$ ,得到所谓学生化残差

$$r_i = \frac{\hat{e}_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n, \quad (3.4.6)$$

这里  $\hat{\sigma}^2 = \text{RSS}/(n - p)$ .即使在  $e \sim N(\mathbf{0}, \sigma^2 I)$  的条件下,  $r_i$  的分布仍然比较复杂,但可以近似地认为  $r_i$  相互独立且服从  $N(0, 1)$  (详细讨论见陈希孺等(1987)).于是我们可以断言当  $e \sim N(\mathbf{0}, \sigma^2 I)$  时,学生化残差  $r_1, \dots, r_n$  近似地看作来自总体  $N(0, 1)$  的一组随机样本.根据正态分布的如下性质:若随机变量  $U \sim N(\mu, \sigma^2)$ , 则

$$P(\mu - 2\sigma < U < \mu + 2\sigma) = 95.4\%.$$

对于现在的情形,  $\mu = 0, \sigma = 1$ , 于是,大约应有 95.4% 的  $r_i$  落在区间  $[-2, 2]$  中.另外,我们可以证明,拟合值向量  $\hat{y}$  与残差  $\hat{e}$  相互独立,因而与学生化残差  $r_1, \dots, r_n$  也独立.所以,如果我们以拟合值  $\hat{y}_i$  为横轴,  $r_i$  为纵轴,那么平面上的点  $(\hat{y}_i, r_i), i = 1, \dots, n$  大致应落在宽度为 4 的水平带  $|r_i| \leq 2$  区域内,且不呈现任何趋势,如图 3.4.1(a).这样的以残差为纵轴而以拟合值  $\hat{y}_i$  或其它量为横轴的图称为残差图,这是回归诊断的一个重要工具.如果残差图具有图 3.4.1(a) 的性状,则我们可以认为,现在我们手头的的数据与假设  $e \sim N(\mathbf{0}, \sigma^2 I)$  没有明显不一致的征兆.我们就可以认为,假设  $e \sim N(\mathbf{0}, \sigma^2 I)$  基本上是合理的.而图 3.4.1 (b) — (d) 显示了误差等方差,即  $\text{Var}(e_i) = \sigma^2, i = 1, \dots, n$  不满足.其中图(b)表

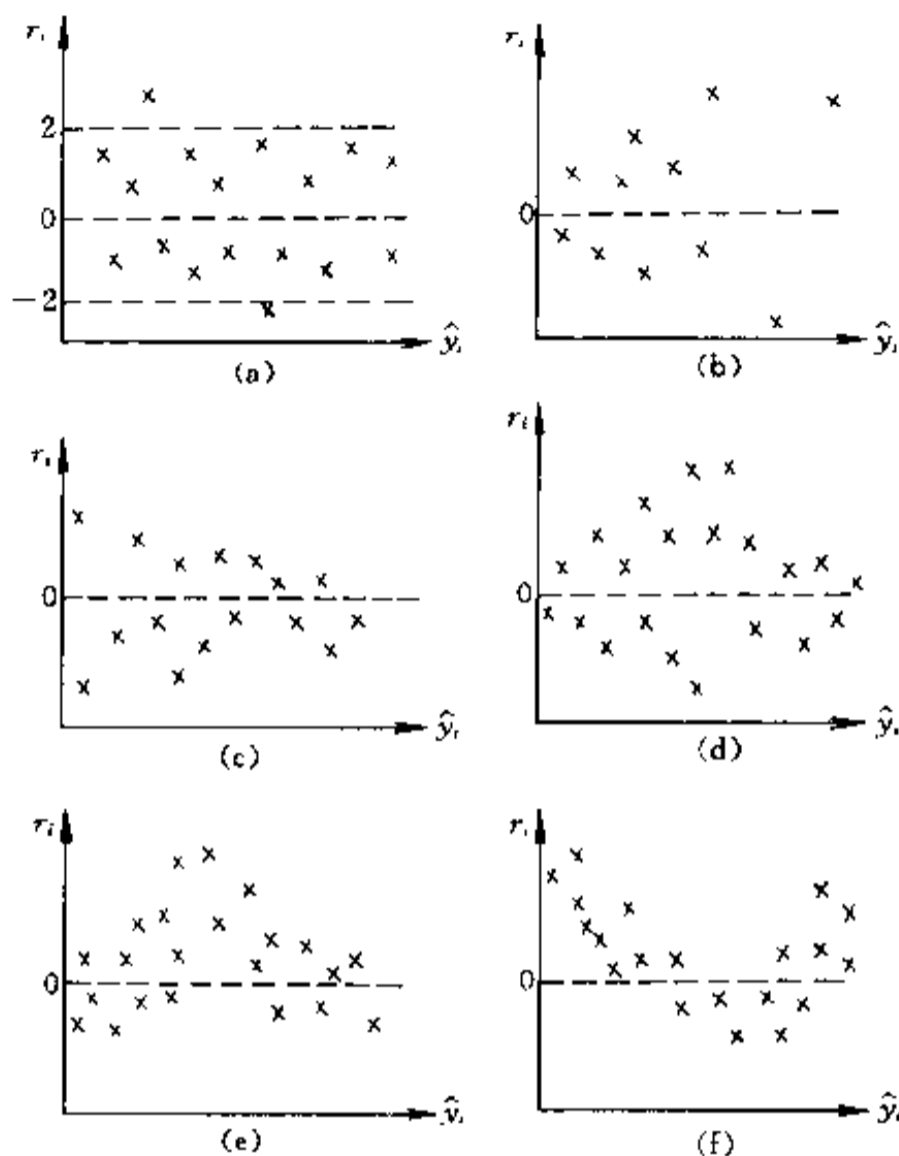


图 3.4.1 学生化残差图

示了误差方差随 $\hat{y}_i$ 的增大而有增加的趋势. 而图(c)所表示的情形正好相反, 即误差方差随 $\hat{y}_i$ 的增大而减少, 但是图(d)表示对较大或较小的 $\hat{y}_i$ , 误差方差偏小, 而对中等大小的 $\hat{y}_i$ , 误差方差偏大. 图(e)和(f)表明回归函数可能是非线性的, 或误差 $e_i$ 之间有一定相关性或漏掉了一个或多个重要的回归自变量. 对于一批实际数据, 这样的残差图究竟反映了哪一种情况, 还需要做进一步的诊断. 这种由一种“症状”可能产生多种“疾病”的情况正是回归诊断的困难所在. 在具体处理时, 和医生治病一样, 临床经验是很重要的.

上面我们讲的是以拟合值 $\hat{y}_i$ 为横坐标的残差图. 为了从不同的角度分析残差, 我们可以做其它一些残差图. 例如, 如果因变量是按时间顺序观测的, 那么 $y_1, \dots, y_n$ 表示了分别在时刻 $t = t_1, \dots, t_n$ 的因变量观测值, 则我们可以取时间 $t$ 或观测序号为 $X$ 轴, 构造 $(t_i, r_i)$ 或 $(i, r_i)$ 的残差图. 又譬如, 我们也可将某个自变量 $X_j$ 取做 $X$ 轴等. 不同的残差图可能从不同角度提供一些有用信息.

从残差图诊断出来可能的“疾病”, 也就是某些假设条件不成立, 我们就需要对问题“对症下药”. 如果有症状使我们怀疑因变量 $Y$ 对自变量的依赖不仅仅是

线性关系,那么我们就可以考虑在回归自变量中增加某些自变量的二次项,如  $X_1^2$ ,或  $X_2^2$  或交叉项  $X_1X_2$  等.至于增添哪些变量的二次项和哪些变量的交叉项,这就要通过对实际问题的分析或实际计算,看其实际效果.若增加二次项  $X_1^2, X_2^2$  和交叉项  $X_1X_2$ ,可以通过引进新变量  $Z_1 = X_1^2, Z_2 = X_2^2, Z_3 = X_1X_2$ ,把问题化成线性回归形式.如果残差图显示了误差方差不相等,我们可以有两种“治疗方案”.其一是对因变量作变换,使变换过的新变量具有近似相等的方差.重要的问题是如何选择所要做的变换.虽然在理论上有一些原则可遵循(参阅陈希孺等(1987)p.122),但在实际应用中还是要靠对具体情况分析,提出一些可选择的变换,然后通过实际计算比较它们的客观效果.另一种方法是应用加权最小二乘估计,这将在 § 3.6 讨论.另外,还有一种因变量的变换,它是从综合角度考虑(即要求对因变量变换过之后,新的因变量关于诸自变量具有线性相依关系,且误差服从正态、等方差、相互独立等)提出的一种“治疗方案”,在实际应用上效果比较好,这就是著名的 Box-Cox 变换,这将在下一节讨论.

为了既简单又能说明问题,下面我们看一个一元线性回归的例子.

**例 3.4.1** 一公司为了研究产品的营销策略,对产品的销售情况进行了调查.设  $Y$  表示某地区该产品的家庭人均购买量(单位:元), $X$  表示家庭人均收入(单位:元).表 3.4.1 记录了 53 个家庭的数据.

表 3.4.1 家庭人均收入数据

$i$	$X(\text{元})$	$Y(\text{元})$	$\hat{y}_i$	$\hat{\varepsilon}_i$	$Z = \sqrt{Y}$	$\hat{\varepsilon}'_i$	$e_i$
1	679	0.790	1.669	-0.879	0.889	1.229	-0.340
2	292	0.440	0.244	0.196	0.663	0.860	-0.197
3	1012	0.560	2.896	-2.336	0.748	1.547	-0.798
4	493	0.790	0.984	-0.194	0.889	1.052	-0.163
5	582	2.700	1.312	1.388	1.643	1.137	0.506
6	1156	3.640	3.426	0.214	1.908	1.684	0.224
7	997	4.730	2.840	1.890	2.175	1.532	0.643
8	2189	9.500	7.230	2.270	3.082	2.668	0.414
9	1097	5.340	3.209	2.131	2.311	1.628	0.683
10	2078	6.850	6.822	0.028	2.617	2.562	0.055
11	1818	5.840	5.864	-0.024	2.417	2.315	0.102
12	1700	5.210	5.430	-0.220	2.283	2.202	0.080
13	747	3.250	1.920	1.330	1.803	1.294	0.509
14	2030	4.430	6.645	-2.215	2.105	2.517	-0.412
15	1643	3.160	5.220	-2.060	1.778	2.148	-0.370
16	414	0.500	0.693	-0.193	0.707	0.977	-0.270
17	354	0.170	0.472	-0.302	0.412	0.920	-0.507
18	1276	1.880	3.868	-1.988	1.371	1.798	-0.427
19	745	0.770	1.912	-1.142	0.877	1.292	-0.415
20	435	1.390	0.771	0.619	1.179	0.997	0.182
21	540	0.560	1.157	-0.597	0.748	1.097	-0.348

续表

$i$	$X(\text{元})$	$Y(\text{元})$	$\hat{y}_i$	$\hat{e}_i$	$Z = \sqrt{Y}$	$\hat{z}_i$	$\bar{e}_i$
22	874	1.560	2.388	-0.828	1.249	1.415	-0.166
23	1543	5.280	4.851	0.429	2.298	2.052	0.245
24	1029	0.640	2.958	-2.318	0.800	1.563	-0.763
25	710	4.000	1.784	2.216	2.000	1.259	0.741
26	1434	0.310	4.450	-4.140	0.557	1.949	-1.392
27	837	4.200	2.251	1.949	2.049	1.380	0.670
28	1748	4.880	5.606	-0.726	2.209	2.248	-0.039
29	1381	3.480	4.255	-0.775	1.865	1.898	-0.033
30	1428	7.580	4.428	3.152	2.753	1.943	0.810
31	1255	2.630	3.791	-1.161	1.622	1.778	-0.156
32	1777	4.990	5.713	-0.723	2.234	2.275	-0.042
33	370	0.590	0.531	0.059	0.768	0.935	-0.167
34	2316	8.190	7.698	0.492	2.862	2.789	0.073
35	1130	4.790	3.330	1.460	2.189	1.659	0.530
36	463	0.510	0.874	-0.364	0.714	1.023	-0.309
37	770	1.740	2.004	-0.264	1.319	1.316	0.003
38	724	4.100	1.835	2.265	2.025	1.272	0.753
39	808	3.940	2.144	1.796	1.985	1.352	0.633
40	790	0.960	2.078	-1.118	0.980	1.335	-0.355
41	783	3.290	2.052	1.238	1.814	1.328	0.486
42	406	0.440	0.664	-0.224	0.663	0.969	-0.306
43	1242	3.240	3.743	-0.503	1.800	1.766	0.034
44	658	2.140	1.592	0.548	1.463	1.209	0.254
45	1746	5.710	5.599	0.111	2.390	2.246	0.144
46	468	0.640	0.892	-0.252	0.800	1.028	-0.228
47	1114	1.900	3.271	-1.371	1.378	1.644	-0.265
48	413	0.510	0.690	-0.180	0.714	0.976	-0.262
49	1787	8.330	5.750	2.580	2.886	2.285	0.601
50	3560	14.940	12.280	2.660	3.865	3.974	-0.109
51	1495	5.110	4.675	0.435	2.261	2.007	0.254
52	2221	3.850	7.348	-3.498	1.962	2.699	-0.736
53	1526	3.930	4.789	-0.859	1.982	2.036	-0.054

应用最小二乘法,求得  $Y$  对  $X$  的一元经验回归方程为

$$\hat{Y} = -0.8313 + 0.003683X.$$

相应的残差  $\hat{e}_i$  和拟合值  $\hat{y}_i$  也列在表 3.4.1 中. 图 3.4.2 是以  $\hat{y}_i$  为横轴, 残差  $\hat{e}_i$  为纵轴的残差图. 直观上容易看出, 残差图从左向右逐渐散开呈漏斗状, 这是误差方差不相等的一个征兆. 考虑对因变量  $Y$  作变换, 先试变换  $Z = Y^{\frac{1}{2}}$ , 得到经验回归方程

$$\hat{Z} = 0.5822 + 0.000953X.$$

计算新的残差  $\tilde{e}_i$ , 残差图画在图 3.4.3, 已无任何明显趋势. 这表明我们所用的变换是合适的. 最后得到的经验回归方程为

$$\begin{aligned}\hat{Y} = \hat{Z}^2 &= (0.5822 + 0.000953X)^2 \\ &= 0.3390 + 0.0011X + 0.00000091X^2.\end{aligned}$$

现在我们讨论回归诊断的第二个问题: 影响分析, 即探查对估计或预测有异常大影响的数据. 为此, 我们先引进一些记号. 用  $\mathbf{y}_{(i)}$ ,  $\mathbf{X}_{(i)}$  和  $\mathbf{e}_{(i)}$  分别表示从  $\mathbf{Y}$ ,  $\mathbf{X}$  和  $\mathbf{e}$  剔除第  $i$  行所得到的向量或矩阵. 从线性回归模型 (3.4.1) 剔除第  $i$  组数据后, 剩余的  $n-1$  组数据的线性回归模型为

$$\mathbf{y}_{(i)} = \mathbf{X}_{(i)}\boldsymbol{\beta} + \mathbf{e}_{(i)}, E(\mathbf{e}_i) = \mathbf{0}, \text{Cov}(\mathbf{e}_{(i)}) = \sigma^2 \mathbf{I}_{n-1}, \quad (3.4.7)$$

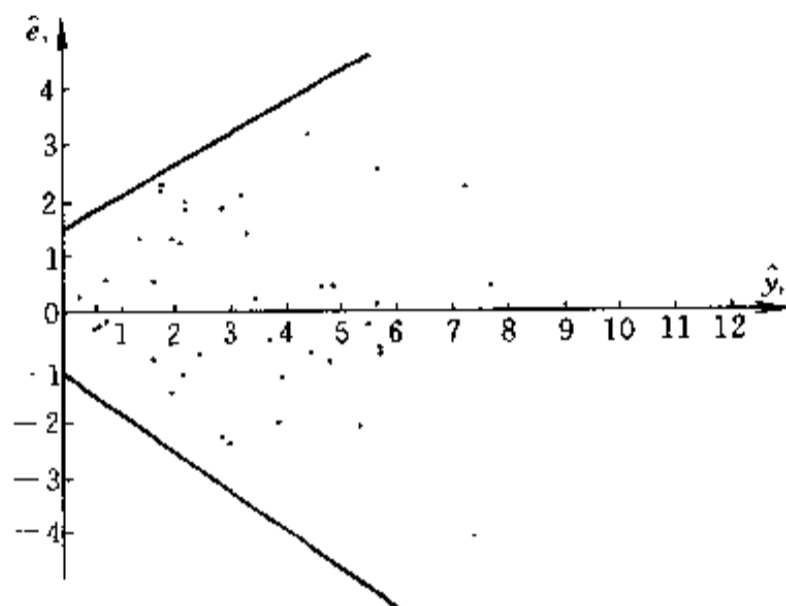


图 3.4.2

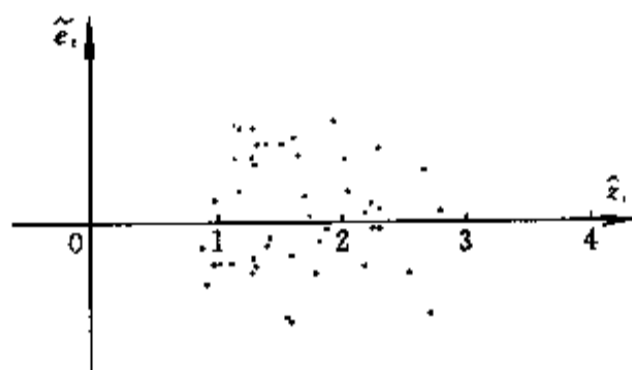


图 3.4.3

将从这个模型求到的  $\boldsymbol{\beta}$  的最小二乘估计记为  $\hat{\boldsymbol{\beta}}_{(i)}$ , 则

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{X}'_{(i)} \mathbf{y}_{(i)}. \quad (3.4.8)$$

很显然, 向量  $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}$  反映了第  $i$  组数据对回归系数估计的影响大小. 但它是一个向量, 应用上不便于使用, 于是需要考虑它的某种数量化函数. Cook 统计量就是其中应用最广泛的一种.

Cook 统计量的定义为

$$D_i = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})' \mathbf{X}' \mathbf{X} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})}{p \hat{\sigma}^2}, \quad i = 1, \dots, n, \quad (3.4.9)$$

这里  $\hat{\sigma}^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2 / (n-p)$ . 于是, 对每一组观测数据, 我们可以有一个数量  $D_i$  来刻画它对回归系数估计影响的大小. 但要从 (3.4.9) 计算  $D_i, i = 1, \dots, n$  很不方便, 它需要计算  $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}_{(1)}, \dots, \hat{\boldsymbol{\beta}}_{(n)}$ , 因而需要计算  $n+1$  个回归, 计算量太大. 下面的定理提供了计算  $D_i$  的简便公式, 它只需要计算完全数据的线性回归

模型(3.4.1).

**定理 3.4.2**

$$D_i = \frac{1}{p} \left( \frac{h_{ii}}{1 - h_{ii}} \right) r_i^2, \quad i = 1, \dots, n \quad (3.4.10)$$

这里  $h_{ii}$  是帽子矩阵  $H = X(X'X)^{-1}X'$  的第  $i$  个对角元,  $r_i$  是学生化残差.

**证明** 设  $A$  为  $n \times n$  可逆阵,  $u$  和  $v$  均为  $n \times 1$  向量. 利用恒等式

$$(A - uv')^{-1} = A^{-1} + \frac{A^{-1}u \cdot v'A^{-1}}{1 - u'A^{-1}v}$$

(这个式子很容易证明, 只需要将右端乘以  $A - uv'$ , 便得到单位阵  $I_n$ ), 有

$$(X'_{(i)}X_{(i)})^{-1} = (X'X - x_i x_i')^{-1} = (X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1 - h_{ii}}, \quad (3.4.11)$$

这里  $x_i'$  为  $X$  的第  $i$  行. 将上式两边右乘  $X'y$ , 并利用

$$X'y = X'_{(i)}y_{(i)} + y_i x_i$$

以及(3.4.8), 我们有

$$\hat{\beta} = \hat{\beta}_{(i)} + y_i (X'_{(i)}X_{(i)})^{-1} x_i - \frac{(X'X)^{-1} x_i (x_i' \hat{\beta})}{1 - h_{ii}} \quad (3.4.12)$$

将(3.4.11)右乘  $x_i$ , 可以得到如下关系式:

$$(X'_{(i)}X_{(i)})^{-1} x_i = \frac{1}{1 - h_{ii}} (X'X)^{-1} x_i.$$

将其代入(3.4.12), 得到

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{\hat{e}_i}{1 - h_{ii}} (X'X)^{-1} x_i. \quad (3.4.13)$$

再代入(3.4.9), 便证明了所要的结论. 定理证毕.

这个定理表明, 在计算 Cook 统计量时, 我们只需要从完全数据的线性回归模型算出学生化残差  $r_i$ , 帽子矩阵的对角元  $h_{ii}$ , 就可以了, 并不必对任何一个不完全数据的线性回归模型(3.4.7)进行计算.

在(3.4.10)中, 若不考虑与  $i$  无关的因子  $\frac{1}{p}$ , 则 Cook 统计量  $D_i$  被分解成两部分, 其中一部分为

$$P_i = \frac{h_{ii}}{1 - h_{ii}},$$

它是  $h_{ii}$  的单调增函数, 因为  $h_{ii}$  度量了第  $i$  组数据  $x_i$  到试验中心  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  的距离(参见陈希孺等(1987)). 因此, 本质上  $P_i$  刻画了第  $i$  组数据距离其它数据

的远近. 而另一部分为  $r_i^2$ . 直观上, 如果一组数据距离试验中心很远, 并且对应的学生化残差又很大, 那么它必定是强影响数据. 但是, 要给 Cook 统计量一个用以判定强影响数据的临界值是很困难的, 在应用上要视具体问题的实际情况而定.

关于这一点, 还有另外一种做法. 由定理 3.2.4 知

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}),$$

利用定理 2.4.1 知

$$\frac{(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{\sigma^2} \sim \chi_p^2.$$

另一方面, 由定理 3.2.4 得

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2,$$

并且这两个  $\chi^2$  变量相互独立. 于是

$$\frac{(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{p\hat{\sigma}^2} \sim F_{p, n-p}. \quad (3.4.14)$$

对给定的  $\alpha$  ( $0 < \alpha < 1$ ), 随机事件

$$\frac{(\hat{\beta} - \beta)' \mathbf{X}'\mathbf{X}(\hat{\beta} - \beta)}{p\hat{\sigma}^2} \leq F_{p, n-p}(\alpha) \quad (3.4.15)$$

发生的概率为  $1 - \alpha$ . 这里  $F_{p, n-p}(\alpha)$  表示自由度为  $p, n-p$  的  $F$  分布的上侧  $\alpha$  分位点. 把  $\beta$  看作变量, (3.4.15) 表示  $p$  维空间中以  $\hat{\beta}$  为中心的椭球, 称为置信椭球. 其置信系数为  $1 - \alpha$ . 在 (3.4.15) 左端用  $\hat{\beta}_{(i)}$  代替  $\hat{\beta}$ , 就得到 Cook 统计量  $D_i$ . 因此, 若  $D_i = F_{p, n-p}(\alpha)$ , 则表明将第  $i$  组数据剔除后,  $\hat{\beta}_{(i)}$  从  $\hat{\beta}$  处移到了  $\beta$  的置信系数为  $1 - \alpha$  的置信椭球上. 这样我们可以借助于置信系数的大小来评价  $D_i$  的大小. 例如, 若  $D_i = F_{p, n-p}(0.80)$ , 则表示第  $i$  组数据剔除后,  $\beta$  的最小二乘估计  $\hat{\beta}_{(i)}$  落在了  $\beta$  的置信系数为  $1 - 0.80 = 20\%$  的置信椭球上.  $D_i$  对应的置信系数愈大, 表明第  $i$  组数据的影响愈大. 但在多数情况下, 对给定的  $D_i$ , 我们难于从  $F$  分布表求到对应的置信系数精确值.

#### 例 3.4.2 智力测试数据.

表 3.4.2 是教育学家测试的 21 个儿童的记录, 其中  $X$  是儿童的年龄 (以月为单位),  $y$  表示某种智力指标, 通过这些数据, 我们要建立智力随年龄变化的关系.

考虑直线回归  $y = \alpha + \beta X + e$ ,  $\alpha$  和  $\beta$  的最小二乘估计分别为  $\hat{\alpha} = 109.87, \hat{\beta}$



$= -1.13$ , 于是经验回归直线为  $\hat{Y} = 109.87 - 1.13X$ . 表 3.4.3 给出了各组数据的残差  $\hat{e}_i$ .

表 3.4.2 智力测试数据

序号	$x$	$y$	序号	$x$	$y$
1	15	95	12	9	96
2	26	71	13	10	83
3	10	83	14	11	84
4	9	91	15	11	102
5	15	102	16	10	100
6	20	87	17	12	105
7	18	93	18	42	57
8	11	100	19	17	121
9	8	104	20	11	86
10	20	94	21	10	100
11	7	113			

表 3.4.3 智力测试数据的诊断统计量

序号	$\hat{e}_i$	$r_i$	$h_i$	$D_i$
1	2.031 0	0.188 8	0.047 9	0.000 9
2	-9.572 1	-0.944 4	0.154 5	0.081 5
3	-15.604 0	-0.821 6	0.062 8	0.071 7
4	-8.730 9	-0.821 6	0.070 5	0.025 6
5	9.031 0	0.839 7	0.047 9	0.017 7
6	-0.334 1	-0.031 5	0.072 6	0.000 0
7	3.412 0	0.318 9	0.058 0	0.003 1
8	2.523 0	0.235 7	0.056 7	0.001 7
9	3.142 1	0.297 2	0.079 9	0.003 8
10	6.665 9	0.628 0	0.072 6	0.015 4
11	11.015 1	1.048 0	0.090 8	0.054 8
12	-3.730 9	-0.351 1	0.070 5	0.004 7
13	-15.604 0	-1.462 3	0.062 8	0.071 7
14	-13.477 0	-1.258 8	0.056 7	0.047 6
15	4.523 0	0.422 5	0.056 7	0.005 4
16	1.396 0	0.130 8	0.062 8	0.000 6
17	8.650 0	0.806 0	0.052 1	0.017 9
18	-5.540 3	-0.851 5	0.651 6	0.678 1
19	30.285 0	2.823 4	0.053 1	0.223 3
20	-11.477 0	-1.072 0	0.056 7	0.034 5
21	1.396 0	0.130 8	0.062 8	0.000 6

从表 3.4.3 看出, 残差  $\hat{e}_{19} = 30.285 0$ ,  $r_{19} = 2.823 4$ , 它们明显地远远大于其它点对应量的绝对值, 这表明第 19 号数据在残差图上会远离其它点, 表现出异常. 然而  $D_{19} = 0.223 3$  却远远小于  $D_{18} = 0.678 1$ , 而其它  $D_i$  值与  $D_{18}$  相比也

十分小,因此,第 18 号数据是一个对回归估计影响很大的数据.对这个数据我们就要格外注意.譬如,检查原始数据的抄录是否有误.如果有误,则需改正后重新计算.不然,需要从原始数据中剔除它.

### § 3.5 Box-Cox 变换

Box-Cox 变换是对回归因变量  $Y$  的如下变换:

$$Y^{(\lambda)} = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln Y, & \lambda = 0, \end{cases} \quad (3.5.1)$$

这里  $\lambda$  是一个待定变换参数.对不同的  $\lambda$ ,所做的变换自然就不同,所以这是一个变换族.它包括了对数变换( $\lambda = 0$ ),平方根变换( $\lambda = 1/2$ )和倒数变换( $\lambda = -1$ )等常用变换.对因变量的  $n$  个观测值  $y_1, \dots, y_n$ ,应用上述变换,我们得到变换后的向量

$$\mathbf{y}^{(\lambda)} = (y_1^{(\lambda)}, \dots, y_n^{(\lambda)})'$$

我们要确定变换参数  $\lambda$ ,使得  $\mathbf{y}^{(\lambda)}$  满足

$$\mathbf{y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}). \quad (3.5.2)$$

这就是说,我们要求通过因变量的变换,使得变换过的向量  $\mathbf{y}^{(\lambda)}$  与回归自变量具有线性相依关系,误差也服从正态分布,误差各分量是等方差且相互独立.因此 Box-Cox 变换是通过参数  $\lambda$  的适当选择,达到对原来数据的“综合治理”,使其满足一个正态线性回归模型的所有假设条件.

我们用极大似然方法来确定  $\lambda$ .因为  $\mathbf{y}^{(\lambda)} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ ,所以对固定的  $\lambda, \boldsymbol{\beta}$  和  $\sigma^2$  的似然函数为

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y}^{(\lambda)} - \mathbf{X}\boldsymbol{\beta}) \right\} J, \quad (3.5.3)$$

这里  $J$  为变换的 Jacobi 行列式

$$J = \prod_{i=1}^n \left| \frac{dy_i^{(\lambda)}}{dy_i} \right| = \prod_{i=1}^n y_i^{\lambda-1}.$$

因此,当  $\lambda$  固定时,  $J$  是不依赖于参数  $\boldsymbol{\beta}$  和  $\sigma^2$  的常数因子.  $L(\boldsymbol{\beta}, \sigma^2)$  的其余部分关于  $\boldsymbol{\beta}$  和  $\sigma^2$  求导数,令其等于零,可以求得  $\boldsymbol{\beta}$  和  $\sigma^2$  的极大似然估计

$$\hat{\boldsymbol{\beta}}(\lambda) = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}^{(\lambda)}, \quad (3.5.4)$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n} \mathbf{y}^{(\lambda)' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y}^{(\lambda)} = \frac{1}{n} \text{RSS}(\lambda, \mathbf{y}^{(\lambda)}),$$

这里残差平方和

$$\text{RSS}(\lambda, \mathbf{y}^{(\lambda)}) = \mathbf{y}^{(\lambda)' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y}^{(\lambda)}.$$

对应的似然函数最大值为

$$\begin{aligned} L_{\max}(\lambda) &= L(\hat{\beta}(\lambda), \hat{\sigma}^2(\lambda)) \\ &= (2\pi e)^{-\frac{n}{2}} \cdot J \cdot \left( \frac{RSS(\lambda, \mathbf{y}^{(\lambda)})}{n} \right)^{\frac{n}{2}}, \end{aligned} \quad (3.5.5)$$

这是  $\lambda$  的一元函数, 通过求它的最大值来确定  $\lambda$ . 因  $\ln x$  是  $x$  的单调函数, 我们的问题可以化为求  $\ln L_{\max}(\lambda)$  的最大值. 对 (3.5.5) 求对数, 略去与  $\lambda$  无关的常数项, 得

$$\begin{aligned} \ln L_{\max}(\lambda) &= -\frac{n}{2} \ln RSS(\lambda, \mathbf{y}^{(\lambda)}) + \ln J \\ &= -\frac{n}{2} \ln \left[ \frac{\mathbf{y}^{(\lambda)'}}{J^{1/n}} (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \frac{\mathbf{y}^{(\lambda)}}{J^{1/n}} \right] \\ &= -\frac{n}{2} \ln RSS(\lambda, \mathbf{z}^{(\lambda)}), \end{aligned} \quad (3.5.6)$$

其中

$$RSS(\lambda, \mathbf{z}^{(\lambda)}) = \mathbf{z}^{(\lambda)' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \mathbf{z}^{(\lambda)}, \quad (3.5.7)$$

$$\mathbf{z}^{(\lambda)} = (z_1^{(\lambda)}, \dots, z_n^{(\lambda)})' = \frac{\mathbf{y}^{(\lambda)}}{J^{1/n}},$$

$$z_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda}{\left(\prod_{i=1}^n y_i\right)^{\frac{\lambda-1}{n}}}, & \lambda \neq 0, \end{cases} \quad (3.5.8)$$

$$\left( \ln y_i \right) \left( \prod_{i=1}^n y_i \right)^{\frac{1}{n}}, \quad \lambda = 0. \quad (3.5.9)$$

(3.5.6) 式对 Box-Cox 变换在计算机上实现带来很大方便, 这是因为为了求  $\ln L_{\max}(\lambda)$  的最大值, 我们只要求残差平方和  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  的最小值. 虽然我们很难找出使  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  达到最小值的  $\lambda$  的解析表达式, 但对一系列的  $\lambda$  给定值, 通过最普通的求最小二乘估计的回归程序, 我们很容易计算出对应的  $RSS(\lambda, \mathbf{z}^{(\lambda)})$ . 画出  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  关于  $\lambda$  的曲线, 从图上可以近似地找出使  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  达到最小值的  $\hat{\lambda}$ .

现在我们把 Box-Cox 变换的具体步骤归纳如下:

1. 对给定的  $\lambda$  值, 计算  $z_i^{(\lambda)}$ . 若  $\lambda \neq 0$ , 用 (3.5.8) 式, 否则用 (3.5.9) 式.
2. 利用 (3.5.7) 式计算残差平方和  $RSS(\lambda, \mathbf{z}^{(\lambda)})$ .
3. 对一系列  $\lambda$  的值, 重复上述步骤, 得到相应的残差平方和  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  的一串值, 以  $\lambda$  为横轴,  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  为纵轴. 作出相应的曲线. 用直观方法, 找出使  $RSS(\lambda, \mathbf{z}^{(\lambda)})$  达到最小值的点  $\hat{\lambda}$ .

4. 利用(3.5.4), 求出  $\hat{\beta}(\hat{\lambda})$ .

**例 3.5.1** 在例3.4.1中, 我们对因变量  $Y$  作了平方根变换, 这相当于选用变换参数  $\lambda = 0.5$ . 应用本节的方法, 我们可以证实作这样的变换是合适的. 表 3.5.1 给出了 12 个不同  $\lambda$  值对应的残差平方和  $RSS(\lambda, z^{(\lambda)})$ , 简单比较可以看出当  $\lambda = 0.5$  时, 残差平方和  $RSS(\lambda, z^{(\lambda)})$  达到最小, 因此我们可以近似地认为 0.5 就是变换参数  $\lambda$  的最优选择.

表 3.5.1

$\lambda$	-2	-1	-0.5	0	0.125	0.25
RSS	34 101.04	986.04	291.59	134.10	119.20	107.21
$\lambda$	0.375	0.5	0.625	0.75	1	2
RSS	100.26	96.95	97.29	101.69	127.87	1 275.56

## § 3.6 广义最小二乘估计

在前面的讨论中, 我们总是假设线性回归模型的误差是等方差且不相关的, 即  $\text{Cov}(e) = \sigma^2 I$ . 虽然在许多情况下, 这个假设总是可以认为近似地成立. 但是, 像在 § 3.4 所指出的, 仍然有许多实际问题, 经过残差分析后, 我们不能认为这些假设是合适的. 它们的误差方差可能不相等, 也可能彼此相关的. 这时误差向量的协方差阵  $\text{Cov}(e) = \sigma^2 \Sigma$ , 这里  $\Sigma$  是一个正定阵, 当然  $\Sigma$  往往包含有未知参数. 为简单计, 在这一节的讨论中, 我们假定  $\Sigma$  是完全已知的. 我们要讨论的线性回归模型具有如下形式:

$$y = X\beta + e, E(e) = 0, \text{Cov}(e) = \sigma^2 \Sigma. \quad (3.6.1)$$

为了求参数的估计, 我们经过适当变换, 把它化成前面讨论过的情形. 既然  $\Sigma$  是正定阵, 于是存在  $n \times n$  正交阵  $P$  使其对角化:

$$\Sigma = P' \Lambda P,$$

这里  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $\lambda_i > 0, i = 1, \dots, n$  是  $\Sigma$  的特征值. 记

$$\Sigma^{-\frac{1}{2}} = P' \text{diag}(\lambda_1^{-\frac{1}{2}}, \dots, \lambda_n^{-\frac{1}{2}}) P.$$

则  $(\Sigma^{-\frac{1}{2}})^2 = \Sigma^{-1}$ , 称  $\Sigma^{-\frac{1}{2}}$  是  $\Sigma^{-1}$  的平方根阵. 用  $\Sigma^{-\frac{1}{2}}$  左乘 (3.6.1), 记  $z = \Sigma^{-\frac{1}{2}} y, U = \Sigma^{-\frac{1}{2}} X, \varepsilon = \Sigma^{-\frac{1}{2}} e$ , 因为  $\text{Cov}(\varepsilon) = \Sigma^{-\frac{1}{2}} \sigma^2 \Sigma \Sigma^{-\frac{1}{2}} = \sigma^2 I$ , 于是我们得到如下线性回归模型:

$$z = U\beta + \varepsilon, E(\varepsilon) = 0, \text{Cov}(\varepsilon) = \sigma^2 I, \quad (3.6.2)$$

这就是我们已经讨论过的情况. 在这个新模型中,  $\beta$  的最小二乘估计为

$$\beta^* = (U'U)^{-1} U'z = (X'\Sigma^{-1}X)^{-1} X'\Sigma^{-1}y. \quad (3.6.3)$$

一般,我们称  $\beta^*$  为  $\beta$  的广义最小二乘估计,也有一些作者称  $\beta^*$  为  $\beta$  的 Gauss-Markov 估计. 下面的定理概括了这个估计的统计性质.

**定理 3.6.1** 对于线性回归模型(3.6.1),

$$(a) E(\beta^*) = \beta;$$

$$(b) \text{Cov}(\beta^*) = \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1};$$

(c) 对任意  $p \times 1$  已知向量  $c$ ,  $c'\beta^*$  为  $c'\beta$  的唯一最小方差无偏估计.

$$\begin{aligned} \text{证明 (a)} \quad E(\beta^*) &= E[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}] \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{X}\beta = \beta. \end{aligned}$$

(b) 利用定理 2.1.3 有

$$\begin{aligned} \text{Cov}\beta^* &= \text{Cov}[(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}] \\ &= \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\Sigma\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}. \end{aligned}$$

(c) 设  $b'y$  是  $c'\beta$  的任一线性无偏估计, 对于模型(3.6.2), 我们有

$$\begin{aligned} c'\beta^* &= c'(U'U)^{-1}U'z, \\ b'y &= b'\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}y = b'\Sigma^{\frac{1}{2}}z, \end{aligned}$$

这就是说, 对变换后的模型而言,  $c'\beta^*$  是  $c'\beta$  的最小二乘估计, 而  $b'y = b'\Sigma^{\frac{1}{2}}z$  是  $c'\beta$  的一个无偏估计, 由定理 3.2.2 知

$$\text{Var}(c'\beta^*) \leq \text{Var}(b'\Sigma^{\frac{1}{2}}z) = \text{Var}(b'y),$$

并且等号成立当且仅当  $c'\beta^* = b'y$ . 定理证毕.

定理 3.6.1(c) 就是一般情况下的 Gauss-Markov 定理. 它表明, 在一般线性回归模型(3.6.1)中, 广义最小二乘估计  $\beta^*$  是最优的. 但是, 如果我们把  $\beta^*$  表达式中的  $\Sigma$  换成单位阵  $I$ , 则得到  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , 称为简单最小二乘估计, 常常简称为最小二乘估计. 容易证明, 对于模型(3.6.1),  $E\hat{\beta} = \beta$ , 即  $\hat{\beta}$  仍然是  $\beta$  的无偏估计. 但这时对任意线性函数  $c'\beta$ ,  $c'\hat{\beta}$  只是  $c'\beta$  的一个无偏估计, 它未必是最优的. 我们称  $c'\beta^*$  和  $c'\hat{\beta}$  分别为  $c'\beta$  的广义最小二乘估计和(简单)最小二乘估计. 根据定理 3.6.1(c), 对一切  $p \times 1$  向量  $c$  有

$$\text{Var}(c'\beta^*) \leq \text{Var}(c'\hat{\beta}).$$

这就是说, 对于一般线性回归模型(3.6.1), 广义最小二乘估计总是优于最小二乘估计.

模型(3.6.1)的最简单例子是因变量的不同观测具有不等方差的情况, 这时

$$\text{Cov}(e) = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_n^2 \end{bmatrix},$$

这里可以有一些  $\sigma_i^2$  彼此相等. 记  $x_1', \dots, x_n'$  分别为设计矩阵  $X$  的  $n$  个行向量, 则容易推出, 这时  $\beta$  的广义最小二乘估计具有如下形式

$$\beta^* = \left[ \sum_{i=1}^n \frac{x_i x_i'}{\sigma_i^2} \right]^{-1} \left[ \sum_{i=1}^n \frac{y_i}{\sigma_i^2} x_i' \right]. \quad (3.6.4)$$

从这个表达式我们可以看出, 两个和式分别是  $x_i x_i'$  和  $y_i x_i'$  的“加权和”, 而所用的“权”都是  $\frac{1}{\sigma_i^2}$ . 因此文献中常常把(3.6.4)定义的  $\beta^*$  称为加权最小二乘估计.

这里讲的是  $\sigma_i^2$  皆已知的情况. 在实际应用中,  $\sigma_i^2$  往往是未知的, 这时我们可以设法求得它们的估计  $\hat{\sigma}_i^2$ , 然后在(3.6.4)中用  $\hat{\sigma}_i^2$  代替  $\sigma_i^2$ . 对于一般线性回归模型, 在实际应用中碰到的问题是确定协方差阵  $\Sigma$  的形式, 但这往往是十分困难的. 一般我们总是从假设  $\Sigma = I$  入手, 求出简单最小二乘估计, 然后通过残差分析, 对误差方差提供一些信息. 另外一种做法是, 从问题本身的专业角度或其它方面, 对误差向量提出一些特殊结构, 这时误差协方差阵就具有特定形式.

**例 3.6.1** 在线性回归模型(3.6.1)中, 若观测向量  $y$  是  $n$  件样品的某项指标, 它们是在  $k$  台仪器上测试得到的. 不妨设前  $n_1$  个是在第一台仪器上测得的, 接下来的  $n_2$  个是从第二台仪器上测得的, 依次类推. 那么我们的模型可以改写为

$$y_{ij} = x_{ij}'\beta + e_{ij}, \quad i = 1, \dots, k; j = 1, \dots, n_i, \quad (3.6.5)$$

其中  $y_{ij}$  表示在第  $i$  台仪器上测量的第  $j$  个样品的指标,  $x_{ij}$  表示对应的样品上自变量的取值,  $e_{ij}$  是对应的试验和测量误差. 假设所有这些测量都是相互独立的, 但不同仪器的测量精度不同. 对这种情况, 下面的假设是合理的.

$$\text{Var}(e_{ij}) = \sigma_i^2, \quad i = 1, \dots, k; j = 1, \dots, n_i.$$

$$\text{Cov}(e_{ij}, e_{i'j'}) = 0, \quad i \neq i', j \neq j'.$$

若记  $e = (e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{k1}, \dots, e_{kn_k})'$ , 则  $e$  的协方差阵具有形式

$$\text{Cov}(e) = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{n_1} & & & 0 \\ & \sigma_2^2 \mathbf{I}_{n_2} & & \\ & & \ddots & \\ 0 & & & \sigma_k^2 \mathbf{I}_{n_k} \end{pmatrix}. \quad (3.6.6)$$

若记

$$\begin{aligned} y_i &= (y_{i1}, \dots, y_{i, n_i})', \\ X_i &= (x_{i1}, \dots, x_{i, n_i})', \\ e_i &= (e_{i1}, \dots, e_{i, n_i})'. \end{aligned} \quad (3.6.7)$$

这时我们的模型具有如下形式:

$$y_i = \mathbf{X}_i\beta + e_i, E(e_i) = \mathbf{0}, \text{Cov}(e_i) = \sigma_i^2 \mathbf{I}_{n_i}, \\ i = 1, \dots, k. \quad (3.6.8)$$

从(3.6.8), 我们得到  $\beta$  的广义最小二乘估计的表达式

$$\beta^* = \left( \sum_{i=1}^k \frac{1}{\sigma_i^2} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^k \frac{1}{\sigma_i^2} \mathbf{X}_i' y_i \right), \quad (3.6.9)$$

这是假定  $\sigma_i^2$  都已已知的情形. 和前面讨论的情况一样, 在实际应用中  $\sigma_i^2$  往往是未知的. 但是对现在的情形, 对固定的  $i$ , 从(3.6.8)我们可以得到  $\sigma_i^2$  的无偏估计

$$\hat{\sigma}_i^2 = \frac{\|y_i - \mathbf{X}_i \hat{\beta}_i\|^2}{n_i - p}, \quad (3.6.10)$$

这里  $\hat{\beta}_i = (\mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' y_i$ , 它是从(3.6.8)的第  $i$  个模型得到的  $\beta$  的最小二乘估计. 将这些估计代入(3.6.9), 便得到  $\beta$  的一个估计

$$\tilde{\beta} = \left( \sum_{i=1}^k \frac{1}{\hat{\sigma}_i^2} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^k \frac{1}{\hat{\sigma}_i^2} \mathbf{X}_i' y_i \right), \quad (3.6.11)$$

称为  $\beta$  的两步估计或可行估计.

我们也可以用迭代的方法求  $\beta$  的其它可行估计. 这就是将(3.6.11)估计  $\tilde{\beta}$  代替(3.6.10)中的  $\hat{\beta}_i$ , 得到  $\sigma_i^2$  的新估计, 然后将它们代入(3.6.11)获得  $\beta$  的新估计. 重复这个过程. 直到相邻两次迭代求到的  $\beta$  的估计相差不多为止. 在很一般情况下, 这个迭代过程总是收敛的.

### §3.7 复共线性

前面我们详细讨论了回归系数的最小二乘估计, 并且证明了它的许多优良性质. 其中最重要的是 Gauss-Markov 定理, 它表明在线性无偏估计类中, 最小二乘估计是唯一的具有最小方差的估计. 正是由于这一点, 最小二乘估计在线性统计模型的估计理论与实际应用中占有绝对重要的地位. 但是随着电子计算机技术的飞速发展, 人们愈来愈多地有能力去处理含较多回归自变量的大型回归问题. 许多应用实践表明, 在这些大型线性回归问题中, 最小二乘估计不总是令人满意. 例如, 有时某些回归系数的估计值的绝对值异常大, 有时回归系数的估计值的符号与问题的实际意义相违背等. 研究结果表明, 产生这些问题的原因之一是回归自变量之间存在着近似线性关系, 称为复共线性(Multicollinearity). 本节我们将讨论复共线性对最小二乘估计的影响以及它的严重程度的度量问题.

我们先引进评价一个估计优劣的标准——均方误差(Mean Squared Errors, 以下简记为 MSE).

设  $\boldsymbol{\theta}$  为  $p \times 1$  未知参数向量,  $\tilde{\boldsymbol{\theta}}$  为  $\boldsymbol{\theta}$  的一个估计. 定义  $\tilde{\boldsymbol{\theta}}$  的均方误差为

$$\begin{aligned} \text{MSE}(\tilde{\boldsymbol{\theta}}) &= E \|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \\ &= E(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}). \end{aligned}$$

它度量了估计  $\tilde{\boldsymbol{\theta}}$  跟未知参数向量  $\boldsymbol{\theta}$  平均偏离的大小. 一个好的估计应该有较小的均方误差.

$$\text{定理 3.7.1} \quad \text{MSE}(\tilde{\boldsymbol{\theta}}) = \text{trCov}(\tilde{\boldsymbol{\theta}}) + \|E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2, \quad (3.7.1)$$

这里  $\text{tr}A$  表示方阵  $A$  的迹, 即  $A$  的对角元之和.

$$\begin{aligned} \text{证明} \quad \text{MSE}(\tilde{\boldsymbol{\theta}}) &= E(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= E[(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}}) + (E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})]'[(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}}) + (E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})] \\ &= E(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})'(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}}) + (E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'(E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\stackrel{d}{=} \Delta_1 + \Delta_2. \end{aligned}$$

因为对任意两个矩阵  $A_{m \times n}, B_{n \times m}$ , 有  $\text{tr}(AB) = \text{tr}(BA)$ , 于是上式第一项

$$\begin{aligned} \Delta_1 &= E\text{tr}(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})'(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}}) \\ &= E\text{tr}(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})' \\ &= \text{tr}E(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - E\tilde{\boldsymbol{\theta}})' \\ &= \text{trCov}(\tilde{\boldsymbol{\theta}}). \end{aligned}$$

而第二项  $\Delta_2 = (E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'(E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \|E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$ . 定理证毕.

若记  $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_p)'$ , 则

$$\Delta_1 = \text{trCov}(\tilde{\boldsymbol{\theta}}) = \sum_{i=1}^p \text{Var}(\tilde{\theta}_i).$$

它是估计  $\tilde{\boldsymbol{\theta}}$  的各分量方差之和. 而第二项

$$\Delta_2 = \|E\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^p (E\tilde{\theta}_i - \theta_i)^2$$

它是估计  $\tilde{\boldsymbol{\theta}}$  的各分量的偏差  $E\tilde{\theta}_i - \theta_i$  的平方和. 这样, 一个估计的均方误差就是由各分量的方差和偏差所决定. 一个好的估计应该有较小的方差和偏差.

现在我们用均方误差这个标准来评价最小二乘估计. 考虑线性回归模型

$$\boldsymbol{y} = \alpha \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{e}, E(\boldsymbol{e}) = \mathbf{0}, \text{Cov}(\boldsymbol{e}) = \sigma^2 \mathbf{I}. \quad (3.7.2)$$

这里假定  $n \times p$  的设计矩阵  $\mathbf{X}$  已经中心化和标准化, 并且秩为  $p$ . 由于设计阵是中心化的, 于是常数项  $\alpha$  和回归系数  $\boldsymbol{\beta}$  的最小二乘估计能够分离开来, 它们分别



具有表达式

$$\hat{\alpha} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

把  $\alpha$  与  $\beta$  的最小二乘估计这样分离开来,对研究回归系数的最小二乘估计的改进带来很大方便.

因为最小二乘估计  $\hat{\beta}$  是无偏估计,因此在  $MSE(\hat{\beta})$  表达式中,  $\Delta_2 = 0$ . 又因  $\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$  于是

$$MSE(\hat{\beta}) = \Delta_1 = \sigma^2 \text{tr}(\mathbf{X}'\mathbf{X})^{-1}. \quad (3.7.3)$$

因为  $\mathbf{X}'\mathbf{X}$  是对称正定阵,于是存在  $p \times p$  正交阵  $\Phi$  将  $\mathbf{X}'\mathbf{X}$  对角化,即

$$\mathbf{X}'\mathbf{X} = \Phi \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{pmatrix} \Phi', \quad (3.7.4)$$

这里  $\lambda_1 \geq \dots \geq \lambda_p$  为  $\mathbf{X}'\mathbf{X}$  的特征值. 记  $\Phi = (\varphi_1, \dots, \varphi_p)$ , 则  $\varphi_1, \dots, \varphi_p$  分别为对应于  $\lambda_1, \dots, \lambda_p$  的标准正交化特征向量. 从(3.7.4)容易验证

$$(\mathbf{X}'\mathbf{X})^{-1} = \Phi \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_p} \end{pmatrix} \Phi', \quad (3.7.5)$$

从这里我们看出  $\lambda_1^{-1}, \dots, \lambda_p^{-1}$  为  $(\mathbf{X}'\mathbf{X})^{-1}$  的特征值. 再利用  $\text{tr} \mathbf{AB} = \text{tr} \mathbf{BA}$  以及  $\Phi'\Phi = \mathbf{I}$ , 得到

$$\text{tr}(\mathbf{X}'\mathbf{X}) = \sum_{i=1}^p \frac{1}{\lambda_i}.$$

代入(3.7.3), 我们证明了

$$MSE(\hat{\beta}) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}. \quad (3.7.6)$$

从这个表达式我们可以看出,如果  $\mathbf{X}'\mathbf{X}$  至少有一个特征根非常小,即非常接近

于零,那么  $MSE(\hat{\boldsymbol{\beta}})$  就会很大.从均方误差的标准来看,这时的最小二乘估计  $\hat{\boldsymbol{\beta}}$  就不是一个好的估计.这一点和 Gauss - Markov 定理并无抵触,因为我们知道, Gauss - Markov 定理仅仅保证了最小二乘估计在线性无偏估计类中的方差最小性.但在  $\mathbf{X}'\mathbf{X}$  至少有一个特征值很小时,这个最小的方差值本身却很大,因而导致了很大的均方误差.

另一方面

$$\begin{aligned} MSE(\hat{\boldsymbol{\beta}}) &= E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= E(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} - 2\boldsymbol{\beta}'\hat{\boldsymbol{\beta}} + \boldsymbol{\beta}'\boldsymbol{\beta}) \\ &= E\|\hat{\boldsymbol{\beta}}\|^2 - \boldsymbol{\beta}'\boldsymbol{\beta}, \end{aligned}$$

于是

$$\begin{aligned} E\|\hat{\boldsymbol{\beta}}\|^2 &= \|\boldsymbol{\beta}\|^2 + MSE(\hat{\boldsymbol{\beta}}) \\ &= \|\boldsymbol{\beta}\|^2 + \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}. \end{aligned} \quad (3.7.7)$$

这就是说,当  $\mathbf{X}'\mathbf{X}$  至少有一个特征根很小时,最小二乘估计  $\hat{\boldsymbol{\beta}}$  的长度平均说来要比真正的未知向量  $\boldsymbol{\beta}$  的长度长得多.这就导致了  $\hat{\boldsymbol{\beta}}$  的某些分量的绝对值太大.

总之,当  $\mathbf{X}'\mathbf{X}$  至少有一个特征值很小时,最小二乘估计不再是一个好的估计.

下面我们来分析,“ $\mathbf{X}'\mathbf{X}$  至少有一个特征值很小”在设计阵  $\mathbf{X}$  本身或回归自变量关系上意味着什么.

记  $\mathbf{X} = (\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(p)})$ , 即  $\mathbf{x}_{(i)}$  为设计阵  $\mathbf{X}$  的第  $i$  列. 设  $\lambda$  为  $\mathbf{X}'\mathbf{X}$  的一个特征值,  $\boldsymbol{\varphi}$  为其对应的特征向量, 其长度为 1, 即  $\boldsymbol{\varphi}'\boldsymbol{\varphi} = 1$ . 若  $\lambda \approx 0$ , 则

$$\mathbf{X}'\mathbf{X}\boldsymbol{\varphi} = \lambda\boldsymbol{\varphi} \approx \mathbf{0}.$$

用  $\boldsymbol{\varphi}'$  左乘上式, 得

$$\boldsymbol{\varphi}'\mathbf{X}'\mathbf{X}\boldsymbol{\varphi} = \lambda\boldsymbol{\varphi}'\boldsymbol{\varphi} = \lambda \approx 0.$$

因为

$$\boldsymbol{\varphi}'\mathbf{X}'\mathbf{X}\boldsymbol{\varphi} = \|\mathbf{X}\boldsymbol{\varphi}\|^2 = \lambda \approx 0,$$

于是, 我们有

$$\mathbf{X}\boldsymbol{\varphi} \approx \mathbf{0}.$$

若记  $\boldsymbol{\varphi} = (c_1, \dots, c_p)'$ , 上式即为

$$c_1\mathbf{x}_{(1)} + \dots + c_p\mathbf{x}_{(p)} \approx \mathbf{0}. \quad (3.7.8)$$

这表明, 设计阵  $\mathbf{X}$  的列向量  $\mathbf{x}_{(1)}, \dots, \mathbf{x}_{(p)}$  之间有近似线性关系 (3.7.8). 如果用

$X_1, \dots, X_p$  分别表示  $p$  个回归自变量, 那么(3.7.8)说明, 从现有的  $n$  组数据看, 回归自变量之间有近似线性关系

$$c_1 X_1 + \dots + c_p X_p \approx 0, \quad (3.7.9)$$

回归设计阵的列向量之间的关系(3.7.8)或等价地回归自变量之间的关系(3.7.9), 称为复共线关系. 相应地, 称设计阵  $\mathbf{X}$  或线性回归模型(3.7.2)存在复共线性, 有时也称设计阵  $\mathbf{X}$  是病态的. 从上面的讨论过程我们知道, “ $\mathbf{X}'\mathbf{X}$  的特征根很小‘等价于’设计阵  $\mathbf{X}$  的列向量之间存在复共线关系”, 并且  $\mathbf{X}'\mathbf{X}$  有几个特征根很小, 设计阵  $\mathbf{X}$  就存在几个复共线关系. 因此, 复共线性是最小二乘估计性质变坏的原因, 对于大型线性回归问题, 也就是回归自变量个数  $p$  比较多的问题, 往往回归自变量之间存在某种程度的复共线性. 因而对于这样的回归问题, 最小二乘估计往往不够理想.

度量复共线性严重程度的一个重要量是方阵  $\mathbf{X}'\mathbf{X}$  的条件数, 定义为

$$k = \frac{\lambda_1}{\lambda_p},$$

也就是  $\mathbf{X}'\mathbf{X}$  的最大特征根与最小特征根之比. 直观上, 条件数刻画了  $\mathbf{X}'\mathbf{X}$  的特征值差异的大小. 从实际应用的经验角度, 一般若  $k < 100$ , 则认为复共线性的程度很小; 若  $100 \leq k \leq 1000$ , 则认为存在中等程度或较强的复共线性; 若  $k > 1000$ , 则认为存在严重的复共线性.

**例 3.7.1** 考虑一个有六个回归自变量的线性回归问题, 原始数据列在表 3.7.1.

表 3.7.1

数据号	Y	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
1	10.006	8.000	1.000	1.000	1.000	0.541	0.099
2	9.737	8.000	1.000	1.000	0.000	0.130	0.070
3	15.087	8.000	1.000	1.000	0.000	2.116	0.115
4	8.422	0.000	0.000	9.000	1.000	-2.397	0.252
5	8.625	0.000	0.000	9.000	1.000	-0.046	0.017
6	16.289	0.000	0.000	9.000	1.000	0.365	1.504
7	5.958	2.000	7.000	0.000	1.000	1.996	-0.865
8	9.313	2.000	7.000	0.000	1.000	0.228	-0.055
9	12.960	2.000	7.000	0.000	1.000	1.380	0.502
10	5.541	0.000	0.000	0.000	10.000	-0.798	-0.399
11	8.756	0.000	0.000	0.000	10.000	0.257	0.101
12	10.937	0.000	0.000	0.000	10.000	0.440	0.432

这里共有 12 组数据. 除第一组外, 自变量  $X_1, \dots, X_6$  的其余 11 组数据满足线性关系

$$X_1 + X_2 + X_3 + X_4 = 10. \quad (3.7.10)$$

将设计阵中心化和标准化,为简单计,仍用  $X_1, \dots, X_6$  表示. 从正态随机数表随机查出的 12 个数  $e_1, \dots, e_{12}$ . 通过理论线性回归关系

$$Y = 10 + 2.0X_1 + 1.0X_2 + 0.2X_3 - 2.0X_4 + 3.0X_5 + 10.0X_6 + e, \quad (3.7.11)$$

算出对应的因变量 12 个观测值, 这些值列在表 3.7.1 的第 1 列. 对于模型 (3.7.11),  $X'X$  为

$$\begin{bmatrix} 1.000 & 0.052 & -0.343 & -0.498 & 0.417 & -0.192 \\ & 1.000 & -0.432 & -0.371 & 0.485 & -0.317 \\ & & 1.000 & -0.355 & -0.505 & 0.494 \\ & & & 1.000 & -0.215 & -0.087 \\ & & & & 1.000 & -0.123 \\ & & & & & 1.000 \end{bmatrix}$$

我们知道, 对中心化和标准化的设计阵  $X$ , 若把回归自变量视为随机变量, 那末  $X'X$  就是回归自变量的相关阵. 从非对角元的绝对值看, 任两个回归自变量之间似乎不存在较严重的线性依赖关系, 而  $X'X$  的六个特征根分别为

$$\begin{aligned} \lambda_1 &= 2.24879, & \lambda_2 &= 1.54615, & \lambda_3 &= 0.92208, \\ \lambda_4 &= 0.79399, & \lambda_5 &= 0.30789, & \lambda_6 &= 0.00111, \end{aligned}$$

于是条件数为

$$k = \frac{\lambda_1}{\lambda_6} = \frac{2.24879}{0.00111} = 2025.94,$$

这个条件数远远大于 1000. 根据我们前面介绍的标准, 模型 (3.7.11) 的设计阵存在严重的复共线性. 因为  $\lambda_6 = 0.00111 \approx 0$ , 算出  $X'X$  对应于  $\lambda_6$  的特征向量

$$\begin{aligned} \varphi' &= (-0.44768, -0.42114, -0.54169, -0.57337, \\ &\quad -0.00605, -0.00217), \end{aligned}$$

因而回归自变量之间有如下复共线关系

$$\begin{aligned} 0.44768X_1 + 0.42114X_2 + 0.54169X_3 + 0.57337X_4 \\ + 0.00605X_5 + 0.00217X_6 \approx 0. \end{aligned}$$

注意到,  $X_5$  和  $X_6$  的系数和前面四个变量的系数相比要小很多. 可以将其略去, 得到

$$0.44768X_1 + 0.42114X_2 + 0.54169X_3 + 0.57337X_4 \approx 0. \quad (3.7.12)$$

我们看到,  $X_1, X_2, X_3$  和  $X_4$  的系数很相近, 于是这个复共线关系大体上反映了原来我们构造数据时所使用的关系 (3.7.10). 因为第一组数据并不满足 (3.7.10). 因此 (3.7.12) 跟 (3.7.10) 不完全相同也是自然的.

这个例子充分显示了本节介绍的寻找复共线性的方法是很有效的. 复共线

关系刻画了回归自变量之间的近似线性关系,找到这些线性关系不仅对我们进一步理解回归自变量之间的相互关系具有重要意义,有时还可以用于减少一些回归自变量,下面两节的例子可以帮助读者理解这一点.

### § 3.8 岭估计

从上节的讨论我们知道,当设计阵存在着复共线关系时,最小二乘估计的性质不够理想,有时甚至很坏.在这种情况下,我们就需要一些新的估计方法.近二十年来,人们提出了许多种新估计,其中在理论上最有影响并且得到广泛应用的,就是岭估计.

对于线性回归模型(3.7.2),回归系数  $\beta$  的岭估计定义为

$$\hat{\beta}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}, \quad (3.8.1)$$

这里  $k > 0$  是可选择参数,称为岭参数或偏参数.当  $k$  取不同的值时,我们得到不同的估计,因此岭估计  $\hat{\beta}(k)$  是一个估计类.当  $k = 0$ ,  $\hat{\beta}(k) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$  就是通常的最小二乘估计.于是严格地讲,最小二乘估计是岭估计类中的一个估计.但是一般情况下,当我们提起岭估计时,总是不包括最小二乘估计.

因为对一切  $k \neq 0$  和  $\beta \neq 0$

$$\begin{aligned} E\hat{\beta}(k) &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'E\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}\beta \\ &\neq \beta, \end{aligned}$$

因此岭估计是有偏估计,这是岭估计与最小二乘估计的一个重要不同之处.由定理 3.7.1 知,一个估计的均方误差由方差和偏差的平方两部分组成,当存在复共线关系时,最小二乘估计虽然仍保持偏差部分为零,但它的方差部分却很大,最终致使它的均方误差很大.我们引进岭估计是以牺牲无偏性,换取方差部分的大幅度减少,最终降低其均方误差,要达到这个目的, $k$  的选取至关重要.

为了证明关于岭估计优良性的一个基本定理,我们需要引进线性回归模型(3.7.2)的典则形式.设  $\lambda_1 \geq \dots \geq \lambda_p$  为  $\mathbf{X}'\mathbf{X}$  的特征根,  $\varphi_1, \dots, \varphi_p$  为对应的标准正交化特征向量.记  $\Phi = (\varphi_1, \dots, \varphi_p)$ , 则  $\Phi$  为  $p \times p$  正交阵.再记  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$ , 于是  $\mathbf{X}'\mathbf{X} = \Phi\Lambda\Phi'$ . 则线性回归模型(3.7.2)可改写为

$$\mathbf{y} = \alpha_0\mathbf{1} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2\mathbf{I}. \quad (3.8.2)$$

这里我们把常数项改记为  $\alpha_0$ ,  $\mathbf{Z} = \mathbf{X}\Phi$ ,  $\boldsymbol{\alpha} = \Phi'\beta$ . 我们称(3.8.2)为线性回归模型的典则形式,  $\boldsymbol{\alpha}$  称为典则回归系数.因为  $\mathbf{X}$  是中心化的,于是  $\mathbf{Z}$  也是中心化的.对典则形式(3.8.2),  $\alpha_0$  和  $\boldsymbol{\alpha}$  的最小二乘估计为

$$\hat{\alpha}_0 = \bar{y},$$

$$\hat{\alpha} = (Z'Z)^{-1}Z'y.$$

注意到  $Z'Z = \Phi'X'X\Phi = \Lambda$ , 因而

$$\begin{aligned}\hat{\alpha} &= \Lambda^{-1}Z'y, \\ \text{Cov}(\hat{\alpha}) &= \sigma^2 \Lambda^{-1}Z'Z\Lambda^{-1} = \sigma^2 \Lambda^{-1}.\end{aligned}$$

容易证明

$$\hat{\alpha} = \Phi'\hat{\beta}, \quad (3.8.3)$$

这里  $\hat{\beta} = (X'X)^{-1}X'y$ . 根据关系式  $\alpha = \Phi'\beta$ , 按定义典则回归系数的岭估计为

$$\hat{\alpha}(k) = \Phi'\hat{\beta}(k), \quad (3.8.4)$$

于是

$$\begin{aligned}\hat{\alpha}(k) &= \Phi'(X'X + kI)^{-1}X'y \\ &= \Phi'(\Phi\Lambda\Phi' + kI)^{-1}\Phi Z'y \\ &= (\Lambda + kI)^{-1}Z'y.\end{aligned} \quad (3.8.5)$$

由(3.8.3)看出, 典则参数  $\alpha$  的最小二乘估计与原来参数  $\beta$  的最小二乘估计之间差一个正交阵, 因而有

$$\text{MSE}(\hat{\alpha}) = \text{MSE}(\hat{\beta}). \quad (3.8.6)$$

从(3.8.4)知, 类似的结论对岭估计也成立, 即

$$\text{MSE}(\hat{\alpha}(k)) = \text{MSE}(\hat{\beta}(k)). \quad (3.8.7)$$

这两个等式很重要, 它对证明岭估计的优良性带来很大方便.

现在我们证明岭估计优良性的基本定理.

**定理 3.8.1** 存在  $k > 0$ , 使得

$$\text{MSE}(\hat{\beta}(k)) < \text{MSE}(\hat{\beta}). \quad (3.8.8)$$

即存在  $k > 0$ , 使得在均方误差意义下, 岭估计优于最小二乘估计.

**证明** 根据(3.8.6)和(3.8.7)知, 只需证明存在  $k > 0$ , 使得

$$\text{MSE}(\hat{\alpha}(k)) < \text{MSE}(\hat{\alpha}). \quad (3.8.9)$$

因为设计阵  $Z$  也是中心化的, 于是  $1'Z = 0$ . 所以

$$\begin{aligned}E\hat{\alpha}(k) &= (\Lambda + kI)^{-1}Z'(\alpha_0\mathbf{1} + Z\alpha) \\ &= (\Lambda + kI)^{-1}Z'Z\alpha \\ &= (\Lambda + kI)^{-1}\Lambda\alpha.\end{aligned}$$

应用定理 2.1.3 有

$$\begin{aligned}\text{Cov}\hat{\alpha}(k) &= \sigma^2(\Lambda + kI)^{-1}Z'Z(\Lambda + kI)^{-1} \\ &= \sigma^2(\Lambda + kI)^{-1}\Lambda(\Lambda + kI)^{-1}.\end{aligned}$$

再由定理 3.7.1, 得到

$$\begin{aligned}\text{MSE}(\hat{\alpha}(k)) &= \text{tr}\text{Cov}\hat{\alpha}(k) + \|E\hat{\alpha}(k) - \alpha\|^2 \\ &= \sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \sum_{i=1}^p \frac{\alpha_i^2}{(\lambda_i + k)^2}\end{aligned} \quad (3.8.10)$$

$$\begin{aligned} &= f_1(k) + f_2(k) \\ &= f(k), \end{aligned}$$

这里  $f_1(k)$  和  $f_2(k)$  分别表示(3.8.10)的第一项和第二项. 对  $k$  求导数, 得

$$f_1'(k) = -2\sigma^2 \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + k)^3}, \quad (3.8.11)$$

$$f_2'(k) = 2k \sum_{i=1}^p \frac{\lambda_i \alpha_i^2}{(\lambda_i + k)^3}. \quad (3.8.12)$$

因为

$$f_1'(0) = -2\sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i^2} < 0,$$

$$f_2'(0) = 0,$$

于是  $f'(0) < 0$ . 但是  $f_1'(k)$  和  $f_2'(k)$  在  $k \geq 0$  时都连续, 所以,  $f'(k)$  在  $k \geq 0$  时也连续. 因而当  $k > 0$  且充分小时  $f'(k) < 0$ . 这就是说,  $f(k) = \text{MSE}(\hat{\alpha}(k))$  在  $k (> 0)$  充分小时, 是  $k$  的单调减函数, 因而存在  $k^* > 0$ , 当  $k \in (0, k^*)$  时, 有  $f(k) < f(0)$ . 但  $f(0) = \text{MSE}(\hat{\alpha})$ . 这就证明了(3.8.9). 定理证毕.

**注 1** 这个定理有很重要的意义, 它为岭估计的实际应用奠定了理论基础. 但是, 从理论证明过程我们知道, 使得(3.8.8)成立的  $k$  依赖于未知参数. 因此, 对找到的一个  $k$ , (3.8.8)并不能对一切的  $\beta$  和  $\sigma^2$  成立. 事实上可以证明, 它只能对相对比较小的  $\beta$  成立(关于这个事实的证明可参阅王松桂(1987), p. 294, 定理 2.2).

**注 2** 对任意  $k > 0$  和  $\|\hat{\beta}\| \neq 0$  总有

$$\|\hat{\beta}(k)\| = \|\hat{\alpha}(k)\| = \|(\Lambda + kI)^{-1}\Lambda\hat{\alpha}\| < \|\hat{\alpha}\| = \|\hat{\beta}\|,$$

这表明, 岭估计  $\hat{\beta}(k)$  的长度总比最小二乘估计  $\hat{\beta}$  的长度小. 因此  $\hat{\beta}(k)$  是对  $\hat{\beta}$  向原点的一种压缩, 所以通常也称  $\hat{\beta}(k)$  是一种压缩估计.

在实际应用中, 岭参数的选择是一个很重要的问题. 但是定理 3.8.1 只是告诉我们使得  $\hat{\beta}(k)$  优于  $\hat{\beta}$  的  $k$  的存在性, 并没有给出具体的算法. 我们自然希望找到使  $\text{MSE}\hat{\beta}(k)$  达到最小值的  $k^*$ . 从(3.8.11)和(3.8.12)容易看出, 这个最优值  $k^*$  应该在方程

$$f'(k) = 2 \sum_{i=1}^p \frac{\lambda_i (k\alpha_i^2 - \sigma^2)}{(\lambda_i + k)^3} = 0 \quad (3.8.13)$$

的根中去找, 这个方程可以化成关于  $k$  的 4 次多项式. 但是这个多项式方程的系数依赖于未知参数  $\alpha_i$  和  $\sigma^2$ , 因而不可能从求解(3.8.13)获得最优值  $k^*$ . 因此统计学家们从别的途径提出了选择参数  $k$  的许多方法. 从计算机模拟比较的结果看, 在这些方法中没有一个方法能够一致地(即对一切参数  $\beta$  和  $\sigma^2$ )优于其它

方法.下面我们介绍目前应用较多的两种方法.

### 1. Hoerl - Kennard 公式

岭估计是由 Hoerl 和 Kennard 于 1970 年提出的,他们所用的选择  $k$  值的公式是

$$\hat{k} = \frac{\hat{\sigma}^2}{\max_i \hat{\alpha}_i^2}. \quad (3.8.14)$$

获得这个公式的想法如下,由(3.8.13)知,如果  $k\alpha_i^2 - \sigma^2 < 0$ ,对  $i=1, \dots, p$  都成立,则  $f'(k) < 0$ .于是取

$$k^* = \frac{\sigma^2}{\max_i \alpha_i^2}, \quad (3.8.15)$$

当  $0 < k < k^*$  时,  $f'(k)$  总是小于 0,因而  $f(k)$  总是  $k$  的单调下降函数,故有  $f(k^*) < f(0)$ ,即  $MSE(\hat{\beta}(k^*)) < MSE(\hat{\beta})$ .在(3.8.15)中,用最小二乘估计  $\hat{\alpha}_i$  和  $\hat{\sigma}^2$  代替  $\alpha_i$  和  $\sigma^2$ ,便得到(3.8.14).

### 2. 岭迹法

岭估计  $\hat{\beta}(k) = (X'X + kI)^{-1}X'y$  是随  $k$  值改变而变化.若记  $\hat{\beta}_i(k)$  为  $\hat{\beta}(k)$  的第  $i$  个分量,它是  $k$  的一元函数.当  $k$  在  $[0, +\infty)$  上变化时,  $\hat{\beta}(k)$  的图形称为岭迹 (ridge trace).选择  $k$  的岭迹法是:将  $\hat{\beta}_1(k), \dots, \hat{\beta}_p(k)$  的岭迹画在同一个图上,根据岭迹的变化趋势选择  $k$  值,使得各个回归系数的岭估计大体上稳定,并且各个回归系数岭估计值的符号比较合理.我们知道,最小二乘估计是使残差平方和达到最小的估计. $k$  愈大,岭估计跟最小二乘估计偏离愈大.因此,它对应的残差平方和也随着  $k$  的增加而增加.当我们用岭迹法选择  $k$  值时,还应考虑使得残差平方和不要上升太多.在实际处理上,上述几点原则有时可能会有些互相不一致,顾此失彼的情况也经常出现,这就要根据不同情况灵活处理.

#### 例 3.8.1 外贸数据分析.

我们所考虑的因变量  $Y$  为进口总额,自变量  $X_1$  为国内总产值,  $X_2$  为存储量,  $X_3$  为总消费量.为了建立  $Y$  对自变量  $X_1, X_2$  和  $X_3$  之间的依赖关系,收集了 11 组数据,列在表 3.8.1.这个问题包含的自变量个数仅仅为 3,很容易从计算机上算出所有可能的最小二乘回归,也就是包含一个自变量  $X_1$  或  $X_2$  或  $X_3$  的回归,和包含两个自变量  $X_1, X_2$  或  $X_1, X_3$  或  $X_2, X_3$  的回归,以及包含全部三个自变量  $X_1, X_2$  和  $X_3$  的回归.其结果列在表 3.8.2.将原始数据中心化和标准化,计算得到

$$X'X = \begin{pmatrix} 1 & 0.026 & 0.997 \\ 0.026 & 1 & 0.036 \\ 0.997 & 0.036 & 1 \end{pmatrix}.$$



表 3.8.1 外贸数据

序号	国内总产值 ( $x_1$ )	存储量 ( $x_2$ )	总消费量 ( $x_3$ )	进口总额 ( $y$ )
1	149.3	4.2	108.1	15.9
2	161.2	4.1	114.8	16.4
3	171.5	3.1	123.2	19.0
4	175.5	3.1	126.9	19.1
5	180.8	1.1	132.1	18.8
6	190.7	2.2	137.7	20.4
7	202.1	2.1	146.0	22.7
8	212.4	5.6	154.1	26.5
9	226.1	5.0	162.3	28.1
10	231.9	5.1	164.3	27.6
11	239.0	0.7	167.6	26.3

表 3.8.2 外贸数据分析问题的所有可能的最小二乘回归

进入回归 的变量	回归系数的最小二乘估计		
	$x_1$	$x_2$	$x_3$
1	0.146	—	—
2	—	0.691	—
3	—	—	0.214
1,2	0.145	0.622	—
1,3	-0.109	—	0.372
2,3	—	0.596	0.212
1,2,3	-0.051	0.587	0.287

再计算出它的三个特征值,分别为  $\lambda_1 = 1.999$ ,  $\lambda_2 = 0.998$ ,  $\lambda_3 = 0.003$ . 于是  $X'X$  的条件数  $\lambda_1/\lambda_3 = 666.333$ , 可见设计矩阵存在中等程度的复共线性. 视  $\lambda_3 \approx 0$ . 对应的特征向量为

$$\varphi_3 = (-0.7070, -0.0070, 0.7072).$$

由 § 3.7 的讨论知,三个自变量之间存在复共线关系

$$-0.7070X_1 - 0.0070X_2 + 0.7072X_3 \approx 0.$$

注意到,自变量  $X_2$  的系数绝对值相对非常小,可视为零,而  $X_1$  和  $X_3$  的系数又近似相等,因此自变量之间的复共线关系可近似地写为  $X_1 = X_3$ . 注意,这里的  $X_1$  和  $X_3$  都是经过中心化和标准化的变量,还原到原来的变量,近似复共线关

系为

$$\frac{X_1 - \bar{x}_1}{s_1} = \frac{X_3 - \bar{x}_3}{s_3}.$$

从表 3.8.1 可以算出

$$\begin{aligned}\bar{x}_1 &= 194.59, & s_1 &= \left[ \sum_{i=1}^{11} (x_{i1} - \bar{x}_1)^2 \right]^{1/2} = 94.87, \\ \bar{x}_3 &= 139.74, & s_3 &= \left[ \sum_{i=1}^{11} (x_{i3} - \bar{x}_3)^2 \right]^{1/2} = 65.25.\end{aligned}$$

代入上式得

$$X_3 = 5.905 + 0.688X_1. \quad (3.8.16)$$

这就是总消费量和国内总产值之间的一个线性依赖关系, 因为  $X$  是中心化和标准化的, 于是  $XX'$  是相关阵, 其中的 0.997 正是  $X_1$  与  $X_3$  的相关系数. 可见,  $X_1$  与  $X_3$  有如此大的相关系数, 和我们找出它们之间的复共线关系 (3.8.16) 这事实是吻合的. 这就是总消费量和国内总产值之间的一个线性关系. 既然自变量之间存在中等程度的复共线性, 我们就采用岭估计来估计回归系数. 对于中心化和标准化的变量, 计算出的岭迹列在表 3.8.3, 对应的岭迹图画在图 3.8.1. 表 3.8.3 的最后一列是岭估计对应的残差平方和. 我们看到, 随着  $k$  的增加, 岭估计的残差平方和也随之增加, 所以残差平方和是岭迹  $k$  的单调增函数. 这是很自然的, 因为最小二乘估计是使残差平方和达到最小的估计. 随着  $k$  的增加, 岭估计跟最小二乘估计偏离就愈大, 因此它的残差平方和自然也就愈大. 从岭迹图上可以看出, 岭迹  $\hat{\beta}_1(k)$  随着  $k$  的增加, 很快增加, 大约在  $k=0.01$  处从负值变为正值. 而  $\hat{\beta}_2(k)$  相对比较稳定, 但  $\hat{\beta}_3(k)$  随着  $k$  的增加, 骤然减少, 大约在 0.04 以后就稳定下来. 总体看来, 我们可以取  $k=0.04$ , 对应的岭估计为

$$\hat{\beta}_1(0.04) = 0.420, \hat{\beta}_2(0.04) = 0.213, \hat{\beta}_3(0.04) = 0.525.$$

各变量的平均值分别为

$$\bar{x}_1 = 194.59, \bar{x}_2 = 3.30, \bar{x}_3 = 139.74, \bar{y} = 21.89.$$

相应的

$$s_1 = 94.87, s_2 = 5.22, s_3 = 65.26, s_y = 14.37.$$

代入经验回归方程

$$\frac{\hat{Y} - \bar{y}}{s_y} = \hat{\beta}_1(0.04) \frac{X_1 - \bar{x}_1}{s_1} + \hat{\beta}_2(0.04) \frac{X_2 - \bar{x}_2}{s_2} + \hat{\beta}_3(0.04) \frac{X_3 - \bar{x}_3}{s_3},$$

化简后得到如下岭回归方程:

$$\hat{Y} = -8.5537 + 0.0635X_1 + 0.5859X_2 + 0.1156X_3.$$

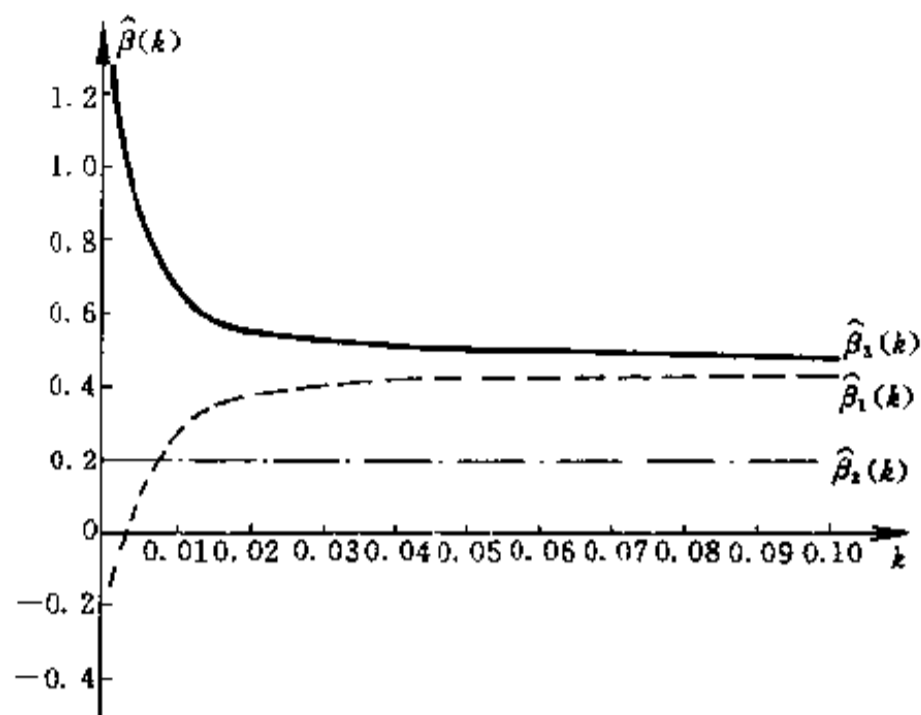


图 3.8.1 外贸数据回归的岭迹

表 3.8.3 外贸数据的岭回归

$k$	$\hat{\beta}_1(k)$	$\hat{\beta}_2(k)$	$\hat{\beta}_3(k)$	RSS
0.000	-0.339	0.213	1.303	1.673
0.001	-0.117	0.215	1.080	1.728
0.002	0.010	0.216	0.952	1.809
0.003	0.092	0.217	0.870	1.881
0.004	0.150	0.217	0.811	1.941
0.005	0.193	0.217	0.768	1.990
0.006	0.225	0.217	0.735	2.031
0.007	0.251	0.217	0.709	2.066
0.008	0.272	0.217	0.687	2.095
0.009	0.290	0.217	0.669	2.120
0.010	0.304	0.217	0.654	2.142
0.020	0.379	0.216	0.575	2.276
0.030	0.406	0.214	0.543	2.352
0.040	0.420	0.213	0.525	2.416
0.050	0.427	0.211	0.513	2.480
0.060	0.432	0.209	0.504	2.548
0.070	0.434	0.207	0.497	2.623
0.080	0.436	0.206	0.491	2.705
0.090	0.436	0.204	0.486	2.794

续表

$k$	$\hat{\beta}_1(k)$	$\hat{\beta}_2(k)$	$\hat{\beta}_3(k)$	RSS
0.100	0.436	0.202	0.481	2.890
0.200	0.426	0.186	0.450	4.236
0.300	0.411	0.173	0.427	6.155
0.400	0.396	0.161	0.408	8.489
0.500	0.381	0.151	0.391	11.1.7
0.600	0.367	0.142	0.376	13.947
0.700	0.354	0.135	0.361	16.9.1
0.800	0.342	0.128	0.348	19.957
0.900	0.330	0.121	0.336	23.047
1.000	0.319	0.115	0.325	26.149

在一般统计软件中,都包含有岭估计计算程序(如目前国内最常用的 SAS 软件,SPSS 软件),只要输入原始数据,计算机就可输出常用的估计量以及有关统计量,同时还可以算出岭迹并作出岭迹图.

### § 3.9 主成分估计

考虑一般线性回归模型

$$\mathbf{y} = \alpha_0 \mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}, \quad (3.9.1)$$

假设  $\mathbf{X}$  设计阵已经中心化,于是常数项  $\alpha_0$  的最小二乘估计  $\hat{\alpha}_0 = \bar{y}$ , 记  $\lambda_1 \geq \dots \geq \lambda_p$  为  $\mathbf{X}'\mathbf{X}$  的特征根,  $\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_p$  为对应的标准正交化特征向量,  $\boldsymbol{\Phi} = (\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_p)$ , 这是一个  $p \times p$  正交阵,再记  $\mathbf{Z} = \mathbf{X}\boldsymbol{\Phi}$ ,  $\boldsymbol{\alpha} = \boldsymbol{\Phi}'\boldsymbol{\beta}$ , 则模型(3.9.1)可变形为

$$\mathbf{y} = \alpha_0 \mathbf{1} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}, \quad (3.9.2)$$

这是我们在上节引进的线性回归典则形式. 称  $\boldsymbol{\alpha}$  为典则参数. 在模型(3.9.2)中,新的设计阵  $\mathbf{Z} = (\mathbf{z}_{(1)}, \dots, \mathbf{z}_{(p)}) = (\mathbf{X}\boldsymbol{\varphi}_1, \dots, \mathbf{X}\boldsymbol{\varphi}_p)$ , 即

$$\mathbf{z}_{(1)} = \mathbf{X}\boldsymbol{\varphi}_1, \dots, \mathbf{z}_{(p)} = \mathbf{X}\boldsymbol{\varphi}_p, \quad (3.9.3)$$

于是  $\mathbf{Z}$  的第  $i$  列  $\mathbf{z}_{(i)}$  是原来  $p$  个自变量的线性组合,其组合系数为  $\mathbf{X}'\mathbf{X}$  的第  $i$  个特征根对应的特征向量  $\boldsymbol{\varphi}_i$ . 因此,  $\mathbf{Z}$  的  $p$  个列就对应于  $p$  个以原来变量的特殊线性组合(即以  $\mathbf{X}'\mathbf{X}$  的特征向量为组合系数)构成的新变量. 在统计学上,称这些新变量为主成分. 排在第 1 列的新变量对应于  $\mathbf{X}'\mathbf{X}$  的最大特征根,于是称为第一主成分,排在第 2 列的就称为第二主成分,依次类推. 因为  $\mathbf{X}$  是中心化的,即  $\mathbf{1}'\mathbf{X} = \mathbf{0}$ , 于是  $\mathbf{1}'\mathbf{Z} = \mathbf{1}'\mathbf{X}\boldsymbol{\Phi} = \mathbf{0}$ . 所以  $\mathbf{Z}$  也是中心化的. 因而  $\mathbf{Z}$  的各列元的平均

值

$$z_j = \frac{1}{n} \sum_{i=1}^n z_{ij} = 0, j = 1, \dots, p. \quad (3.9.4)$$

由(3.9.3)可得

$$z'_{(i)} z_{(i)} = \boldsymbol{\varphi}'_i \mathbf{X}' \mathbf{X} \boldsymbol{\varphi}_i = \lambda_i. \quad (3.9.5)$$

结合(3.9.4)知

$$\sum_{i=1}^n (z_{ij} - \bar{z}_j)^2 = z'_{(i)} z_{(i)} = \lambda_i, \quad i = 1, \dots, p.$$

于是  $\mathbf{X}'\mathbf{X}$  的第  $i$  个特征根  $\lambda_i$  就度量了第  $i$  个主成分取值变动大小. 当设计阵  $\mathbf{X}$  存在复共线关系时, 有一些  $\mathbf{X}'\mathbf{X}$  的特征值很小, 不妨假设  $\lambda_{r+1}, \dots, \lambda_p \approx 0$ . 这时后面的  $p-r$  个主成分取值变动就很小, 再结合(3.9.4)(即它们的均值都为零), 因而这些主成分取值近似为零. 因此, 在用主成分作为新的回归自变量时, 这后面的  $p-r$  个主成分对因变量的影响就可以忽略掉, 故可将它们从回归模型中剔除. 用最小二乘法做剩下的  $r$  个主成分的回归, 然后再变回到原来的自变量, 就得到了主成分回归.

现在将上述思想具体化. 记  $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ , 对  $\boldsymbol{\Lambda}, \boldsymbol{\alpha}, \mathbf{Z}$  和  $\boldsymbol{\Phi}$  做分块:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_2 \end{bmatrix},$$

其中  $\boldsymbol{\Lambda}_1$  为  $r \times r$  矩阵,

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix},$$

其中  $\boldsymbol{\alpha}_1$  为  $r \times 1$  向量,

$$\mathbf{Z} = (\mathbf{Z}_1 : \mathbf{Z}_2),$$

其中  $\mathbf{Z}_1$  为  $n \times r$  矩阵,

$$\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1 : \boldsymbol{\Phi}_2),$$

其中  $\boldsymbol{\Phi}_1$  为  $p \times r$  矩阵, 代入(3.9.2)并剔除  $\mathbf{Z}_2 \boldsymbol{\alpha}_2$  项得到回归模型

$$\mathbf{y} = \alpha_0 \mathbf{1} + \mathbf{Z}_1 \boldsymbol{\alpha}_1 + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}, \quad (3.9.6)$$

这个新的回归模型就是在剔除了后面  $p-r$  个对因变量影响较小的主成分后得到的. 因此, 事实上我们是利用主成分进行了一次回归自变量的选择. 对模型(3.9.6)应用最小二乘法, 得到  $\alpha_0$  和  $\boldsymbol{\alpha}_1$  的最小二乘估计:

$$\hat{\alpha}_0 = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$\hat{\boldsymbol{\alpha}}_1 = (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 \mathbf{y} = \boldsymbol{\Lambda}_1^{-1} \mathbf{Z}'_1 \mathbf{y}.$$

前面我们从模型中剔除了后面  $p-r$  个主成分, 这相当于用  $\hat{\boldsymbol{\alpha}}_2 = \mathbf{0}$  去估计  $\boldsymbol{\alpha}_2$ .

利用关系  $\beta = \Phi\alpha$ , 可以获得原来参数  $\beta$  的估计

$$\begin{aligned}\hat{\beta} &= \Phi \begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = (\Phi_1, \Phi_2) \begin{pmatrix} \hat{\alpha}_1 \\ \mathbf{0} \end{pmatrix} \\ &= \Phi_1 \Lambda_1^{-1} Z_1' y = \Phi_1 \Lambda_1^{-1} \Phi_1' X' y.\end{aligned}\quad (3.9.7)$$

这就是  $\beta$  的主成分估计.

从上面的讨论, 我们可以把获得主成分估计的方法归纳如下:

(a) 做正交变换  $Z = X\Phi$ , 获得新的自变量, 称为主成分.

(b) 做回归自变量选择: 剔除对应的特征值比较小的那些主成分.

(c) 将剩余的主成分对  $Y$  做最小二乘回归, 再返回到原来的参数, 便得到因变量对原始自变量的主成分回归.

在前面的讨论中, 我们从直观分析的角度引进了主成分估计. 一个重要的问题是, 主成分估计有什么良好性质? 不难证明, 它也是有偏估计. 事实上, 由 (3.9.7) 知

$$E(\hat{\beta}) = (\Phi_1 : \Phi_2) \begin{pmatrix} \alpha_1 \\ \mathbf{0} \end{pmatrix} = \Phi_1 \alpha_1,$$

但

$$\beta = \Phi\alpha = \Phi_1 \alpha_1 + \Phi_2 \alpha_2,$$

可见, 一般说来  $E(\hat{\beta}) \neq \beta$ , 于是主成分估计也是有偏估计. 对于有偏估计, 我们应该用均方误差作为度量其优劣的标准. 下面的定理证明了, 在一定条件下主成分估计比最小二乘估计有较小的均方误差.

**定理 3.9.1** 当设计阵存在复共线关系时, 适当选择保留的主成分个数可致主成分估计比最小二乘估计有较小的均方误差, 即

$$MSE(\hat{\beta}) < MSE(\hat{\beta}').$$

**证明** 利用前面的记号, 假设  $X'X$  的后面  $p-r$  个特征根  $\lambda_{r+1}, \dots, \lambda_p$  很接近于 0, 我们利用 (3.9.7) 和定理 3.7.1, 有

$$\begin{aligned}MSE(\hat{\beta}) &= MSE \begin{pmatrix} \hat{\alpha} \\ \mathbf{0} \end{pmatrix} \\ &= \text{tr Cov} \begin{pmatrix} \hat{\alpha} \\ \mathbf{0} \end{pmatrix} + \left\| E \begin{pmatrix} \hat{\alpha} \\ \mathbf{0} \end{pmatrix} - \alpha \right\|^2 \\ &= \sigma^2 \text{tr}(\Lambda_1^{-1}) + \|\alpha_2\|^2.\end{aligned}$$

因为

$$MSE(\hat{\beta}') = \sigma^2 \text{tr}(\Lambda^{-1}),$$

所以

$$MSE(\boldsymbol{\beta}) = MSE(\hat{\boldsymbol{\beta}}) + (\|\boldsymbol{\alpha}_2\|^2 - \sigma^2 \text{tr} \mathbf{A}_2^{-1}).$$

于是

$$MSE(\boldsymbol{\beta}) < MSE(\hat{\boldsymbol{\beta}}),$$

当且仅当

$$\|\boldsymbol{\alpha}_2\|^2 < \sigma^2 \text{tr} \mathbf{A}_2^{-1} = \sigma^2 \sum_{i=r+1}^p \frac{1}{\lambda_i}. \quad (3.9.8)$$

因为我们假定  $\mathbf{X}'\mathbf{X}$  的后面  $p-r$  个特征值接近于 0, 于是上式右端很大, 故不等式(3.9.8)成立. 定理得证.

**注** 因为  $\boldsymbol{\alpha}_2 = \Phi_2' \boldsymbol{\beta}$ , 于是变回到原来参数, (3.9.8)可变形为

$$\left( \frac{\boldsymbol{\beta}}{\sigma} \right)' \Phi_2 \Phi_2' \left( \frac{\boldsymbol{\beta}}{\sigma} \right) \leq \text{tr} \mathbf{A}_2^{-1}, \quad (3.9.9)$$

这就是说, 仅当  $\boldsymbol{\beta}$  和  $\sigma^2$  满足(3.9.9)时, 主成分估计才能比最小二乘估计有较小的均方误差. (3.9.9)表示了参数空间中(视  $\frac{\boldsymbol{\beta}}{\sigma}$  为参数)一个中心在原点的椭球.

于是从(3.9.9)我们可以得到如下直观的结论:

(a) 对固定的参数  $\boldsymbol{\beta}$  和  $\sigma^2$ , 当  $\mathbf{X}'\mathbf{X}$  的后面的  $p-r$  个特征值很小时, 主成分估计比最小二乘估计有较小的均方误差.

(b) 对给定的  $\mathbf{X}'\mathbf{X}$ , 也就是固定的  $\mathbf{A}_2$ , 对相对比较小的  $\frac{\boldsymbol{\beta}}{\sigma}$ , 主成分估计比最小二乘估计有较小的均方误差.

在主成分估计应用中, 有一个重要问题就是如何选择保留的主成分个数. 通常有两种方法, 其一是保留对应的特征根相对比较大的那些主成分. 其二是选择  $r$ , 使得  $\sum_{i=1}^r \lambda_i$  与全部  $p$  个特征值之和  $\sum_{i=1}^p \lambda_i$  的比值(称这个比值为前  $r$  个主成分的贡献率)达到预先给定值, 譬如 75% 或 80% 等.

需要说明一点, 主成分作为原来变量的线性组合, 是一种“人造变量”, 一般并不具有任何实际含义, 特别当回归自变量具有不同度量单位时更是如此. 例如在研究农作物产量与气候条件、生产条件的关系问题中, 假定  $X_1$  和  $X_2$  分别表示该作物生长期日平均气温和降雨量, 它们的度量单位分别是  $1^\circ\text{C}$  (摄氏度) 和 mm(毫米), 而  $X_3$  表示单位面积上化学肥料的施用量, 单位是 kg(公斤). 这时主成分作为这些变量的线性组合, 它们的单位就什么都不是了, 更谈不上其实际意义. 当然也存在一些实际问题, 自变量都是同一类型的物理量, 它们具有相同的度量单位, 并且它们的主成分具有十分明显的实际解释.

**例 3.9.1** (续例 3.8.1)外贸数据分析问题.

在例 3.8.1 中, 我们已经对这批数据作了统计分析, 并且求出了回归系数的

岭估计,现在我们求它的主成分估计.  $\mathbf{X}'\mathbf{X}$  的三个特征值分别为

$$\lambda_1 = 1.999, \lambda_2 = 0.998, \lambda_3 = 0.003,$$

它们对应的三个标准正交化特征向量分别为

$$\varphi'_1 = (0.7063, 0.0435, 0.7065),$$

$$\varphi'_2 = (-0.0357, 0.9990, -0.0258),$$

$$\varphi'_3 = (-0.7070, -0.0070, 0.7072).$$

三个主成分分别为

$$z_1 = 0.7063X_1 + 0.0435X_2 + 0.7065X_3,$$

$$z_2 = -0.0357X_1 + 0.9990X_2 - 0.0258X_3,$$

$$z_3 = -0.7070X_1 - 0.0070X_2 + 0.7072X_3.$$

注意这里  $X_1$ 、 $X_2$  和  $X_3$  是中心化和标准化后的变量. 因为  $\lambda_3 \approx 0$ , 且前两个主成分的贡献率

$$\sum_{i=1}^2 \lambda_i / \sum_{i=1}^3 \lambda_i = 0.999 = 99.9\%.$$

因此, 我们剔除第三个主成分, 只保留前两个主成分, 它们的回归系数的最小二乘估计分别为  $\hat{\alpha}_1 = 0.6900$ ,  $\hat{\alpha}_2 = 0.1913$ . 还原到原来变量, 得到经验回归方程

$$\hat{Y} = -9.1057 + 0.0727X_1 + 0.6091X_2 + 0.1062X_3.$$

表 3.9.1 给出了主成分估计、岭估计和最小二乘估计. 总起来讲, 主成分估计和岭估计比较相近. 而跟最小二乘估计相比, 复共线关系 (3.8.16) 所包含的  $X_1$  和  $X_3$  的回归系数变化较大, 并且  $X_1$  的回归系数的符号也发生了变化.

表 3.9.1 外贸数据分析问题的三种估计

变量	常数项	$x_1$	$x_2$	$x_3$
主成分估计 ( $r=2$ )	-9.1057	0.0727	0.6091	0.1062
最小二乘估计	-10.1300	-0.0514	0.5869	0.2868
岭估计 ( $k=0.04$ )	-8.5537	0.0635	0.5859	0.1156

### 习题三

3.1 假设我们要求出四个物体的重量  $\mu_1, \mu_2, \mu_3$  和  $\mu_4$ . 一种方法是将每个物体称  $k$  次, 譬如称 5 次, 然后求平均. 假定称量的误差的方差都为  $\sigma^2$ . 用  $y_{ij}$  表示第  $i$  个物体第  $j$  次称重时得到的重量,  $i=1, 2, 3, 4; j=1, \dots, 5$ .

(1) 试写出相应的线性模型.

(2) 求出  $\mu_i$  的最小二乘估计  $\hat{\mu}_i$ , 这些估计正是每个物体 5 次称重的平均值.



(3) 计算  $\text{Var}(\hat{\beta}_i)$ .

3.2 对上题讨论的称重问题,另一种称法是组合法,方法是这样的:在天平的两个称盘上分别放上这四个物体中的几个物体,并在其中的一个称盘上加上砝码使之达到平衡.这样便有一个线性回归模型

$$Y = \mu_1 X_1 + \mu_2 X_2 + \mu_3 X_3 + \mu_4 X_4 + e,$$

其中  $Y$  为使天平达到平衡所需的砝码重量.我们约定,如果砝码放在左边称盘上则  $Y$  应为负值.  $X_i$  的值为 0, 1 或 -1. 0 表示在这次称重时,第  $i$  个物体没有被称; 1 和 -1 分别表示该物体放在左边和右边称盘上.回归系数  $\mu_i$  就是第  $i$  个物体的重量,我们总共称了四次,其结果如下表:

$Y$	$X_1$	$X_2$	$X_3$	$X_4$
20.2	1	1	1	1
8.0	1	-1	1	-1
9.7	1	1	-1	-1
1.9	1	-1	-1	1

(1) 试用线性回归模型表示这些称重数据;

(2) 验证设计矩阵  $X$  满足  $X'X = 4I_4$ , 并计算物体重量  $\mu_i$  的最小二乘估计  $\hat{\beta}_i$ ;

(3) 假设模型误差的方差为  $\sigma^2$ , 证明  $\text{Var}(\hat{\beta}_i) = \sigma^2/4$ ;

(4) 如果这些物体是用上题的方法分别称重,  $\mu_i$  的估计要达到这样的精度:  $\text{Var}(\hat{\beta}_i) = \sigma^2/4$ , 需要称多少次?

3.3 写出过原点的一元、二元线性回归模型, 并分别求出回归系数的最小二乘估计.

3.4 为考察某种维尼纶纤维的耐水性能, 安排了一批试验, 测得甲醛浓度  $x$  及“缩醛化度”  $y$  的数据如下:

$x$	18	20	22	24	26	28	30
$y$	26.86	28.35	28.75	28.87	29.75	30.00	30.36

从实际经验及理论可知二者近似线性关系. 求  $y$  关于  $x$  的经验回归方程, 并画出原始数据及回归直线的图形.

3.5 在动物学研究中, 有时需要找出某种动物的体积与重量的关系, 因为重量相对容易测量, 而测量体积比较困难. 我们可以利用重量预测体积的值. 下面是某种动物的 18 个随机样本的体重  $X$  (kg) 与体积  $Y$  ( $10^{-3} \text{m}^3$ ) 的数据:

$X$	17.1	10.5	13.8	15.7	11.9	10.4	15.0	16.0	17.8	15.8
$Y$	16.7	10.4	13.5	15.7	11.6	10.2	14.5	15.8	17.6	15.2
$X$	15.1	12.1	18.4	17.1	16.7	16.5	15.1	15.1		
$Y$	14.8	11.9	18.3	16.7	16.6	15.9	15.1	14.5		

(1) 画出数据散点图, 并求回归直线  $y = \hat{\beta}_0 + \hat{\beta}_1 x$ ;

(2) 对体重  $X_0 = 15.3$  的这种动物, 试预测它的体积  $Y_0$ .

$$\begin{aligned}
 3.6 \quad & \text{设 } y_1 = \theta + e_1, \\
 & y_2 = 2\theta - \varphi + e_2, \\
 & y_3 = \theta + 2\varphi + e_3.
 \end{aligned}$$

其中  $\theta, \varphi$  是未知参数,  $E(e_i) = 0, \text{Var}(e_i) = \sigma^2, i = 1, 2, 3$ . 且  $e_1, e_2, e_3$  相互独立.

(1) 求  $\theta, \varphi$  的最小二乘估计  $\hat{\theta}, \hat{\varphi}$ ;

(2) 求  $\text{Cov}\begin{pmatrix} \hat{\theta} \\ \hat{\varphi} \end{pmatrix}$ .

$$3.7 \quad \text{设 } \begin{cases} y_i = \theta + e_i, i = 1, \dots, m, \\ y_{m+i} = \theta + \varphi + e_{m+i}, i = 1, \dots, m, \\ y_{2m+i} = \theta - 2\varphi + e_{2m+i}, i = 1, \dots, n, \end{cases}$$

其中  $\theta, \varphi$  是未知参数, 各  $e_i$  相互独立, 且服从  $N(0, \sigma^2)$ .

(1) 写出设计矩阵  $X$ ;

(2) 求  $\theta, \varphi$  的最小二乘估计  $\hat{\theta}$  和  $\hat{\varphi}$ ;

(3) 证明当  $m = 2n$  时,  $\hat{\theta}$  与  $\hat{\varphi}$  不相关.

3.8 对于正态线性模型

$$y = X\beta + e, e \sim N(0, \sigma^2 I_n),$$

证明  $\beta$  的最小二乘估计与极大似然估计是一致的.

3.9 设  $y_1, \dots, y_n$  是取自  $N(\theta, \sigma^2)$  的独立同分布样本, 求  $\theta$  的最小方差线性无偏估计  $\hat{\theta}$ , 并求  $\text{Var}(\hat{\theta})$ .

3.10 设  $y = X\beta + e, E(e) = 0, \text{Cov}(e) = \sigma^2 I_n, X$  是  $n \times p$  设计矩阵, 其秩为  $p$ . 将  $X, \beta$  分块成

$$X\beta = (X_1 X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

(1) 证明  $\beta_2$  的最小二乘估计  $\hat{\beta}_2$  由下式给出:

$$\hat{\beta}_2 = [X_2' X_2 - X_2' X_1 (X_1' X_1)^{-1} X_1' X_2]^{-1} [X_2' y - X_2' X_1 (X_1' X_1)^{-1} X_1' y];$$

(2) 求  $\text{Cov}(\hat{\beta}_2)$ .

3.11 对于线性回归模型  $y = X\beta + e$ , 假设  $X$  的第 1 列的元全为 1, 证明:

$$(1) \sum_{i=1}^n (y_i - \hat{y}_i) = 0;$$

$$(2) \sum_{i=1}^n \hat{y}_i (y_i - \hat{y}_i) = 0,$$

其中  $\hat{y}_i$  是拟合值向量  $y = X\hat{\beta}$  的第  $i$  个分量.

3.12 设  $y = \beta + e, E(e) = 0, \text{Cov}(e) = \sigma^2 I_n$ . 直接用 Lagrange 乘子法证明: 在约束条件  $A\beta = 0$  下, 使  $\|y - \beta\|^2$  达到极小的  $\beta$  值为  $\hat{\beta} = (I_n - A'(AA')^{-1}A)y$  (其中  $A$  是已知的  $q \times n$  矩阵, 其秩为  $q$ ).

3.13 设  $y_i \sim N(i\theta, i^2\sigma^2), i=1, \dots, n$ , 且相互独立, 求  $\theta$  的最小方差线性无偏估计  $\hat{\theta}$ , 并求  $\text{Var}(\hat{\theta})$ .

3.14 证明推论 3.2.3.

3.15 研究用电高峰时居民家庭每小时的用电量  $Y$  与每月总用电量  $X$  之间的关系. 53 户居民某月用电记录如下表:

用户	X	Y	用户	X	Y
1	679	0.79	28	1748	4.88
2	292	0.44	29	1381	3.48
3	1012	0.56	30	1428	7.58
4	493	0.79	31	1255	2.63
5	582	2.70	32	1777	4.99
6	1156	3.64	33	370	0.59
7	997	4.73	34	2316	8.19
8	2189	9.50	35	1130	4.79
9	1097	5.34	36	463	0.51
10	2078	6.85	37	770	1.74
11	1818	5.84	38	724	4.10
12	1700	5.21	39	808	3.94
13	747	3.25	40	790	0.96
14	2030	4.43	41	783	3.29
15	1643	3.16	42	406	0.44
16	414	0.50	43	1242	3.24
17	354	0.17	44	658	2.14
18	1276	1.88	45	1746	5.71
19	745	0.77	46	468	0.64
20	435	1.39	47	1114	1.90
21	540	0.56	48	413	0.51
22	874	1.56	49	1787	8.33
23	1543	5.28	50	3560	14.94
24	1029	0.64	51	1495	5.11
25	710	4.00	52	2221	3.85
26	1434	0.31	53	1526	3.93
27	837	4.20			

试用计算机完成下面统计分析:

- (1) 应用最小二乘法求经验回归方程;
- (2) 以拟合值  $\hat{y}_i$  为横坐标, 残差  $\hat{e}_i$  为纵坐标, 作残差图, 分析 Gauss - Markov 假设对本例的适用性;
- (3) 考虑因变量的变换  $U = Y^{\frac{1}{2}}$ , 再对新变量  $U$  和  $X$  重复(1)和(2)的统计分析;
- (4) 将 Box - Cox 变换应用到本例, 计算变换参数  $\lambda$  的值, 并做讨论.

3.16 在林业工程中,研究树干的体积  $Y$  与离地面一定高度的树干直径  $X_1$  和树干高度  $X_2$  之间的关系具有重要的实用意义,因为这种关系使我们能够用简单方法从  $X_1$  和  $X_2$  的值去估计一棵树的体积,进而估计一片森林的木材储量.下表是一组观测数据:

$X_1$ (直径)	$X_2$ (高)	$Y$ (体积)	$X_1$ (直径)	$X_2$ (高)	$Y$ (体积)
8.3	70	10.3	12.9	85	33.8
8.6	65	10.3	13.3	86	27.4
8.8	63	10.2	13.7	71	25.7
10.5	72	16.4	13.8	64	24.9
10.7	81	18.8	14.0	78	34.5
10.8	83	19.7	14.2	80	31.7
11.0	66	15.6	14.5	74	36.3
11.0	75	18.2	16.0	72	38.3
11.1	80	22.6	16.3	77	42.6
11.2	75	19.9	17.3	81	55.4
11.3	79	24.2	17.5	82	55.7
11.4	76	21.0	17.9	80	58.3
11.4	76	21.4	18.0	80	51.5
11.7	69	21.3	18.0	80	51.0
12.0	75	19.1	20.6	87	77.0
12.9	74	22.2			

试用计算机完成下面的统计分析:

(1) 先假设  $Y$  与  $X_1$  和  $X_2$  有如下线性关系:  $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$ , 做最小二乘分析, 并做相应的残差图. 试计算 Box-Cox 变换参数  $\lambda$  的值.

(2) 对(1)中计算出的变换参数  $\lambda$  值, 做相应的 Box-Cox 变换, 并对变换后的因变量做对  $X_1$  和  $X_2$  的最小二乘回归. 并做残差图.

3.17 在上题中, 我们把树干的体积  $Y$  看成  $X_1$  与  $X_2$  的线性函数, 这只是 - 种近似. 因为实际上树干可以近似地看成圆柱或圆锥, 于是考虑  $Y$  与  $X_1^2$  和  $X_2$  的线性回归可能更合理一些, 对线性回归关系  $Y = \alpha + \beta_1 X_1^2 + \beta_2 X_2 + e$  重复上题的统计分析.

3.18 对于线性回归模型(3.1.6), 如果真正的协方差阵  $\text{Cov}(e) = \sigma^2 V$ .

(1) 证明此时最小二乘估计  $\hat{\beta} = (X'X)^{-1} X'y$  仍然是  $\beta$  的一个无偏估计.

(2) 证明  $\text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} X'VX (X'X)^{-1}$ .

(3) 记  $\hat{\sigma}^2 = y'(I - X(X'X)^{-1} X')y / (n - p)$  证明

$$E(\hat{\sigma}^2) = \frac{\sigma^2}{n - p} \text{tr}[V(I_n - X(X'X)^{-1} X')].$$

3.19 做了10次试验得观测数据如下:

$y$	16.3	16.8	19.2	18.0	19.5	20.9	21.1	20.9	20.3	22.0
$X_1$	1.1	1.4	1.7	1.7	1.8	1.8	1.9	2.0	2.3	2.4
$X_2$	1.1	1.5	1.8	1.7	1.9	1.8	1.8	2.1	2.4	2.5

(1) 若以  $X_1, X_2$  为回归自变量, 问它们之间是否存在复共线关系?

(2) 试用岭迹法求  $y$  关于  $X_1, X_2$  的岭回归方程, 并画出岭迹图.

3.20 对某种商品的销量  $Y$  进行调查, 并考虑有关的四个因素:  $X_1$ ——居民可支配收入,  $X_2$ ——该商品的平均价格指数,  $X_3$ ——该商品的社会保有量,  $X_4$ ——其它消费品平均价格指数. 下面是调查数据:

序号	$X_1$	$X_2$	$X_3$	$X_4$	$Y$
1	82.9	92.0	17.1	94.0	8.4
2	88.0	93.0	21.3	96.0	9.6
3	99.9	96.0	25.1	97.0	10.4
4	105.3	94.0	29.0	97.0	11.4
5	117.7	100.0	34.0	100.0	12.2
6	131.0	101.0	40.0	101.0	14.2
7	148.2	105.0	44.0	104.0	15.8
8	161.8	112.0	49.0	109.0	17.9
9	174.2	112.0	51.0	111.0	19.6
10	184.7	112.0	53.0	111.0	20.8

利用主成分方法建立  $Y$  与  $X_1, X_2, X_3, X_4$  的回归方程.

## 第四章 假设检验与预测

在上一章,我们讨论了回归参数的几种估计方法,依据这些方法得到回归系数的估计,就可以建立经验回归方程.但是,所建立的经验回归方程是否真正刻画了因变量和自变量之间实际的依赖关系呢?这一方面,也许是最重要的方面,是要把经验方程拿到实践中去考察;另一方面,我们可以做统计假设检验,这叫做回归方程的显著性检验.另外,我们往往还希望研究因变量是否真正依赖一个或几个特定的自变量,这就导致了相应的回归系数显著性检验.在本章的前三节,我们将讨论这些检验问题.在第四节,将研究奇异值的检验.最后一节讨论在给定了回归自变量的情况下,如何预测对应的因变量的值.

### § 4.1 一般线性假设

考虑正态线性回归模型

$$y = X\beta + e, e \sim N(0, \sigma^2 I), \quad (4.1.1)$$

其中  $X$  为  $n \times p$  设计阵,其秩为  $p$ . 本节讨论比较一般的线性假设

$$H: A\beta = b \quad (4.1.2)$$

的检验问题,这里  $A$  为  $m \times p$  矩阵,其秩为  $m$ ;  $b$  为  $m \times 1$  已知向量.在接下来两节的讨论中读者将会看到,实际应用中许多感兴趣的问题都可归结为形如(4.1.2)的假设的检验问题.

我们先提出检验方法的基本思想.对模型(4.1.1)应用最小二乘法,对应的残差平方和为

$$RSS = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'(I - X(X'X)^{-1}X')y, \quad (4.1.3)$$

这里  $\hat{\beta} = (X'X)^{-1}X'y$  (参阅(3.2.6)和定理3.2.3),它反映了实际数据与模型(4.1.1)拟合的程度.  $RSS$  愈小表示数据与模型拟合得愈好.现在在模型(4.1.1)上附加线性假设(4.1.2),再应用最小二乘法,获得约束最小二乘估计

$$\hat{\beta}_H = \hat{\beta} - (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(A\hat{\beta} - b), \quad (4.1.4)$$

(参阅(3.3.7)式)相应的残差平方和

$$RSS_H = (y - X\hat{\beta}_H)'(y - X\hat{\beta}_H). \quad (4.1.5)$$

很明显,加了约束条件(4.1.2),模型参数  $\beta$  的变化范围缩小了.因而残差平方和  $RSS_H$  就要变大,于是总有  $RSS_H \geq RSS$ . 如果真正的参数确实满足约束条件(4.1.2),那末加上约束条件跟不加约束条件本质上是一样的.这时对无约束模型和有约束模型,数据拟合的程度也应该一样.因而刻画拟合程度的残差平方和之差  $RSS_H - RSS$  应该比较小.反过来,若真正的参数不满足(4.1.2),则  $RSS_H - RSS$  倾向于比较大,因此,当  $RSS_H - RSS$  比较大时,我们就拒绝假设(4.1.2),不然就接受它.在统计学上当我们谈到一个量大小时,往往有一个比较标准.对现在的情况,我们把比较的标准取为  $RSS$ . 于是用统计量  $(RSS_H - RSS)/RSS$  的大小来决定是接受假设(4.1.2),还是拒绝(4.1.2).

**定理 4.1.1** 对于正态线性回归模型(4.1.1)

- (a)  $RSS/\sigma^2 \sim \chi_{n-p}^2$ ;
- (b) 若假设(4.1.2)成立,则  $(RSS_H - RSS)/\sigma^2 \sim \chi_m^2$ ;
- (c)  $RSS$  与  $RSS_H - RSS$  相互独立;
- (d) 当假设(4.1.2)成立时

$$F_H = \frac{(RSS_H - RSS)/m}{RSS/(n-p)} \sim F_{m, n-p}, \quad (4.1.6)$$

这里  $F_{m,n}$  表示自由度为  $m, n$  的  $F$  分布.

**证明** (a) 这条结论在定理 3.2.4 已证明过.

(b) 根据(3.3.10),我们有

$$\|y - X\hat{\beta}_H\|^2 = \|y - X\hat{\beta}\|^2 + \|X(\hat{\beta} - \hat{\beta}_H)\|^2,$$

也就是

$$RSS_H = RSS + (\hat{\beta} - \hat{\beta}_H)' X' X (\hat{\beta} - \hat{\beta}_H). \quad (4.1.7)$$

将  $\hat{\beta}_H$  的表达式(4.1.4)代入上式,整理得

$$RSS_H - RSS = (A\hat{\beta} - b)' (A(X'X)^{-1}A')^{-1} (A\hat{\beta} - b). \quad (4.1.8)$$

因为  $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ . 根据定理 2.3.2, 有

$$A\hat{\beta} - b \sim N(A\beta - b, \sigma^2 A(X'X)^{-1}A').$$

当假设(4.1.2)成立时,

$$A\hat{\beta} - b \sim N(0, \sigma^2 A(X'X)^{-1}A').$$

因为  $A$  的秩为  $m$ , 所以  $A(X'X)^{-1}A'$  是可逆阵. 对(4.1.8)式应用定理 2.4.1, 便证明了

$$\frac{RSS_H - RSS}{\sigma^2} \sim \chi_m^2.$$

(c) 因为

$$\begin{aligned} A\hat{\beta} - b &= A(X'X)^{-1}X'(X\beta + e) - b \\ &= A(X'X)^{-1}X'e + (A\beta - b), \end{aligned}$$

代入(4.1.8)得

$$\begin{aligned} RSS_H - RSS &= e'X(X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}A(X'X)^{-1}X'e \\ &\quad + (A\beta - b)'(A(X'X)^{-1}A')^{-1}A(X'X)^{-1}X'e + \Delta \\ &= e'Me + c'e + \Delta, \end{aligned}$$

其中

$$\begin{aligned} M &= X(X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}A(X'X)^{-1}X', \\ c' &= (A\beta - b)'(A(X'X)^{-1}A')^{-1}A(X'X)^{-1}X', \\ \Delta &= (A\beta - b)'(A(X'X)^{-1}A')^{-1}(A\beta - b). \end{aligned}$$

这里  $\Delta$  是不包含误差向量  $e$  的非随机项. 若记  $N = I - X(X'X)^{-1}X'$ , 注意到  $NX = 0$ , 于是(4.1.3)变形为

$$RSS = e'Ne.$$

因此, 为了证明  $RSS_H - RSS$  与  $RSS$  相互独立, 只需证明  $e'Me$  和  $c'e$  都跟  $e'Ne$  独立. 因为  $e \sim N(0, \sigma^2 I)$ , 根据定理 2.4.5 和 2.4.6, 只需证明  $NM = 0$  和  $c'N = 0$ . 而根据  $M, N$  和  $c$  的定义及  $NX = 0$ , 这两条是很容易验证的. 结论得证.

(d) 本结论是前三条的直接推论, 定理证毕.

(4.1.6)式给出了线性假设  $H: A\beta = b$  的检验统计量. 对给定的水平  $\alpha$ , 记  $F_{m, n-p}(\alpha)$  为相应的  $F$  分布的上侧  $\alpha$  分位点, 即

$$P(F_{m, n-p} \leq F_{m, n-p}(\alpha)) = 1 - \alpha,$$

那么当  $F_H > F_{m, n-p}(\alpha)$  时, 我们拒绝线性假设  $H$ , 否则就接受  $H$ . 这个检验的水平为  $\alpha$ .

检验统计量表达式(4.1.6)比较容易在计算机上实现. 我们知道, 计算  $RSS$  的简便公式是

$$RSS = y'y - \hat{\beta}'X'y. \quad (4.1.9)$$

而  $RSS_H$  也可以用同样公式来计算. 因为我们只要把约束条件,  $A\beta = b$  “融入”到原来的模型, 就化成了一个无约束的回归模型, 称其为约简模型. 例如, 把模型(4.1.1)写成分量形式

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{i, p-1}\beta_{p-1} + e_i. \quad (4.1.10)$$

若要检验  $\beta_1 = \beta_2 = \beta_3$ , 这时线性假设具有形式  $A\beta = 0$ , 其中

$$A = \begin{pmatrix} 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 \end{pmatrix}.$$



将  $\beta_1 = \beta_2 = \beta_3$  代入(4.1.10)得到

$$y_i = \beta_0 + (x_{i1} + x_{i2} + x_{i3})\beta_1 + x_{i4}\beta_4 + \cdots + x_{i,p-1}\beta_{p-1} + e_i.$$

这个不带约束的新模型,也就是约简模型,等价于原来的带约束条件的模型.约简模型的参数向量  $\alpha = (\beta_0, \beta_1, \beta_4, \cdots, \beta_{p-1})$ ,而设计阵是将原来的设计阵的第2、3、4列相加而得到的  $n \times (p-2)$ 矩阵  $\tilde{X}$ ,对约简模型应用最小二乘法,得到最小二乘估计

$$\hat{\alpha} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y.$$

其残差平方和为

$$RSS_H = y'y - \hat{\alpha}'\tilde{X}'y, \quad (4.1.11)$$

和(4.1.9)有完全相同的形式.这些计算工作非常适于在计算机上完成.

利用(4.1.8)我们可以得到  $F$  检验统计量的另一个表达式

$$F_H = \frac{(\mathbf{A}\hat{\beta} - \mathbf{b})'(\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{b})/m}{RSS/(n-p)}, \quad (4.1.12)$$

跟(4.1.6)相比,  $F_H$  的这个表达式更多地用在理论讨论或  $m \leq 2$  的情形.

#### 例 4.1.1 假设

$$y_1 = \beta_1 + e_1,$$

$$y_2 = 2\beta_1 - \beta_2 + e_2,$$

$$y_3 = \beta_1 + 2\beta_2 + e_3,$$

其中  $e = (e_1, e_2, e_3)' \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I}_3)$ . 我们欲检验  $H: \beta_1 = \beta_2$ . 将观测数据写成线性回归模型

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

线性假设  $H$  等价于

$$(1 \quad -1) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = 0.$$

于是,对现在的情形  $\mathbf{A} = (1, -1)$ ,  $n = 3$ ,  $p = 2$ ,  $m = 1$ . 容易求出  $\beta = (\beta_1, \beta_2)'$  的最小二乘估计

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{6}(y_1 + 2y_2 + y_3) \\ \frac{1}{5}(-y_2 + y_3) \end{pmatrix}.$$

从(4.1.9)得到残差平方和

$$RSS = y'y - \hat{\beta}'X'y = \sum_{i=1}^3 y_i^2 - 6\hat{\beta}_1^2 - 5\hat{\beta}_2^2.$$

为了求在假设  $H: \beta_1 = \beta_2$  下的残差平方和  $RSS_H$ . 我们将  $\beta_1 = \beta_2$  (记它们的公共值为  $\alpha$ ) 代入原来模型, 得到约简模型

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \alpha + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}.$$

从这个模型容易得到  $\alpha$  的最小二乘估计 (也就是  $\beta_1$  和  $\beta_2$  的约束最小二乘估计)

$$\hat{\alpha} = \hat{\beta}_H = \frac{1}{11}(y_1 + y_2 + 3y_3)$$

于是  $RSS_H = (y - \hat{\beta}_H \mathbf{1}_3)'(y - \hat{\beta}_H \mathbf{1}_3) = (y - \hat{\alpha} \mathbf{1}_3)'(y - \hat{\alpha} \mathbf{1}_3)$ .

由(4.1.6)得到  $H$  的检验统计量  $F_H$ . 当  $H$  成立时, 此  $F_H \sim F_{1,1}$ . 作为练习读者可以利用(4.1.12)导出  $F_H$ .

#### 例 4.1.2 同一模型检验.

假设我们对因变量  $Y$  和自变量  $X_1, \dots, X_{p-1}$  有两批观测数据. 对第一批数据, 有线性回归模型

$$y_i = \beta_0^{(1)} + \beta_1^{(1)} x_{i1} + \dots + \beta_{p-1}^{(1)} x_{i,p-1} + e_i, i = 1, \dots, n_1;$$

而对第二批数据, 也有线性回归模型

$$y_i = \beta_0^{(2)} + \beta_1^{(2)} x_{i1} + \dots + \beta_{p-1}^{(2)} x_{i,p-1} + e_i, i = n_1 + 1, \dots, n_1 + n_2,$$

其中, 所有误差  $e_i$  都相互独立, 且服从  $N(0, \sigma^2)$ . 现在的问题是, 考察这两批数据所反映的因变量  $Y$  与自变量  $X_1, \dots, X_{p-1}$  之间的依赖关系是不是完全一样. 也就是要检验模型中的系数是否完全相等, 即检验  $\beta_i^{(1)} = \beta_i^{(2)}, i = 0, 1, \dots, p-1$ .

这个问题具有广泛的应用背景. 例如, 这两批数据可以是同一公司在两个不同时间段上的数据,  $Y$  是反映公司经济效益的某项指标, 而自变量  $X_1, \dots, X_{p-1}$  是影响公司效益的内在和外在因素. 那么我们所要做的检验就是考察公司效益指标对诸因素的依赖关系在两个时间段上是否有了变化, 也就是所谓经济结构的变化. 又譬如, 在生物学研究中, 有很多试验花费时间比较长, 而为了保证结论的可靠性, 又必须做一定数量的试验. 为此, 很多试验要分配在几个试验室同时进行. 这时, 前面讨论的两批数据就可以看作是来自两个不同试验室的观测数据, 而我们检验的目的是考察两个试验室所得结论有没有差异. 类似的例子还可以举出很多.

为了要导出所需的检验统计量, 我们首先把上面的两个模型写成矩阵形式

$$y_1 = X_1 \beta_1 + e_1, e_1 \sim N(0, \sigma^2 I_{n_1}),$$

$$y_2 = X_2 \beta_2 + e_2, e_2 \sim N(0, \sigma^2 I_{n_2}),$$

其中所含矩阵的意义是不言自明的. 将它们合并, 便得到如下模型:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & \mathbf{0} \\ \mathbf{0} & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}, \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N(\mathbf{0}, \sigma^2 I_{n_1+n_2}). \quad (4.1.13)$$

我们要检验的假设为

$$H: (I_p \quad \vdots \quad -I_p) \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \mathbf{0}. \quad (4.1.14)$$

若记  $\hat{\beta}_1, \hat{\beta}_2$  为从模型(4.1.13)得到的最小二乘估计, 则

$$\begin{aligned} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} &= \begin{bmatrix} X_1' X_1 & \mathbf{0} \\ \mathbf{0} & X_2' X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1' & \mathbf{0} \\ \mathbf{0} & X_2' \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \\ &= \begin{bmatrix} (X_1' X_1)^{-1} & \mathbf{0} \\ \mathbf{0} & (X_2' X_2)^{-1} \end{bmatrix} \begin{bmatrix} X_1' y_1 \\ X_2' y_2 \end{bmatrix} \\ &= \begin{bmatrix} (X_1' X_1)^{-1} X_1' y_1 \\ (X_2' X_2)^{-1} X_2' y_2 \end{bmatrix}. \end{aligned}$$

于是

$$\hat{\beta}_1 = (X_1' X_1)^{-1} X_1' y_1, \quad \hat{\beta}_2 = (X_2' X_2)^{-1} X_2' y_2.$$

应用公式(4.1.9), 得到残差平方和

$$RSS = y_1' y_1 + y_2' y_2 - \hat{\beta}_1' X_1' y_1 - \hat{\beta}_2' X_2' y_2. \quad (4.1.15)$$

为了求约束条件(4.1.14)下的残差平方和  $RSS_H$ , 我们应用前面说过的把约束条件融入模型的方法, 当(4.1.14)成立时  $\beta_1 = \beta_2$ , 记它们的公共值为  $\beta$ , 代入(4.1.12), 得到约简模型

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + e,$$

从这个模型求到的  $\beta$  的无约束最小二乘估计就是原来模型在约束(4.1.14)下的约束最小二乘估计

$$\hat{\beta}_H = (X_1' X_1 + X_2' X_2)^{-1} (X_1' y_1 + X_2' y_2).$$

对应的残差平方和

$$RSS_H = y_1' y_1 + y_2' y_2 - \hat{\beta}_H' (X_1' y_1 + X_2' y_2). \quad (4.1.16)$$

从(4.1.15)和(4.1.16)得到

$$RSS_H - RSS = \hat{\beta}_1' X_1' y_1 + \hat{\beta}_2' X_2' y_2 - \hat{\beta}_H' (X_1' y_1 + X_2' y_2)$$

$$= (\hat{\beta}_1 - \hat{\beta}_H)' X_1' y_1 + (\hat{\beta}_2 - \hat{\beta}_H)' X_2' y_2.$$

至此,我们求到了检验统计量

$$F_H = \frac{(RSS_H - RSS)/p}{RSS/(n_1 + n_2 - 2p)}$$

中的分子与分母的具体表达式. 据此,我们可以对假设  $\beta_1 = \beta_2$  做出检验. 对给定的水平  $\alpha$ , 若  $F > F_{p, n_1+n_2-2p}(\alpha)$ , 则拒绝原假设, 即认为两批数据不服从同一个线性回归模型. 否则, 我们认为它们服从同一个线性回归模型.

## § 4.2 回归方程的显著性检验

我们将正态线性回归模型(4.1.1)写成分量形式

$$y_i = \beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1} + e_i, e_i \sim N(0, \sigma^2), i = 1, \cdots, n. \quad (4.2.1)$$

所谓回归方程的显著性检验, 就是检验假设: 所有回归系数都等于零, 即检验

$$H: \beta_1 = \cdots = \beta_{p-1} = 0. \quad (4.2.2)$$

如果我们检验的结论是拒绝原假设  $H$ , 这意味着我们接受断言: 至少有一个  $\beta_i \neq 0$ . 当然也可能所有  $\beta_i$  都不等于零. 换句话说, 我们认为  $Y$  线性依赖于至少某一个自变量  $X_i$ , 也可能线性依赖于所有自变量  $X_1, \cdots, X_{p-1}$ . 如果检验的结论是接受原假设  $H$ , 这意味着我们接受断言: 所有  $\beta_i = 0$ , 即我们可以认为, 相对于误差而言, 所有自变量对因变量  $Y$  的影响是不重要的.

很明显, 假设(4.2.2)是上节一般线性假设(4.1.2)中  $A = (\mathbf{0}, I_{p-1})$ ,  $b = \mathbf{0}$  的特殊形式. 因此定理 4.1.1 所给出的检验统计量(4.1.6)可以直接应用在这里. 下面我们所要做的只是对现在的特殊情形, 导出检验统计量的简单形式, 并解释其统计意义.

将假设(4.2.2)代入模型(4.2.1), 得到约简模型

$$y_i = \beta_0 + e_i, i = 1, \cdots, n. \quad (4.2.3)$$

很容易求出,  $\beta_0$  的最小二乘估计为  $\beta_0^* = \bar{y}$ , 于是对应的残差平方和

$$RSS_H = y'y - \beta_0^* \mathbf{1}'y = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (4.2.4)$$

这里  $\mathbf{1}$  表示所有分量全为 1 的  $n$  维向量. 在回归分析中, 总是把这个特殊的残差平方和称为总平方和 (Total Sum of Squares, 简记为 TSS). 这是因为约简模型(4.2.3)不包含任何回归自变量, 残差平方和  $RSS_H$  完全是  $n$  个观测数据  $y_1, \cdots, y_n$  的变动平方和.

对于原来模型(4.2.1),我们知道残差平方和

$$RSS = y'y - \hat{\beta}'X'y.$$

于是

$$RSS_H - RSS = \hat{\beta}'X'y - \beta_0^*1'y. \quad (4.2.5)$$

它是由于在模型(4.2.3)中引进回归自变量之后所引起的残差平方和减少量,称为回归平方和,简记为  $SS_{\text{回}}$ ,根据定理 4.1.1,我们可以得到检验假设(4.2.2)的统计量,注意到对现在的情形,  $m = p - 1$ ,于是检验统计量为

$$F_{\text{回}} = \frac{SS_{\text{回}}/(p-1)}{RSS/(n-p)}. \quad (4.2.6)$$

当原假设(4.2.2)成立时,  $F_{\text{回}} \sim F_{p-1, n-p}$ . 对给定的水平  $\alpha$ , 当  $F_{\text{回}} > F_{p-1, n-p}(\alpha)$  时,我们拒绝原假设  $H_0$ , 否则就接受  $H_0$ .

现在我们从方差分析的角度对检验统计量  $F_{\text{回}}$  做一些解释. 根据前面的讨论, (4.2.5)中的  $RSS_H$  就是通常的总平方和  $TSS$ , 于是我们将那里的  $RSS_H$  用  $TSS$  代替. 而右端就是  $SS_{\text{回}}$ , 于是(4.2.5)可写为

$$TSS = RSS + SS_{\text{回}}. \quad (4.2.7)$$

这就是说,我们把总平方和分解成两部分. 一部分是回归平方和  $SS_{\text{回}} = \hat{\beta}'X'y - \beta_0^*1'y$ , 它反映了回归自变量对因变量变动平方和的贡献. 另一部分是残差平方和  $RSS$ , 它是误差的影响, 这里误差包括试验的随机误差和模型误差, 后者是指重要回归自变量的遗漏, 模型的非线性等, 因此, 检验统计量(4.2.6)是把回归平方和与试验误差相比较, 当回归平方和相对试验误差比较大时, 我们就拒绝原假设. 通常我们把每个平方和除以相应的自由度, 称为均方和, 并列成下面的方差分析表.

表 4.2.1 方差分析表

方差源	平方和	自由度	均方	F 比	$P(F > F_{\text{回}})$
回归	$SS_{\text{回}}$	$p-1$	$SS_{\text{回}}/(p-1)$	$F_{\text{回}}$	
误差	$RSS$	$n-p$	$RSS/(n-p)$		
总计	$TSS$	$n-1$			

表 4.2.1 的最后一列  $P(F > F_{\text{回}})$  表示自由度为  $p-1, n-p$  的  $F$  变量取值大于  $F_{\text{回}}$  的概率, 这个值的给出是充分利用现代电子计算机的高速计算能力的结果. 不难看出, 当  $F_{\text{回}} > F_{p-1, n-p}$  时拒绝原假设, 等价于

$$p = P(F_{p-1, n-p} > F_{\text{回}}) < \alpha.$$

于是, 当表中概率  $p < \alpha$  时, 我们拒绝原假设, 有了表 4.2.1 的最后一列, 我们就

省去了查  $F$  分布表的麻烦.

需要强调的是,如果经过检验,结论是接受原假设  $H: \beta_1 = \dots = \beta_{p-1} = 0$ . 这意思是说,和模型的各种误差比较起来,诸自变量对  $Y$  的影响是不重要的. 这里可能有两种情况. 其一是,模型的各种误差太大,因而即使回归自变量对  $Y$  有一定影响,但相比这较大的模型误差,也不算大. 对这种情况,我们就要想办法缩小误差. 这包括从分析问题的专业背景入手,检查是否漏掉了重要自变量,或  $Y$  对某些回归自变量有非线性相依关系等. 其二是,回归自变量对  $Y$  的影响确实很小. 对这种情况,我们就要放弃建立  $Y$  对诸自变量的线性回归.

#### 例 4.2.1 煤净化问题(取自 Myers(1986)).

表 4.2.2 给出了煤净化过程的一组数据. 这里因变量  $Y$  为净化后煤溶液中所含杂质的重量,这是衡量净化效率的指标,  $X_1$  表示输入净化过程的溶液所含的煤与杂质的比,  $X_2$  是溶液的 pH 值,  $X_3$  表示溶液流量. 试验者的目的是通过一组试验数据,建立净化效率  $Y$  与三个因素  $X_1, X_2$  和  $X_3$  的经验关系. 通过控制某些自变量来提高净化效率. 考虑线性回归模型

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e,$$

应用最小二乘法,得到经验回归方程

$$\hat{Y} = 397.087 - 110.750X_1 + 15.583X_2 - 0.058X_3.$$

表 4.2.2 煤净化数据

编号	$x_1$	$x_2$	$x_3$	$y$
1	1.50	6.00	1 315	243
2	1.50	6.00	1 315	261
3	1.50	9.00	1 890	244
4	1.50	9.00	1 890	285
5	2.00	7.50	1 575	202
6	2.00	7.50	1 575	180
7	2.00	7.50	1 575	183
8	2.00	7.50	1 575	207
9	2.50	9.00	1 315	216
10	2.50	9.00	1 315	160
11	2.50	6.00	1 890	104
12	2.50	6.00	1 890	110

为了检验回归方程的显著性,即检验  $\beta_1 = \beta_2 = \beta_3 = 0$ ,我们计算出下面的方差分析表.

表 4.2.3 煤净化问题的方差分析表

方差源	平方和	自由度	均方	$F_{\text{回}}$	$P(F > F_{\text{回}})$
回归	31 156.02	3	10 385.33	23.82	0.000 2
误差	3 486.89	8	435.85		
总计	34 642.92	11			

给定水平  $\alpha = 0.05$ ,  $F_{3,8}(0.05) = 4.07$ . 算出的  $F$  统计量为  $23.82 > F_{3,8}(0.05)$ , 于是我们拒绝原假设(从表中  $P(F > F_{\text{回}})$  的数值为 0.000 2 也可直接得到同样结论), 认为  $Y$  对  $X_1, X_2$  和  $X_3$  有一定依赖关系. 从经验回归方程可以看出,  $X_2$  的回归系数估计值为正数, 于是, 当其它回归自变量取值不变, 而当  $X_2$  (溶液中的 pH 值) 增加时, 因变量  $Y$  (净化后的溶液中所含杂质) 也随之增加. 与此相反,  $X_1$  的回归系数估计值为负数, 则说明若其它回归自变量保持不变, 当  $X_1$  增加时, 因变量  $Y$  却在减少. 这与  $X_1$  的实际意义有些矛盾. 因为从直观上讲, 如果其它回归自变量保持不变. 当输入溶液中杂质含量在增加时, 净化后的溶液中杂质不太可能呈减少趋势. 正是由于出现了回归系数估计值与直观相抵触的现象, 于是我们还不能认为所得到的经验回归方程刻画了变量之间的真正关系. 这时, 我们应该进一步分析出现这种现象的原因, 并采取适当的修正措施. 例如, 我们可以检查原始数据的记录是否有误差, 也可以用 § 3.4 介绍的方法计算每组数据的 Cook 统计量  $D_i$ , 看是否有少数数据具有过大影响(见例 4.4.1). 但是, 从直观上看, 我们认为对现在含有三个回归自变量的问题, 只有 12 组数据, 显然这个数量太少. 因此, 我们首先应该考虑增加试验次数, 获得更多数据. 然后重新计算回归系数的估计.

### § 4.3 回归系数的显著性检验

回归方程的显著性检验是对线性回归方程的一个整体性检验. 如果我们检验的结果是拒绝原假设, 这意味着因变量  $Y$  线性地依赖于自变量  $X_1, \dots, X_{p-1}$  这个回归自变量的整体. 但是, 这并不排除  $Y$  并不依赖于其中某些自变量, 即某些  $\beta_i$  可能等于零. 于是在回归方程显著性检验被拒绝之后, 我们还需要对每个自变量逐一做显著性检验, 即对固定的  $i, 1 \leq i \leq p-1$  做如下检验:

$$H_i: \beta_i = 0. \quad (4.3.1)$$

这个假设作为(4.1.2)的一种特殊情况, 利用定理 4.1.1 可以获得所需的检验.

我们把这种推导方法作为一个习题留给读者. 下面我们给出一种更直接的方法.

对于模型(4.1.1),  $\beta$  的最小二乘估计为  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . 根据定理 3.2.4 知

$$\hat{\beta} \sim N(\beta, \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}).$$

若记  $\mathbf{C}_{p \times p} = (c_{ij}) = (\mathbf{X}'\mathbf{X})^{-1}$ , 则有

$$\hat{\beta}_i \sim N(\beta_i, \sigma^2 c_{ii}). \quad (4.3.2)$$

于是当  $H_i$  成立时

$$\frac{\hat{\beta}_i}{\sigma \sqrt{c_{ii}}} \sim N(0, 1).$$

因为  $RSS/\sigma^2 \sim \chi_{n-p}^2$ , 且与  $\hat{\beta}_i$  相互独立, 根据  $t$  分布的定义, 有

$$t_i = \frac{\hat{\beta}_i}{\sqrt{c_{ii}} \hat{\sigma}} \sim t_{n-p}, \quad (4.3.3)$$

这里  $\hat{\sigma}^2 = RSS/(n-p)$ ,  $t_{n-p}$  表示自由度为  $n-p$  的  $t$  分布. 对给定的水平  $\alpha$ , 当

$$|t_i| > t_{n-p} \left( \frac{\alpha}{2} \right) \quad (4.3.4)$$

时, 我们拒绝原假设  $H_i$ ; 否则, 我们接受原假设  $H_i$ .

从(4.3.2)我们可以看出, 回归系数  $\beta_i$  的最小二乘估计  $\hat{\beta}_i$  的方差  $\text{Var}(\hat{\beta}_i) = \sigma^2 c_{ii}$ . 文献中常把  $\text{Var}(\hat{\beta}_i)^{1/2} = \sigma \sqrt{c_{ii}}$  称为  $\hat{\beta}_i$  的标准误差. 它的一个估计为  $\hat{\sigma} \sqrt{c_{ii}}$ , 因此(4.3.3)所给出的  $t$  检验统计量就是回归系数最小二乘估计  $\hat{\beta}_i$  与其标准误差估计的商. 在许多统计软件中(例如 SAS 软件等), 都给出如下回归系数估计表, 有的称其为方差分析表(见表 4.3.1). 利用表中第 4 列给出的  $t_i$  值, 我们就可以利用(4.3.4)做相应的检验.

表 4.3.1 回归系数的方差分析表

变量或系数	最小二乘估计	标准误差估计	$t_i$	$P(t_{n-p} >  t_i )$
常数项( $\beta_1$ )	$\hat{\beta}_1$	$\hat{\sigma} \sqrt{c_{11}}$	$t_1$	
$\beta_2$	$\hat{\beta}_2$	$\hat{\sigma} \sqrt{c_{22}}$	$t_2$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$\beta_p$	$\hat{\beta}_p$	$\hat{\sigma} \sqrt{c_{pp}}$	$t_p$	

在表 4.3.1 的最后一列,  $P(t_{n-p} > |t_i|)$  是自由度为  $n-p$  的  $t$  分布取值大于  $|t_i|$  的概率, 这些值的给出是充分利用现代计算机高速计算能力的结果. 显然(4.3.4)等价于



$$p_i = P(t_{n-p} > |t_i|) < \frac{\alpha}{2}.$$

于是当  $p_i < \frac{\alpha}{2}$  时,我们就拒绝原假设  $\beta_i = 0$ . 因此,有了表 4.3.1 的最后一列,我们就不需要再查  $t$  分布表.

如果我们经过检验,接受原假设  $\beta_i = 0$  时,我们就认为回归自变量  $X_i$  对因变量  $Y$  无显著的影响,因而可以将其从回归方程中剔除. 将这个回归自变量从回归方程中剔除之后,剩余变量的回归系数的估计也随之发生变化. 将  $Y$  对剩余的回归自变量重新做回归. 然后再检验其余回归系数是否为零. 再剔除经检验认为对  $Y$  无显著影响的变量,这样的过程一直继续下去,直到对所有的自变量,经检验都认为对  $Y$  有显著的影响为止. 对回归系数做显著性检验的过程,事实上也是回归自变量的选择过程. 关于回归自变量的选择,我们将在下一章进一步详细讨论.

#### 例 4.3.1 (续例 4.2.1)

对例 4.2.1 所讨论的问题,三个回归系数对应的  $t_i$  值分别是  $t_1 = -7.50$ ,  $t_2 = -3.17$ ,  $t_3 = -2.27$ . 对这个例子,  $n = 12$ ,  $p = 4$ ,  $n - p = 8$ , 给定  $\alpha = 0.05$ ,  $t_8(0.025) = 2.3060$ . 对  $i = 1, 2$  有  $|t_i| > t_8(0.025)$ . 因此,我们在水平  $\alpha = 0.05$ , 对每一个回归系数的单独检验,都接受  $\beta_i \neq 0$ . 也就是说,我们认为这两个自变量对净化效率都有重要影响.

表 4.3.2 煤净化问题回归系数检验的方差分析表

变 量	回归系数估计值	标准误差	$t_i$	$P(t_8 >  t_i )$
$X_1$	-110.750	14.762	-7.502	0.0001
$X_2$	15.583	4.921	3.167	0.0133
$X_3$	0.058	0.026	2.27	0.0526

## § 4.4 异常点检验

在统计学中,异常点(outlier)是泛指在一组数据中,与它们的主体不是来自同一分布的那些少数点. 几何直观上,异常点的“异常”之处就是它们远离数据组的主体. 这里度量“远离”的指标,当然在不同的地方有所不同. 在回归分析中,我们总是用残差  $\hat{e}_i$  或  $r_i$  来度量第  $i$  组数据的所谓“远离”程度. 对一组数据,如果它的残差  $\hat{e}_i$  或  $r_i$  相对很大,那么我们就称这组数据为异常点. 何谓“很大”,这里当然存在一个检验问题. 这就是本节我们所要讨论的问题.

为了方便讨论,我们先把正态线性回归模型(4.1.1)改写成如下分量形式:

$$y_i = x_i' \boldsymbol{\beta} + e_i, e_i \sim N(0, \sigma^2), i = 1, \dots, n, \quad (4.4.1)$$

这里  $e_i, i = 1, \dots, n$  相互独立. 如果第  $j$  组数据  $(x_j', y_j)$  是一个异常点, 那么它的残差就很大. 发生这种情况的原因是均值  $E y_j$  发生了非随机漂移  $\eta: E y_j = x_j' \boldsymbol{\beta} + \eta$ . 这样我们有一个新的模型

$$\begin{cases} y_i = x_i' \boldsymbol{\beta} + e_i, & i \neq j \\ y_j = x_j' \boldsymbol{\beta} + \eta + e_j, & e_i \sim N(0, \sigma^2), \end{cases} \quad (4.4.2)$$

记  $\mathbf{d}_j = (0, \dots, 0, 1, 0, \dots, 0)'$ , 这是一个  $n$  维向量, 它的第  $j$  个元素为 1, 其余元素皆为 0. 将模型 (4.4.2) 写成矩阵形式

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{d}_j \eta + \mathbf{e}, \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (4.4.3)$$

模型 (4.4.2) 和 (4.4.3) 称为均值漂移线性回归模型. 要判定  $(x_j', y_j)$  不是异常点, 等价于检验假设  $H: \eta = 0$ .

为了导出所要的检验统计量, 我们需要求出模型 (4.4.3) 中参数  $\boldsymbol{\beta}$  和  $\eta$  的最小二乘估计. 记这些估计分别为  $\boldsymbol{\beta}^*$  和  $\eta^*$ . 显然假设  $\eta = 0$  成立时  $\boldsymbol{\beta}$  的最小二乘估计就是  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ .

**定理 4.4.1** 对均值漂移线性回归模型 (4.4.3),  $\boldsymbol{\beta}$  和  $\eta$  的最小二乘估计分别为

$$\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}_{(j)}$$

和

$$\eta^* = \frac{1}{1 - h_{jj}} \hat{e}_j,$$

其中  $\hat{\boldsymbol{\beta}}_{(j)}$  为从非均值漂移线性回归模型 (4.1.1) 剔除第  $j$  组数据后得到的  $\boldsymbol{\beta}$  的最小二乘估计.  $\mathbf{H} = (h_{ij}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ ,  $h_{jj}$  为  $\mathbf{H}$  的第  $j$  个对角元.  $\hat{e}_j$  为从模型 (4.4.1) 导出的第  $j$  个残差.

**证明** 显然  $\mathbf{d}_j' \mathbf{y} = y_j, \mathbf{d}_j' \mathbf{d}_j = 1$ . 记  $\mathbf{X} = (x_1, \dots, x_n)'$ , 则  $\mathbf{X}' \mathbf{d}_j = x_j$ . 于是根据定义

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\beta}^* \\ \eta^* \end{bmatrix} &= \begin{bmatrix} \mathbf{X}' \\ \mathbf{d}_j' \end{bmatrix} (\mathbf{X} \mathbf{d}_j)^{-1} \begin{bmatrix} \mathbf{X}' \\ \mathbf{d}_j' \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \mathbf{X}'\mathbf{X} & x_j \\ x_j' & 1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ y_j \end{bmatrix}. \end{aligned}$$

根据分块矩阵的逆矩阵公式 (见附录 1), 以及  $h_{jj} = x_j' (\mathbf{X}'\mathbf{X})^{-1} x_j$ ,

$$\begin{bmatrix} \boldsymbol{\beta}^* \\ \eta^* \end{bmatrix} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} + \frac{1}{1 - h_{jj}} (\mathbf{X}'\mathbf{X})^{-1} x_j x_j' (\mathbf{X}'\mathbf{X})^{-1} & -\frac{1}{1 - h_{jj}} (\mathbf{X}'\mathbf{X})^{-1} x_j \\ -\frac{1}{1 - h_{jj}} x_j' (\mathbf{X}'\mathbf{X})^{-1} & \frac{1}{1 - h_{jj}} \end{pmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ y_j \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \hat{\beta} + \frac{1}{1-h_{jj}}(X'X)^{-1}x_jx_j'\hat{\beta} - \frac{1}{1-h_{jj}}(X'X)^{-1}x_jy_j \\ -\frac{1}{1-h_{jj}}x_j'\hat{\beta} + \frac{1}{1-h_{jj}}y_j \end{bmatrix} \\
&= \begin{bmatrix} \hat{\beta} - \frac{1}{1-h_{jj}}(X'X)^{-1}x_j\hat{e}_j \\ \frac{1}{1-h_{jj}}\hat{e}_j \end{bmatrix}.
\end{aligned}$$

再根据公式(3.4.13),知命题得证.

这个定理给出了一个很重要的事实:如果因变量的第  $j$  个观测值发生均值漂移,那么在相应的均值漂移回归模型中,回归系数的最小二乘估计恰等于在原来模型中剔除第  $j$  组数据后,所获得的最小二乘估计.

现在我们应用定理 4.1.1,来求检验  $H: \eta = 0$  的检验统计量.注意到,对现在的情形,在约束条件  $\eta = 0$  下,模型(4.4.3)就化为模型(4.4.1),于是

$$\begin{aligned}
RSS_H &= \text{模型(4.1.1)无约束情形的残差平方和} \\
&= y'y - \hat{\beta}'X'y.
\end{aligned}$$

而模型(4.4.3)的无约束残差平方和

$$RSS = y'y - \beta^*X'y - \eta^*d'y. \quad (4.4.4)$$

利用定理 4.4.1 得

$$\begin{aligned}
RSS_H - RSS &= (\beta^* - \hat{\beta})'X'y + \eta^*d'y \\
&= -\frac{1}{1-h_{jj}}\hat{e}_jx_j'\hat{\beta} + \frac{\hat{e}_jy_j}{1-h_{jj}} \\
&= \frac{\hat{e}_j^2}{1-h_{jj}}, \quad (4.4.5)
\end{aligned}$$

这里  $\hat{e}_j = y_j - x_j'\hat{\beta}$  为第  $j$  组数据的残差.

利用  $\beta^*$  和  $\eta^*$  的具体表达式将(4.4.4)作进一步化简:

$$\begin{aligned}
RSS &= y'y - \hat{\beta}'X'y + \frac{\hat{e}_jy_j}{1-h_{jj}} - \frac{\hat{e}_jy_j}{1-h_{jj}} \\
&= (n-p)\hat{\sigma}^2 - \frac{\hat{e}_j^2}{1-h_{jj}},
\end{aligned}$$

其中  $\hat{\sigma}^2 = \frac{\|y - X\hat{\beta}\|^2}{n-p}$ . 根据定理 4.4.1,所求的检验统计量

$$F = \frac{RSS_H - RSS}{RSS/(n-p-1)}$$

$$\begin{aligned}
 &= \frac{\frac{\hat{e}_j^2}{1-h_{jj}}}{\frac{(n-p)\hat{\sigma}^2}{n-p-1}} = \frac{\hat{e}_j^2}{(n-p-1)(1-h_{jj})} \\
 &= \frac{(n-p-1)r_j^2}{n-p-r_j^2},
 \end{aligned}$$

这里

$$r_j = \frac{\hat{e}_j}{\hat{\sigma} \sqrt{1-h_{jj}}},$$

这是学生化残差,于是我们证明了如下事实.

**定理 4.4.2** 对于均值漂移线性回归模型(4.4.3),如果假设  $H: \eta = 0$  成立,则

$$F_j = \frac{(n-p-1)r_j^2}{n-p-r_j^2} \sim F_{1,n-p-1}.$$

据此,我们就得到如下检验:对给定的  $\alpha (0 < \alpha < 1)$ ,若

$$F_j = \frac{(n-p-1)r_j^2}{n-p-r_j^2} > F_{1,n-p-1}(\alpha), \quad (4.4.6)$$

则判定第  $j$  组数据  $(x'_j, y_j)$  为异常点.当然,这个结论可能是错的,也就是说,  $(x'_j, y_j)$  可能不是异常点,而被误判为异常点.但我们犯这种错误的概率只有  $\alpha$ .事先我们可以把它控制得很小.

显然,根据  $t$  分布和  $F$  分布的关系,我们也可以利用  $t$  检验法完成上面的检验.若定义

$$t_j = F_j^{1/2} = \left[ \frac{(n-p-1)r_j^2}{n-p-r_j^2} \right]^{1/2}.$$

则对给定的  $\alpha$ ,当

$$|t_j| > t_{n-p-1}\left(\frac{\alpha}{2}\right)$$

时,我们拒绝假设:  $\eta = 0$ .即判定第  $j$  组数据  $(x'_j, y_j)$  为异常点.

**例 4.4.1** (续例 4.2.1 和例 4.3.1)

对于例 4.2.1 和例 4.3.1 所讨论的煤净化问题,现在我们检验 12 组数据中是否有异常点.表 4.4.1 给出了各组数据的有关量.对现在情形,  $n = 12$ ,  $p = 4$ ,  $n - p - 1 = 7$ ,对给定水平  $\alpha = 0.05$ ,

表 4.4.1 煤净化数据的诊断统计量

编 号	$y_i$	$\hat{y}_i$	$\hat{e}_i$	$h_i$	$t_i$
1	243.00	247.81	-4.808	0.450 1	-0.292 3
2	261.00	247.81	13.192	0.450 1	0.835 9
3	244.00	261.04	-17.040	0.466 0	-1.137 2
4	285.00	261.04	23.960	0.466 0	1.766 5
5	202.00	200.65	1.348	0.083 8	0.063 1
6	180.00	200.65	-20.652	0.083 8	-1.038 5
7	183.00	200.65	-17.752	0.083 8	-0.869 8
8	207.00	200.65	6.348	0.083 8	0.299 0
9	216.00	183.81	32.192	0.450 1	2.869 5
10	160.00	183.81	-23.808	0.450 1	-1.714 1
11	104.00	103.54	0.460	0.460 0	0.028 2
12	110.00	103.54	6.460	0.466 0	0.400 6

$$t_7(0.025) = 2.364 6.$$

从表 4.1.1 最后一列可以看出只有  $t_9 = 2.869 5$  超过这个值. 于是, 我们认为第 9 号数据为异常点.

异常点的检验是一个很复杂的问题. 首先, 我们要确定异常点的个数. 如果只有一个异常点, 那么可以应用定理 4.4.2 来检验. 如果有多个异常点, 我们就不能应用这个定理去逐个检验, 而需要多个点的同时检验. 虽然我们可以毫无困难的把定理 4.4.2 推广到多个异常点的检验情形(参考陈希孺等(1987)), 但是严重的问题往往出现在异常点个数的确定上面. 如果所假设的个数小于实际个数, 那末可能由于未被怀疑的异常点的存在而产生掩盖现象, 使得真正的异常点检验不出来. 如果我们所假设的异常点个数大于实际个数, 则可能把正常点误判为异常点. 因此, 关于多个异常点的检验目前仍然是统计学家深入研究的一个问题.

## § 4.5 因变量的预测

所谓预测, 乃是对给定的回归自变量的值, 预测对应的回归因变量所可能取的值. 这是回归分析最重要的应用之一, 因为在线性回归模型中, 回归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, 由于试验或生产等方面的费用或花费时间长的原因, 我们在有了经验回归模型之后, 希望对一些感兴趣的试

验、生产条件不真正去做试验,就对相应的因变量的取值能够作出估计和分析,因此,预测就常常显得十分必要.

对于线性回归模型

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + e_i, \quad i = 1, \dots, n, \quad (4.5.1)$$

这里不要求误差  $e_1, \dots, e_n$  服从正态分布,只要求满足 Gauss-Markov 假设,  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$ ,  $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i, p-1})'$ ,  $i = 1, \dots, n$ . 假设我们要预测  $\mathbf{x}_0 = (1, x_{01}, \dots, x_{0, p-1})'$  所对应的因变量值  $y_0$ , 并且  $y_0$  可表为

$$y_0 = \mathbf{x}_0' \boldsymbol{\beta} + e_0 \quad (4.5.2)$$

的形式,这里  $e_0$  与  $e_1, \dots, e_n$  不相关. 为预测  $y_0$ , 我们分别去预测均值部分  $\mathbf{x}_0' \boldsymbol{\beta}$  和误差部分  $e_0$ . 若用  $\hat{\boldsymbol{\beta}}$  表示  $\boldsymbol{\beta}$  的最小二乘估计, 那末我们自然用  $\mathbf{x}_0' \hat{\boldsymbol{\beta}}$  去估计  $\mathbf{x}_0' \boldsymbol{\beta}$ . 误差  $e_0$  是均值为零的随机变量, 我们就用零作为预测值. 于是  $y_0$  的预测值为  $\mathbf{x}_0' \hat{\boldsymbol{\beta}}$ , 记之为  $\hat{y}_0$ , 即

$$\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}, \quad (4.5.3)$$

这就是  $y_0$  的点预测.

预测(4.5.3)具有以下性质

1.  $\hat{y}_0$  是  $y_0$  的无偏预测. 这里“无偏”的含义是预测量与被预测量具有相同的均值. 因为在预测问题中, 被预测量是一个随机变量, 因此这里关于“无偏性”的定义不同于前面在参数估计讨论中的无偏性.

这个事实的证明是容易的. 因为  $E\mathbf{x}_0' \hat{\boldsymbol{\beta}} = \mathbf{x}_0' \boldsymbol{\beta}$ , 于是  $E\hat{y}_0 = \mathbf{x}_0' \boldsymbol{\beta} = Ey_0$ .

2. 在  $y_0$  的一切线性无偏预测中,  $\hat{y}_0$  具有最小方差.

这是 Gauss-Markov 定理(即定理 3.2.2)的直接结果. 事实上假设  $\mathbf{a}'\mathbf{y}$  是  $y_0$  的任一线性无偏预测, 则  $E\mathbf{a}'\mathbf{y} = Ey_0 = \mathbf{x}_0' \boldsymbol{\beta}$ , 因此  $\mathbf{a}'\mathbf{y}$  可以看作  $\mathbf{x}_0' \boldsymbol{\beta}$  的一个线性无偏估计. 而预测  $\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$  也可看作是  $\mathbf{x}_0' \boldsymbol{\beta}$  的一个线性无偏估计. 根据 Gauss-Markov 定理, 总有

$$\text{Var}(\mathbf{a}'\mathbf{y}) \geq \text{Var}(\mathbf{x}_0' \hat{\boldsymbol{\beta}}).$$

这就证明了我们的结论.

这里需要特别强调的是, 虽然从形式上讲,  $y_0$  的预测量  $\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$  与参数函数  $\mu_0 = \mathbf{x}_0' \boldsymbol{\beta}$  的最小二乘估计  $\hat{\mu}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}}$  完全相同, 但它们的实际意义是不同的. 如果我们引进预测偏差  $d_1 = \hat{y}_0 - y_0$  和估计偏差  $d_2 = \hat{\mu}_0 - \mu_0$  并计算它们的方差, 就可清楚这一点. 因为  $e_0$  与  $e_1, \dots, e_n$  不相关, 所以  $y_0$  与  $\hat{\boldsymbol{\beta}}$  也不相关. 因而

$$\text{Var}(d_1) = \text{Var}(y_0) + \text{Var}(\hat{y}_0)$$

$$\begin{aligned}
&= \sigma^2 + \text{Var}(\mathbf{x}'_0 \hat{\boldsymbol{\beta}}) \\
&= \sigma^2 + \sigma^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0.
\end{aligned} \tag{4.5.4}$$

另一方面

$$\begin{aligned}
\text{Var}(d_2) &= \text{Var}(\hat{\mu}_0 - \mathbf{x}'_0 \hat{\boldsymbol{\beta}}) \\
&= \text{Var}(\mathbf{x}'_0 \hat{\boldsymbol{\beta}}) \\
&= \sigma^2 \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0.
\end{aligned}$$

因此,总有  $\text{Var}(d_1) > \text{Var}(d_2)$ . 这样的差别来源于被预测量  $y_0$  是随机变量,而被估计量  $\mathbf{x}'_0 \hat{\boldsymbol{\beta}}$  是非随机变量.

在应用上,有时区间预测更为人们所关心.所谓区间预测就是找一个区间,使得被预测量落在这个区间内的概率达到预先给定的值.在我们讨论区间预测时,需要假定模型误差服从正态分布,即假设  $e_1, \dots, e_n$  以及  $e_0$  都有分布  $N(0, \sigma^2)$ . 这时预测偏差具有性质:

$$\hat{y}_0 - y_0 \sim N(0, \sigma^2(1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0)) \tag{4.5.5}$$

和

$$\hat{y}_0 - y_0 \text{ 与 } \hat{\sigma}^2 \text{ 相互独立.} \tag{4.5.6}$$

事实上

$$\hat{y}_0 - y_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} - y_0.$$

因为  $\mathbf{x}'_0 \hat{\boldsymbol{\beta}}$  和  $y_0$  都服从正态分布,且  $\mathbf{x}'_0 \hat{\boldsymbol{\beta}}$  只依赖于  $y_1, \dots, y_n$ , 它们都与  $y_0$  独立,因而  $\mathbf{x}'_0 \hat{\boldsymbol{\beta}}$  与  $y_0$  独立,于是  $\mathbf{x}'_0 \hat{\boldsymbol{\beta}}$  与  $y_0$  之差也服从正态分布. (4.5.5) 的剩余部分可由  $\hat{y}_0$  的预测无偏性和 (4.5.4) 得出.

因为  $\hat{\sigma}^2$  只依赖于  $y_1, \dots, y_n$ , 所以它也与  $y_0$  独立. 为了证明 (4.5.6), 我们只需证明  $\hat{\sigma}^2$  与  $\hat{y}_0$  独立. 但是,注意到  $\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}}$ , 由定理 3.2.4 知  $\hat{\sigma}^2$  与  $\hat{\boldsymbol{\beta}}$  独立, 于是  $\hat{\sigma}^2$  就与  $\hat{y}_0$  独立. (4.5.6) 得证.

由 (4.5.5), 我们有

$$\frac{\hat{y}_0 - y_0}{\sigma \sqrt{1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}} \sim N(0, 1).$$

但是

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-p},$$

于是

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma} \sqrt{1 + \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0}} \sim t_{n-p}.$$

因而对给定的  $\alpha$ , 有

$$P\left\{\frac{|\hat{y}_0 - y_0|}{\hat{\sigma} \sqrt{1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}} \leq t_{n-p}\left(\frac{\alpha}{2}\right)\right\} = 1 - \alpha.$$

由此可得到  $y_0$  的概率为  $1 - \alpha$  的预测区间为

$$\left[\hat{y}_0 - t_{n-p}\left(\frac{\alpha}{2}\right)\hat{\sigma} \sqrt{1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}, \hat{y}_0 + t_{n-p}\left(\frac{\alpha}{2}\right)\hat{\sigma} \sqrt{1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}\right].$$

#### 例 4.5.1 一元线性回归模型.

考虑一元线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n,$$

这里  $e_i \sim N(0, \sigma^2)$  且相互独立. 对  $x_0$  处因变量  $Y$  对应值  $y_0$  的预测为

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0,$$

这里  $\hat{\beta}_0$  和  $\hat{\beta}_1$  为  $\beta_0$  和  $\beta_1$  的最小二乘估计. 而概率为  $1 - \alpha$  的预测区间的端点为

$$\hat{y}_0 \pm t_{n-2}\left(\frac{\alpha}{2}\right)\hat{\sigma} \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]^{1/2},$$

这里  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  是自变量的平均值, 也就是试验中心位置. 我们可以看出, 预测区间的长度在  $x_0 = \bar{x}$  时达到最小. 而当  $x_0$  离  $\bar{x}$  愈远, 预测区间就愈长.

#### 例 4.5.2(续例 4.2.1)

对于煤净化问题, 我们在前面的讨论中, 已经建立了经验回归方程. 现在假设我们要考察  $x_1 = 1.5, x_2 = 7.50, x_3 = 1.315$  这一试验条件下的净化效率, 那么, 取  $\alpha = 0.05$ .

$$\begin{aligned} \hat{y}_0 &= 397.087 - 110.750 \times 1.5 + 15.583 \times 7.50 + 1.315 \times (-0.058) \\ &= 271.564, \end{aligned}$$

$$\hat{\sigma} = (435.862)^{1/2} = 20.88,$$

$$n - p = 12 - 4 = 8,$$

$$t_8(0.025) = 2.3060,$$

于是  $y_0$  的包含概率为 0.95 的预测区间为  $[215.756, 326.609]$ .

### 习题四

4.1 设  $y_1 = \beta_1 + e_1,$

$$y_2 = 2\beta_1 - \beta_2 + e_2,$$

$$y_3 = \beta_1 + 2\beta_2 + e_3,$$



这里,  $e = (e_1, e_2, e_3)' \sim N_3(\mathbf{0}, \sigma^2 \mathbf{I})$ . 利用(4.1.12)导出检验  $H_0: \beta_1 = 2\beta_2$  的统计量.

4.2 设  $y_{1i} \sim N(\beta_{10} + \beta_{11}x_{1i}, \sigma^2)$ ,  $i = 1, \dots, n$ ,

$y_{2i} \sim N(\beta_{20} + \beta_{21}x_{2i}, \sigma^2)$ ,  $i = 1, \dots, m$

且相互独立, 导出检验  $H_0: \beta_{11} = \beta_{21}$  的统计量.

4.3 设  $y_i = \beta_0 + \beta_1 x_i + e_i$ ,  $i = 1, \dots, n$ .

$e_i$  相互独立且服从  $N(0, \sigma^2)$ . 对于假设  $H_0: \beta_0 = 0$  (也就是回归直线经过原点), 导出检验统计量.

4.4 对线性模型  $y = \mathbf{X}\boldsymbol{\beta} + e$ ,  $e \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I})$  回归系数的显著性检验  $H_0: \beta_i = 0$ ,  $i = 1, \dots, p-1$ , 看作(4.1.2)的一种特例, 利用定理 4.1.1 导出检验统计量.

4.5 对习题 3.4,

(1) 做回归方程及回归系数的显著性检验, 并列出的方差分析表, 取显著性水平  $\alpha = 0.05$ ;

(2) 对  $x_0 = 27.5$ , 求  $y_0$  的包含概率为 0.95 的预测区间.

4.6 从空中对地面上的一四边形的四个角  $\theta_1, \theta_2, \theta_3, \theta_4$  进行测量, 得测量值分别为  $Y_1, Y_2, Y_3, Y_4$ .

(1) 将此问题表示成线性模型的形式;

(2) 对于假设  $H_0$ : 四边形是平行四边形导出检验统计量.

4.7 在回归分析计算中, 常对数据进行变换:

$$\tilde{y}_i = \frac{y_i - c_0}{d_0}, \quad i = 1, \dots, n,$$

$$\tilde{x}_{ik} = \frac{x_{ik} - c_k}{d_k}, \quad k = 1, \dots, p-1, i = 1, \dots, n,$$

其中  $c_0, c_k, d_0 \neq 0, d_k \neq 0$  是适当选取的常数.

(1) 试找出由原始数据和变换后的数据得到的回归系数的最小二乘估计、残差平方和、回归平方和、总平方和之间的关系.

(2) 若用  $F$  及  $\tilde{F}$  分别表示变换前后对回归方程做显著性检验所用的  $F$  统计量, 证明  $F = \tilde{F}$ .

4.8 对于一元线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n,$$

$E(e_i) = 0, \text{Var}(e_i) = \sigma^2, e_i$  互不相关. 记  $\hat{\beta}_0$  和  $\hat{\beta}_1$  分别为  $\beta_0$  和  $\beta_1$  的最小二乘估计.

(1) 记  $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$  为因变量  $y$  在  $x_0$  处值  $y_0$  的预测值,  $\hat{y}_0 - y_0$  称为预测偏差, 证明

$$\text{Var}(\hat{y}_0 - y_0) = \sigma^2 \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right];$$

(2) 记  $\mu_0 = \beta_0 + \beta_1 x_0$  为因变量  $y$  在  $x_0$  处取值  $y_0$  的均值, 即  $E y_0 = \mu_0$ . 用  $\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$  估计  $\mu_0$ , 证明

$$\text{Var}(\hat{\mu}_0) = \sigma^2 \left[ \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

4.9 假设回归直线通过原点,即一元线性回归模型为

$$y_i = \beta x_i + e_i, \quad i = 1, \dots, n,$$

$E(e_i) = 0, \text{Var}(e_i) = \sigma^2, e_i$  互不相关.

(1) 写出  $\beta$  和  $\sigma^2$  的最小二乘估计;

(2) 记因变量  $Y$  在  $x_0$  处的值  $y_0$  的预测值为  $\hat{y}_0 = \hat{\beta} x_0$ , 求  $\text{Var}(\hat{y}_0 - y_0)$ ;

(3) 记  $\mu_0 = \beta x_0, \hat{\mu}_0 = \hat{\beta} x_0$ , 求  $\text{Var}(\hat{\mu}_0)$ .

## 第五章 回归方程的选择

我们在第三章和第四章分别讨论了线性回归模型的估计方法和检验问题,但应用回归分析处理实际问题时,首先要解决的问题——回归方程的选择,则还没有完全涉及.对回归模型而言,所谓回归方程的选择包含两方面的内容.一是选择回归方程的类型,即判断是用线性回归模型,还是用非线性回归模型来处理欲解决的实际问题.统计上称之为回归模型的线性性检验.由于这一内容超出了本书讨论的范围,这里不再多叙.二是在模型选定后,自变量的选择问题.当我们根据经验的、专业的、统计的方法,确定因变量以及对其可能有影响的自变量适合一个线性回归模型之后,这时回归方程的选择就成为回归自变量的选择.通常,为全面起见,在做回归分析时,人们把各种与因变量有关或可能有关的自变量引进回归方程,其结果是把一些对因变量影响很小的,有些甚至没有影响的自变量也包含在回归方程中,从而使计算量大为增加,并导致回归模型参数的估计和对因变量预报的精度下降.此外,对有些实际问题,某些自变量的观测数据的获得代价昂贵.如果这些自变量本来就对因变量影响不大,或者根本就没有影响,而我们又不加选择地将其包罗进回归方程,势必造成观测数据收集和模型应用的费用不必要地加大.因此,在应用回归分析解决实际问题时,从与因变量保持线性关系的自变量集合中,选择一个“最优”的自变量子集是非常重要的.本章将围绕回归自变量选择,介绍一些自变量选择的准则,以及相应的求“最优”自变量子集的计算方法.

### § 5.1 评价回归方程的标准

假设根据经验或专业理论,确定一切可能对因变量  $Y$  有影响的自变量共有  $p-1$  个,记为  $X_1, \dots, X_{p-1}$ ,它们与  $Y$  一起适合线性回归模型(3.1.1),即

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + e, \quad (5.1.1)$$

其中  $e$  为误差项,含义如(3.1.1)式.当我们获得了  $n$  组观测数据  $(x_{i1}, \dots, x_{i,p-1}, y_i), i = 1, 2, \dots, n$  之后,便有

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + e_i, \quad i = 1, 2, \dots, n, \quad (5.1.2)$$

其中误差项  $e_i (i = 1, 2, \dots, n)$  满足(3.1.3)式,即

$$Ee_i = 0, \text{Var}(e_i) = \sigma^2, \text{Cov}(e_i, e_j) = 0, i \neq j.$$

如果模型(5.1.1)就是真实模型,这时统计建模的任务就是利用第三章和第四章所介绍的统计方法,通过(5.1.2)式估计未知参数 $\beta_1, \dots, \beta_{p-1}$ 和 $\sigma^2$ ,并对所估计的回归方程作统计分析.根据需要,用所估的回归方程对因变量 $Y$ 进行预测.如果自变量中包含了某些对因变量 $Y$ 虽有影响,但获得观测数据的代价太昂贵,或者从经验看,某些自变量似乎对 $Y$ 有影响,但统计分析(比如用第四章介绍的回归系数假设检验)表明其无显著影响,这时,我们就有必要从全部自变量中精心挑选一些,比如,前 $q-1$ 个自变量 $X_1, \dots, X_{q-1}$ ,成为一个“最优”自变量子集.于是,相应的“最优”回归方程为

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{q-1} X_{q-1} + e. \quad (5.1.3)$$

为叙述方便,我们称(5.1.3)式为选模型,而称(5.1.1)式为全模型.现在,我们将面临三个问题:

- (1) 回归方程(5.1.1)减少自变量后,对回归方程的估计有什么影响?
- (2) 回归方程(5.1.1)减少自变量后,对因变量 $Y$ 的预测有什么影响?
- (3) 如何评价所选择的回归方程是“最优”的?

为简单起见,我们以含两个自变量的回归方程为例来回答上述问题(1)和(2),并假定常数项 $\beta_0 = 0$ .

设全模型(5.1.1)为二元回归模型,即

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i, \quad i = 1, 2, \dots, n, \quad (5.1.4)$$

而选模型为一元回归模型

$$y_i = \beta_1 x_{i1} + e_i, \quad i = 1, 2, \dots, n. \quad (5.1.5)$$

由第三章的讨论知,对选模型(5.1.5), $\beta_1$ 的最小二乘估计为

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_{i1} y_i}{\sum_{i=1}^n x_{i1}^2}. \quad (5.1.6)$$

对全模型(5.1.4), $\beta_1$ 的最小二乘估计为

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n x_{i2}^2 \sum_{i=1}^n x_{i1} y_i - \sum_{i=1}^n x_{i1} x_{i2} \sum_{i=1}^n x_{i2} y_i}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2}. \quad (5.1.7)$$

如果全模型是正确的,则

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^n x_{i1} E y_i}{\sum_{i=1}^n x_{i1}^2}$$

$$\begin{aligned}
&= \frac{\sum_{i=1}^n x_{i1}(\beta_1 x_{i1} + \beta_2 x_{i2})}{\sum_{i=1}^n x_{i1}^2} \\
&= \beta_1 + \beta_2 \frac{\sum_{i=1}^n x_{i1} x_{i2}}{\sum_{i=1}^n x_{i1}^2}, \quad (5.1.8)
\end{aligned}$$

$$\begin{aligned}
\text{Var}(\hat{\beta}_1) &= E(\hat{\beta}_1 - E\hat{\beta}_1)^2 \\
&= E \left[ \frac{\sum_{i=1}^n x_{i1} y_i}{\sum_{i=1}^n x_{i1}^2} - \beta_1 - \beta_2 \frac{\sum_{i=1}^n x_{i1} x_{i2}}{\sum_{i=1}^n x_{i1}^2} \right]^2 \\
&= E \left[ \frac{\sum_{i=1}^n x_{i1} e_i}{\sum_{i=1}^n x_{i1}^2} \right]^2 \\
&= \sigma^2 / \sum_{i=1}^n x_{i1}^2. \quad (5.1.9)
\end{aligned}$$

而由第三章的讨论有

$$E\tilde{\beta}_1 = \beta_1,$$

$$\begin{aligned}
\text{Var}(\tilde{\beta}_1) &= \frac{\sigma^2 \sum_{i=1}^n x_{i2}^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n x_{i1}^2} + \frac{\sigma^2 \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2}{\sum_{i=1}^n x_{i1}^2 \left[ \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2 \right]}. \quad (5.1.10)
\end{aligned}$$

由(5.1.8)~(5.1.10)三式可见,当全模型(5.1.4)正确时选模型(5.1.5)的参数  $\beta_1$  的最小二乘估计  $\hat{\beta}_1$  是有偏的,其偏差为

$$\text{Bias}(\hat{\beta}_1) = \beta_2 \sum_{i=1}^n x_{i1} x_{i2} / \sum_{i=1}^n x_{i1}^2. \quad (5.1.11)$$

但注意到(5.1.10)式右端第二项是非负的,从而

$$\text{Var}(\hat{\beta}_1) \leq \text{Var}(\tilde{\beta}_1),$$

即选模型(5.1.5)的回归系数最小二乘估计的方差减少了.

由于  $\hat{\beta}_1$  是  $\beta_1$  的有偏估计, 这时  $\text{Var}(\hat{\beta}_1)$  不能作为衡量估计精度的标准, 而更合理的衡量标准是用均方误差, 其定义为

$$\text{MSE}(\hat{\beta}_1) = E(\hat{\beta}_1 - \beta_1)^2.$$

由(5.1.6), (5.1.8)和(5.1.9)可得

$$\begin{aligned} \text{MSE}(\hat{\beta}_1) &= E(\hat{\beta}_1 - \beta_1)^2 \\ &= E[(\hat{\beta}_1 - E\hat{\beta}_1) + (E\hat{\beta}_1 - \beta_1)]^2 \\ &= E(\hat{\beta}_1 - E\hat{\beta}_1)^2 + (E\hat{\beta}_1 - \beta_1)^2 \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_{i1}^2} + \beta_2^2 \left[ \frac{(\sum_{i=1}^n x_{i1}x_{i2})^2}{\sum_{i=1}^n x_{i1}^2} \right]. \end{aligned} \quad (5.1.12)$$

注意到  $\text{MSE}(\hat{\beta}_1) = \text{Var}(\hat{\beta}_1)$ , 及在全模型(5.1.4)下,  $\beta_2$  的最小二乘估计为

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}y_i - \sum_{i=1}^n x_{i1}x_{i2} \sum_{i=1}^n x_{i1}y_i}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2} \quad (5.1.13)$$

且

$$\text{Var}(\hat{\beta}_2) = \frac{\sigma^2 \sum_{i=1}^n x_{i1}^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2}, \quad (5.1.14)$$

如果进一步假定

$$\text{Var}(\hat{\beta}_2) \geq \beta_2^2, \quad (5.1.15)$$

比较(5.1.10)和(5.1.12)两式即得

$$\text{MSE}(\hat{\beta}_1) \leq \text{MSE}(\hat{\beta}_1). \quad (5.1.16)$$

我们自然要问条件(5.1.15)合理吗? 下面对此作一简要分析. 由(5.1.14)知, (5.1.15)式等价于

$$\frac{\beta_2^2 \left[ \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2 \right]}{\sum_{i=1}^n x_{i1}^2} \leq \sigma^2.$$

在上式中用  $\bar{\beta}_2$  和  $\hat{\sigma}^2$  (在全模型下误差方差估计)代替  $\beta_2$  和  $\sigma^2$ , 并记

$$c_{22} = \frac{\sum_{i=1}^n x_{i1}^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2},$$

则有

$$\frac{\hat{\beta}_2^2}{\sigma^2 c_{22}} \leq 1, \quad (5.1.17)$$

由于  $c_{22}$  恰好是矩阵

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} \\ \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 \end{bmatrix}^{-1}$$

中对角线上第二个元,由第四章讨论知,(5.1.17)式正是检验假设  $H_0: \beta_2 = 0$  所对应的检验统计量. 不过,在通常的 5% 的显著性水平下,所得的临界值比 (5.1.17) 式给出的 1 大. 因此,条件 (5.1.15) 确实在一定程度上刻画了自变量对因变量影响很小. 上面的讨论表明,即使全模型是正确的,若丢掉的自变量对因变量影响确实很小,这时选模型给定的估计具有较小的均方误差.

现在我们来讨论减少自变量后,对因变量  $Y$  的预测值的影响. 假设我们欲在点  $\mathbf{x} = (x_1, x_2)'$  处预测因变量的值,

$$y = \beta_1 x_1 + \beta_2 x_2 + e.$$

由 § 4.5 的讨论知,在全模型 (5.1.4) 下,  $y$  的预测值为

$$\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2,$$

其预测偏差为

$$\tilde{e} = y - \hat{y} = y - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2.$$

而在选模型 (5.1.5) 下,  $y$  的预测值为

$$\hat{y} = \hat{\beta}_1 x_1,$$

其预测偏差为

$$\hat{e} = y - \hat{\beta}_1 x_1.$$

由 § 4.5 的讨论又知,当全模型 (5.1.4) 正确时,  $E\tilde{e} = 0$ , 即  $\hat{y}$  是无偏预测值,而此时对选模型 (5.1.5), 由 (5.1.8) 式有

$$\begin{aligned} E(\hat{e}) &= [E(\beta_1 - \hat{\beta}_1)]x_1 + \beta_2 x_2 \\ &= \beta_2 \left[ x_2 - \left( \frac{\sum_{i=1}^n x_{i1}x_{i2}}{\sum_{i=1}^n x_{i1}^2} \right) x_1 \right]. \end{aligned} \quad (5.1.18)$$

可见,只要  $\beta_2 \neq 0$ , 则预测  $\hat{y}$  就是有偏的. 类似于 (5.1.16) 式的讨论, 我们应该用均方预测误差来衡量预测的优劣, 其定义为

$$MSEP(\hat{y}) = E(y - \hat{y})^2 = E(\hat{e}^2).$$

注意到  $e$  与  $e_1, \dots, e_n$  不相关,  $E\tilde{e} = 0$ , 由定理 3.2.1 有

$$\begin{aligned}
\text{Var}(\bar{e}) &= E(\bar{e}^2) = E(y - \bar{y})^2 \\
&= E[e + (\beta_1 - \hat{\beta}_1)x_1 + (\beta_2 - \hat{\beta}_2)x_2]^2 \\
&= \sigma^2 \left[ 1 + (x_1, x_2) \begin{pmatrix} \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} \\ \sum_{i=1}^n x_{i1}x_{i2} & \sum_{i=1}^n x_{i2}^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right]. \quad (5.1.19)
\end{aligned}$$

由(5.1.8)和(5.1.12)式,有

$$\begin{aligned}
E(\hat{e}^2) &= E(y - \hat{y})^2 \\
&= E[(\beta_1 - \hat{\beta}_1)x_1 + \beta_2x_2 + e]^2 \\
&= [E(\beta_1 - \hat{\beta}_1)^2]x_1^2 + Ee^2 + \beta_2^2x_2^2 \\
&\quad - 2\beta_2x_1x_2E(\hat{\beta}_1 - \beta_2) \\
&= \sigma^2 \left[ 1 + \frac{x_1^2}{\sum_{i=1}^n x_{i1}^2} \right] + \left[ \frac{\sum_{i=1}^n x_{i1}x_{i2}}{\sum_{i=1}^n x_{i1}^2} x_1\beta_2 - x_2\beta_2 \right]^2,
\end{aligned}$$

从而再由(5.1.8)式得

$$\begin{aligned}
\text{Var}(\hat{e}) &= E(\hat{e} - E\hat{e})^2 = E\hat{e}^2 - (E\hat{e})^2 \\
&= \sigma^2 \left[ 1 + \frac{x_1^2}{\sum_{i=1}^n x_{i1}^2} \right]. \quad (5.1.20)
\end{aligned}$$

由(5.1.19)和(5.1.20)有

$$\begin{aligned}
[\text{Var}(\bar{e}) - \text{Var}(\hat{e})]/\sigma^2 &= \frac{(\sum_{i=1}^n x_{i2}^2)x_1^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2} \\
&\quad - 2 \frac{(\sum_{i=1}^n x_{i1}x_{i2})x_1x_2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2} \\
&\quad + \frac{(\sum_{i=1}^n x_{i1}^2)x_2^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - (\sum_{i=1}^n x_{i1}x_{i2})^2} - \frac{x_1^2}{\sum_{i=1}^n x_{i1}^2}
\end{aligned}$$



$$= \frac{\left[ \left( \sum_{i=1}^n x_{i1} x_{i2} \right) x_1 - \left( \sum_{i=1}^n x_{i1}^2 \right) x_2 \right]^2}{\sum_{i=1}^n x_{i1}^2 \left[ \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2 \right]} \geq 0. \quad (5.1.21)$$

由(5.1.18)~(5.1.21)还可得

$$\begin{aligned} \text{MSEP}(\hat{y}) - \text{MSEP}(\hat{y}) &= \text{Var}(\hat{e}) - \text{Var}(\hat{e}) - (E\hat{e})^2 \\ &= \frac{\sigma^2 \left[ \left( \sum_{i=1}^n x_{i1} x_{i2} \right) x_1 - \left( \sum_{i=1}^n x_{i1}^2 \right) x_2 \right]^2}{\sum_{i=1}^n x_{i1}^2 \left[ \sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2 \right]} \\ &\quad - \beta_2^2 \frac{\left[ \left( \sum_{i=1}^n x_{i1} x_{i2} \right) x_1 - \left( \sum_{i=1}^n x_{i1}^2 \right) x_2 \right]^2}{\left( \sum_{i=1}^n x_{i1}^2 \right)^2} \\ &= \frac{\left[ \left( \sum_{i=1}^n x_{i1} x_{i2} \right) x_1 - \left( \sum_{i=1}^n x_{i1}^2 \right) x_2 \right]^2}{\sum_{i=1}^n x_{i1}^2} \left[ \frac{\sigma^2 \sum_{i=1}^n x_{i1}^2}{\sum_{i=1}^n x_{i1}^2 \sum_{i=1}^n x_{i2}^2 - \left( \sum_{i=1}^n x_{i1} x_{i2} \right)^2} - \beta_2^2 \right], \end{aligned}$$

这样,在条件(5.1.15)下,由(5.1.14)有

$$\text{MSEP}(\hat{y}) - \text{MSEP}(\hat{y}) \geq 0 \quad (5.1.22)$$

这个式子表明,用了选模型,预测的精度(用均方预测误差来度量)反而提高了.

综上所述,我们有如下结论:

(1) 即使全模型(5.1.4)是正确的,减少一些自变量后,总使选模型(5.1.5)的回归系数的最小二乘估计和因变量的预测值的方差减少,但以牺牲估计和预测值的无偏性为代价.

(2) 尽管全模型(5.1.4)是正确的,而且减少的那些自变量对因变量  $Y$  确有影响,譬如  $\beta_2 \neq 0$ ,但是,若  $\beta_2$  难于准确地估计(用条件(5.1.15)来刻画),丢掉这些自变量,可使选模型(5.1.5)的回归系数的最小二乘估计和因变量的预测的精度提高(由(5.1.16)和(5.1.22)来刻画).

由此可见,在回归方程中,丢掉那些对因变量  $Y$  影响不大,或虽有影响,但难于观测的自变量是有利的. 换句话说,我们需要对自变量集合  $\mathcal{A} = \{X_1, \dots, X_{p-1}\}$  的各种不同的自变量子集进行比较分析,从中挑选一个“最优”自变量子集出来,拟合回归模型. 这里的“最优”是相对于某种自变量选择准则而言的. 下面我们回到模型(5.1.2)和(5.1.3),介绍两种从数据与模型拟合优劣或预测精

度出发建立的变量选择准则.

由 § 3.2 的讨论知,残差平方和  $RSS$  的大小反映了实际数据与理论模型的偏离程度,是评价回归方程的一个重要标准.一般说来, $RSS$  愈小,数据与模型拟合得愈好.记选模型(5.1.3)下的残差平方和为  $RSS_q$ ,则由定理 3.2.3 有

$$RSS_q = \mathbf{y}'(\mathbf{I} - \mathbf{X}_q(\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q) \mathbf{y}, \quad (5.1.23)$$

其中

$$\mathbf{y} = (y_1, \dots, y_n)',$$

$$\mathbf{X}_q = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1,q-1} \\ 1 & x_{21} & \cdots & x_{2,q-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{n,q-1} \end{pmatrix}.$$

当在选模型(5.1.3)中再增加一个自变量,记作  $x_q$ ,相应的设计矩阵记为  $\mathbf{X}_{q+1} = (\mathbf{X}_q, \mathbf{x}_q)$ ,其中  $\mathbf{x}_q = (x_{1q}, \dots, x_{nq})'$ ,这时残差平方和为

$$RSS_{q+1} = \mathbf{y}'(\mathbf{I} - \mathbf{X}_{q+1}(\mathbf{X}'_{q+1} \mathbf{X}_{q+1})^{-1} \mathbf{X}'_{q+1}) \mathbf{y}, \quad (5.1.24)$$

$$\text{记 } \mathbf{a} = (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q \mathbf{x}_q, \mathbf{b}^{-1} = \mathbf{x}'_q \mathbf{x}_q - \mathbf{x}'_q \mathbf{X}_q (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q \mathbf{x}_q,$$

$$\mathbf{c} = -\mathbf{a}\mathbf{b}.$$

利用附录 1 关于分块矩阵求逆公式有

$$(\mathbf{X}'_{q+1} \mathbf{X}_{q+1})^{-1} = \begin{pmatrix} \mathbf{X}'_q \mathbf{X}_q & \mathbf{X}'_q \mathbf{x}_q \\ \mathbf{x}'_q \mathbf{X}_q & \mathbf{x}'_q \mathbf{x}_q \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} (\mathbf{X}'_q \mathbf{X}_q)^{-1} + \mathbf{a}\mathbf{b}\mathbf{a}' & \mathbf{c} \\ \mathbf{c}' & \mathbf{b} \end{pmatrix}.$$

于是

$$\mathbf{X}_{q+1}(\mathbf{X}'_{q+1} \mathbf{X}_{q+1})^{-1} \mathbf{X}'_{q+1} = (\mathbf{X}_q, \mathbf{x}_q) \begin{pmatrix} \mathbf{X}'_q \mathbf{X}_q & \mathbf{X}'_q \mathbf{x}_q \\ \mathbf{x}'_q \mathbf{X}_q & \mathbf{x}'_q \mathbf{x}_q \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_q \\ \mathbf{x}'_q \end{pmatrix}$$

$$= \mathbf{X}_q (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q + \mathbf{X}_q \mathbf{a}\mathbf{b}\mathbf{a}' \mathbf{X}'_q + \mathbf{x}_q \mathbf{c} \mathbf{X}'_q$$

$$+ \mathbf{X}_q \mathbf{c} \mathbf{x}'_q + \mathbf{x}_q \mathbf{b} \mathbf{x}'_q,$$

从而

$$\mathbf{X}_{q+1}(\mathbf{X}'_{q+1} \mathbf{X}_{q+1})^{-1} \mathbf{X}'_{q+1} - \mathbf{X}_q (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q$$

$$= \mathbf{b}(\mathbf{X}_q \mathbf{a} - \mathbf{x}_q)(\mathbf{X}_q \mathbf{a} - \mathbf{x}_q)' \geq 0.$$

再由(5.1.23)和(5.1.24)两式得

$$RSS_{q+1} \leq RSS_q,$$

此即当自变量子集扩大时, $RSS_q$  随之减少.如果按 § 3.2 的“ $RSS$  愈小愈好”的原则,选入回归方程的自变量越多越好.这样,“最优”自变量子集应取  $\mathcal{A}$ .可见,残差平方和还不能直接用作选择自变量的准则.

由于  $RSS_q$  是随着  $q$  增大而下降, 为了防止选取过多的自变量, 一种常见的作法是, 对残差平方和  $RSS_q$  乘上一个随  $q$  增加而上升的函数, 作为惩罚因子. 于是, 我们定义

$$RMS_q = \frac{1}{n-q} RSS_q. \quad (5.1.25)$$

由  $RSS_q$  的定义知,  $RMS_q$  是平均残差平方和, 其图形大致为图 5.1.1. 由图 5.1.1 可知, 随着  $q$  的增加,  $RMS_q$  先是减少, 而后稳定下来, 最后又增加. 之所以这样, 是因为随着  $q$  的增加, 尽管惩罚因子  $(n-q)^{-1}$  增大了, 但此时  $RSS_q$  减

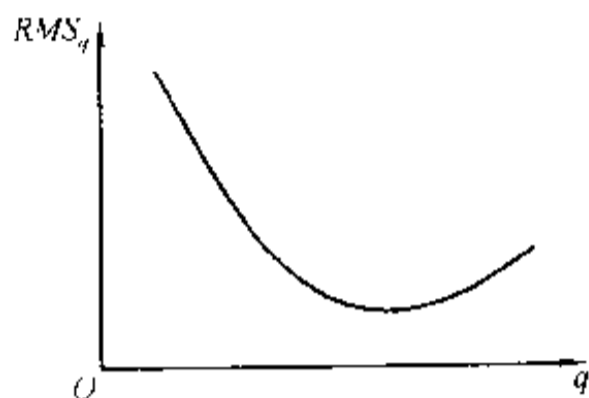


图 5.1.1  $RMS_q$  变化曲线

少很多, 故总的效果为  $RMS_q$  是减少的. 当自变量个数增加到一定程度, 虽然  $RSS_q$  还是减少的, 但尚不足以抵消惩罚因子  $(n-q)^{-1}$  的增加, 最终导致了  $RMS_q$  的增加. 因此, 因子  $(n-q)^{-1}$  确实体现了对自变量个数的增加所施加的惩罚.

依  $RMS_q$  的性质, 我们按“ $RMS_q$  愈小愈好”的原则选取自变量子集, 并称其为“平均残差平方和”准则, 或简称为  $RMS_q$  准则.

$RMS_q$  准则是从数据与模型拟合优劣的角度出发导出的, 由于评价回归方程优劣还可以从对因变量的预测精度来考察, 从而, 也可以从预测精度出发导出另一种类型的准则.

对选模型(5.1.3), 在任一点  $\mathbf{x} = (\mathbf{x}'_q, \mathbf{x}'_r)'$  处, 其中  $\mathbf{x}_q$  表示  $\mathbf{x}$  的前  $q$  个分量, 并且约定第一个分量为 1. 我们用  $\hat{y} = \mathbf{x}'_q \hat{\boldsymbol{\phi}}_q$  预测  $y = \mathbf{x}'\boldsymbol{\beta} + e$  的值, 其中  $\hat{\boldsymbol{\phi}}_q$  为  $\boldsymbol{\phi}_q = (\hat{\beta}_0, \dots, \hat{\beta}_{q-1})'$  的最小二乘估计. 此时

$$MSEP(\hat{y}) = E(\hat{y} - y)^2$$

可作为度量预测精度的指标. 对  $MSEP(\hat{y})$  作分解得

$$MSEP(\hat{y}) = E[\mathbf{x}'_q \hat{\boldsymbol{\phi}}_q - E(\mathbf{x}'_q \hat{\boldsymbol{\phi}}_q) + E(\mathbf{x}'_q \hat{\boldsymbol{\phi}}_q) - y]^2$$

$$= \text{Var}(\mathbf{x}'_q \hat{\boldsymbol{\phi}}_q) + (Ee^\wedge)^2. \quad (5.1.26)$$

由上式可见,预报精度取决于两个方面:一是预报值方差的大小,二是预报值与真值偏离程度.

以“ $MSEP(\hat{y})$ 愈小愈好”为出发点导出的准则,称为  $C_p$  准则,它由 Mallows 于 1964 年提出,其定义为

$$C_p = \frac{RSS_q}{\hat{\sigma}^2} - (n - 2q), \quad (5.1.27)$$

其中  $\hat{\sigma}^2$  为在全模型下误差方差的估计.  $C_p$  准则按“ $C_p$  愈小愈好”的原则选取自变量子集.

除上面介绍的  $RMS_q$  准则和  $C_p$  准则外,还有一些其它的自变量选择准则. 我们介绍其中的 AIC 准则.

极大似然原理是统计学中估计参数的一种重要方法. 日本统计学家 Akaike 把这个方法加以修正,于 1974 年提出了一种较为一般的模型选择准则,称为 Akaike 信息量准则,简称 AIC 准则. 它可以表达为

“使

$$AIC = -2\ln(\text{模型似然度}) + 2(\text{模型自由参数个数}) \quad (5.1.28)$$

达到极小的那组参数是最优的参数选择.”

这是一个很广泛的准则. 它已被用于统计学的许多领域. 下面讨论如何把它用于回归模型自变量的选择.

在选模型(5.1.3)中,假设误差  $\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , 则参数  $\boldsymbol{\varphi}_q = (\beta_0, \dots, \beta_{q-1})'$  和  $\sigma_q^2$  的似然函数为

$$L(\boldsymbol{\varphi}_q, \sigma_q^2 | \mathbf{y}) = (2\pi\sigma_q^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_q^2} \sum_{i=1}^n \left(y_i - \sum_{j=0}^{q-1} \beta_j x_{ij}\right)^2\right\},$$

其中  $x_{i0} \equiv 1, i = 1, 2, \dots, n$ , 而对数似然函数为

$$\ln L(\boldsymbol{\varphi}_q, \sigma_q^2 | \mathbf{y}) = -\frac{n}{2} \ln(2\pi\sigma_q^2) - \frac{1}{2\sigma_q^2} \sum_{i=1}^n \left(y_i - \sum_{j=0}^{q-1} \beta_j x_{ij}\right)^2, \quad (5.1.29)$$

根据极大似然原理,容易求得  $\boldsymbol{\varphi}_q$  和  $\sigma_q^2$  的极大似然估计为

$$\hat{\boldsymbol{\phi}}_q = (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q \mathbf{y},$$

$$\hat{\sigma}_q^2 = \frac{RSS_q}{n} = \frac{1}{n} \mathbf{y}' (\mathbf{I} - \mathbf{X}_q (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q) \mathbf{y}.$$

把它们代入(5.1.29)式,得对数函数的最大值为

$$\ln L(\hat{\boldsymbol{\phi}}_q, \hat{\sigma}_q^2 | \mathbf{y}) = \left[-\frac{n}{2} + \frac{n}{2} \ln\left(\frac{n}{2\pi}\right)\right] - \frac{n}{2} \ln(RSS_q).$$

上式中略去与  $q$  无关的项,按照(5.1.28)式, AIC 统计量为

$$AIC = n \ln(RSS_q) + 2q. \quad (5.1.30)$$

对所有可能的回归自变量的子集,计算相应的  $AIC$  之值.

$AIC$  准则为:选择使(5.1.30)式达到最小的自变量子集为最优回归子集.

关于  $AIC$  准则,还有一些不同形式的推广,其中常用的一种推广为计算统计量

$$BIC = n \ln(RSS_q) + 2q \ln n, \quad (5.1.31)$$

$BIC$  准则可表述为:选择使(5.1.31)式达到最小的自变量子集为最优回归子集.

## § 5.2 计算所有可能的回归

对于一个线性回归模型,设有  $p-1$  个自变量  $x_1, \dots, x_{p-1}$ , 且这  $p-1$  个自变量的任一子集都可以和因变量  $Y$  建立一个线性回归方程. 为了全面了解因变量和自变量之间的结构关系,寻找最优的回归方程,一个自然的想法是,把  $p-1$  个自变量的所有可能的组合,一一与  $Y$  建立线性回归方程,然后按前一节介绍的准则,逐一进行比较,从中挑选出最优的回归方程. 对  $p-1$  个自变量的线性回归问题,所有可能的回归有  $2^{p-1}-1$  个,它们是:

只含一个自变量的回归有  $C_{p-1}^1 = p-1$  个;

只含两个自变量的回归有  $C_{p-1}^2 = \frac{(p-1)(p-2)}{2}$  个;

.....

包含全部  $p-1$  个自变量的回归有  $C_{p-1}^{p-1} = 1$  个.

当  $p$  很大时,比如  $p=11$ ,则需要计算的回归方程的个数有  $2^{10}-1=1023$  个. 可见,计算所有可能的自变量子集的回归,不仅计算量很大,而且这时误差积累也就成为一个不可忽略的问题. 因此,我们必须设计一个非常合理的计算次序和有效的计算方法,使得从一个自变量子集到另一个自变量子集所需要的计算量比较少,并把误差积累控制在一个适当的范围内. 下面介绍的  $S_p$  序列方法给出了一种计算次序.

把  $p$  个变量  $x_1, \dots, x_p$  的子集用一个  $p$  维向量  $(u_1, \dots, u_p)$  来表示,其中  $u_i = 0$ , 或  $1$ . 当  $u_i = 0$ , 则表示变量子集中不含变量  $x_i$ ; 当  $u_i = 1$  时,则表示变量  $x_i$  在子集中. 由于  $p$  维向量  $(u_1, \dots, u_p)$  的每个分量  $u_i$  只有两种可能取值,故这  $p$  维向量所有不同取值的可能情形有  $2^p - 1$  种,即它可以用来表示变量  $x_1, \dots, x_p$  的  $2^p$  个变量子集(其中包括不含任何变量的集合). 例如,当  $p=2$  时,向量  $(u_1, u_2)$  可表示变量  $x_1, x_2$  的四个子集:

$(0,0)$ ——不含  $x_1, x_2$  的子集,即模型仅有常数项;

(1,0)——仅含  $x_1$  的子集;

(0,1)——仅含  $x_2$  的子集;

(1,1)——含  $x_1, x_2$  的子集.

又如,当  $p=3$  时,向量  $(u_1, u_2, u_3)$  可表示变量  $x_1, x_2, x_3$  的八个子集:

(0,0,0)——不含  $x_1, x_2, x_3$  的子集;

(1,0,0)——仅含  $x_1$  的子集;

(0,1,0)——仅含  $x_2$  的子集;

(0,0,1)——仅含  $x_3$  的子集;

(1,1,0)——仅含  $x_1, x_2$  的子集;

(1,0,1)——仅含  $x_1, x_3$  的子集;

(0,1,1)——仅含  $x_2, x_3$  的子集;

(1,1,1)——包含  $x_1, x_2, x_3$  的子集.

显然,我们可以把  $(u_1, u_2)$  的四种可能情形和  $(u_1, u_2, u_3)$  的八种可能情形,分别看作是二维平面上单位正方形的四个顶点和三维空间中单位立方体的八个顶点.这时,设计计算次序,就相当于找一条路线,使它从  $(0,0)$  或  $(0,0,0)$  出发,沿正方形的边,或立方体的边,不重复地通过每个顶点.例如,对  $p=2$ ,这样的路线之一为:

$$(0,0) \longrightarrow (1,0) \longrightarrow (1,1) \longrightarrow (0,1).$$

对  $p=3$ ,这样的路线之一为:

$$(0,0,0) \longrightarrow (1,0,0) \longrightarrow (1,1,0) \longrightarrow (0,1,0) \longrightarrow (0,1,1) \longrightarrow (1,1,1) \longrightarrow (1,0,1) \longrightarrow (0,0,1).$$

按这样的路线计算回归子集,有如下两个特点:

(1) 每个回归子集恰好出现一次;

(2) 相邻的两个回归子集仅有一个变量不同.

依上述计算路线,我们定义  $S_p$  序列如下:以“ $i$ ”表示引入变量  $x_i$ ,“ $-i$ ”表示剔除变量  $x_i$ ,则

$$S_2 = \{1, 2, -1\}$$

表示第一次引入变量  $x_1$ ,计算回归

$$Y = \beta_0 + \beta_1 x_1 + e;$$

第二次再引入  $x_2$ ,计算回归

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e;$$

第三次剔除  $x_1$ ,计算回归

$$Y = \beta_0 + \beta_2 x_2 + e.$$

可见  $S_2$  序列恰对应前述  $p=2$  的计算子集回归的路线.同理,与  $p=3$  的情形相

对应的  $S_3$  序列为

$$S_3 = \{1, 2, -1, 3, 1, -2, -1\}.$$

若记  $T_i$  为  $S_i$  的反序并改变符号所得的序列, 比如  $T_2 = \{1, -2, -1\}$ ,  $T_3 = \{1, 2, -1, -3, 1, -2, -1\}$ , 则我们有

$$S_2 = \{S_1, 2, T_1\},$$

$$S_3 = \{S_2, 3, T_2\},$$

其中  $S_1 = \{1\}$ .

一般地, 对  $p = 1, 2, \dots, S_p$  序列构造方法如下:

变量个数	$S_p$ 序列
1	1
2	1    2    -1
3	$\underbrace{1 \quad 2 \quad -1}_{S_2} \quad 3 \quad \underbrace{1 \quad -2 \quad -1}_{T_2}$
4	$\underbrace{1 \quad 2 \quad -1 \quad 3 \quad 1 \quad -2 \quad -1}_{S_3} \quad 4 \quad \underbrace{1 \quad 2 \quad -1 \quad -3 \quad 1 \quad -2 \quad -1}_{T_3}$
.....	
$p$	$S_{p-1} \quad p \quad T_{p-1}$

当一个回归问题的所有可能自变量被确定后, 我们就可以利用  $S_p$  序列的方法确定所有可能的回归子集. 例如对  $p = 4$ , 即含有四个回归变量  $x_1, x_2, x_3, x_4$ . 由  $S_4$  序列的定义, 所有可能的回归子集及计算回归方程的次序为

$$\begin{aligned}
 &x_1 \quad x_1x_2 \quad x_2 \quad x_2x_3 \quad x_2x_3x_1 \quad x_3x_1 \\
 &x_3 \quad x_3x_4 \quad x_3x_4x_1 \quad x_3x_4x_1x_2 \quad x_3x_4x_2 \\
 &x_4x_2 \quad x_4x_2x_1 \quad x_4x_1 \quad x_4
 \end{aligned}$$

计算所有可能回归的方式还有:

1. 字典式: 按字典编排次序计算所有可能回归. 该计算方式的优点是占用计算机内存小, 当  $p$  较大时, 也可考虑此方式.

2. 自然式: 按自变量下标的自然顺序计算所有可能回归. 该计算方式需要占用的计算机存储量较大.

表 5.2.1 列出了  $p = 4$  时, 上述两种计算方式计算所有可能回归的顺序.

表 5.2.1

序 号	字典式	自然式
1	$x_1$	$x_1$
2	$x_1 x_2$	$x_2$
3	$x_1 x_2 x_3$	$x_3$
4	$x_1 x_2 x_3 x_4$	$x_4$
5	$x_1 x_2 x_4$	$x_1 x_2$
6	$x_1 x_3$	$x_1 x_3$
7	$x_1 x_3 x_4$	$x_1 x_4$
8	$x_1 x_4$	$x_2 x_3$
9	$x_2$	$x_2 x_4$
10	$x_2 x_3$	$x_3 x_4$
11	$x_2 x_3 x_4$	$x_1 x_2 x_3$
12	$x_2 x_4$	$x_1 x_2 x_4$
13	$x_3$	$x_1 x_3 x_4$
14	$x_3 x_4$	$x_2 x_3 x_4$
15	$x_4$	$x_1 x_2 x_3 x_4$

有了计算所有可能回归的方式,还必须有有效的算法.否则,计算所有可能回归是难以实现的.在计算所有可能回归和 § 5.4 的逐步回归时,我们常用如下的矩阵消去变换来实现.

设  $A = (a_{ij})$  是  $n \times n$  矩阵.若  $a_{ii} \neq 0$ , 定义一个新的方阵  $B = (b_{ij})$ , 其中

$$\begin{cases} b_{ii} = 1/a_{ii}, \\ b_{ij} = a_{ij}/a_{ii}, \quad j \neq i, j = 1, \dots, n, \\ b_{ji} = -a_{ji}/a_{ii}, \quad j \neq i, j = 1, \dots, n \\ b_{kl} = a_{kl} - a_{ki}a_{il}/a_{ii}, \quad k \neq i, l \neq i, k, l = 1, \dots, n. \end{cases}$$

称由  $A$  到  $B$  的这种变换为以  $a_{ii}$  为枢轴的消去变换,记为  $B = T_i A$ .

例如,对  $i=1$ ,  $B = T_1 A$  为

$$B = \begin{pmatrix} a_{11}^{-1} & a_{12}/a_{11} & \cdots & a_{1n}/a_{11} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{21}/a_{11} & & & \\ \vdots & & & a_{ij} - a_{i1}a_{1j}/a_{11} \\ -a_{n1}/a_{11} & & & \end{pmatrix}.$$

由定义易证明消去变换具有如下性质:

(1)  $T_i T_i A = A$ , 即对矩阵  $A$  连续施行两次以同一个对角元为枢轴的消去变换,其结果是没有作消去变换.



(2)  $T_i T_j \mathbf{A} = T_j T_i \mathbf{A}$  (可交换性).

(3) 将  $n$  阶方阵  $\mathbf{A}$  分块为

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

其中  $\mathbf{A}_{11}$  为  $q$  阶方阵, 则

$$T_1 T_2 \cdots T_q \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \\ -\mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \end{pmatrix}.$$

现在, 我们把消去变换用于线性回归模型的计算.

考虑线性回归模型

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, E(\mathbf{e}) = \mathbf{0}, \text{Cov}(\mathbf{e}) = \sigma^2 \mathbf{I}, \quad (5.2.1)$$

其中  $\mathbf{y}$  是  $n$  维观察向量,  $\mathbf{X}$  是  $n \times p$  设计矩阵,  $\mathbf{X}'\mathbf{X}$  的秩为  $p$ ,  $\mathbf{e}$  为随机误差向量. 把  $\mathbf{X}$  分块为  $\mathbf{X} = (\mathbf{X}_q, \mathbf{X}_r)$ ,  $\boldsymbol{\beta} = (\boldsymbol{\varphi}'_q, \boldsymbol{\varphi}'_r)'$ , 记

$$\mathbf{B} = \mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}'_q \mathbf{X}_q & \mathbf{X}'_q \mathbf{X}_r \\ \mathbf{X}'_r \mathbf{X}_q & \mathbf{X}'_r \mathbf{X}_r \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}.$$

构造增广矩阵

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{X}'\mathbf{y} \\ \mathbf{y}'\mathbf{X} & \mathbf{y}'\mathbf{y} \end{pmatrix}, \quad (5.2.2)$$

由消去变换性质(3), 对增广矩阵  $\mathbf{A}$  作消去变换, 我们有

$$T_1 T_2 \cdots T_q \mathbf{A} = \begin{pmatrix} \mathbf{B}_{11}^{-1} & * & \mathbf{B}_{11}^{-1} \mathbf{X}'_q \mathbf{y} \\ * & * & * \\ * & * & \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_q \mathbf{B}_{11}^{-1} \mathbf{X}'_q \mathbf{y} \end{pmatrix},$$

其中“\*”表示没有写出的一些子块,  $\mathbf{B}_{11}^{-1} \mathbf{X}'_q \mathbf{y}$  正是选模型

$$\mathbf{y} = \mathbf{X}_q \boldsymbol{\varphi}_q + \mathbf{e} \quad (5.2.3)$$

的参数  $\boldsymbol{\varphi}_q$  的最小二乘估计, 而  $\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_q \mathbf{B}_{11}^{-1} \mathbf{X}'_q \mathbf{y}$  正是该选模型对应的残差平方和  $\text{RSS}_q$ .

类似地, 如果我们欲求选模型

$$Y = \beta_0 + \beta_{i_1} X_{i_1} + \cdots + \beta_{i_q} X_{i_q} + e$$

$\boldsymbol{\varphi}_q = (\beta_{i_1}, \cdots, \beta_{i_q})'$  的最小二乘估计和残差平方和  $\text{RSS}_q$ , 则只要对增广矩阵  $\mathbf{A}$  连续施行消去变换  $T_{i_1}, T_{i_2}, \cdots, T_{i_q}$  即可. 这时, 矩阵

$$\mathbf{A}^* = (a_{ij}^*) = T_{i_1} T_{i_2} \cdots T_{i_q} \mathbf{A}$$

的第  $i_1, i_2, \cdots, i_q$  行和第  $i_1, i_2, \cdots, i_q$  列交叉处的元组成的子阵, 就是矩阵  $\mathbf{A}$  的相应子阵的逆,  $a_{i_1 p+1}^*, \cdots, a_{i_q p+1}^*$  为选模型参数  $\boldsymbol{\varphi}_q$  的最小二乘估计.

进一步, 假定变量  $x_{i_1}, \cdots, x_{i_{q-1}}$  已选入回归模型, 如果下一步欲把变量  $x_{i_q}$ ,

$l \neq i_1, \dots, i_{q-1}$  引入回归方程, 只需对  $A^*$  再进行一次  $T_l$  消去变换. 如果下一步要把变量  $x_{i_l}$  从选模型中剔除, 根据消去变换的性质(1), 只需对矩阵  $A^*$  再进行  $T_{i_l}$  变换, 依此做法, 消去变换的次序可按  $S_p$  序列中的元素(不计符号)进行. 例如, 对  $S_3 = \{1, 2, -1, 3, 1, -2, -1\}$ , 作消去变换的次序依次为  $T_1, T_2, T_1, T_3, T_1, T_2, T_1$ . 每作一次变换, 将增广矩阵最后一列的前  $p$  个元中对应于子集下标的那些元和最后一列的最后一个元存储起来, 前者是相应的选模型的回归系数的最小二乘估计, 后者是相应的选模型的残差平方和. 当按  $S_3$  序列进行完消去变换后, 可按上节介绍的评价回归方程的准则中的某一个, 譬如, 按照  $C_p$  准则, 取使  $C_p$  统计量最小的一个选模型作为“最优”子集回归.

### § 5.3 计算最优子集回归

所谓线性回归方程的选择, 就是从对因变量  $Y$  有影响或可能有影响的自变量集合  $\mathcal{A} = \{x_1, \dots, x_{p-1}\}$  中, 依据某种选择变量的准则, 选择一个“最优”的自变量子集  $\{x_{i_1}, \dots, x_{i_{q-1}}\}$ , 它与  $Y$  一起构成如下线性回归方程

$$Y = \beta_0 + \beta_{i_1} x_{i_1} + \dots + \beta_{i_{q-1}} x_{i_{q-1}} + e.$$

如何从  $\{x_1, \dots, x_{p-1}\}$  中挑选出这个“最优”的自变量子集呢? 由 § 5.2 的讨论知, 我们可以计算  $\{x_1, \dots, x_{p-1}\}$  的所有可能的子集与  $Y$  的线性回归, 选定某种自变量选择准则, 计算相应的统计量的值, 对其逐一进行比较, 按照该准则选择自变量, 确定最优自变量子集. 譬如, 我们用  $RMS_q$  准则, 按 § 5.2 给出的计算所有可能回归的方法, 求出每一个子集回归对应的  $RMS_q$  值, 因  $RMS_q$  准则是按  $RMS_q$  值愈小愈好的原则选择自变量子集, 故比较这些  $RMS_q$  值, 其中  $RMS_q$  值最小者所对应的自变量子集即为所求.

为了便于比较, 我们对回归自变量子集引进秩的概念. 设有  $p-1$  个自变量, 对给定的  $q, 1 \leq q-1 \leq p-1$ , 包含  $q-1$  个自变量的子集共有  $k = C_{p-1}^{q-1}$  个, 设有了某种选择自变量的准则  $U$ , 它归结为对每个包含  $q-1$  个自变量的子集计算统计量  $U_{q-1}$  的值, 按  $U_{q-1}$  值的大小将这  $k$  个自变量子集排序. 如果准则是  $U_{q-1}$  值愈小愈好, 则把其中具有最小  $U_{q-1}$  值的那个自变量子集称为秩为 1, 具有次最小的称为秩为 2. 余类推. 如果准则是  $U_{q-1}$  愈大愈好, 则把其中具有最大  $U_{q-1}$  值的那个自变量子集称为秩为 1, 等等. 可见, 秩反映了在该准则下, 这  $k$  个自变量子集的优劣. 把具有秩为 1 的自变量子集的  $U_{q-1}$  ( $q-1=1, 2, \dots, p-1$ ) 值的图形称为  $U_{q-1}$  图.

下面我们讨论一个例子, 它在回归分析的文献中常被用来讨论自变量的选择和计算方法.

## 例 5.3.1 Hald 水泥问题

这组数据取自 Hald, A. 的著作“Statistical Theory with Engineering Applications(1952)”. 问题是考察含有如下四种化学成分:

$x_1$ —— $3\text{CaO}\cdot\text{Al}_2\text{O}_3$  的含量(%);

$x_2$ —— $3\text{CaO}\cdot\text{SiO}_2$  的含量(%);

$x_3$ —— $4\text{CaO}\cdot\text{Al}_2\text{O}_3\cdot\text{Fe}_2\text{O}_3$  的含量(%);

$x_4$ —— $2\text{CaO}\cdot\text{SiO}_2$  的含量(%)

的某种水泥,每一克所释放出的热量  $y$  与这四种成分含量之间的关系. 数据共  $n = 13$  组,列在表 5.3.1 中.

表 5.3.1 Hald 水泥问题数据

序号	$x_1$	$x_2$	$x_3$	$x_4$	$y$
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

这个回归问题有四个自变量,共有  $C_4^1 + C_4^2 + C_4^3 + C_4^4 - 1 = 15$  个不同的自变量子集. 按  $S_p$  序列方式,用消去变换计算这 15 个不同自变量子集与  $y$  的线性回归,所得的参数最小二乘估计和  $RMS_q$  的值分别列在表 5.3.2 和表 5.3.3. 对  $q=2,3,4,5$ ,图 5.3.1 画出了  $RMS_q$  图.

表 5.3.2 Hald 水泥问题参数最小二乘估计

模型中的自变量	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
$x_1$	81.479	1.869			
$x_1, x_2$	52.577	1.468	0.662		
$x_2$	57.424		0.789		
$x_2, x_3$	72.075		0.731	-1.008	
$x_1, x_2, x_3$	48.194	1.696	0.657	0.250	

续表

模型中的自变量	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
$x_1 x_4$	72.349	2.312		0.494	
$x_3$	110.203			-1.256	
$x_3 x_4$	131.282			-1.200	-0.724
$x_1 x_3 x_4$	111.684	1.052		-0.410	-0.643
$x_1 x_2 x_3 x_4$	62.405	1.551	0.510	0.102	-0.144
$x_2 x_3 x_4$	203.642		-0.923	-1.448	1.557
$x_2 x_4$	94.160		0.311		0.457
$x_1 x_2 x_4$	71.648	1.452	0.416		-0.237
$x_1 x_4$	103.097	1.440			-0.614
$x_4$	117.568				-0.738

表 5.3.3 Hald 水泥问题的  $RMS_q$  值

模型中的自变量	$RMS_q$	模型中的自变量	$RMS_q$
$x_1$	115.026 4	$x_1 x_3 x_4$	8.201 7
$x_1 x_2$	5.790 4 *	$x_1 x_2 x_3 x_4$	5.982 9 *
$x_2$	82.394 2	$x_2 x_3 x_4$	5.648 5
$x_2 x_3$	41.544 3	$x_2 x_4$	86.888 0
$x_1 x_2 x_3$	5.345 6	$x_1 x_2 x_4$	5.330 3 * *
$x_1 x_3$	122.707 3	$x_1 x_4$	7.476 2
$x_3$	176.302 9	$x_4$	80.851 5 *
$x_3 x_4$	17.573 8		

从表 5.3.3 或图 5.3.1 可以看出,  $RMS_q$  在  $\{x_1, x_2, x_4\}$  达到最小值, 为 5.330 3. 因此最优子集回归为

$$y = 71.648 + 1.452x_1 + 0.416x_2 - 0.237x_4 \quad (5.3.1)$$

现在, 我们来讨论用  $C_p$  准则选择自变量子集. 首先我们不加证明地叙述  $C_p$  统计量的几个性质(其证明参见陈希孺等(1987). p. 175).

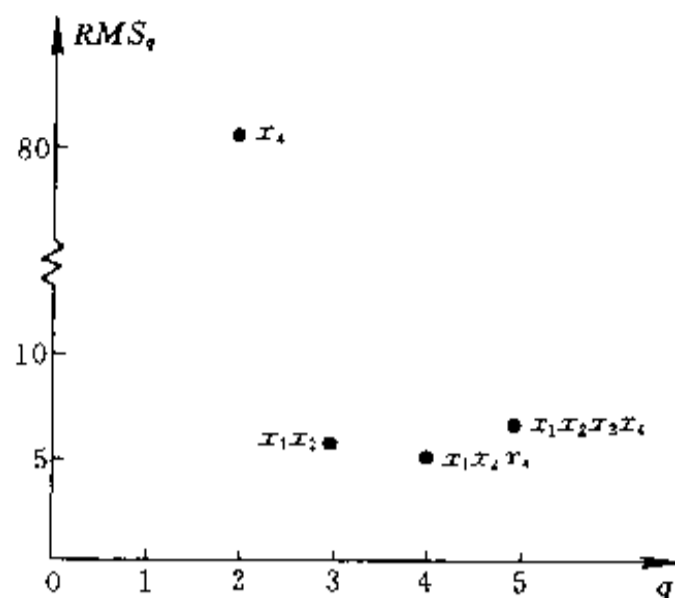
**性质 1** 若全模型(5.2.1)的误差  $e \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , 则对选模型(5.2.3)的  $C_p$ , 有

$$E(C_p) = q - [2r - (n - p)\lambda^2]/(n - p - 2), \quad (5.3.2)$$

其中  $r = p - q$ ,  $\lambda^2 = \boldsymbol{\varphi}'_r \mathbf{B}^{-1} \boldsymbol{\varphi}_r / \sigma^2$ ,  $\mathbf{B}^{-1} = \mathbf{X}'_r \mathbf{X}_r - \mathbf{X}'_r \mathbf{X}_q (\mathbf{X}'_q \mathbf{X}_q)^{-1} \mathbf{X}'_q \mathbf{X}_r$ .

**性质 2** 在性质 1 的条件下, 当  $\boldsymbol{\varphi}_r = \mathbf{0}$  时, 有

$$E(C_p) = q + 2r/(n - p - 2). \quad (5.3.3)$$

图 5.3.1 Hald 水泥问题  $RMS_e$  图

**性质 3** 在性质 1 的条件下, 若  $\boldsymbol{\varphi}_r = \mathbf{0}$ , 则

$$C_p = q + r(F_r - 1), \quad (5.3.4)$$

其中  $F_r \sim F_{r, n-p}$ .

由性质 2 可知, 若选模型 (5.2.3) 是正确的 (即  $\boldsymbol{\varphi}_r = \mathbf{0}$ ), 且  $n-p$  比较大, 则有

$$E(C_p) \approx q.$$

这表明, 对于正确的选模型, 点  $(q, C_p)$  落在平面直角坐标系的第一象限角平分线附近. 如果选模型不正确, 即  $\boldsymbol{\varphi}_r \neq \mathbf{0}$ , 则由性质 1 可知, 当  $n-p$  较大时,

$$E(C_p) \approx q + \lambda^2 > q.$$

这表明, 当选模型不正确时, 点  $(q, C_p)$  将会向第一象限角平分线上方移动. 因此, 点  $(q, C_p)$  在第一象限的位置使我们对“最优”自变量子集有一个直观的判断. 我们把由点  $(q, C_p)$  构成的散点图称为  $C_p$  图. 依  $C_p$  准则, 自变量选择准则是, 选点  $(q, C_p)$  最接近第一象限角平分线且  $C_p$  值最小的自变量子集. 下面通过 Hald 水泥问题说明  $C_p$  图的使用.

### 例 5.3.2 Hald 水泥问题(续)

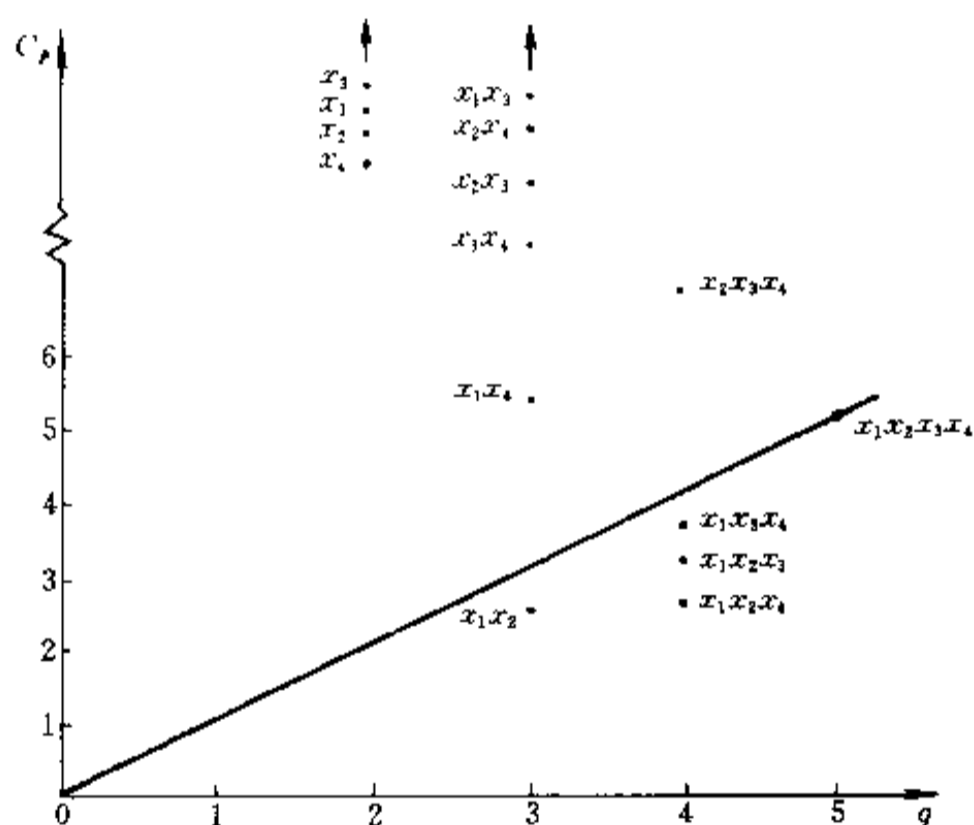
对 Hald 水泥问题, 表 5.3.4 给出了所有可能回归模型的  $C_p$  值, 图 5.3.2 为 Hald 水泥问题的  $C_p$  图.

由表 5.3.4 知,  $C_p$  的最小值对应的变量子集为  $\{x_1, x_2\}$ , 其值为 2.68. 对三个变量的子集, 子集  $\{x_1, x_2, x_3\}$  对应的点  $(4, 3.04)$  最接近第一象限的角平分线, 而  $\{x_1, x_2, x_4\}$  对应的  $C_p$  值 3.02 为最小 (相对于三个变量的子集而言). 如果没有别的附加考虑, 在  $C_p$  准则下,  $\{x_1, x_2\}$  是最优自变量子集. 因此, 最优子集回归为

$$y = 52.577 + 1.468x_1 + 0.662x_2.$$

表 5.3.4 Hald 水泥问题  $C_p$  值

模型中的变量	$C_p$	模型中的变量	$C_p$
$x_1$	202.55	$x_1 x_3 x_4$	3.50
$x_1 x_2$	2.68	$x_1 x_2 x_3 x_4$	5.00
$x_2$	142.49	$x_2 x_3 x_4$	7.34
$x_2 x_3$	62.44	$x_2 x_4$	138.23
$x_1 x_2 x_3$	3.04	$x_1 x_2 x_4$	3.02
$x_1 x_3$	198.10	$x_1 x_4$	5.50
$x_3$	315.16	$x_4$	138.73
$x_3 x_4$	22.37		

图 5.3.2 Hald 水泥问题的  $C_p$  图

由例 5.3.1 和例 5.3.2 的讨论可见,在不同的选择自变量准则下,与之相应的“最优”自变量子集也不尽相同.自然要问:我们应选哪个准则下的“最优”自变量子集呢?由表 5.3.3 和表 5.3.4 可见,子集  $\{x_1, x_2\}$  具有最小的  $C_p$  值 2.68 和次最小的  $RMS_q$  值 5.790 4,而子集  $\{x_1, x_2, x_4\}$  具有次最小的  $C_p$  值 3.02 和最小的  $RMS_q$  值 5.330 3.对这两个子集,从其相应准则的统计量值来看,由  $\{x_1, x_2\}$  增加一个自变量  $x_4$  后,其值下降或上升不多这个事实看,很可能  $x_4$  和  $x_1$  或  $x_2$  有很高的相关性.从表 5.3.5 给出的相关系数矩阵可以发现,  $x_4$  与  $x_2$  有高度的负相关.这就解释了为什么增加  $x_4$  之后,  $C_p$  上升不多和  $RMS_q$  下降不多的原因.总起来看,我们应该选择子集  $\{x_1, x_2\}$ .

表 5.3.5 Hald 水泥数据样本相关阵

	$x_1$	$x_2$	$x_3$	$x_4$	$y$
$x_1$	1.000				
$x_2$	0.229	1.000			
$x_3$	-0.824	-0.139	1.000		
$x_4$	-0.245	-0.973	0.030	1.000	
$y$	0.731	0.816	-0.535	-0.821	1.000

## 例 5.3.3 Gorman-Toman 车辙数据

Gorman-Toman 车辙数据是一个含六个因素(自变量)的回归问题. 因变量为

$$y = \ln(\text{车辙深度改变量} / \text{车轮碾过百万次});$$

自变量为

$$x_1 = \ln(\text{沥青粘度}),$$

$$x_2 = \text{路面沥青的比例},$$

$$x_3 = \text{路底部沥青比例},$$

$$x_4 \text{——示性变量,用以把数据分成两组},$$

$$x_5 \text{——路面空疏比例},$$

$$x_6 \text{——路面好坏}.$$

共有 31 组原始数据,列在表 5.3.6. 对这个问题共有  $2^6 - 1 = 63$  个子集回归. 表 5.3.7 是按自然式次序计算出的所有可能子集回归的自变量及对应的  $RMS_q$  和  $C_p$  统计量的值,这两个准则都选择子集  $\{x_1, x_2, x_4, x_6\}$  为最优子集. 图 5.3.3 和图 5.3.4 分别为  $RMS_q$  图和  $C_p$  图.

表 5.3.6 Gorman-Toman 车辙数据

观测值 序号	$Y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$
1	0.829 30	0.447 16	4.68	4.87	-1.00	8.4	4.916
2	1.113 94	0.146 13	5.19	4.50	-1.00	6.5	4.563
3	1.168 79	0.146 13	4.82	4.73	-1.00	7.9	5.321
4	1.100 37	0.518 51	4.85	4.76	-1.00	8.3	4.865
5	0.916 45	0.230 45	4.86	4.95	-1.00	8.4	3.776
6	1.028 16	0.462 40	5.16	4.45	-1.00	7.4	4.397
7	0.862 13	0.568 20	4.82	5.05	-1.00	6.8	4.867
8	1.102 78	0.230 45	4.86	4.70	-1.00	8.6	4.828

续表

观测值 序号	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
9	1.099 68	-0.036 21	4.78	4.84	-1.00	6.7	4.865
10	1.313 87	0.167 49	5.16	4.76	-1.00	7.7	4.034
11	0.553 88	0.778 15	4.57	4.82	-1.00	7.4	5.450
12	0.845 10	0.633 47	4.61	4.65	-1.00	6.7	4.853
13	1.418 30	-0.221 85	5.07	5.10	-1.00	7.5	4.257
14	1.067 07	0.255 37	4.66	5.09	-1.00	8.2	5.144
15	0.884 79	0.778 15	5.42	4.41	-1.00	5.8	3.718
16	1.088 14	0.643 45	5.01	4.74	-1.00	7.1	4.715
17	-0.119 19	1.944 48	4.97	4.66	1.00	6.5	4.625
18	0.130 33	1.792 39	4.01	4.72	1.00	8.0	4.977
19	0.158 36	1.698 97	4.96	4.90	1.00	6.8	4.322
20	0.204 12	1.763 43	5.20	4.70	1.00	8.2	5.087
21	0.041 39	1.954 24	4.80	4.60	1.00	6.6	5.971
22	0.070 58	1.819 54	4.98	4.69	1.00	6.4	4.647
23	0.079 18	2.146 13	5.35	4.76	1.00	7.3	5.115
24	0.251 81	2.380 21	5.04	4.80	1.00	7.8	5.939
25	0.142 67	2.623 25	4.80	4.80	1.00	7.4	5.916
26	0.327 90	2.698 97	4.83	4.60	1.00	6.7	5.471
27	0.481 49	2.255 27	4.66	4.72	1.00	7.2	4.602
28	0.585 03	2.431 36	4.67	4.50	1.00	6.3	5.043
29	0.119 19	2.230 45	4.72	4.70	1.00	6.8	5.075
30	0.096 91	1.991 23	5.00	5.07	1.00	7.2	4.334
31	0.301 02	1.544 07	4.70	4.80	1.00	7.7	5.705

表 5.3.7 Gorman - Toman 数据的所有可能回归及  $RMS_q$  和  $C_p$  值

变量子集	$RMS_q$	$C_p$	变量子集	$RMS_q$	$C_p$
$x_1$	0.055 1 *	20.43 *	$x_1 x_3$	0.056 4	21.90
$x_2$	0.980 8	817.00	$x_1 x_4$	0.054 8	20.53
$x_3$	0.968 8	806.70	$x_1 x_5$	0.056 2	21.68
$x_4$	0.138 3	92.00	$x_1 x_6$	0.054 2	20.00
$x_5$	0.901 5	748.80	$x_2 x_3$	0.956 8	770.00
$x_6$	0.835 5	692.00	$x_2 x_4$	0.114 6	70.23
$x_1 x_2$	0.046 9 *	13.98 *	$x_2 x_5$	0.892 0	716.10



续表

变量子集	$RMS_q$	$C_p$	变量子集	$RMS_q$	$C_p$
$x_2x_6$	0.865 0	693.70	$x_3x_5x_6$	0.749 3	577.30
$x_3x_4$	0.136 7	88.60	$x_4x_5x_6$	0.131 9	82.70
$x_3x_5$	0.930 7	748.30	$x_1x_2x_3x_4$	0.044 7	13.51
$x_3x_6$	0.832 6	666.8	$x_1x_2x_3x_5$	0.048 2	16.17
$x_4x_5$	0.131 4	84.20	$x_1x_2x_3x_6$	0.037 0	7.56
$x_4x_6$	0.141 8	92.80	$x_1x_2x_4x_5$	0.056 8	11.32
$x_5x_6$	0.727 7	575.50	$x_1x_2x_4x_6$	0.032 7 * *	4.26 * *
$x_1x_2x_3$	0.048 6	15.95	$x_1x_2x_5x_6$	0.036 9	7.45
$x_1x_2x_4$	0.044 0	12.22	$x_1x_3x_4x_5$	0.056 8	22.86
$x_1x_2x_5$	0.046 5	14.29	$x_1x_3x_4x_6$	0.056 2	22.36
$x_1x_2x_6$	0.035 6 *	5.67 *	$x_1x_3x_5x_6$	0.056 8	22.84
$x_1x_3x_4$	0.056 7	22.42	$x_1x_4x_5x_6$	0.056 0	22.24
$x_1x_3x_5$	0.056 8	22.53	$x_2x_3x_4x_5$	0.097 1	53.92
$x_1x_3x_6$	0.055 4	21.37	$x_2x_3x_4x_6$	0.100 9	56.83
$x_1x_4x_5$	0.055 5	21.46	$x_2x_3x_5x_6$	0.777 8	579.10
$x_1x_4x_6$	0.054 4	20.61	$x_2x_4x_5x_6$	0.103 8	59.09
$x_1x_5x_6$	0.056 0	21.87	$x_3x_4x_5x_6$	0.136 0	83.90
$x_2x_3x_4$	0.104 1	58.38	$x_1x_2x_3x_4x_5$	0.063 9 *	13.24
$x_2x_3x_5$	0.913 6	709.00	$x_1x_2x_3x_4x_6$	1.137 7	5.31 *
$x_2x_3x_6$	0.862 5	668.00	$x_1x_2x_3x_5x_6$	1.123 8	9.43
$x_2x_4x_5$	0.100 6	57.56	$x_1x_2x_4x_5x_6$	0.160 4	5.51
$x_2x_4x_6$	0.116 3	70.21	$x_1x_3x_4x_5x_6$	1.045 8	23.65
$x_2x_5x_6$	0.749 4	557.40	$x_2x_3x_4x_5x_6$	0.969 2	54.38
$x_3x_4x_5$	0.134 9	85.10	$x_1x_2x_3x_4x_5x_6$	0.033 7 *	7.00
$x_3x_4x_6$	0.140 2	89.36			

需要注意的是,这个最优自变量子集中含示性变量  $x_4$ ,它只取 1 或 -1,如果我们建立回归方程的目的在于因变量与自变量的结构间的关系的话,最优子集回归方程为

$$y = 0.548 0 - 0.026 5x_1 + 0.047 7x_2 - 0.386 9x_4 - 0.029 7x_6$$

如果建立回归方程的目的在于预测的话, $x_4$  无法取值.对此情形的处理,一是考察原始数据的获得过程,搞清把数据分成两组的原因,以决定如何对待  $x_4$ ; 是在不含自变量  $x_4$  的子集中,按自变量选择准则确定最优子集.由表 5.3.7 可知,

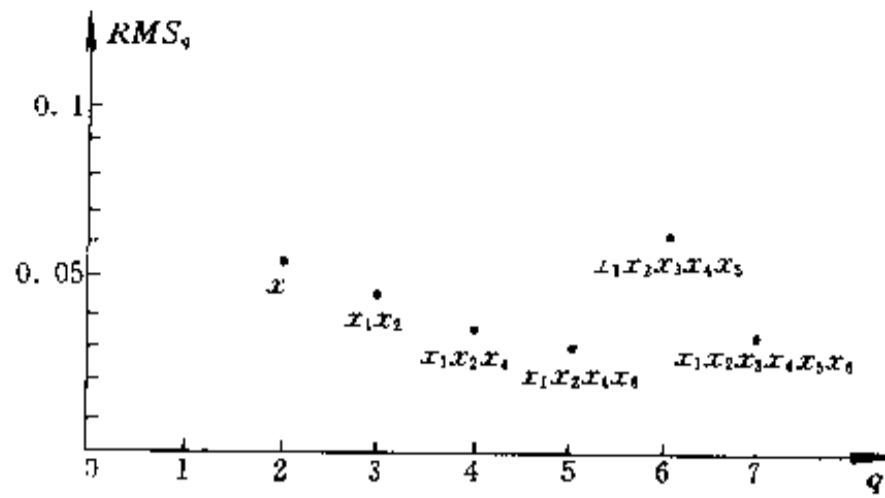


图 5.3.3 Gorman - Toman 数据  $RMS_q$  图

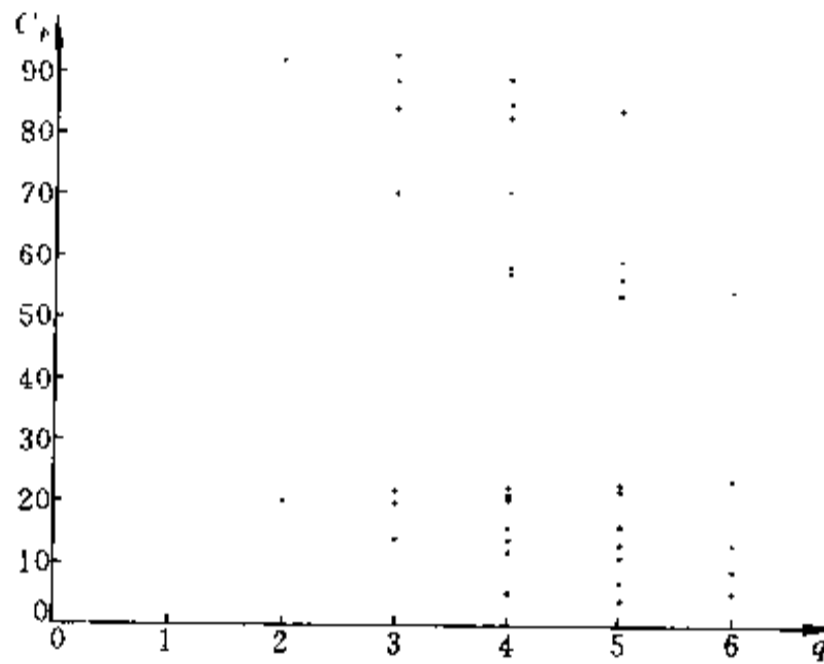


图 5.3.4 Gorman - Toman 数据  $C_p$  图

在不含  $x_4$  的自变量子集中,  $RMS_q$  和  $C_p$  两个准则都选择  $\{x_1, x_2, x_6\}$  为最优子集. 于是, 可用于预测的最优子集回归为

$$y = 1.539 - 0.3627x_1 + 0.0174x_2 - 0.1221x_6.$$

**例 5.3.4** Hald 水泥问题(续)

按(5.1.30)式计算 Hald 水泥问题各变量子集对应的  $AIC$  值, 结果列于表 5.3.8.

**表 5.3.8 Hald 水泥问题  $AIC$  值**

自变量子集	$x_1$	$x_2$	$x_3$	$x_4$	$x_1x_2$	$x_1x_3$	$x_1x_4$
$AIC$ 值	95.995 0	91.653 5	102.939 4	91.407 8	58.003 3	97.700 1	61.325 1
自变量子集	$x_2x_3$	$x_2x_4$	$x_3x_4$	$x_1x_2x_3$	$x_1x_2x_4$	$x_2x_3x_4$	$x_1x_3x_4$
$AIC$ 值	83.620 5	93.212 7	72.436 0	57.725 2	57.687 9	58.441 7	63.290 0
自变量子集	$x_1x_2x_3x_4$						
$AIC$ 值	59.819 7						

由表中结果可见,最小的 AIC 准则值为 57.687 9,对应的变量子集为  $\{x_1, x_2, x_4\}$ . 由前面的讨论知,在  $RMS_q$  准则下,  $\{x_1, x_2, x_4\}$  也是最优回归变量子集. AIC 准则与它的结论一致. 而在  $C_p$  准则下,这也是接近最优的:

### 例 5.3.5 Gorman-Toman 车辙数据问题(续)

按照 AIC 统计量公式(5.1.30)式,计算由 Gorman-Toman 车辙数据问题的六个自变量产生的所有可能子集对应的 AIC 值,结果列于表 5.3.9.

表 5.3.9 Gorman-Toman 车辙数据的 AIC 值

变量子集	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_1 x_2$
AIC 值	17.580 3	106.836 1	106.454 5	46.108 9*	104.222 6	101.865 6	13.534 3*
变量子集	$x_1 x_3$	$x_1 x_4$	$x_1 x_5$	$x_1 x_6$	$x_2 x_3$	$x_2 x_4$	$x_2 x_5$
AIC 值	19.252 3	18.360 2	19.142 2	18.018 9	107.017 2	41.230 6	104.843 2
变量子集	$x_2 x_6$	$x_3 x_4$	$x_3 x_5$	$x_3 x_6$	$x_4 x_5$	$x_4 x_6$	$x_5 x_6$
AIC 值	103.890 4	46.697 2	106.159 8	102.706 9	45.471 4	47.832 7	98.532 3
变量子集	$x_1 x_2 x_3$	$x_1 x_2 x_4$	$x_1 x_2 x_5$	$x_1 x_2 x_6$	$x_1 x_3 x_4$	$x_1 x_5 x_6$	$x_1 x_3 x_6$
AIC 值	15.550 3	12.467 8	14.180 9	5.900 6*	20.328 9	20.583 6	19.609 9
变量子集	$x_1 x_4 x_5$	$x_1 x_4 x_6$	$x_1 x_5 x_6$	$x_2 x_3 x_4$	$x_2 x_3 x_5$	$x_2 x_3 x_6$	$x_2 x_4 x_5$
AIC 值	19.665 8	19.045 2	19.943 8	39.163 8	106.497 1	104.712 8	38.103 6
变量子集	$x_2 x_4 x_6$	$x_2 x_5 x_6$	$x_3 x_4 x_5$	$x_3 x_4 x_6$	$x_3 x_5 x_6$	$x_4 x_5 x_6$	$x_1 x_2 x_3 x_4$
AIC 值	42.599 3	100.355 4	47.198 5	48.393 1	100.351 2	46.501	13.829 7
变量子集	$x_1 x_2 x_3 x_5$	$x_1 x_2 x_3 x_6$	$x_1 x_2 x_4 x_5$	$x_1 x_2 x_4 x_6$	$x_1 x_2 x_5 x_6$	$x_1 x_3 x_4 x_5$	$x_1 x_3 x_4 x_6$
AIC 值	16.166 7	7.969 0	21.256 2	4.139 2*	7.885 1	21.256 2	20.926 9
变量子集	$x_1 x_3 x_5 x_6$	$x_1 x_4 x_5 x_6$	$x_2 x_3 x_4 x_5$	$x_2 x_3 x_4 x_6$	$x_2 x_3 x_5 x_6$	$x_2 x_4 x_5 x_6$	$x_3 x_4 x_5 x_6$
AIC 值	21.256 2	20.816 4	37.878 5	39.068 6	102.381 1	39.947 0	48.322 8
变量子集	$x_1 x_2 x_3 x_4 x_5$	$x_1 x_2 x_3 x_4 x_6$	$x_1 x_2 x_3 x_5 x_6$	$x_1 x_2 x_4 x_5 x_6$	$x_1 x_3 x_4 x_5 x_6$		
AIC 值	25.737 5	115.000 3	114.619 2	54.268 4*	112.389 2		
变量子集	$x_2 x_3 x_4 x_5 x_6$	$x_1 x_2 x_3 x_4 x_5 x_6$					
AIC 值	110.031 2	6.687 2					

由表中所列结果可见, AIC 统计量的最小值为 4.139 2, 对应的回归子集为  $\{x_1, x_2, x_4, x_6\}$ . 在不含示性变量  $x_4$  的子集中, AIC 值的最小值对应的变量子集为  $\{x_1, x_2, x_6\}$ . 这些结论与例 5.3.3 是一致的.

## § 5.4 逐步回归

在 § 5.3 中, 我们讨论了利用选择自变量的准则, 通过计算所有可能子集回.

归,寻找最优子集回归的方法.当自变量个数很多(比如超过40个)时,对这样的大型回归问题,用§5.3的方法,其计算量是很大的.目前,有另外一些不计算所有可能子集回归的变量选择算法,其中所谓逐步回归算法是应用最普遍的.

逐步回归法的基本思想是,将变量一个一个引入,引入变量的条件是其偏回归平方和经检验是显著的.同时,每引入一个新变量,对已入选方程的老变量逐个进行检验,将经检验认为不显著的变量剔除,以保证所得自变量子集中的每个变量都是显著的.此过程经若干步直到不能再引入新变量为止.这时,回归方程中所有自变量对因变量 $y$ 都是显著的,而不在回归方程中的变量对 $y$ 都是经检验不显著的.由此可见,逐步回归法选择变量的过程包含两个基本步骤,一是从回归方程中剔除经检验不显著的变量,二是引入新变量到回归方程中.据此,我们可以得出剔除和引入变量的一般方法.

假设在某一步已入选回归方程的自变量为 $x_{i_1}, \dots, x_{i_q}, q \leq p-1$ .现在我们考察这 $q$ 个自变量是否有要剔除的.为叙述方便,不妨假设这 $q$ 个入选的自变量就是前 $q$ 个,即 $x_1, \dots, x_q$ ,且考察 $x_q$ 是否被剔除.这时,所要讨论的回归模型为

$$y = \mathbf{X}_{q+1} \boldsymbol{\varphi}_{q+1} + e, \quad (5.4.1)$$

其中 $\mathbf{X}_{q+1} = (\mathbf{X}_q, x_q)$ ,  $\boldsymbol{\varphi}_{q+1} = (\boldsymbol{\varphi}_q', \beta_q)'$ ,  $\mathbf{x}_q = (x_{1q}, \dots, x_{nq})'$ ,  $\mathbf{X}_q$ 和 $\boldsymbol{\varphi}_q$ 的定义如§5.1.显然,要确定 $x_q$ 是否被剔除,从假设检验角度出发,相当于检验如下假设

$$H_{\text{剔}}: \beta_q = 0 \quad (5.4.2)$$

是否被接受.由§4.3的讨论知,用于检验假设(5.4.2)的检验统计量 $t_q$ ,当(5.4.2)成立时, $t_q \sim t_{n-q-1}$ (见(4.3.3)),这等价于 $t_q^2 \sim F_{1, n-q-1}$ .将 $t_q^2$ 改记为 $F_{\text{剔}}(q)$ ,它有如下等价形式

$$F_{\text{剔}}(q) = (n - q - 1) \frac{\hat{\beta}_q^2(\mathbf{x}_q' \mathbf{N}_q \mathbf{x}_q)}{\text{RSS}_{q+1}}, \quad (5.4.3)$$

其中 $\hat{\beta}_q$ 和 $\text{RSS}_{q+1}$ 为模型(5.4.1)的回归系数 $\beta_q$ 的最小二乘估计和残差平方和,

$$\mathbf{N}_q = \mathbf{I} - \mathbf{X}_q' (\mathbf{X}_q' \mathbf{X}_q)^{-1} \mathbf{X}_q. \quad (5.4.4)$$

(注:在(5.4.3)的证明中要利用此附录中的结果(A.4.3).)如果我们记 $\text{RSS}_q$ 表示相应于回归模型

$$y = \mathbf{X}_q \boldsymbol{\varphi}_q + e \quad (5.4.5)$$

的残差平方和,则变量 $x_q$ 的偏回归平方和定义为

$$\text{RSS}_{q-1} - \text{RSS}_q = \hat{\beta}_q^2(\mathbf{X}_q' \mathbf{N}_q \mathbf{X}_q).$$

对给定的显著性水平  $\alpha$ , 相应的临界值为  $F_{1, n-q-1}(\alpha)$ , 若

$$F_{\text{删}}(q) \leq F_{1, n-q-1}(\alpha), \quad (5.4.6)$$

则从回归方程(5.4.1)中剔除变量  $x_q$ .

在剔除回归自变量的过程中, 我们通常要计算回归方程(5.4.1)中所有变量  $x_j (j=1, 2, \dots, q)$  相对应的  $F_{\text{删}}(j)$ , 其定义

$$F_{\text{删}}(j) = (N - q - 1) \frac{\hat{\beta}_j^2(\mathbf{x}_j' \mathbf{N}_j \mathbf{x}_j)}{\text{RSS}_{q+1}}, \quad j = 1, 2, \dots, q, \quad (5.4.7)$$

其中  $\hat{\beta}_j$  和  $\text{RSS}_{q+1}$  是回归模型(5.4.1)中回归系数  $\beta_j$  的最小二乘估计和残差平方和,  $\mathbf{x}_j = (x_{1j}, \dots, x_{nj})'$

$$\mathbf{N}_j = \mathbf{I} - \mathbf{X}_{(j)} (\mathbf{X}_{(j)}' \mathbf{X}_{(j)})^{-1} \mathbf{X}_{(j)}'$$

$$\mathbf{X}_{(j)} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1j-1} & x_{1j+1} & \cdots & x_{1q} \\ 1 & x_{21} & \cdots & x_{2j-1} & x_{2j+1} & \cdots & x_{2q} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nj-1} & x_{nj+1} & \cdots & x_{nq} \end{pmatrix}.$$

比较这些  $F_{\text{删}}(j), j=1, \dots, q$ , 不妨设其中最小者为  $F_{\text{删}}(q)$ , 即

$$F_{\text{删}}(q) = \min\{F_{\text{删}}(1), \dots, F_{\text{删}}(q)\},$$

对给定的显著性水平  $\alpha$ , 如果

$$F_{\text{删}}(q) > F_{1, n-q-1}(\alpha), \quad (5.4.8)$$

则表明  $x_1, \dots, x_q$  中没有需要剔除的变量. 如果(5.4.6)式成立, 则表明变量  $x_q$  对  $y$  的作用不显著, 可以剔除. 对剩下的变量  $x_1, \dots, x_{q-1}$  重复上述过程, 直到没有变量可剔除, 然后转入考察能否引入新变量.

下面讨论如何引入新变量. 为方便计, 假设已入选  $q-1$  个变量  $x_1, \dots, x_{q-1}$ , 不在回归方程的变量记为  $x_q, \dots, x_{p-1}$ , 此时,  $x_1, \dots, x_{q-1}$  与  $Y$  的回归为

$$y = \mathbf{X}_q \boldsymbol{\varphi}_q + e. \quad (5.4.9)$$

现在考虑把  $x_q, \dots, x_{p-1}$  逐个引入回归方程. 这相当于把变量  $x_q, \dots, x_{p-1}$  的观测数据  $\mathbf{x}_i = (x_{1i}, \dots, x_{mi})', i=q, \dots, p-1$ , 逐个添加到回归模型(5.4.9)中. 比如考虑引入变量  $x_q$ , (5.4.9)式变为

$$y = (\mathbf{X}_q, \mathbf{x}_q) \begin{pmatrix} \boldsymbol{\varphi}_q \\ \beta_q \end{pmatrix} + e. \quad (5.4.10)$$

显然, 变量  $x_q$  能否引入回归方程, 相当于检验假设

$$H_{\text{引}}: \beta_q = 0 \quad (5.4.11)$$

是否被拒绝. 和检验(5.4.2)完全一样, 对假设(5.4.11)进行检验的统计量为

$$F_{\text{引}}(q) = (n - q - 1) \frac{\hat{\beta}_q^2(\mathbf{x}_q' \mathbf{N}_q \mathbf{x}_q)}{\text{RSS}_{q+1}},$$

公式中的各量的定义如(5.4.3).

类似于剔除变量的过程,我们要计算所有待引入的变量  $x_i (i = q, \dots, p-1)$  相对应的  $F_{\beta_i}(i)$ , 其定义为

$$F_{\beta_i}(i) = (n - q - 1) \frac{\hat{\beta}_i^2(\mathbf{x}_i' \mathbf{N}_q \mathbf{x}_i)}{RSS_{q+1}}, i = q, \dots, p-1, \quad (5.4.12)$$

其中  $\hat{\beta}_i$  和  $RSS_{q+1}$  为回归模型

$$y = (\mathbf{X}_q, \mathbf{x}_i) \begin{bmatrix} \phi_q \\ \beta_i \end{bmatrix} + e, i = q, \dots, p-1$$

的回归系数  $\beta_i$  的最小二乘估计和残差平方和. 比较这些  $F_{\beta_i}(i) (i = q, \dots, p-1)$ , 不妨设其中最大者为  $F_{\beta_i}(q)$ , 即

$$F_{\beta_i}(q) = \max\{F_{\beta_i}(q), \dots, F_{\beta_i}(p-1)\}.$$

对给定的显著性水平  $\alpha$ , 如果

$$F_{\beta_i}(q) < F_{1, n-q-1}(\alpha), \quad (5.4.13)$$

则  $x_q, \dots, x_{p-1}$  都不能引入回归方程, 选择变量过程结束. 如果

$$F_{\beta_i}(q) \geq F_{1, n-q-1}(\alpha), \quad (5.4.14)$$

则将变量  $x_q$  引入回归方程, 然后用  $(\mathbf{X}_q, x_q)$  代替(5.4.10)中的  $\mathbf{X}_q$ , 再逐个考察剩余变量  $x_{q+1}, \dots, x_{p-1}$ , 直到没有变量可选入回归方程为止.

基于上述讨论, 我们通过一个例子把逐步回归的计算步骤作一总结.

**例 5.4.1** 继续讨论例5.3.1 中Hald 水泥问题, 其逐步回归计算步骤如下:

步骤 1 对 Hald 水泥问题的数据进行中心化处理, 得到

$$y = 95.42, \bar{x}_1 = 7.46, \bar{x}_2 = 48.15, \bar{x}_3 = 11.77, \bar{x}_4 = 30.00.$$

$$\mathbf{A}^{(0)} = \begin{pmatrix} 415.23 & 251.08 & -372.62 & -290 & 775.96 \\ 251.08 & 2905.69 & -166.54 & -3041 & 2292.96 \\ -372.62 & -166.54 & 492.31 & 38 & -618.23 \\ -290 & -3041 & 38 & 3362 & -2481.70 \\ 775.96 & 2292.96 & -618.23 & -2481.70 & 2715.76 \end{pmatrix}.$$

这里的  $\mathbf{A}^{(0)}$  即为 § 5.2 的增广矩阵(5.2.2). 检验的显著水平  $\alpha = 0.05$  或  $0.10$ , 本例取  $\alpha = 0.10$ .

步骤 2 引入第一个自变量. 这时, 我们考察一元回归方程

$$y = \beta_0 + \beta_i x_i + e, i = 1, 2, 3, 4.$$

由 § 5.2 的讨论知, 若记

$$\mathbf{A}^{(1)} = T_r \mathbf{A}^{(0)} \stackrel{d}{=} (a_{ij}^{(1)}),$$

则有

$$\hat{\beta}_i = a_{i5}^{(1)} = a_{i5}^{(0)} / a_{ii}^{(0)}, i = 1, 2, 3, 4,$$

$$RSS_1(i) = a_{55}^{(1)}(i) = a_{55}^{(0)} - [a_{i5}^{(1)}]^2 / a_{ii}^{(0)}, i = 1, 2, 3, 4.$$

变量  $x_i$  的偏回归平方和  $p_i^{(1)}$  为

$$p_i^{(1)} = [a_{i5}^{(0)}]^2 / a_{ii}^{(0)}, i = 1, 2, 3, 4.$$

经计算得

$$p_1^{(1)} = 1450.07, p_2^{(1)} = 1809.42,$$

$$p_3^{(1)} = 776.36, p_4^{(1)} = 1831.90,$$

且

$$p_4^{(1)} = \max\{p_1^{(1)}, p_2^{(1)}, p_3^{(1)}, p_4^{(1)}\}.$$

对变量  $x_4$ , 其相应的检验统计量的值为

$$F_{\frac{1}{9}}^{(1)}(4) = \frac{11 \cdot p_4^{(1)}}{RSS_1} = 22.80$$

由于  $F_{1,11}(0.1) = 3.23$ , 故  $F_{\frac{1}{9}}^{(1)}(4) > F_{1,11}(0.1)$ . 引入变量  $x_4$ , 并对  $A^{(0)}$  施行 (4,4) 为枢轴的消去变换, 得

$$A^{(1)} = \begin{pmatrix} 390.22 & -11.23 & -369.34 & 0.09 & 561.89 \\ -11.23 & 155.04 & -132.17 & 0.91 & 48.20 \\ -369.34 & -132.17 & 491.88 & -0.01 & -590.18 \\ -0.09 & -0.91 & 0.01 & 0.0003 & -0.74 \\ 561.89 & 48.20 & -590.18 & 0.74 & 883.87 \end{pmatrix}.$$

步骤3 由于回归方程只引入变量  $x_4$ , 此步不考虑剔除变量的情况. 计算不在回归方程中的变量  $x_1, x_2, x_3$  的偏回归平方和

$$p_i^{(2)} = [a_{i5}^{(1)}]^2 / a_{ii}^{(1)}, i = 1, 2, 3$$

得  $p_1^{(2)} = 809.08, p_2^{(2)} = 14.98, p_3^{(2)} = 708.12.$

故

$$p_1^{(2)} = \max\{p_1^{(2)}, p_2^{(2)}, p_3^{(2)}\}.$$

检验  $x_1$  能否被引入回归方程的统计量的值为

$$F_{\frac{1}{9}}^{(2)}(1) = \frac{10 \cdot p_1^{(2)}}{RSS_2} = 108.18,$$

而  $F_{1,10}(0.1) = 3.28$ , 从而  $F_{\frac{1}{9}}^{(2)}(1) > F_{1,10}(0.1)$ . 引入变量  $x_1$ , 对  $A^{(1)}$  施行 (1,1) 为枢轴的消去变换得

$$A^{(2)} = \begin{pmatrix} 2.56 \times 10^{-3} & -0.03 & -0.95 & 2.2 \times 10^{-3} & 1.44 \\ 0.03 & 154.72 & -142.79 & 0.91 & 63.38 \\ 0.95 & -142.79 & 142.30 & 0.70 & -58.35 \\ 2.21 \times 10^{-3} & -0.91 & -0.70 & 3.17 \times 10^{-4} & -0.61 \\ -1.44 & 64.38 & 58.35 & 0.61 & 74.76 \end{pmatrix}.$$

步骤4 由习题1知:此步不可能剔除变量,故考虑能否从变量  $x_2, x_3$  中引进新变量,计算它们的偏回归平方和

$$p_i^{(3)} = [a_{i5}^{(2)}]^2 / a_{ii}^{(2)}, i = 2, 3$$

得  $p_2^{(3)} = 26.79, p_3^{(3)} = 23.93.$

故  $p_2^{(3)} = \max\{p_2^{(3)}, p_3^{(3)}\}.$

检验  $x_2$  能否被引入的统计量的值为

$$F_{\frac{1}{9}}^{(3)}(2) = \frac{9 \cdot p_2^{(3)}}{RSS_3} = 5.03.$$

由于  $F_{1,9}(0.1) = 3.36$ , 故  $F_{\frac{1}{9}}^{(3)}(2) > F_{1,9}(0.1)$ , 引入变量  $x_2$ . 对  $A^{(2)}$  施行(2,2)为枢轴的消去变换得

$$A^{(3)} = \begin{pmatrix} 2.57 \times 10^{-3} & 1.86 \times 10^{-4} & -0.97 & 3.89 \times 10^{-4} & 1.45 \\ 1.86 \times 10^{-4} & 6.46 \times 10^{-3} & -0.92 & 5.86 \times 10^{-3} & 0.42 \\ 0.97 & 0.92 & 10.50 & 0.91 & 1.07 \\ 3.89 \times 10^{-4} & 5.86 \times 10^{-3} & -0.91 & 5.63 \times 10^{-3} & -0.24 \\ -1.45 & -0.42 & 1.07 & 0.24 & 49.97 \end{pmatrix}.$$

步骤5 从  $A^{(3)}$  出发,考虑能否从回归方程中剔除变量.计算已选入回归方程的变量  $x_1, x_2, x_4$  的偏回归平方和

$$p_i^{(4)} = [a_{i5}^{(3)}]^2 / a_{ii}^{(3)}, i = 1, 2, 4$$

得  $p_1^{(4)} = 818.09, p_2^{(4)} = 26.79, p_4^{(4)} = 10.23.$

显然  $p_4^{(4)} = \min\{p_1^{(4)}, p_2^{(4)}, p_4^{(4)}\}.$

检验  $x_4$  能否被剔的统计量的值为

$$F_{\frac{1}{9}}^{(4)}(4) = \frac{9 \cdot p_4^{(4)}}{RSS_3} = 1.92,$$

而  $F_{1,9}(0.1) = 3.36$ , 故  $F_{\frac{1}{9}}^{(4)}(4) < F_{1,9}(0.1)$ , 因此,从回归方程中剔除  $x_4$ . 对  $A^{(3)}$  施行以(4,4)为枢轴的消去变换得

$$A^{(4)} = \begin{pmatrix} 2.54 \times 10^{-3} & -2.19 \times 10^{-4} & -0.91 & -0.07 & 1.47 \\ -2.19 \times 10^{-4} & 3.63 \times 10^{-4} & 0.021 & -1.04 & 0.66 \\ 0.81 & -0.02 & 156.68 & -161.08 & 39.17 \\ -0.09 & 1.04 & -161.08 & 177.51 & -41.99 \\ -1.47 & -0.66 & 39.17 & 41.99 & 57.90 \end{pmatrix}.$$

步骤6 从  $A^{(4)}$  出发,考虑能否再从回归方程中剔除变量.计算已在回归方程中的变量  $x_1, x_2$  的偏回归平方和

$$p_i^{(5)} = [a_{i5}^{(4)}]^2 / a_{ii}^{(4)}, i = 1, 2$$



得  $p_1^{(5)} = 850.75, p_2^{(5)} = 1200.$

我们有  $p_1^{(5)} = \min\{p_1^{(5)}, p_2^{(5)}\},$

求得检验变量  $x_1$  能否被剔除的统计量的值为

$$F_{\text{剔}}^{(5)}(1) = \frac{10 \cdot p_1^{(5)}}{RSS_2} = 146.93.$$

而  $F_{1,10}(0.1) = 3.28.$  显然,  $F_{\text{剔}}^{(5)}(1) > F_{1,10}(0.1).$  故变量  $x_1, x_2$  均不能从回归方程中剔除.

步骤 7 从  $A^{(4)}$  出发, 考虑能否再引入新变量. 计算不在回归方程的变量  $x_3, x_4$  的偏回归平方和

$$p_i^{(6)} = [a_{i5}^{(4)}] / a_{55}^{(4)}, i = 3, 4$$

得  $p_3^{(6)} = 9.79, p_4^{(6)} = 10.23,$

且  $p_4^{(6)} = \max\{p_3^{(6)}, p_4^{(6)}\}.$

求得检验变量  $x_4$  能否引入回归方程的统计量的值为

$$F_{\text{引}}^{(6)}(4) = \frac{10 \cdot p_4^{(6)}}{RSS_3} = 146.93.$$

显然,  $F_{\text{引}}^{(6)}(4) > F_{1,10}(0.1) = 3.28,$  故不能再引入变量.

至此, 逐步回归过程结束. 所得最优回归方程及相关的统计量的值为

$$y = 52.67 + 1.47x_1 + 0.66x_2,$$

$$\text{残差平方和: } RSS_{\text{残}} = 57.90,$$

$$\text{回归平方和: } RSS_{\text{回}} = 2657.86,$$

$$\text{复相关系数: } R = 0.99,$$

$$F \text{ 值: } F = 229.52.$$

至于与回归系数估计有关的统计量的值, 在此略去. 这些在标准的统计软件包 (如 SAS, SPSS 等) 都会列出.

在结束本节前, 我们再介绍两种逐步型选元法: 向前法和前后法.

向前法: 向前法的思想是变量由少到多, 每次增加一个, 直至没有可引入的变量为止. 具体实施方法如下.

步骤 1 对  $p-1$  个自变量  $x_1, \dots, x_{p-1}$ , 分别同因变量  $y$  建立一元回归方程

$$y = \beta_0 + \beta_i x_i + e, i = 1, 2, \dots, p-1.$$

计算变量  $x_i$  相应的回归系数的  $F$  检验统计量的值, 记为  $F_1^{(1)}, \dots, F_{p-1}^{(1)}$ , 取其中最大者, 记为  $F_{i_1}^{(1)}$ , 即

$$F_{i_1}^{(1)} = \max\{F_1^{(1)}, \dots, F_{p-1}^{(1)}\}.$$

对给定的显著性水平  $\alpha$ , 记相应的临界值为  $F_{\alpha}^{(1)}$ , 若  $F_{i_1}^{(1)} \geq F_{\alpha}^{(1)}$ , 则将  $x_{i_1}$  引入回归方程.

步骤 2 建立因变量  $y$  与自变量子集  $\{x_{i_1}, x_1\}, \dots, \{x_{i_1}, x_{i_1-1}\}, \{x_{i_1}, x_{i_1+1}\}, \dots, \{x_{i_1}, x_{p-1}\}$  的二元回归方程, 共有  $p-2$  个. 计算变量  $x_k (k \neq i_1)$  的回归系数  $F$  检验的统计量值, 记为  $F_k^{(2)} (k \neq i_1)$ , 选其中最大者, 记为  $F_{i_2}^{(2)}$ , 即

$$F_{i_2}^{(2)} = \max \{ F_1^{(2)}, \dots, F_{i_1-1}^{(2)}, F_{i_1+1}^{(2)}, \dots, F_{p-2}^{(2)} \}.$$

对给定的显著性水平  $\alpha$ , 记其相应的临界值为  $F_{\alpha}^{(2)}$ , 若  $F_{i_2}^{(2)} \geq F_{\alpha}^{(2)}$ , 则引进变量  $x_{i_2}$ ; 否则, 终止变量引进过程.

步骤 3 考虑变量子集  $\{x_{i_1}, x_{i_2}, x_k\}, (k \neq i_1, i_2)$  与因变量的回归, 重复步骤 2.

依此方法重复进行, 每次从未引入回归方程的自变量中选取一个, 直到经  $F$  检验, 没有变量可引入为止.

#### 例 5.4.2 Hald 水泥问题(续).

对例 5.3.1 讨论的 Hald 水泥问题数据, 应用向前法的结果列于表 5.4.1, 取显著性水平为  $\alpha = 0.10$ , 选取的变量子集为  $\{x_1, x_2, x_4\}$ , 从例 5.3.1 和例 5.3.2 知, 若以  $RMS_c$  准则, 该子集是最优回归子集, 而且在  $C_p$  准则下, 它也是接近最优的. 向前法所得的最优回归方程为

$$y = 71.6483 + 1.4519x_1 + 0.4161x_2 - 0.2365x_4.$$

表 5.4.1 Hald 水泥问题的向前法

步号	所选变量	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	$x_4$	117.5679				-0.7382
2	$x_1, x_4$	103.0974	1.4400			-0.6140
3	$x_1, x_2, x_4$	71.6483	1.4519	0.4161		-0.2365

向前法的一个优点是计算量比 § 5.3 介绍的计算所有可能回归的算法要少得多, 譬如,  $p = 11$ , 即有 10 个变量, 计算所有可能回归算法需要计算 1023 个回

归, 而向前法最多只需计算  $\sum_{i=1}^{10} i = \frac{10 \times 11}{2} = 55$  个回归. 但此方法却有一个明显

缺点. 就是由于各自变量之间可能存在着相关关系, 因此后续变量的选入可能会使前面已选入的自变量变得不重要, 而向前法又不考虑从已选变量中剔除不重要的变量, 这样最后得到的“最优”回归方程可能包含一些对  $Y$  影响不大的自变量.

向后法: 向后法与向前法正好相反. 它是先将全部自变量选入回归模型, 然

后逐个剔除对残差平方和贡献较小的自变量. 具体做法如下.

步骤 1 建立  $p-1$  个自变量  $x_1, \dots, x_{p-1}$  与因变量  $y$  的全模型, 计算各变量  $x_i$  相应的回归系数的  $F$  检验统计量的值  $F_1^{(1)}, \dots, F_{p-1}^{(1)}$ , 选其中最小者, 记为  $F_{i_1}^{(1)}$ , 即

$$F_{i_1}^{(1)} = \min\{F_1^{(1)}, \dots, F_{p-1}^{(1)}\}.$$

对给定的显著性水平  $\alpha$ , 记相应的临界值为  $F_{\alpha}^{(1)}$ , 若  $F_{i_1}^{(1)} \leq F_{\alpha}^{(1)}$ , 则从回归方程中剔除  $x_{i_1}$ ; 否则, 选择变量过程结束.

步骤 2 对剩下的  $p-2$  个变量  $x_1, \dots, x_{i_1-1}, x_{i_1+1}, \dots, x_{p-1}$ , 重复步骤 1. 依次进行, 直到经  $F$  检验, 没有变量可剔除为止.

### 例 5.4.3 Hald 水泥问题(续).

对例 5.4.1 的 Hald 水泥数据, 应用向后法, 并取显著性水平  $\alpha = 0.10$ , 其计算结果列于表 5.4.2, 由表中结果可见, 向后法选择的最优回归方程为

$$y = 52.5773 + 1.4683x_1 + 0.6623x_2.$$

它与  $C_p$  准则(见例 5.3.2)的结论是一致的, 也与逐步回归法的结论是一致的.

表 5.4.2 Hald 水泥问题的向后法

步号	所选变量	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
1	$x_1 x_2 x_3 x_4$	62.4054	1.5511	0.5102	0.1019	-0.1441
2	$x_1 x_2 x_4$	71.6483	1.4519	0.4161		-0.2365
3	$x_1 x_2$	52.5773	1.4683	0.6623		

比较本节介绍的三种方法可见, 逐步回归法本质上是向前法, 但吸收了向后法的作法. 从长期应用实践看, 一般情况下, 逐步回归方法所给出的回归方程比较好, 加之计算量少, 又有较为成熟的计算程序, 因此, 到目前为止仍是被广泛采用的变量选择计算方法.

## 习题五

5.1 考虑  $Y$  与  $x_1, x_2, x_3, x_4$  的逐步回归, 由  $A^{(0)} = \begin{pmatrix} X'X & X'y \\ y'X & y'y \end{pmatrix}$  出发, 第一步引入  $x_{i_1}$ , 记  $A^{(1)} = S_{i_1} A^{(0)}$ , 第二步引入  $x_{i_2}$ , 记  $A^{(2)} = S_{i_2} S_{i_1} A^{(0)}$ , 试证明第三步不可能剔除变量.

5.2 假设全模型为

$$y_i = \beta_0 + \beta_1 x_{i_1} + \beta_2 x_{i_2} + e_i, \quad i = 1, 2, \dots, n,$$

选模型为

$$y_i = \beta_0 + \beta_1 x_{i_1} + e_i, \quad i = 1, 2, \dots, n,$$

其中  $Ee_i = 0, Ee_i^2 = \sigma^2, \text{Cov}(e_i, e_j) = 0, i \neq j$ . 若  $\hat{\sigma}^2$  是在选模型下,  $\sigma^2$  的最小二乘估计. 当全模型是正确时, 求  $E(\hat{\sigma}^2)$ , 并问从计算结果能得出什么结论?

5.3 对某地区18年某种消费品销售情况数据, 试用  $RMS_q, C_p$  和  $AIC'$  准则, 建立子集回归模型.

- $y$ : 消费品的销售额(百万元),  
 $x_1$ : 居民可支配收入(元),  
 $x_2$ : 该类消费品的价格指数(%),  
 $x_3$ : 其它消费品平均价格指数(%).

某地 18 年某种消费品销售数据

$y$	$x_1$	$x_2$	$x_3$
7.8	81.2	85.0	87.0
8.4	82.9	92.0	94.0
8.7	83.2	91.5	95.0
9.0	85.9	92.9	95.5
9.6	88.0	93.0	96.0
10.3	99.0	96.0	97.0
10.6	102.0	95.0	97.5
10.9	105.3	95.6	98.0
11.3	117.7	98.9	101.2
12.3	126.4	101.5	102.5
13.5	131.2	102.0	104.0
14.2	148.0	105.0	105.9
14.9	153.0	106.0	109.5
15.9	161.0	109.0	111.0
18.5	170.0	112.0	
19.5	174.0	112.5	112.0
19.9	185.0	113.0	112.3
20.5	189.0	114.0	113.0

5.4 对习题5.3的数据做逐步回归、向前法回归和向后法回归, 并对三种方法所给出的选模型做出评价.

5.5 下表给出10名女中学生体重、胸围、胸围之呼吸差及肺活量数据, 试用逐步回归、向前法和向后法建立子集回归模型.

- $y$ : 肺活量(ml),  $x_1$ : 体重(kg),  $x_2$ : 胸围(cm),  
 $x_3$ : 胸围之呼吸差(cm).

$y$	$x_1$	$x_2$	$x_3$
1600	35	69	0.7
2600	40	74	2.5
2100	40	64	2.0
2650	42	74	3.0
2400	37	72	1.1
2200	45	68	1.5
2750	43	78	4.3
1600	37	66	2.0
2750	44	70	3.2
2500	42	65	3.0

5.6 对习题5.5的数据,试计算  $RMS_q$ ,  $C_p$  和  $AIC$  值,并由挑选最优子集回归,对你的结论做出评价.

5.7 对下表中数据考虑如下方法选择子集回归模型:

- (1) 用向前法建立子集回归模型;
- (2) 用逐步回归法建立子集回归模型;
- (3) 应用所有可能子集回归法建立子集回归模型,计算  $RMS_q$  和  $AIC$  值,你推荐哪一个子集回归模型,为什么?

$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
271.8	783.35	33.53	40.55	16.66	13.20
264.0	748.45	36.50	36.19	16.46	14.11
238.8	684.45	34.66	37.31	17.66	15.68
230.7	827.80	33.13	32.52	17.50	10.53
251.6	860.45	35.75	33.71	16.40	11.00
257.9	875.15	34.46	34.14	16.28	11.31
263.9	909.45	34.60	34.85	16.06	11.96
266.5	905.55	35.38	35.89	15.93	12.58
229.1	756.00	35.85	33.53	16.60	10.66
239.3	769.35	35.68	33.79	16.41	10.85
258.0	793.50	35.35	34.72	16.17	11.41
257.6	801.65	35.04	35.22	15.92	11.91
267.3	819.65	34.07	36.50	16.04	12.85
267.0	808.55	32.20	37.60	16.19	13.58
259.6	774.95	34.32	37.89	16.62	14.21

续表

$y$	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
240.4	711.85	31.08	37.71	17.37	15.56
227.2	694.85	35.73	37.00	18.12	15.83
196.0	638.10	34.11	36.76	18.53	16.41
278.7	774.55	34.79	34.62	15.54	13.10
272.3	757.90	35.77	35.40	15.70	13.63
267.4	753.35	36.44	35.96	16.45	14.51
254.5	704.70	37.82	36.26	17.62	15.38
224.7	666.80	35.07	36.34	18.12	16.10
181.5	568.55	35.26	35.90	19.05	16.73
227.5	653.10	35.56	31.84	16.51	10.58
253.6	704.05	35.73	33.16	16.02	11.28
263.0	709.60	36.46	33.83	15.89	11.91
265.8	726.90	36.26	34.89	15.83	12.65
263.8	697.15	37.20	36.27	16.71	14.06

## 第六章 方差分析模型

在第一章里,我们已经通过实例引进了方差分析模型.在工业、农业、经济、生物、医学等领域的很多实际问题都可以用这样的模型来描述.因此,方差分析模型是很重要的一类线性模型.本章将讨论这种模型的统计分析方法——方差分析.

### § 6.1 单因素方差分析

方差分析渊源于农业田间试验.假定某个农业试验基地引进了  $a$  种小麦品种,在进行大面积种植之前,先进行小范围的试验种植,以便从中挑选出最适合本地区的优良品种.将一块大田划分为面积相等的几个小块,其中  $n_1$  块种植第 1 种小麦,  $n_2$  块种植第 2 种小麦等,  $n_1 + \cdots + n_a = n$ . 因为我们的目的是比较小麦的品种,因此我们所感兴趣的只是小麦品种这一个因素.所有其它因素,例如施肥量、浇水等对这  $n$  块田都控制在相同状态下,也就是说,从播种到收割的全过程中,我们对这  $n$  块田都给予完全相同的“待遇”,所不同的仅仅是小麦品种.

在这个例子中,我们感兴趣的因素只有一个,即小麦品种.每个具体的品种,都称为小麦品种这个因素的一个“水平”.现在有  $a$  个不同品种.于是小麦品种这个因素一共有  $a$  个水平.因此这是单因素  $a$  个水平的问题.

记  $y_{ij}$  为种植第  $i$  个品种的小麦的第  $j$  块田的产量,  $i = 1, \cdots, a; j = 1, \cdots, n_i$ . 对固定的  $i$ ,  $y_{i1}, y_{i2}, \cdots, y_{in_i}$  分别是种植第  $i$  种小麦的  $n_i$  块田的产量.因为除了一些随机误差之外,这  $n_i$  块田的一切生产条件完全一样,因此它们可以看作来自一个正态总体的随机样本,这个正态总体的均值只与  $i$  有关.记这个均值为  $\mu_i$  (如表 6.1.1 所示),也就是说,  $y_{i1}, \cdots, y_{in_i}$  都相互独立,且

$$y_{ij} \sim N(\mu_i, \sigma^2), j = 1, \cdots, n_i, \quad (6.1.1)$$

因此要比较  $a$  个小麦品种的问题就归结为比较  $a$  个正态总体的均值  $\mu_1, \cdots, \mu_a$  的问题.

考虑一般的单因素方差分析问题,我们称所考虑的因素为因素  $A$ , 假定它有  $a$  个水平,我们的目的是比较这  $a$  个水平的差异.假定对第  $i$  个水平,我们所感兴趣的指标观测值如表(6.1.1).将其改写成如下形式:

表 6.1.1 单因素方差分析问题

水 平	总 体	样 本
1	$N(\mu_1, \sigma^2)$	$y_{11} \quad y_{12} \quad \cdots \quad y_{1n_1}$
2	$N(\mu_2, \sigma^2)$	$y_{21} \quad y_{22} \quad \cdots \quad y_{2n_2}$
$\vdots$	$\vdots$	$\vdots \quad \vdots \quad \quad \quad \vdots$
$a$	$N(\mu_a, \sigma^2)$	$y_{a1} \quad y_{a2} \quad \cdots \quad y_{an_a}$

$$\begin{cases} y_{ij} = \mu_i + e_{ij}, \\ e_{ij} \sim N(0, \sigma^2), \end{cases} \quad i = 1, \cdots, a, j = 1, \cdots, n_i, \quad (6.1.2)$$

其中  $\mu_i$  是第  $i$  个总体的均值,  $e_{ij}$  是相应的试验误差, 比较因素  $A$  的  $a$  个水平的差异归结为比较这  $a$  个总体的均值. 记

$$\mu = \frac{1}{n} \sum_{i=1}^a n_i \mu_i, \quad n = \sum_{i=1}^a n_i,$$

$$\alpha_i = \mu_i - \mu,$$

这里  $\mu$  为所有样本均值  $E(y_{ij})$  的总平均.  $\alpha_i$  为第  $i$  个水平对指标  $Y$  的效应, 反映第  $i$  水平下的均值与总平均的差异. 不难验证,  $\sum_{i=1}^a n_i \alpha_i = 0$ . 因为  $\mu_i = \mu + \alpha_i$ , 于是(6.1.2)可以写成

$$\begin{cases} y_{ij} = \mu + \alpha_i + e_{ij}, \\ e_{ij} \sim N(0, \sigma^2), \\ \sum_{i=1}^a n_i \alpha_i = 0, \end{cases} \quad (6.1.3)$$

这就是单因素方差分析模型. 写成矩阵形式即为

$$\begin{cases} \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \\ \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n), \\ \mathbf{h}'\boldsymbol{\beta} = 0, \end{cases} \quad (6.1.4)$$

其中

$$\mathbf{y}_{n \times 1} = (y_{11}, \cdots, y_{1n_1}, y_{21}, \cdots, y_{2n_2}, \cdots, y_{an_a})',$$



$$\mathbf{X}_{n \times (a+1)} = \begin{pmatrix} 1 & 1 & & & & & & \\ \vdots & \vdots & & & & & & \\ 1 & 1 & & & & & & \\ & 1 & & 1 & & & & \\ \vdots & \vdots & & \vdots & & & & \\ 1 & & & 1 & & & & \\ \vdots & & & \vdots & & & & \\ & & & & \ddots & & & \\ 1 & & & & & & 1 & \\ \vdots & & & & & & \vdots & \\ 1 & & & & & & 1 & \end{pmatrix} \begin{matrix} \left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} n_1 \text{ 行} \\ \\ \left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} n_2 \text{ 行} \\ \vdots \\ \left. \begin{matrix} \\ \\ \\ \end{matrix} \right\} n_a \text{ 行} \\ \vdots \end{matrix},$$

$$\boldsymbol{\beta}_{(a+1) \times 1} = (\mu, \alpha_1, \alpha_2, \dots, \alpha_a)',$$

$$\mathbf{e}_{n \times 1} = (e_{11}, \dots, e_{1n_1}, e_{21}, \dots, e_{2n_2}, \dots, e_{an_a})',$$

$$\mathbf{h}_{(a+1) \times 1} = (0, n_1, n_2, \dots, n_a)',$$

可见,单因素方差分析模型是一个带约束条件  $\mathbf{h}'\boldsymbol{\beta} = 0$  的线性模型.

对此模型检验因素 A 的  $a$  个水平下的均值是否有显著差异,即检验假设

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a,$$

这等价于检验

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0.$$

如果  $H_0$  被拒绝,则说明因素 A 的各水平的效应之间有显著的差异.在小麦品种的例子中,就是  $a$  种小麦品种之间有显著差异.下面我们将导出  $H_0$  的检验统计量.

以  $\bar{y}$  表示所有  $y_{ij}$  的总平均值,即

$$\bar{y} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}.$$

考虑统计量

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2,$$

称  $SS_T$  为总的离差平方和,简称为总平方和.它反映全部试验数据之间的差异.对其进行分解:

$$\begin{aligned} SS_T &= \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot} + \bar{y}_{i\cdot} - \bar{y})^2 \end{aligned}$$

$$= \sum_{i=1}^a \sum_{j=1}^{n_i} [(y_{ij} - \bar{y}_{i.})^2 + 2(y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}) + (\bar{y}_{i.} - \bar{y})^2],$$

其中  $\bar{y}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$ , 为第  $i$  水平下的样本均值. 由于

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}) = 0,$$

所以有

$$\begin{aligned} SS_T &= \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y})^2 \\ &= SS_E + SS_A, \end{aligned} \quad (6.1.5)$$

其中

$$SS_E = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2, \quad (6.1.6)$$

$$SS_A = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y})^2. \quad (6.1.7)$$

这里  $SS_E$  表示了随机误差的影响. 因为对固定的  $i$ , 如(6.1.1)所示, 所有观察值  $y_{i1}, \dots, y_{in_i}$  都是来自同一个正态总体  $N(\mu_i, \sigma^2)$  的样本. 因此, 它们之间的差异完

全是由于随机误差所致. 而  $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$  是这  $n_i$  个数据的变动平方和, 正是它们差异大小的度量. 将  $a$  组这样的变动平方和相加, 就得到了  $SS_E$ , 通常称  $SS_E$  为误差平方和或组内平方和.

将  $SS_A$  改写为

$$SS_A = \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y})^2, \quad (6.1.8)$$

因为  $\bar{y}_{i.}$  为第  $i$  个总体的样本均值, 它是第  $i$  个总体均值  $\mu_i$  的估计, 因此  $a$  个总体均值  $\mu_1, \dots, \mu_a$  之间差异愈大, 这些样本均值  $\bar{y}_{1.}, \dots, \bar{y}_{a.}$  之间的差异也就愈倾

向于大. 平方和  $\sum_{i=1}^a (\bar{y}_{i.} - \bar{y})^2$  正是这种差异大小的度量. 而(6.1.8)中每一项

$(\bar{y}_{i.} - \bar{y})^2$  前面的系数  $n_i$  反映了第  $i$  个总体的样本大小在平方和  $SS_A$  中的作用.

总起来,  $SS_A$  是  $a$  个总体均值  $\mu_1, \dots, \mu_a$  差异大小的度量, 通常称为组间平方和或因素  $A$  的平方和. (6.1.5) 是很重要的平方和分解公式, 它将总平方和按其来源分解成两部分. 一部分是  $SS_E$ , 即误差平方和, 是由随机误差引起的. 另一部分是

$SS_A$ , 即因素 A 的平方和, 是由因素 A 的各水平的差异引起的.

从(6.1.1)知

$$S_i^2 = \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2 / (n_i - 1) \quad (6.1.9)$$

是来自第  $i$  个总体  $N(\mu_i, \sigma^2)$  的样本  $y_{i1}, \dots, y_{in_i}$  的样本方差, 因而, 它是  $\sigma^2$  的一个无偏估计, 即

$$ES_i^2 = \sigma^2. \quad (6.1.10)$$

于是

$$\begin{aligned} E(SS_E) &= \sum_{i=1}^a E \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2 \\ &= \sum_{i=1}^a (n_i - 1) \sigma^2 \\ &= (n - a) \sigma^2, \end{aligned}$$

此式表明  $SS_E/(n - a)$  是  $\sigma^2$  的一个无偏估计.

另一方面

$$\begin{aligned} E(SS_A) &= E \left[ \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y})^2 \right] \\ &= E \left[ \sum_{i=1}^a n_i (\bar{y}_{i.} - \bar{y} - \alpha_i + \alpha_i)^2 \right] \\ &= \sum_{i=1}^a n_i [E(\bar{y}_{i.} - \bar{y} - \alpha_i)^2 + \alpha_i^2] \\ &= \sum_{i=1}^a n_i \left( \frac{\sigma^2}{n_i} - \frac{\sigma^2}{n} \right) + \sum_{i=1}^a n_i \alpha_i^2 \\ &= (a - 1) \sigma^2 + \sum_{i=1}^a n_i \alpha_i^2, \end{aligned}$$

所以有

$$E \left( \frac{SS_A}{a - 1} \right) = \sigma^2 + \frac{1}{a - 1} \sum_{i=1}^a n_i \alpha_i^2.$$

由这个式子也可以看出,  $SS_A/(a - 1)$  反映了各水平效应的影响. 当  $H_0$  成立时,  $SS_A/(a - 1)$  是  $\sigma^2$  的一个无偏估计. 所以, 从直观上看, 若  $H_0$  为真, 则比值

$$\frac{SS_A/(a - 1)}{SS_E/(n - a)}$$

将接近于 1; 而当  $H_0$  不成立时, 它将倾向于比较大. 这就启发我们通过比较  $SS_A$

与  $SS_E$  的大小来检验  $H_0$ . 记统计量

$$F = \frac{SS_A/(a-1)}{SS_E/(n-a)}.$$

下面将导出当  $H_0$  成立时  $F$  的分布.

注意到  $\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$  是总体  $N(\mu_i, \sigma^2)$  的样本方差的  $n_i - 1$  倍, 由例 2.4.1 可得

$$\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 / \sigma^2 \sim \chi_{(n_i-1)}^2, i = 1, \dots, a.$$

又由样本  $y_{ij}$  的独立性知, 此  $a$  项平方和相互独立, 因为独立  $\chi^2$  变量之和仍为  $\chi^2$  变量(见定理 2.4.3), 于是

$$SS_E / \sigma^2 = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 / \sigma^2 \sim \chi_{\sum_{i=1}^a (n_i-1)}^2 = \chi_{n-a}^2. \quad (6.1.11)$$

应用著名的 Cochran 定理(见项可风等(1989), p. 22)可以证明, 当  $H_0$  成立时,  $SS_A / \sigma^2 \sim \chi_{a-1}^2$ , 并且与  $SS_E$  相互独立. 因此, 当  $H_0$  成立时

$$F = \frac{SS_A/(a-1)}{SS_E/(n-a)} \sim F_{a-1, n-a}. \quad (6.1.12)$$

于是  $F$  可以作为  $H_0$  的检验统计量. 对给定的显著性水平  $\alpha$ , 若  $F > F_{a-1, n-a}(\alpha)$ , 我们就拒绝原假设, 认为因素  $A$  的  $a$  个水平效应有显著差异. 相反, 若  $F \leq F_{a-1, n-a}(\alpha)$ , 我们接受原假设, 认为因素  $A$  的  $a$  个水平效应没有明显差异. 通常把计算结果列成表 6.1.2 的形式, 称为方差分析表.

表 6.1.2 单因素方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	$SS_A$	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F = \frac{MS_A}{MS_E}$
误差	$SS_E$	$n-a$	$MS_E = \frac{SS_E}{n-a}$	
总和	$SS_T$	$n-1$		

在此表中, 我们依照惯例, 把每个平方和除以相应的自由度称为均方, 例如因素  $A$  的平方和  $SS_A$  除以它的自由度  $a-1$ , 得到因素  $A$  的均方, 记为  $MS_A$ .

**例 6.1.1** 设有三个小麦品种, 经试种得每公顷产量数据如表 6.1.3(单位:  $\text{kg}/\text{hm}^2$ )

表 6.1.3 小麦品种试验数据

试验号 品种	1	2	3	4	5
1	4 350	4 650	4 080	4 275	
2	4 125	3 720	3 810	3 960	3 930
3	4 695	4 245	4 620		

问:不同品种的小麦产量之间有无显著的差异?

在这项试验中考虑的因素只有一个,即“品种”,它取 3 个不同的水平.利用表 6.1.3 中的数据计算可得

全部数据的总平均值  $\bar{y} = 4\ 205$ ,

三个品种各自数据的平均值

$\bar{y}_{1.} = 4\ 339, n_1 = 4,$

$\bar{y}_{2.} = 3\ 909, n_2 = 5,$

$\bar{y}_{3.} = 4\ 520, n_3 = 3,$

总平方和  $SS_T = 1\ 186\ 800.00$  其自由度  $n - a = 12 - 3 = 9$ .

因素平方和  $SS_A = 807\ 311.25,$

误差平方和  $SS_E = 379\ 488.75,$

均方值  $MS_A = 807\ 311.25/2 = 403\ 655.62,$

$MS_E = 379\ 488.75/9 = 42\ 165.42,$

$F_{\text{比}} = 403\ 655.62/42\ 165.42 = 9.57.$

将计算结果填入方差分析表中得表 6.1.4.

表 6.1.4 小麦品种的方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	807 311.25	2	403 655.62	9.57
误差	379 488.75	9	42 165.42	
总和	1 186 800.00	11		

查表得  $F_{2,9}(0.05) = 4.26, F_{2,9}(0.01) = 8.02$ , 由于  $9.57 > 8.02$ , 可认为品种效应具有显著的差异.

如果  $F$  检验的结论是拒绝原假设, 则表明从现在掌握的数据看, 我们有理由认为因素 A 的  $a$  个水平效应有显著差异, 也就是说,  $\mu_1, \dots, \mu_a$  不完全相同. 这时, 我们还需要对每一对  $\mu_i$  和  $\mu_j$  之间的差异程度作出估计. 这就要对效应之差  $\mu_i - \mu_j$  作区间估计.

从假设(6.1.1), 我们知道  $\bar{y}_i$  是来自总体  $N(\mu_i, \sigma^2)$  的  $n_i$  个样本的均值, 根

据正态总体样本均值的性质知

$$y_{i\cdot} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right), \quad i = 1, \dots, a,$$

并且  $\bar{y}_{i\cdot}$  与  $\bar{y}_{j\cdot}$  ( $i \neq j$ ) 相互独立. 再由正态分布的性质(推论 2.3.3), 容易证明

$$\bar{y}_{i\cdot} - \bar{y}_{j\cdot} \sim N\left(\mu_i - \mu_j, \left(\frac{1}{n_i} + \frac{1}{n_j}\right)\sigma^2\right), \quad (6.1.13)$$

因而

$$U = \frac{(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - (\mu_i - \mu_j)}{\sigma \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim N(0, 1). \quad (6.1.14)$$

记

$$\hat{\sigma}^2 = \frac{SS_E}{n-a},$$

由(6.1.11)知

$$\frac{(n-a)\hat{\sigma}^2}{\sigma^2} = \frac{SS_E}{\sigma^2} \sim \chi_{n-a}^2. \quad (6.1.15)$$

再由正态总体样本均值与样本方差的独立性可推出,  $U$  和  $\hat{\sigma}^2$  相互独立. 因此, 根据  $t$  分布的定义, 从(6.1.14)和(6.1.15)可得

$$\frac{(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - (\mu_i - \mu_j)}{\hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t_{n-a}.$$

对给定的  $\alpha$ , 随机事件

$$\left| \frac{(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - (\mu_i - \mu_j)}{\hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \right| \leq t_{n-a}\left(\frac{\alpha}{2}\right)$$

发生的概率为  $1 - \alpha$ , 其中  $t_{n-a}\left(\frac{\alpha}{2}\right)$  为自由度为  $n - a$  的  $t$  分布的上侧  $\frac{\alpha}{2}$  分位点. 因此对固定的  $i, j, \mu_i - \mu_j$  的置信系数  $1 - \alpha$  的置信区间为

$$\left[ (\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-a}\left(\frac{\alpha}{2}\right), (\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) + \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-a}\left(\frac{\alpha}{2}\right) \right]. \quad (6.1.16)$$

如果这个区间包含零, 则表明我们可以以概率  $1 - \alpha$  断言  $\mu_i$  与  $\mu_j$  没有显著差异. 如果整个区间落在零的左边, 则我们以概率  $1 - \alpha$  断言  $\mu_i$  小于  $\mu_j$ . 相反, 如果整个区间落在零的右边, 则我们以概率  $1 - \alpha$  断言  $\mu_i$  大于  $\mu_j$ .

#### 例 6.1.2 (续例 6.1.1)

由于在例 6.1.1 中检验结果是显著的, 所以我们还需进一步对每一对  $\mu_i$  和

$\mu_i$  之差  $\mu_i - \mu_j$  作区间估计. 利用公式 (6.1.16), 取水平  $\alpha = 0.05$ , 查表得

$t_{n-\alpha}\left(\frac{\alpha}{2}\right) = t_9(0.025) = 2.262$ . 得到  $\mu_i - \mu_j$  的置信区间分别为

$$\mu_1 - \mu_2 \in [118.17, 741.34],$$

$$\mu_1 - \mu_3 \in [-536.01, 173.51],$$

$$\mu_2 - \mu_3 \in [-950.21, -271.79].$$

可见第一个区间整个在零点的右边, 所以我们以概率 95% 断言  $\mu_1$  大于  $\mu_2$ . 第二个区间包含零点, 所以我们只能从点估计上看  $\mu_3$  大于  $\mu_1$ . 第三个区间完全位于零点左边, 所以我们以概率 95% 断言  $\mu_3$  大于  $\mu_2$ .

对于每一对固定的  $i, j$ , 我们用 (6.1.16) 构造出置信系数为  $1 - \alpha$  的置信区间. 但是对多个这样的置信区间, 它们联合起来的置信系数就不再是  $1 - \alpha$ . 为了说明这个问题, 我们需要下面的 Bonferroni 不等式.

假设  $E_i, i = 1, \dots, m$  为  $m$  个随机事件,  $P(E_i) = 1 - \alpha, i = 1, \dots, m$ , 则

$$P\left(\bigcap_{i=1}^m E_i\right) = 1 - P\left(\overline{\bigcap_{i=1}^m E_i}\right) = 1 - P\left(\bigcup_{i=1}^m \bar{E}_i\right) \geq 1 - \sum_{i=1}^m P(\bar{E}_i) = 1 - m\alpha,$$

即

$$P\left(\bigcap_{i=1}^m E_i\right) \geq 1 - m\alpha.$$

这就是著名的 Bonferroni 不等式. 这个不等式说明,  $m$  个事件若每个单独发生的概率为  $1 - \alpha$ , 那么它们同时发生概率不再是  $1 - \alpha$ , 而是大于或等于  $1 - m\alpha$ , 它可能比  $1 - \alpha$  小得多, 为了使它们同时发生的概率不低于  $1 - \alpha$ , 一个办法是把每个事件发生的概率提高到  $1 - \frac{\alpha}{m}$ , 即  $P(E_i) = 1 - \frac{\alpha}{m}$ , 此时我们有

$$P\left(\bigcap_{i=1}^m E_i\right) \geq 1 - \alpha$$

应用这个思想, 我们可以构造  $m$  个形如  $\mu_i - \mu_j$  的效应之差的的同时置信区间. 事实上, 对每个  $\mu_i - \mu_j$  应用 (6.1.16) 构造置信系数  $1 - \frac{\alpha}{m}$  的置信区间

$$\left[ (\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) - \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-\alpha}\left(\frac{\alpha}{2m}\right), (\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) + \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} t_{n-\alpha}\left(\frac{\alpha}{2m}\right) \right]. \quad (6.1.17)$$

那么这  $m$  个  $\mu_i - \mu_j$  同时分别落在这  $m$  个置信区间的置信系数为  $1 - \alpha$ .

### 例 6.1.3 (续例 6.1.2)

下面我们利用公式 (6.1.17), 取  $\alpha = 0.05$ , 计算出  $\mu_i - \mu_j$  的置信系数为  $1 - \alpha$  的同时置信区间为

$$\mu_1 - \mu_2 \in [25.691, 833.809],$$

$$\mu_1 - \mu_3 \in [-641.291, 278.791],$$

$$\mu_2 - \mu_3 \in [-1050.883, -171.117],$$

即我们以 95% 的概率说明上面三个式子同时成立. 从同时置信区间及前面点估计的结果来看, 品种 3 应视为最好, 品种 1 其次.

## § 6.2 两因素方差分析

在一项实际试验中, 往往有这样的情况, 研究者本想考察某个因素对指标的影响, 但是由于客观条件的限制, 还有个别因素不可能在所有试验中把它们控制在完全相同的状态. 譬如, 在上节一开始讨论的小麦品种的例子中, 试验者要研究的是“小麦品种”这一因素对产量的影响. 但在实际中可能会出现这样的情况, 很难找到一大块田土质肥沃程度完全一样. 因此“土质”就成为另一个因素不可避免地进入了试验, 导致了两因素的试验问题. 在农业试验中解决这个问题的方法是采用所谓区组设计. 它的做法是, 先把一块田分成若干块, 譬如  $b$  块, 使得每块田的土质肥沃程度基本上保持一样. 在试验设计中, 称这种块为区组, 然后把每一个区组又分成若干小块, 称为试验单元. 现在有  $a$  种小麦品种, 方便的方法就是把每个区组分  $a$  个试验单元. 在每一个试验单元上种植一种小麦. 若用  $y_{ij}$  表示在第  $j$  个区组中种植第  $i$  种小麦的那个试验单元的产量, 则  $y_{ij}$  就可表为

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad i = 1, \dots, a, j = 1, \dots, b, \quad (6.2.1)$$

这里  $\mu$  称为总平均,  $\alpha_i$  为第  $i$  种小麦品种的效应,  $\beta_j$  为第  $j$  个区组的效应,  $e_{ij}$  为随机误差.

上面我们讨论的只是导致两因素问题的一种情况. 在实际应用中, 更多的是, 研究者所感兴趣的问题本身就是两因素的. 例如, 在一项工业试验中, 影响产品质量的因素是反应温度和反应压力. 试验者的目的是选择最好的生产条件, 若反应温度有  $a$  个水平, 反应压力有  $b$  个水平. 记  $y_{ij}$  为在反应温度处于第  $i$  个水平和反应压力处于第  $j$  个水平时产品质量的指标值, 那末  $y_{ij}$  也有表达式 (6.2.1). 如果影响产品质量的因素不止反应温度和反应压力这两个, 还有反应时间和催化剂种类两个因素, 这样共有四个因素. 当我们把任两个因素控制在某一状态而研究剩余两个因素对产品质量影响时, 同样导致了一个两因素问题.

考虑一般的两因素试验问题, 将这两个因素分别记为  $A$  和  $B$ . 假定因素  $A$  有  $a$  个不同的水平, 记为  $A_1, \dots, A_a$ , 而因素  $B$  有  $b$  个不同的水平, 记为  $B_1, \dots, B_b$ . 在因素  $A$  和  $B$  的各个水平的组合下做  $c$  次试验. 设  $y_{ijk}$  为在水平组合  $(A_i, B_j)$  下第  $k$  次试验的指标值. 对固定的  $i$  和  $j$ ,  $y_{ij1}, y_{ij2}, \dots, y_{ijc}$  都是在水平组合



$(A_i, B_j)$  下的指标观测值, 我们可以把它们看成来自一个正态总体的样本, 这个正态总体的均值只与  $i, j$  有关, 记这个均值为  $\mu_{ij}$ . 于是  $y_{ij1}, y_{ij2}, \dots, y_{ijc}$  都相互独立, 且

$$y_{ijk} \sim N(\mu_{ij}, \sigma^2), k = 1, \dots, c. \quad (6.2.2)$$

将这些数据列成表, 如表 6.2.1.

表 6.2.1 两因素方差分析问题数据

因素 B 各水平 因素 A 各水平	$B_1$	$B_2$	...	$B_b$
$A_1$	$y_{111} y_{112} \dots y_{11c}$	$y_{121} y_{122} \dots y_{12c}$	...	$y_{1b1} y_{1b2} \dots y_{1bc}$
$A_2$	$y_{211} y_{212} \dots y_{21c}$	$y_{221} y_{222} \dots y_{22c}$	...	$y_{2b1} y_{2b2} \dots y_{2bc}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
$A_a$	$y_{a11} y_{a12} \dots y_{a1c}$	$y_{a21} y_{a22} \dots y_{a2c}$	...	$y_{ab1} y_{ab2} \dots y_{abc}$

我们可以将(6.2.2)改写成如下形式:

$$\begin{cases} y_{ijk} = \mu_{ij} + e_{ijk}, \\ i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, \\ e_{ijk} \sim N(0, \sigma^2). \end{cases} \quad (6.2.3)$$

为了做统计分析, 我们需要将均值  $\mu_{ij}$  做适当的分解, 为此, 引入

$$\mu = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b \mu_{ij},$$

$$\bar{\mu}_{i\cdot} = \frac{1}{b} \sum_{j=1}^b \mu_{ij},$$

$$\bar{\mu}_{\cdot j} = \frac{1}{a} \sum_{i=1}^a \mu_{ij},$$

$$\alpha_i = \bar{\mu}_{i\cdot} - \mu, \quad i = 1, \dots, a,$$

$$\beta_j = \bar{\mu}_{\cdot j} - \mu, \quad j = 1, \dots, b,$$

$$\gamma_{ij} = \mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \mu,$$

其中  $\mu$  为总平均,  $\alpha_i$  为因素 A 的水平  $A_i$  的效应,  $\beta_j$  为因素 B 的水平  $B_j$  的效应.  $\gamma_{ij}$  的意义不是很明显, 我们把它改写为

$$\begin{aligned} \gamma_{ij} &= \mu_{ij} - (\bar{\mu}_{i\cdot} - \mu) - (\bar{\mu}_{\cdot j} - \mu) - \mu \\ &= (\mu_{ij} - \mu) - \alpha_i - \beta_j, \end{aligned}$$

其中  $\mu_{ij} - \mu$  反映了水平组合  $(A_i, B_j)$  对指标值的效应. 在许多情况下, 水平组合

$(A_i, B_j)$ 的这种效应并不等于水平  $A_i$  的效应  $\alpha_i$  和  $B_j$  的效应  $\beta_j$  之和. 我们把效应  $\mu_{ij} - \mu$  减去  $A_i$  的效应  $\alpha_i$  和  $B_j$  的效应  $\beta_j$  所得到的差  $\gamma_{ij}$  称为  $A_i$  和  $B_j$  的交互效应. 通常把因素  $A$  和  $B$  对试验指标的交互效应设想为某一因素的效应. 称这个因素为  $A$  与  $B$  的交互作用, 记为  $A \times B$ , 不难验证

$$\sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{i=1}^a \sum_{j=1}^b \gamma_{ij} = 0.$$

引进了上述记号之后, 就有  $\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$ , 于是我们可以把(6.2.3)写成如下形式:

$$\begin{cases} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \\ i = 1, \dots, a, j = 1, \dots, b, k = 1, \dots, c, \\ e_{ijk} \sim N(0, \sigma^2), \text{ 且相互独立,} \end{cases} \quad (6.2.4)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \quad \sum_{i=1}^a \sum_{j=1}^b \gamma_{ij} = 0, \end{array} \right.$$

这就是两因素方差分析模型. 下面分两种情况来讨论.

### 一、无交互效应的情形

假设  $\gamma_{ij} = 0, i = 1, 2, \dots, a, j = 1, 2, \dots, b$ , 即不存在交互效应. 为简单计, 我们只考虑每种水平组合下试验次数为  $c = 1$  的情形. 对  $c > 1$  的情形, 统计分析方法完全相同. 因为  $\gamma_{ij} = 0$ , 于是

$$\mu_{ij} = \mu + \alpha_i + \beta_j, \quad i = 1, \dots, a, \quad j = 1, \dots, b.$$

此时, 模型(6.2.4)可写成

$$\begin{cases} y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \\ i = 1, \dots, a, j = 1, \dots, b, \\ e_{ij} \sim N(0, \sigma^2), \text{ 且相互独立,} \\ \sum_{i=1}^a \alpha_i = 0, \quad \sum_{j=1}^b \beta_j = 0, \end{cases} \quad (6.2.5)$$

这就是无交互效应的两因素方差分析模型. 我们看到(6.2.1)与此处  $y_{ij}$  的分解式具有相同的形式. 我们的目的是要考查因素  $A$  或  $B$  各水平对指标的影响有无显著差异, 这归结为对假设

$$H_1: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

或

$$H_2: \beta_1 = \beta_2 = \dots = \beta_b = 0,$$

的检验. 下面采用与单因素方差分析模型类似的方法导出检验统计量.

记

$$\bar{y} = \frac{1}{ab} \sum_{i=1}^a \sum_{j=1}^b y_{ij},$$

$$\bar{y}_{i\cdot} = \frac{1}{b} \sum_{j=1}^b y_{ij},$$

$$y_{\cdot j} = \frac{1}{a} \sum_{i=1}^a y_{ij},$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y})^2,$$

其中  $SS_T$  为全部试验数据的总变差,称为总平方和,对其进行分解

$$\begin{aligned} SS_T &= \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y} + \bar{y}_{i\cdot} - \bar{y} + \bar{y}_{\cdot j} - \bar{y})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2 + \sum_{i=1}^a b(\bar{y}_{i\cdot} - \bar{y})^2 \\ &\quad + \sum_{j=1}^b a(\bar{y}_{\cdot j} - \bar{y})^2 \\ &= SS_E + SS_A + SS_B, \end{aligned}$$

读者可以验证,在上述平方和分解中交叉项均为 0. 其中

$$SS_E = \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y})^2,$$

$$SS_A = b \sum_{i=1}^a (\bar{y}_{i\cdot} - \bar{y})^2,$$

$$SS_B = a \sum_{j=1}^b (\bar{y}_{\cdot j} - \bar{y})^2.$$

我们先来看  $SS_A$  的统计意义. 因为  $\bar{y}_{i\cdot}$  是水平  $A_i$  下所有观测值的平均, 所以  $\sum_{i=1}^a (\bar{y}_{i\cdot} - \bar{y})^2$  反映了  $\bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \dots, \bar{y}_{a\cdot}$  差异的程度. 这种差异是由于因素 A 的不同水平所引起的, 因此  $SS_A$  称为因素 A 的平方和. 类似地,  $SS_B$  称为因素 B 的平方和. 至于  $SS_E$  的意义不甚明显, 我们可以这样来理解: 因为

$$SS_E = SS_T - SS_A - SS_B, \quad (6.2.6)$$

在我们所考虑的两因素问题中, 除了因素 A 和 B 之外, 剩余的再没有其它系统性因素的影响, 因此从总平方和中减去  $SS_A$  和  $SS_B$  之后, 剩下的数据变差只能归入随机误差, 故  $SS_E$  反映了试验的随机误差.

有了总平方和的分解式

$$SS_T = SS_E + SS_A + SS_B$$

以及各个平方和的统计意义,我们就可以明白,假设  $H_1$  (或  $H_2$ ) 的检验统计量应取为  $SS_A$  (或  $SS_B$ ) 与  $SS_E$  的比.

和一元方差分析相类似,可以证明,当  $H_1$  成立时,  $SS_A/\sigma^2 \sim \chi_{a-1}^2$ , 并且与  $SS_E$  相互独立,而  $SS_E/\sigma^2 \sim \chi_{(a-1)(b-1)}^2$ . 于是当  $H_1$  成立时

$$F_A = \frac{SS_A/(a-1)}{SS_E/(a-1)(b-1)} \sim F_{a-1, (a-1)(b-1)}, \quad (6.2.7)$$

这个统计量与(6.1.12)非常相似,只是分母的自由度不同. 它可以用来检验假设  $H_1$ . 对给定的水平  $\alpha$ , 当  $F_A > F_{a-1, (a-1)(b-1)}(\alpha)$  时, 我们拒绝原假设, 认为因子  $A$  的  $a$  个水平的效应具有显著差异.

完全类似地, 当  $H_2$  成立时

$$F_B = \frac{SS_B/(b-1)}{SS_E/(a-1)(b-1)} \sim F_{b-1, (a-1)(b-1)}, \quad (6.2.8)$$

它可以用来检验  $H_2$ .

与单因素情形一样, 我们可以写出方差分析表, 如表 6.2.2 所示.

表 6.2.2 无交互效应的两因素方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	$SS_A$	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_A = \frac{MS_A}{MS_E}$
因素 B	$SS_B$	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_B = \frac{MS_B}{MS_E}$
误差	$SS_E$	$(a-1)(b-1)$	$MS_E = \frac{SS_E}{(a-1)(b-1)}$	
总和	$SS_T$	$ab-1$		

**例 6.2.1** 一种火箭使用了四种燃料、三种推进器, 进行射程试验. 对于每种燃料与每种推进器的组合作一次试验, 得到试验数据如表 6.2.3. 问各种燃料之间及各种推进器之间有无显著差异?

表 6.2.3 火箭试验数据

燃料 A	推进器 B		
	$B_1$	$B_2$	$B_3$
$A_1$	58.2	56.2	65.3
$A_2$	49.1	54.1	51.6
$A_3$	60.1	70.9	39.2
$A_4$	75.8	58.2	48.7

这是一个双因素试验, 且不考虑交互作用. 记“燃料”为因素  $A$ , 它有 4 个水

平. 水平效应为  $\alpha_i, i=1, 2, 3, 4$ . “推进器”为因素  $B$ , 它有 3 个水平. 水平效应为  $\beta_j, j=1, 2, 3$ . 我们在显著性水平  $\alpha=0.05$  下检验

$$H_1: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0,$$

$$H_2: \beta_1 = \beta_2 = \beta_3 = 0.$$

由表中数据计算可得

全部数据的总平均值  $\bar{y} = 54$ ;

因素  $A$  的各水平下的数据平均值  $\bar{y}_{\cdot 1} = 59.9, \bar{y}_{\cdot 2} = 51.6, \bar{y}_{\cdot 3} = 56.7, \bar{y}_{\cdot 4} = 60.9$ ;

因素  $B$  的各水平下的数据平均值  $\bar{y}_{1 \cdot} = 60.8, \bar{y}_{2 \cdot} = 59.9, \bar{y}_{3 \cdot} = 51.2$ .

总平方和  $SS_T = 1113.42$ ,

因素  $A$  的平方和  $SS_A = 157.59$ ,

因素  $B$  的平方和  $SS_B = 223.85$ ,

误差平方和  $SS_E = 731.96$ .

将计算结果填入方差分析表 6.2.2 中得表 6.2.4.

表 6.2.4 火箭试验的方差分析表

方差来源	平方和	自由度	均方	F 比
因素 A	157.59	3	52.53	$F_A = 0.4306$
因素 B	223.85	2	111.93	$F_B = 0.9175$
误差	731.98	6	122.00	
总和	1113.42	11		

因为  $F_{3,6}(0.05) = 4.76 > F_A = 0.4306$ , 所以接受  $H_1$ . 又因为  $F_{2,6}(0.05) = 5.14 > F_B = 0.9175$ , 所以接受  $H_2$ . 检验结果即各种燃料和各种推进器之间的差异对于火箭射程无显著影响.

如果经过  $F_A$  检验,  $H_1$  被拒绝, 那么在这种情况下, 我们认为因素  $A$  的  $a$  个水平效应  $\alpha_1, \dots, \alpha_a$  不全相同. 和一元方差分析一样, 此时我们希望比较  $\alpha_i$  的大小, 这就要做  $\alpha_i - \alpha_t$  的区间估计.

因为  $y_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ , 利用  $\sum_{j=1}^b \beta_j = 0$  及正态分布的性质, 可以证明

$$y_{i \cdot} \sim N\left(\mu + \alpha_i, \frac{\sigma^2}{b}\right), \quad i = 1, \dots, a,$$

于是

$$y_{i \cdot} - y_{t \cdot} \sim N\left(\alpha_i - \alpha_t, \frac{2}{b}\sigma^2\right). \quad (6.2.9)$$

采用与一元方差分析完全相同的讨论, 并用

$$\hat{\sigma}^2 = \frac{SS_E}{(a-1)(b-1)}$$

作为  $\sigma^2$  的估计,可以得到对固定的  $i, t, \alpha_i - \alpha_t$  的置信系数为  $1 - \alpha$  的置信区间为

$$\left[ \bar{y}_{i.} - \bar{y}_{t.} - \hat{\sigma} \sqrt{\frac{2}{b} t_{(a-1)(b-1)} \left( \frac{\alpha}{2} \right)}, (\bar{y}_{i.} - \bar{y}_{t.}) + \hat{\sigma} \sqrt{\frac{2}{b} t_{(a-1)(b-1)} \left( \frac{\alpha}{2} \right)} \right]. \quad (6.2.10)$$

如果这个区间包含零,则表明我们以概率  $1 - \alpha$  断言  $\alpha_i$  与  $\alpha_t$  没有显著差异.如果整个区间落在零的左边,则我们以概率  $1 - \alpha$  断言  $\alpha_i$  大于  $\alpha_t$ .相反,如果整个区间落在零的右边,则我们以概率  $1 - \alpha$  断言  $\alpha_i$  小于  $\alpha_t$ .

请注意,(6.2.10)是一个效应之差  $\alpha_i - \alpha_t$  的置信系数为  $1 - \alpha$  的置信区间.类似于单因素方差分析情形,不难推出, $m$  个效应之差  $\alpha_i - \alpha_t$  的置信系数为  $1 - \alpha$  的同时置信区间为

$$\left[ (\bar{y}_{i.} - \bar{y}_{t.}) - \hat{\sigma} \sqrt{\frac{2}{b} t_{(a-1)(b-1)} \left( \frac{\alpha}{2m} \right)}, (\bar{y}_{i.} - \bar{y}_{t.}) + \hat{\sigma} \sqrt{\frac{2}{b} t_{(a-1)(b-1)} \left( \frac{\alpha}{2m} \right)} \right]. \quad (6.2.11)$$

如果经过  $F_B$  检验,假设  $H_2$  被拒绝,用与上面完全类似的方法,可以建立  $\beta_j - \beta_q$  的置信系数为  $1 - \alpha$  的置信区间

$$\left[ (\bar{y}_{.j} - \bar{y}_{.q}) - \hat{\sigma} \sqrt{\frac{2}{a} t_{(a-1)(b-1)} \left( \frac{\alpha}{2} \right)}, (\bar{y}_{.j} - \bar{y}_{.q}) + \hat{\sigma} \sqrt{\frac{2}{a} t_{(a-1)(b-1)} \left( \frac{\alpha}{2} \right)} \right]. \quad (6.2.12)$$

类似于(6.2.11), $m$  个效应之差  $\beta_j - \beta_q$  的置信系数为  $1 - \alpha$  的同时置信区间可以将上式中的  $\alpha$  改为  $\frac{\alpha}{m}$  而得到.

## 二、关于交互效应的检验

在无交互效应时,对因素  $A, B$  各水平的每种组合只进行一次试验,即  $c = 1$ ,仍然可以进行统计分析.而当要考虑因素  $A, B$  间的交互作用  $A \times B$  时,在各水平组合下需要做重复试验.设每种组合下试验次数均为  $c (c > 1)$ .此时对应的统计模型就是(6.2.4).在这样的模型中,效应  $\alpha_i$  并不能反映水平  $A_i$  的优劣.这是因为在交互效应存在的情况下,因子水平  $A_i$  的优劣还与因子  $B$  的水平有关系.对不同的  $B_j, A_i$  的优劣也不相同.因此,对这样的模型,检验  $\alpha_1 = \dots = \alpha_a = 0$

与检验  $\beta_1 = \cdots = \beta_b = 0$  都是没有实际意义的. 然而一个重要的检验问题是交互效应是否存在的检验, 即检验

$$H_3: \gamma_{ij} = 0, \quad i = 1, \cdots, a, \quad j = 1, \cdots, b.$$

与前面方法类似, 引进

$$\bar{y} = \frac{1}{abc} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c y_{ijk},$$

$$\bar{y}_{j\cdot} = \frac{1}{c} \sum_{k=1}^c y_{ijk},$$

$$\bar{y}_{i\cdot\cdot} = \frac{1}{bc} \sum_{j=1}^b \sum_{k=1}^c y_{ijk},$$

$$\bar{y}_{\cdot j\cdot} = \frac{1}{ac} \sum_{i=1}^a \sum_{k=1}^c y_{ijk}.$$

作平方和分解, 得

$$\begin{aligned} SS_T &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y}_{j\cdot} + \bar{y}_{i\cdot\cdot} - \bar{y} + \bar{y}_{\cdot j\cdot} - \bar{y} + \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y}_{j\cdot})^2 + bc \sum_{i=1}^a (\bar{y}_{i\cdot\cdot} - \bar{y})^2 + \\ &\quad ac \sum_{j=1}^b (\bar{y}_{\cdot j\cdot} - \bar{y})^2 + c \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y})^2 \\ &= SS_E + SS_A + SS_B + SS_{A \times B}, \end{aligned}$$

其中

$$SS_E = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^c (y_{ijk} - \bar{y}_{j\cdot})^2,$$

$$SS_A = bc \sum_{i=1}^a (\bar{y}_{i\cdot\cdot} - \bar{y})^2,$$

$$SS_B = ac \sum_{j=1}^b (\bar{y}_{\cdot j\cdot} - \bar{y})^2,$$

$$SS_{A \times B} = c \sum_{i=1}^a \sum_{j=1}^b (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y})^2,$$

称  $SS_E$  为误差平方和,  $SS_A$  为因素 A 的平方和(或行间平方和),  $SS_B$  为因素 B 的平方和(或列间平方和),  $SS_{A \times B}$  为交互作用的平方和(或格间平方和).

与前面讨论的方法相类似,可以证明,当  $H_3$  成立时

$$F_{A \times B} = \frac{SS_{A \times B} / (a-1)(b-1)}{SS_E / ab(c-1)} \sim F_{(a-1)(b-1), ab(c-1)} \quad (6.2.13)$$

据此统计量,可以检验  $H_3$ .

如果经检验  $H_3$  被接受,我们就认为交互效应不存在.这时我们可以进一步检验因子  $A$  的各水平的效应是否有差异,也可以检验因子  $B$  的各水平效应是否有差异,这就回到了本节第一部分讨论的情形.

本节讨论的几种检验可以归纳成表 6.2.5.

表 6.2.5 关于交互效应的两因素方差分析表

方差来源	平方和	自由度	均 方	F 比
因素 A	$SS_A$	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_A = \frac{MS_A}{MS_E}$
因素 B	$SS_B$	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_B = \frac{MS_B}{MS_E}$
交互效应 (A × B)	$SS_{A \times B}$	$(a-1)(b-1)$	$MS_{A \times B} = \frac{SS_{A \times B}}{ab(c-1)}$	$F_{A \times B} = \frac{MS_{A \times B}}{MS_E}$
误差	$SS_E$	$ab(c-1)$		
总 和	$SS_T$	$abc-1$		

## § 6.3 正交试验设计与方差分析

前面介绍了一个或两个因素的试验,由于因素较少,我们可以对不同因素的所有可能的水平组合做试验,这叫做全面试验.当因素较多时,虽然理论上仍可采用前面的方法进行全面试验后再做相应的方差分析,但是在实际中有时会遇到试验次数太多的问题.如三因素四水平的问题,所有不同水平的组合有  $4^3 = 64$  种,在每一种组合下只进行一次试验,也需做 64 次.如果考虑更多的因素及水平,则全面试验的次数可能会大得惊人.因此在实际应用中,对于多因素做全面试验是不现实的.于是我们考虑是否可以选择其中一部分组合进行试验,这就用到试验设计方法选择合理的试验方案,使得试验次数不多,但也能得到比较满意的结果.

### 一、用正交表安排试验

正交表是一系列规格化的表格,每个表都有一个记号,如  $L_8(2^7)$ ,  $L_9(3^4)$



等,见表 6.3.1 和表 6.3.2.以  $L_9(3^4)$  为例, $L$  表示正交表,9 是正交表的行数,表示需要做的试验次数.4 是正交表的列数,表示最多可以安排的因素的个数.3 是因素水平数,表示此表可以安排三水平的试验.

表 6.3.1 正交表  $L_8(2^7)$ 

列号 试验号	1	2	3	4	5	6	7
1	1	1	1	2	2	1	2
2	2	1	2	2	1	1	1
3	1	2	2	2	2	2	1
4	2	2	1	2	1	2	2
5	1	1	2	1	1	2	2
6	2	1	1	1	2	2	1
7	1	2	1	1	1	1	1
8	2	2	2	1	2	1	2

表 6.3.2 正交表  $L_9(3^4)$ 

列号 试验号	1	2	3	4
1	1	1	3	2
2	2	1	1	1
3	3	1	2	3
4	1	2	2	1
5	2	2	3	3
6	3	2	1	2
7	1	3	1	3
8	2	3	2	2
9	3	3	3	1

从表 6.3.1 和 6.3.2 可见,  $L_9(3^4)$  有 9 行,4 列,表中由数字 1,2,3 组成.  $L_8(2^7)$  有 8 行,7 列,表中由数字 1,2 组成.

正交表的特点:

(a) 每列中数字出现的次数相同,如  $L_9(3^4)$  表每列中数字 1,2,3 均出现三次.  $L_8(2^7)$  表每列中数字 1,2 均出现四次.

(b) 任取两列数字的搭配是均衡的,如  $L_9(3^4)$  表里每两列中 (1,1), (1,2), ..., (3,3), 九种组合各出现一次.  $L_8(2^7)$  表里每两列中 (1,1), (1,2), (2,1) 和 (2,2) 各出现两次.

这种均衡性是一般正交表构造的特点,它使得根据正交表安排的试验,其试验结果具有很好的可比性,易于进行统计分析.

用正交表安排试验时,根据因素和水平个数的多少以及试验工作量的大小来考虑选用哪张正交表,下面举例说明.

**例 6.3.1** 为提高某种化学产品的转化率(%),考虑三个有关因素:反应温度  $A(^{\circ}\text{C})$ ,反应时间  $B(\text{min})$  和使用催化剂的含量  $C(\%)$ . 各因素选取三个水平,如表 6.3.3 所示.

表 6.3.3 转化率试验因素水平表

水 平	因 素	温度 A	时间 B	催化剂含量 C
1		80	90	5
2		85	120	6
3		90	150	7

如果做全面试验,则需  $3^3 = 27$  次.若用正交表  $L_9(3^4)$ ,仅做 9 次试验.将三个因素  $A, B, C$  分别放在  $L_9(3^4)$  表的任意三列上,如将  $A, B$  分别放在第 1, 2 列上,  $C$  放在第 4 列上.将表中  $A, B, C$  所在的三列上的数字 1, 2, 3 分别用相应的因素水平去替代,得 9 次试验方案.以上工作称为表头设计.再将 9 次试验结果转化率数据列于表上(见表 6.3.4),并在表上进行计算.

表 6.3.4 转化率试验的正交表

因素 试验号	反应温度 A	反应时间 B	催化剂含量 C	转化率
1	80(1)	90(1)	6(2)	31
2	85(2)	90(1)	5(1)	54
3	90(3)	90(1)	7(3)	38
4	80(1)	120(2)	5(1)	53
5	85(2)	120(2)	7(3)	49
6	90(3)	120(2)	6(2)	42
7	80(1)	150(3)	7(3)	57
8	85(2)	150(3)	6(2)	62
9	90(3)	150(3)	5(1)	64
$K_1$	141	123	171	
$K_2$	165	144	135	
$K_3$	144	183	144	
$k_1$	47	41	57	
$k_2$	55	48	45	
$k_3$	48	61	48	

表中各列的  $K_1, K_2, K_3$  值分别是对应因素第一, 二, 三水平的试验指标值之和.如因素  $A, K_1 = 31 + 53 + 57 = 141$ , 它是在九次试验中,所有  $A$  在第一水平(即  $80^\circ\text{C}$ )时试验所得转化率之和.类似地  $K_2 = 54 + 49 + 62 = 165$  和  $K_3 = 38 + 42 + 64 = 144$  分别是所有  $A$  在第二水平(即  $85^\circ\text{C}$ )和在第三水平(即  $90^\circ\text{C}$ )时试验所得转化率之和.各列的  $k_1, k_2, k_3$  分别是本列的  $K_1, K_2, K_3$  分别除以 3 得到的平均转化率.如对  $A$  有  $k_1 = 141/3 = 47, k_2 = 165/3 = 55, k_3 = 144/3 = 48$ .

在这个试验中,指标转化率是愈高愈好,经过直观比较各因素的  $k_1, k_2$  和  $k_3$ ,我们看出,对因素  $A$ ,最高平均转化率是 55%,它出现在第二水平  $A_2$ ,对因素  $B$ ,最高平均转化率是 61%,它出现在第三水平  $B_3$ .最后对因素  $C$ ,最高平均转化率是 57%它出现在第一水平  $C_1$ .因此,从现在的九次试验看,最好的试验

条件应是水平组合 $(A_2, B_3, C_1)$ ,也就是温度  $85^\circ\text{C}$ ,反应时间  $150\text{min}$ ,催化剂含量  $5\%$ .需要注意的是,这个试验水平的组合,是已经做过的几次试验中没有出现过的.它是否真正符合客观实际,还需要通过试验或生产实际来验证.

## 二、正交试验的方差分析

以上正交试验数据也可用线性模型来描述.以  $a_i, b_j, c_k$  分别表示  $A_i, B_j, C_k$  水平的效应,  $\mu$  为总平均,  $y_i$  为第  $i$  次试验结果,则例 6.3.1 可用下面的线性模型来描述:

$$\left\{ \begin{array}{l} y_1 = \mu + a_1 + b_1 + c_2 + \varepsilon_1, \\ y_2 = \mu + a_2 + b_1 + c_1 + \varepsilon_2, \\ y_3 = \mu + a_3 + b_1 + c_3 + \varepsilon_3, \\ y_4 = \mu + a_1 + b_2 + c_1 + \varepsilon_4, \\ y_5 = \mu + a_2 + b_2 + c_3 + \varepsilon_5, \\ y_6 = \mu + a_3 + b_2 + c_2 + \varepsilon_6, \\ y_7 = \mu + a_1 + b_3 + c_3 + \varepsilon_7, \\ y_8 = \mu + a_2 + b_3 + c_2 + \varepsilon_8, \\ y_9 = \mu + a_3 + b_3 + c_1 + \varepsilon_9, \\ \varepsilon_i \sim N(0, \sigma^2), \text{且相互独立}, i = 1, \dots, 9, \\ \sum_{i=1}^3 a_i = 0, \sum_{j=1}^3 b_j = 0, \sum_{k=1}^3 c_k = 0. \end{array} \right. \quad (6.3.1)$$

对此模型考虑如下三种假设的检验问题:

$$\begin{aligned} H_1: & a_1 = a_2 = a_3 = 0, \\ H_2: & b_1 = b_2 = b_3 = 0, \\ H_3: & c_1 = c_2 = c_3 = 0. \end{aligned} \quad (6.3.2)$$

若  $H_1$  成立,则说明因素  $A$  的三个水平对指标  $y$  的影响无显著差异.类似地,若  $H_2$ (或  $H_3$ )成立,则表示因素  $B$ (或  $C$ )的三个水平对指标  $y$  的影响无显著差异.导出检验统计量的方法与前两节相类似,下面讨论最一般的情况.

若在正交表中总的试验次数为  $n$ ,  $n$  次试验的结果分别记为  $y_1, y_2, \dots, y_n$ . 设因素有  $m$  个,每个因素取  $a$  个水平,每个水平作了  $r$  次试验,则  $n = ra$ ,总平方和为

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2. \quad (6.3.3)$$

记  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ , 为试验数据的总平均. 与前面两节讨论相类似, 可将  $SS_T$  分解为

$$SS_T = SS_E + SS_1 + SS_2 + \cdots + SS_m, \quad (6.3.4)$$

其中  $SS_E$  为误差平方和,  $SS_i$  为第  $i$  个因素的平方和. 下面给出分解式(6.3.4)中各平方和的具体计算公式.

先计算各因素平方和. 例如计算因素  $A$  的平方和, 设将因素  $A$  安排在正交表的第  $l$  列上, 此时将试验看作单因素  $A$  的试验. 用  $y_{ij}$  表示因素  $A$  的第  $j$  水平的第  $i$  个试验值,  $i=1, \dots, r, j=1, \dots, a$ , 则

$$\sum_{i=1}^r \sum_{j=1}^a y_{ij} = \sum_{i=1}^n y_i.$$

于是由单因素方差分析知

$$SS_A = r \sum_{j=1}^a (K_j^A - \bar{y})^2, \quad (6.3.5)$$

其中  $K_j^A$  表示因素  $A$  的第  $j$  水平的试验值之和(从正交表中的第  $l$  列上可以很容易算出).

在(6.3.5)中的  $A$  代表所考虑的任何因素. 如前所设共有  $m$  个这样的因素, 所以我们可得到  $m$  个平方和  $SS_1, SS_2, \dots, SS_m$ . 和(6.2.6)一样, 我们用下式计算  $SS_E$ :

$$SS_E = SS_T - \sum_{i=1}^m SS_i.$$

以上完成了平方和分解. 记  $f_T$  和  $f_E$  分别为总平方和及误差平方和的自由度, 而用  $f_i$  表示第  $i$  个因子平方和的自由度. 则各平方和的自由度分别为

$$f_T = \text{总试验次数} - 1 = n - 1,$$

$$f_i = \text{因素水平数} - 1 = a - 1, \quad i = 1, \dots, m,$$

$$f_E = f_T - \text{各因素自由度之和}$$

$$= f_T - \sum_{i=1}^m f_i = n - m(a - 1) - 1.$$

与前面两节的讨论相类似, 可以证明, 当第  $i$  个因素的各水平效应相等时

$$F_i = \frac{SS_i/f_i}{SS_E/f_E} \sim F_{a-1, n-m(a-1)-1}.$$

于是  $F_i$  可以用作检验第  $i$  个因素诸水平对试验指标  $y$  的影响有无显著差异的统计量. 统计分析的结果可列成方差分析表, 如表 6.3.5 所示.

表 6.3.5 正交试验设计的方差分析表

方差来源	平方和	自由度	均 方	F 比
因素 1	$SS_1$	$a - 1$	$MS_1 = SS_1 / (a - 1)$	$F_1 = MS_1 / MS_E$
因素 2	$SS_2$	$a - 1$	$MS_2 = SS_2 / (a - 1)$	$F_2 = MS_2 / MS_E$
⋮	⋮	⋮	⋮	⋮
因素 $m$	$SS_m$	$a - 1$	$MS_m = SS_m / (a - 1)$	$F_m = MS_m / MS_E$
误 差	$SS_E$	$n - m(a - 1) - 1$	$MS_E = SS_E / [n - m(a - 1) - 1]$	
总 和	$SS_T$	$n - 1$		

例6.3.2 (续例 6.3.1)对正交试验进行方差分析.

总平方和

$$SS_T = \sum_{i=1}^9 y_i - 9 \times y,$$

$$\sum_{i=1}^9 y_i^2 = 31^2 + 54^2 + \cdots + 64^2 = 23\,484,$$

$$y = \frac{1}{9}(31 + 54 + \cdots + 64) = \frac{1}{9} \times 450 = 50.$$

各平方和分别为

$$SS_T = 23484 - 9 \times 50 = 984,$$

$$SS_A = \frac{1}{3}(141^2 + 165^2 + 144^2) - \frac{1}{9} \times 450^2 = 114,$$

$$SS_B = \frac{1}{3}(123^2 + 144^2 + 183^2) - \frac{1}{9} \times 450^2 = 618,$$

$$SS_C = \frac{1}{3}(171^2 + 135^2 + 144^2) - \frac{1}{9} \times 450^2 = 234,$$

$$SS_E = 984 - 114 - 618 - 234 = 18.$$

各平方和相应的自由度分别为

$$f_T = 9 - 1 = 8, \quad f_A = f_B = f_C = 3 - 1 = 2,$$

$$f_E = 9 - 3(3 - 1) - 1 = 2.$$

将计算结果列于方差分析表中,如表 6.3.6 所示.

表 6.3.6 转化率试验方差分析表

方差来源	平方和	自由度	均方	F 比
A	114	2	57	6.33
B	618	2	309	34.33
C	234	2	117	13.00
误差	18	2	9	
总和	984	8		

查  $F$  分布表得临界值,  $F_{2,2}(0.05) = 19, F_{2,2}(0.1) = 9$ . 可见因素  $B, C$  的各水平对指标值  $y$  的影响有显著差异, 而因素  $A$  的各水平对  $y$  的影响无显著差异.

### 习题六

6.1 对单因素方差分析模型(6.1.3), 证明因子效应  $\alpha_i$  满足  $\sum_{i=1}^a n_i \alpha_i = 0$ .

6.2 对单因素方差分析模型(6.1.3), 证明下面两个假设是等价的:

$$H_0: \mu_1 = \cdots = \mu_a,$$

$$H'_0: \alpha_1 = \cdots = \alpha_a = 0.$$

6.3 利用正态分布的性质(推论2.3.3)证明(6.1.13).

6.4 现有工厂  $A, B, C$  生产同一型号的电池, 为比较其质量, 从各厂的产品中随机抽取6只电池, 经测试得其寿命(h)如下

生产厂	电池寿命					
A	40	48	38	42	45	46
B	26	34	30	28	32	33
C	39	40	48	50	50	52

(1) 将数据表成方差分析模型(6.1.2)的形式;

(2) 在显著性水平  $\alpha = 0.05$  下检验三厂生产的电池的平均寿命有无显著差异? 列出方差分析表;

(3) 记  $\mu_A, \mu_B$  和  $\mu_C$  分别为三厂生产的电池的平均寿命, 写出均值之差  $\mu_A - \mu_B, \mu_A - \mu_C, \mu_B - \mu_C$  的 95% 同时置信区间.

6.5 对两因素方差分析模型(6.2.5), 证明其最后两个等式:  $\sum_{i=1}^a \alpha_i = 0, \sum_{j=1}^b \beta_j = 0$ .

6.6 设四名工人  $W_i, i = 1, 2, 3, 4$  分别操作机床  $A_1, A_2, A_3$  各一天, 生产同样产品, 其日产量如下表(单位: 件).

机 床	工 人			
	$W_1$	$W_2$	$W_3$	$W_4$
$A_1$	50	47	47	53
$A_2$	63	54	57	58
$A_3$	52	42	41	48

(1) 将数据表成方差分析模型(6.2.1)的形式;

(2) 在显著性水平  $\alpha=0.05$  下, 检验四名工人的日加工量有无显著差异?

(3) 在显著性水平  $\alpha=0.05$  下, 检验三台机床对日加工量有无显著差异?

6.7 利用正态分布的性质(推论2.3.3)证明(6.2.9).

6.8 对两因素方差分析模型(6.2.5), 导出效应差  $\alpha_i - \alpha_j$  的置信系数为  $1 - \alpha$  的置信区间(6.2.10).

6.9 下表给出某个化学试验在三种浓度(%), 四种温度( $^{\circ}\text{C}$ )下的得率(%)的观测值:

浓 度	温 度			
	10	24	38	52
2	14, 10	11, 11	13, 19	10, 12
4	9, 7	10, 8	7, 11	6, 10
6	5, 11	13, 14	12, 13	14, 10

(1) 将数据表成方差分析模型(6.2.1)的形式;

(2) 在显著性水平  $\alpha=0.05$  下检验浓度对产品得率的影响有无显著差异?

(3) 在显著性水平  $\alpha=0.05$  下检验温度对产品得率的影响有无显著差异?

## \* 第七章 其它线性回归模型

### § 7.1 引言

在前面几章中,我们对线性回归模型的参数估计、假设检验以及统计性质的讨论,都是基于如下的基本假设:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{ip-1} + e_i, \quad i = 1, \cdots, n, \quad (7.1.1)$$

其中误差项  $e_i, i = 1, \cdots, n$  满足下面的 Gauss-Markov 假设:

$$E(e_i) = 0, \text{Var}(e_i) = \sigma^2, \text{Cov}(e_i, e_j) = 0, i \neq j, \quad (7.1.2)$$

我们称具有条件(7.1.2)的回归模型(7.1.1)为经典线性回归模型.但是在实际中遇到的问题往往比较复杂,有时难于满足(7.1.2)的假定.比如

(1)  $\text{Var}(e_i) = \sigma^2$  (等方差性)不成立,即误差项具有不等方差,统计上称作异方差性.

(2)  $\text{Cov}(e_i, e_j) = 0 (i \neq j)$  不成立,即误差项存在自相关性.

(3) 有时我们考虑的实际问题涉及  $M$  个回归方程,具有如下形式

$$y_{it} = \beta_{0i} + \beta_{1i} x_{it} + e_{it} \quad (7.1.3)$$

$$i = 1, \cdots, M; \quad t = 1, \cdots, n,$$

其中,  $y_{it}$  表示第  $i$  个回归方程中的因变量的观测值,  $x_{it}$  表示第  $i$  个回归方程中自变量的观测值,  $e_{it}$  为第  $i$  个回归方程的误差项.记

$$\mathbf{y}_i = (y_{1i}, \cdots, y_{ni})', \quad \boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i})', \quad \mathbf{e}_i = (e_{1i}, \cdots, e_{ni})',$$

$$\mathbf{X}_i = \begin{bmatrix} 1 & x_{1i} \\ \vdots & \vdots \\ 1 & x_{ni} \end{bmatrix}, \quad i = 1, \cdots, M,$$

则(7.1.3)式可写成

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & & & 0 \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ & & & 0 \\ & & & & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_M \end{bmatrix}. \quad (7.1.4)$$



对(7.1.4)式中的误差项作如下假定

$$E(e_i) = \mathbf{0} \quad E(e_i e_j') = \sigma_{ij} \mathbf{I}_n, \quad i, j = 1, \dots, M, \quad (7.1.5)$$

我们称满足(7.1.3)和(7.1.5)的回归方程组为半相依回归模型(seemingly unrelated regression model). 对半相依回归模型而言,它也不满足经典回归模型的假定.

上述三种情况都是针对模型(7.1.1)的误差结构或者是模型的结构而言的,把它们的每一个作为模型(7.1.1)的限制条件,都可视为经典回归模型的推广.我们还可以考虑对自变量  $x_{ij}$  和因变量  $y_i$  的限制条件上推广模型(7.1.1).

(4) 由于人为的,或仪器的,或测量手段的原因,自变量  $x_i$  ( $i=1, \dots, p-1$ ) 和因变量  $y_i$  的观测值难于精确地获得,而是含有随机误差.以一元回归模型为例,可表述为

$$\begin{cases} y_i = \beta_0 + \beta_1 x_i + e_i, \\ X_i = x_i + \varepsilon_i, \end{cases} \quad (7.1.6)$$

其中  $x_i$  不能被直接观测到,而只能观测到  $X_i$ .  $e_i$  和  $\varepsilon_i$  是随机误差,通常假定  $(X_i, e_i, \varepsilon_i)$ ,  $i=1, \dots, n$  是独立随机向量,服从正态分布  $N((x_i, 0, 0)', \text{diag}(\sigma_x^2, \sigma_e^2, \sigma_\varepsilon^2))$ . 统计上称这类模型为误差变量模型(Errors-in-Variable Model).

(5) 模型(7.1.1)对因变量的假定之一是:若记  $\mu_i = E(y_i)$ , 则  $\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1}$ . 如果存在一个严格增的可微函数  $g$ , 使得

$$\begin{cases} g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1}, \\ \mu_i = E(y_i), \quad i = 1, \dots, n, \end{cases} \quad (7.1.7)$$

则(7.1.7)式也是对模型(7.1.1)的一种推广. 统计上称这类模型为广义线性模型. 当然,它只是广义线性模型中的一类模型. 如果  $g(p) = \ln[p/(1-p)]$ ,  $0 < p < 1$ , 则(7.1.7)即为应用很广泛的一类模型——logistic 模型.

本书不打算讨论(7.1.6)和(7.1.7)这两类模型的统计分析. 有兴趣的读者可参阅 Fuller(1987)、陈希孺和王松桂(1987). 本章将介绍一些常用的线性回归模型的建模方法及相应的统计性质.

## § 7.2 具有异方差误差的线性回归模型

考虑如下线性回归模型:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{ip-1} + e_i, \quad (7.2.1)$$

当误差项  $e_i$  ( $i=1, \dots, n$ ) 满足

$$E(e_i) = 0, \text{Cov}(e_i, e_j) = 0, (i \neq j), \text{Var}(e_i) = \sigma_i^2 > 0, \quad (7.2.2)$$

且  $\sigma_1^2, \dots, \sigma_n^2$  不全相等, 我们称模型(7.2.1)具有异方差性. 在第三章中, 我们已经讨论过这种情形下, 模型(7.2.1)参数的估计问题. 利用广义最小二乘估计, 我们可以获得其参数  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$  的估计. 但是在处理实际问题时, 我们常常要先回答模型是否具有异方差性这个问题. 只有知道了模型是否具有异方差性, 才能决定使用什么统计方法来估计模型(7.2.1)中的未知参数  $\boldsymbol{\beta}$ .

记  $\mathbf{y} = (y_1, \dots, y_n)'$ ,  $\mathbf{e} = (e_1, \dots, e_n)'$ .

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p-1} \\ 1 & x_{21} & \cdots & x_{2p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np-1} \end{pmatrix}, \mathbf{\Lambda} = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ 0 & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix},$$

把模型(7.2.1)写成矩阵形式, 有

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \quad (7.2.3)$$

模型(7.2.3)是否具有异方差性可表述为如下统计假设检验:

$$\begin{cases} H_0: \text{Cov}(\mathbf{e}, \mathbf{e}) = \sigma^2 \mathbf{I}_n, \\ H_1: \text{Cov}(\mathbf{e}, \mathbf{e}) = \mathbf{\Lambda}. \end{cases} \quad (7.2.4)$$

在处理统计假设(7.2.4)时, 我们常假定异方差  $\sigma_i^2$  具有如下形式:

$$\sigma_i^2 = \sigma(\alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_i), \quad (7.2.5)$$

其中  $\sigma(\cdot)$  是一个可微非负函数,  $\sigma(\alpha_0) > 0$ , 且  $\sigma'(\alpha_0) > 0$  (这里  $\sigma'(\alpha)$  表示  $\sigma(\alpha)$  的导数),  $\alpha_0$  和  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)'$  为待估参数,  $\mathbf{Z}_i = (z_{i1}, \dots, z_{iq})'$  为影响  $e_i$  方差变化的因素, 它可以是  $x_{ij}$ ,  $i, j = 1, \dots, p-1$  的函数, 也可以是别的变量. 如果取  $\sigma(\cdot)$  为线性函数, 即

$$\sigma_i^2 = \alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_i,$$

当  $\mathbf{Z}_i = (x_{i1}^2, \dots, x_{i,p-1}^2)'$  时, (7.2.5) 式成为

$$\sigma_i^2 = \alpha_0 + \alpha_1 x_{i1}^2 + \cdots + \alpha_{p-1} x_{i,p-1}^2. \quad (7.2.6)$$

当  $\mathbf{Z}_i = (|x_{i1}|, \dots, |x_{i,p-1}|)'$  时, (7.2.5) 式成为

$$\sigma_i^2 = \alpha_0 + \alpha_1 |x_{i1}| + \cdots + \alpha_{p-1} |x_{i,p-1}|. \quad (7.2.7)$$

如果

$$\sigma_i^2 = \exp\{\alpha_0 + \boldsymbol{\alpha}'\mathbf{Z}_i\},$$

则当  $\mathbf{Z}_i = (x_{i1}, \dots, x_{i,p-1})'$  时, (7.2.5) 式成为

$$\sigma_i^2 = \exp\left\{\alpha_0 + \sum_{j=1}^p \alpha_j x_{ij}\right\}. \quad (7.2.8)$$

这时, 对具有形如(7.2.6)—(7.2.8)式的异方差结构的线性回归模型(7.2.1), 假设检验(7.2.4)又可表述为

$$\begin{cases} H_0: \alpha_i = 0, & i = 1, \dots, q; \\ H_1: \alpha_1, \dots, \alpha_q \text{ 不全为零.} \end{cases} \quad (7.2.9)$$

以下我们在  $e_i$  服从正态分布的假设下, 针对假设检验问题(7.2.9)来讨论如何构造检验统计量.

我们的统计量的构造主要是基于拉格朗日乘子检验原理. 在统计分析中, 对参数进行假设检验, 构造检验统计量的方法通常有三种: 一种是以参数估计的渐近正态性为基础的 Wald 方法, 例如常见的  $t$  检验,  $F$  检验等都属于这类方法. 另一种是似然比方法. 第三种方法就是下面将要介绍的拉格朗日乘子法, 它是将零假设条件看成一个约束条件, 通过对有约束的极大似然函数的一阶偏导数进行检验, 对参数假设作出判断. 在某些模型检验中, 拉格朗日乘子检验方法不仅能带来计算上的方便, 而且所构造的检验统计量与上述两种方法所构造的统计量具有相同的渐近性质.

拉格朗日乘子检验.

设  $x_1, \dots, x_n$  是来自某个分布族的样本. 该分布族中的分布除参数  $\theta = (\theta_1, \dots, \theta_p)'$  之外都是已知的. 记  $L(\theta)$  为样本  $(x_1, \dots, x_n)$  的对数似然函数. 考虑如下关于  $\theta$  的假设检验问题:

$$H_0: h_j(\theta) = 0, \quad j = 1, \dots, k, k < p. \quad (7.2.10)$$

将上述零假设中的  $k$  个方程视为关于参数  $\theta$  的  $k$  个约束方程, 利用拉格朗日乘子法, 求  $L(\theta)$  在约束条件(7.2.10)下的极大似然估计, 等价于求如下函数的极大值问题:

$$f(\theta, \lambda_1, \dots, \lambda_k) = L(\theta) + \sum_{j=1}^k \lambda_j h_j(\theta). \quad (7.2.11)$$

关于  $\theta, \lambda_1, \dots, \lambda_k$ , 求函数(7.2.11)的一阶偏导数, 并令其为零, 得

$$\frac{\partial f}{\partial \theta_i} = \frac{\partial L(\theta)}{\partial \theta_i} + \sum_{j=1}^k \lambda_j \frac{\partial h_j(\theta)}{\partial \theta_i} = 0, \quad i = 1, \dots, p. \quad (7.2.12)$$

$$\frac{\partial f}{\partial \lambda_i} = h_i(\theta) = 0, \quad i = 1, \dots, k. \quad (7.2.13)$$

令

$$\mathbf{D}(\theta) = \begin{bmatrix} \frac{\partial L(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial L(\theta)}{\partial \theta_p} \end{bmatrix}, \quad \mathbf{H}(\theta) = \begin{bmatrix} \frac{\partial h_1(\theta)}{\partial \theta_1} & \dots & \frac{\partial h_1(\theta)}{\partial \theta_p} \\ \vdots & & \vdots \\ \frac{\partial h_k(\theta)}{\partial \theta_1} & \dots & \frac{\partial h_k(\theta)}{\partial \theta_p} \end{bmatrix},$$

$\lambda = (\lambda_1, \dots, \lambda_k)'$ . 则(7.2.12)和(7.2.13)式可写成如下矩阵方程:

$$\begin{aligned} D(\boldsymbol{\theta}) + H(\boldsymbol{\theta})'\lambda &= \mathbf{0}, \\ h_i(\boldsymbol{\theta}) &= 0, \quad i = 1, \dots, k. \end{aligned} \quad (7.2.14)$$

若零假设  $H_0$  成立, 则  $\boldsymbol{\theta}$  的有约束极大似然估计  $\hat{\boldsymbol{\theta}}_{H_0}$  近似于  $\boldsymbol{\theta}$  的无约束极大似然估计  $\hat{\boldsymbol{\theta}}_{ML}$ , 从而  $D(\hat{\boldsymbol{\theta}}_{H_0})$  将近似于零向量. 在适当的条件下, 可以证明.

$$C_n^{-\frac{1}{2}} D(\hat{\boldsymbol{\theta}}_{H_0}) \xrightarrow{d} N(0, W(\boldsymbol{\theta})),$$

式中

$$W(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} C_n^{-\frac{1}{2}} I(\boldsymbol{\theta}) C_n^{-\frac{1}{2}},$$

其中  $I(\boldsymbol{\theta}) = E(-\partial^2 L(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}')$  是 Fisher 信息矩阵,  $C_n$  是适当选取的矩阵序列. 根据上述结论, 我们可以构造如下的拉格朗日检验统计量:

$$LM = D(\hat{\boldsymbol{\theta}}_{H_0})' I(\hat{\boldsymbol{\theta}}_{H_0})^{-1} D(\hat{\boldsymbol{\theta}}_{H_0}). \quad (7.2.15)$$

可以证明, 当  $H_0$  成立时, 在通常的极大似然估计渐近理论条件下,

$$LM \xrightarrow{d} \chi_k^2, \text{ 当 } n \rightarrow \infty, \quad (7.2.16)$$

式中  $\chi_k^2$  是自由度为  $k$  的  $\chi^2$  分布. 从而对给定的显著性水平  $\alpha$ , 不难确定  $H_0$  的否定域.

在存在多余参数的情形下, 记多余参数为  $\mathbf{v} = (v_1, \dots, v_q)'$ , 需要修正表达式 (7.2.15). 令  $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}', \mathbf{v}')'$ , 考虑如下假设检验问题:

$$H_0: h_j(\tilde{\boldsymbol{\theta}}) = 0, \quad j = 1, \dots, k.$$

按前面所述的方法求得 (7.2.14) 式, 根据  $\boldsymbol{\theta}$  和  $\mathbf{v}$  的维数, 把  $D(\tilde{\boldsymbol{\theta}})$  和  $I(\tilde{\boldsymbol{\theta}})$  写成分块向量和分块矩阵的形式, 即

$$D(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} D_1(\tilde{\boldsymbol{\theta}}) \\ D_2(\tilde{\boldsymbol{\theta}}) \end{bmatrix}, \quad I(\tilde{\boldsymbol{\theta}}) = \begin{bmatrix} I_{11}(\tilde{\boldsymbol{\theta}}) & I_{12}(\tilde{\boldsymbol{\theta}}) \\ I_{21}(\tilde{\boldsymbol{\theta}}) & I_{22}(\tilde{\boldsymbol{\theta}}) \end{bmatrix},$$

由 (7.2.12) 和 (7.2.14) 式, 可知  $D_2(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ , 从而由 (7.2.15) 式和分块矩阵求逆公式, 当零假设  $H_0$  成立时,  $LM$  统计量 (7.2.15) 式为

$$LM_1 = D_1(\tilde{\boldsymbol{\theta}}_{H_0})' [I_{11}(\tilde{\boldsymbol{\theta}}_{H_0}) - I_{12}(\tilde{\boldsymbol{\theta}}_{H_0})' I_{22}^{-1}(\tilde{\boldsymbol{\theta}}_{H_0}) I_{21}(\tilde{\boldsymbol{\theta}}_{H_0})]^{-1} D_1(\tilde{\boldsymbol{\theta}}_{H_0}). \quad (7.2.17)$$

式中  $\tilde{\boldsymbol{\theta}}_{H_0}$  为参数  $\tilde{\boldsymbol{\theta}}$  的有约束极大似然估计. 同样, 在适当条件下, 对 (7.2.17) 式, 有 (7.2.16) 式成立.

现在我们来讨论 (7.2.9) 式提出的假设检验问题.

记  $\boldsymbol{\theta} = (\alpha_0, \alpha_1, \dots, \alpha_q)'$ ,  $\mathbf{v} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$ , 即把回归系数  $\beta_0, \beta_1, \dots, \beta_{p-1}$  看作是多余参数, 这是因为我们此时只关心  $\alpha_i$  的统计行为. 于是在备择假设  $H_1$  下,  $\mathbf{y} = (y_1, \dots, y_n)'$  的对数似然函数为

$$L(\hat{\theta}) = m_0 - \frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^n \frac{e_i^2}{\sigma_i^2}, \quad (7.2.18)$$

式中  $m_0$  为常数, 对(7.2.18)式关于参数  $\theta$  和  $\mathbf{v}$  求偏导数可得

$$\begin{aligned} \frac{\partial L(\hat{\theta})}{\partial \theta} &= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} + \frac{1}{2} \sum_{i=1}^n \frac{e_i^2}{\sigma_i^4} \frac{\partial \sigma_i^2}{\partial \theta} \\ &= \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \left[ \frac{e_i^2}{\sigma_i^2} - 1 \right], \end{aligned} \quad (7.2.19)$$

$$\frac{\partial^2 L(\hat{\theta})}{\partial \theta \partial \theta'} = \frac{1}{2} \sum_{i=1}^n \left\{ \left[ \frac{e_i^2}{\sigma_i^2} - 1 \right] \frac{\partial}{\partial \theta'} \left[ \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \right] - \frac{e_i^2}{\sigma_i^6} \frac{\partial \sigma_i^2}{\partial \theta} \frac{\partial \sigma_i^2}{\partial \theta'} \right\}. \quad (7.2.20)$$

$$\begin{aligned} \frac{\partial^2 L(\hat{\theta})}{\partial \theta \partial \mathbf{v}'} &= \frac{1}{2} \sum_{i=1}^n \left\{ \left[ \frac{e_i^2}{\sigma_i^2} - 1 \right] \frac{\partial}{\partial \mathbf{v}'} \left[ \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \right] \right. \\ &\quad \left. + \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \left[ \frac{2e_i}{\sigma_i^2} \frac{\partial e_i}{\partial \mathbf{v}'} - \frac{e_i^2}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \mathbf{v}'} \right] \right\}. \end{aligned} \quad (7.2.21)$$

注意到  $E(e_i) = 0$ ,  $E(e_i^2) = \sigma_i^2$ ,  $\sigma_i^2$  仅是  $\theta$  的函数, 我们有

$$E \left( - \frac{\partial^2 L(\hat{\theta})}{\partial \theta \partial \theta'} \right) = \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^4} \frac{\partial \sigma_i^2}{\partial \theta} \frac{\partial \sigma_i^2}{\partial \theta'}, \quad (7.2.22)$$

$$E \left( - \frac{\partial^2 L(\hat{\theta})}{\partial \theta \partial \mathbf{v}'} \right) = 0. \quad (7.2.23)$$

为简明起见, 我们现以形如(7.2.6)式的异方差结构来讨论对假设  $H_0$  的 LM 检验.

记  $\mathbf{v}_i = (1, x_{i1}^2, \dots, x_{iq}^2)'$ ,  $g_i = \frac{e_i^2}{\sigma_i^2} - 1$ ,  $i = 1, \dots, n$ , 由于在(7.2.6)式的异方差结构下,

$$\frac{\partial \sigma_i^2}{\partial \theta} = \mathbf{v}_i,$$

于是由(7.2.19)和(7.2.22)两式有

$$\mathbf{D}(\hat{\theta}) = \frac{\partial L(\hat{\theta})}{\partial \theta} = \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^2} \mathbf{v}_i g_i, \quad (7.2.24)$$

$$\mathbf{I}_{11}(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma_i^4} \mathbf{v}_i \mathbf{v}_i'. \quad (7.2.25)$$

这样我们可按如下步骤实施对假设(7.2.9)中  $H_0$  的检验.

步骤1 在  $H_0$  下, 对参数  $\mathbf{v} = (\beta_0, \beta_1, \dots, \beta_{p-1})'$  进行极大似然估计, 即求回归模型(7.2.1)的极大似然估计. 记为  $\hat{\mathbf{v}}_{ML}$ , 并定义

$$\hat{e}_i = y_i - \hat{\beta}_{0ML} - \hat{\beta}_{1ML}x_{i1} - \cdots - \hat{\beta}_{(p-1)ML}x_{i(p-1)},$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{e}_i^2.$$

步骤 2 在  $H_0$  下, 注意到  $\sigma_i^2$  是一个常数, 记其为  $\sigma^2$ , 它的极大似然估计由步骤 1 给出. 令

$$\hat{g}_i = (\hat{e}_i^2 / \hat{\sigma}^2) - 1, \quad i = 1, \dots, n.$$

$$\mathbf{V}_n = (v_1, \dots, v_n), \quad \mathbf{G}_n = (\hat{g}_1, \dots, \hat{g}_n)',$$

则由(7.2.24)和(7.2.25)两式有

$$\mathbf{D}(\tilde{\boldsymbol{\theta}}_{H_0}) = \frac{1}{2} \sum_{i=1}^n \frac{1}{\hat{\sigma}^2} v_i \hat{g}_i = \frac{1}{2\hat{\sigma}^2} \mathbf{V}_n' \mathbf{G}_n,$$

$$\mathbf{I}_{11}(\tilde{\boldsymbol{\theta}}_{H_0}) = \frac{1}{2} \sum_{i=1}^n \frac{1}{\hat{\sigma}^2} v_i v_i' = \frac{1}{2\hat{\sigma}^2} \mathbf{V}_n' \mathbf{V}_n,$$

其中  $\tilde{\boldsymbol{\theta}}_{H_0} = (\boldsymbol{\theta}', v_{ML}')'$ .

步骤 3 由(7.2.17)式及步骤 2 中两式, 可得所求的 LM 检验统计量

$$\begin{aligned} LM &= \mathbf{D}(\tilde{\boldsymbol{\theta}}_{H_0})' \mathbf{I}_{11}^{-1}(\tilde{\boldsymbol{\theta}}_{H_0}) \mathbf{D}(\tilde{\boldsymbol{\theta}}_{H_0}) \\ &= \frac{1}{2} \mathbf{G}_n' \mathbf{V}_n (\mathbf{V}_n' \mathbf{V}_n)^{-1} \mathbf{V}_n' \mathbf{G}_n. \end{aligned} \quad (7.2.26)$$

由(7.2.16)和(7.2.17)两式知, 当  $n \rightarrow \infty$  时

$$LM \xrightarrow{d} \chi_{q+1}' \quad (7.2.27)$$

于是, 给定显著性水平, 我们就可以根据(7.2.27)式确定  $H_0$  下的否定域.

对形如(7.2.7)和(7.2.8)式的异方差结构的 LM 检验统计量, 可用同样的方法得出, 在此从略.

下面通过两个例子来说明 LM 检验的功效.

**例 7.2.1** 考虑如下一元线性回归模型:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad (7.2.28)$$

其中  $x_i$  由如下方式生成:

$$\text{I: } \begin{cases} x_i = \xi_i, \\ \xi_i \sim U(1, 31), \quad i = 1, \dots, n. \end{cases}$$

$$\text{II: } \begin{cases} x_i = \epsilon_i, \\ \ln(\epsilon_i) \sim N(3, 1), \quad i = 1, \dots, n. \end{cases}$$

式中  $U(1, 31)$  为区间  $(1, 31)$  上的均匀分布,  $e_i$  ( $i = 1, \dots, n$ ) 是独立正态  $N(0, \sigma_i^2)$  随机误差, 且  $\sigma_i^2$  具有形式

$$(i) \sigma_i^2 = \sigma^2, \quad i = 1, \dots, n,$$

$$(ii) \sigma_i^2 = \sigma^2 x_i^2, \quad i = 1, \dots, n,$$

$$(iii) \sigma_i^2 = \sigma^2 x_i, \quad i = 1, \dots, n.$$

选择样本容量  $n = 20, 30, 40, 50, 60, 70, 80$ , 每种不同的样本容量选择, 模拟重复 1 000 次. 由于异方差结构中仅含有一个未知参数, 故按 LM 检验步骤中 (7.2.27) 式知: 在模型 (7.2.28) 下, LM 检验统计量的渐近分布为  $\chi_1^2$  分布. 对显著性水平 5% 和 10%, 每次模拟计算相应的 LM 检验统计量的值, 并与给定的显著性水平下的临界值相比, 若该值大于临界值, 则拒绝零假设  $H_0$ , 记录下 1 000 次模拟中拒绝  $H_0$  的次数的百分比.

模拟过程如下:

(1) 对给定的  $n$ , 用计算机产生三组随机数  $e_i \sim N(0, 1)$ ,  $\xi'_i \sim U(0, 1)$  和  $\varepsilon'_i \sim N(0, 1)$ ,  $i = 1, \dots, n$ . 这一点很容易做到, 因为计算机的系统软件配置都装备了  $N(0, 1)$  和  $U(0, 1)$  随机数的发生器.

(2) 生成  $x_i$ . 对模型 I, 由变换  $\xi_i = 15\xi'_i + \frac{17}{2}$  可知  $\xi_i \sim U(1, 31)$ . 于是

$$x_i = 15\xi'_i + \frac{17}{2}, \quad i = 1, \dots, n.$$

对模型 II, 由变换  $\varepsilon_i = \exp(\varepsilon'_i + 3)$  知,  $\varepsilon_i$  服从对数正态分布, 且  $\ln(\varepsilon_i) \sim N(3, 1)$ . 故

$$x_i = \exp(\varepsilon'_i + 3), \quad i = 1, \dots, n.$$

(3) 由  $x_i$  和  $e_i$  ( $i = 1, \dots, n$ ) 生成  $y_i$ .

按照模型

$$y_i = \beta_0 + \beta_1 x_i + \sigma e_i,$$

$$y_i = \beta_0 + \beta_1 x_i + (\sigma x_i) e_i,$$

$$y_i = \beta_0 + \beta_1 x_i + (\sigma \sqrt{x_i}) e_i,$$

即可生成所需要的模拟数据  $y_i$ , 式中的  $\beta_0, \beta_1$  和  $\sigma$  要事先给定, 比如  $\beta_0 = 1$ ,  $\beta_1 = 0.5, \sigma = 2$ .

(4) 用模拟数据  $(y_i, x_i)$ ,  $i = 1, \dots, n$  计算所要的 LM 检验统计量的值, 并按 LM 检验原理给出判断, 记录下拒绝与否, 拒绝  $H_0$  为 1, 接受  $H_0$  为 0.

(5) 将上述 (1) ~ (4) 步骤独立地重复 1 000 次, 其中拒绝  $H_0$  的次数记为  $n_1$ , 则  $n_1/n$  即为 1 000 次重复模拟中, 检验的结果.

按以上五步, 对  $n = 20, 30, \dots, 80$ , 进行数值模拟, 结果列于表 7.2.1—7.2.3.

表 7.2.1 异方差结构(i)的模拟结果

样本容量	I		II	
	0.10	0.05	0.10	0.05
20	12.6	3.5	6.9	3.9
30	8.2	3.9	7.5	4.3
40	10.2	5.3	7.7	4.1
50	9.4	4.1	9.7	5.1
60	8.7	3.9	11.0	5.6
70	9.3	4.6	9.8	5.5
80	9.1	3.7	9.0	3.9

表 7.2.2 异方差结构(ii)的模拟结果

样本容量	I		II	
	0.10	0.05	0.10	0.05
20	55.0	33.6	95.4	93.8
30	98.7	93.8	99.9	99.7
40	99.8	99.4	100.0	100.0
50	100.0	99.9	100.0	100.0
60	99.6	98.2	100.0	100.0
70	100.0	100.0	100.0	100.0
80	100.0	100.0	100.0	100.0

表 7.2.3 异方差结构(iii)的模拟结果

样本容量	I		II	
	0.10	0.05	0.10	0.05
20	21.9	7.8	69.8	60.2
30	73.7	48.9	89.3	83.9
40	81.9	68.3	94.8	90.7
50	92.4	82.8	98.2	96.1
60	98.1	95.1	99.8	98.5
70	99.7	98.5	99.4	98.2
80	100.0	99.8	99.8	99.8

由表 7.2.1—7.2.3 可见;



(i) 如果零假设  $H_0$  成立, 即不存在异方差, 表 7.2.1 的结果表明 1 000 次模拟中,  $H_0$  被拒绝的百分比都接近给定的显著性水平.

(ii) 对异方差结构(ii), 表 7.2.2 的结果表明, 当样本容量  $n \geq 30$  时, 1 000 次模拟中,  $H_0$  被拒绝的百分比都大于或接近(仅 I 的 0.05 显著性水平情形)  $1 - \delta$  (给定显著性水平).

(iii) 对异方差结构(iii), 当样本容量  $n < 40$  时, 检验的功效较差, 当样本容量  $\geq 40$ ,  $H_0$  被拒绝的百分比大多数接近或大于  $1 - \delta$ .

因此, LM 检验对  $n \geq 40$  是有效的.

例 7.2.2 考虑如下二元线性回归模型:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i,$$

其中  $e_i$  服从独立  $N(0, \sigma_i^2)$ , 异方差结构具有形式

$$\sigma_i^2 = \exp\{\alpha_0 + \alpha_1 \ln(x_{i1}) + \alpha_2 \ln(x_{i2})\},$$

式中  $x_{i1}$  和  $x_{i2}$  满足

$$x_{i1} = \xi_i \sim U(1, 31), x_{i2} = \epsilon_i, \ln(\epsilon_i) \sim N(3, 1).$$

对  $\alpha_1$  和  $\alpha_2$  的如下取值:

(i)  $\alpha_1 = \alpha_2 = 0$ , 即  $\sigma_i^2 = \sigma^2$ ,

(ii)  $\alpha_1 = 2, \alpha_2 = 0$ , 即  $\sigma_i^2 = \sigma^2 x_{i1}^2$ ,

(iii)  $\alpha_1 = 0, \alpha_2 = 2$ , 即  $\sigma_i^2 = \sigma^2 x_{i2}^2$ ,

(iv)  $\alpha_1 = \alpha_2 = 1$ , 即  $\sigma_i^2 = \sigma^2 x_{i1} x_{i2}$ ,

同例 7.2.1 的模拟方式一样, 并把结果列于表 7.2.4 和 7.2.5 中.

表 7.2.4 显著性水平 5% 下模拟结果(1 000 次重复中拒绝  $H_0$  的%)

样本容量	$\alpha_1$	0	2	0	1
	$\alpha_2$	0	0	2	1
20		2.8	67.4	47.2	41.1
30		3.9	77.8	91.4	75.4
40		3.1	94.8	99.9	95.2
50		4.6	97.2	100.0	98.1
60		4.8	100.0	100.0	94.6
70		5.3	99.8	100.0	99.8
80		4.6	100.0	100.0	100.0

表 7.2.5 显著性水平 10% 下模拟结果(1 000 次重复中拒绝  $H_0$  的%)

样本容量	$\alpha_1$	0	2	0	1
	$\alpha_2$	0	0	2	1
20		7.1	86.5	63.3	63.4
30		8.9	90.1	96.9	86.7
40		6.5	99.3	99.9	97.6
50		8.7	98.9	100.0	99.2
60		9.1	100.0	100.0	97.3
70		8.8	100.0	100.0	100.0
80		7.3	100.0	100.0	100.0

由表 7.2.4 和表 7.2.5 的结果可见:

(i) 如果零假设是真的, 则  $LM$  检验的功效是令人满意的.

(ii) 当样本容量  $n \geq 40$  时,  $LM$  检验的功效仍然是令人满意.

因此, 当  $n \geq 40$ , 可用  $LM$  检验来判断线性回归模型是否存在异方差性. 当  $n < 40$  时, 应视对异方差结构的假定, 确定是否可用  $LM$  检验. 若假定异方差结构是平方函数的形式, 则当  $n \geq 20$ ,  $LM$  检验都是有效的.

最后我们通过一个实际例子来介绍如何分析和解决具有异方差结构的回归模型.

**例 7.2.3** 表 7.2.6 给出了 27 组来自一实际问题的数据. 对表中的数据, 我们作  $(x, y)$  的图形 (见图 7.2.1). 由图 7.2.1 可见, 我们有充分理由对  $(x, y)$  拟合如下的二元线性回归模型

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, 27, \quad (7.2.29)$$

其最小二乘结果列于表 7.2.7.

表 7.2.6

序 号	$y$	$x$
1	30	294
2	32	247
3	37	267
4	44	358
5	47	423
6	49	311
7	56	450
8	62	534
9	68	438
10	78	697

续表

序 号	$y$	$x$
11	80	688
12	84	630
13	88	709
14	97	627
15	100	615
16	109	999
17	114	1022
18	117	1015
19	106	700
20	128	850
21	130	980
22	160	1 025
23	97	1 021
24	180	1 200
25	112	1 250
26	210	1 500
27	135	1 650

表 7.2.7 模型(7.2.29)参数估计结果

变 量	参数估计值	标准误差的估计 $s.e$	$t$ 值
$x$	0.1150	0.0011	9.30
常 数	14.448	9.562	1.51
$s = 21.729$	$R^2 = 0.776$		

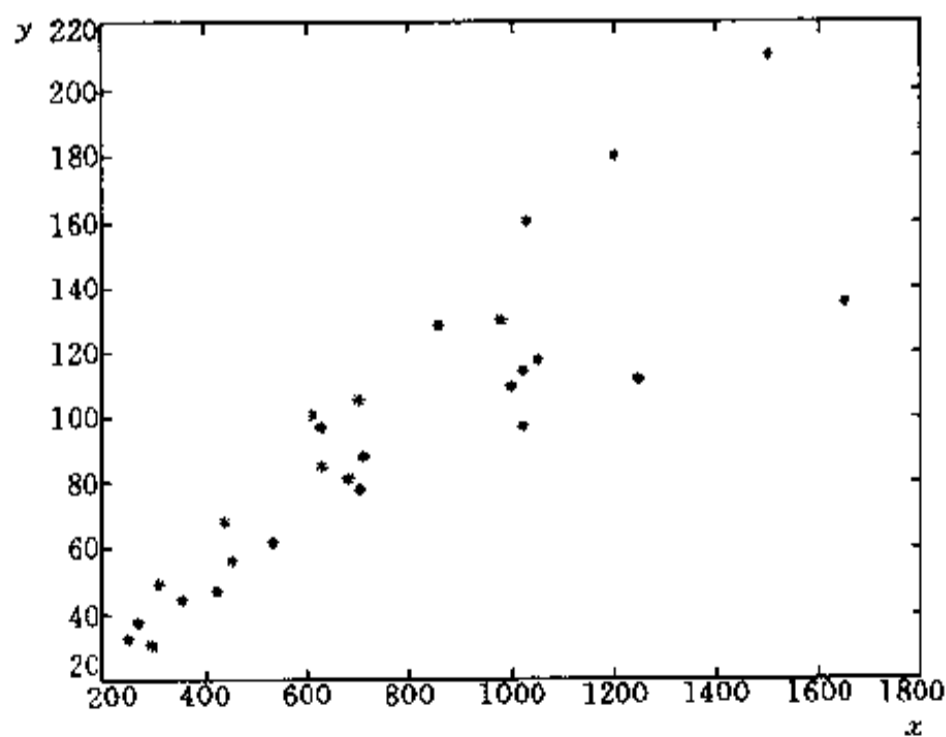
图 7.2.1  $(x, y)$  的散点图

表 7.2.7 中的一些记号的含义为:

$$s.e.: \quad s.e.(\hat{\beta}_1) = s / \left[ \sum_{i=1}^{27} (x_i - \bar{x})^2 \right]^{\frac{1}{2}},$$

$$s.e.(\hat{\beta}_0) = s \left[ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{27} (x_i - \bar{x})^2} \right]^{\frac{1}{2}},$$

它们是估计量  $\hat{\beta}_1$  和  $\hat{\beta}_0$  的标准误差的估计。

$$s: \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2,$$

$s^2$  是  $\sigma^2$  的估计。

$$R^2 = 1 - \frac{\sum_{i=1}^{27} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{27} (y_i - \bar{y})^2},$$

它是复相关系数(或称判定系数)。

残差  $\hat{e}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, i=1, \dots, 27$  的散点图由图 7.2.2 给出, 由图 7.2.2 可见, 随着  $i$  的增大,  $\hat{e}_i$  的点在平面上以  $x=0$  为对称轴呈喇叭形状扩大, 提示有异方差存在的可能。进一步作  $(x_i, \tilde{e}_i)$  在平面上的散点图(图 7.2.3), 其中  $\tilde{e}_i$  为标准化残差。我们发现, 从趋势上,  $\tilde{e}_i$  随  $x_i$  的增加而增加。一般地, 如果残差所散布的点的区域随  $x$  的增加而扩大, 误差的方差也随  $x$  增加而增加。如果这个区域随  $x$  的增加而收缩, 则误差方差随  $x$  的增加而减小; 如果残差落在两条平行于  $x$  轴的线之间, 可粗略地认为无异方差性。在许多工业、经济和生物学的应用问题中, 不等的误差方差常常遇到, 而且标准化残差趋向于随自变量的增加而增加。

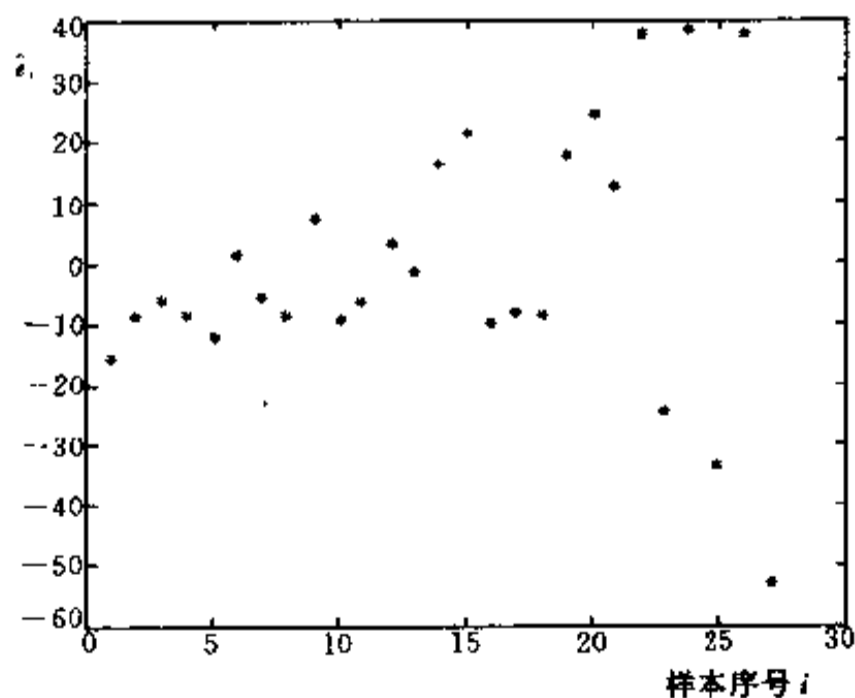
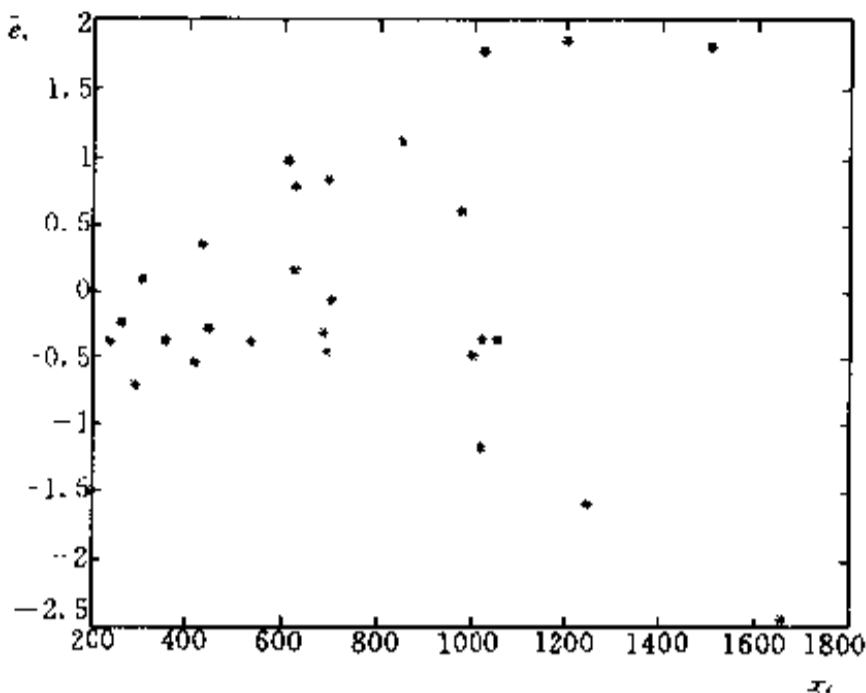


图 7.2.2  $\hat{e}_i$  的散点图

是否真的存在异方差性呢? 我们用  $LM$  检验方法来检验。此时假定

图 7.2.3  $(x_i, \hat{e}_i)$  的散点图

$$\sigma_i^2 = \sigma^2(1 + \alpha x_i^2).$$

按 LM 检验的步骤算得  $LM = 21.8461$ . 由于  $\sigma_i^2$  的结构中含有两个未知参数, 故 LM 检验统计量的渐近分布为  $\chi_2^2$  分布. 在显著性水平 0.05 下, 临界值为  $\chi_2^2(0.05) = 5.99$ . 显然  $LM > \chi_2^2(0.05)$ . 故认为模型 (7.2.29) 的异方差性是显著的. 于是我们可按第三章介绍的广义最小二乘估计去求参数  $\beta_0, \beta_1$  的估计. 估计结果列于表 7.2.8.

表 7.2.8 广义最小二乘估计的结果

变 量	参数估计值	标准误差的估计 $s.e$	$t$ 值
$x$	0.120 6	8.990 6	13.091 2
常数	3.964 1	0.158 8	0.844 4
RSS = 0.863 4		$R^2 = 0.999 7$	

残差散点图由图 7.2.4 给出. 图中  $\hat{e}_i$  的散布状况说明  $\hat{e}_i$  是随机分布的, 即异方差性已被消除. 表 7.2.8 也表明广义最小二乘估计的拟合效果很好.

对表 7.2.6 中的数据, 计算  $x$  和  $y$  的样本均值和样本均方差, 有

$$\bar{x} = 760.392 6, \sigma_x = 377.270 5, \bar{y} = 94.444 4, \sigma_y = 45.007 1.$$

故有

$$\sigma_x/\bar{x} = 0.496 0, \sigma_y/\bar{y} = 0.467 5.$$

这表明数据的离散程度较大. 对这类数据, 进行数据分析时, 常作一些变换. 我们对表 7.2.6 的数据  $y$  作对数变换, 并拟合如下二元线性回归模型

$$\ln(y_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i. \quad (7.2.30)$$

拟合结果列于表 7.2.9. 拟合残差散点图由图 7.2.5 给出.

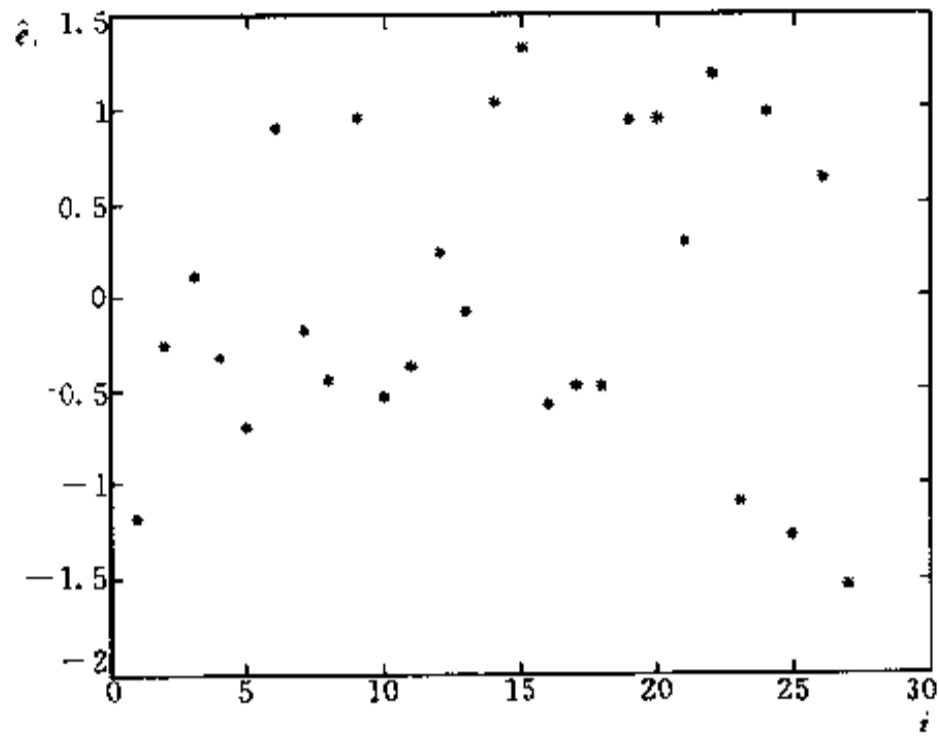
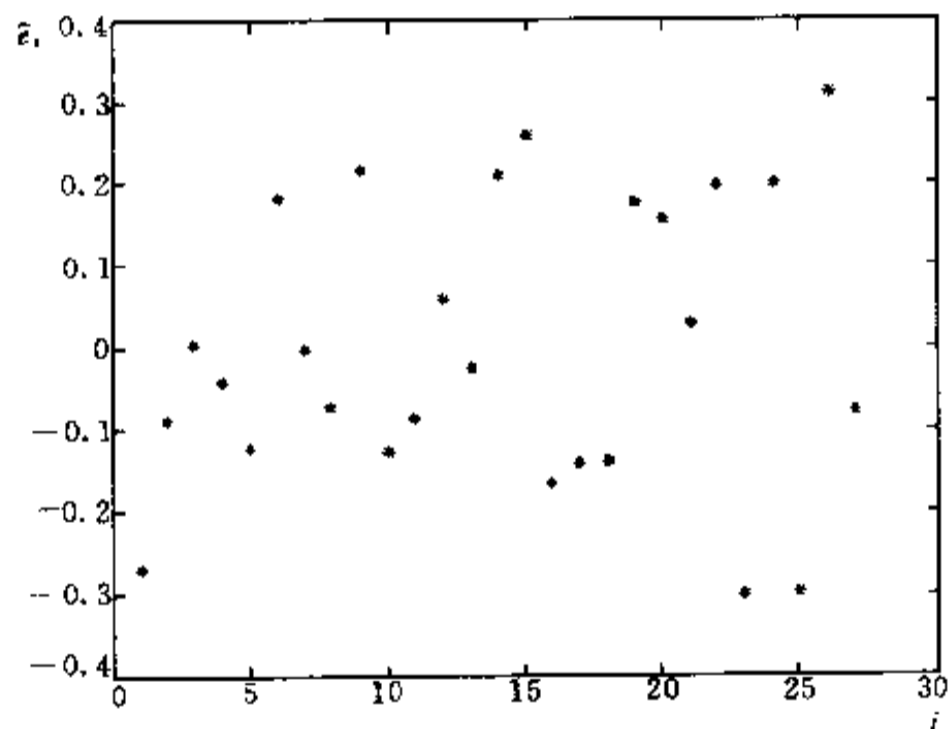
图 7.2.4  $\hat{e}_i$  的散点图图 7.2.5  $\hat{e}_i$  的散点图

图 7.2.5 表明,  $\hat{e}_i$  的散布状态是随机的. 因此模型 (7.2.30) 不存在异方差性. 表 7.2.9 的结果表明: 拟合效果也是好的, 但有些指标较广义最小二乘估计差.

通过这个例子, 我们看到了在处理实际问题时, 统计方法使用的多样性. 但是评判最终结果的好坏只能从所研究的实际问题所处领域的有关理论来做出.

表 7.2.9 模型(7.2.30)的拟合结果

变 量	参数估计值	标准误差估计 s.e	t 值
$x$	0.003 1	$0.399 \times 10^{-3}$	7.80
$x^2$	$0.11 \times 10^{-5}$	$0.220 \times 10^{-6}$	-4.93
常数	2.852	0.157	18.20
	$s=0.182$	$R^2=0.886$	

### § 7.3 具有自回归误差的线性回归模型

在 § 7.1 中我们曾经指出,对经典线性回归模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + e_i, i = 1, \cdots, n, \quad (7.3.1)$$

实际问题有时难于满足其对误差项  $e_i$  的全部假定. 当  $\text{Cov}(e_i, e_j) \neq 0, i \neq j$  时, 即误差项的不相关性不满足, 我们称误差项存在自相关性. 如何处理这类统计问题的建模呢? 通常的处理方法是对误差项拟合一个宽平稳自回归模型 (AR 模型), 比如一阶自回归 (AR(1)) 模型

$$e_i = \varphi e_{i-1} + \varepsilon_i, \quad |\varphi| < 1, \quad (7.3.2)$$

其中  $\varepsilon_i, i = 1, \cdots, n$  是独立同分布的随机变量, 且  $E(\varepsilon_i) = 0, E(\varepsilon_i^2) = \sigma_\varepsilon^2 < \infty$ . 因此作为准备知识, 本节先来扼要地介绍宽平稳 AR 模型的建模. 关于这方面的文献, 读者可参阅安鸿志等 (1983) 和常学将等 (1993).

设  $\{\xi_t, t = 1, 2, \cdots\}$  是一个随机变量序列, 如果

$$\begin{cases} E(\xi_t) = \mu (\text{常数}) \\ \text{Cov}(\xi_t, \xi_{t-k}) = \gamma_k \quad k = 0, 1, 2, \cdots \end{cases} \quad (7.3.3)$$

则称  $\{\xi_t, t = 1, 2, \cdots\}$  是一个宽平稳序列, 或称为是宽平稳的.  $\gamma_k$  称为  $\{\xi_t, t = 1, 2, \cdots\}$  的  $k$  阶自协方差,  $\{\gamma_k\}$  称为  $\{\xi_t, t = 1, 2, \cdots\}$  的自协方差函数.

$$\rho_k = \gamma_k / \gamma_0, k = 1, 2, \cdots \quad (7.3.4)$$

称为  $\{\xi_t, t = 1, 2, \cdots\}$  的  $k$  阶自相关,  $\{\rho_k\}$  称为  $\{\xi_t, t = 1, 2, \cdots\}$  的自相关函数.

显然, 如果随机序列  $\{e_i, i = 1, 2, \cdots\}$  的均值为零, 方差有限, 且  $\text{Cov}(e_i, e_j) = 0, i \neq j$ , 则  $\{e_i, i = 1, 2, \cdots\}$  是一个宽平稳序列.

如果  $\{\xi_t, t = 1, 2, \cdots\}$  满足如下的随机差分方程:

$$\xi_t = \varphi \xi_{t-1} + \varepsilon_t, \quad |\varphi| < 1, \quad (7.3.5)$$

其中  $\{\varepsilon_t, t = 1, 2, \cdots\}$  是独立同分布的随机序列,  $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty$ , 则称  $\{\xi_t, t = 1, 2, \cdots\}$  服从一阶自回归模型, 简记为 AR(1).

对 AR(1)模型(7.3.5)反复递推,可得

$$\begin{aligned}\xi_t &= \varphi \xi_{t-1} + \varepsilon_t \\ &= \varphi(\varphi \xi_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \varphi^2 \xi_{t-2} + \varphi \varepsilon_{t-1} + \varepsilon_t \\ &= \dots \\ &= \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \dots.\end{aligned}\quad (7.3.6)$$

不难验证

$$\begin{aligned}E(\xi_t) &= 0, \\ \gamma_0 &= \text{Var}(\xi_t) = \sigma_\varepsilon^2(1 + \varphi^2 + \varphi^4 + \dots), \\ \gamma_k &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \varphi^i \varphi^{k+i}, k \geq 0.\end{aligned}$$

由于  $|\varphi| < 1$ , 故而

$$\gamma_0 = \sigma_\varepsilon^2 / (1 - \varphi^2) \quad (7.3.7)$$

是有限的, 而且

$$\begin{aligned}\gamma_k &= \varphi^k \sigma_\varepsilon^2 / (1 - \varphi^2) \\ &= \varphi^k \gamma_0, \quad k \geq 0,\end{aligned}\quad (7.3.8)$$

因此, 自相关函数为

$$\rho_k = \varphi^k. \quad (7.3.9)$$

由(7.3.8)式可见, 由(7.3.5)式定义的随机序列  $\{\xi_t\}$  是宽平稳的.

对(7.3.5)式两边乘以  $\xi_{t-k}$ , 并取数学期望, 得到

$$E\xi_t \xi_{t-k} = \varphi E\xi_{t-1} \xi_{t-k} + E\varepsilon_t \xi_{t-k}, \quad k > 0.$$

由(7.3.6)式知,  $E\varepsilon_t \xi_{t-k} = 0$ . 于是上式可写为

$$\gamma_k = \varphi \gamma_{k-1}, \quad k > 0, \quad (7.3.10)$$

(7.3.10)式称为 Yule-Walker 方程.

当我们获得 AR(1)模型的  $n$  个样本观测值  $\xi_1, \dots, \xi_n$  后, 我们通常以

$$\hat{\gamma}_k = \frac{1}{n-k+1} \sum_{i=k+1}^n \xi_i \xi_{i-k}, \quad k = 0, 1, 2, \dots, n-1 \quad (7.3.11)$$

作为  $\gamma_k$  的估计, 而以

$$\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0, \quad k = 1, 2, \dots, n-1 \quad (7.3.12)$$

作为  $\rho_k$  的估计.

在(7.3.10)式中, 用估计量代替  $\gamma_k$  得到

$$\hat{\gamma}_1 = \varphi \hat{\gamma}_0,$$

由此得到  $\varphi$  的估计



$$\hat{\varphi} = \hat{\gamma}_1 / \hat{\gamma}_0 = \sum_{i=2}^n \xi_i \xi_{i-1} / \sum_{i=1}^n \xi_i^2, \quad (7.3.13)$$

我们称  $\hat{\varphi}$  为  $\varphi$  的矩估计. 实际上, 它还是  $\varphi$  的最小二乘估计.

如果  $\{\xi_t, t=1, 2, \dots\}$  满足随机差分方程

$$\xi_t = \varphi_1 \xi_{t-1} + \dots + \varphi_p \xi_{t-p} + \varepsilon_t, \quad (7.3.14)$$

其中  $\{\varepsilon_t, t=1, 2, \dots\}$  如(7.3.5)式中假定, 系数  $\varphi_1, \dots, \varphi_p$  满足

$$1 - \varphi_1 Z + \dots + \varphi_p Z^p \neq 0, \quad |Z| \leq 1, \quad (7.3.15)$$

则称  $\{\xi_t, t=1, 2, \dots\}$  服从  $p$  阶自回归模型, 简记为  $AR(p)$ .

由时间序列分析知识, 在(7.3.15)下, (7.3.14)可以表示为

$$\xi_t = \sum_{j=0}^p \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=1}^{\infty} |\psi_j| < \infty. \quad (7.3.16)$$

类似于  $AR(1)$  模型, 对  $AR(p)$  模型(7.3.14)我们有

$$E\xi_t = 0,$$

$$\gamma_k = \sigma_\varepsilon^2 \sum_{j=1}^{\infty} \psi_j \psi_{k+j}, \quad k = 0, 1, 2, \dots. \quad (7.3.17)$$

因此, 由(7.3.16)式知,  $\gamma_k$  是有限的. 从而由(7.3.14)定义的  $\{\xi_t, t=1, 2, \dots\}$  是一个宽平稳序列. (7.2.15)式称为  $AR(p)$  模型的平稳性条件.  $p$  称为  $AR$  模型的阶数.

对(7.3.14)式两边乘以  $\xi_{t-k}, k > 0$ , 并取数学期望, 得到

$$\gamma_k = \varphi_1 \gamma_{k-1} + \dots + \varphi_p \gamma_{k-p} + E\varepsilon_t \xi_{t-k}, \quad k > 0.$$

由(7.3.16)式有  $E\varepsilon_t \xi_{t-k} = 0$ . 于是得到  $AR(p)$  模型的 Yule-Walker 方程

$$\gamma_k = \varphi_1 \gamma_{k-1} + \dots + \varphi_p \gamma_{k-p}, \quad k > 0, \quad (7.3.18)$$

或等价地

$$\rho_k = \varphi_1 \rho_{k-1} + \dots + \varphi_p \rho_{k-p}, \quad k > 0. \quad (7.3.19)$$

设  $\xi_1, \dots, \xi_n$  是  $AR(p)$  模型(7.3.14)的  $n$  个样本观测值, 由此可得  $\gamma_k$  和  $\rho_k$  的估计如(7.3.11)和(7.3.12). 在(7.3.18)和(7.3.19)式中用估计量代替  $\gamma_k$  和  $\rho_k$ , 得到

$$\hat{\gamma}_k = \varphi_1 \hat{\gamma}_{k-1} + \dots + \varphi_p \hat{\gamma}_{k-p}, \quad k = 1, 2, \dots, p \quad (7.3.20)$$

或

$$\hat{\rho}_k = \varphi_1 \hat{\rho}_{k-1} + \dots + \varphi_p \hat{\rho}_{k-p}, \quad k = 1, 2, \dots, p \quad (7.3.21)$$

记  $\hat{\boldsymbol{\gamma}} = (\hat{\rho}_1, \dots, \hat{\rho}_p)'$ ,  $\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \dots, \hat{\varphi}_p)'$ ,

$$\hat{\mathbf{R}} = \begin{pmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{1-p} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{2-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \dots & 1 \end{pmatrix},$$

则(7.3.21)式可写为

$$\hat{\mathbf{R}}\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\gamma}}.$$

由时间序列分析知识,  $\hat{\mathbf{R}}$  是可逆的. 于是可解得

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{R}}^{-1}\hat{\boldsymbol{\gamma}}, \quad (7.3.22)$$

称其为  $AR(p)$  模型(7.3.14)参数的矩估计.

由(7.3.14)式, 我们还可求  $\boldsymbol{\phi}$  的最小二乘估计  $\hat{\boldsymbol{\phi}}_L$ . 令

$$\mathbf{y}_n = (\xi_{p+1}, \dots, \xi_n)',$$

$$\mathbf{X}_n = \begin{bmatrix} \xi_p & \dots & \xi_1 \\ \xi_{p+1} & \dots & \xi_2 \\ \vdots & & \vdots \\ \xi_{n-1} & \dots & \xi_{n-p} \end{bmatrix},$$

则  $\boldsymbol{\phi}$  的最小二乘估计为

$$\hat{\boldsymbol{\phi}}_L = (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{y}_n. \quad (7.3.23)$$

可以证明(见常学将等(1993)), 当样本量很大时,  $\hat{\boldsymbol{\phi}}$  和  $\hat{\boldsymbol{\phi}}_L$  是渐近相等的. 但在有限样本下,  $\hat{\boldsymbol{\phi}}$  的估计精度总比  $\hat{\boldsymbol{\phi}}_L$  差. 以下通过两个  $AR(2)$  模型的数值模拟来比较  $LS$  估计和矩估计的精度.

$$\text{模型 I: } \xi_t = 0.8\xi_{t-1} - 0.64\xi_{t-2} + \varepsilon_t,$$

$$\text{模型 II: } \xi_t = 0.4\xi_{t-1} - 0.55\xi_{t-2} + \varepsilon_t.$$

其中  $\{\varepsilon_t, t=1, 2, \dots, n\}$  是独立同  $N(0, 1)$  分布. 取样本容量  $n=100, 300, 500$ . 对样本量的每种选择, 模拟重复 1 000 次.

模拟过程如下:

(1) 取初值  $\xi_{-1} = \xi_0 = 0$ . 对给定的样本容量  $n$ , 用计算机产生  $n+100$  个正态  $N(0, 1)$  随机数.

为什么要多产生 100 个正态  $N(0, 1)$  随机数呢? 这主要是为了消除确定性初值  $\xi_{-1} = \xi_0 = 0$  的影响. 一般地, 在产生随机性时间序列时, 对驱动模型的误差序列  $\{\varepsilon_t\}$ , 产生  $n+100$  个所要求的随机数, 并由此通过模型递推产生所要的  $\xi_t$ . 使用时丢掉前 100 个, 而只用后  $n$  个数据, 这样即可保证  $\xi_t$  的随机性.

(2) 由递推关系

$$\xi_t = 0.8\xi_{t-1} - 0.64\xi_{t-2} + \varepsilon_t,$$

初值  $\xi_0 = \xi_{-1} = 0$  及  $\varepsilon_t$  产生  $n+100$  个  $\xi_t$ , 用后  $n$  个  $\xi_t$  作为模型 I 的模拟数据.

(3) 对所得的  $\xi_t, t=1, 2, \dots, n$  (注意这时的  $\xi_t$ , 实际上是  $\xi_{100+t}$ ), 按(7.3.22)和(7.3.23), 计算自回归系数的矩估计和最小二乘估计, 记作  $\hat{\boldsymbol{\phi}}_m^{(i)}$ ,  $\hat{\boldsymbol{\phi}}_{LS}^{(i)}$ .

(4) 将(1)—(3)步独立地重复 1 000 次, 计算样本均值和样本均方误差.

(5) 对不同的样本容量  $n$  的选择, 重复(1)—(4)步.

我们把所得的结果列于表 7.3.1.

表 7.3.1

模 型	n = 100		n = 300		n = 500	
	LS 估计	矩估计	LS 估计	矩估计	LS 估计	矩估计
I	0.791 6 (0.076 7)	0.775 2 (0.079 5)	0.796 2 (0.042 7)	0.791 0 (0.043 8)	0.798 3 (0.033 7)	0.795 5 (0.037 9)
	0.632 3 (0.080 5)	-0.615 2 (0.083 2)	-0.636 3 (0.045 7)	-0.630 9 (0.046 2)	-0.637 6 (0.033 3)	-0.634 3 (0.033 8)
II	0.398 2 (0.083 4)	0.391 3 (0.082 4)	0.398 9 (0.047 3)	0.396 9 (0.047 3)	0.399 5 (0.036 7)	0.398 2 (0.036 8)
	-0.541 8 (0.084 1)	-0.530 6 (0.085 0)	-0.548 0 (0.048 3)	-0.544 2 (0.048 3)	-0.549 0 (0.037 0)	-0.546 6 (0.037 1)

表中数字分别为参数估计值的平均值和估计误差的均方根,其定义为

$$0.791\ 6 = \frac{1}{1\ 000} \sum_{i=1}^{1\ 000} \hat{\varphi}_1^{(i)},$$

$$(0.076\ 7) = \left[ \frac{1}{1\ 000} \sum_{i=1}^{1\ 000} (\hat{\varphi}_1^{(i)} - 0.8)^2 \right]^{\frac{1}{2}}.$$

尽管矩估计的精度不如最小二乘估计,但在 AR 模型建模中,人们还是偏爱矩估计.这是因为矩估计和最小二乘估计相差不大,更重要在于矩估计有如下的递推算法.

在获得观测样本  $\xi_1, \dots, \xi_n$  后,我们对样本拟合 AR( $p$ )模型,

$$\xi_t = \hat{\varphi}_{p1}\xi_{t-1} + \dots + \hat{\varphi}_{pp}\xi_{t-p} + \varepsilon_t,$$

$\varepsilon_t$  是零均值,方差为  $\hat{\sigma}_\varepsilon^2(p)$  的误差序列,其中系数  $\hat{\varphi}_{p1}, \dots, \hat{\varphi}_{pp}$  可由如下 Durbin-Levinson 递推算法求得.

$$\left\{ \begin{array}{l} \hat{\varphi}_{11} = \hat{\rho}_1, \\ \hat{\sigma}_\varepsilon^2(1) = \hat{\gamma}_0(1 - \hat{\varphi}_{11}^2), \\ \hat{\varphi}_{kk} = \left[ \hat{\gamma}_k - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j} \hat{\gamma}_{k-j} \right] / \hat{\sigma}_\varepsilon^2(k-1), \\ \begin{cases} \hat{\varphi}_{k1} \\ \vdots \\ \hat{\varphi}_{kk-1} \end{cases} = \begin{cases} \hat{\varphi}_{k-11} \\ \vdots \\ \hat{\varphi}_{k-1,k-1} \end{cases} - \hat{\varphi}_{kk} \begin{cases} \varphi_{k-1,k-1} \\ \vdots \\ \hat{\varphi}_{k-11} \end{cases}, \\ \hat{\sigma}_\varepsilon^2(k) = \hat{\sigma}_\varepsilon^2(k-1)(1 - \hat{\varphi}_{kk}^2), \\ k = 2, 3, \dots, p. \end{array} \right. \quad (7.3.24)$$

递推公式(7.3.24)还为 AR 模型的识别提供了一个关于残差方差的递推算法.

对 AR 模型来说,模型识别主要是确定自回归阶数  $p$  一个有效的统计方法是第五章介绍的 AIC 类准则.

定义统计量

$$BIC(k) = \ln \hat{\sigma}_e^2(k) + \frac{k \ln n}{n}, \quad (7.3.25)$$

关于  $k$  极小化(7.2.25)式,使其达到最小的  $\hat{k}$  就是所求的阶  $p$  的估计,即

$$BIC(\hat{k}) = \min_{1 \leq k \leq P} BIC(k),$$

其中  $P$  是一个给定的大于真阶  $p$  的常数.

对模型 I,取  $p=10$ ,先用递推公式(7.3.24)计算出  $\hat{\sigma}_e^2(k), k=1,2,\dots,10$ . 把它们代入统计量  $BIC(k)$ ,并求这 10 个  $BIC(k)$  中的最小值,相应的  $k$  就是所求的阶估计.我们选择样本容量  $n=100,300,500$ .对每一样本量的选择,模拟重复 1 000 次,结果列于表 7.3.2.

表 7.3.2

样本容量 $n$	阶的估计频数									
	1	2	3	4	5	6	7	8	9	10
100	0	973	25	2	0	0	0	0	0	0
300	0	982	18	0	0	0	0	0	0	0
500	0	989	11	0	0	0	0	0	0	0

由表中结果可见,用 BIC 准则定阶的效果是令人满意的.

作为本节的结束,我们用一组实际数据来拟合一个 AR 模型.

例 7.3.1 下面的数据(表 7.3.3)是美国 1975 年第一季度到 1982 年第四季度啤酒产量.单位:百万桶.

表 7.3.3

36.140 0	44.600 0	44.150 0	35.720 0	36.190 0	44.630 0	46.950 0
36.900 0	39.660 0	49.720 0	44.490 0	36.540 0	41.440 0	49.070 0
48.980 0	39.590 0	44.290 0	50.090 0	48.420 0	41.390 0	46.110 0
53.440 0	53.000 0	42.520 0	44.610 0	55.180 0	52.240 0	41.660 0
47.840 0	54.270 0	52.310 0	42.030 0			

表 7.3.3 中数据依时间变化的图形如图 7.3.1.

由图 7.3.1 可见,随时间的变化,数据  $\xi_t$  呈现周期变化,并有明显的增加趋势.对这样的数据,一般对  $\xi_t$  作如下的变化:

$$\Delta \xi_t = \xi_t - \xi_{t-1},$$

称作对时间序列  $\{\xi_t\}$  作一阶差分.记  $y_t = \Delta \xi_t$ ,图 7.3.2 是  $y_t$  随时间变化的图形.

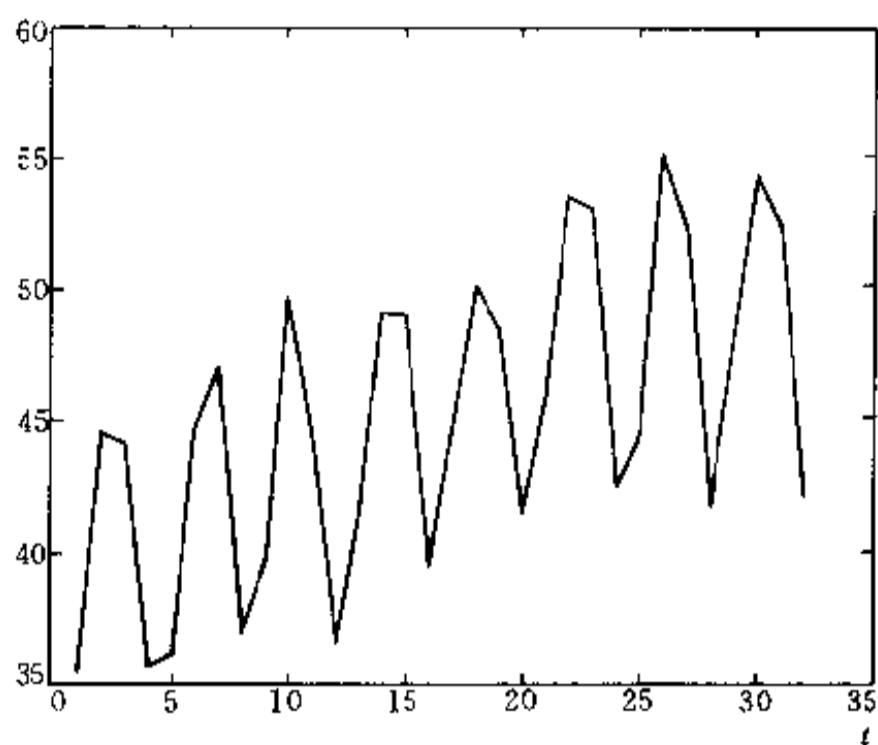


图 7.3.1

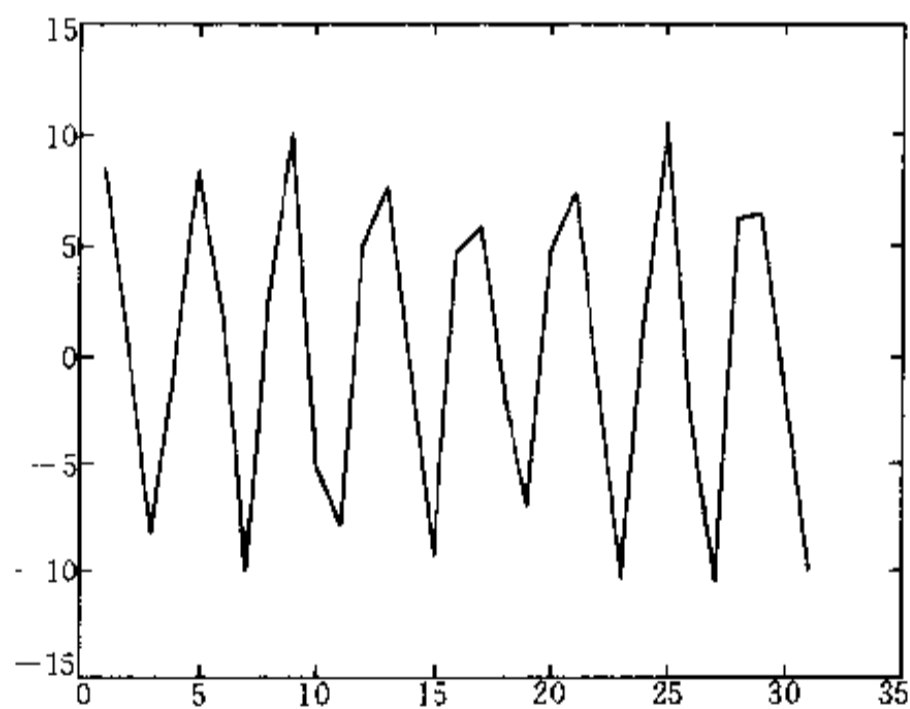


图 7.3.2

图 7.3.2 表明数据  $\xi_t$  经过一阶差分后趋势性被去掉了. 由于前面讨论 AR 模型建模的过程是在  $E(\xi_t) = 0$  的前提下进行的. 而一般地, 实际数据未必满足这一点. 或者说, 我们拟合的 AR 模型应为

$$\xi_t = \varphi_0 + \varphi_1 \xi_{t-1} + \cdots + \varphi_p \xi_{t-p} + \varepsilon_t. \quad (7.3.26)$$

对上式两边取数学期望, 并由宽平稳性可得

$$E\xi_t = \varphi_0 + \varphi_1 E\xi_{t-1} + \cdots + \varphi_p E\xi_{t-p},$$

亦即

$$\mu = \varphi_0 + \varphi_1\mu + \cdots + \varphi_p\mu. \quad (7.3.27)$$

再由(7.3.26)与(7.3.27)式得

$$\xi_t - \mu = \varphi_1(\xi_{t-1} - \mu) + \cdots + \varphi_p(\xi_{t-p} - \mu) + \varepsilon_t.$$

令  $\eta_t = \xi_t - \mu$ , 则上式可写为

$$\eta_t = \varphi_1\eta_{t-1} + \cdots + \varphi_p\eta_{t-p} + \varepsilon_t,$$

且  $E\eta_t = 0$ .

对经过一阶差分后的数据  $y_t$ , 计算样本均值得

$$\bar{y} = 0.1900.$$

对  $y_t$  作如下变换,

$$\eta_t = y_t - 0.1900, \quad t = 1, \cdots, 31. \quad (7.3.28)$$

这时  $\{\eta_t\}$  就是零均值的. 对  $\{\eta_t\}$ , 用前面介绍的  $BIC$  准则来定阶. 取  $P = 10$ . 先由 Durbin-Levinson 递推算法求得  $\hat{\sigma}_e^2(k)$ ,  $k = 1, 2, \cdots, 10$ . 然后再由(7.3.25)求  $BIC(k)$  的值. 下面是由数据求出的 10 个  $BIC$  值:

$$\begin{array}{ccccc} 3.9263 & 2.6774 & 2.7202 & 2.7854 & 2.8892 \\ 2.9993 & 3.1100 & 3.2128 & 3.3236 & 3.3886. \end{array}$$

可见  $BIC(2)$  最小, 故  $\{\eta_t\}$  服从  $AR(2)$  模型.

Durbin-Levinson 递推算法关于参数  $\varphi$  的结果如表 7.3.4.

表 7.3.4

-0.0338	0	0	0	0	0	0
-0.0629	-0.8621	0	0	0	0	0
-0.2838	-0.8782	-0.2563	0	0	0	0
-0.2297	-0.6928	-0.1964	0.2112	0	0	0
-0.2121	-0.7092	-0.2542	0.1920	-0.0835	0	0
-0.2099	0.7141	-0.2477	0.2102	-0.0781	0.0256	0
-0.2098	-0.7145	-0.2467	0.2090	-0.0813	0.0247	-0.0045
-0.2094	-0.7165	-0.2399	0.1916	-0.0607	0.0845	0.0130
-0.2120	-0.7169	-0.2426	0.1935	-0.0667	0.0920	0.0355
-0.2187	-0.7360	-0.2501	0.1740	-0.0526	0.0511	0.0868
0	0	0				
0	0	0				
0	0	0				
0	0	0				
0	0	0				
0	0	0				
0	0	0				
0.0837	0	0				
0.0902	0.0314	0				
0.2418	0.0762	0.2114				

于是, 由表 7.3.4 中第二行的非零数即为我们所求的参数矩估计值. 故

$$\varphi_1 = -0.0629, \quad \varphi_2 = -0.8621,$$

这样对  $\eta_t$  拟合的  $AR(2)$  模型为

$$\eta_t = 0.0629\eta_{t-1} - 0.8621\eta_{t-2} + \varepsilon_t.$$

再由(7.3.28)式得

$$y_t = 0.3589 - 0.0629y_{t-1} - 0.8621y_{t-2} + \varepsilon_t.$$

再由  $y_t$  的定义得到

$$\xi_t = 0.3589 + 0.9371\xi_{t-1} - 0.7992\xi_{t-2} + 0.8621\xi_{t-3} + \varepsilon_t.$$

我们也可以在定阶后,用最小二乘法来求模型的参数估计.

## § 7.4 具有一阶自回归误差的线性回归模型

具有一阶自回归( $AR(1)$ )误差的线性回归模型可表述为

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_{p-1} x_{tp-1} + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \cdots, n, \end{cases} \quad (7.4.1)$$

其中  $\{\varepsilon_t, t = 1, \cdots, n\}$  是独立同分布的随机误差,  $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty$ , 自回归系数满足  $|\varphi| < 1$ .

对模型(7.4.1)的统计推断主要包含如下内容:

- (i) 模型参数  $\beta_0, \beta_1, \cdots, \beta_{p-1}, \varphi, \sigma_\varepsilon^2$  的估计;
- (ii) 随机误差项的自相关性检验.

本节先来讨论(i). 首先考虑具有  $AR(1)$  误差的一元线性回归模型, 即

$$\begin{cases} y_t = \beta_0 + \beta_1 x_t + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \cdots, n. \end{cases} \quad (7.4.2)$$

由(7.4.2)式, 我们有

$$\varphi y_{t-1} = \varphi \beta_0 + \varphi \beta_1 x_{t-1} + \varphi e_{t-1},$$

用此式去减(7.4.2)式, 经整理可得

$$y_t = \beta_0(1 - \varphi) + \varphi y_{t-1} + \beta_1 x_t - \varphi \beta_1 x_{t-1} + \varepsilon_t.$$

令  $\alpha_0 = \beta_0(1 - \varphi), \alpha_1 = \varphi, \alpha_2 = \beta_1, \alpha_3 = -\varphi \beta_1$ , 则有

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + \varepsilon_t, \quad t = 2, 3, \cdots, n, \quad (7.4.3)$$

注意, 经这样变形后, 样本量减少一个.

由于  $\{\varepsilon_t, t = 2, 3, \cdots, n\}$  是独立同分布的, 且  $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty$ , 故(7.4.3)式满足经典线性回归模型的假定. 由第三章介绍的线性回归模型参数的最小二乘估计法, 我们可容易地求得模型(7.4.3)中未知参数的最小二乘估计  $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$  和  $\hat{\sigma}_\varepsilon^2$ . 由此可得模型(7.4.2)中未知参数的估计

$$\hat{\beta}_{0L} = \frac{\hat{\alpha}_0}{1 - \hat{\alpha}_1}, \quad \hat{\beta}_{1L} = \hat{\alpha}_2, \quad \hat{\varphi}_L = \hat{\alpha}_1.$$

我们称由上述方法获得的参数估计为  $\beta_0, \beta_1, \varphi$  和  $\sigma_\varepsilon^2$  的最小二乘估计.

**例 7.4.1** 考虑如下具有 AR(1) 误差的一元线性回归模型

$$\begin{cases} y_t = 1.5 + 0.9x_t + e_t, \\ e_t = 0.67e_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \end{cases} \quad (7.4.4)$$

其中  $\{\varepsilon_t, t = 1, \dots, n\}$  为独立同分布  $N(0, 6.25)$  随机误差.  $x_t$  由如下三个随机模型生成.

$$\text{I: } x_t = 15u_t, t = 1, \dots, n;$$

$$\text{II: } x_t = e^{0.04t} + u_t, t = 1, \dots, n;$$

$$\text{III: } x_t = \sin(3t) + u_t, t = 1, \dots, n;$$

其中  $u_t$  服从  $[0, 1]$  上均匀分布. 相应于不同的  $x_t$ , (7.4.4) 生成三个不同的线性回归模型. 对每个回归模型, 产生样本量分别为  $n = 50, 100, 200$  的样本各 1 000 组, 用最小二乘法估计模型中的参数, 计算这 1 000 次模拟所得估计的样本均值和与真值  $\beta_1 = 0.9, \varphi = 0.67$  和  $\sigma_\varepsilon^2 = 6.25$  的样本均方误差.

模拟过程如下:

(1) 对给定的  $n$ , 用计算机正态随机数发生器产生  $n + 100$  个正态  $N(0, 1)$  随机数  $\varepsilon_t, t = 1, 2, \dots, n + 100$ .

(2) 取初值  $e_0 = 0$ , 由递推公式

$$e_t = 0.67e_{t-1} + \varepsilon_t,$$

生成  $n + 100$  个 AR(1) 模型的数据, 取这组数据的后  $n$  个作为模拟数据. 仍记为  $e_t, t = 1, \dots, n$ .

(3) 用计算机均匀分布随机数发生器产生  $n$  个  $[0, 1]$  上均匀分布随机数. 分别按模型 I, II 和 III 生成所要的模拟数据  $x_t, t = 1, \dots, n$ .

(4) 按模型

$$y_t = 1.5 + 0.9x_t + e_t$$

生成  $y_t, t = 1, \dots, n$ . 并对数据  $(y_t, x_t)$  用最小二乘法拟合, 得参数估计值.

(5) 将上述步骤(1)–(4)独立地重复 1 000 次, 并计算这 1 000 次所得的估计值的样本均值和与真值的均方误差.

在模拟结果中, 我们没有计算估计值  $\beta_{0L}$  与  $\beta_0 = 1.5$  的样本均方误差, 原因是  $\beta_{0L}$  不是直接由最小二乘估计所得, 而且每次估计  $\hat{\alpha}_0$  和  $\hat{\alpha}_1$  都有估计误差. 所以计算  $\beta_{0L}$  与  $\beta_0 = 1.5$  的均方误差势必会扩大. 实际模拟也证实了这一点. 模拟结果列于表 7.4.1.



表 7.4.1

模型	参数	n = 50		n = 100		n = 200	
		平均值	均方误差	平均值	均方误差	平均值	均方误差
I	$\beta_0$	1.521 4		1.469 0		1.465 0	
	$\beta_1$	0.898 6	0.007 0	0.901 5	0.003 6	0.900 4	0.001 8
	$\varphi$	0.607 1	0.018 5	0.640 7	0.007 6	0.655 8	0.003 0
	$\sigma_e^2$	5.752 9	1.856 8	5.992 6	0.841 7	6.140 4	0.428 9
II	$\beta_0$	1.545 9		1.532 1		1.510 0	
	$\beta_1$	0.902 6	0.732 6	0.875 9	0.356 5	0.887 6	0.175 4
	$\varphi$	0.577 1	0.023 6	0.626 3	0.008 7	0.649 6	0.003 4
	$\sigma_e^2$	5.680 7	1.707 5	5.976 0	0.878 5	6.121 6	0.412 0
III	$\beta_0$	1.471 2		1.456 3		1.495 9	
	$\beta_1$	0.923 4	0.957 4	0.915 3	0.428 1	0.895 1	0.205 5
	$\varphi$	0.604 5	0.019	0.643 5	0.006 8	0.656 2	0.003 3
	$\sigma_e^2$	5.788 0	1.596 2	5.990 0	0.821 6	6.098 9	0.386 6

由表 7.4.1 的结果可以看出,  $\varphi$  的最小二乘估计是可用的. 对模型 I,  $\beta_1$  的估计效果是令人满意的. 但对模型 II 和 III,  $\beta_1$  估计的效果较差,  $\sigma_e^2$  的估计效果也差, 但随着样本量的逐渐增加, 估计的精度得到改善. 因此, 寻找更为有效的参数估计是必要的.

记  $e = (e_1, \dots, e_n)'$ , 则由 (7.3.2) 式,  $e_t$  的平稳性, 并注意到自协方差函数的对称性, 即  $\gamma_k = \gamma_{-k}$ , 我们有

$$\begin{aligned} \Sigma = \text{Cov}(e) &= \begin{pmatrix} Ee_1^2 & Ee_1e_2 & \cdots & Ee_1e_n \\ Ee_2e_1 & Ee_2^2 & \cdots & Ee_2e_n \\ \vdots & \vdots & & \vdots \\ Ee_ne_1 & Ee_ne_2 & \cdots & Ee_n^2 \end{pmatrix} \\ &= \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix}. \end{aligned}$$

再由 (7.3.8) 式知, 对 AR(1) 模型, 自协方差函数为

$$\gamma_k = \varphi^k \gamma_0,$$

于是

$$\Sigma = \gamma_0 \begin{pmatrix} 1 & \varphi & \varphi^2 & \cdots & \varphi^{n-1} \\ \varphi & 1 & \varphi & \cdots & \varphi^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \varphi^{n-1} & \varphi^{n-2} & \varphi^{n-3} & \cdots & 1 \end{pmatrix}.$$

由矩阵论知识又知,  $\Sigma$  是正定对称矩阵且

$$\Sigma^{-1} = \begin{pmatrix} 1 & -\varphi & 0 & \cdots & 0 & 0 & 0 \\ -\varphi & 1+\varphi^2 & -\varphi & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\varphi & 1+\varphi^2 & -\varphi \\ 0 & 0 & 0 & \cdots & 0 & -\varphi & 1 \end{pmatrix}.$$

记

$$\Gamma = \begin{pmatrix} \sqrt{1-\varphi^2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -\varphi & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -\varphi & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\varphi & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\varphi & 1 \end{pmatrix},$$

则不难验证

$$\Sigma^{-1} = \Gamma\Gamma.$$

对样本观测值  $(y_t, x_t), t=1, \dots, n$  做如下变换:

$$y_1^* = \sqrt{1-\varphi^2} y_1, \quad x_1^* = \sqrt{1-\varphi^2} x_1,$$

$$y_t^* = y_t - \varphi y_{t-1}, \quad x_t^* = x_t - \varphi x_{t-1}, t=2, 3, \dots, n.$$

则(7.4.2)式改写为

$$\begin{cases} y_1^* = \sqrt{1-\varphi^2} \beta_0 + \beta_1 x_1^* + \sqrt{1-\varphi^2} e_1, \\ y_2^* = (1-\varphi) \beta_0 + \beta_1 x_2^* + \varepsilon_2, \\ \vdots \\ y_n^* = (1-\varphi) \beta_0 + \beta_1 x_n^* + \varepsilon_n. \end{cases} \quad (7.4.5)$$

注意到  $e_1, \varepsilon_2, \dots, \varepsilon_n$  是均值为 0, 方差为  $\sigma_\varepsilon^2$  的不相关的随机误差, 故(7.4.5)式满足第三章经典线性回归模型的假定. 令  $y^* = (y_1^*, \dots, y_n^*)'$ ,

$$\mathbf{X}^* = \begin{bmatrix} \sqrt{1-\varphi^2} & x_1^* \\ 1-\varphi & x_2^* \\ \vdots & \vdots \\ 1-\varphi & x_n^* \end{bmatrix},$$

则  $\beta_0$  和  $\beta_1$  及  $\sigma_\varepsilon^2$  的最小二乘估计分别为

$$\boldsymbol{\beta}_{GL}^* = \begin{bmatrix} \beta_{0GL}^* \\ \beta_{1GL}^* \end{bmatrix} = (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{y}^*, \quad (7.4.6)$$

和

$$\sigma_{\varepsilon GL}^{2*} = \frac{1}{n} \sum_{i=1}^n (y_i^* - \beta_{0GL}^* - \beta_{1GL}^* x_i^*)^2. \quad (7.4.7)$$

实际上, (7.4.6) 和 (7.4.7) 式还可写成

$$\boldsymbol{\beta}_{GL}^* = (\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y}, \quad (7.4.8)$$

$$\sigma_{\varepsilon GL}^{2*} = \frac{1}{n} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_{GL}^*)' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_{GL}^*), \quad (7.4.9)$$

其中  $\mathbf{y} = (y_1, \dots, y_n)'$ ,

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

这正是我们第三章介绍的广义最小二乘估计. 但不要忘记在估计量 (7.4.6) 和 (7.4.7) 中还有一个未知的自回归系数  $\varphi$  需要估计. 因此, 对模型 (7.4.2) 实施广义最小二乘估计要分两步进行.

步骤 1 估计自回归系数  $\varphi$ . 这里有两种估计方法.

方法 I: 用前面介绍的最小二乘估计法, 求得  $\varphi$  的最小二乘估计  $\hat{\varphi}_L$ . 由例 7.3.1 可见, 对自回归系数  $\varphi$  最小二乘估计的精度是令人满意的.

方法 II: (i) 对一元线性回归模型做最小二乘估计, 求出估计  $\tilde{\beta}_0, \tilde{\beta}_1$ .

(ii) 令

$$\hat{e}_t = y_t - \tilde{\beta}_0 - \tilde{\beta}_1 x_t, \quad t = 1, \dots, n.$$

则

$$\hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2,$$

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}.$$

(iii) 由(7.2.13)式知,  $\varphi$  可由如下估计量来估计

$$\tilde{\varphi} = \hat{\gamma}_1 / \hat{\gamma}_0.$$

步骤2 用  $\varphi$  的估计量  $\hat{\varphi}_L$  或  $\tilde{\varphi}$  代替  $\Sigma^{-1}$  中的  $\varphi$ , 记作  $\hat{\Sigma}_L$  或  $\tilde{\Sigma}$ . 于是所求的参数  $\beta_0, \beta_1$  和  $\sigma_e^2$  的广义最小二乘估计为

$$\hat{\beta}_{GL} = \begin{pmatrix} \hat{\beta}_{0GL} \\ \hat{\beta}_{1GL} \end{pmatrix} = (\mathbf{X}' \hat{\Sigma}_L^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\Sigma}_L^{-1} \mathbf{y}, \quad (7.4.10)$$

和

$$\hat{\sigma}_{\varepsilon GL}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \hat{\beta}_{GL})' \hat{\Sigma}_L^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}_{GL}), \quad (7.4.11)$$

或

$$\tilde{\beta}_{GL} = \begin{pmatrix} \tilde{\beta}_{0GL} \\ \tilde{\beta}_{1GL} \end{pmatrix} = (\mathbf{X}' \tilde{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}' \tilde{\Sigma}^{-1} \mathbf{y}, \quad (7.4.12)$$

和

$$\tilde{\sigma}_{\varepsilon GL}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \tilde{\beta}_{GL})' \tilde{\Sigma}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\beta}_{GL}). \quad (7.4.13)$$

我们称上述广义最小二乘法为两步广义最小二乘估计.

下面我们通过两个例子来比较最小二乘估计(LSE)和两步广义最小二乘估计(GLSE)的精度, 以及两种不同的估计  $\varphi$  的方法所得的广义最小二乘估计的精度.

**例 7.4.2** 考虑例 7.4.1 中的三个具有 AR(1) 误差的一元线性回归模型. 选择样本容量  $n = 50, 100, 200$ , 每一种样本容量的选择产生 1 000 组数据. 对每组数据, 同时用最小二乘法求得  $\hat{\beta}_{0L}, \hat{\beta}_{1L}, \hat{\varphi}_L$  和  $\hat{\sigma}_{\varepsilon L}^2$ . 并把  $\hat{\varphi}_L$  作为  $\Sigma^{-1}$  中  $\varphi$  的估计, 按两步广义最小二乘估计(7.4.10)和(7.4.11), 求得  $\hat{\beta}_{0GL}, \hat{\beta}_{1GL}$  和  $\hat{\sigma}_{\varepsilon GL}^2$ . 计算估计值的平均值, 估计值与真参数值误差均方根. 结果列于表 7.4.2. 由表 7.4.2 的结果可知两步广义最小二乘估计确实改善了估计精度. 特别是  $\beta_0$  的估计.

表 7.4.2

模型	参数	$n = 50$		$n = 100$		$n = 200$	
		LSE	GLSE	LSE	GLSE	LSE	GLSE
I	$\beta_0$	1.544 2	1.559 1 (1.369 0)	1.471 7	1.492 5 (0.692 8)	1.457 9	1.511 5 (0.329 5)
	$\beta_1$	0.902 0 (0.007 0)	0.899 0 (0.004 8)	0.900 6 (0.003 5)	0.899 8 (0.002 3)	0.901 4 (0.001 8)	0.900 2 (0.001 1)
	$\varphi$	0.607 5 (0.017 7)	0.607 5 (0.017 7)	0.639 8 (0.006 6)	0.639 8 (0.006 6)	0.657 7 (0.002 9)	0.657 7 (0.002 9)
	$\sigma_e^2$	5.794 2 (1.646 0)	5.973 1 (1.586 1)	5.948 8 (0.825 7)	6.084 5 (0.799 0)	6.129 3 (0.406 7)	6.177 4 (0.401 3)
II	$\beta_0$	1.378 8	1.492 7 (4.010 9)	1.483 1	1.497 0 (1.019 0)	1.492 9	1.497 8 (0.359 6)
	$\beta_1$	0.902 1 (0.731 8)	0.909 9 (0.200 3)	0.907 5 (0.361 4)	0.901 4 (0.002 3)	0.889 9 (0.187 0)	0.900 0 (0.000 0)
	$\varphi$	0.575 0 (0.024 6)	0.575 0 (0.024 6)	0.625 9 (0.008 6)	0.625 9 (0.008 6)	0.647 5 (0.003 3)	0.647 5 (0.003 3)
	$\sigma_e^2$	5.706 4 (1.699 9)	5.856 0 (1.595 8)	5.985 9 (0.789 9)	6.051 7 (0.759 6)	6.079 4 (0.374 5)	6.112 2 (0.363 8)
III	$\beta_0$	1.503 3	1.502 0 (1.088 6)	1.457 4	1.472 5 (0.567 8)	1.470 9	1.478 6 (0.287 6)
	$\beta_1$	0.922 4 (0.908 2)	0.909 5 (0.093 6)	0.912 3 (0.402 1)	0.901 5 (0.039 8)	0.897 9 (0.207 1)	0.897 8 (0.019 9)
	$\varphi$	0.607 3 (0.017 7)	0.607 3 (0.017 7)	0.640 7 (0.006 8)	0.640 7 (0.006 8)	0.656 7 (0.003 2)	0.656 7 (0.003 2)
	$\sigma_e^2$	5.759 5 (1.585 7)	5.910 0 (1.489 4)	5.978 4 (0.836 5)	6.046 7 (0.807 3)	6.155 8 (0.400 5)	6.190 8 (0.396 8)

例 7.4.3 (续例 7.4.2) 仍然选择样本容量  $n = 50, 100, 200$ , 对每种样本容量产生 1 000 组数据, 对每组数据同时用两种广义最小二乘估计求参数  $\beta_0$ ,  $\beta_1$ ,  $\varphi$  和  $\sigma_e^2$  的估计. 我们记以方法 II 获得  $\varphi$  的估计而求的  $\hat{\beta}_{GL}$  和  $\hat{\sigma}_{eGL}^2$  为 GLSE1, 以方法 I 获得  $\varphi$  的估计而求得的  $\hat{\beta}_{GL}$  和  $\hat{\sigma}_{eGL}^2$  为 GLSE2. 模拟结果列于表 7.4.3.

由表 7.4.3 可知两种估计的精度相当,  $GLSE2$  稍优于  $GLSE1$ .

表 7.4.3

模型	参数	$n = 50$		$n = 100$		$n = 200$	
		$GLSE1$	$GLSE2$	$GLSE1$	$GLSE2$	$GLSE1$	$GLSE2$
I	$\beta_0$	1.522 0 (1.346 4)	1.518 5 (1.351 4)	1.492 9 (0.701 6)	1.493 7 (0.699 5)	1.479 3 (0.352 7)	1.479 4 (0.352 9)
	$\beta_1$	0.896 4 (0.004 9)	0.896 4 (0.008 4)	0.898 2 (0.002 3)	0.898 2 (0.002 3)	0.900 0 (0.001 2)	0.900 0 (0.001 2)
	$\varphi$	0.578 1 (0.022 1)	0.604 2 (0.018 8)	0.621 7 (0.008 7)	0.634 6 (0.007 7)	0.646 3 (0.003 3)	0.653 2 (0.003 1)
	$\sigma_e^2$	5.957 7 (1.657 5)	5.942 4 (1.652 4)	6.069 0 (0.758 9)	6.065 0 (0.758 9)	6.150 3 (0.376 5)	6.149 1 (0.376 7)
II	$\beta_0$	1.503 9 (4.217 0)	1.512 0 (4.203 2)	1.458 8 (1.060 8)	1.457 8 (1.059 5)	1.484 4 (0.385 4)	1.484 6 (0.385 5)
	$\beta_1$	0.901 3 (0.221 8)	0.899 3 (0.219 6)	0.901 2 (0.002 2)	0.901 2 (0.002 2)	0.900 0 (0.000 0)	0.900 0 (0.000 0)
	$\varphi$	0.559 1 (0.026 4)	0.581 5 (0.022 1)	0.612 5 (0.009 7)	0.622 2 (0.008 7)	0.644 7 (0.003 8)	0.649 9 (0.003 6)
	$\sigma_e^2$	5.840 3 (1.708 3)	5.832 3 (1.709 4)	6.041 6 (0.795 8)	6.040 7 (0.795 8)	6.152 6 (0.392 4)	6.152 3 (0.392 5)
III	$\beta_0$	1.498 2 (1.149 9)	1.496 8 (1.152 2)	1.487 5 (0.545 6)	1.487 4 (0.545 6)	1.471 3 (0.303 6)	1.471 3 (0.303 7)
	$\beta_1$	0.903 1 (0.092 0)	0.902 9 (0.092 1)	0.902 5 (0.040 9)	0.902 5 (0.040 9)	0.892 9 (0.020 2)	0.892 9 (0.020 2)
	$\varphi$	0.597 1 (0.018 2)	0.610 2 (0.017 4)	0.629 3 (0.007 8)	0.635 7 (0.007 6)	0.651 3 (0.003 2)	0.654 6 (0.003 1)
	$\sigma_e^2$	5.888 7 (1.621 9)	5.885 7 (1.621 6)	5.991 6 (0.861 5)	5.991 0 (0.861 6)	6.172 0 (0.408 7)	6.171 9 (0.408 6)

从以上讨论,我们可以发现,用两步广义最小二乘估计确实能改善参数  $\beta_0$ 、 $\beta_1$  和  $\sigma_e^2$  的估计精度,但这主要取决于  $\varphi$  的估计精度.而两步广义最小二乘估计对此没有贡献.如何提高  $\varphi$  的估计精度呢?我们可采用如下的循环算法来达到这一目的.并称由此得到的参数估计为循环广义最小二乘估计,其步骤如下:

步骤 1 用方法 II 获得两步广义最小二乘估计  $\tilde{\beta}_{GL}^{(1)}$ , 并由此得到  $RSS^{(1)}$ .

步骤 2 令

$$\tilde{e}_t^{(1)} = y_t - \tilde{\beta}_{0GL} - \tilde{\beta}_{1GL}x_t, \quad t = 1, \dots, n.$$

并计算

$$\tilde{\varphi}^{(1)} = \frac{\sum_{t=2}^n \tilde{e}_t^{(1)} \tilde{e}_t^{(2)}}{\sum_{t=1}^n (\tilde{e}_t^{(1)})^2}.$$

步骤 3 用  $\tilde{\varphi}^{(1)}$  作为  $\Sigma^{-1}$  中  $\varphi$  的估计, 再求广义最小二乘估计, 得  $\tilde{\beta}_{GL}^{(2)}$ , 及  $RSS^{(2)}$ .

步骤 4 如果  $|RSS^{(2)} - RSS^{(1)}| < 0.0001$ , 则  $\tilde{\beta}_{GL}^{(2)}$ ,  $\tilde{\varphi}^{(1)}$  和  $\tilde{\sigma}_{eGL}$  即为所求的估计. 否则重复步骤 2 和步骤 3, 直到精度达到要求为止.

下面我们来讨论具有  $AR(1)$  误差的多元线性回归模型 (7.4.1) 的参数估计, 即

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp-1} + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \end{cases} \quad (7.4.14)$$

其中  $\{\varepsilon_t, t = 1, \dots, n\}$  是零均值, 方差为  $\sigma_\varepsilon^2$  的独立同分布随机误差.

### 1. 最小二乘估计

类似于 (7.4.3) 式的导出, 由 (7.4.14) 式我们有

$$\varphi y_{t-1} = \varphi \beta_0 + \varphi \beta_1 x_{(t-1)1} + \dots + \varphi \beta_{p-1} x_{(t-1)p-1} + \varphi e_{t-1},$$

用它减去 (7.4.14) 式, 经整理得

$$\begin{aligned} y_t &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 x_{t1} + \dots + \alpha_p x_{tp-1} + \alpha_{p+1} x_{(t-1)1} \\ &\quad + \dots + \alpha_{2p-1} x_{(t-1)p-1} + \varepsilon_t \quad t = 2, 3, \dots, n, \end{aligned} \quad (7.4.15)$$

其中  $\alpha_0 = \beta_0(1 - \varphi)$ ,  $\alpha_1 = \varphi$ ,  $\alpha_i = \beta_{i-1}$ ,  $i = 2, \dots, p$ .

$$\alpha_{p+i} = -\varphi \beta_i, \quad i = 1, 2, \dots, p-1.$$

对 (7.4.15) 式, 利用样本观测值实施最小二乘估计, 即可得  $\hat{\alpha}_{iL}$ ,  $i = 0, 1, \dots, 2p-1$ , 和  $\hat{\sigma}_{eL}^2$ . 再令

$$\hat{\beta}_{0L} = \frac{\hat{\alpha}_{0L}}{1 - \hat{\alpha}_{1L}}, \quad \hat{\beta}_{iL} = \hat{\alpha}_{i+1L}, \quad i = 1, \dots, p-1, \quad \hat{\varphi}_L = \hat{\alpha}_{1L}.$$

### 2. 两步广义最小二乘估计

方法 I: 步骤 1. 用最小二乘估计求得  $\hat{\varphi}_L$ .

步骤 2. 用  $\hat{\varphi}_L$  作为  $\Sigma^{-1}$  中  $\varphi$  的估计, 记作  $\hat{\Sigma}_L^{-1}$ , 令

$$\mathbf{y} = (y_1, \dots, y_n)',$$

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{np-1} \end{pmatrix}.$$

则所求的广义最小二乘估计为

$$\hat{\boldsymbol{\beta}}_{GL} = \begin{pmatrix} \hat{\beta}_{0GL} \\ \hat{\beta}_{1GL} \\ \vdots \\ \hat{\beta}_{p-1GL} \end{pmatrix} = (\mathbf{X}' \hat{\boldsymbol{\Sigma}}_L^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Sigma}}_L^{-1} \mathbf{y}, \quad (7.4.16)$$

$$\hat{\sigma}_{eGL}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{GL})' \hat{\boldsymbol{\Sigma}}_L^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}_{GL}). \quad (7.4.17)$$

方法 II 步骤 1 对多元线性回归模型(7.4.14)实施最小二乘法获得参数  $\beta_0, \beta_1, \dots, \beta_{p-1}$  的估计, 记作  $\tilde{\beta}_0, \tilde{\beta}_1, \dots, \tilde{\beta}_{p-1}$ . 计算

$$\hat{e}_t = y_t - \tilde{\beta}_0 - \tilde{\beta}_1 x_{t1} - \dots - \tilde{\beta}_{p-1} x_{t,p-1}, t = 1, \dots, n.$$

$$\hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n \hat{e}_t^2,$$

$$\hat{\gamma}_1 = \frac{1}{n} \sum_{t=2}^n \hat{e}_t \hat{e}_{t-1},$$

由此得到  $\varphi$  的估计

$$\tilde{\varphi} = \hat{\gamma}_1 / \hat{\gamma}_0.$$

步骤 2 用  $\tilde{\varphi}$  作为  $\boldsymbol{\Sigma}^{-1}$  中  $\varphi$  的估计, 同方法 I 的步骤 2, 可得  $\beta_0, \beta_1$  和  $\sigma_e^2$  的广义最小二乘估计

$$\tilde{\boldsymbol{\beta}}_{GL} = \begin{pmatrix} \tilde{\beta}_{0GL} \\ \tilde{\beta}_{1GL} \\ \vdots \\ \tilde{\beta}_{p-1GL} \end{pmatrix} = (\mathbf{X}' \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{y}, \quad (7.4.18)$$

$$\tilde{\sigma}_{eGL}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}}_{GL})' \tilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{y} - \mathbf{X} \tilde{\boldsymbol{\beta}}_{GL}). \quad (7.4.19)$$

下面我们通过一个模拟例子来比较具有 AR(1) 误差的两元线性回归 LSE 和两步 GLSE 的估计精度.

**例 7.4.4** 考虑如下的线性回归模型



$$\begin{cases} y_t = 1.5 + 0.9x_{1t} + 0.54x_{2t} + e_t, \\ e_t = 0.67e_{t-1} + \varepsilon_t, \end{cases}$$

$\varepsilon_t, t = 1, \dots, n$  是独立同分布的正态  $N(0, 6.25)$  随机误差,  $x_{1t}$  和  $x_{2t}$  由如下三组随机模型生成:

$$\begin{aligned} \text{I:} & \begin{cases} x_{1t} = 5u_t, \\ x_{2t} = 2.7v_t. \end{cases} \\ \text{II:} & \begin{cases} x_{1t} = e^{9.01t} + u_t, \\ x_{2t} = 1 + 0.025t + v_t. \end{cases} \\ \text{III:} & \begin{cases} x_{1t} = \sin(3t) + u_t, \\ x_{2t} = 1 + 0.02t + v_t. \end{cases} \end{aligned}$$

其中  $u_t (t = 1, \dots, n)$  和  $v_t (t = 1, \dots, n)$  都是独立同分布的随机变量,  $u_t$  服从  $[0, 1]$  中均匀分布,  $v_t$  服从  $N(0, 1)$  分布. 样本容量  $n = 50, 100, 200$ , 模拟重复次数为 1 000. 对每种样本容量的选择, 计算 1 000 次, 运算所得的 1 000 个估计值的平均值和均方误差. 类似于例 7.3.1. 这里没有给出  $\beta_1$  的最小二乘估计的均方误差. 模拟结果列于表 7.4.4.

从表 7.4.4 可见, 结论与前面的例子是一致的, 即两步广义最小二乘估计的精度高于最小二乘估计.

表 7.4.4

模型	参数	n = 50		n = 100		n = 200	
		LSE	GLSE	LSE	GLSE	LSE	GLSE
I	$\beta_0$	1.505 2	1.490 7 (1.250 9)	1.509 0	1.504 9 (0.641 0)	1.483 4	1.516 9 (0.330 5)
	$\beta_1$	0.915 0 (0.067 4)	0.916 5 (0.045 2)	0.893 9 (0.030 8)	0.893 3 (0.020 0)	0.898 0 (0.015 6)	0.896 1 (0.010 4)
	$\beta_2$	0.536 9 (0.020 6)	0.538 3 (0.012 8)	0.539 9 (0.009 3)	0.546 1 (0.005 6)	0.536 4 (0.004 4)	0.544 1 (0.003 1)
	$\varphi$	0.609 1 (0.018 2)	0.609 1 (0.018 2)	0.635 9 (0.007 6)	0.635 9 (0.007 6)	0.656 1 (0.003 1)	0.656 1 (0.003 1)
	$\sigma_e^2$	5.487 8 (1.975 0)	5.783 3 (1.693 9)	5.852 8 (0.864 0)	5.992 0 (0.798 4)	6.071 7 (0.388 4)	6.137 3 (0.370 8)

续表

模型	参数	$n = 50$		$n = 100$		$n = 200$	
		LSE	GLSE	LSE	GLSE	LSE	GLSE
II	$\beta_0$	1.472 7	1.505 1 (4.486 0)	1.384 2	1.468 0 (2.989 8)	1.497 6	1.498 6 (1.183 2)
	$\beta_1$	0.856 6 (1.341 5)	0.850 8 (1.049 7)	0.878 4 (0.490 1)	0.866 7 (0.457 0)	0.883 5 (0.180 4)	0.895 6 (0.068 0)
	$\beta_2$	0.541 9 (0.148 0)	0.544 7 (0.105 1)	0.534 8 (0.054 9)	0.542 0 (0.013 2)	0.538 9 (0.027 9)	0.539 6 (0.014 6)
	$\varphi$	0.587 6 (0.021 8)	0.587 6 (0.021 8)	0.627 5 (0.008 0)	0.627 5 (0.008 0)	0.646 1 (0.003 5)	0.646 1 (0.003 5)
	$\sigma_e^2$	5.465 0 (1.928 3)	5.786 5 (1.652 1)	5.859 0 (0.866 0)	5.995 3 (0.792 4)	6.092 4 (0.398 1)	6.156 6 (0.371 4)
III	$\beta_0$	1.396 1	1.530 0 (1.225 8)	1.506 4	1.494 6 (0.734 3)	1.463 0	1.480 4 (0.444 4)
	$\beta_1$	0.927 9 (0.931 0)	0.905 0 (0.087 0)	0.906 1 (0.428 7)	0.898 9 (0.041 8)	0.901 4 (0.196 0)	0.899 8 (0.019 7)
	$\beta_2$	0.553 3 (0.139 1)	0.544 5 (0.020 0)	0.538 3 (0.052 8)	0.539 9 (0.015 1)	0.540 5 (0.019 8)	0.541 2 (0.015 4)
	$\varphi$	0.612 2 (0.017 4)	0.612 2 (0.017 4)	0.633 9 (0.007 6)	0.633 9 (0.007 6)	0.650 9 (0.003 4)	0.650 9 (0.003 4)
	$\sigma_e^2$	5.483 4 (1.842 8)	5.777 5 (1.538 6)	5.870 4 (0.845 1)	6.010 0 (0.778 5)	6.073 1 (0.406 0)	6.137 0 (0.387 3)

现在我们来讨论模型(7.4.1)的几种参数估计的性质. 令  $y = (y_1, \dots, y_n)'$ ,  $e = (e_1, \dots, e_n)'$ ,

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np-1} \end{bmatrix}.$$

则(7.4.1)式可以写成如下向量形式

$$y = \mathbf{X}\beta + e. \quad (7.4.20)$$

如果我们忽略掉 AR(1) 误差, 而用最小二乘法估计参数  $\beta = (\beta_0, \beta_1, \dots, \beta_{p-1})'$ , 并记其为  $\hat{\beta}_{LS}$ . 则由第三章知

$$\hat{\beta}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y. \quad (7.4.21)$$

于是由(7.4.20)可得

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{LS}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{e}) \\ &= \boldsymbol{\beta}, \end{aligned}$$

即  $\hat{\boldsymbol{\beta}}_{LS}$  仍是  $\boldsymbol{\beta}$  的无偏估计, 且估计的误差协方差阵为

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta})' &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}\mathbf{e}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= \frac{\sigma_{\varepsilon}^2}{1 - \varphi^2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \end{aligned} \quad (7.4.22)$$

其中

$$E(e_i^2) = \frac{\sigma_{\varepsilon}^2}{1 - \varphi^2}.$$

当模型(7.4.1)中自回归系数已知时,  $\boldsymbol{\beta}$  的广义最小二乘估计为

$$\boldsymbol{\beta}^* = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}. \quad (7.4.23)$$

从而有

$$\begin{aligned} E(\boldsymbol{\beta}^*) &= E[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})] \\ &= \boldsymbol{\beta} + (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}E(\mathbf{e}) \\ &= \boldsymbol{\beta}. \end{aligned}$$

即  $\boldsymbol{\beta}^*$  也是  $\boldsymbol{\beta}$  的无偏估计, 且估计的误差协方差阵为

$$\begin{aligned} E(\boldsymbol{\beta}^* - \boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})' &= E[(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{e}\mathbf{e}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}] \\ &= \frac{\sigma_{\varepsilon}^2}{1 - \varphi^2}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}. \end{aligned} \quad (7.4.24)$$

为了比较估计(7.4.21)和(7.4.23)的优劣, 我们采用估计的均方误差准则, 即

$$MSE(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

把上式做进一步分解可得

$$\begin{aligned} MSE(\hat{\boldsymbol{\theta}}) &= E(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= E[(\hat{\boldsymbol{\theta}} - E\hat{\boldsymbol{\theta}}) + (E\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]'[(\hat{\boldsymbol{\theta}} - E\hat{\boldsymbol{\theta}}) + (E\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})] \\ &= E(\hat{\boldsymbol{\theta}} - E\hat{\boldsymbol{\theta}})'(\hat{\boldsymbol{\theta}} - E\hat{\boldsymbol{\theta}}) + \|E\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \\ &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\theta}})) + \|E\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2. \end{aligned}$$

当  $\hat{\boldsymbol{\theta}}$  为  $\boldsymbol{\theta}$  的无偏估计时, 上式成为

$$MSE(\hat{\boldsymbol{\theta}}) = \text{tr}(\text{Cov}(\hat{\boldsymbol{\theta}})). \quad (7.4.25)$$

**定理 7.4.1** 在模型(7.4.1)的假定下

$$MSE(\boldsymbol{\beta}^*) \leq MSE(\hat{\boldsymbol{\beta}}_{LS}), \quad (7.4.26)$$

即在均方误差意义下,  $\boldsymbol{\beta}^*$  优于  $\hat{\boldsymbol{\beta}}_{LS}$ .

**证明** 注意到  $\hat{\boldsymbol{\beta}}_{LS}$  和  $\boldsymbol{\beta}^*$  都是  $\boldsymbol{\beta}$  的无偏估计, 由(7.4.25)式得到

$$\begin{aligned} MSE(\hat{\boldsymbol{\beta}}_{LS}) &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{LS})) \\ &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{LS}) - \text{Cov}(\boldsymbol{\beta}^*)) + \text{tr}(\text{Cov}(\boldsymbol{\beta}^*)) \\ &= MSE(\boldsymbol{\beta}^*) + \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{LS}) - \text{Cov}(\boldsymbol{\beta}^*)). \end{aligned}$$

因此为证明(7.4.26)式, 只需证明

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{LS}) - \text{Cov}(\boldsymbol{\beta}^*) \geq 0. \quad (7.4.27)$$

由(7.4.22)和(7.4.24)有

$$\begin{aligned} &\text{Cov}(\hat{\boldsymbol{\beta}}_{LS}) - \text{Cov}(\boldsymbol{\beta}^*) \\ &= \frac{\sigma_\varepsilon^2}{1 - \varphi^2} [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}] \\ &= \frac{\sigma_\varepsilon^2}{1 - \varphi^2} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{\frac{1}{2}} (\mathbf{I} - \mathbf{P}) \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \end{aligned}$$

其中  $\mathbf{P} = \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-\frac{1}{2}}$ .

由于  $\mathbf{P}'\mathbf{P} = \mathbf{P}$ , 即  $\mathbf{P}$  为幂等阵, 从而由矩阵论知识知

$$\mathbf{I} - \mathbf{P} \geq 0.$$

注意到  $|\varphi| < 1$ , 便得到(7.4.27)式, 定理证毕.

定理 7.4.1 表明, 对具有 AR(1)误差的线性回归模型(7.4.1), 应采用广义最小二乘估计法求参数  $\boldsymbol{\beta}$  的估计.

下面的定理给出了两种两步最小二乘估计的误差协方差阵的表达式. 由于这个定理涉及较多的时间序列分析的知识, 我们只叙述结果, 而不给出证明. 有兴趣的读者可参阅文献 King 等(1987).

**定理 7.4.2** 在模型(7.4.1)中, 进一步假定  $\varepsilon_t$  是正态的, 则  $\hat{\boldsymbol{\beta}}_{GL}$  和  $\tilde{\boldsymbol{\beta}}_{GL}$  都是  $\boldsymbol{\beta}$  的无偏估计, 且估计的误差协方差阵均为

$$\frac{\sigma_\varepsilon^2}{1 - \varphi^2} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} + \frac{\sigma_\varepsilon^2}{n\varphi^2} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} + o(n^{-2}), \quad (7.4.28)$$

其中  $\boldsymbol{\Omega} = \boldsymbol{\Sigma} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'$ .

比较(7.4.28)和(7.4.24)两式, 我们有

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{GL}) - \text{Cov}(\boldsymbol{\beta}^*)$$

$$= \frac{\sigma_\varepsilon^2}{n\varphi} (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} + o(n^{-2}).$$

注意到

$$\begin{aligned} \boldsymbol{\Sigma} - \mathbf{X}'(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X} \\ = \boldsymbol{\Sigma}^{\frac{1}{2}}[\mathbf{I} - \boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-\frac{1}{2}}]\boldsymbol{\Sigma}^{\frac{1}{2}} \end{aligned}$$

是半正定矩阵,从而有

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{GL}) - \text{Cov}(\boldsymbol{\beta}^*) \geq 0.$$

类似于定理 7.4.1 的证明,有

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{GL}) \geq \text{MSE}(\boldsymbol{\beta}^*).$$

同理有

$$\text{MSE}(\tilde{\boldsymbol{\beta}}_{GL}) \geq \text{MSE}(\boldsymbol{\beta}^*).$$

这表明,当  $\varphi$  未知时,用其估计量  $\hat{\varphi}_{LS}$  和  $\tilde{\varphi}_{LS}$  代替  $\boldsymbol{\Sigma}^{-1}$  中的  $\varphi$  后,所得的两步广义最小二乘估计的均方误差将变大.

从例 7.4.1—例 7.4.4 我们看到,  $\varphi$  的估计精度对回归参数估计精度有影响.而理论上分析这一点有困难.下面我们对不同的  $\varphi$  值计算如下称为相对效率的量,以衡量不同  $\varphi$  值下不同参数估计方法的优劣.

$$\begin{aligned} m_1 &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{LS}))/\text{tr}(\text{Cov}(\boldsymbol{\beta}^*)), \\ m_2 &= \text{tr}(\text{Cov}(\hat{\boldsymbol{\beta}}_{GL}))/\text{tr}(\text{Cov}(\boldsymbol{\beta}^*)), \\ m_1^* &= \text{tr}(\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{LS}))/\text{tr}(\hat{\text{Cov}}(\boldsymbol{\beta}^*)), \\ m_2^* &= \text{tr}(\hat{\text{Cov}}(\tilde{\boldsymbol{\beta}}_{GL}))/\text{tr}(\hat{\text{Cov}}(\boldsymbol{\beta}^*)), \\ m_3^* &= \text{tr}(\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{GL}))/\text{tr}(\hat{\text{Cov}}(\boldsymbol{\beta}^*)), \end{aligned}$$

其中  $\text{Cov}(\boldsymbol{\beta}^*)$ ,  $\text{Cov}(\hat{\boldsymbol{\beta}}_{LS})$  和  $\text{Cov}(\hat{\boldsymbol{\beta}}_{GL})$  分别由 (7.4.22) 式、(7.4.24) 式和 (7.4.28) 式给出.  $\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{LS})$ ,  $\hat{\text{Cov}}(\boldsymbol{\beta}^*)$ ,  $\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{GL})$  和  $\hat{\text{Cov}}(\tilde{\boldsymbol{\beta}}_{GL})$  为对应的参数估计值的样本协方差,比如

$$\hat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{LS}) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_{0LS}^{(i)} - \beta_0)^2 & \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_{0LS}^{(i)} - \beta_0)(\hat{\beta}_{1LS}^{(i)} - \beta_1) \\ \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_{0LS}^{(i)} - \beta_0)(\hat{\beta}_{1LS}^{(i)} - \beta_1) & \frac{1}{n} \sum_{i=1}^n (\hat{\beta}_{1LS}^{(i)} - \beta_1)^2 \end{pmatrix},$$

其中  $(\hat{\beta}_{0LS}^{(i)}, \hat{\beta}_{1LS}^{(i)})'$  为第  $i$  次估计.我们所采用的模型分别为

$$\text{模型 I: } \begin{cases} y_t = 1.5 + 0.9x_t + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \end{cases}$$

其中  $\varphi$  的取值分别为 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95.  $\{\varepsilon_t, t=1, \dots, n\}$  为独立同分布  $N(0, 1)$  随机误差.  $x_t$  由如下三个模型生成:

$$\text{I - (i)} \quad x_t = 1 + 0.025t + u_t,$$

$$\text{I - (ii)} \quad x_t = e^{0.01t} + u_t,$$

$$\text{I - (iii)} \quad x_t = \sin(3t) + u_t,$$

其中  $\{u_t, t=1, \dots, n\}$  是独立同服从  $[0, 1]$  上均匀分布.

### 模型 II

$$\begin{cases} y_t = 1.5 + 0.9x_{t1} + 0.54x_{t2} + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \end{cases}$$

其中  $x_{ti} (i=1, 2)$  由如下两组模型生成:

$$\text{II - (i)} \quad \begin{cases} x_{t1} = e^{0.01t} + u_t, \\ x_{t2} = 1 + 0.025t + v_t, \end{cases}$$

$$\text{II - (ii)} \quad \begin{cases} x_{t1} = \sin(3t) + u_t, \\ x_{t2} = 1 + 0.025t + v_t, \end{cases}$$

其中  $\{\varepsilon_t\}, \{u_t\}$  和  $\varphi$  如模型 I 假定.  $\{v_t\}$  为独立同分布  $N(0, 1)$  随机误差.

选择样本容量  $n = 25, 50, 100, 150, 200$ , 对每种不同的样本容量选择, 模拟重复 1 000 次, 计算  $m_i$  的均值. 对  $\hat{\text{Cov}}(\hat{\beta}_{LS}), \hat{\text{Cov}}(\hat{\beta}_{GL})$  和  $\hat{\text{Cov}}(\bar{\beta}_{GL})$ , 对不同的样本容量选择, 先产生 1 000 组数据, 计算估计的样本协方差, 然后将此过程重复 1 000 次, 计算  $m_i^*$  的均值. 模拟结果表明:

(i) 当  $\varphi < 0.2$  时,  $\hat{\beta}_{LS}$  的均方误差比  $\hat{\beta}_{GL}$  和  $\bar{\beta}_{GL}$  小. 因此对这种情形, 我们建议, 忽略掉 AR(1) 误差存在, 采用最小二乘估计  $\hat{\beta}_{LS}$ .

(ii) 当  $\varphi > 0.9$  时, 误差结构已近似于时间序列中的一类特殊模型——求和模型. 对这种模型在处理上与 AR 模型有较大的差别, 有关参数估计的统计性质也与 AR 模型有较大的差别. 因此对这种情形, 我们不推荐两步广义最小二乘估计.

(iii) 当  $0.2 \leq \varphi \leq 0.9$  时,  $\hat{\beta}_{GL}$  和  $\bar{\beta}_{GL}$  比  $\hat{\beta}_{LS}$  的均方误差小, 而且随着样本量的增加越来越小. 因此在这种情形, 两步广义最小二乘估计可获得理想结果. 另外, 在大多数情况下,  $\hat{\beta}_{GL}$  的均方误差小于  $\bar{\beta}_{GL}$ . 这与前面的数值模拟结果是一致的.

当  $-1 < \varphi < 0$  时, 模拟结果与  $0 < \varphi < 1$  类似. 至此, 我们已对具有 AR(1) 误差的线性回归模型的参数估计作了详尽的讨论. 但在实际中, 首要问题是如何发现模型具有 AR 误差.

## § 7.5 对一阶自回归误差的假设检验

对具有一阶自回归(AR(1))误差的线性回归模型(7.4.1)即

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_{p-1} x_{tp-1} + e_t, \\ e_t = \varphi e_{t-1} + \varepsilon_t, \quad t = 1, \cdots, n, \end{cases} \quad (7.5.1)$$

其中  $\{\varepsilon_t, t = 1, \cdots, n\}$  是独立同分布的随机误差,  $E(\varepsilon_t) = 0, E(\varepsilon_t^2) = \sigma_\varepsilon^2 < \infty, |\varphi| < 1$ .

在模型(7.5.1)下, 自相关性的检验可通过检验如下的统计假设来实现:

$$\begin{cases} \text{零假设 } H_0: \varphi = 0, \\ \text{备择假设 } H_1: \varphi \neq 0. \end{cases} \quad (7.5.2)$$

对上述统计假设, 常用的检验方法是  $D-W$  方法, 它是 Durbin 和 Watson 于 1951 年提出的, 其相应检验统计量称为  $d$  统计量, 定义为

$$d = \frac{\sum_{t=2}^n (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^n \hat{e}_t^2}, \quad (7.5.3)$$

式中  $\hat{e}_t, t = 1, \cdots, n$  是由普通最小二乘估计所得的残差估计, 即

$$\hat{e}_t = y_t - \hat{\beta}_{OLS} - \hat{\beta}_{1LS} x_{t1} - \cdots - \hat{\beta}_{(p-1)LS} x_{tp-1},$$

其中  $\hat{\beta}_{LS}$  的记号如 § 7.4.

由(7.5.3)式并注意到  $e_t$  的宽平稳性, 我们有

$$\begin{aligned} d &\approx 2 - 2 \frac{\sum_{t=2}^n \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^n \hat{e}_t^2} \\ &= 2(1 - \hat{\varphi}). \end{aligned}$$

由于  $-1 < \hat{\varphi} < 1$ , 因此

$$0 < d < 4.$$

当  $\varphi = 0$  时,

$$d \approx 2.$$

为了检验(7.5.2), 我们需要求出  $d$  统计量的分布. 但是  $d$  的分布依赖于设计阵  $\mathbf{X}$ , 因此我们不仅无法得到  $d$  的精确分布, 就是渐近分布也很难求得. 在给定的显著性水平下, Durbin 和 Watson 给出了  $d$  的上界  $d_U$  和下界  $d_L$  的表(见附表), 按如下方法作出决策:

(i) 如果  $d < d_L$ , 则拒绝  $H_0: \varphi = 0$ , 而认为  $\varphi > 0$  是显著的.

(ii) 如果  $d > d_U$ , 则不拒绝  $H_0: \varphi = 0$ .

(iii) 如果  $d_L \leq d \leq d_U$ , 检验无结论.

对负自相关情形, 在实际问题中遇到不多. 如果需要检验假设  $H_0: \varphi = 0$ , 与  $H_1: \varphi < 0$ , 则可用统计量

$$\bar{d} = 4 - d, \quad (7.5.4)$$

式中  $d$  由(7.5.3)式定义. 决策方法如下:

(i) 如果  $\bar{d} < d_L$ , 则拒绝  $H_0: \varphi = 0$ , 而认为  $\varphi < 0$  是显著的.

(ii) 如果  $\bar{d} > d_U$ , 则不拒绝  $H_0: \varphi = 0$ .

(iii) 如果  $d_L \leq \bar{d} \leq d_U$ , 检验无结论.

我们可以通过图 7.5.1 来形象地说明  $D-W$  检验的否定域.

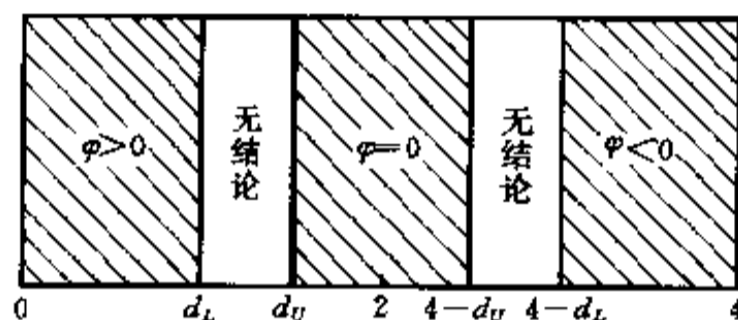


图 7.5.1

对显著性水平为 0.05 和 0.01, 书末给出了详细的  $D-W$  检验的上、下界值的表.

**例 7.5.1** 一个果汁饮料公司希望利用该公司 1960 年至 1979 年这 20 年的饮料销售量和广告费用的数据, 研究广告对销售量的作用. 数据列于表 7.5.1.

表 7.5.1 果汁饮料公司数据

年份	序号	年度销售量( $y_t$ )	广告费用(1 000\$)( $x_t$ )
1960	1	3 083	75
1961	2	3 149	78
1962	3	3 218	80
1963	4	3 239	82
1964	5	3 295	84
1965	6	3 374	88
1966	7	3 475	93
1967	8	3 569	97
1968	9	3 597	99
1969	10	3 725	104
1970	11	3 794	109
1971	12	3 959	115
1972	13	4 043	120



续表

年份	序号	年度销售量( $y_t$ )	广告费用(1 000\$)( $x_t$ )
1973	14	4 194	127
1974	15	4 318	135
1975	16	4 493	144
1976	17	4 683	153
1977	18	4 850	161
1978	19	5 005	170
1979	20	5 236	182

对表 7.5.1 中的数据拟合如下二元线性回归:

$$y_t = \beta_0 + \beta_1 x_t + e_t. \quad (7.5.5)$$

结果总结于表 7.5.2.

表 7.5.2 模型(7.5.5)估计结果

变量	估计值	标准误差估计 $s.e$	$t$ 值
$x$	20.091	0.142 8	140.71
常数	1 608 508	17 022 3	94.49
$RSS = 421.548 5$		$R^2 = 0.999 1$	

残差散点图由图 7.5.2 给出. 从图中残差的分布可见, 它们有一个有限的先向上, 然后再向下的漂移. 因此我们怀疑残差有自相关结构.

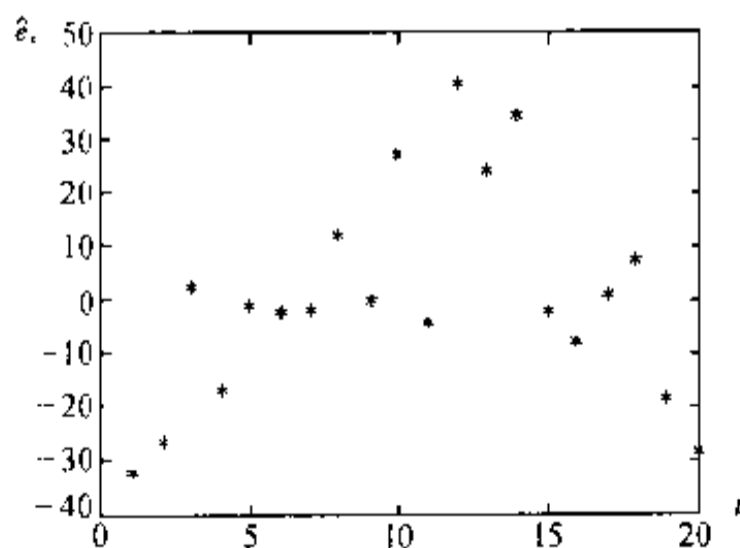


图 7.5.2  $\hat{e}_t$  的散点图

按(7.5.3)式计算  $d$  统计量的值,得  $\hat{d} = 1.08$ . 如果我们选择显著性水平为 0.05,则由附录 3,对应于  $n = 20$  和一个回归变量可查得  $d_L = 1.20, d_U = 1.41$ . 由于  $\hat{d} < d_L$ ,故拒绝  $H_0$ ,而认为“误差具有正的自相关”是显著的.由  $\hat{\varphi}$  的估计式,我们有

$$\hat{\varphi} = 0.3531.$$

按 § 7.4,这时我们可以用两步广义最小二乘估计去求  $\beta_0$  和  $\beta_1$  的估计.

**例 7.5.2(续例 7.5.1)** 例 7.5.1 是考虑广告对饮料销售的作用.实际上,一个地区饮料的销售量与当地的居民人数有很大的关系.因此,在本例中,我们把当地居民人口也作为一个回归自变量,数据列于表 7.5.3.

表 7.5.3 果汁饮料公司数据

年份	序号	年度销售量( $y_t$ )	年度广告费用( $x_{1t}$ , (100 0\$))	居民人数( $x_{2t}$ , (人))
1960	1	3 083	75	825 000
1961	2	3 149	78	830 445
1962	3	3 218	80	838 750
1963	4	3 239	82	842 940
1964	5	3 295	84	846 315
1965	6	3 374	88	852 240
1966	7	3 475	93	960 760
1967	8	3 569	97	865 925
1968	9	3 597	99	871 640
1969	10	3 725	104	877 745
1970	11	3 794	109	886 520
1971	12	3 959	115	894 500
1972	13	4 043	120	900 400
1973	14	4 194	127	904 005
1974	15	4 318	135	908 525
1975	16	4 493	144	912 160
1976	17	4 683	153	917 630
1977	18	4 850	161	922 220
1978	19	5 005	170	925 910
1979	20	5 236	182	929 610

对表 7.5.3 的数据,拟合如下二元线性回归:

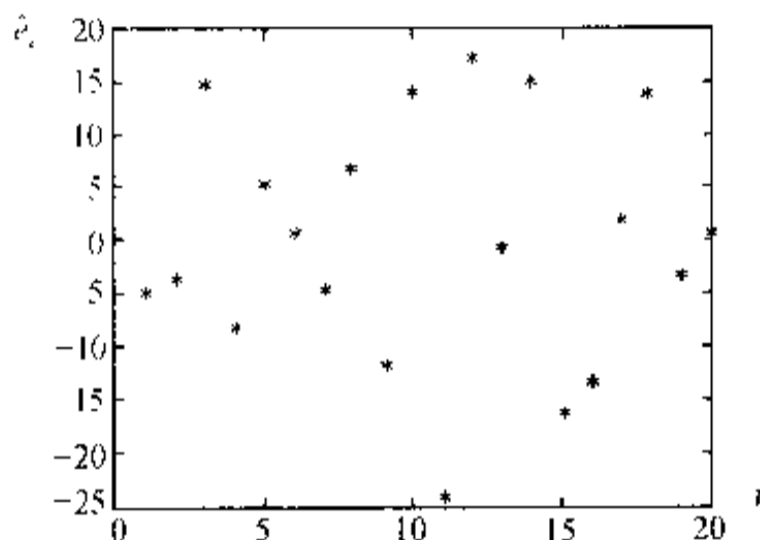
$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + e_t, \quad (7.5.6)$$

结果列于表 7.5.4.

表 7.5.4 模型(7.5.6)估计结果

变量	估计值	标准误差估计 $s.e$	$t$ 值
$x_1$	18.434	0.2915	63.23
$x_2$	0.002	0.0003	5.93
常数	320.340	217.3278	1.47
RSS = 145.3408		$R^2 = 0.9997$	

残差散点图见图 7.5.3.

图 7.5.3  $\hat{e}_i$  的散点图

从图中  $\hat{e}_i$  的分布看, 例 7.5.1 中那种先增后降的特征明显地消失了. 计算  $d$  统计量的值得

$$\hat{d} = 1.4470.$$

显然  $\hat{d} > d_U$ . 因此我们不能拒绝  $H_0$ , 即  $\hat{e}_i$  不存在正自相关. 又因  $\hat{d} < 4 - d_U$ , 故  $\hat{e}_i$  不存在负自相关. 于是我们认为  $\varphi = 0$  在显著性水平为 0.05 下是显著的. 由  $\varphi$  的估计式可得

$$\hat{\varphi} = 0.1812.$$

按 § 7.4 的讨论, 这时对模型(7.5.6)进行普通最小二乘估计的结果是可信的.

比较例 7.5.1 和例 7.5.2 发现, 当我们根据实际问题的背景, 增加对  $y$  有重要影响的回归自变量, 有可能消除误差的自相关性. 这表明模型自相关性产生的原因可能是由于丢失回归自变量而引起的.

## § 7.6 半相依线性回归模型

在本节之前,我们主要讨论具有如下结构的线性回归模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + e_i, \quad i = 1, \cdots, n,$$

其中误差满足下面的假设之一:

(1)  $e_i, i = 1, \cdots, n$  不相关,且  $E(e_i) = 0, E(e_i^2) = \sigma^2 < \infty$ ;

(2)  $e_i, i = 1, \cdots, n$  不相关,且  $E(e_i) = 0, E(e_i^2) = \sigma_i^2 < \infty$ ,

而  $\sigma_i^2, i = 1, \cdots, n$  不全相等;

(3)  $e_i$  服从  $AR(p)$  模型,即具有自相关误差.

在一些实际问题中,我们的研究往往要涉及多个以某种统计关系相联系在一起的线性回归模型.下面通过一个实际例子来引入这类模型.

考虑如下三个对数线性商品需求模型

$$\begin{cases} \ln q_{1i} = \beta_{10} + \beta_{11} \ln p_{1i} + \beta_{12} \ln y_i + e_{1i}, \\ \ln q_{2i} = \beta_{20} + \beta_{21} \ln p_{2i} + \beta_{22} \ln y_i + e_{2i}, \\ \ln q_{3i} = \beta_{30} + \beta_{31} \ln p_{3i} + \beta_{32} \ln y_i + e_{3i}, \end{cases} \quad i = 1, \cdots, n. \quad (7.6.1)$$

在(7.6.1)式中,我们假定第  $k$  类商品的需求量  $q_{ki}$  依赖于自身的价格  $p_{ki}$  和国民的个人收入  $y_i$ . 一般说来,  $q_{ki}$  还依赖于别的类商品的价格.这一点在模型(7.6.1)式中没有得到直接反映.

记  $\beta_k = (\beta_{k0}, \beta_{k1}, \beta_{k2})', e_k = (e_{k1}, e_{k2}, \cdots, e_{kn})'$ ,

$$y_k = \begin{pmatrix} \ln q_{k1} \\ \ln q_{k2} \\ \vdots \\ \ln q_{kn} \end{pmatrix}, \quad X_k = \begin{pmatrix} 1 & \ln p_{k1} & \ln y_1 \\ 1 & \ln p_{k2} & \ln y_2 \\ \vdots & \vdots & \vdots \\ 1 & \ln p_{kn} & \ln y_n \end{pmatrix}, \quad k = 1, 2, 3.$$

则模型(7.6.1)可写成如下形式:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} X_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & X_3 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}. \quad (7.6.2)$$

对模型(7.6.2),我们做如下假定:

(1)  $E(e_{ki}) = 0, \quad i = 1, \cdots, n; \quad k = 1, 2, 3. \quad (7.6.3)$

(2)

$$\begin{cases} \text{Var}(e_{1i}) = \sigma_{11}, \\ \text{Var}(e_{2i}) = \sigma_{22}, \\ \text{Var}(e_{3i}) = \sigma_{33}, \end{cases} \quad i = 1, \dots, n. \quad (7.6.4)$$

$$(3) \quad \text{Cov}(e_{ji}, e_{ki}) = E(e_{ji}e_{ki}) = \sigma_{jk}, \quad (7.6.5)$$

$$j, k = 1, 2, 3; \quad i = 1, \dots, n,$$

$$(4) \quad \text{Cov}(e_{j_1i_1}, e_{k_2i_2}) = 0, \quad i_1 \neq i_2, j, k = 1, 2, 3. \quad (7.6.6)$$

这样在误差假定(7.6.3)、(7.6.4)和(7.6.6)下,(7.6.1)式中的每个方程都是一个经典线性回归模型,而误差假定(7.6.5)又把这三个经典线性回归模型联系在一起,故而模型(7.6.2)在(7.6.3)—(7.6.6)下得名半相依线性回归模型.实际上,假定(7.6.5)在一定程度上反映了别的类商品的价格对  $q_{ki}$  的影响.

进一步,我们还可以把假定(7.6.3)—(7.6.6)写得更为简洁.

记  $e = (e'_1, e'_2, e'_3)'$ , 则有

$$E(e) = \mathbf{0},$$

$$V = E(ee') = \begin{pmatrix} Ee_1e'_1 & Ee_1e'_2 & Ee_1e'_3 \\ Ee_2e'_1 & Ee_2e'_2 & Ee_2e'_3 \\ Ee_3e'_1 & Ee_3e'_2 & Ee_3e'_3 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \sigma_{13}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \sigma_{23}\mathbf{I} \\ \sigma_{31}\mathbf{I} & \sigma_{32}\mathbf{I} & \sigma_{33}\mathbf{I} \end{pmatrix}$$

$$= \Sigma \otimes \mathbf{I}, \quad (7.6.7)$$

式中符号“ $\otimes$ ”为矩阵的 Kronecker 乘积(见附录 1),

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix}.$$

推而广之,我们可考虑更一般形式的半相依线性回归模型,即

$$\begin{cases} y_{1i} = \beta_{10} + \beta_{11}x_{1i,1} + \dots + \beta_{1p_1}x_{1i,p_1} + e_{1i}, \\ y_{2i} = \beta_{20} + \beta_{21}x_{2i,1} + \dots + \beta_{2p_2}x_{2i,p_2} + e_{2i}, \\ \vdots \\ y_{Mi} = \beta_{M0} + \beta_{M1}x_{Mi,1} + \dots + \beta_{Mp_M}x_{Mi,p_M} + e_{Mi}, \end{cases} \quad i = 1, \dots, n.$$

用矩阵方式可把上述  $M$  个线性回归方程写成

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & & & 0 \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ 0 & & & \mathbf{X}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix}, \quad (7.6.8)$$

其中  $\mathbf{y}_k = (y_{k1}, \dots, y_{kn})'$ ,  $\mathbf{e}_k = (e_{k1}, \dots, e_{kn})'$ ,

$$\boldsymbol{\beta}_k = (\beta_{k0}, \beta_{k1}, \dots, \beta_{kp_k})',$$

$$\mathbf{X}_k = \begin{pmatrix} 1 & x_{k1,1} & \cdots & x_{k1,p_k} \\ 1 & x_{k2,1} & \cdots & x_{k2,p_k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{kn,1} & \cdots & x_{kn,p_k} \end{pmatrix},$$

$$k = 1, \dots, M,$$

记  $\mathbf{e} = (e_1', \dots, e_M')$ , 并假定

$$E(\mathbf{e}) = \mathbf{0}, \mathbf{V} = E(\mathbf{e}\mathbf{e}') = \boldsymbol{\Sigma} \otimes \mathbf{I}, \text{ 且 } \boldsymbol{\Sigma} > \mathbf{0}, \quad (7.6.9)$$

其中

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ \vdots & \vdots & & \vdots \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{pmatrix}.$$

(7.6.8)和(7.6.9)组成的模型就是本节将要讨论的半相依线性回归模型. 进一步,如果我们记

$$\mathbf{y} = (y_1', y_2', \dots, y_m')', \boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \dots, \boldsymbol{\beta}_M')',$$

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & & & 0 \\ & \mathbf{X}_2 & & \\ & & \ddots & \\ 0 & & & \mathbf{X}_M \end{pmatrix},$$

则(7.6.8)式又可写为

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}.$$

在获得观测样本  $(\mathbf{y}_k, \mathbf{X}_k)_{n \times (p_k+2)}$ ,  $k = 1, \dots, M$  后,如果我们忽略  $\boldsymbol{\Sigma}$  不是对角阵这一事实,即可获得最小二乘估计

$$\hat{\beta}_{UL} = (X'X)^{-1}X'y, \quad (7.6.10)$$

并容易算得

$$E(\hat{\beta}_{UL}) = \beta, \quad (7.6.11)$$

$$\text{Cov}(\hat{\beta}_{UL}) = (X'X)^{-1}X'VX(X'X)^{-1}. \quad (7.6.12)$$

由我们前面章节所介绍的知识知, 鉴于  $V(>0)$  为非对角阵, 即  $\sigma_{jk} \neq 0, j \neq k$ , 因而应该用广义最小二乘估计法求参数  $\beta$  的估计. 当  $V$  已知时, 我们有

$$\begin{aligned} \hat{\beta}_{UG} &= (X'V^{-1}X)^{-1}X'V^{-1}y \\ &= (X'(\Sigma^{-1} \otimes I)X)^{-1}X'(\Sigma^{-1} \otimes I)y. \end{aligned} \quad (7.6.13)$$

容易验证

$$E(\hat{\beta}_{UG}) = \beta, \quad (7.6.14)$$

$$\begin{aligned} \text{Cov}(\hat{\beta}_{UG}) &= (X'V^{-1}X)^{-1} \\ &= (X'(\Sigma^{-1} \otimes I)X)^{-1}. \end{aligned} \quad (7.6.15)$$

因此, 由(7.6.11)和(7.6.14)两式知  $\hat{\beta}_{UL}$  和  $\hat{\beta}_{UG}$  都是  $\beta$  的无偏估计. 但是, 如果记

$$A = I - V^{-\frac{1}{2}}X(X'V^{-1}X)^{-1}X'V^{-\frac{1}{2}},$$

则由(7.6.12)和(7.6.15)两式得

$$\begin{aligned} \text{Cov}(\hat{\beta}_{UL}) - \text{Cov}(\hat{\beta}_{UG}) &= (X'X)^{-1}X'VX(X'X)^{-1} - (X'V^{-1}X)^{-1} \\ &= (X'X)^{-1}X'V^{\frac{1}{2}}AV^{\frac{1}{2}}X(X'X)^{-1} \geq 0. \end{aligned}$$

于是, 从均方误差准则的角度看

$$\begin{aligned} &MSE(\hat{\beta}_{UL}) - MSE(\hat{\beta}_{UG}) \\ &= E(\hat{\beta}_{UL} - \beta)'(\hat{\beta}_{UL} - \beta) - E(\hat{\beta}_{UG} - \beta)'(\hat{\beta}_{UG} - \beta) \\ &= \text{tr}(\text{Cov}(\hat{\beta}_{UL})) - \text{tr}(\text{Cov}(\hat{\beta}_{UG})) \geq 0, \end{aligned}$$

即在均方误差准则下,  $\hat{\beta}_{UG}$  优于  $\hat{\beta}_{UL}$ .

我们再来研究一下模型(7.6.8)的结构. 显然, 如果  $\sigma_{jk} = 0, j \neq k$ , 则

$$V = \begin{bmatrix} \sigma_{11} & & & 0 \\ & \sigma_{22} & & \\ & & \ddots & \\ 0 & & & \sigma_{MM} \end{bmatrix} \otimes I,$$

从而

$$\mathbf{V}^{-1} = \begin{pmatrix} \frac{1}{\sigma_{11}} & & & 0 \\ & \frac{1}{\sigma_{22}} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_{MM}} \end{pmatrix} \otimes \mathbf{I}.$$

注意到设计矩阵  $\mathbf{X}$  的特殊结构, 我们有

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{UG} &= \begin{pmatrix} \frac{1}{\sigma_{11}} (\mathbf{X}'_1 \mathbf{X}_1) & & & 0 \\ & \frac{1}{\sigma_{22}} (\mathbf{X}'_2 \mathbf{X}_2) & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_{MM}} (\mathbf{X}'_M \mathbf{X}_M) \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sigma_{11}} \mathbf{X}'_1 y_1 \\ \frac{1}{\sigma_{22}} \mathbf{X}'_2 y_2 \\ \vdots \\ \frac{1}{\sigma_{MM}} \mathbf{X}'_M y_M \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & & & 0 \\ & \mathbf{X}'_2 \mathbf{X}_2 & & \\ & & \ddots & \\ 0 & & & \mathbf{X}'_M \mathbf{X}_M \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'_1 y_1 \\ \mathbf{X}'_2 y_2 \\ \vdots \\ \mathbf{X}'_M y_M \end{pmatrix} \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= \hat{\boldsymbol{\beta}}_{UL}. \end{aligned}$$

这一结果是不令人惊奇的. 实际上, 当  $\sigma_{jk} = 0 (j \neq k)$  时, 模型 (7.6.8) 是由  $M$  个“不相关”的经典线性回归模型组成的回归方程. 因此, 在  $e$  服从正态分布  $N(\mathbf{0}, \text{diag}(\sigma_{11}, \sigma_{12}, \dots, \sigma_{MM}) \otimes \mathbf{I})$  的假定下,  $\hat{\boldsymbol{\beta}}_{UL}$  就是  $\boldsymbol{\beta}$  的一致最小方差无偏估计, 故应和广义最小二乘估计是一致的.

如果假定 (7.6.8) 式中的  $M$  个回归模型都有相同的回归变量, 即

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{X}_2 = \dots = \mathbf{X}_M = \mathbf{Z}, \\ p_1 &= p_2 = \dots = p_M = p, \end{aligned} \quad (7.6.16)$$

则不论  $\boldsymbol{\Sigma}$  是否为对角阵, 我们都有  $\hat{\boldsymbol{\beta}}_{UG} = \hat{\boldsymbol{\beta}}_{UL}$ .

事实上, 在假定 (7.6.16) 下, 利用矩阵 Kronecker 乘积的运算, 我们有

$$\mathbf{X} = \begin{pmatrix} \mathbf{Z} & & 0 \\ & \ddots & \\ 0 & & \mathbf{Z} \end{pmatrix} = \mathbf{I} \otimes \mathbf{Z},$$

这样, 由 (7.6.10) 式可得



$$\begin{aligned}
\hat{\beta}_{UL} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= ((\mathbf{I} \otimes \mathbf{Z})'(\mathbf{I} \otimes \mathbf{Z}))^{-1}(\mathbf{I} \otimes \mathbf{Z})'\mathbf{y} \\
&= (\mathbf{I} \otimes (\mathbf{Z}'\mathbf{Z}))^{-1}(\mathbf{I} \otimes \mathbf{Z}')\mathbf{y} \\
&= (\mathbf{I} \otimes (\mathbf{Z}'\mathbf{Z})^{-1})(\mathbf{I} \otimes \mathbf{Z}')\mathbf{y} \\
&= (\mathbf{I} \otimes (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y}.
\end{aligned}$$

而由(7.6.13)式,我们有

$$\begin{aligned}
\hat{\beta}_{UG} &= ((\mathbf{I} \otimes \mathbf{Z}')(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})(\mathbf{I} \otimes \mathbf{Z}))^{-1}(\mathbf{I} \otimes \mathbf{Z}')(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y} \\
&= (\boldsymbol{\Sigma} \otimes (\mathbf{Z}'\mathbf{Z})^{-1})(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Z}')\mathbf{y} \\
&= (\mathbf{I} \otimes (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y},
\end{aligned}$$

故  $\hat{\beta}_{UG} = \hat{\beta}_{UL}$ .

这一结论在一定程度上反映了半相依回归模型(7.6.8)的特征.

如果上述两种情形都不成立,  $\hat{\beta}_{UG}$  和  $\hat{\beta}_{UL}$  的差别受什么影响呢? 下面我们来研究一个  $M=2$ , 且  $p_1 = p_2 = 1$  的半相依回归模型, 即

$$\begin{cases} y_{1i} = \beta_{10} + \beta_{11}x_{1i} + e_{1i}, \\ y_{2i} = \beta_{20} + \beta_{21}x_{2i} + e_{2i}, \end{cases} \quad i = 1, \dots, n,$$

且

$$\begin{cases} E(e_{1i}) = 0, E(e_{2i}^2) = \sigma_{kk}^2, \\ E(e_{1i}e_{2i}) = \sigma_{12} \neq 0, \\ \text{Cov}(e_{j_1}, e_{k_2}) = 0, i_1 = i_2; j, k = 1, 2. \end{cases}$$

这时, 有

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{pmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{pmatrix},$$

记

$$\mathbf{y}_k = \begin{pmatrix} y_{k1} \\ \vdots \\ y_{kn} \end{pmatrix}, \quad \mathbf{X}_k = \begin{pmatrix} 1 & x_{k1} \\ \vdots & \vdots \\ 1 & x_{kn} \end{pmatrix}, \quad \boldsymbol{\beta}_k = \begin{pmatrix} \beta_{k0} \\ \beta_{k1} \end{pmatrix}, \quad k = 1, 2.$$

则

$$\begin{aligned}
\hat{\beta}_{UG} &= \begin{pmatrix} \hat{\beta}_{1UG} \\ \hat{\beta}_{2UG} \end{pmatrix} \\
&= \begin{pmatrix} \sigma_{22}(\mathbf{X}'_1\mathbf{X}_1) & -\sigma_{12}(\mathbf{X}'_1\mathbf{X}_2) \\ -\sigma_{12}(\mathbf{X}'_2\mathbf{X}_1) & \sigma_{11}(\mathbf{X}'_2\mathbf{X}_2) \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{22}(\mathbf{X}'_1\mathbf{y}_1) & -\sigma_{12}(\mathbf{X}'_1\mathbf{y}_2) \\ -\sigma_{12}(\mathbf{X}'_2\mathbf{y}_1) & \sigma_{11}(\mathbf{X}'_2\mathbf{y}_2) \end{pmatrix}.
\end{aligned}$$

如果我们再记

$$\bar{y}_k = \frac{1}{n} \sum_{i=1}^n y_{ki}, \quad \bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ki},$$

$$m_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{ik} - \bar{x}_k),$$

则由(7.6.15)式,可计算得  $\hat{\beta}_{11UG}$  和  $\hat{\beta}_{21UG}$  的协方差矩阵为

$$\begin{aligned} \text{Cov} \begin{pmatrix} \hat{\beta}_{11UG} \\ \hat{\beta}_{21UG} \end{pmatrix} &= \begin{pmatrix} \text{Var}(\hat{\beta}_{11UG}) & \text{Cov}(\hat{\beta}_{11UG}, \hat{\beta}_{21UG}) \\ \text{Cov}(\hat{\beta}_{11UG}, \hat{\beta}_{21UG}) & \text{Var}(\hat{\beta}_{21UG}) \end{pmatrix} \\ &= (\sigma_{11}\sigma_{22} - \sigma_{12}^2) \begin{pmatrix} \sigma_{22}m_{11} & -\sigma_{12}m_{12} \\ -\sigma_{12}m_{12} & \sigma_{11}m_{22} \end{pmatrix}^{-1}, \end{aligned}$$

由此得到

$$\text{Var}(\hat{\beta}_{11UG}) = \frac{(\sigma_{11}\sigma_{22} - \sigma_{12}^2)\sigma_{11}m_{22}}{\sigma_{11}\sigma_{22}m_{11}m_{22} - \sigma_{12}^2m_{12}^2}. \quad (7.6.17)$$

如果令

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}},$$

表示  $e_{1i}$  和  $e_{2i}$  的相关系数,而令

$$\gamma_{12} = \frac{m_{12}}{\sqrt{m_{11}m_{22}}}$$

表示  $x_{1i}$  和  $x_{2i}$  的样本相关系数,则(7.6.17)式又可表为

$$\text{Var}(\hat{\beta}_{11UG}) = \frac{\sigma_{11}}{m_{11}} \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 \gamma_{12}^2}. \quad (7.6.18)$$

用类似的方法可算得

$$\text{Var}(\hat{\beta}_{11UL}) = \frac{\sigma_{11}}{m_{11}}.$$

于是

$$c = \frac{\text{Var}(\hat{\beta}_{11UG})}{\text{Var}(\hat{\beta}_{11UL})} = \frac{1 - \rho_{12}^2}{1 - \rho_{12}^2 \gamma_{12}^2} \leq 1. \quad (7.6.19)$$

由(7.6.19)式可见,当  $\rho_{12} = 0$ , 即  $\sigma_{12} = 0$ , 或  $\gamma_{12} = 1$ , 即  $X_1 = X_2$ , 我们有  $c = 1$ . 这正是我们前面的结论. 对  $0 < \rho_{12} < 1$  和  $0 \leq \gamma_{12} < 1$ . 表 7.6.1 列出了  $c$  的值. 从表中结果可见,  $\rho_{12}$  越大,  $\text{Var}(\hat{\beta}_{11UG})$  相对于  $\text{Var}(\hat{\beta}_{11UL})$  就越小. 而对固定的  $\rho_{12}$ ,  $\gamma_{12}$  越小,  $\text{Var}(\hat{\beta}_{11UG})$  相对于  $\text{Var}(\hat{\beta}_{11UL})$  也越小.

表 7.6.1 (7.6.19)式中  $c$  的值

$\hat{y}_{12}$	$\rho_{12}^2$				
	0	0.100 0	0.200 0	0.300 0	0.400 0
0	1.000 0	0.900 0	0.800 0	0.700 0	0.600 0
0.100 0	1.000 0	0.909 1	0.816 3	0.721 6	0.625 0
0.200 0	1.000 0	0.918 4	0.833 3	0.744 7	0.652 2
0.300 0	1.000 0	0.927 8	0.851 1	0.769 2	0.681 8
0.400 0	1.000 0	0.937 5	0.869 6	0.795 5	0.714 3
0.500 0	1.000 0	0.947 4	0.888 9	0.823 5	0.750 0
0.600 0	1.000 0	0.957 4	0.909 1	0.853 7	0.789 5
0.700 0	1.000 0	0.967 7	0.930 2	0.886 1	0.833 3
0.800 0	1.000 0	0.978 3	0.952 4	0.921 1	0.882 4
0.900 0	1.000 0	0.989 0	0.975 6	0.958 9	0.937 5
1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0
$\hat{y}_{12}$	$\rho_{12}^2$				
	0.500 0	0.600 0	0.700 0	0.800 0	0.900 0
0	0.500 0	0.400 0	0.300 0	0.200 0	0.100 0
0.100 0	0.526 3	0.425 5	0.322 6	0.217 4	0.109 9
0.200 0	0.555 6	0.454 5	0.348 8	0.238 1	0.122 0
0.300 0	0.588 2	0.487 8	0.379 7	0.263 2	0.137 0
0.400 0	0.625 0	0.526 3	0.416 7	0.294 1	0.156 3
0.500 0	0.666 7	0.571 4	0.461 5	0.333 3	0.181 8
0.600 0	0.714 3	0.625 0	0.517 2	0.384 6	0.217 4
0.700 0	0.769 2	0.689 7	0.588 2	0.454 5	0.270 3
0.800 0	0.833 3	0.769 2	0.681 8	0.555 6	0.357 1
0.900 0	0.909 1	0.869 6	0.810 8	0.714 3	0.526 3
1.000 0	1.000 0	1.000 0	1.000 0	1.000 0	1.000 0

当  $\Sigma$  未知时, 必须用其估计来代替(7.6.13)中的  $\Sigma$ , 从而获得参数  $\beta$  的估计. 我们可按如下步骤进行:

步骤 1 用观测样本  $(y_k, X_k)_{n \times (p_k+2)}$ ,  $k=1, \dots, M$ , 对(7.6.8)式中每个回归模型分别进行最小二乘估计, 得  $\beta_k$  的最小二乘估计.

$$\hat{\beta}_{kL} = (X_k' X_k)^{-1} X_k' y_k, \quad k=1, \dots, M.$$

步骤 2 以下式估计  $e_{ki}$ :

$$\hat{e}_{ki} = y_{ki} - \mathbf{X}_{ki} \hat{\boldsymbol{\beta}}_{kL}, \quad i = 1, \dots, n; k = 1, \dots, M, \quad (7.6.20)$$

式中  $\mathbf{X}_{ki}$  为  $\mathbf{X}_i$  的第  $i$  行向量  $(1, x_{ki,1}, \dots, x_{ki,p_k-1})$ . 这时  $\sigma_{jk}$  可用下式估计:

$$\hat{\sigma}_{jk} = \frac{1}{n} \sum_{i=1}^n \hat{e}_{ji} \hat{e}_{ki}, \quad j, k = 1, \dots, M.$$

步骤 3 以

$$\hat{\boldsymbol{\Sigma}} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1M} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2M} \\ \vdots & \vdots & & \vdots \\ \hat{\sigma}_{M1} & \hat{\sigma}_{M2} & \cdots & \hat{\sigma}_{MM} \end{pmatrix}$$

作为  $\boldsymbol{\Sigma}$  的估计, 代入(7.6.13)式得

$$\hat{\boldsymbol{\beta}}_{UGL} = (\mathbf{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{X})^{-1} \mathbf{X}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{y}. \quad (7.6.21)$$

可以证明(见安鸿志等(1983)), 如果  $e_{ki}$  的概率分布是对称的, 则

$$E(\hat{\boldsymbol{\beta}}_{UGL}) = \boldsymbol{\beta}. \quad (7.6.22)$$

如果进一步假定  $(e_{1i}, \dots, e_{Mi})$  独立且服从向量正态分布  $N(\mathbf{0}, \boldsymbol{\Sigma})$ , 且  $\boldsymbol{\Sigma} > 0$ . 则

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{UGL}) = \text{Cov}(\hat{\boldsymbol{\beta}}_{UG}) + o\left(\frac{1}{n}\right). \quad (7.6.23)$$

(7.6.22)和(7.6.23)表明, 如果对误差的分布作适当限制,  $\hat{\boldsymbol{\beta}}_{UGL}$  具有较好的统计性质, 下面通过一个数值模拟例子来比较  $\hat{\boldsymbol{\beta}}_{UL}$ ,  $\hat{\boldsymbol{\beta}}_{UG}$  和  $\hat{\boldsymbol{\beta}}_{UGL}$  的估计精度.

**例 7.6.1** 考虑如下含有两个方程的半相依线性回归模型:

$$\begin{cases} y_{1i} = 1 + 1.56x_{1i,1} + 1.38x_{1i,2} + e_{1i}, \\ y_{2i} = 1 + 1.72x_{2i,1} + 1.14x_{2i,2} + e_{2i}, \end{cases} \quad (7.6.24)$$

其中  $(e_{1i}, e_{2i}), i = 1, \dots, n$  独立且服从二维正态分布  $N(\mathbf{0}, \boldsymbol{\Sigma})$ , 这里

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1.9154 & 1.9189 \\ 1.9189 & 3.5149 \end{pmatrix}.$$

$(e_{1i}, e_{2i})$  的生成方法为, 先由计算机产生两个相互独立的正态  $N(0, 1)$  随机数  $\varepsilon_{1i}, \varepsilon_{2i}, i = 1, \dots, n$ , 然后作如下线性变换:

$$\begin{cases} e_{1i} = 0.77\varepsilon_{1i} + 1.15\varepsilon_{2i}, \\ e_{2i} = 1.82\varepsilon_{1i} + 0.45\varepsilon_{2i}, \end{cases}$$

即可得到所要的二维正态  $N(\mathbf{0}, \boldsymbol{\Sigma})$  的样本.  $x_{1i,j}, x_{2i,j} (j = 1, 2)$  由如下模型生

成:

(1)

$$\begin{cases} x_{1i,1} = 1 + 0.025i + u_i, \\ x_{1i,2} = \sin(3i) + v_i, \end{cases}$$

(2)

$$\begin{cases} x_{2i,1} = 1 + 0.025i + u_i, \\ x_{2i,2} = \exp(0.01i) + v_i, \end{cases}$$

式中  $u_i (i = 1, \dots, n)$  独立且服从  $N(0, 1)$  分布,  $v_i (i = 1, \dots, n)$  独立且服从  $[0, 1]$  上中心化的均匀分布.

我们选择样本容量  $n = 15, 25, 50, 100, 150, 200$ , 对每种样本容量的选择, 模拟重复 1000 次, 计算这 1000 次模拟中参数估计的样本均值和样本均方误差. 结果列于表 7.6.2 中. 模拟结果表明, 在均方误差意义下,  $\hat{\beta}_{UGL}$  优于  $\hat{\beta}_{UL}$ , 而且随着样本容量  $n$  的增大,  $\hat{\beta}_{UGL}$  的估计精度越来越接近  $\hat{\beta}_{UG}$ . 如何衡量这种接近程度呢? 类似于 § 7.4, 我们定义如下相对效率:

$$c_1 = \frac{\text{tr}(\text{Cov}(\hat{\beta}_{UL}))}{\text{tr}(\text{Cov}(\hat{\beta}_{UG}))},$$

$$c_2 = \frac{\text{tr}(\text{Cov}(\hat{\beta}_{UGL}))}{\text{tr}(\text{Cov}(\hat{\beta}_{UG}))},$$

$c_1$  或  $c_2$  愈接近于 1, 表明  $\hat{\beta}_{UL}$  或  $\hat{\beta}_{UGL}$  与  $\hat{\beta}_{UG}$  愈接近. 我们用  $\hat{\text{Cov}}(\hat{\beta}_{UL})$ 、 $\hat{\text{Cov}}(\hat{\beta}_{UG})$  和  $\hat{\text{Cov}}(\hat{\beta}_{UGL})$  表示  $\hat{\beta}_{UL}$ 、 $\hat{\beta}_{UG}$  和  $\hat{\beta}_{UGL}$  的样本协方差, 其定义为 (以  $\hat{\beta}_{UL}$  为例)

$$\hat{\text{Cov}}(\hat{\beta}_{UL}) = \left( \frac{1}{l} \sum_{i=1}^l (\hat{\beta}_{ijUL}^{(i)} - \hat{\beta}_{ij})(\hat{\beta}_{kmUL}^{(i)} - \hat{\beta}_{km}) \right),$$

$$1 \leq i, k \leq 2, \quad 0 \leq j, m \leq 2.$$

于是  $c_1$  和  $c_2$  的估计可由下式得到

$$\hat{c}_1 = \frac{\text{tr}(\hat{\text{Cov}}(\hat{\beta}_{UL}))}{\text{tr}(\hat{\text{Cov}}(\hat{\beta}_{UG}))},$$

$$C_2 = \frac{\text{tr}(\widehat{\text{Cov}}(\hat{\beta}_{UGL}))}{\text{tr}(\widehat{\text{Cov}}(\hat{\beta}_{UG}))}$$

表 7.6.3 列出了对  $n = 15, 25, 50, 75, 100$  的模拟结果, 其中  $l = 100$ , 模拟重复 1 000 次.

表 7.6.2 三种估计的比较

样本容量	参数	$\hat{\beta}_{UG}$		$\hat{\beta}_{UL}$		$\hat{\beta}_{UGL}$	
		均值	均方误差	均值	均方误差	均值	均方误差
$n = 15$		1.012 3	0.567 6	1.022 2	0.730 6	1.010 9	0.583 6
		1.570 1	0.276 0	1.561 8	0.407 7	1.571 1	0.291 0
		0.789 8	0.100 7	0.797 6	0.215 3	0.790 1	0.112 8
		1.074 3	2.797 8	1.056 0	5.390 5	1.077 6	3.178 0
		0.852 8	1.444 6	0.875 4	3.213 6	0.848 7	1.694 8
		1.083 6	0.275 5	1.081 1	0.294 0	1.086 4	0.277 6
$n = 25$		0.983 5	0.374 3	0.974 8	0.509 0	0.985 5	0.383 5
		1.574 0	0.163 8	1.580 9	0.236 6	1.572 6	0.168 5
		0.794 6	0.078 4	0.784 5	0.155 6	0.795 8	0.082 2
		1.010 9	1.655 7	1.029 1	2.979 9	1.011 8	1.759 0
		0.862 9	0.689 7	0.850 9	1.500 8	0.863 1	0.745 6
		1.108 5	0.172 9	1.109 4	0.179 8	1.108 2	0.173 1
$n = 50$		1.001 0	0.200 7	0.985 2	0.276 6	0.997 6	0.204 0
		1.559 6	0.057 9	1.569 2	0.086 3	1.561 7	0.058 5
		0.808 2	0.031 2	0.798 0	0.065 0	0.806 2	0.031 8
		0.994 8	0.652 6	0.982 2	1.056 6	0.991 2	0.673 2
		0.872 3	0.176 6	0.882 2	0.351 2	0.874 8	0.182 2
		1.099 4	0.074 0	1.097 2	0.077 5	1.099 1	0.074 2
$n = 100$		1.014 0	0.113 1	1.013 0	0.141 9	1.014 1	0.113 5
		1.552 7	0.018 6	1.553 1	0.024 3	1.552 6	0.018 7
		0.799 7	0.012 7	0.798 9	0.023 4	0.799 7	0.012 8
		1.030 0	0.255 0	1.024 6	0.344 6	1.030 8	0.257 6
		0.862 4	0.038 9	0.865 4	0.073 3	0.862 0	0.039 8
		1.095 3	0.029 8	1.094 4	0.036 2	1.095 5	0.030 0
$n = 150$		1.009 5	0.079 8	1.003 3	0.089 5	1.008 6	0.079 8
		1.558 9	0.007 8	1.561 1	0.009 0	1.559 2	0.007 8
		0.802 6	0.007 7	0.798 7	0.012 5	0.802 0	0.007 8
		1.003 2	0.147 2	0.997 6	0.184 2	1.001 4	0.147 5
		0.870 9	0.020 9	0.873 4	0.035 0	0.872 0	0.021 1
		1.099 0	0.018 6	1.098 5	0.023 7	1.098 4	0.018 8
$n = 200$		1.001 5	0.056 1	1.000 1	0.058 1	1.001 4	0.056 0
		1.558 4	0.003 7	1.558 8	0.003 9	1.558 5	0.003 7
		0.800 9	0.004 8	0.799 1	0.007 4	0.801 1	0.004 8
		1.002 6	0.104 8	0.995 6	0.133 8	1.002 8	0.105 2
		0.868 1	0.014 6	0.874 7	0.026 0	0.867 9	0.014 7
		1.099 2	0.008 6	1.094 3	0.013 1	1.099 4	0.008 6

表 7.6.3 估计比较

样本容量	$\hat{c}_1$	$\hat{c}_2$
15	3.916 0	1.585 2
25	3.675 9	1.171 1
50	3.088 9	1.062 2
75	2.093 1	1.044 0
100	1.973 7	1.024 8

表中结果显示,  $\hat{\beta}_{UGL}$  在样本容量  $n \geq 25$  时,  $\hat{c}_2$  已相当接近于 1, 于是估计精度已相当令人满意. 因此, 对半相依回归模型,  $\hat{\beta}_{UGL}$  是非常有效的.

对于一般半相依回归模型, 王松桂(1988)提出了一种新的估计. 它不仅具有一定的优良性质, 而且便于在计算机上实现.

以上讨论主要是针对如何根据观测样本来估计半相依回归模型的参数, 现在我们来讨论半相依回归模型的假设检验问题.

对半相依回归模型(7.6.8), 使用者最关心的一个假设检验问题是:

原假设  $H_0: \sigma_{jk} = 0, j \neq k, j, k = 1, \dots, M.$

备择假设  $H_1: \sigma_{jk}$  中至少有一个不为零.

如果对给定的显著性水平,  $H_0$  被接受, 则意味着我们只需对每个回归方程分别进行建模即可. 否则, 若原假设  $H_0$  被拒绝, 则需要采用一个半相依回归模型.

检验统计量是依据拉格朗日乘子检验原理(见 § 7.2)来构造的. 这里略去推导过程, 而只把检验  $H_0$  的步骤总结如下.

步骤 1 根据样本  $(y_k, \mathbf{X}_k), k = 1, \dots, M$ , 用最小二乘法拟合半相依回归模型(7.6.8), 得估计  $\hat{\beta}_{UL}$ . 由此计算误差  $e_{ki}$  的估计值

$$\hat{e}_{ki} = y_{ki} - \mathbf{X}_{ki} \hat{\beta}_{UL},$$

$$k = 1, \dots, M; \quad i = 1, \dots, n.$$

步骤 2 计算误差的样本协方差

$$\hat{\sigma}_{jk} = \frac{1}{n} \sum_{i=1}^n \hat{e}_{ji} \hat{e}_{ki}, \quad j, k = 1, \dots, M.$$

令

$$\hat{\gamma}_{jk}^2 = \frac{\hat{\sigma}_{jk}^2}{\hat{\sigma}_{jj} \hat{\sigma}_{kk}}, \quad j, k = 1, \dots, M.$$

步骤 3 计算检验统计量

$$\hat{\lambda}_n = n \sum_{j=2}^M \sum_{k=1}^{j-1} \hat{\gamma}_{jk}^2. \quad (7.6.25)$$

在  $H_0$  下可以证明,  $\hat{\lambda}_n$  渐近于  $\chi_{\frac{M(M-1)}{2}}^2$  分布, 即自由度为  $M(M-1)/2$  的  $\chi^2$  分布. 因此在给定的显著性水平  $\alpha$  下, 可以用这个渐近分布来检验  $H_0$ . 若拒绝  $H_0$ , 则认为  $\sigma_{jk}$  中有一些值不为零.

最后, 我们再回到本节开头提出的商品需求模型.

**例 7.6.2** 表 7.6.4 是三种商品的需求量  $q_i$ , 价格  $p_i$  和收入的数据.

表 7.6.4 商品需求量模型数据

$p_1$	$p_2$	$p_3$	$y$	$q_1$	$q_2$	$q_3$
10.763	4.474	6.629	487.648	11.632	13.194	45.770
13.033	10.836	13.774	364.877	12.029	2.181	13.393
9.244	5.856	4.063	541.037	8.916	5.586	104.819
4.605	14.010	3.868	760.343	33.908	5.231	137.269
13.045	11.417	14.922	421.746	4.561	10.930	15.914
7.706	8.755	14.318	578.214	17.594	11.854	23.667
7.405	7.317	4.794	561.734	18.842	17.045	62.057
7.519	6.360	3.768	301.470	11.637	2.682	52.262
8.764	4.188	8.089	379.636	7.645	13.008	31.916
13.511	1.996	2.708	478.855	7.881	19.623	123.026
4.943	7.268	12.901	433.741	9.614	6.534	26.255
8.360	5.839	11.115	525.702	9.067	9.397	35.540
5.721	5.160	11.220	513.067	14.070	13.188	32.487
7.225	9.145	5.810	408.666	15.474	3.340	45.838
6.617	5.034	5.516	192.061	3.041	4.716	26.867
14.219	5.926	3.707	462.621	14.096	17.141	43.325
6.769	8.187	10.125	312.659	4.118	4.695	24.330
7.769	7.193	2.471	400.848	10.489	7.639	107.017
9.804	13.315	8.976	392.215	6.231	9.089	23.407
11.063	6.874	12.883	377.724	6.458	10.346	18.254
6.535	15.533	4.115	343.552	8.736	3.901	54.895
11.063	4.477	4.962	301.599	5.158	4.350	45.360
4.016	9.231	6.294	294.112	16.618	7.371	25.318
4.759	5.907	8.298	365.032	11.342	6.507	32.852
5.483	7.077	9.638	256.125	2.903	3.770	22.154
7.890	9.942	7.122	184.798	3.138	1.360	20.575
8.460	7.043	4.157	359.084	15.315	6.497	44.205
6.195	4.142	10.040	629.378	22.240	10.963	44.443
6.743	3.369	15.459	306.527	10.012	10.140	13.251
11.977	4.806	6.172	347.488	3.982	8.637	41.845

对数据考虑拟合如下半相依线性回归模型



$$\ln q_{ki} = \beta_{k0} + \beta_{k1} \ln p_{ki} + \beta_{k2} \ln y_i + \varepsilon_{ki},$$

$$k = 1, 2, 3; \quad i = 1, \dots, 30.$$

首先,我们由表 7.6.4 中的数据求最小二乘估计,得

$$\hat{\beta}_{10} = -5.71, \quad \hat{\beta}_{11} = -0.566, \quad \hat{\beta}_{12} = 1.434,$$

$$\hat{\beta}_{20} = -3.64, \quad \hat{\beta}_{21} = -0.648, \quad \hat{\beta}_{22} = 1.144,$$

$$\hat{\beta}_{30} = 0.274, \quad \hat{\beta}_{31} = -0.964, \quad \hat{\beta}_{32} = 0.871.$$

其次,我们利用统计量  $\hat{\lambda}_n$  (见(7.6.25)式)检验假设

$$H_0: \sigma_{12} = \sigma_{31} = \sigma_{32} = 0.$$

经计算得

$$\hat{\gamma}_{12}^2 = 0.0143, \quad \hat{\gamma}_{31}^2 = 0.3701, \quad \hat{\gamma}_{32}^2 = 0.2400.$$

于是

$$\hat{\lambda}_n = 30(0.0143 + 0.3701 + 0.24) = 18.73.$$

取显著性水平  $\alpha = 0.05$ , 由于  $M = 3$ , 故  $\hat{\lambda}_n$  渐近于  $\chi_3^2$  分布, 查  $\chi^2$  分布表得  $\chi_3^2(0.05) = 7.81$ , 显然,  $\hat{\lambda}_n > \chi_3^2(0.05)$ , 所以拒绝  $H_0$ , 即认为  $\sigma_{12}, \sigma_{31}$  和  $\sigma_{32}$  不全为零是统计显著的.

对数据用广义最小二乘法进行拟合, 得参数估计

$$\hat{\beta}_{10G} = -4.57, \quad \hat{\beta}_{11G} = -0.909, \quad \hat{\beta}_{12G} = 1.452,$$

$$\hat{\beta}_{20G} = -3.19, \quad \hat{\beta}_{21G} = -0.865, \quad \hat{\beta}_{22G} = 1.137,$$

$$\hat{\beta}_{30G} = 0.352, \quad \hat{\beta}_{31G} = -0.999, \quad \hat{\beta}_{32G} = 0.869.$$

最后,我们还需考察模型估计结果是否与有关的经济理论符合. 由估计结果可见, 价格变量的回归系数都是负的, 这意味着商品价格越高, 人们对该商品的需求量越少. 即在其它条件不变时, 商品需求量与其自身价格成反比, “薄利多销”正是这种关系的一个形象比喻. 收入变量的回归系数全都是正的, 这表明, 收入越高, 人们对商品的需求越高. 因此, 估计结果符合经济的有关理论. 至此, 建模工作结束, 我们可以根据需要, 利用所得模型去对商品的需求量进行预报等.

## 附录 1 关于矩阵的若干基础知识

为了方便读者,下面我们罗列了本书中用到的矩阵代数的部分事实.除了定理 A.2.1 之外,其余都给出了详细证明.

### 1. 正交阵

设  $\Phi$  为  $n \times n$  实矩阵,满足  $\Phi' \Phi = I_n$ ,则称  $\Phi$  为正交阵.若记  $\Phi = (\varphi_1, \dots, \varphi_n)$ ,则从  $\Phi' \Phi = I_n$  可知

$$\varphi_i' \varphi_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

因此,正交阵的  $n$  个列向量是相互正交的,并且长度均为 1.

若  $\Phi$  为正交阵,从定义容易证明: $\Phi' = \Phi^{-1}$ .即对正交阵,它的转置阵就是它的逆矩阵.

设  $x = (x_1, \dots, x_n)'$  为  $n \times 1$  向量.记  $\|x\| = \left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}}$ ,表示向量  $x$  的长度.设  $\Phi$  为正交阵,记  $y = \Phi x$ ,则

$$\|y\| = \|\Phi x\| = (x' \Phi' \Phi x)^{\frac{1}{2}} = (x' x)^{\frac{1}{2}} = \|x\|.$$

也就是说,在正交变换  $y = \Phi x$  下, $y$  与  $x$  具有相同的长度.

### 2. 实对称阵、半正定阵、正定阵

设  $A = (a_{ij})_{n \times n}$ ,元  $a_{ij}$  全是实数,且  $A' = A$ ,则称  $A$  为实对称阵.下面是关于实对称阵的最重要的性质之一.

**定理 A.2.1** 设  $A$  为  $n \times n$  实对称阵,则

(a)  $A$  的所有特征值都是实数;

(b) 存在正交阵  $\Phi$ ,使得

$$\Phi' A \Phi = \text{diag}(\lambda_1, \dots, \lambda_n), \quad (\text{A.2.1})$$

这里  $\lambda_1, \dots, \lambda_n$  为  $A$  的特征值, $\Phi$  的列为对应的标准正交化特征向量.

记  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ , (A.2.1) 可以改写为

$$A = \Phi \Lambda \Phi'. \quad (\text{A.2.2})$$

实对称矩阵的这种分解是非常有用的.

在实对称矩阵中,半正定阵是一个特别重要的矩阵类.

设  $A$  为  $n \times n$  实对称阵,若对任意  $n$  维实向量  $x$ ,均有  $x' A x \geq 0$ ,则称  $A$  为

半正定阵,记为  $A \geq 0$ . 若进一步假设  $x'Ax = 0$ , 当且仅当  $x = 0$ , 则称  $A$  为正定阵, 记为  $A > 0$ .

**定理 A.2.2** 设  $A$  为  $n \times n$  实对称阵.

(a) 若  $A \geq 0$ , 则  $A$  的所有特征值均为非负数;

(b) 若  $A > 0$ , 则  $A$  的所有特征值均为正数.

**证明** (a) 设  $\lambda_i$  为  $A$  的任一特征值,  $\varphi_i$  为对应的特征向量, 不妨设  $\|\varphi_i\| = 1$ . 从特征值和特征向量的定义, 知  $A\varphi_i = \lambda_i\varphi_i$ . 用  $\varphi_i'$  左乘得

$$\varphi_i' A \varphi_i = \lambda_i \varphi_i' \varphi_i = \lambda_i \|\varphi_i\|^2 = \lambda_i, \quad (\text{A.2.3})$$

因为  $A$  为半正定阵, 所以  $\varphi_i' A \varphi_i \geq 0$ . 从 (A.2.3) 知,  $\lambda_i \geq 0$ .

(b) 若  $A$  为正定阵, 于是  $\varphi_i' A \varphi_i > 0$ . 从 (A.2.3) 知,  $\lambda_i > 0$ , 证毕.

若  $A \geq 0$ , 根据定理 A.2.2 知, 它的所有特征值  $\lambda_i \geq 0$ . 于是算术平方根  $\lambda_i^{1/2}$  都是实数. 记

$$A^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_n^{1/2}).$$

定义

$$A^{1/2} = \Phi A^{1/2} \Phi', \quad (\text{A.2.4})$$

称  $A^{1/2}$  为  $A$  的平方根阵. 此因

$$(A^{1/2})^2 = \Phi A^{1/2} \Phi' \Phi A^{1/2} \Phi' = \Phi A \Phi' = A. \quad (\text{A.2.5})$$

显然,  $A^{1/2} \geq 0$ .

如果  $A > 0$ , 则不难证明  $A^{1/2} > 0$ . 因此, 我们可以求  $A^{1/2}$  的逆矩阵, 记之为  $A^{-1/2}$ , 即  $A^{-1/2} = (A^{1/2})^{-1}$ . 利用  $\Phi$  为正交阵, 从 (A.2.4) 可以推出

$$A^{-1/2} = \Phi A^{-1/2} \Phi',$$

其中

$$A^{-1/2} = \text{diag}(\lambda_1^{-1/2}, \dots, \lambda_n^{-1/2}).$$

**推论 A.2.1** 设  $A$  为  $n \times n$  半正定阵, 则存在矩阵  $B$ , 使得  $A = BB'$ .

**证明** 分两种情况. 第一种情况, 对矩阵  $B$  的阶数不加任何限制, 则我们可以取  $B = A^{1/2}$ , 这时  $B$  还是对称阵.

第二种情况, 设  $\text{rank}(A) = r$ , 如果我们要求  $B$  为  $n \times r$  矩阵. 这时定义  $\Lambda_r = \text{diag}(\lambda_1, \dots, \lambda_r)$ , 这里  $\lambda_i > 0, i = 1, \dots, r$  为  $A$  的非零特征值.  $\Phi_1$  为  $\Phi$  的前  $r$  列, 它们为对应于  $\lambda_1, \dots, \lambda_r$  的标准正交化特征向量. 记  $B = \Phi_1 \Lambda_r^{1/2}$ , 则由 (A.2.2) 知,  $A = BB'$ . 证毕.

注. 推论中的第二种情况, 称为矩阵  $A$  的满秩分解.

### 3. 幂等阵

设  $A$  为方阵, 且满足  $A^2 = A$ , 则称  $A$  为幂等阵. 下面是一些幂等阵的例子.

(a) 单位阵  $I_n$  为  $n$  阶幂等阵.

(b) 设  $n$  阶方阵  $A$  具有形式:

$$A = \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

则  $A$  为幂等阵.

(c) 设  $n$  阶方阵  $A$  具有形式:

$$A = \Phi \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi',$$

其中  $\Phi$  为  $n$  阶正交阵, 则不难验证  $A$  为幂等阵.

(d) 设  $A$  为  $n \times m$  阵, 其秩为  $m$ , 记

$$P_A = A(A'A)^{-1}A', \quad (\text{A.3.1})$$

则  $P_A$  是幂等阵, 而且是对称阵. 这个矩阵很重要, 称为正交投影阵. 它是向矩阵  $A$  的列向量张成的子空间  $\mu(A)$  上的正交投影阵. 很明显, 矩阵  $N = I - P_A$  也是对称幂等阵, 因而也是正交投影阵. 它是向  $\mu(A)$  的正交补空间上的正交投影阵. 特别当  $A = \mathbf{1}_n$ , 即  $n$  个分量全为 1 的  $n$  维向量时,  $N = I - P_A = I - \frac{1}{n} \mathbf{1}\mathbf{1}'$ , 这个对称幂等阵在本书中多次用到.

**定理 A.3.1** (a) 若  $A$  为对称幂等阵, 则  $A$  的特征值只能是 1 或 0.

(b) 若  $A$  为实对称的幂等阵, 且秩为  $r$ , 则存在正交阵  $\Phi$ , 使得

$$A = \Phi \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi'. \quad (\text{A.3.2})$$

**证明** (a) 设  $\lambda$  为  $A$  的任一特征值,  $\varphi$  为对应的特征向量. 从  $A\varphi = \lambda\varphi$  和  $A^2 = A, A' = A$  可推知

$$\lambda\varphi'\varphi = \varphi'A\varphi = \varphi'A^2\varphi = (A\varphi)'(A\varphi) = \lambda^2\varphi'\varphi.$$

从而得  $\lambda(\lambda - 1) = 0$ , 因此特征值  $\lambda$  只能是 0 或 1.

(b) 这一结论可以从 (A.2.1) 直接推出.

**定理 A.3.2** 设  $A$  为对称幂等阵, 则  $\text{rank}(A) = \text{tr}(A)$ .

**证明** 设  $\text{rank}(A) = r$ . 根据定理 A.3.1,  $A$  具有分解式 (A.3.2). 于是

$$\begin{aligned} \text{rank}(A) &= \text{rank} \left[ \Phi \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi' \right] \\ &= \text{rank} \begin{pmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \\ &= r. \end{aligned}$$

在上面第二个等号处, 我们利用了矩阵的秩的如下性质: 一个矩阵的秩在乘了可逆方阵后保持不变. 另一方面, 利用  $\text{tr}AB = \text{tr}BA$  以及  $\Phi'\Phi = I$ , 从 (A.3.2) 得

$$\operatorname{tr} \mathbf{A} = \operatorname{tr} \Phi \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Phi' = \operatorname{tr} \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = r.$$

定理得证.

#### 4. 分块矩阵的逆矩阵

设  $\mathbf{A} > 0$ , 将其分块为

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

则它的逆矩阵

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{A}_{22}^{-1} \end{pmatrix} \quad (\text{A. 4. 1})$$

$$= \begin{pmatrix} \mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} & \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \end{pmatrix}, \quad (\text{A. 4. 2})$$

这里  $\mathbf{A}_{22}^{-1} = \mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12}$ ,  $\mathbf{A}_{11}^{-1} = \mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21}$ .

**证明** 因为一个矩阵的逆矩阵是唯一的, 为了证明(A. 4. 1)的右端是  $\mathbf{A}^{-1}$ , 只需验证(A. 4. 1)的右端与  $\mathbf{A}$  的乘积是单位阵.

作为一个应用, 下面我们给出一个很有趣的例子. 设  $\mathbf{A} = \mathbf{X}'\mathbf{X}$ ,  $\mathbf{X} = (x_1, \dots, x_p)$  为  $n \times p$  矩阵, 其秩为  $p$ . 则从(A. 4. 2)可以推出

$$(\mathbf{X}'\mathbf{X})^{-1} \text{ 的 } (i, i) \text{ 元} = \frac{1}{x_i' \mathbf{N}_i x_i}, \quad (\text{A. 4. 3})$$

这里  $\mathbf{N}_i = \mathbf{I} - \mathbf{X}_{(i)} (\mathbf{X}'_{(i)} \mathbf{X}_{(i)})^{-1} \mathbf{X}'_{(i)}$ ,  $\mathbf{X}_{(i)}$  为从  $\mathbf{X}$  中删去第  $i$  列后剩下的矩阵.

我们只对  $i=1$ , 证明(A. 4. 3). 对于一般的  $i$ , 我们只要经过适当的行、列变换就可以化成  $i=1$  的情形. 将  $\mathbf{X}$  改写为  $\mathbf{X} = (x_1, \mathbf{X}_{(1)})$ , 于是

$$\mathbf{A} = \mathbf{X}'\mathbf{X} = \begin{pmatrix} x_1' x_1 & x_1' \mathbf{X}_{(1)} \\ \mathbf{X}'_{(1)} x_1 & \mathbf{X}'_{(1)} \mathbf{X}_{(1)} \end{pmatrix}.$$

应用(A. 4. 2). 我们所求的  $(\mathbf{X}'\mathbf{X})^{-1}$  的  $(1, 1)$  元为  $\mathbf{A}_{11}^{-1}$ , 而此时的

$$\begin{aligned} \mathbf{A}_{11}^{-1} &= x_1' x_1 - x_1' \mathbf{X}_{(1)} (\mathbf{X}'_{(1)} \mathbf{X}_{(1)})^{-1} \mathbf{X}'_{(1)} x_1 \\ &= x_1' \mathbf{N}_1 x_1. \end{aligned}$$

于是(A. 4. 3)的  $i=1$  的情形得证.

#### 5. 矩阵的 Kronecker 乘积

设  $\mathbf{A} = (a_{ij})$  和  $\mathbf{B} = (b_{ij})$  分别为  $m \times n$  和  $p \times q$  矩阵, 定义

$$C = A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix},$$

这是一个  $mp \times nq$  的矩阵,称为  $A$  与  $B$  的 Kronecker 乘积.

Kronecker 乘积具有下列性质:

- (1)  $(A_1 + A_2) \otimes B = (A_1 \otimes B) + (A_2 \otimes B)$ ,  
 $A \otimes (B_1 + B_2) = (A \otimes B_1) + (A \otimes B_2)$ .
- (2)  $(\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B)$ , 其中  $\alpha$  和  $\beta$  均为数.
- (3)  $(A_1 \otimes B_1)(A_2 \otimes B_2) = (A_1 A_2) \otimes (B_1 B_2)$ .
- (4)  $(A \otimes B)' = A' \otimes B'$ .

## 附录 2 本书部分例题常用统计软件包 计算机输出结果

在回归分析的大量计算工作中,计算机及相应的统计软件已经成为不可缺少的工具.许多常用统计软件包的问世为我们进行统计分析工作提供了极大的帮助.下面我们仅对本书中的部分例题给出 SAS(统计分析软件包)或 SPSS(社会统计软件包)的计算机输出结果,以帮助读者对统计软件包的使用有初步的印象.使用不同软件所得到的计算输出结果形式有所不同.另外不同程序的软件所控制的舍入误差也略有不同.

### (一) 智力数据的统计分析

对于例 3.4.2,下面是 SAS 输出结果.

Model: MODEL1

Dependent Variable: Y

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	1	1604.08089	1604.08089	13.202	0.0018
Error	19	2308.58578	121.50451		
C Total	20	3912.66667			
		Root MSE	11.02291	R-square	0.4100
		Dep Mean	93.66667	Adj R-sq	0.3789
		C.V.	11.76823		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob >  T
INTERCEP	1	109.873841	5.06780177	21.681	0.0001
X	1	-1.126989	0.31017209	-3.633	0.0018

Obs	Dep Var	Predict	Std Err		Std Err	Student				Cook's
	Y	Value	Predict	Residual	Residual	Residual	-2	-1	0 1 2	D
1	95.0000	92.9690	2.413	2.0310	10.756	0.189				0.001
2	71.0000	80.5721	4.333	-9.5721	10.136	-0.944	*			0.081
3	83.0000	98.6040	2.763	-15.6040	10.671	-1.462	**			0.072
4	91.0000	99.7309	2.928	-8.7309	10.627	-0.822	*			0.026
5	102.0	92.9690	2.413	9.0310	10.756	0.840			*	0.018
6	87.0000	87.3341	2.970	-0.3341	10.615	-0.031				0.000
7	93.0000	89.5880	2.654	3.4120	10.699	0.319				0.003
8	100.0	97.4770	2.624	2.5230	10.706	0.236				0.002
9	104.0	100.9	3.115	3.1421	10.574	0.297				0.004
10	94.0000	87.3341	2.970	6.6659	10.615	0.628			*	0.015
11	113.0	102.0	3.321	11.0151	10.511	1.048			**	0.055
12	96.0000	99.7309	2.928	-3.7309	10.627	-0.351				0.005
13	83.0000	98.6040	2.763	-15.6040	10.671	-1.462	**			0.072
14	84.0000	97.4770	2.624	-13.4770	10.706	-1.259	**			0.048
15	102.0	97.4770	2.624	4.5230	10.706	0.422				0.005
16	100.0	98.6040	2.763	1.3960	10.671	0.131				0.001
17	105.0	96.3500	2.516	8.6500	10.732	0.806			*	0.018
18	57.0000	62.5403	8.898	-5.5403	6.506	-0.852	*			0.678
19	121.0	90.7150	2.539	30.2850	10.727	2.823			****	0.223
20	86.0000	97.4770	2.624	-11.4770	10.706	-1.072	**			0.035
21	100.0	98.6040	2.763	1.3960	10.671	0.131				0.001

Sum of Residuals                    0  
Sum of Squared Residuals        2308.5858  
Predicted Resid SS(Press)        2850.5261

以上结果给出了回归系数的最小二乘估计、对回归方程的显著性检验(F 检验)的方差分析表、对回归系数的显著性检验( $t$  检验)的方差分析表、在 21 个观测点处的拟合值向量  $\hat{y}$  (即表中的 Predict Value)、学生化残差(Student Residual)、诊断统计量  $D_i$  (即表中的 Cook's D) 及学生化残差图(参见下页图)。

下面是用 SPSS 对例 3.4.2 的输出结果, 内容与上面的 SAS 结果相当, 只是形式有所不同, 请读者自行对照。

18 Mar 99 SPSS for MS WINDOWS Release 6.0

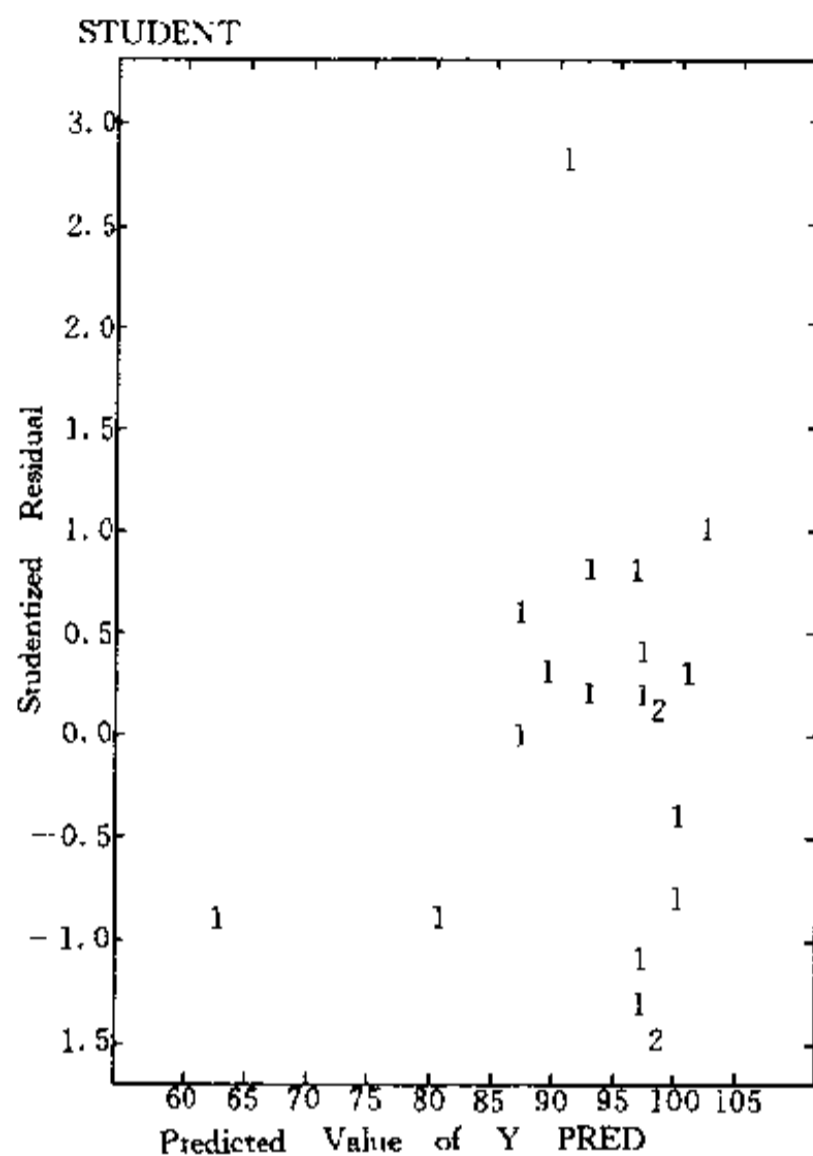
\* \* \* \* MULTIPLE REGRESSION \* \* \* \*

Listwise Deletion of Missing Data

Equation Number 1    Dependent Variable.    Y

Block Number 1. Method: Enter    X





Variable(s) Entered on Step Number

1.. X

Multiple R	.64029
R Square	.40997
Adjusted R Square	.37892
Standard Error	11.02291

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	1604.08089	1604.08089
Residual	19	2308.58578	121.50451

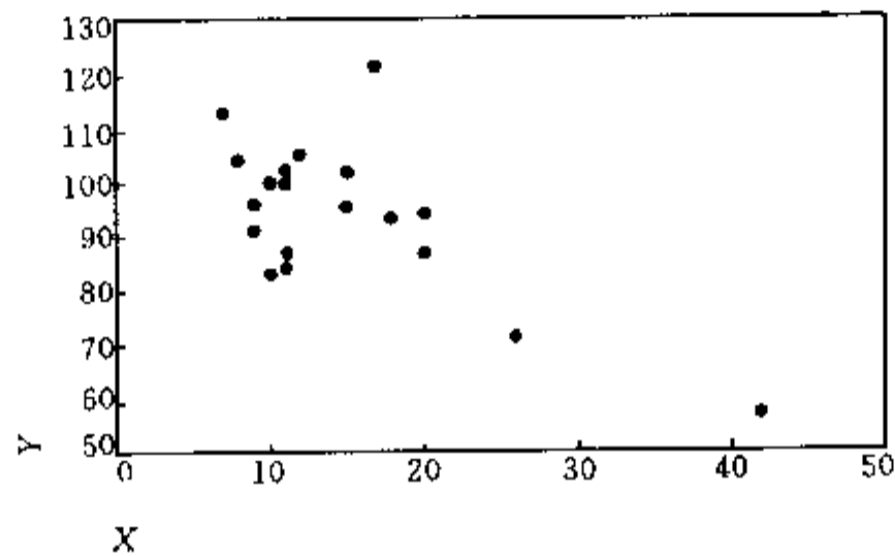
F= 13.20182      Signif F= .0018

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
X	-1.126989	.310172	-.640290	-3.633	.0018
(Constant)	109.873841	5.067802		21.681	.0000

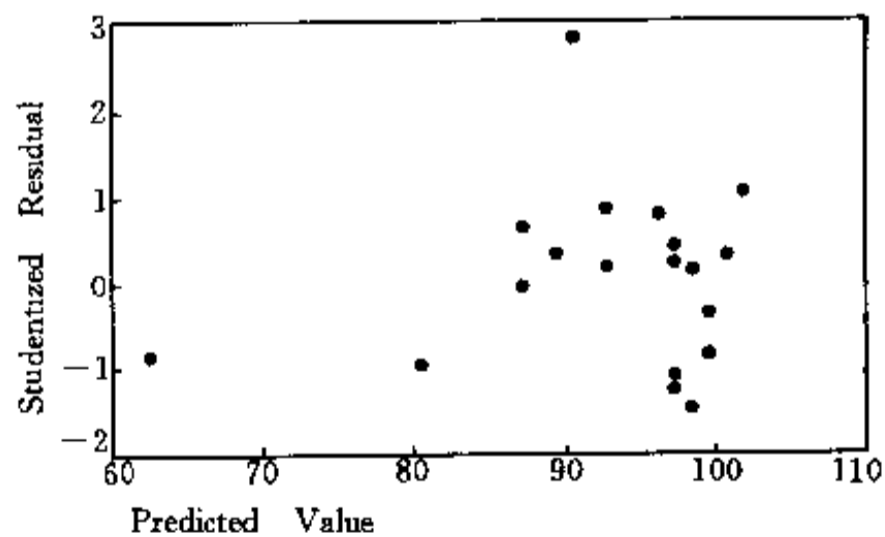
End Block Number 1 All requested variables entered.

18 Mar 99 SPSS for MS WINDOWS Release 6.0



Hi-Res Chart # 1: Scatter of y x

18 Mar 99 SPSS for MS WINDOWS Release 6.0



Hi-Res Chart # 2: Scatter of sre \_1 pre \_1

(二) 煤净化数据的统计分析

对于例 4.2.1, 例 4.3.1, 下面是 SAS 输出结果.

Model: MODEL1

Dependent Variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	31156.02427	10385.34142	23.827	0.0002
Error	8	3486.89240	435.86155		
C Total	11	34642.91667			

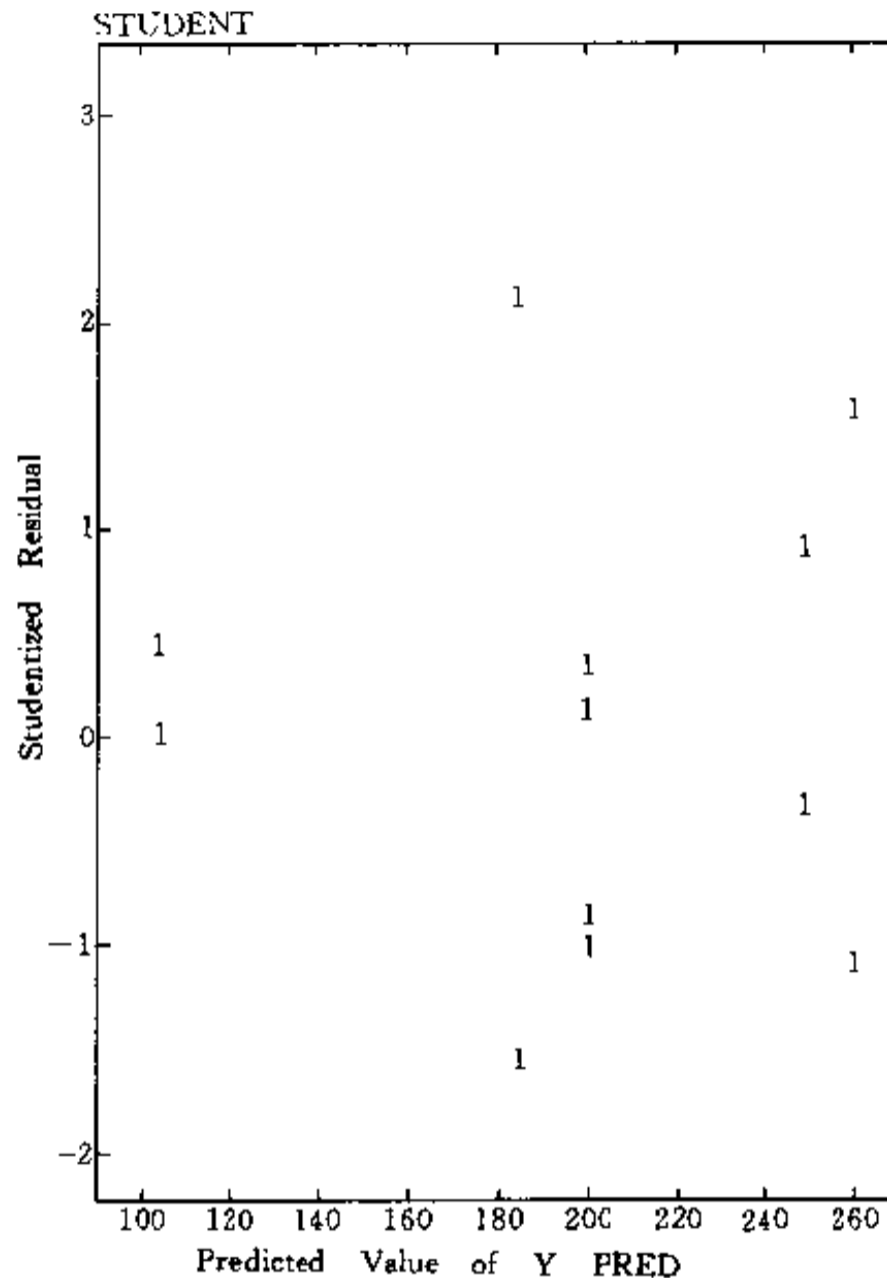
Root MSE	20.87730	R-square	0.8993
Dep Mean	199.58333	Adj R-sq	0.8616
C.V.	10.46044		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter = 0	Prob> T
INTERCEP	1	397.087383	62.75675582	6.327	0.0002
X1	1	-110.750000	14.76247862	-7.502	0.0001
X2	1	15.583333	4.92082621	3.167	0.0133
X3	1	-0.058292	0.02563482	-2.274	0.0526

Obs	Dep Var Y	Predict Value	Std Err Predict	Residual	Std Err Residual	Student Residual	-2 -1 0 1 2	Cook's D
1	243.0	247.8	14.007	-4.8080	15.481	-0.311		0.020
2	261.0	247.8	14.007	13.1920	15.481	0.852	*	0.149
3	244.0	261.0	14.252	-17.0400	15.256	-1.117	**	0.272
4	285.0	261.0	14.252	23.9600	15.256	1.571	***	0.538
5	202.0	200.7	6.045	1.3480	19.983	0.067		0.000
6	180.0	200.7	6.045	-20.6520	19.983	-1.033	**	0.024
7	183.0	200.7	6.045	-17.6520	19.983	-0.883	*	0.018
8	207.0	200.7	6.045	6.3480	19.983	0.318		0.002
9	216.0	183.8	14.007	32.1920	15.481	2.079	****	0.885
10	160.0	183.8	14.007	-23.8080	15.481	-1.538	***	0.484
11	104.0	103.5	14.252	0.4600	15.256	0.030		0.000
12	110.0	103.5	14.252	6.4600	15.256	0.423		0.039

Sum of Residuals	0
Sum of Squared Residuals	3486.8924
Predicted Resid SS(Press)	10062.6903



以上结果给出了例 4.2.1, 例 4.3.1 中的最小二乘估计, 对回归方程的显著性检验 ( $F$  检验) 的方差分析表, 对回归系数的显著性检验 ( $t$  检验) 的方差分析表, 部分诊断统计量的数据, 如  $D_i$  统计量 (Cook's  $D$ )、学生化残差、拟合值向量等及学生化残差图。

### (三) 小麦品种试验数据的方差分析

下面是对例 6.1.1, 用 SAS 输出的单因素方差分析表。

The Data of wheat

Analysis of Variance Procedure

Class Level Information

Class	Levels	Values
KIND	3	1 2 3

Number of observations in data set = 12

The Data of wheat

Analysis of Variance Procedure

Dependent Variable: WHEAT

Source	DF	Sum of Squares	Mean Square	F Value	Pr>F
Model	2	807311.25000000	403655.62500000	9.57	0.0059
Error	9	379488.75000000	42165.41666667		
Corrected Total	11	1186800.00000000			

R-Square	C. V.	Root MSE	WHEAT Mean
0.680242	3.599337	205.34219407	5705.00000000

Source	DF	Anova SS	Mean Square	F Value	Pr>F
KIND	2	807311.25000000	403655.62500000	9.57	0.0059

## 附录3 Durbin - Watson 统计量的上、下界值表

5%显著性水平下 Durbin - Watson 统计量的上、下界值表

(表中  $k$  为回归变量个数)

$n$	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

1%显著性水平下 Durbin-Watson 统计量的上、下界值表  
(表中  $k$  为回归变量个数)

$n$	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Durbin-Watson 统计量的上、下界值表

样本量	显著性水平	回归变量个数									
		1		2		3		4		5	
		$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	.01	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
	.025	.95	1.23	.83	1.40	.71	1.61	.59	1.84	.48	2.09
	.05	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
20	.01	.95	1.15	.86	1.27	.77	1.41	.63	1.57	.60	1.74
	.025	1.08	1.28	.99	1.41	.89	1.55	.79	1.70	.70	1.87
	.05	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99
25	.01	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
	.025	1.13	1.34	1.10	1.43	1.02	1.54	.94	1.65	.86	1.77
	.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89
30	.01	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
	.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	.98	1.73
	.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
	.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
	.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
	.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
80	.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
	.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
100	.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
	.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78



## 参 考 文 献

- 1 王松桂. 线性模型的理论及其应用. 合肥: 安徽教育出版社, 1987
- 2 王松桂. 线性回归系统回归系数的一种新估计. 中国科学, A 辑(1988) 10, 1033—1040
- 3 韦博成, 鲁国斌, 史建清. 统计诊断引论. 南京: 东南大学出版社, 1991
- 4 安鸿志, 陈兆国, 杜金观, 潘一民. 时间序列的分析与应用. 北京: 科学出版社, 1983
- 5 陈希孺, 王松桂. 近代回归分析. 合肥: 安徽教育出版社, 1987
- 6 陈希孺. 概率论与数理统计. 合肥: 中国科技大学出版社, 1992
- 7 项可风, 吴启光. 试验设计与数据分析. 上海: 上海科学技术出版社, 1989
- 8 常学将, 陈敏, 王明生. 时间序列分析. 北京: 高等教育出版社, 1993
- 9 Cook R D, Weisberg S. Residuals and Influence in Regression. New York: Chapman and Hall, 1982
- 10 Fuller W A. Measurement Errors Models. New York: Wiley and Sons Inc., 1987
- 11 King M L, Giles D E. Specification Analysis in the Linear Model. New York: Routledge and Kegan Paul Inc. 1987
- 12 Myers R. Classical and Modern Regression with Applications. Boston: PWS Publishers, 1986
- 13 SAS. SAS/STAT User's Guide. Ver. 6, Vol. 2, 4th Ed. SAS Institute Inc, 1994
- 14 Tong Y L. The Multivariate Normal Distribution. Beijing: Springer-Verlag, 1990
- 15 Wang S G, Chow S C. Advanced Linear Models. New York: Marcel Dekker Inc., 1994

[ G e n e r a l I n f o r m a t i o n ]

书名 = 面向 2 1 世纪课程教材 线性统计模型——线性回归与方差分析

作者 = 王松桂 陈敏 陈立萍

页数 = 2 3 6

S S 号 = 1 0 5 4 0 1 1 7

出版日期 = 1 9 9 9 年 0 9 月 第 1 版

前言  
目录  
目录

第一章	引论
§ 1.1	线性回归模型
§ 1.2	方差分析模型
§ 1.3	应用概述
	习题一
第二章	随机向量
§ 2.1	均值向量与协方差阵
§ 2.2	随机向量的二次型
§ 2.3	正态随机向量
§ 2.4	$\chi^2$ 分布
	习题二
第三章	回归参数的估计
§ 3.1	最小二乘估计
§ 3.2	最小二乘估计的性质
§ 3.3	约束最小二乘估计
§ 3.4	回归诊断
§ 3.5	Box-Cox 变换
§ 3.6	广义最小二乘估计
§ 3.7	复共线性
§ 3.8	岭估计
§ 3.9	主成分估计
	习题三
第四章	假设检验与预测
§ 4.1	一般线性假设
§ 4.2	回归方程的显著性检验
§ 4.3	回归系数的显著性检验
§ 4.4	异常点检验
§ 4.5	因变量的预测
	习题四
第五章	回归方程的选择
§ 5.1	评价回归方程的标准
§ 5.2	计算所有可能的回归
§ 5.3	计算最优子集回归
§ 5.4	逐步回归
	习题五
第六章	方差分析模型
§ 6.1	单因素方差分析
§ 6.2	两因素方差分析
§ 6.3	正交试验设计与方差分析
	习题六
* 第七章	其它线性回归模型
§ 7.1	引言
§ 7.2	具有异方差误差的线性回归模型
§ 7.3	具有自回归误差的线性回归模型
§ 7.4	具有一阶自回归误差的线性回归模型
§ 7.5	对一阶自回归误差的假设检验
§ 7.6	半相依线性回归模型
附录 1	关于矩阵的若干基础知识
附录 2	本书部分例题常用统计软件包计算机输出结果

附录3 Durbin-Watson统计量的上、下界值表  
参考文献