线性和广义线性混合模型及其统计诊断

费　宇　陈　飞
喻达磊　韩俊林　著

科学出版社

线性和广义线性混合模型
及其统计诊断

费　宇　陈　飞　喻达磊　韩俊林　著

科　学　出　版　社
北　京

内 容 简 介

本书系统介绍线性混合模型和广义线性混合模型的基本理论和方法. 主要包括两类模型的参数估计、假设检验、置信区域和统计诊断问题. 重点是两类模型的统计诊断分析, 采用数据删除方法研究两类模型影响点的探测问题, 基于EM算法中的Q函数, 来构建影响度量——广义Cook统计量, 解决了一般方差结构的两类混合模型统计诊断的困难. 而且, 获得的影响度量有很好的统计意义, 能够方便地用于全参数(均值参数与方差参数)和部分参数(均值参数或方差参数)的诊断分析.

本书可以作为统计专业高年级本科生及研究生的教材和参考书, 也可以作为数学、生物、医学和经济等领域教师和研究人员的参考书.

图书在版编目(CIP)数据

线性和广义线性混合模型及其统计诊断/费宇等著. —北京：科学出版社, 2013.3

ISBN 978-7-03-036479-1

Ⅰ. ①线… Ⅱ. ①费… Ⅲ. ①线性模型–研究 ②线性模型–统计分析（数学） Ⅳ. ①O212

中国版本图书馆 CIP 数据核字 (2013) 第 012694 号

责任编辑: 李 欣 赵彦超 / 责任校对: 李 影
责任印制: 钱玉芬 / 封面设计: 陈 敬

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码: 100717
http://www.sciencep.com

中国科学院印刷厂印刷

科学出版社发行 各地新华书店经销

*

2013 年 3 月第 一 版 开本: B5(720×1000)
2013 年 3 月第一次印刷 印张: 11 1/4
字数：208 000

定价：48.00 元

(如有印装质量问题, 我社负责调换)

前　　言

线性混合模型和广义线性混合模型是两类重要的应用广泛的模型, 近年来关于两类模型的参数估计、假设检验、置信区域和统计诊断问题方面的研究有了很多新的进展. 本书试图向读者系统介绍近几年来关于两类模型的研究结果, 特别是两类模型和统计诊断分析是本书的重要内容.

第 1 章简要介绍普通线性模型、广义线性模型、线性混合模型和广义线性混合模型的定义, 统计诊断的含义和基本方法, 还给出了一些矩阵代数和矩阵微商法则, 方便以后应用.

第 2 章介绍线性混合模型的定义、常见类型. 在正态假定下讨论了线性混合模型参数的最大似然估计和限制最大似然估计, 在非正态假定下讨论了方差分量参数的估计方法, 接着研究了该模型的假设检验和置信区域理论, 最后介绍线性混合模型随机效应的预测问题和模型选择方法.

第 3 章研究线性混合模型的统计诊断. 先从一般似然函数出发讨论独立方差结构下线性混合模型的影响点探测问题, 然后从 EM 算法中的 Q 函数出发来构建影响度量, 来研究非独立方差结构下线性混合模型的统计诊断, 获得了两个重要的基于 Q 函数的广义 Cook 统计量, 它们有很好的统计意义, 能够方便地用于全参数 (均值参数与方差参数) 和部分参数 (均值参数或方差参数) 的统计诊断分析.

第 4 章介绍广义线性混合模型的定义、模型参数估计 (包括基于 EM 算法的最大似然估计和基于条件似然的参数估计), 讨论了估计量的大样本性质, 然后研究了参数的区间估计和假设检验问题, 最后介绍了广义线性混合模型的选择准则.

第 5 章研究广义线性混合模型的统计诊断. 首先在一般似然函数框架下研究模型的影响点探测问题, 然后从 EM 算法中的 Q 函数出发构造影响度量, 来讨论模型的统计诊断分析, 最后介绍了模型扰动的选择问题.

本书内容是国家自然科学基金项目 “线性混合模型和广义线性混合模型均值和方差协方差结构的同时拟合及其统计诊断”(项目编号：11061036) 的部分研究成果, 同时, 本书的出版也获得国家自然科学基金数学天元基金项目 “充分降维理论

中基于分布加权思想的压缩估计”(项目编号：11126297)、云南财经大学统计学博士点建设基金的支持, 在此表示诚挚的感谢.

作　者

2012 年 10 月于昆明

目　　录

插 图 目 录

表 格 目 录

第 1 章　引　　论

1.1　线性模型简介

1.1.1　普通线性模型

定义 1.1　如下模型通常用来描述 y 与 x 之间的随机线性关系

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon, \tag{1.1}$$

其中 $x_1, \cdots, x_k$ 是非随机的自变量, y 是因变量, β_0 是常数项, $\beta_1, \cdots, \beta_k$ 是回归系数, ε 是随机误差项.

假设对 $y, x_1, \cdots, x_k$ 进行了 n 次观测, 得到 n 组观测值 $y_i, x_{1i}, \cdots, x_{ki} (i = 1, \cdots, n)$, 它们满足关系式

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i, \tag{1.2}$$

引入矩阵记号, 记

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} \\ 1 & x_{12} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{kn} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

则模型 (1.2) 可以写成如下形式

$$y = X\beta + \varepsilon, \tag{1.3}$$

其中 y 是 $n \times 1$ 观测向量, X 是 $n \times (k+1)$ 已知设计阵, ε 是随机误差向量, β 是未知参数向量.

如果模型 (1.3) 满足条件：(1) $E(\varepsilon) = 0$, (2) $\mathrm{var}(\varepsilon) = \sigma^2 I$, (3) $x_1, \cdots, x_k$ 不相关, 则称模型 (1.3) 为普通线性回归模型 (ordinary linear regression model).

进一步, 如果模型的随机误差项服从正态分布, 即 $\varepsilon \sim N(0, \sigma^2 I)$, 则称模型

(1.3) 为普通正态线性回归模型.

下面给一个例子说明.

例 1.1(普通线性回归模型) 考虑变量 y 随变量 $x_1,\cdots,x_k$ 变化的情况, 随机观测了 n 组观测值 $(y_i,x_{1i},\cdots,x_{ki})(i=1,\cdots,n)$, 这里 y_i 是 y 的第 i 个观测值, x_{ji} 是 $x_j(j=1,\cdots,k)$ 的第 i 个观测值, 影响 y_i 取值的主要因素是 $x_{1i},\cdots,x_{ki}$. 此外, 随机误差也要考虑, 于是可以将 y 随 x 变化的情况写出如下形式

$$y_i=f(x_{1i},\cdots,x_{ki})+\varepsilon_i\quad(i=1,\cdots,n).\tag{1.4}$$

如果函数 f 是线性函数, 即 $f(x_{1i},\cdots,x_{ki})=\beta_0+\beta_1x_{1i}+\cdots+\beta_kx_{ki}$, 则模型 (1.4) 就是一个线性回归模型. 如果变量 $x_1,\cdots,x_k$ 相互独立, 随机误差项 ε_i 相互独立, 均值为 0, 方差为 σ^2, 则该模型是一个普通线性回归模型.

1.1.2 广义线性模型

模型 (1.3) 是一般的线性模型, 可以处理连续型变量, 即 y 变量是连续型变量的情况, 但在实际中, 很多变量不是连续的, 例如调查顾客是否购买了某种商品, 这里的响应变量 y 是二值变量, 比如 $y=1$ 表示购买了该商品, $y=0$ 表示没有购买该商品. 对于 y 是二值变量或其他非连续变量的情况, 不能直接采用一般线性模型 (1.3) 进行分析, 可以采用本节定义的广义线性模型 (generalized linear model, GLM).

定义 1.2 广义线性模型 (McCulloch et al., 2002) 是一般线性模型的推广, 其定义由以下三个部分组成:

(1) 随机成分 (random components): 设 $y_1,\cdots,y_n$ 是来自于指数分布族的随机样本, 即 y_i 的密度函数为

$$f(y_i,\alpha_i,\phi)=\exp\left\{\frac{y_i\alpha_i-b(\alpha_i)}{a_i(\phi)}+c_i(y_i,\phi)\right\},\tag{1.5}$$

其中 ϕ 是散度参数 (dispersion parameter), $a_i(\cdot),b(\cdot)$ 和 $c_i(\cdot)$ 是已知函数.

(2) 系统成分 (system components): 对于第 i 个响应 y_i, 以下系统成分称为线性预测项 (linear predictor)

$$\eta_i=x_i^{\mathrm{T}}\beta=\sum_{j=1}^{k}x_{ij}\beta_j\quad(i=1,\cdots,n),\tag{1.6}$$

它是协变量 x_j 的线性组合.

(3) 关联函数 (link function)：设 $\mu_i = E(y_i)$ 是 y_i 的期望, 而 $g(\cdot)$ 是单调可微函数, 将随机成分的期望 μ_i 与系统成分联结起来, 即

$$g(\mu_i) = \eta_i = x_i^{\mathrm{T}}\beta \quad (i = 1, \cdots, n). \tag{1.7}$$

引入矩阵符号, 记

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}_{n\times 1}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{bmatrix}_{n\times k},$$

则关联函数可以写成以下矩阵形式

$$g(\mu) = \eta = X\beta, \tag{1.8}$$

其中 β 是未知参数.

以上定义的广义线性模型是一类广泛的线性模型, 实际中应用广泛的概率单位模型 (probit model) 和逻辑斯谛模型 (logistic model) 就属于广义线性模型.

例 1.2 (概率单位模型)　设 y_i 服从参数为 p_i 的伯努利分布 (Bernoulli distribution), 即 $y_i \sim B(p_i)(i = 1, \cdots, n)$, 则 $\mu_i = E(y_i) = p_i > 0$, 采用正态关联函数, 即 $g(\mu_i) = \Phi^{-1}(p_i) = x_i^{\mathrm{T}}\beta$, 其中 $\Phi(\cdot)$ 是正态分布的分布函数, 此模型称为概率单位模型.

例 1.3 (逻辑斯谛模型)　设 y_i 服从参数为 p_i 的伯努利分布, 即 $y_i \sim B(p_i)(i = 1, \cdots, n)$, 则 $\mu_i = E(y_i) = p_i > 0$, 采用逻辑关联函数, 即 $g(\mu_i) = \mathrm{logit}(p_i) = \log\dfrac{p_i}{1-p_i} = \eta_i = x_i^{\mathrm{T}}\beta$, 此模型称为逻辑斯谛模型.

1.1.3　线性混合模型

线性模型 $y = X\beta + \varepsilon$ 中, 回归系数 β 是未知常数, 但实际中, 在某些情况下, 系数视为随机的更合理. 比如在医学研究中, 随机观测了 m 个病人的某个指标的测量值, 每个病人都作了若干次重复测量, 用 y_{ij} 记第 i 个病人的第 j 次观测值 $(i = 1, \cdots, m; j = 1, \cdots, n_i)$, 显然, 第 i 个病人的 n_i 次观测值是相关的, 假定 u_i 是第 i 个病人的随机效应 (random effect), 则可以采用以下线性混合模型 (linear mixed model, LMM) 来拟合 y_{ij}, 即

$$y_{ij} = x_{ij}^{\mathrm{T}}\beta + u_i + \varepsilon_{ij} \quad (i = 1, \cdots, m; j = 1, \cdots, n_i),$$

其中 x_{ij} 是已知协变量, β 是未知回归系数, u_i 是未知随机效应, 一般假定 u_i 是相互独立同分布的随机变量, 均值为 0, 方差为 σ_u^2; 而 ε_{ij} 是随即误差项, 均值为 0, 方差为 σ_ε^2.

定义 1.3 一般的线性混合模型 (Laid et al., 1982) 定义为

$$y = X\beta + Zu + \varepsilon, \tag{1.9}$$

其中 y 是响应观测向量, X 是已知协变量矩阵, β 是未知回归系数, 通常称为固定效应, Z 是已知矩阵, u 是随机效应, ε 是随机误差, 一般假定 u 与 ε 相互独立.

进一步, 不失一般性, 可以假定 $E(u) = 0, E(\varepsilon) = 0$, 而 $\mathrm{var}(u) = G, \mathrm{var}(\varepsilon) = R$, 如果假定 u 和 ε 均服从正态分布, 即 $u \sim N(0, G)$, $\varepsilon \sim N(0, R)$, 称模型为正态线性混合模型.

线性混合模型是一类应用非常广泛的模型, 可以用来拟合多种复杂数据, 比如纵向数据和面版数据等, 线性混合模型 (1.9) 包含了许多常用模型, 比如方差分量模型, 含协变量的两向随机效应模型、纵向模型、增长曲线模型等 (具体参见 2.2 节), 下面仅给一个例子说明.

例 1.4 (单向随机效应模型) 医学实验中, 通常要比较 k 种药治疗某种疾病的效果, 药效度量指标为 Y, 通常采用双盲实验法, 假设随机抽取了 $n = mk$ 个病人, 分为 k 组, 每组有 m 个人, 分别服用 k 种药, 记 y_{ij} 是服用第 i 种药的第 j 个病人的药效测量值, 则 y_{ij} 可以表示为

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (i = 1, \cdots, k; j = 1, \cdots, m), \tag{1.10}$$

其中 μ 称为总平均, α_i 表示第 i 种药的效应, ε_{ij} 是随机误差.

在这个问题中, 我们感兴趣的因子是药品, 它有 k 个不同的品种, 称为因子的水平或处理, 模型 (1.10) 称为单向分类模型 (或单因素方差分析模型). 引入矩阵记号, 记 $X = \mathbf{1}_n, \beta = \mu$, 且

$$y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1m} \\ y_{21} \\ \vdots \\ y_{km} \end{bmatrix}_{n\times 1}, \quad Z = \begin{bmatrix} \mathbf{1}_m & & & \\ & \mathbf{1}_m & & \\ & & \ddots & \\ & & & \mathbf{1}_m \end{bmatrix}_{n\times k}, \quad u = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}_{k\times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1m} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{km} \end{bmatrix}_{n\times 1},$$

则上述模型可以表示为 $y = X\beta + Zu + \varepsilon$, 即线性混合模型 (1.9) 的形式.

1.1.4 广义线性混合模型

广义线性混合模型 (generalized linear mixed model, GLMM) 是广义线性模型 (GLM) 和线性混合模型 (LMM) 的有机结合, 可以用来拟合离散型非独立的一类数据, 这类数据无法单独运用广义线性模型或线性混合模型来处理, 下面先给出广义线性混合模型的定义, 然后给一个例子说明.

定义 1.4 广义线性模型 (Jiang, 2007) 的定义由以下三个部分组成：

(1) 随机成分 (random components)：设给定随机向量 $u = (u_1, \cdots, u_m)^{\mathrm{T}}$, 响应变量 $y_1, \cdots, y_m$ 是来自于指数分布族的变量, 其密度函数为

$$f(y_i, \alpha_i, \phi|u) = \exp\left\{\frac{y_i\alpha_i - b(\alpha_i)}{a_i(\phi)} + c_i(y_i, \phi)\right\}, \tag{1.11}$$

其中 ϕ 是散度参数 (dispersion parameter), $a_i(\cdot), b(\cdot)$ 和 $c_i(\cdot)$ 是已知函数.

(2) 系统成分 (system components)：对于第 i 个响应 y_i, 以下系统成分称为线性预测项 (linear predictor)

$$\eta_i = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}u \quad (i = 1, \cdots, n), \tag{1.12}$$

其中 x_i 和 z_i 是已知向量, 而 β 和 u 是未知的固定效应和随机效应.

(3) 关联函数 (link function)：在给定 u 的条件下, y_i 的条件期望是 μ_i, 即 $\mu_i = E(y_i|u)$, 而 $g(\cdot)$ 是单调可微函数, 将随机成分的期望 μ_i 与系统成分 η_i 联结起来, 即

$$g(\mu_i) = \eta_i = x_i^{\mathrm{T}}\beta + z_i^{\mathrm{T}}u_i \quad (i = 1, \cdots, n). \tag{1.13}$$

引入矩阵符号, 记

$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}_{n\times 1}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} x_1^{\mathrm{T}} \\ x_2^{\mathrm{T}} \\ \vdots \\ x_n^{\mathrm{T}} \end{bmatrix}_{n\times k},$$

$$Z = \begin{bmatrix} z_1^{\mathrm{T}} \\ z_2^{\mathrm{T}} \\ \vdots \\ z_n^{\mathrm{T}} \end{bmatrix}_{n\times m}, \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix}_{m\times 1},$$

则关联函数可以写成以下矩阵形式

$$g(\mu) = \eta = X\beta + Zu. \tag{1.14}$$

例 1.5 (广义逻辑斯谛模型) 医学研究经常涉及到配对二值数据, 比如想了解癌症病人在主要的癌症治疗中心是否比在一般的社区医院能接受更有效的治疗, 我们不能简单地比较两个地方癌症治疗的有效率, 因为两个地方病人的数目差别很大, 比如, 癌症治疗中心可能提供较差的治疗, 原因是他们要治疗最难的癌症. 一个可能的方法是采用配对设计：根据治疗的时间、方法和病人的年龄, 我们从癌症治疗中心随机抽取一个病人与从社区医院抽取的一个病人配成一对, 假设响应变量是 90 天内肿瘤大小有没有缩小, 用 $y_{ij} = 1$ 表示缩小了, 用 $y_{ij} = 0$ 表示没有缩小, 这里 $i = 1, \cdots, m$ 表示病人对别; 而 j 表示医院类型, $j = 1$ 表示癌症治疗中心, $j = 2$ 表示一般社区医院; 协变量 x_{ij} 表示病人来自于哪种医院, $x_{i1} = 0$ 表示来自癌症治疗中心, $x_{i2} = 1$ 表示来自一般社区医院; 记 $p_{ij} = P(y_{ij} = 1|u_i)$, 其中 u_i 表示第 i 个病人的随机效应. 于是, 给定 u_i 条件下, y_{ij} 服从参数为 p_{ij} 的伯努利分布, 即 $y_{ij}|u_i \sim B(p_{ij})(i = 1, \cdots, m; j = 1, 2)$, 显然 $\mu_{ij} = E(y_{ij}|u_i) = p_{ij} > 0$, 采用逻辑关联函数, 即 $g(\mu_{ij}) = \mathrm{logit}(p_{ij}) = \log \dfrac{p_{ij}}{1 - p_{ij}} = \eta_{ij} = \alpha + \beta x_{ij} + u_i$, 这样得到的广义线性混合模型称为广义逻辑斯谛模型.

1.2 统计诊断概述

1.2.1 统计诊断的含义

统计模型是统计分析的重要手段, 采用统计模型拟合数据, 进行统计预测和推断是数据分析的常用方法, 只有模型是合理的情况下相应的统计预测和推断才是有效的, 而模型是否合理需要进行统计检验和统计诊断分析, 因此, 统计诊断是数据分析的重要环节. 统计诊断的主要任务是通过诊断统计量检测用既定模型 (postulated model) 拟合观测数据的合理性, 具体说来, 统计诊断主要检测两类数据点：一是异常点 (outlier), 即严重偏离既定模型的数据点; 二是强影响点 (influential case), 即对统计推断 (比如参数估计、假设检验等) 的结果有重要影响的点.

需要指出的是, 异常点很可能是强影响点, 强影响点也很可能是异常点; 但二者之间的联系不是必然的, 即异常点也可能不是强影响点, 反之亦然. 异常点与强影响

点之间的关系是一个比较复杂的有争议的问题, 它与数据的实际背景有很大关系, 关于二者关系的讨论可以参阅 Beckman 和 Cook(1983), Cook 等 (1982), Chatterjee 和 Hadi(1988), 韦博成等 (1991), Pan 和 Fang(2002).

统计诊断方法总是结合数据点进行分析, 研究它们对于统计推断的影响, 所以也把这个过程称为影响分析 (influence analysis).

1.2.2 统计诊断的主要方法

统计诊断有两种主要方法：点删除 (case deletion) 方法和局部影响分析 (local influence analysis) 方法, 下面简要介绍这两种方法.

1. *点删除方法*

点删除方法也称数据删除方法 (Cook, 1977; Cook et al., 1982), 它通过比较点删除模型与原模型相应统计量之间的差异, 进行统计诊断分析, 它是统计诊断的最基本的方法, 适用于线性回归模型和其他更复杂的统计模型.

这里以线性回归模型简单说明点删除方法的含义. 给定一组数据集 $X = \{x_1, \cdots, x_n\}$, 假设 X 对应参数模型 $M(\theta)$, 即 $X \sim M(\theta)$. 参数 θ 的一个估计为 $\hat{\theta}$(MLE 或 LS 估计等), 研究数据点 x_i 对 $\hat{\theta}$ 的影响, 考虑删除 x_i 前后估计量的变化, 设删除 x_i 后 θ 的估计为 $\hat{\theta}_{[i]}$, 构造某种合适的“距离”D_i, 用来度量 $\hat{\theta}_{[i]}$ 与 $\hat{\theta}$ 之间的“差异”, D_i 通常称为诊断统计量. 最常用就是如下定义的 Cook 距离

$$D_i(M) = (\hat{\theta}_{[i]} - \hat{\theta})^{\mathrm{T}} M (\hat{\theta}_{[i]} - \hat{\theta}), \tag{1.15}$$

其中 M 是权矩阵 (Cook et al., 1982).

如果 x_i 是一个正常的点, 则 $\hat{\theta}_{[i]}$ 与 $\hat{\theta}$ 应该相差不大, 从而 D_i 会比较小; 如果 D_i 比很大, 则 $\hat{\theta}_{[i]}$ 与 $\hat{\theta}$ 相差很大, 说明 x_i 的存在与否对 θ 的估计值有重大影响, 即 x_i 对 $\hat{\theta}$ 有很大影响, 那么 x_i 可能是异常点或强影响点. 本书采用的统计诊断方法是点删除方法, 主要使用类似 (1.15) 的 Cook 距离作为诊断统计量.

2. *局部影响分析*

局部影响分析是由 Cook(1986) 基于似然函数提出的一种统计诊断方法, 它用于评价既定模型假定有微小变动对统计推断产生的局部影响. Cook(1986) 引入了扰动 (perturbation) 概念, 把异常点和强影响点归结为“比其他点受到更大扰动的

点”. 局部影响分析方法是一种应用广泛的统计诊断分析方法, 在一定条件下, 前面提到的点删除诊断方法可以视为局部影响分析方法的特例. 下面以似然距离为例, 简要介绍局部影响分析方法和相应的统计诊断方法.

给定一组数据集 $(x_i, y_i)(i = 1, \cdots, n)$, 设 $Y = (y_1, \cdots, y_n)^{\mathrm{T}}$ 的分布密度为 $p(y, \theta)$, 对数似然函数为 $L(\theta), \theta \in \Theta$, 参数 θ 的最大似然估计为 $\hat{\theta}$, 以上模型记为 M. 假定模型 M 受到扰动变为 $M(\omega)$, $\omega = (\omega_1, \cdots, \omega_q)^{\mathrm{T}} \in \Omega$ 是 q 维扰动向量, 设相应的对数似然函数为 $L(\theta|\omega)$, 相应的 θ 的 MLE 为 $\hat{\theta}(\omega)$. 假定存在 $\omega = \omega_0 \in \Omega$, 使 $M(\omega_0) = M$ 是未受到扰动的原模型, 于是有 $L(\theta|\omega_0) = L(\theta)$, $\hat{\theta}(\omega_0) = \hat{\theta}$, $L(\hat{\theta}) = L(\hat{\theta}(\omega_0)|\omega_0)$. 为了评价扰动对统计推断的影响, 对模型 M 和其扰动模型 $M(\omega)$, 最大似然估计 $\hat{\theta}$ 与 $\hat{\theta}(\omega)$ 之间的似然距离定义为

$$LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}(\omega))], \tag{1.16}$$

其中 $L(\hat{\theta})$ 是 $L(\theta)$ 在 Θ 上的全局最大值, 因此 $LD(\omega)$ 表示扰动前后这两个最大值的改变量, 而且 $LD(\omega) \geqslant 0$, $LD(\omega)$ 越大, 说明扰动对于估计量 $\hat{\theta}$ 的影响越大. 在一定正则条件下, 参数 θ 的渐近似然置信域为

$$C(\theta) = \{\theta : 2[L(\hat{\theta}) - L(\theta)] \leqslant \chi^2_{1-\alpha}(p)\}, \tag{1.17}$$

因此, 若 $LD(\omega)$ 很大, 使 $\hat{\theta}(\omega)$ 落在置信域之外, 则说明 $\hat{\theta}(\omega)$ 离 θ 很远, 也离 $\hat{\theta}$ 很远, 即 $\hat{\theta}(\omega)$ 与 $\hat{\theta}$ 之间的差异很大, 从而认为扰动对估计量的影响很大.

1.3 预 备 知 识

本节给出本书后面章节会用到的一些预备知识, 比如矩阵代数和矩阵微商法则等, 这些结果的证明可以参阅 Pan 和 Fang(2002), Searle 等 (1992), Vonesh 和 Chinchilli(1997), 王松桂等 (2004).

1.3.1 矩阵代数

定义 1.5 (矩阵的 Kronecker 乘积) 设 $A = (a_{ij})_{m\times n}, B = (b_{ij})_{p\times q}$, 定义矩阵 $C = (a_{ij}B)_{mp\times nq}$, 称为 A 和 B 的 Kronecker 乘积, 记为 $C = A \otimes B$, 即

$$C = A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}_{mp\times nq}. \tag{1.18}$$

定义 1.6 (矩阵的向量化运算) 设 $A_{m\times n}=(a_1,a_2,\cdots,a_n)$, 称 $mn\times 1$ 向量 $\mathrm{vec}(A)=(a_1^{\mathrm{T}},a_2^{\mathrm{T}},\cdots,a_n^{\mathrm{T}})^{\mathrm{T}}$ 为矩阵 A 的按列拉直运算, 即将矩阵 A 向量化变为一个向量.

由定义 1.5 和 1.6 容易得以下公式

$$\mathrm{vec}(ABC)=(A\otimes C^{\mathrm{T}})\mathrm{vec}(B). \tag{1.19}$$

引理 1.1 (Pan et al., 2002) 设 A 和 B 是 $p\times p$ 和 $q\times q$ 非奇异矩阵, C 是 $p\times q$ 矩阵, 而 D 是 $q\times p$ 矩阵, 那么

$$(A+CBD)^{-1}=A^{-1}-A^{-1}CB(B+BDA^{-1}CB)BDA^{-1} \tag{1.20}$$

引理 1.2 (Pan et al., 2002) 设 A 是 $p\times p$ 非奇异矩阵有如下分解

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

其中 A_{11},A_{12},A_{21} 和 A_{22} 分别是 $k\times k, k\times(p-k),(p-k)\times k$ 和 $(p-k)\times(p-k)$ 子阵. 假定 A_{11} 和 A_{22} 非奇异, 记 $A_{11.2}=A_{11}-A_{12}A_{22}^{-1}A_{21}$ 和 $A_{22.1}=A_{22}-A_{21}A_{11}^{-1}A_{12}$, 则

$$\begin{aligned} A^{-1}&=\begin{pmatrix} A_{11.2}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22.1}^{-1} \end{pmatrix} \\ &=\begin{pmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}A_{22.1}^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}A_{22.1}^{-1} \\ -A_{22}^{-1}A_{21}A_{11.2}^{-1} & A_{22.1}^{-1} \end{pmatrix} \end{aligned} \tag{1.21}$$

引理 1.3 对于分块行向量 $a=(a_1,\cdots,a_m)$ 与分块对角矩阵 $B=\mathrm{diag}\,(B_1,\cdots,B_m)$, 有 $aBa^{\mathrm{T}}=a_1B_1a_1^{\mathrm{T}}+\cdots+a_mB_ma_m^{\mathrm{T}}$, 这里假定涉及的矩阵乘法都有意义.

证明

$$\begin{aligned} aBa^{\mathrm{T}}&=[a_1,\cdots,a_m]\begin{bmatrix} B_1 & & \\ & \ddots & \\ & & B_m \end{bmatrix}\begin{bmatrix} a_1^{\mathrm{T}} \\ \vdots \\ a_m^{\mathrm{T}} \end{bmatrix} \\ &=[a_1B_1,\cdots,a_mB_m]\begin{bmatrix} a_1^{\mathrm{T}} \\ \vdots \\ a_m^{\mathrm{T}} \end{bmatrix} \end{aligned}$$

$$= a_1 B_1 a_1^{\mathrm{T}} + \cdots + a_m B_m a_m^{\mathrm{T}}.$$

1.3.2 矩阵微商

定义 1.7 (矩阵微商) 设 X 为 $m \times n$ 矩阵, $y = f(X)$ 是 X 的一个实值函数, 下列矩阵

$$\frac{\partial y}{\partial X} \triangleq \begin{bmatrix} \dfrac{\partial y}{\partial x_{11}} & \dfrac{\partial y}{\partial x_{12}} & \cdots & \dfrac{\partial y}{\partial x_{1n}} \\ \dfrac{\partial y}{\partial x_{21}} & \dfrac{\partial y}{\partial x_{22}} & \cdots & \dfrac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial y}{\partial x_{m1}} & \dfrac{\partial y}{\partial x_{m2}} & \cdots & \dfrac{\partial y}{\partial x_{mn}} \end{bmatrix}_{m \times n} \tag{1.22}$$

称为 y 对 X 的微商. 设 $F(X)$ 是 X 的矩阵函数, 则 $F(X)$ 关于 X 的微商定义为 $\dfrac{\partial F(X)}{\partial X} = \dfrac{\partial \mathrm{vec}(F(X))}{\partial \mathrm{vec}(X)}$.

引理 1.4 (Vonesh et al., 1997) 设 $g(x)$ 是 $p \times 1$ 向量 x 的函数, 则对于常数对称矩阵 A, 有

$$\frac{\partial \{y - g(x)\}^{\mathrm{T}} A \{y - g(x)\}}{\partial x} = -2D(x)^{\mathrm{T}} A \{y - g(x)\}, \tag{1.23}$$

其中 $D(x) = \partial g(x)/\partial x^{\mathrm{T}}$.

引理 1.5 (Searle et al., 1992) 设 A 是 x 的矩阵函数, 记 $|A|$ 为矩阵 A 的行列式, 则

$$\frac{\partial \log |A|}{\partial x} = \mathrm{tr}\left(A^{-1}\frac{\partial A}{\partial x}\right), \quad \frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1}, \tag{1.24}$$

其中 $\mathrm{tr}(\cdot)$ 表示矩阵的迹.

引理 1.6 (矩阵函数乘积求导公式) 设 X 为 $m \times n$ 阶实矩阵, $F(X)$ 和 $G(X)$ 分别为 $p \times q$ 阶和 $q \times r$ 阶实矩阵, 其元素均为 X 的函数, 则

$$\frac{\partial F(X)G(X)}{\partial X} = \frac{\partial F(X)}{\partial X}(I_p \otimes G(X)) + \frac{\partial G(X)}{\partial X}(F(X)^{\mathrm{T}} \otimes I_r). \tag{1.25}$$

引理 1.7 (复合矩阵函数求导公式) 设 $F(G(X))$ 是两个矩阵函数的复合函数, 则有

$$\frac{\partial F(G(X))}{\partial X} = \frac{\partial G(X)}{\partial X}\frac{\partial F(G)}{\partial G}. \tag{1.26}$$

引理 1.8 (矩阵正则逆求导公式)

$$\frac{\partial X^{-1}}{\partial X} = -(X^{-1})^{\mathrm{T}} \otimes X^{-1}. \tag{1.27}$$

引理 1.9 (矩阵行列式求导公式)

$$\frac{\partial |X|}{\partial \{X\}} = (X^{\mathrm{T}})^{-1}|X|. \tag{1.28}$$

第 2 章　线性混合模型

2.1　模 型 简 介

在现实问题中, 人们常使用下述线性模型来刻画变量之间的随机线性关系

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon,$$

其中, y 是因变量, $x_1, \cdots, x_k$ 是自变量, ε 是随机误差. $\beta_1, \cdots, \beta_k$ 是回归系数, 刻画了自变量对因变量的影响大小. 这里, $\beta_0, \beta_1, \cdots, \beta_k$ 均为非随机的常数. 然而, 在实际问题中, 对 Y 产生影响的因素未必一定以固定的方式作用于 Y, 在模型中以线性形式引入这些因素时, 变量前面的系数就应当是随机的, 而非固定值. 此时, 模型的形式就推广为本章所要讨论的线性混合模型 (linear mixed model). 为了方便, 给出其样本形式

$$y = X\beta + Zu + \varepsilon, \tag{2.1}$$

其中, y 是因变量的观测向量, X 是协变量的设计矩阵 (已知), β 是未知回归系数向量 (非随机), 称为固定效应, Z 是已知矩阵. u 是一个随机向量, 称为随机效应向量, ε 是随机误差向量, u 与 ε 相互独立, 均无法观测.

上述线性混合模型中, 关于随机效应 u 和随机误差 ε 的分布, 根据不同的先验信息, 可设置不同的假定. 不失一般性, 总假定 u 和 ε 的期望均为 0, 事实上, 若记 $\mu_u = E(u)$, 注意到 $ZU = Z\mu_u + Z(U - \mu_u)$, 故可将 u 中心化, 而将其期望归并到固定效应部分. 此外, 通常假定 u 和 ε 的协方差矩阵有限, 记作 $G \doteq \mathrm{var}(u)$ 和 $R \doteq \mathrm{var}(\varepsilon)$. G 和 R 的结构可根据先验的信息进一步具体化, 通常会包含未知参数, G 和 R 中包含的未知参数组成的向量记作 ϑ. 关于 u 和 ε 的分布类型, 常常假定 u 和 ε 均服从正态分布, 此时, 称模型为正态 (高斯) 线性混合模型; 若对 u 和 ε 不作正态假定, 则称模型为非正态线性混合模型.

在线性混合模型 (2.1) 式中, 若记 $\varepsilon^* = Zu + \varepsilon$, 则 $E(\varepsilon^*) = 0$, (2.1) 可以表述为 $y = X\beta + \varepsilon^*$, 将 ε^* 视为随机误差项, 则 (2.1) 式也可以视为一个线性模型, 具有协

方差结构 $\text{var}(y) = ZGZ^{\mathrm{T}} + R$. 若加上正态假设, 则线性混合模型 (2.1) 可等价地表述为 $y \sim N(X\beta, ZGZ^{\mathrm{T}} + R)$.

在这个意义上, 线性混合模型可视作具有特殊协方差结构的线性模型. 但是, 用 (2.1) 式描述线性混合模型有其独特的优势. 随机效应项的引入, 使得 (2.1) 式对复杂数据结构的刻画更为清晰和直观. 此外, 在许多场合, 随机效应本身的预测也是人们关心的问题.

通过恰当地设定随机效应设计阵 Z, 以及随机效应 u 和误差 ε 的协方差 (即 $\text{var}(u)$ 和 $\text{var}(\varepsilon)$) 结构的形式, (2.1) 式可以刻画多种复杂的数据结构 (如：纵向数据、面板数据等), 下面仅举一例加以说明.

例 2.1 (面板数据模型) 研究一个指标 Y 的变化规律, 随机抽取 m 个个体进行 T 个时刻的观测. 第 i 个个体第 t 个时刻该指标的观测记为 y_{it}, x_{it} 是 p 个协变量相应的取值 (这是一个非随机可观测的 $p \times 1$ 向量). 在此问题中, 影响 y_{it} 取值的因素除了协变量之外, 当然还应当考虑随机误差. 如果假定随机误差 ε_{it} 是独立同分布, 那么, 仅仅用 ε_{it} 来刻画 y_{it} 中所包含的随机因素很可能是不够充分的. 事实上, $y_{it}(t = 1, \cdots, T)$ 都是来自第 i 个个体的观测, 只是观测时刻不同, 因此, 简单地认定它们之间相互独立是不合理的. 由于 $y_{it}(t = 1, \cdots, T)$ 来自的第 i 个个体是随机抽取的, 我们有理由认为, 除了随机误差外, 它们可能受到同一个随机因素的影响, 称之为随机个体效应, 记作 u_i. 将上述内容归纳为一个模型, 即有

$$y_{it} = x_{it}^{\mathrm{T}}\beta + u_i + \varepsilon_{it}, \quad i = 1, \cdots, m; \quad t = 1, \cdots, T,$$

其中, 随机误差 $\varepsilon_{it}(i = 1, \cdots, m; t = 1, \cdots, T)$ 独立同分布, 均值为 0, 方差为 σ_ε^2(未知); 随机个体效应 $u_i(i = 1, \cdots, m)$ 独立同分布, 均值为 0, 方差为 σ_u^2(未知); 所有 u_i 与 ε_{it} 相互独立. 易见, 该模型是模型 (2.1) 的一种特殊形式, 具体而言, 取 $y = (y_{11}, \cdots, y_{1T}, \cdots, y_{m1}, \cdots, y_{mT})^{\mathrm{T}}$,

$$X = (x_{11}, \cdots, x_{1T}, \cdots, x_{m1}, \cdots, x_{mT})^{\mathrm{T}}, \quad Z = I_m \otimes \mathbf{1}_T,$$

$$u = (u_1, \cdots, u_m)^{\mathrm{T}}, \quad \varepsilon = (\varepsilon_{11}, \cdots, \varepsilon_{1T}, \cdots, \varepsilon_{m1}, \cdots, \varepsilon_{mT})^{\mathrm{T}},$$

其中, $\mathbf{1}_T$ 为元素全为 1 的 T 维列向量. 等价地, 也可以认为, 该模型刻画了 y 的均值结构和协方差结构. 其中, 均值结构为 $E(y) = X\beta$, 而协方差结构为 $\text{var}(y) = \sigma_u^2 ZZ^{\mathrm{T}} + \sigma_\varepsilon^2 I_{mT}$, 这里, ZZ^{T} 可具体化为 $I_m \otimes J_T$, 其中 J_T 表示元素全为 1 的 T 阶方阵.

上述例子中, 当协变量设计阵 X 选取为某种特殊形式时, 模型也可视为方差分析模型, 但与一般的固定效应方差分析模型不同的是, 它含有随机效应. Laird 和 Ware(1982) 曾使用线性混合模型研究空气污染对肺功能的影响. 在该研究中, 设定了 5 个观测时刻, 分别处在正常空气质量条件下、空气污染警报期和接下来的 3 个星期, 随机抽取了 200 个在校儿童, 观测每个儿童在上述 5 个观测时刻的第一秒用力呼气量 (FEV_1). 研究的一个重要目的是获知空气污染警报期, FEV_1 是否有所下降. 形如例 2.1 中的模型被用于该问题的研究. 特别地, 取 x_{it} 为第 t 个元素为 1, 其余元素均为 0 的 T 维向量 ($T=5$), 取 $\beta=(\beta_1,\cdots,\beta_T)^{\mathrm{T}}$, 则 β_t 就代表了在第 t 个观测时刻, FEV_1 的平均取值. 问题归结为研究 $\beta_t(t=1,\cdots,5)$ 是否全部相同, 抑或呈现出某种趋势. 应用这个含随机效应的方差分析模型, Laird 和 Ware 得到了这样的结论：较之空气质量正常时期, 在空气污染警报期及之后的观测时刻, FEV_1 在平均意义上有所下降.

上述例子仅就线性混合模型的特点和适用背景略作说明. 事实上, 线性混合模型涵盖着多种类型的模型, 适用于纵向数据、面板数据等多种数据类型, 被广泛应用于生物、医学、经济等领域. 关于该模型的理论研究, 目前已有多部专著, 如：Greert 和 Greert(2000), Rao 和 Kleffe(1988), 以及 Searle(1992). 此外, Jiang(2007) 对于线性混合模型下的理论研究工作近期进展作了较为完整的回顾.

2.2 线性混合模型的常见类型

在线性混合模型 (2.1) 下, 随机效应设计阵 Z、随机效应 u 和随机误差 ε 的协方差结构 (即 $\mathrm{var}(u)$ 和 $\mathrm{var}(\varepsilon)$ 的结构), 需要根据所研究的问题和数据的特征慎重选择. 选取不同的 Z、$\mathrm{var}(u)$ 和 $\mathrm{var}(\varepsilon)$, 将得到不同类型的模型. 本节介绍两种典型的线性混合模型.

2.2.1 方差分量模型

一般的方差分量模型 (variance component model) 定义如下

$$y=X\beta+Z_1u_1+\cdots+Z_su_s+Z_{s+1}u_{s+1}, \tag{2.2}$$

其中, y 为观测向量, $X,Z_1,\cdots,Z_{s+1}$ 为已知的非随机矩阵, β 为未知参数向量, $u_1,\cdots,u_{s+1}$ 均为随机向量, 相互独立. $E(u_i)=0$, $\mathrm{var}(u_i)=\sigma_i^2I_{t_i}(i=1,\cdots,s+1)$,

$\sigma_i^2(i=1,\cdots,s+1)$ 为未知参数, 称为方差分量 (variance component).

若取 $Z_{s+1}=I$, 记 $\varepsilon=u_{s+1}$, $Z=(Z_1,\cdots,Z_s)$, $u=(u_1^{\mathrm{T}},\cdots,u_s^{\mathrm{T}})^{\mathrm{T}}$, 则方差分量模型可写作线性混合模型的标准形式 (2.1).

例 2.2 (含协变量的两向随机效应模型) 现研究一个指标 Y 与非随机协变量 $X=(X_1,\cdots,X_p)^{\mathrm{T}}$ 之间的关系. 随机抽取 m 个个体, 并随机抽取 T 个时刻, 在每个时刻对所有个体的指标 Y 和非随机协变量向量 X 进行观测, 分别记其观测值为 y_{it} 和 x_{it}. 则在建模过程中, 不加验证地认定 $y_{it}(i=1,\cdots,m;t=1,\cdots,T)$ 相互独立显然是不合适的. 事实上, 数据 y_{it} 之间至少可能存在两种相关, 即来自同一个个体的数据之间的相关和来自同一时刻的数据之间的相关. 因此, 用下述模型来刻画数据的基本规律不失合理性

$$y_{it}=x_{it}^{\mathrm{T}}\beta+\xi_i+\eta_t+\varepsilon_{it},\quad i=1,\cdots,m;\quad t=1,\cdots,T, \tag{2.3}$$

其中, $\xi_i\sim N(0,\sigma_1^2)$ 代表个体效应, $\eta_t\sim N(0,\sigma_2^2)$ 代表时间效应, x_{it} 是非随机协变量向量的样本, 随机误差 $\varepsilon_{it}\sim N(0,\tau^2)$. $\xi_i,\eta_t,\varepsilon_{it},(i=1,\cdots,m;t=1,\cdots,T)$ 之间全部相互独立. 在此模型中, 来自第 i 个个体的数据 $y_{it}(t=1,\cdots,T)$ 共享同一个个体效应 ξ_i, 而来自第 t 个时刻的数据 $y_{it}(i=1,\cdots,m)$ 则共享同一个时间效应 η_t, 从而考虑了数据之间可能存在的上述两种相关性. 需要说明的是, 上述模型的设定并不意味着我们事先认定随机个体效应或随机时间效应一定存在. 事实上, 若 $\sigma_1^2=0$, 则意味着随机个体效应不存在; 若 $\sigma_2^2=0$, 则意味着随机时间效应不存在. 判断这两个随机效应是否存在, 实际上就是假设检验问题 $H_{01}:\sigma_1^2=0$ 和 $H_{02}:\sigma_2^2=0$.

易见, 模型 (2.3) 是方差分量模型 (2.2) 的一种特殊形式, 其中, X, y 和 u_3 的形式分别同例 2.1 中的 X, y 和 ε, 而 $Z_1=I_m\otimes\mathbf{1}_T$, $Z_2=\mathbf{1}_m\otimes I_T$, $Z_3=I_{mT}$, $u_1=(\xi_1,\cdots,\xi_m)^{\mathrm{T}}$, $u_2=(\eta_1,\cdots,\eta_T)^{\mathrm{T}}$. 在该模型中, 若 $\xi_i(i=1,\cdots,m)$ 或 $\eta_t(t=1,\cdots,T)$ 的方差退化为 0, 则模型退化为单向随机效应模型. 此外, 恰当地选取 $x_{it}(i=1,\cdots,m;t=1,\cdots,T)$ 的取值, 则模型 (2.3) 也可能是一个同时包含随机效应和固定效应的方差分析模型 (如例 2.1).

2.2.2 纵向模型

在实际问题中, 人们可能需要处理这样的数据：它们形成自然的分组, 各组数据之间相互独立, 而组内数据之间则可能存在相关. 纵向数据 (longitudinal data) 就

常常具备这样的性质：假设随机抽取 m 个个体, 对第 i 个个体在 T_i 个时刻进行观测, 如果这 m 个个体之间相互独立, 则可将个体视为上述组别, 而个体在不同时刻的观测值之间存在相关性是很常见的情形. 具备上述特征的数据, 其统计规律可以用下述 Datta 和 Lahiri(2000) 提出的模型加以刻画：

$$y_i = X_i\beta + Z_i\alpha_i + \varepsilon_i, \quad i = 1, \cdots, m, \tag{2.4}$$

其中, y_i 是来自第 i 个组的观测向量; X_i 和 Z_i 是已知的非随机矩阵; β 是待估的回归系数向量; α_i 是第 i 个组的随机效应; ε_i 是第 i 个组的随机误差; α_i, ε_i $(i = 1, \cdots, m)$ 相互独立, $E(\alpha_i) = 0$, $E(\varepsilon_i) = 0$, $\mathrm{var}(\alpha_i) = G_i$, $\mathrm{var}(\varepsilon_i) = R_i$ $(i = 1, \cdots, m)$. $y = (y_1^{\mathrm{T}}, \cdots, y_m^{\mathrm{T}})^{\mathrm{T}}$ 的协方差结构

$$\mathrm{var}(y) = \mathrm{diag}(Z_1G_1Z_1^{\mathrm{T}} + R_1, \cdots, Z_mG_mZ_m^{\mathrm{T}} + R_m),$$

这里, $\mathrm{diag}(A_1, \cdots, A_m)$ 表示以 $A_1, \cdots, A_m$ 为主对角块的分块对角矩阵.

易见, 模型 (2.4) 是线性混合模型 (2.1) 的特例. 在 (2.1) 式中取

$$y = (y_1^{\mathrm{T}}, \cdots, y_m^{\mathrm{T}})^{\mathrm{T}}, \quad X = (X_1^{\mathrm{T}}, \cdots, X_m^{\mathrm{T}})^{\mathrm{T}}, \quad \alpha = (\alpha_1^{\mathrm{T}}, \cdots, \alpha_m^{\mathrm{T}})^{\mathrm{T}},$$

$$Z = \mathrm{diag}(Z_1, \cdots, Z_m), \quad \varepsilon = (\varepsilon_1^{\mathrm{T}}, \cdots, \varepsilon_m^{\mathrm{T}})^{\mathrm{T}},$$

即得模型 (2.4). 该模型虽然称为纵向模型, 但其适用范围并不局限于纵向数据, 对于组间独立的分组数据通常都是适用的.

根据先验信息, 可以进一步设定协方差矩阵 G_i 和 R_i 的结构, 但 G_i 和 R_i 通常包含未知参数, 记 G_i 和 $R_i (i = 1, \cdots, m)$ 中包含的未知参数组成的向量为 ϑ. Z_i, G_i 和 R_i 的形式选择, 就构成了对 y_i 的协方差结构的具体假定. 此外, 通过选取 X_i 的形式, 也可以进一步对纵向模型 (2.4) 的均值结构作出较为具体的假定, 下面举一个例子以作说明.

例 2.3 (增长曲线模型) 增长曲线模型主要被用来刻画某个指标的变化趋势和规律, 比如：儿童体重随着时间推移的变化规律; 某个指标随着空间位置变动的变化趋势; 某项人体体征指标随着使用药物剂量变化而产生的变化等. 该模型一直较受关注, 许多文献对其加以讨论, 如：Potthoff 和 Roy(1964), Grizzle 和 Allen(1969), Rao(1973), Kshirsagan 和 Smith(1995). 这里仅讨论增长曲线模型的一个特例, 以说明纵向模型的一些特点. 假设对 m 个个体, 在 k 个时刻进行观测, 记 y_{it} 为第 i 个

个体在第 t 个时刻的观测值, 模型的形式为

$$y_{it} = \xi_i + z_{it}^{\mathrm{T}}\eta_i + \varepsilon_{it}, \quad i = 1, \cdots, m; \quad t = 1, \cdots, k, \tag{2.5}$$

其中, ε_{it} 为随机误差, 由两部分组成：一部分刻画序列相关 (这里假设是一阶自回归相关), 另一部分代表独立同分布的随机波动. $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ik})^{\mathrm{T}}$ 的期望为 0, 协方差阵为

$$R_\varepsilon \mathrel{\hat{=}} \mathrm{cov}(\varepsilon_i) = \frac{\sigma_1^2}{1-\phi^2}\begin{pmatrix} 1 & \phi & \cdots & \phi^{n-1} \\ \phi & 1 & \cdots & \phi^{n-2} \\ \vdots & \vdots & & \vdots \\ \phi^{n-1} & \phi^{n-2} & \cdots & 1 \end{pmatrix} + \sigma_2^2 I;$$

此外, z_{it} 为第 i 个个体在第 t 个时刻处协变量向量 (非随机) 的观测值; ξ_i 为随机截距, η_i 为随机回归系数向量. $\alpha_i^* \mathrel{\hat{=}} (\xi_i, \eta_i^{\mathrm{T}})^{\mathrm{T}}$ 满足 $E(\alpha_i) = \mu_\alpha$, $G_\alpha \mathrel{\hat{=}} \mathrm{var}(\alpha_i) > 0$(不假定 G_α 具有特殊结构, 如对角阵等). 另, 假定 α_i 与 ε_i 相互独立, 且对 $i \neq j$, α_i 与 α_j 独立, ε_i 与 ε_j 独立.

易见, 该增长曲线模型是纵向模型的一个特例. 事实上, 在模型 (2.4) 中, 取 $X_i = Z_i = ((1, z_{i1}^{\mathrm{T}})^{\mathrm{T}}, \cdots, (1, z_{ik}^{\mathrm{T}})^{\mathrm{T}})^{\mathrm{T}}$, $\beta = \mu_\alpha$, $\alpha_i = (\xi_i, \eta_i^{\mathrm{T}})^{\mathrm{T}} - \mu_\alpha$, $G_i = G_\alpha$, $R_i = R_\varepsilon$, $i = 1, \cdots, m$, 即得 (2.5) 式.

本节介绍了线性混合模型的两种常见类型 —— 方差分量模型和纵向模型. 这两种类型并不完全相斥, 有些模型既是方差分量模型, 又可视为纵向模型, 例如：下述单向随机效应模型

$$y_{ij} = \mu + \xi_i + \varepsilon_{ij}, \quad i = 1, \cdots, m; \quad j = 1, \cdots, k, \tag{2.6}$$

其中, μ 是总平均, 为非随机常数; ξ_i 为随机个体效应, ε_{ij} 为随机误差. 现假定 $\xi_i (i = 1, \cdots, m)$ 相互独立同分布, $E(\xi_i) = 0$, $\mathrm{var}(\xi_i) = \sigma_\xi^2$, ε_{ij} $(i = 1, \cdots, m; j = 1, \cdots, k)$ 相互独立同分布, $E(\varepsilon_{ij}) = 0$, $\mathrm{var}(\varepsilon_{ij}) = \sigma_\varepsilon^2$. 所有 ξ_i, ε_{ij} 之间均相互独立. 上述单向随机效应模型 (2.6) 式显然是一个方差分量模型, 只需在 (2.2) 式中取 $X = \mathbf{1}_{mk}$, $\beta = \mu$, $Z_1 = I_m \otimes \mathbf{1}_k$, $u_1 = (\xi_1, \cdots, \xi_m)^{\mathrm{T}}$, $Z_2 = I$, $u_2 = (\varepsilon_{11}, \cdots, \varepsilon_{1k}, \cdots, \varepsilon_{m1}, \cdots, \varepsilon_{mk})^{\mathrm{T}}$; 同时, (2.6) 式也是一个纵向模型, 只需在 (2.4) 中取 $X_i = \mathbf{1}_k$, $\beta = \mu$, $Z_i = \mathbf{1}_k$, $\alpha_i = \xi_i$, $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ik})^{\mathrm{T}}$ $(i = 1, \cdots, m)$.

另一方面, 方差分量模型和纵向模型又各有特点：方差分量模型的一个显著特点是出现在模型中的所有随机效应均互不相关; 纵向模型的显著特点则是模型

的分组性质和各个组之间的独立性, 这使得纵向模型较适用于数据自然分组的情形. 这两种模型一般而言不能相互涵盖, 例如, 例 2.2 中所列两向分类随机效应模型就无法视作纵向模型, 因为无论以个体或时间对数据分组, 组间的独立性均会受到破坏, 这种破坏来自于分属不同组的观测在另一个方向上共有的随机效应. 又如, 例 2.3 中所列增长曲线模型也无法视为方差分量模型, 因为无论是随机效应向量 $\alpha_i \triangleq (\xi_i, \eta_i^{\mathrm{T}})^{\mathrm{T}}$ 还是随机误差向量 $\varepsilon_i \triangleq (\varepsilon_{i1}, \cdots, \varepsilon_{ik})^{\mathrm{T}}$, 其分量之间均无法保证互不相关, 这与方差分量模型的假设相悖.

线性混合模型涵盖甚广, 包括的模型形式灵活多样, 这里不再一一列举. 线性混合模型形式的设定, 本质上是根据先验信息, 对因变量观测向量的均值结构和协方差结构作出具体假设. 既然是假设, 当然有可能存在偏误, 而在参数模型下, 不预作假设显然是不现实的, 例如, 若不对因变量观测向量 y 的协方差结构作出假定, 则 y 的协方差矩阵中的元素都是未知的待估参数, 其数量之大显然超过了样本量所能支持的程度. 因此, 根据先验信息选择模型具体形式, 但不过分信任初始选择的模型形式, 重视模型的检验、诊断和模型选择, 不失为明智的策略.

2.3 参 数 估 计

参数估计是线性混合模型统计推断的基本内容和重要环节. 本节将介绍几种常见的估计方法. 正如 2.1 节中提到的, 线性混合模型是线性模型 $y = X\beta + e$ $(E(e) = 0, \mathrm{var}(e) = \Sigma)$ 的一个特例 (观测向量 y 的协方差矩阵具有特定的结构). 因此, 若其观测向量的协方差矩阵已知, 则对于固定效应回归系数 β 的估计理论, 可以纳入线性模型回归系数的估计理论体系 (包括估计方法和估计的性质等内容), 细节可参看相关文献, 如：王松桂等 (2004).

2.3.1 最大似然估计

1. 最大似然估计原理及求取方法

最早将最大似然估计 (maximum likelihood estimate) 方法应用于混合模型的是 Hartley 和 Rao(1967). 在线性混合模型 (2.1) 中, 若假设随机效应和随机误差都服从正态分布, 即 $\varepsilon \sim N(0, R)$, $u \sim N(0, G)$, 则 $y \sim N_n(X\beta, V)$, 其中 $V \triangleq ZGZ^{\mathrm{T}} + R$. 故有因变量观测向量 y 的联合密度为

$$f(y) = (2\pi)^{-n/2}|V|^{-1/2}\exp\left\{-\frac{1}{2}(y-X\beta)^{\mathrm{T}}V^{-1}(y-X\beta)\right\}. \tag{2.7}$$

故有, 对数似然函数为

$$l(\beta,\vartheta|y) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|V| - \frac{1}{2}(y-X\beta)^{\mathrm{T}}V^{-1}(y-X\beta), \tag{2.8}$$

其中, ϑ 是 V 中包含的未知参数组成的向量. 根据最大似然估计 (MLE) 的定义, β 和 ϑ 的最大似然估计应使得上述对数似然函数达到最大, 即

$$(\hat{\beta}^{\mathrm{T}}, \hat{\vartheta}^{\mathrm{T}})^{\mathrm{T}} = \arg\max_{\beta,\vartheta} l(\beta,\vartheta|y). \tag{2.9}$$

注意到使得 $l(\beta,\vartheta|y)$ 达到最大的一个必要条件是 $(y-X\beta)^{\mathrm{T}}V^{-1}(y-X\beta) = [V^{-1/2}(y-X\beta)]^{\mathrm{T}}[V^{-1/2}(y-X\beta)]$ 达到最小, 即在 $V^{-1/2}X$ 的列空间上寻求一个向量使得 $V^{-1/2}y - V^{-1/2}X\beta$ 的欧氏模达到最小. 这意味着 $V^{-1/2}X\hat{\beta}$ 是 $V^{-1/2}y$ 在 $V^{-1/2}X$ 的列空间上的投影, 故根据投影的性质有

$$(V^{-1/2}y - V^{-1/2}X\hat{\beta})^{\mathrm{T}}(V^{-1/2}X) = 0, \quad \text{i.e.,} \quad X^{\mathrm{T}}V^{-1}X\hat{\beta} = X^{\mathrm{T}}V^{-1}y, \tag{2.10}$$

解方程立得

$$\hat{\beta}(\vartheta) = (X^{\mathrm{T}}V^{-1}X)^{-1}X^{\mathrm{T}}V^{-1}y. \tag{2.11}$$

当 X 列不满秩时, 应该在上式中将 $(X^{\mathrm{T}}V^{-1}X)^{-1}$ 替换为广义逆 $(X^{\mathrm{T}}V^{-1}X)^{-}$. 虽然 X 列不满秩时满足方程 (2.10) 的 $\hat{\beta}$ 不唯一, 但 $V^{-1/2}X\hat{\beta}$ 却是唯一的, 与广义逆的选择无关. 将 $\hat{\beta}(\vartheta) = (X^{\mathrm{T}}V^{-1}X)^{-1}X^{\mathrm{T}}V^{-1}y$ 代入 $l(\beta,\vartheta|y)$ 中, 记

$$l^*(\vartheta|y) \triangleq l(\hat{\beta}(\vartheta),\vartheta|y). \tag{2.12}$$

如此, 则最大似然估计的求解问题归结为求取 $l^*(\vartheta|y)$ 的最大值点：

$$\hat{\vartheta} = \arg\max_{\vartheta} l^*(\vartheta|y). \tag{2.13}$$

对于该最大值问题, 通常很难获得显式解, 但可以使用数值求解方法解决该问题. 下面介绍较为常用的牛顿–拉弗森 (Newton-Raphson) 方法.

$l^*(\vartheta|y)$ 的极大值点应当满足方程

$$\frac{\partial l^*(\vartheta|y)}{\partial\vartheta} = 0. \tag{2.14}$$

选定初值 $\vartheta^{(0)}$, 将上式左端在 $\vartheta^{(0)}$ 处泰勒 (Taylor) 展开得

$$\left.\frac{\partial l^*(\vartheta|y)}{\partial\vartheta}\right|_{\vartheta=\vartheta^{(0)}}+\left.\frac{\partial^2 l^*(\vartheta|y)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}\right|_{\vartheta=\vartheta^{(0)}}(\vartheta-\vartheta^{(0)})\approx 0, \tag{2.15}$$

以 ϑ 为未知量, 解方程得

$$\vartheta^{(1)}=\vartheta^{(0)}-\left[\left.\frac{\partial^2 l^*(\vartheta|y)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}\right|_{\vartheta=\vartheta^{(0)}}\right]^{-1}\left.\frac{\partial l^*(\vartheta|y)}{\partial\vartheta}\right|_{\vartheta=\vartheta^{(0)}}. \tag{2.16}$$

牛顿–拉弗森迭代方法通过下述迭代公式实现：

$$\vartheta^{(k+1)}=\vartheta^{(k)}-\left[\left.\frac{\partial^2 l^*(\vartheta|y)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}\right|_{\vartheta=\vartheta^{(k)}}\right]^{-1}\left.\frac{\partial l^*(\vartheta|y)}{\partial\vartheta}\right|_{\vartheta=\vartheta^{(k)}}. \tag{2.17}$$

其中, $\vartheta^{(k)}$ 是方程 (2.14) 的第 k 步迭代近似解. 初始值 $\vartheta^{(0)}$ 给定后, 即可据 (2.17) 得到 $\vartheta^{(1)}$, 进而得到 $\vartheta^{(2)},\cdots$, 直至某一步的迭代值 $\vartheta^{(l)}$ 满足预设的收敛精度要求 (精度通常通过比较前后两次或多次迭代值之间的差异确定), 则取该 $\vartheta^{(l)}$ 为方程 (2.14) 的数值解, 亦即近似地取 ϑ 的 MLE$\hat{\vartheta}=\vartheta^{(l)}$. 这里, (2.17) 右端的第二项 $\left[\partial^2 l^*(\vartheta|y)/\partial\vartheta\partial\vartheta^{\mathrm{T}}\big|_{\vartheta=\vartheta^{(k)}}\right]^{-1}\partial l^*(\vartheta|y)/\partial\vartheta\big|_{\vartheta=\vartheta^{(k)}}$ 可以视作在求取第 $k+1$ 步迭代值 $\vartheta^{(k+1)}$ 时, 对 $\vartheta^{(k)}$ 的修正项. 当 $\vartheta^{(k)}$ 渐趋于 0 时, $\partial l^*(\vartheta|y)/\partial\vartheta\big|_{\vartheta=\vartheta^{(k)}}$ 渐趋于 0, 从而修正项也渐趋于 0. 一阶偏导向量 $\partial l^*(\vartheta|y)/\partial\vartheta$ 和二阶偏导矩阵 $\partial^2 l^*(\vartheta|y)/\partial\vartheta\partial\vartheta^{\mathrm{T}}$ 的求取可以通过矩阵微商等运算公式实现, 这些矩阵理论细节可参看本书预备知识及相关文献, 如：倪国熙 (1982), Bellman(1970), 王松桂等 (2004). 下面求取 (2.17) 中需要用到的几个矩阵微商. 为了表述简洁, 在不会发生误解的前提下, 将对一些符号使用简写, 如：将 $l^*(\vartheta|y)$ 简记为 l^* 或 $l^*(\vartheta)$. 首先, 由链锁法则 (chain rule), 有 $\dfrac{\partial l^*(\vartheta)}{\partial\vartheta}=\dfrac{\partial\hat{\beta}(\vartheta)}{\partial\vartheta}\left.\dfrac{\partial l(\beta,\vartheta)}{\partial\beta}\right|_{\beta=\hat{\beta}(\vartheta)}+\left.\dfrac{\partial l(\beta,\vartheta)}{\partial\vartheta}\right|_{\beta=\hat{\beta}(\vartheta)}$, 再由 $\left.\dfrac{\partial l(\beta,\vartheta)}{\partial\beta}\right|_{\beta=\hat{\beta}(\vartheta)}=0$, 立得

$$\frac{\partial l^*(\vartheta)}{\partial\vartheta}=\left.\frac{\partial l(\beta,\vartheta)}{\partial\vartheta}\right|_{\beta=\hat{\beta}(\vartheta)}. \tag{2.18}$$

上式中, 左右两端对 ϑ 求导, 再次使用链锁法则得 $\dfrac{\partial^2 l^*(\vartheta)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}=\dfrac{\partial\hat{\beta}(\vartheta)}{\partial\vartheta}\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\vartheta\partial\beta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}+\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\vartheta\partial\vartheta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}$. 另由 $0=\dfrac{\partial\left.\dfrac{\partial l(\beta,\vartheta)}{\partial\beta}\right|_{\beta=\hat{\beta}(\vartheta)}}{\partial\vartheta}=\dfrac{\partial\hat{\beta}(\vartheta)}{\partial\vartheta}\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\beta\partial\beta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}+\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\beta\partial\vartheta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}$, 立得 $\dfrac{\partial\hat{\beta}(\vartheta)}{\partial\vartheta}=-\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\beta\partial\vartheta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}\left(\left.\dfrac{\partial^2 l(\beta,\vartheta)}{\partial\beta\partial\beta^{\mathrm{T}}}\right|_{\beta=\hat{\beta}(\vartheta)}\right)^{-1}$. 将其代