

因子平方和的正交对照分解及其应用*

孙法省 刘民千

(南开大学数学科学学院统计学系, 天津 300071)

(E-mail: mqliu@nankai.edu.cn)

摘 要 本文把 a 水平因子的平方和分解成相互正交的 $a - 1$ 个对照的平方和, 这样总变差平方和就可以分解成 a 个部分 (包括残差项). 然后又将该分解方法推广到了多因子的情形, 并通过因子平方和的分解找到了多因子交互效应对应的对照向量, 这使得多水平因子交互效应的计算和解释更加容易, 也为方差分析带来了更多的方便. 最后给出了几个应用示例.

关键词 正交对照; 因子平方和; 总变差平方和; 方差分析; 正交表

MR(2000) 主题分类 62J10; 62K15

中图分类 O212.6

1 引言

方差分析模型是应用非常广泛的一类线性统计模型. 这种模型多有一定的试验设计背景, 因而也称 试验设计模型. 对于这种模型, 有两种常用的统计分析方法: 第一种方法是将数据总变差平方和按其来源 (各种因子、交互作用及随机误差) 进行分解, 得到各因子、交互效应及误差平方和, 接下来的统计分析是基于各因子平方和与误差平方和大小的比较, 这种方法叫平方和分解法^[1,2]. 第二种方法则是把方差分析模型看作一般的线性统计模型去分析^[3-5]. 本文将采用另一种方法, 即将数据总变差平方和分解成相互正交的多个对照的平方和, 这给方差分析带来了方便. 另外, 通过研究交互效应对应的对照平方和及对照向量可以更方便地解释和计算多水平因子之间的交互效应, 使多水平情形交互效应的解释与计算和两水平情形类似.

本文后面的内容安排如下: 第 2 节讨论单因子情形因子平方和的正交对照分解; 第 3 节则把该正交对照分解推广到多因子情形, 主要以两个因子为例来说明如何利用正交对照分解因子平方和, 进而分解总变差平方和; 第 4 节是应用示例, 分别考虑二水平、三水平及混水平情形两因子平方和的正交对照分解; 最后一节是小结.

2 单因子情形

我们首先考虑单因子情形因子平方和的正交对照分解. 令 1_n 为元素全为 1 的 $n \times 1$ 向量, 称满足 $c'1_n = 0$ 的向量 c 为 $n \times 1$ 对照向量. 下面这个引理是因子平方和分解的基础.

引理 1 任给 $\alpha \in R^n$, 存在单位向量 $c_1, c_2, \dots, c_{n-1} \in V_{n-1} = \{c \mid c'1_n = 0\}$ (V_{n-1} 称为对照向量空间), 使得

$$\|\alpha\|^2 = (c'_1\alpha)^2 + (c'_2\alpha)^2 + \dots + (c'_{n-1}\alpha)^2 + (c'_n\alpha)^2,$$

本文 2007 年 4 月 6 日收到, 2007 年 7 月 9 日收到修改稿.

* 国家自然科学基金 (10671099 号) 和高等学校博士学科点专项科研基金 (20050055038 号) 资助项目.

其中 $c_n = (1/\sqrt{n}, 1/\sqrt{n}, \dots, 1/\sqrt{n})'$, $\|\alpha\|^2 = \alpha'\alpha$.

证 取 $c_1, c_2, \dots, c_{n-1}$ 为 V_{n-1} 中的一组标准正交基, 则

$$c'_i c_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad \text{且 } c'_i c_n = 0, \quad i = 1, 2, \dots, n-1, \quad j = 1, 2, \dots, n-1.$$

记 $A = (c_1, c_2, \dots, c_n)'$, $\beta = A\alpha = (c_1, c_2, \dots, c_n)'\alpha = (c'_1\alpha, c'_2\alpha, \dots, c'_n\alpha)'$, 则显然 A 为正交阵, 且 $\|\beta\|^2 = \|A\alpha\|^2 = \|\alpha\|^2$. 另一方面 $\|\beta\|^2 = \beta'\beta = (c'_1\alpha)^2 + (c'_2\alpha)^2 + \dots + (c'_{n-1}\alpha)^2 + (c'_n\alpha)^2$. 于是 $\|\alpha\|^2 = (c'_1\alpha)^2 + (c'_2\alpha)^2 + \dots + (c'_{n-1}\alpha)^2 + (c'_n\alpha)^2$. 证毕.

定义 1 对线性统计模型

$$\begin{cases} y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \\ \text{诸 } \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \end{cases} \quad i = 1, 2, \dots, a, \quad j = 1, 2, \dots, n_i, \quad (1)$$

称因子诸效应的一个线性组合 $c = \sum_{i=1}^a c_i \tau_i$ 为一个对照, 如果它的系数满足约束 $\sum_{i=1}^a c_i = 0$, 也即向量 $(c_1, c_2, \dots, c_a)'$ 为对照向量. 若两个对照对应的对照向量正交则称这两个对照正交 (见 [1,2,4,5]).

由 $\mu_i = \mu + \tau_i$ 可知对照又可表示为 $c = \sum_{i=1}^a c_i \mu_i$. 由于 $\hat{\mu}_i = \bar{y}_i$ 是 μ_i 的一个无偏估计 (在本文中凡是涉及到 y 的记号, 横代表求平均, 点代表求和, 如 $\bar{y}_i = (y_{i1} + y_{i2} + \dots + y_{in_i})/n_i$), 所以 $\sum_{i=1}^a c_i \bar{y}_i$ 就是对照 c 的一个无偏估计. 由模型的假定可知, 在等重复情况, 不妨设重复次数为 m , 该估计服从正态分布 $N(c, \sigma^2 \sum_{i=1}^a c_i^2/m)$; 在不等重复情况, 它服从正态分布 $N(c, \sigma^2 \sum_{i=1}^a c_i^2/n_i)$. 因此下面的对照平方的定义^[1]是合理的.

定义 2 对照 c 的平方 SS_c 定义为

$$SS_c = \begin{cases} \frac{\left(\sum_{i=1}^a c_i \bar{y}_i\right)^2}{\sum_{i=1}^a c_i^2/m}, & \text{等重复情况;} \\ \frac{\left(\sum_{i=1}^a c_i \bar{y}_i\right)^2}{\sum_{i=1}^a c_i^2/n_i}, & \text{不等重复情况.} \end{cases}$$

注 当 $\sum_{i=1}^a c_i \mu_i = 0$ 成立时, SS_c/σ^2 服从自由度为 1 的 χ^2 分布.

定理 1 设单因子 A 的线性统计模型为 (1), 其中 $n_1 = \dots = n_a = m$, 则因子 A 的平方和 $SS_A = m \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2$ 可以分解为 $a-1$ 个相互正交对照的平方之和.

证 令 $\alpha = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_a)'$. 由引理 1 可知, 存在单位向量 $c_i = (c_{i1}, c_{i2}, \dots, c_{ia})' \in V_{a-1}$, $i = 1, 2, \dots, a-1$, 使得 $\|\alpha\|^2 = \sum_{i=1}^{a-1} (c'_i \alpha)^2 + (c'_a \alpha)^2$, 其中 $c_a = (1/\sqrt{a}, \dots, 1/\sqrt{a})'$.

于是 $\|\alpha\|^2 - (c'_a \alpha)^2 = \sum_{i=1}^{a-1} (c'_i \alpha)^2$, 即

$$\sum_{i=1}^a \bar{y}_i^2 - \frac{1}{a} \left(\sum_{i=1}^a \bar{y}_i \right)^2 = \sum_{i=1}^{a-1} (c'_i \alpha)^2.$$

从而

$$\begin{aligned} SS_A &= m \sum_{i=1}^a (\bar{y}_i - \bar{y}_{..})^2 = m \left(\sum_{i=1}^a \bar{y}_i^2 - \frac{1}{a} \left(\sum_{i=1}^a \bar{y}_i \right)^2 \right) \\ &= m \sum_{i=1}^{a-1} (c'_i \alpha)^2. \end{aligned}$$

又

$$(c'_i \alpha)^2 = \left(\sum_{j=1}^a c_{ij} \bar{y}_j \right)^2 = \left(\sum_{j=1}^a c_{ij} \bar{y}_j \right)^2 / \sum_{j=1}^a c_{ij}^2,$$

其中后一个等号利用 c_i 为单位向量得到. 所以

$$SS_A = \sum_{i=1}^{a-1} \frac{\left(\sum_{j=1}^a c_{ij} \bar{y}_j \right)^2}{\sum_{j=1}^a c_{ij}^2 / m} = \sum_{i=1}^{a-1} SS_{c_i}, \quad (2)$$

其中后一等号由对照平方的定义得到. 证毕.

上面证明了有 a 个水平的因子 A 的平方和可以分解成 $a-1$ 个相互正交对照的平方和, 这 $a-1$ 个对照可以根据不同需要进行选择. 例如, 若某试验者想检验一下两种新生产方法和原来的旧方法哪一种更好, 则可设生产方法这个因子为 A , 它有 3 个水平, 其中水平 1 为原来的生产方法, 水平 2 和 3 分别代表两种新方法. 对本问题可取正交对照为 $c_1 = (-2, 1, 1)' / \sqrt{6}$ 和 $c_2 = (0, -1, 1)' / \sqrt{2}$, 其中对照 c_1 可用来比较原来方法的效应和两个新方法的平均效应是否相等, 而对照 c_2 可用来比较两个新方法的效应是否相等.

基于定理 1, 有 a 个水平的单因子 A 的方差分析可以归结为以下几步:

(i) 找到感兴趣的 $a-1$ 个单位正交对照向量, 由 (2) 式计算出 SS_A ;

(ii) 由 $SS_e = SS_T - SS_A$ 计算出残差平方和 SS_e , 其中 $SS_T = \sum_{i=1}^a \sum_{j=1}^m (y_{ij} - \bar{y}_{..})^2$ 为总

变差平方和;

(iii) 进行有关的分析.

由定理 1, $SS_T = SS_A + SS_e = \sum_{i=1}^{a-1} \frac{\left(\sum_{j=1}^a c_{ij} \bar{y}_j \right)^2}{\sum_{j=1}^a c_{ij}^2 / m} + SS_e$. 若记

$$SS_e = y' A_a y, \quad m \left(\sum_{j=1}^a c_{ij} \bar{y}_j \right)^2 = y' A_i y,$$

又知 $SS_T = y'y - y' \frac{1}{am} 1_{am} 1'_{am} y$, 则

$$y'y = y' \frac{1}{am} 1_{am} 1'_{am} y + \sum_{i=1}^{a-1} y' A_i y + y' A_a y.$$

令 $\frac{1}{am} 1_{am} 1'_{am} = A_0$, 则有

$$y'y = \sum_{i=0}^a y' A_i y, \quad A_i A_j = \begin{cases} 0, & i \neq j, \\ A_i, & i = j, \end{cases}$$

$$\text{rk}(A_i) = \begin{cases} 1, & i = 0, 1, \dots, a-1, \\ am-a, & i = a. \end{cases}$$

这样总平方和按其来源就分解成了三部分: 截距平方、因子对照平方和、误差平方和, 并且三部分是相互独立的.

3 多因子情形

为了把定理 1 的结果推广到多因子情形, 我们需要下面的引理.

引理 2 设 $c_i = (c_{i1}, c_{i2}, \dots, c_{in})'$, $i = 1, 2, \dots, a$, 为 a 个两两正交的单位对照向量, $s_j = (s_{j1}, s_{j2}, \dots, s_{jm})'$, $j = 1, 2, \dots, b$, 为另外的 b 个两两正交的单位对照向量, 记下面三类向量

$$V_1: D_i = \frac{1}{\sqrt{m}} (c_{i1}, \dots, c_{i1}, c_{i2}, \dots, c_{i2}, \dots, c_{in}, \dots, c_{in})' \quad (\text{括号内每个 } c_{ij} \text{ 重复 } m \text{ 次});$$

$$V_2: T_j = \frac{1}{\sqrt{n}} (s_{j1}, \dots, s_{jm}, s_{j1}, \dots, s_{jm}, \dots, s_{j1}, \dots, s_{jm})' \quad (\text{括号内每个 } s_j \text{ 重复 } n \text{ 次});$$

$$V_3: L_{j+(i-1)b} = (c_{i1}s_{j1}, \dots, c_{i1}s_{jm}, c_{i2}s_{j1}, \dots, c_{i2}s_{jm}, \dots, c_{in}s_{j1}, \dots, c_{in}s_{jm})',$$

其中 $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$, 则这 $a + b + ab$ 个向量是两两正交的单位对照向量.

证 下面的 (i) 和 (ii) 证明它们为单位对照向量, 而两两正交性则由 c_i 和 s_j 都是两两正交的单位对照向量易得.

(i) 由 c_i, s_j 的定义易证对 $\forall \alpha \in V_1 \cup V_2$, 有 $\|\alpha\| = 1$, 且 $\alpha' 1_{mn} = 0$.

(ii) 对 $\alpha \in V_3$, 不妨设 $\alpha = L_{j+(i-1)b}$ ($i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$), 则由 c_i 和 s_j 都是单位对照向量, 易知

$$\|\alpha\|^2 = (c_{i1}^2 + c_{i2}^2 + \dots + c_{in}^2)(s_{j1}^2 + s_{j2}^2 + \dots + s_{jm}^2) = 1,$$

$$\alpha' 1_{mn} = (c_{i1} + c_{i2} + \dots + c_{in})(s_{j1} + s_{j2} + \dots + s_{jm}) = 0.$$

证毕.

由该引理我们可知 $V_1 \cup V_2 \cup V_3$ 中的向量是两两正交的单位对照向量, 于是我们可以把定理 1 中关于单因子平方和分解的结果推广到两因子的情形. 事实上, 我们容易利用上面的方法把定理 1 推广到多于两个因子的情形. 下面我们以两个因子为例来说明如何利用正交对照来分解因子平方和, 进而分解总变差平方和. 设两个因子的线性统计模型为

$$\begin{cases} y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, & i = 1, \dots, a, \quad j = 1, \dots, b, \quad k = 1, \dots, m, \\ \text{诸 } \varepsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2), \end{cases} \quad (3)$$

其中两个因子分别记为 A, B , 它们分别有 a 与 b 个水平. 具体步骤如下:

第1步 找到我们感兴趣的对照向量 $c_i = (c_{i1}, c_{i2}, \dots, c_{ia})'$ 与 $s_j = (s_{j1}, s_{j2}, \dots, s_{jb})'$. 由引理2得到单位正交对照向量

$$\begin{aligned} D_i &= \frac{1}{\sqrt{b}}(c_{i1}, \dots, c_{i1}, c_{i2}, \dots, c_{i2}, \dots, c_{ia}, \dots, c_{ia})', \\ T_j &= \frac{1}{\sqrt{a}}(s_{j1}, \dots, s_{jb}, s_{j1}, \dots, s_{jb}, \dots, s_{j1}, \dots, s_{jb})', \\ L_{j+(i-1)(b-1)} &= (c_{i1}s_{j1}, \dots, c_{i1}s_{jb}, c_{i2}s_{j1}, \dots, c_{i2}s_{jb}, \dots, c_{ia}s_{j1}, \dots, c_{ia}s_{jb})', \end{aligned}$$

其中 $i = 1, 2, \dots, a-1$, $j = 1, 2, \dots, b-1$.

第2步 计算出 SS_A 与 SS_B . 由定理1,

$$\begin{aligned} SS_A &= b m \sum_{i=1}^{a-1} (c_{i1}\bar{y}_{1..} + c_{i2}\bar{y}_{2..} + \dots + c_{ia}\bar{y}_{a..})^2 \\ &= m \sum_{i=1}^{a-1} \left(\frac{1}{\sqrt{b}}\right)^2 (c_{i1}\bar{y}_{11.} + \dots + c_{i1}\bar{y}_{1b.} + \dots + c_{ia}\bar{y}_{a1.} + \dots + c_{ia}\bar{y}_{ab.})^2 \\ &= m \sum_{i=1}^{a-1} (D'_i y^*)^2, \end{aligned} \quad (4)$$

其中 $y^* = (\bar{y}_{11.}, \bar{y}_{12.}, \dots, \bar{y}_{1b.}, \dots, \bar{y}_{a1.}, \bar{y}_{a2.}, \dots, \bar{y}_{ab.})'$. 同理可得

$$SS_B = m \sum_{j=1}^{b-1} (T'_j y^*)^2. \quad (5)$$

第3步 计算出 SS_{AB} . 现假设有一 ab 个水平的单因子 G , 在每个水平下重复做 m 次试验, 并设在第 $j + (i-1)b$ 个水平上的效应平均值为 $\bar{y}_{ij.}$, $i = 1, 2, \dots, a$, $j = 1, 2, \dots, b$. 由引理2知可取因子 G 的单位正交对照向量为 $D_i, T_j, L_{j+(i-1)(b-1)}$. 于是由定理1可得

$$\begin{aligned} SS_G &= \sum_{i=1}^a \sum_{j=1}^b m(\bar{y}_{ij.} - \bar{y}_{...})^2 \\ &= m \sum_{i=1}^{a-1} (D'_i y^*)^2 + m \sum_{j=1}^{b-1} (T'_j y^*)^2 + m \sum_{i=1}^{a-1} \sum_{j=1}^{b-1} (L'_{j+(i-1)(b-1)} y^*)^2. \end{aligned}$$

于是由(4)式和(5)式,

$$\sum_{i=1}^a \sum_{j=1}^b m(\bar{y}_{ij.} - \bar{y}_{...})^2 = SS_A + SS_B + m \sum_{i=1}^{a-1} \sum_{j=1}^{b-1} (L'_{j+(i-1)(b-1)} y^*)^2.$$

又

$$\sum_{i=1}^a \sum_{j=1}^b m(\bar{y}_{ij.} - \bar{y}_{...}) = SS_A + SS_B + SS_{AB} \quad (\text{参见 [1]}),$$

所以

$$SS_{AB} = m \sum_{i=1}^{a-1} \sum_{j=1}^{b-1} (L'_{j+(i-1)(b-1)} y^*)^2. \quad (6)$$

初看上面的等式感觉有些不可思议, 但是由下面的解释我们就会认为 (6) 式是很符合情理的. 由

$$\begin{aligned} L'_{j+(i-1)(b-1)} y^* &= c_{i1} s_{j1} \bar{y}_{11} + \cdots + c_{i1} s_{jb} \bar{y}_{1b} + c_{i2} s_{j1} \bar{y}_{21} + \cdots + c_{i2} s_{jb} \bar{y}_{2b} \\ &\quad + \cdots + c_{ia} s_{j1} \bar{y}_{a1} + \cdots + c_{ia} s_{jb} \bar{y}_{ab} \\ &= c_{i1} (s_{j1} \bar{y}_{11} + \cdots + s_{jb} \bar{y}_{1b}) + c_{i2} (s_{j1} \bar{y}_{21} + \cdots + s_{jb} \bar{y}_{2b}) \\ &\quad + \cdots + c_{ia} (s_{j1} \bar{y}_{a1} + \cdots + s_{jb} \bar{y}_{ab}), \end{aligned}$$

可以看出, 表达式 $(s_{j1} \bar{y}_{11} + \cdots + s_{jb} \bar{y}_{ib})$ 是在因子 A 的第 i 个水平下因子 B 的诸水平的一个对照, 从而该式可理解为 B 在 A 的第 i 个水平下的一个条件主效应, 从而 $L'_{j+(i-1)(b-1)} y^*$ 的含义就是对 B 在 A 的每个水平下的条件主效应再求对照, 即是 AB 的一个交互效应.

第 4 步由 $SS_e = SS_T - SS_A - SS_B - SS_{AB}$ 得到残差平方和 SS_e , 其中

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^m (y_{ijk} - \bar{y}_{...})^2.$$

第 5 步 进行相关的分析.

利用表 1 进行上述步骤会方便一些. 表 1 中每个因子下各向量列与观测均值向量的内积的平方和再乘以重复次数 m , 就是该因子的平方和. 例如: $SS_B = m \sum_{j=1}^{b-1} (T'_j y^*)^2$. 而表 1 中每个因子下面的各向量列与观测均值向量的内积就是该列对应的效应. 例如: D_1 对应的效应是 $D'_1 y^*$.

表 1 两因子平方和的对照向量分解

A				B				AB				观测均值
D_1	D_2	$\cdots$	D_{a-1}	T_1	T_2	$\cdots$	T_{b-1}	L_1	L_2	$\cdots$	$L_{(a-1)(b-1)}$	y^*
d_{11}	d_{21}	$\cdots$	$d_{(a-1),1}$	t_{11}	t_{21}	$\cdots$	$t_{(b-1),1}$	l_{11}	l_{21}	$\cdots$	$l_{(a-1)(b-1),1}$	$\bar{y}_{11}$
d_{12}	d_{22}	$\cdots$	$d_{(a-1),2}$	t_{12}	t_{22}	$\cdots$	$t_{(b-1),2}$	l_{12}	l_{22}	$\cdots$	$l_{(a-1)(b-1),2}$	$\bar{y}_{12}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$d_{1,ab}$	$d_{2,ab}$	$\cdots$	$d_{(a-1),ab}$	$t_{1,ab}$	$t_{2,ab}$	$\cdots$	$t_{(b-1),ab}$	$l_{1,ab}$	$l_{2,ab}$	$\cdots$	$l_{(a-1)(b-1),ab}$	$\bar{y}_{ab}$

注 对线性统计模型

$$y^* = X\beta + e, \quad e \sim N_n(0, \sigma^2 I),$$

其中设计阵 $X = (D_1, \cdots, D_{a-1}, T_1, \cdots, T_{b-1}, L_1, \cdots, L_{(a-1)(b-1)})$, 因为 X 的列是两两正交的单位向量, β 的最小二乘估计为

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

由此可知, 在这样选择的设计阵下所得到的最小二乘估计就是各因子所对应的效应.

4 应用示例

本节我们看三个例子, 分别考虑二水平、三水平及混水平情形两因子平方和的正交对照分解.

例 1 (二水平情形) 设二水平因子 A, B 的线性统计模型为 (3), 其中 $a = 2, b = 2$. 取 $c = s = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})'$, 则由引理 2 得

$$\begin{aligned} D &= \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)' = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2} \right)', \\ T &= \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)' = \left(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2} \right)', \\ L &= \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2} \right)'. \end{aligned}$$

对应于表 1, 我们下面的表 2.

表 2 二水平因子平方和的对照向量分解

A	B	AB	观测均值
D	T	L	y^*
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\bar{y}_{11.}$
$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{2}$	$\bar{y}_{12.}$
$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{2}$	$\bar{y}_{21.}$
$-\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\bar{y}_{22.}$

该表中, 每个因子下面的向量与观测均值的内积就是该因子的主效应, 例如:

$$D'y^* = \frac{1}{2}(\bar{y}_{11.} + \bar{y}_{12.}) - \frac{1}{2}(\bar{y}_{21.} + \bar{y}_{22.}) = \bar{y}_{1..} - \bar{y}_{2..}$$

就是因子 A 的主效应, 它的平方和为

$$SS_A = m(D'y^*)^2 = \frac{m}{4}(\bar{y}_{11.} + \bar{y}_{12.} - \bar{y}_{21.} - \bar{y}_{22.})^2.$$

同理可以计算 B, AB 的效应和因子 (效应) 平方和.

例 2 (三水平情形) 设三水平因子 A, B 的线性统计模型为 (3), 其中 $a = 3, b = 3$. 取

$$\begin{aligned} c_1 = s_1 &= \frac{1}{\sqrt{2}}(1, 0, -1) = \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right)', \\ c_2 = s_2 &= \frac{1}{\sqrt{6}}(1, -2, 1) = \left(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)', \end{aligned}$$

则

$$\begin{aligned} D_1 &= \frac{1}{\sqrt{3}} \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, 0, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right)', \\ D_2 &= \frac{1}{\sqrt{3}} \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)', \\ T_1 &= \frac{1}{\sqrt{3}} \left(\frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right)', \\ T_2 &= \frac{1}{\sqrt{3}} \left(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}, \frac{1}{\sqrt{6}} \right)', \\ L_1 &= \frac{1}{2}(1, 0, -1, 0, 0, 0, -1, 0, 1)', \end{aligned}$$

$$L_2 = \frac{1}{\sqrt{2}\sqrt{6}}(1, -2, 1, 0, 0, 0, -1, 2, -1)',$$

$$L_3 = \frac{1}{\sqrt{2}\sqrt{6}}(1, 0, -1, -2, 0, 2, 1, 0, -1)',$$

$$L_4 = \frac{1}{\sqrt{6}\sqrt{6}}(1, -2, 1, -2, 4, -2, 1, -2, 1)'.$$

该三水平因子平方和的对照向量分解见表 3.

表 3 三水平因子平方和的对照向量分解

A		B		AB				观测均值
D_1	D_2	T_1	T_2	L_1	L_2	L_3	L_4	y^*
1	1	1	1	1	1	1	1	$\bar{y}_{11}$
1	1	0	-2	0	-2	0	-2	$\bar{y}_{12}$
1	1	-1	1	-1	1	-1	1	$\bar{y}_{13}$
0	-2	1	1	0	0	-2	-2	$\bar{y}_{21}$
0	-2	0	-2	0	0	0	4	$\bar{y}_{22}$
0	-2	-1	1	0	0	2	-2	$\bar{y}_{23}$
-1	1	1	1	-1	-1	1	1	$\bar{y}_{31}$
-1	1	0	-2	0	2	0	-2	$\bar{y}_{32}$
-1	1	-1	1	1	-1	-1	1	$\bar{y}_{33}$

注意的是表 3 中的每一列对照向量还要除以它的长度. 由表 3 我们可以清楚地看出此种对照的选法对应着线性二次系统 (见 [8]). 类似于例 1 中的方法可以求出各因子的效应以及效应的平方和.

例 3 (混水平情形) 设混水平因子 A, B 的线性统计模型为 (3), 其中 $a = 4, b = 2$. 可认为原来一共有三个独立的二水平因子 1, 2, 3, 见表 4, 其中 A 是由前两个因子 1, 2 及其交互效应列 12 替换而来^[6-8], B 为第三个因子 3. 表 4 中括号内为替换前的因子, 该表中除 A 列以外的其它 7 个因子列构成一个饱和正交表 (见 [7, 8]). 由前面的分析, SS_A 可以分解为三个相互正交对照的平方和, 且这三个对照与 $B, 13, 23, 123$ 列对应的对照正交, 于是这三个对照就是 1, 2, 12 列对应的对照, 因而 $SS_A = SS_1 + SS_2 + SS_{12}$, 该结果可以用来简化 SS_A 的计算.

表 4 混水平因子正交表

试验号	A	(1	2	12)	B(3)	13	23	123
1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	1	1	1
3	1	0	1	1	0	0	1	1
4	1	0	1	1	1	1	0	0
5	2	1	0	1	0	1	0	1
6	2	1	0	1	1	0	1	0
7	3	1	1	0	0	1	1	0
8	3	1	1	0	1	0	0	1

5 小结

本文给出了对方差分析模型中的因子平方和进行正交对照分解的方法, 并结合例子进行了分析解释. 从上面的例子可以看出该方法给方差分析带来了许多方便, 无论是多水平的还是混水平的. 另外利用正交对照生成的设计阵求出的最小二乘估计就是我们所关心的效应, 所以在实际建模中可以事先选定这样的设计阵. 多水平情形交互效应的计算也和两

水平情形一样,通过因子主效应的对照向量,可以直接计算出交互效应的对照向量,进而求出相应的交互效应,并且这种计算方法有助于对交互效应的理解.

致谢 作者衷心感谢审稿人提出的宝贵建议.

参 考 文 献

- [1] Wang W Z, Mao S S, Zeng L R. Design and Analysis of Experiments. Beijing: Higher Education Press, 2004
- [2] Morrison D F. Applied Linear Statistical Methods. New Jersey: Prentice-Hall, 1983
- [3] Wang S G, Shi J H, Yin S J, Wu M X. Introduction to Linear Models. Beijing: Science Press, 2004
- [4] Lindman H R. Analysis of Variance in Experimental Design. New York: Springer-Verlag, 1992
- [5] Hocking R R. Methods and Applications of Linear Models: Regression and the Analysis of Variance. New York: John Wiley & Sons, 1996
- [6] Addelman S. Orthogonal Main-effect Plans for Asymmetrical Factorial Experiments. *Technometrics*, 1962, 4: 21-46
- [7] Hedayat A S, Sloane N J A, Stufken J. Orthogonal Arrays: Theory and Applications. New York: Springer-Verlag, 1999
- [8] Wu C F J, Hamada M. Experiments: Planning, Analysis, and Parameter Design Optimization. New York: John Wiley & Sons, 2000

Orthogonal Contrasts Decomposition for Factor Sum of Squares and its Application

SUN FASHENG LIU MINQIAN

(Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin 300071)

(E-mail: mqliu@nankai.edu.cn)

Abstract In this paper, the sum of squares of an a -level factor is decomposed to $a-1$ sums of squares of mutually orthogonal contrasts, such that the total variation sum of squares is decomposed to a parts including the residual sum of squares. Then this decomposition method is generalized to the multi-factor case, and the orthogonal contrast vectors corresponding to multi-factor interactions are obtained via this decomposition, which makes the computation and explanation of the interactions between multi-level factors easier, and also brings more convenience to the analysis of variance. Some application examples are provided at last.

Key words orthogonal contrast; factor sum of squares; total variation sum of squares; analysis of variance; orthogonal array

MR(2000) Subject Classification 62J10; 62K15

Chinese Library Classification O212.6