

文章编号: 1000-5862(2012)01-0071-04

饱和正交表列效应的线性约束条件检验

罗 纯¹, 李 玮², 潘长缘³

(1. 上海应用技术学院理学院, 上海 200235; 2. 江西师范大学数学与信息科学学院, 江西 南昌 330022;
3. 依格斯医疗科技有限公司生物统计部, 上海 200021)

摘要: 采用矩阵像分解技术和零效应搜索法来讨论饱和模型的数据分析, 对零成分导出了相应的线性约束条件. 通过饱和正交表列效应的线性约束条件的问题进行适当变换, 使其等同于一个子成分的显著性检验问题, 再利用零成分搜索法判断其是否为零成分, 可以完满地解决该问题.

关键词: 正交表; 约束条件; 投影矩阵; 矩阵像; 饱和模型

中图分类号: O 212.6

文献标志码: A

关于参数线性约束条件的检验, 是统计学家非常关心的问题之一, Tukey 在 20 世纪 80 年代发表了这方面的专著. 文献[1]综合讨论了正交表列效应的约束条件检验的一般形式, 文献[2]针对不饱和正交表进行了专门论述. 本文重点讨论正交饱和模型, 针对同列中效应参数的一般线性约束条件给出了检验方案, 然后给出了模拟实例.

1 零成分搜索法

在进行正交设计数据分析时, 采用矩阵像技术^[3-9]比传统方法更加有效. 文献[1]首次提出了零成分搜索法, 它是解决正交饱和模型的一种有效方法. 主效应模型中各列的效应之和恒为0, 这是一个不需要检验的线性约束条件. 为了叙述简洁起见, 本文将不再讨论正交饱和模型中的参数估计等与约束条件无关的问题.

1.1 正交饱和模型

考虑饱和正交表 $L_n(v_1 \cdots v_m) = (b_{ij})_{n \times m}$ 的主效应模型:

$$y_i = \mu + \beta_{1b_{i1}} + \cdots + \beta_{mb_{im}} + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2), \quad (1)$$

其中 $i = 1, \cdots, n$. 简记成向量形式为

$$Y = \theta_0 + \theta_1 + \cdots + \theta_m + \varepsilon. \quad (2)$$

假设饱和正交表 $L_n(v_1 \cdots v_m)$ 各列对应的矩阵像依次记为 A_1, \cdots, A_m , 记总均值列对应的矩阵像为

A_0 , 这些矩阵像的有关定义参见文献[3-4]. 由矩阵像的基本定义, 其对应向量模型有分解式:

$$I_n = A_0 + A_1 + \cdots + A_m, \quad (3)$$

且对应向量模型, 容易得出, 对应的自由度分解式及矩阵像秩的分解式相同, 均为

$$n = 1 + (v_1 - 1) + \cdots + (v_m - 1). \quad (4)$$

利用饱和正交表的矩阵像, 可以很方便求出总平方和 $SS_T = Y^T(I_n - A_0)Y$, 各列平方和 $SS_j = Y^T A_j Y$, $j = 1, \cdots, m$.

1.2 零成分搜索具体步骤

零成分搜索法在文献[1]中已经做了详细论述, 这里只简述其具体步骤:

步骤 1 对饱和正交表的秩 $(v_j - 1)$ 大于 1 的矩阵像, 其可分解成 $(v_j - 1)$ 个投影矩阵(秩为 1). 具体做法如下: 找出矩阵像 A_j 对应特征值为 1 的正交规格化特征向量 $\zeta_{j1}, \cdots, \zeta_{j, v_j - 1}$, 从而有 $A_j = \zeta_{j1} \zeta_{j1}^T + \cdots + \zeta_{j, v_j - 1} \zeta_{j, v_j - 1}^T$.

令 $A_{j,x} = \zeta_{jx} \zeta_{jx}^T, x = 1, \cdots, v_j - 1$, 因此 $A_{j1}, \cdots, A_{j, v_j - 1}$ 均为秩为 1 的投影矩阵, 对模型参数 θ_j , 记 $\theta_{j,x} = A_{j,x} \theta_j$, 称其为 θ_j 的子成分, 而称 $A_{j,x}$ 为设计 $L_n(v_1 \cdots v_m)$ 的第 j 列的子矩阵像.

步骤 2 容易得到如下分解式:

$$\tau_n = I_n - A_0 = A_{1,1} + \cdots + A_{1, v_1 - 1} + \cdots + A_{m,1} + \cdots + A_{m, v_m - 1},$$

收稿日期: 2011-09-20

基金项目: 教育部高等学校博士学科点专项基金(44K55050)资助项目.

作者简介: 罗 纯(1966-), 男, 安徽合肥人, 副教授, 博士研究生, 主要从事概率论与数理统计方面的研究.

其中 $A_{1,1}, \dots, A_{1,v_1-1}, \dots, A_{m,1}, \dots, A_{m,v_m-1}$ 均为秩为 1 的子矩阵像, 共 $(n-1)$ 个.

根据这些子矩阵像, 可以得到平方和 $SS_{j,x} =$

$$Y^T A_{j,x} Y, x=1, \dots, v_j-1, j=1, \dots, m, \text{ 且 } SS_j = \sum_{x=1}^{v_j-1} SS_{j,x},$$

$j=1, \dots, m$. 由柯卡伦定理, 各平方和之间相互独立;

步骤 3 对这 $(n-1)$ 个子成分, 可以构造 $(n-2)$ 个 W 统计量. 利用这些 W 统计量, 找出其中的零成分. 不妨假设其中 u 个子成分为零成分(假设为 $\theta_1, \dots, \theta_u$), 其对应的子平方和记为 TT_1, \dots, TT_u , 那么其余 $n-1-u$ 个子成分都是非零成分(假设为 $\theta_{u+1}, \dots, \theta_{n-1}$), 其对应的子平方和记为 $TT_{u+1}, \dots, TT_{n-1}$;

步骤 4 利用 $\hat{\sigma}^2 = (TT_1 + \dots + TT_u)/u$ 来估计方差. 针对其它 $n-1-u$ 个非零成分 θ_v (对应的子平方和为 TT_v), 考虑假设检验问题

$$H_0: \theta_v = 0, H_1: \theta_v \neq 0, \quad (5)$$

构造 F 统计量: $F_v = \frac{TT_v}{(TT_1 + \dots + TT_u)/u}, v=u+1, \dots, n-1$; 在原假设成立的条件下 F_v 服从 $F(1, u)$ 分布. 利用该性质可以判断该子成分是否为零成分.

2 线性约束条件的导出

前文已经通过构造 W 统计量和 F 统计量的方法找出所有不显著的子成分.

针对存在不显著子成分的列, 可以增加一个线性约束条件. 不妨假设第 j 列的子成分 θ_{jx} (其对应的矩阵像为 $A_{j,x} = \zeta_{jx} \zeta_{jx}^T$, 其对应的子平方和为 TT_v) 是不显著的子成分, 视为 0, 即 $\theta_{jx} = A_{jx} \theta_j = \zeta_{jx} \zeta_{jx}^T \theta_j = 0$, 从而, $\zeta_{jx}^T \theta_j = 0$ 或者 $\theta_j^T \zeta_{jx} = 0$. 此即所求约束条件.

3 一般线性约束条件的检验

应用工作者在试验前可能对列效应的一些信息有所了解, 或者在试验后由参数的估计值有可能推断某个约束条件成立, 这些先验信息对试验设计来说很有意义. 因此, 寻求检验这些约束条件是否成立的方法是有价值的.

讨论饱和正交表第 j 列列效应的约束条件的假设检验问题:

$$H_0: C_j^T \beta_j = 0; H_1: C_j^T \beta_j \neq 0, \quad (6)$$

其中 C_j 为 v_j 维非零列向量, 假设 $C_j = (c_{j1}, \dots, c_{jv_j})^T$.

首先对约束条件进行如下变换.

(i) 将 $C_j^T \beta_j = 0, C_j = (c_{j1}, \dots, c_{jv_j})^T$ 变换为 $L_j^T \theta_j = 0, L_j = (l_{j1}, \dots, l_{jv_j})^T$ 型的约束条件. 这里选取最简单的变换方法: 如果在 θ_j 的所有行中 β_{jx} 首次出现, 则 L_j 对应行的元素选取为 β_j 中 β_{jx} 所在行对应的 C_j 的元素, 否则为 0.

(ii) 假设 $\zeta_{j1}, \dots, \zeta_{j,v_j-1}$ 为矩阵像 A_j 的特征值为 1 的一组正交规格化特征向量, 不妨将其扩充到 R^n 空间上的一组正交规格化基 $\zeta_{j1}, \dots, \zeta_{j,v_j-1}, \zeta_{j,v_j}, \dots, \zeta_{jn}$. 故存在一组不全为 0 的实数 k_1, \dots, k_n , 使得 $L_j = k_1 \zeta_{j1} + \dots + k_n \zeta_{jn}$. 实际上, 对该等式两边同时左乘 ζ_{ji}^T , 就可以得到 $k_i = \zeta_{ji}^T L_j, i=1, \dots, n$.

由于 θ_j 的最小二乘无偏估计为 $\hat{\theta}_j = A_j Y$, 即 $\theta_j = E \hat{\theta}_j = A_j E Y$, 又由于 A_j 是投影矩阵, 所以 $\theta_j = A_j \theta_j$.

令 $SS = \theta_j^T L_j L_j^T \theta_j = \theta_j^T A_j L_j L_j^T A_j \theta_j$, 将 $L_j = k_1 \zeta_{j1} + \dots + k_n \zeta_{jn}$ 代入该式, 由于 $A_j = \zeta_{j1} \zeta_{j1}^T + \dots + \zeta_{j,v_j-1} \zeta_{j,v_j-1}^T$, 因此有

$$TT = \theta_j^T A_j \left(\sum_{i=1}^{v_j-1} k_i \zeta_{ji} \right) \left(\sum_{i=1}^{v_j-1} k_i \zeta_{ji} \right)^T A_j \theta_j = \theta_j^T \left(\sum_{i=1}^{v_j-1} k_i \zeta_{ji} \right) \left(\sum_{i=1}^{v_j-1} k_i \zeta_{ji} \right)^T \theta_j.$$

令 $\eta_1 = \sum_{i=1}^{v_j-1} k_i \zeta_{ji}$, 检验 $\eta_1^T \theta_j = 0$ 与检验 $L_j^T \theta_j = 0$

等价.

(iii) 如果 k_1, \dots, k_{v_j-1} 这组值全为 0, 那么 $SS=0$ 恒成立, 也就有 $L_j^T \theta_j = 0$ 恒成立, 此时问题已经解决; 如果 k_1, \dots, k_{v_j-1} 至少有 1 个不为 0, 那么将 η_1 扩张成 A_j 的特征值为 1 的特征向量空间的一组基 $\eta_1, \dots,$

η_{v_j-1} . 对这组基进行规格正交化, 得到 $\eta_1^* = \frac{\eta_1}{\sqrt{\eta_1^T \eta_1}}, \eta_2^*, \dots, \eta_{v_j-1}^*$.

对该组规格正交化特征向量 $\eta_1^*, \eta_2^*, \dots, \eta_{v_j-1}^*$, 利用零成分搜索法判断 TT 对应的子成分 $\eta_1^{*T} \theta_j$ 是否显著可以考察约束条件 $\eta_1^{*T} \theta_j = 0$ 是否成立, 其与检验

$\eta_1^T \Theta_j = 0$ 等价, 进而与检验 $C_j^T \beta_j = 0$ 等价.

4 模拟分析

为了对一般线性约束条件的检验方法有更为直观地认识, 下面将给出一个具体的模拟实例.

例 1 取饱和正交表 $L_9(3^4) =$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 3 & 3 & 3 \\ 2 & 1 & 2 & 3 \\ 2 & 2 & 3 & 1 \\ 2 & 3 & 1 & 2 \\ 3 & 1 & 3 & 2 \\ 3 & 2 & 1 & 3 \\ 3 & 3 & 2 & 1 \end{pmatrix},$$

统计分析模型写成向量形式为

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \\ \mu \end{pmatrix} +$$

$$\begin{pmatrix} \beta_{11} \\ \beta_{11} \\ \beta_{11} \\ \beta_{12} \\ \beta_{12} \\ \beta_{12} \\ \beta_{13} \\ \beta_{13} \\ \beta_{13} \end{pmatrix} + \begin{pmatrix} \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \end{pmatrix} + \begin{pmatrix} \beta_{31} \\ \beta_{32} \\ \beta_{33} \\ \beta_{32} \\ \beta_{33} \\ \beta_{31} \\ \beta_{33} \\ \beta_{31} \\ \beta_{32} \end{pmatrix} + \begin{pmatrix} \beta_{41} \\ \beta_{42} \\ \beta_{43} \\ \beta_{43} \\ \beta_{42} \\ \beta_{42} \\ \beta_{43} \\ \beta_{43} \\ \beta_{41} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{pmatrix}, \text{ 简记为 } Y = \Theta_0 +$$

$\Theta_1 + \Theta_2 + \Theta_3 + \Theta_4 + \varepsilon$. 模型约束条件:

$$\beta_{11} + \beta_{12} + \beta_{13} = 0, \beta_{21} + \beta_{22} + \beta_{23} = 0,$$

$$\beta_{31} + \beta_{32} + \beta_{33} = 0, \beta_{41} + \beta_{42} + \beta_{43} = 0.$$

记

$$\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})^T, \beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})^T,$$

$$\beta_3 = (\beta_{31}, \beta_{32}, \beta_{33})^T, \beta_4 = (\beta_{41}, \beta_{42}, \beta_{43})^T.$$

先随机取定一组参数(效应参数选取时为了更好地考察模型效果, 选取参数时应该满足如下原则: 应该满足基本的约束条件), 再产生 9 个随机数 $\varepsilon_i \sim N(0, \sigma^2)$, 如果该因子为活动因子, 那么其绝对值应该大于 3σ ; 如果该因子为不活动因子, 那么可以取该参数值为 0. 由上述模型可得 9 个观测值. 再由 9 个观测值对参数做估计并且进行方差分析, 并利用上述的约束条件检验方法, 对给定的线性约束条件进行检验. 随机得到一组参数值和随机数值如表 1、表 2 所示.

这里希望检验线性约束条件 $2\beta_{11} = \beta_{12}$ 是否成立, 即若令 $C_1 = (2, -1, 0)^T$, 则约束条件为 $C_1^T \beta_1 = 0$.

不妨取 $L_1 = (2, 0, 0, -1, 0, 0, 0, 0)^T$, 从而检验 $C_1^T \Theta_1 = 0$ 与检验 $C_1^T \beta_1 = 0$ 等价.

再进一步将问题转化, 直到变成一个零成分搜索的问题, 可以得到零成分搜索表(见表 3).

由表 3 可知, 如果选取显著性水平为 0.05, 找到 2 个零成分, 分别对应到第 1 列和第 2 列. 第一个零成分对应的恰是线性约束条件所对应的子成分. 从而, 在统计学意义上, 该约束条件可以认为是成立的.

表 1 总均值及各列效应值

μ	β_1	β_2	β_3	β_4
47	$(3, 6, -9)^T$	$(10, 8, -18)^T$	$(6, 8, -14)^T$	$(8, 12, -20)^T$

表 2 随机数及观测值($\sigma^2=4$)

ε_i	0.472 4	0.616 6	-0.920 1	1.975 5	0.189 0	-2.136 0	-4.069 5	2.340 5	-1.374 8
y_i	74.472 4	78.616 6	-2.920 1	52.975 5	55.189 0	50.864 0	41.930 5	34.340 5	34.625 2

表 3 零成分搜索表

子成分序号	1	2	3	4	5	6	7	8
子平方和	0.1	0.3	294.2	342.9	437.4	714.6	1 261.9	1 650.9
对应列	1	2	4	3	1	3	4	2
W 统计量	0	0.544 6	1.120 2	0.031 1	0.228 8	0.905 4	1.967 3	1.234 4
W 临界值(0.05)	0	3.964 9	0.911 8	2.866 0	5.182 7	7.686 0	12.171 8	24.128 6
P 值	0	0.687 9	0.001 3	0.926 8	0.741 9	0.461 6	0.388 1	0.711 3
是否零成分	Yes	Yes	No	No	No	No	No	No

5 参考文献

- [1] 潘长缘, 陈雪平, 张应山. 正交表列效应的约束条件检验 [J]. 山西大学学报: 自然科学版, 2008, 31(3): 380-384.
- [2] 马海南, 潘长缘, 张应山. 不饱和正交表列效应的约束条件检验 [J]. 数学的实践与认识, 2010, 40(20): 96-101.
- [3] 张应山. 多边矩阵理论 [M]. 北京: 中国统计出版社, 1993.
- [4] 张应山. 正交表的数据分析及其构造 [D]. 上海: 华东师范大学, 2006.
- [5] Zhang Yingshan, Lu Yiqiang, Pang Shanqi. Orthogonal arrays obtained by orthogonal decomposition of projection matrices [J]. Statistica Sinica, 1999, 9(2): 595-604.
- [6] 潘长缘. 正交平衡区组设计 [D]. 上海: 华东师范大学, 2009.
- [7] 游晓锋, 丁树良, 刘红云. 缺失数据的估计方法及应用 [J]. 江西师范大学学报: 自然科学版, 2011, 35(3): 325-330.
- [8] Wang Xiaodi, Tang Yincai, Chen Xueping, et al. Design of experiment in global sensitivity analysis based on ANOVA high-dimension model representation [J]. Communications in Statistics-Simulation and Computation, 2010, 39(6): 1183-1195.
- [9] Wang Xiaodi, Tang Yincai, Zhang Yingshan. Orthogonal arrays for the estimation of global sensitivity indices based on ANOVA high-dimension model representation [J]. Communications in Statistics-Simulation and Computation, 2011, 40(9): 1324-1341.

Test of Linear Constraint Equation in Column Effect of Saturated Orthogonal Arrays

LUO Chun¹, LI Wei², PAN Chang-yuan²

- (1. College of Sciences, Shanghai College of Applied Technology, Shanghai 200235, China;
2. College of Mathematics and Informatics, Jiangxi Normal University, Nanchang Jiangxi 330022, China;
3. Biostatistics Department, Excel PharmaStudies Incorporated, Shanghai 200021, China)

Abstract: Data analysis in orthogonal saturated model is discussed by doing decomposition to matrix images and Procedure of Searching Zero-Decomposition. And the linear constraint equation for some zero-decomposition is deduced. By transforming the test problem of general linear constraint equation to significant test of a decomposition of column effect and applying Procedure of Searching Zero-Decomposition, this problem is perfectly solved at last.

Key words: orthogonal arrays; constraint equation; projection matrices; matrix images; saturated model

(责任编辑: 曾剑锋)