应用概率统计　第二十七卷
第五期　2011年10月

Chinese Journal of Applied Probability
and Statistics Vol.27 No.5 Oct. 2011

单次多水平正交试验中效应的显著性检验 *

陈　颖

(上海财经大学统计与管理学院, 上海, 200433)

《应用概率统计》版权所用

摘　要

方差分析自试验设计诞生以来一直是用于分析试验中各因子是否显著的统计方法, 对于正交试验设计而言其更是唯一的分析方法. 然而, 当正交表各列放满了被考虑的各个因子及其交互作用并且各条件组合下只能进行一次试验时, 方差分析中的误差项将恒等于0, 从而方差分析不再能用于对此试验设计的分析. 对此, 本文针对使用多水平完备正交表的单次正交试验, 提出了一种新的统计分析方法. 示例表明: 本文提出的检验法不仅解决了方差分析无法胜任的问题, 而且在表头设计有空白列从而方差分析仍能实施时, 其比方差分析具有更大的局部功效.

关键词: 饱和试验设计, 多水平, 单次试验, 完备正交表.

学科分类号: O212.6, O213.

§1. 引　　言

正交试验设计被广泛应用于农业、工业和医学制药等各个领域, 其中使用正交表$L_n(q^m)$的正交试验占了很大一部分. 这里, n是正交表的行数, m是正交表的列数, $q\,(q\geq 2)$是水平数. 在实践中, 由于受到试验材料、时间和经费等因素的限制, 或为了节约试验成本, 各条件组合下往往只能做一次试验. 例如, 在某一焊接试验中, 据专家研究影响焊接质量的可能因子有两个: A(焊盘形状)和B(焊料成分), 每个因子都取3个水平, 由于试验材料非常昂贵, 再加上项目预算的限制, 只能使用正交表$L_9(3^4)$进行9次试验. 这就是一个单次3水平正交试验. 又如, 在医学临床试验中, 为发现影响关节炎治疗效果的重要因子和最佳治疗方案考虑2个5水平的因子: A(药物治疗)和B(运动治疗), 用正交表$L_{25}(5^6)$安排试验. 由于要找到25个以上病情相似的病人非常困难再加上人力物力的限制, 从而导致无法实施重复试验, 这是一个单次5水平正交试验.

上述两个试验中, 若两因子的交互作用A×B可能存在, 则在表头设计时各因子及其交互作用将放满正交表各列. 记$SS_i=(n/q)\sum\limits_{l=1}^{q}(\overline{T}_{il}-\overline{y})^2$, $i=1,2,\ldots,m$为正交表第i列的平方和, 其中$\overline{T}_{il}$为第i列第l水平均值, $\overline{y}$为n次试验结果的总平均. 那么许多教科书中都介绍过的一个结论是: $SS_1+SS_2+\cdots+SS_m=SS_T$, 其中$SS_T=\sum\limits_{j=1}^{n}(y_j-\overline{y})^2$是$n$个试验数据的总平方和[1]. 这时, 方差分析中的误差平方和$SS_E=SS_T-(SS_1+SS_2+\cdots+SS_m)\equiv 0$.

* 上海财经大学"211工程"三期重点学科建设项目和上海市自然科学基金资助(11ZR1412000).
本文2009年10月18日收到, 2011年2月17日收到修改稿.

《应用概率统计》版权所用

由此可见, 自从R. Fisher创立试验设计以来, 一直是标准分析方法的方差分析此时不再适用. 对于上述问题, 有人提出用因子水平均值图和贡献率分析来考察各个因子的重要程度. 但这些方法都无法告知人们哪些因子是统计意义上显著的.

大量的实践经验表明, 在筛选试验阶段只有少部分因子或效应是显著的[2], 而在通常情况下, 所考虑的因子及其交互作用各分量也不会全都是显著的. 在此前提下, 本文针对上述问题提出了一种新的统计分析方法. 示例表明, 此检验方法不仅可解决方差分析无法胜任的问题, 而且在表头设计有空白列从而方差分析仍能实施时, 其要比方差分析有更大的局部功效. 本文只限讨论使用完备正交表$L_n(q^m)$的情况, 其中$n=q^k\,(q>2,\,k=2,3,\ldots)$, 且$m$, n和q满足$m(q-1)=n-1$. 而关于$q=2$时的无重复试验的正交试验设计分析方法, 文献中已有很丰富的内容, 对这些方法的回顾和存在的问题总结可参见[3].

§2. 检验统计量及其应用示例

设正交试验中使用的正交表为完备正交表$L_n(q^m)$, 表头设计时被考虑的各个因子及其交互作用不发生混杂但可占满正交表的各列, 各水平组合下只进行了一次试验, 试验误差相互独立均服从正态分布$N(0,\sigma^2)$. 易知, 在各列效应(具体含义见2.2节)均为0的情况下, SS_i/σ^2, $i=1,2,\ldots,m$相互独立且均服从自由度为$q-1$的卡方分布.

记$MS_i=SS_i/(q-1)$, $i=1,2,\ldots,m$为正交表第i列的均方差. 若已知有k列(不妨设为第$i_1,i_2,\ldots,i_k$列)效应不等于零, 则这k列的均方差期望值将比其它零效应列的要大(在以后各小节中将分不同情况论述); 从而

$$\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)$$

可能是所有$\binom{m}{k}$个类似比值(任取k列的均方差之均值比上剩余各列均方差之均值)中最大的. 但实际中, 我们往往不知道k是多少, 只能知道$k\le r\,(1\le r<m)$. 为了能合理比较$\sum_{k=1}^{r}\binom{m}{k}$个此种比值的大小, 先对其进行变换:

$$F_{k(q-1),(m-k)(q-1)}\Big(\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)\Big),$$

其中$F_{k(q-1),(m-k)(q-1)}(\cdot)$是分子分母自由度分别为$k(q-1)$和$(m-k)(q-1)$的$F$分布的分布函数. 当正交表各列效应均为0时,

$$\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)$$

服从分子分母自由度分别为$k(q-1)$和$(m-k)(q-1)$的F分布, 从而

$$F_{k(q-1),(m-k)(q-1)}\Big(\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)\Big)$$

均服从$(0,1)$区间上的均匀分布. 而当有k^*列(不妨设为第$i_1,i_2,\ldots,i_{k^*}$列)效应非零时,

$$F_{k^*(q-1),(m-k^*)(q-1)}\Big(\sum_{j=1}^{k^*}MS_{i_j}/k^*\Big/\sum_{j\notin\{i_1,\ldots,i_{k^*}\}}MS_j/(m-k^*)\Big)$$

可能是所有$\sum_{k=1}^{r}\binom{m}{k}$个值中最大的. 因此, 可用

$$\max_{1\le k\le r}\max_{\{i_1,\ldots,i_k\}\in P_k}\Big\{F_{k(q-1),(m-k)(q-1)}\Big(\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)\Big)\Big\}$$

作为检验H_0: "全部列的效应均为零"的检验统计量, 其中P_k是具有$\{1,2,\ldots,m\}$中k个元素的所有子集所组成的集合. 当上述检验统计量的值大于临界值时, 拒绝H_0. 容易证明:

$$\begin{aligned}&F_{k(q-1),(m-k)(q-1)}\Big(\sum_{i=m-k+1}^{m}MS_{(i)}/k\Big/\sum_{i=1}^{m-k}MS_{(i)}/(m-k)\Big)\\=\;&\max_{\{i_1,\ldots,i_k\}\in P_k}\Big\{F_{k(q-1),(m-k)(q-1)}\Big(\sum_{j=1}^{k}MS_{i_j}/k\Big/\sum_{j\notin\{i_1,\ldots,i_k\}}MS_j/(m-k)\Big)\Big\},\end{aligned}$$

其中$MS_{(1)},MS_{(2)},\ldots,MS_{(m)}$是$MS_1,MS_2,\ldots,MS_m$的次序统计量. 所以实际中使用如下$\mathrm{Max}U_r^{(q)}$检验统计量:

$$\mathrm{Max}U_r^{(q)}=\max_{1\le k\le r}(MU_k),\tag{2.1}$$

其中

$$MU_k=F_{k(q-1),(m-k)(q-1)}\Big(\sum_{i=m-k+1}^{m}MS_{(i)}/k\Big/\sum_{i=1}^{m-k}MS_{(i)}/(m-k)\Big),$$

r是预估的正交表中具有非零效应的最大列数. 使用$\mathrm{Max}U_r^{(q)}$检验是否存在显著性因子或交互作用的步骤如下:

(1) 选定全局显著性水平$\alpha\,(0<\alpha<1)$. 取一正整数r, 使其不小于正交表中真实效应为非零的列数; 若无相关信息可作判断依据, 则取$r=m-1$.

(2) 计算$\mathrm{Max}U_r^{(q)}$; 若$\mathrm{Max}U_r^{(q)}\le c_{m,r;\alpha}^{(q)}$, 则表明所考虑的全部因子和交互作用在全局显著性水平α下均不显著, 停止检验过程.

(3) 若$\mathrm{Max}U_r^{(q)}>c_{m,r;\alpha}^{(q)}$, 则表明存在非零效应列. 此时必存在正整数$k^*\,(1\le k^*\le r)$使得$MU_{k^*}=\mathrm{Max}U_r^{(q)}$, 将平方和(或均方差)最大的$k^*$列看成是非零效应列. 若某非零效应列在表头设计时放置的是某因子, 则认为此因子是显著的; 若某非零效应列对应的是某交互作用(或其中一个分量), 则认为此交互作用是显著的.

这里, $c_{m,r;\alpha}^{(q)}$是相应检验的临界值. 其在全部主效应和交互作用为0的条件下, 满足

$$\mathrm{P}\{\mathrm{Max}U_r^{(q)}>c_{m,r;\alpha}^{(q)}\}=\alpha,\qquad(0<\alpha<1).$$

一些常用的临界值$c_{m,r;\alpha}^{(q)}$列在附录中. 其中, 一些3水平正交表的相关临界值$c_{m,r;\alpha}^{(3)}$取之[4]. 附录中的有关4、5水平正交表的临界值$c_{m,r;\alpha}^{(q)}$是在H_0成立的条件下通过1万次随机模拟

《应用概率统计》版权所用

《应用概率统计》版权所用

获得的, 其模拟过程是这样进行的: 1) 产生一个容量为n的服从标准正态分布的简单随机样本观测值. 2) 将1)中产生的n个值分别作为正交表$L_n(q^m)$的n行的观测值, 再计算均方差$MS_1, MS_2, \ldots, MS_m$. 3) 根据定义(2.1)计算统计量$\mathrm{Max}U_r^{(q)}$的值. 4) 重复上述步骤1)~3) 1万次, 得1万个$\mathrm{Max}U_r^{(q)}$的模拟观测值. 5) 将所得的1万个$\mathrm{Max}U_r^{(q)}$观测值从小到大排列, 其中排在第$[10000(1-\alpha)]$位的值即为$c_{m,r;\alpha}^{(q)}$的估计值. 事实上, 这里的$\mathrm{Max}U_r^{(q)}$检验是Chen和Kunert提出的$\mathrm{Max}U_r$检验[5]在多水平正交设计中的推广.

关于$\mathrm{Max}U_r^{(q)}$的具体使用, 将在以下各小节中介绍. 分成无交互作用、有交互作用和存在空白列三种情况. 后文将分别通过若干示例演示在上述三种不同情况下该检验的具体使用过程和展示其可行性, 同时还通过随机模拟给出了其局部功效. 关于$\mathrm{Max}U_r^{(q)}$检验的全面功效, 还有待进一步研究.

2.1 无交互作用情况

不失一般性, 考虑3水平正交设计: 4个3水平因子A、B、C和D依次放满了正交表$L_9(3^4)$的各列, 各个因子间不存在交互作用, 各水平组合下只进行了一次试验. 相应的统计模型为:

$$y_{ijkl} = \mu + a_i + b_j + c_k + d_l + \varepsilon_{ijkl}, \qquad i,j,k,l = 1,2,3, \tag{2.2}$$

其中, y_{ijkl}, $i,j,k,l = 1,2,3$是因子A取第i水平、因子B取第j水平、因子C取第k水平、因子D取第l水平时的试验结果. μ是总均值, a_i, b_i, c_i, d_i, $i = 1,2,3$分别是因子A、B、C和D的主效应, 其分别满足约束$\sum\limits_{i=1}^{3} a_i = 0$, $\sum\limits_{i=1}^{3} b_i = 0$, $\sum\limits_{i=1}^{3} c_i = 0$, $\sum\limits_{i=1}^{3} d_i = 0$. ε_{ijkl}, $i,j,k,l = 1,2,3$是试验误差, 其相互独立服从均值为0方差为σ^2的正态分布. 需要检验的全局原假设是

$$H_0 : a_i, b_i, c_i, d_i, \qquad i = 1,2,3, \text{ 全为0}. \tag{2.3}$$

若H_0被拒绝, 再判断A、B、C和D中哪些因子效应非零.

在模型(2.2)的条件下, 易知SS_i, $i = 1,2,3,4$的数学期望分别为$\mathsf{E}(SS_1) = 3\sum\limits_{i=1}^{3} a_i^2 + 2\sigma^2$, $\mathsf{E}(SS_2) = 3\sum\limits_{i=1}^{3} b_i^2 + 2\sigma^2$, $\mathsf{E}(SS_3) = 3\sum\limits_{i=1}^{3} c_i^2 + 2\sigma^2$, $\mathsf{E}(SS_4) = 3\sum\limits_{i=1}^{3} d_i^2 + 2\sigma^2$. 实际中, 特别是在试验初始阶段, 工程师往往知道: 不会所有的因子效应都非零, 但不知道哪些因子其效应为0. 在全部考虑的因子中至少有一个因子效应为0的情况是常常满足的. 若原假设(2.3)不成立, 至少有1个而最多有$m-1=3$个非零效应因子, 则此时检验统计量$\mathrm{Max}U_r^{(q)}$的值将比在(2.3)成立时要大, 因此我们可用$\mathrm{Max}U_r^{(q)}$来检验(2.3)是否成立, 并在拒绝H_0之后按照前述法则进一步确定各显著性因子.

例 1 一个使用正交表$L_9(3^4)$只进行了单次试验的3水平4因子的正交试验, 其表头设计和试验结果列在表1中. 这时, $m = 4$, $n = 9$, $q = 3$, 各列的自由度为$q-1 = 2$. 由于4个因子放满了正交表全部各列, 没有剩余自由度可用于误差方差的估计, 所以方差分析无法进行.

《应用概率统计》版权所用

表1 正交表$L_9(3^4)$及其表头设计和试验结果

表头设计	A	B	C	D	
试验号 列号	1	2	3	4	y_{ijkl}
1	1	1	1	1	8.467843
2	1	2	2	2	3.7613059
3	1	3	3	3	-0.536417
4	2	1	2	3	0.143804
5	2	2	3	1	3.236313
6	2	3	1	2	-1.590383
7	3	1	3	2	-3.774542
8	3	2	1	3	-7.074682
9	3	3	2	1	-4.93260
MS_i	64.54	11.91	0.081	17.07	

表2 $\mathrm{Max}U_r^{(q)}$检验计算结果(例1)

k	MU_k	$\mathrm{Max}U_r^{(3)}$	r
1	0.9700615	0.99740	3
2	0.9549506		
3	0.997400		

假设现有经验和专业知识不能判断可能的非零效应因子的最大个数, 故取$r = m - 1$. 然后计算$\mathrm{Max}U_r^{(q)}$, 其结果列在表2中. 对给定的全局显著性水平$\alpha = 0.05$, 查表可得$c_{m,r;\alpha}^{(q)} = 0.9954914$. 由于$\mathrm{Max}U_r^{(q)} = 0.99740 > c_{m,r;\alpha}^{(q)}$, 故拒绝原假设, 认为存在非零效应列. 从表2可见, $MU_3 = \mathrm{Max}U_r^{(q)}$, 所以非零效应列的列数等于3. 其中均方差最大的3列, 即第1、2、4列具有非零效应. 再据表头设计, 即得相应的因子A、B和D是显著的.

事实上, 例1中的试验结果是按照模型(2.2)通过随机模拟产生的. 其中, $\mu = 0$, $a_1 = 4$, $a_2 = 0$, $a_3 = -4$, $b_1 = 2$, $b_2 = 0$, $b_3 = -2$, $c_1 = c_2 = c_3 = 0$, $d_1 = 3$, $d_2 = 0$, $d_3 = -3$, $\varepsilon_{ijkl} \overset{\text{i.i.d.}}{\sim} N(0,1)$, $i,j,k,l = 1,2,3$. 在全局显著性水平$\alpha = 0.05$下, $\mathrm{Max}U_r^{(q)}$正确地识别出了全部非零效应因子.

2.2 有交互作用情况

不失一般性, 考虑5水平正交设计: 2个5水平因子A和B, 以及交互作用A × B放满了正交表$L_{25}(5^6)$的各列, 各水平组合下只进行了一次试验. 相应的统计模型为

$$y_{ij} = \mu + a_i + b_j + ab_{ij} + \varepsilon_{ij}, \qquad i,j = 1,2,\ldots,5, \tag{2.4}$$

《应用概率统计》版权所用

其中, y_{ij}, $i,j=1,2,\ldots,5$是因子A取第i水平、因子B取第j水平时的试验结果. μ是总均值, a_i,b_i, $i=1,2,\ldots,5$分别是因子A和B的主效应, 其分别满足约束$\sum\limits_{i=1}^{5}a_i=0$, $\sum\limits_{i=1}^{5}b_i=0$. ab_{ij}, $i,j=1,2,\ldots,5$是因子A和B的交互效应, 其满足约束$\sum\limits_{i=1}^{5}ab_{ij}=0$, $j=1,2,\ldots,5$, $\sum\limits_{j=1}^{5}ab_{ij}=0$, $i=1,2,\ldots,5$. $\varepsilon_{ij}\overset{\text{i.i.d.}}{\sim}N(0,\sigma^2)$, $i,j=1,2,\ldots,5$. 需要检验的全局原假设是

$$H_0:a_i,b_j,ab_{ij},\qquad i,j=1,2,\ldots,5,\ \text{全为0}. \tag{2.5}$$

若H_0被拒绝, 再判断A、B和$A\times B$中哪些效应非零.

表3 正交表$L_{25}(5^6)$及其表头设计和试验结果

表头设计	A	B	A × B	A × B	A × B	A × B	
试验号 列号	1	2	3	4	5	6	y_{ij}
1	1	1	1	1	1	1	3.1791054
2	1	2	2	2	2	2	2.49015
3	1	3	3	3	3	3	1.2665402
4	1	4	4	4	4	4	1.3101373
5	1	5	5	5	5	5	-1.440634
6	2	1	2	3	4	5	-1.497213
7	2	2	3	4	5	1	5.818685
8	2	3	4	5	1	2	1.0603289
9	2	4	5	1	2	3	-0.476091
10	2	5	1	2	3	4	0.6092975
11	3	1	3	5	2	4	0.3619429
12	3	2	4	1	3	5	-3.569327
13	3	3	5	2	4	1	-3.232852
14	3	4	1	3	5	2	2.3542692
15	3	5	2	4	1	3	2.2106234
16	4	1	4	2	5	3	-1.878346
17	4	2	5	3	1	4	-4.780231
18	4	3	1	4	2	5	0.2574972
19	4	4	2	5	3	1	2.2993385
20	4	5	3	1	4	2	0.1553509
21	5	1	5	4	3	2	-0.665268
22	5	2	1	5	4	3	-0.072242
23	5	3	2	1	5	4	-2.058318
24	5	4	3	2	1	5	-3.249216
25	5	5	4	3	2	1	-3.45545
MS	9.21	0.72	11.20	7.78	1.02	8.03	

《应用概率统计》版权所用

表头设计如表3所示, 易知在模型(2.4)的条件下, SS_i, $i=1,2,\ldots,6$的数学期望分别为

$$\mathsf{E}(SS_1)=5\sum_{i=1}^{5}a_i^2+4\sigma^2,\qquad \mathsf{E}(SS_2)=5\sum_{i=1}^{5}b_i^2+4\sigma^2,$$

$$\begin{aligned}
\mathsf{E}(SS_3) &= \frac{1}{5}[(ab_{11}+ab_{25}+ab_{34}+ab_{43}+ab_{52})^2+(ab_{12}+ab_{21}+ab_{35}+ab_{44}+ab_{53})^2]\\
&\quad+\frac{1}{5}[(ab_{13}+ab_{22}+ab_{31}+ab_{45}+ab_{54})^2+(ab_{14}+ab_{23}+ab_{32}+ab_{41}+ab_{55})^2]\\
&\quad+\frac{1}{5}(ab_{15}+ab_{24}+ab_{33}+ab_{42}+ab_{51})^2+4\sigma^2,\\
\mathsf{E}(SS_4) &= \frac{1}{5}[(ab_{11}+ab_{24}+ab_{32}+ab_{45}+ab_{53})^2+(ab_{12}+ab_{25}+ab_{33}+ab_{41}+ab_{54})^2]\\
&\quad+\frac{1}{5}[(ab_{13}+ab_{21}+ab_{34}+ab_{42}+ab_{55})^2+(ab_{14}+ab_{22}+ab_{35}+ab_{43}+ab_{51})^2]\\
&\quad+\frac{1}{5}(ab_{15}+ab_{23}+ab_{31}+ab_{44}+ab_{52})^2+4\sigma^2,\\
\mathsf{E}(SS_5) &= \frac{1}{5}[(ab_{11}+ab_{23}+ab_{35}+ab_{42}+ab_{54})^2+(ab_{12}+ab_{24}+ab_{31}+ab_{43}+ab_{55})^2]\\
&\quad+\frac{1}{5}[(ab_{13}+ab_{25}+ab_{32}+ab_{44}+ab_{51})^2+(ab_{14}+ab_{21}+ab_{33}+ab_{45}+ab_{52})^2]\\
&\quad+\frac{1}{5}(ab_{15}+ab_{22}+ab_{34}+ab_{41}+ab_{53})^2+4\sigma^2,\\
\mathsf{E}(SS_6) &= \frac{1}{5}[(ab_{11}+ab_{22}+ab_{33}+ab_{44}+ab_{55})^2+(ab_{12}+ab_{23}+ab_{34}+ab_{45}+ab_{51})^2]\\
&\quad+\frac{1}{5}[(ab_{13}+ab_{24}+ab_{35}+ab_{41}+ab_{52})^2+(ab_{14}+ab_{25}+ab_{31}+ab_{42}+ab_{53})^2]\\
&\quad+\frac{1}{5}(ab_{15}+ab_{21}+ab_{32}+ab_{43}+ab_{54})^2+4\sigma^2.
\end{aligned}$$

令

$$\begin{cases}
5c_1=ab_{11}+ab_{25}+ab_{34}+ab_{43}+ab_{52}\\
5c_2=ab_{12}+ab_{21}+ab_{35}+ab_{44}+ab_{53}\\
5c_3=ab_{13}+ab_{22}+ab_{31}+ab_{45}+ab_{54}\\
5c_4=ab_{14}+ab_{23}+ab_{32}+ab_{41}+ab_{55}\\
5c_5=ab_{15}+ab_{24}+ab_{33}+ab_{42}+ab_{51}
\end{cases},\quad
\begin{cases}
5d_1=ab_{11}+ab_{24}+ab_{32}+ab_{45}+ab_{53}\\
5d_2=ab_{12}+ab_{25}+ab_{33}+ab_{41}+ab_{54}\\
5d_3=ab_{13}+ab_{21}+ab_{34}+ab_{42}+ab_{55}\\
5d_4=ab_{14}+ab_{22}+ab_{35}+ab_{43}+ab_{51}\\
5d_5=ab_{15}+ab_{23}+ab_{31}+ab_{44}+ab_{52}
\end{cases},$$

$$\begin{cases}
5e_1=ab_{11}+ab_{23}+ab_{35}+ab_{42}+ab_{54}\\
5e_2=ab_{12}+ab_{24}+ab_{31}+ab_{43}+ab_{55}\\
5e_3=ab_{13}+ab_{25}+ab_{32}+ab_{44}+ab_{51}\\
5e_4=ab_{14}+ab_{21}+ab_{33}+ab_{45}+ab_{52}\\
5e_5=ab_{15}+ab_{22}+ab_{34}+ab_{41}+ab_{53}
\end{cases},\quad
\begin{cases}
5f_1=ab_{11}+ab_{22}+ab_{33}+ab_{44}+ab_{55}\\
5f_2=ab_{12}+ab_{23}+ab_{34}+ab_{45}+ab_{51}\\
5f_3=ab_{13}+ab_{24}+ab_{35}+ab_{41}+ab_{52}\\
5f_4=ab_{14}+ab_{25}+ab_{31}+ab_{42}+ab_{53}\\
5f_5=ab_{15}+ab_{21}+ab_{32}+ab_{43}+ab_{54}
\end{cases}.$$

《应用概率统计》版权所用

则有

$$\mathsf{E}(SS_3)=5\sum_{i=1}^{5}c_i^2+4\sigma^2,\qquad \mathsf{E}(SS_4)=5\sum_{i=1}^{5}d_i^2+4\sigma^2,$$
$$\mathsf{E}(SS_5)=5\sum_{i=1}^{5}e_i^2+4\sigma^2,\qquad \mathsf{E}(SS_6)=5\sum_{i=1}^{5}f_i^2+4\sigma^2.$$

注意到模型(2.4)中交互效应满足的约束条件, 故可得$\sum_{i=1}^{5}c_i=0$, $\sum_{i=1}^{5}d_i=0$, $\sum_{i=1}^{5}e_i=0$, $\sum_{i=1}^{5}f_i=0$. 另外, 还可得到如下方程组:

$$\frac{1}{5}\times\begin{pmatrix}
1&-1&0&0&-1&-2&-1&-1&0&-1&0&1&0&-1&1&0\\
0&1&-1&0&1&0&-1&0&-1&-1&-2&-1&0&0&-1&1\\
0&0&1&-1&0&1&0&-1&1&0&0&-1&-1&-1&-1&-2\\
1&1&1&2&1&1&2&1&1&2&1&1&2&1&1&1\\
1&0&-1&0&0&0&-1&1&0&1&-1&0&-1&-1&-2&-1\\
0&1&0&-1&-1&-1&-1&-2&0&0&1&-1&1&0&0&-1\\
1&1&2&1&2&1&1&1&1&1&1&2&1&2&1&1\\
-1&0&0&1&-1&1&0&0&-2&-1&-1&-1&-1&0&1&0\\
1&0&0&-1&0&0&1&-1&-1&-1&-1&-2&0&1&0&-1\\
1&2&1&1&1&1&1&2&2&1&1&1&1&1&2&1\\
-1&0&1&0&-2&-1&-1&-1&-1&1&0&0&-1&0&0&1\\
0&-1&0&1&1&-1&0&0&0&-1&1&0&-1&-2&-1&-1\\
2&1&1&1&1&2&1&1&1&1&2&1&1&1&1&2\\
-1&1&0&0&-1&0&1&0&-1&0&0&1&-2&-1&-1&-1\\
0&-1&1&0&0&-1&0&1&-1&-2&-1&-1&1&-1&0&0\\
0&0&-1&1&-1&-1&-2&-1&1&0&-1&0&0&1&-1&0
\end{pmatrix}\begin{pmatrix}ab_{11}\\ab_{12}\\ab_{13}\\ab_{14}\\ab_{21}\\ab_{22}\\ab_{23}\\ab_{24}\\ab_{31}\\ab_{32}\\ab_{33}\\ab_{34}\\ab_{41}\\ab_{42}\\ab_{43}\\ab_{44}\end{pmatrix}=\begin{pmatrix}c_1\\c_2\\c_3\\c_4\\d_1\\d_2\\d_3\\d_4\\e_1\\e_2\\e_3\\e_4\\f_1\\f_2\\f_3\\f_4\end{pmatrix}. \tag{2.6}$$

可以验证方程组(2.6)的系数矩阵是非奇异的, 因此交互效应ab_{ij}全为0的充要条件是c_i, d_i, e_i, f_i, $i=1,2,\ldots,5$全为0. 所以检验假设(2.5)等价于检验如下假设

$$H_0:\frac{\mathsf{E}MS_1}{\sigma^2}=\frac{\mathsf{E}MS_2}{\sigma^2}=\cdots=\frac{\mathsf{E}MS_m}{\sigma^2}=1, \tag{2.7}$$

其中, $\mathsf{E}MS_i$, $i=1,2,\ldots,m$是MS_i, $i=1,2,\ldots,m$的数学期望. 这里c_i, d_i, e_i, f_i, $i=1,2,\ldots,5$分别称为第3、4、5、6列的效应. 另外, 只要c_i, d_i, e_i, f_i, $i=1,2,\ldots,5$不全为0, 交互效应, ab_{ij}就不全为0, 也即存在交互作用. 而当ab_{ij}不全为0(即存在交互作用)时, c_i(或d_i, 或e_i, 或f_i), $i=1,2,\ldots,5$可能全为0, 这时MS_3(或MS_4, 或MS_5, 或MS_6)可用于估计σ^2. 当原假设(2.7)不成立, 而至少有1列最多有$m-1=5$列效应不全为0时, 检验统计量$\mathrm{Max}U_r^{(q)}$的值就会偏大, 所以可用$\mathrm{Max}U_r^{(q)}$检验(2.7)也即检验(2.5)是否成立.

例 2 一正交试验设计考虑2个5水平因子A和B, 交互作用可能存在, 使用正交表$L_{25}(5^6)$安排试验, 其表头设计见表3. 由于条件限制, 每一试验条件组合下只能进行1次试验. 这里2个因子及其交互作用放满了正交表全部各列, 不再有多余自由度可用于误差方差的估计, 所以方差分析仍无法进行.

《应用概率统计》版权所用

为了考察真实的非零主效应和交互效应能否被正确识别出, 我们按照模型(2.4)通过随机模拟产生试验结果. 其中, $\mu = 0$, 主效应$a_1 = 2$, $a_2 = 1$, $a_3 = 0$, $a_4 = -1$, $a_5 = -2$, $b_1 = b_2 = \cdots = b_5 = 0$, 交互效应为

$$(ab_{ij}) = \begin{pmatrix} 2 & 1 & 0 & -1 & -2 \\ 0 & 2 & 1 & -3 & 0 \\ 0 & -3 & -2 & 3 & 2 \\ -2 & -3 & 2 & 3 & 0 \\ 0 & 3 & -1 & -2 & 0 \end{pmatrix};$$

$\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0,1)$, $i,j = 1,2,\ldots,5$. 模拟的试验结果见表3中的最后一列.

此例中, $m = 6$, $n = 25$, $q = 5$, 各列的自由度为$q-1 = 4$, 仍假定对可能非零效应的列数一无所知, 故取$r = m - 1 = 5$. $\mathrm{Max}U_r^{(q)}$的计算结果列在表4中.

表4 $\mathrm{Max}U_r^{(q)}$检验计算结果(例2)

k	MU_k	$\mathrm{Max}U_r^{(5)}$	r
1	0.8801972	0.9988763	5
2	0.9283058		
3	0.9652082		
4	0.9988763		
5	0.9821424		

对给定的全局显著性水平$\alpha = 0.05$, 查表可得$c_{6,5;\alpha}^{(5)} = 0.9985537$. 由于$\mathrm{Max}U_r^{(q)} = 0.9988763 > c_{6,5;\alpha}^{(5)}$, 故拒绝原假设, 认为存在非零效应列. 从表4可见, $MU_4 = \mathrm{Max}U_r^{(q)}$, 所以非零效应列的列数等于4. 将其中均方差最大的4列, 即第1、3、4、6列认为是具有非零效应的. 据表头设计, 第1列上放的是因子A, 故因子A显著. 而第3、4、6列都是交互作用A × B的分量, 故交互作用A × B也显著. 此结论与真实的模型完全一致.

有时, 在正交试验中会采取忽略交互作用将其作为误差估计的方法来进行方差分析. 此例中, 若采用这一方法, 认为交互作用A × B不存在, 而将其所在的第3到第6列用于误差的估计, 则相应检验的F比分别为: $F_A = 1.315$, $F_B = 0.103$; 而$\alpha = 0.10$时的临界值$F_{0.10}(4,16) = 2.3327$. 事实上, 在显著性水平$\alpha = 0.30$下, 此方差分析都不能判定因子A是显著的! 由此可见, 在这种情况下将交互作用忽略作为误差来检验因子的主效应是不可行的.

2.3 有空白列的情况

有时, 单次多水平的正交设计也会存在少量的空白列. 此时, 空白列可作为误差的估计, 从而方差分析是可以实施的. 这里我们要举一个例子说明: 此时不仅仍可应用$\mathrm{Max}U_r^{(q)}$检验, 而且$\mathrm{Max}U_r^{(q)}$检验还比相应的方差分析更灵敏.

例 3 考虑一个4水平正交设计: 4个4水平因子A、B、C和D分别依次放在正交表$L_{16}(4^5)$的前4列上, 各因子间不存在交互作用, 其表头设计见表5. 同样, 我们用随机模拟方法产生试验结果, 相应数据见表5中的最后一列. 尽管每一试验条件组合下只能进行1次试验, 但由于存在空白列(第5列), 此时方差分析仍能进行, 相应的方差分析表见表6. 从方差分析表可见, 在0.10显著性水平下, 只有因子B是显著的, 其余因子均不显著.

表5 正交表$L_{16}(4^5)$及其表头设计和试验结果

表头设计	A	B	C	D		
试验号 列号	1	2	3	4	5	y_{ij}
1	1	1	1	1	1	3.9441878
2	1	2	2	2	2	1.7801828
3	1	3	3	3	3	-1.442954
4	1	4	4	4	4	-2.395088
5	2	1	2	3	4	2.6724882
6	2	2	1	4	3	1.9768728
7	2	3	4	1	2	0.0604903
8	2	4	3	2	1	-1.26167
9	3	1	3	4	2	0.6188127
10	3	2	4	3	1	-0.216986
11	3	3	1	2	4	-1.752362
12	3	4	2	1	3	-0.946037
13	4	1	4	2	3	1.955827
14	4	2	3	1	4	1.0241185
15	4	3	2	4	1	-2.494579
16	4	4	1	3	2	-4.885305
MS	3.30	18.88	0.21	3.08	0.67	

表6 相应的方差分析表(例3)

来源	SS	d.f.	MS	F比	p值
A	9.9045336	3	3.3015112	4.9473	0.110954
B	56.64064	3	18.880213	28.2921	0.010598
C	0.6390341	3	0.2130114	0.3192	0.813312
D	9.2467532	3	3.0822511	4.6188	0.120425
误差	2.0019945	3	0.6673315		
总和	78.43296	15			

《应用概率统计》版权所用

《应用概率统计》版权所用

此例用$\mathrm{Max}U_r^{(q)}$进行分析时, $m=5$, $n=16$, $q=4$, 各列的自由度为$q-1=3$. 由于有一空白列, 故取$r=m-1=4$ (若有2列空白列, 则取$r=m-2=3$). $\mathrm{Max}U_r^{(q)}$的计算结果列在表7中.

表7 $\mathrm{Max}U_r^{(q)}$检验计算结果(例3)

k	MU_k	$\mathrm{Max}U_r^{(4)}$	r
1	0.9988226	0.9990638	4
2	0.9971800		
3	0.9990638		
4	0.9915815		
5	0.9821424		

对给定的全局显著性水平$\alpha=0.05$, 查表可得$c_{5,4;\alpha}^{(4)}=0.997699$. 由于$\mathrm{Max}U_r^{(q)}=0.9990638>c_{5,4;\alpha}^{(4)}$, 故拒绝原假设, 认为存在非零效应列. 从表7可见, $MU_3=\mathrm{Max}U_r^{(q)}$, 故非零效应列的列数等于3. 将其中均方差最大的3列, 即第1、2、4列认为是具有非零效应的. 再据表头设计, 即得相应的因子A、B和D是显著的. 此结论与真实的模型完全一致.

事实上, 此例中用于产生试验结果的真实模型是

$$y_{ijkl}=\mu+a_i+b_j+c_k+d_l+\varepsilon_{ijkl},\qquad i,j,k,l=1,2,3,4,$$

其中, y_{ijkl}, $i,j,k,l=1,2,3,4$是因子A取第i水平、因子B取第j水平、因子C取第k水平、因子D取第l水平时的试验结果. 总均值$\mu=0$; 因子A、B、C和D的主效应分别为$a_1=a_2=1$, $a_3=a_4=-1$, $b_1=b_2=2$, $b_3=b_4=-2$, $c_1=c_2=c_3=c_4=0$, $d_1=d_2=1$, $d_3=d_4=-1$; 其分别满足约束$\sum\limits_{i=1}^{4}a_i=0$, $\sum\limits_{i=1}^{4}b_i=0$, $\sum\limits_{i=1}^{4}c_i=0$, $\sum\limits_{i=1}^{4}d_i=0$. ε_{ijkl}, $i,j,k,l=1,2,3,4$是试验误差, 其相互独立服从均值为0方差为1的正态分布.

由于单个样本无法反映检验的功效大小, 因此在例3的模型条件及其参数设置下, 重复进行模拟产生了10万个样本以比较$\mathrm{Max}U_r^{(q)}$检验和方差分析的功效大小. 为了公平地比较这两个检验, 必须将它们犯第一类错误的概率控制在相同水平. 但是, $\mathrm{Max}U_r^{(q)}$检验的是全局原假设

$$H_0: a_i,b_i,c_i,d_i,\qquad i=1,2,3,4,\ \text{全为0}. \tag{2.8}$$

而此例中方差分析分别检验了如下4个原假设:

$$\begin{aligned}
&H_{0A}: a_i,\qquad i=1,2,3,4,\ \text{全为0};\\
&H_{0B}: b_i,\qquad i=1,2,3,4,\ \text{全为0};\\
&H_{0C}: c_i,\qquad i=1,2,3,4,\ \text{全为0};\\
&H_{0D}: d_i,\qquad i=1,2,3,4,\ \text{全为0}.
\end{aligned}$$

《应用概率统计》版权所用

注意到, (2.8)成立的充要条件是H_{0A}、H_{0B}、H_{0C}和H_{0D}同时成立. 因此, 当将方差分析检验$H_{0A}\sim H_{0D}$的显著性水平设置在0.05时, 其错误拒绝全局原假设(2.8)的概率要大于0.05. 所以为了将方差分析的全局显著性水平控制在0.05, 必须将其每个F检验的显著性水平调整在小于0.05的水平上. 由于方差分析中的各个F检验统计量间不是相互独立的, 理论上精确计算其错误拒绝全局原假设(2.8)的概率较困难, 因此通过尝试逐步减小每个F检验的显著性水平用随机模拟的方法来确定方差分析的全局显著性水平. 经尝试, 当将每个F检验的显著性水平设置为0.022时, 方差分析的全局显著性水平为0.05 (10万次模拟结果). 在例3的参数设置下, Max$U_r^{(q)}$检验与方差分析的功效模拟比较结果见表8, 其中功效是这样定义的: 效应非零的因子被正确识别出的比例的期望值[6].

表8 Max$U_r^{(q)}$检验与方差分析的局部功效模拟结果(基于10万次)

检验方法	功效	识别出的显著性因子包含全部真实非零效应因子的比例(%)	识别出的显著性因子正好等于全部真实非零效应因子的比例(%)
Max$U_r^{(q)}$	0.49400	40.754	37.405
方差分析	0.39068	16.475	14.280

从表8可见, 在例3的参数设置条件下, Max$U_r^{(q)}$检验的功效要比方差分析的功效大10个百分点. 由于实际中漏掉重要因子将会导致今后优化过程的失败, 因此在模拟研究中还考察了这样的检验性能: 识别出的显著性因子包含全部真实非零效应因子的比例. 10万次模拟结果表明, Max$U_r^{(q)}$检验的相应性能要比方差分析的高24个百分点. 此外, 在模拟研究中还同时考察了: 识别出的显著性因子正好等于全部真实非零效应因子的比例; 也就是: 正确识别出全部非零因子但又不把效应为零的因子误判为显著的比例, 这是最理想的情况. 相应的模拟结果列在表8的最后一列, 从中可见Max$U_r^{(q)}$正确识别的比例要比方差分析的高23个百分点.

§3. 结 论

由于成本、时间效率、试验条件等因素的关系, 实际中有时一个正交试验的每一水平组合条件下只能进行一次试验. 只进行单次试验的正交试验数据分析的困难在于: 正交表中各列往往被考察的各因子及其交互作用所占满, 从而无剩余自由度可用于误差方差的估计, 使得方差分析无法实施. 而在分析单次多水平($q>2$)正交试验时遇到的另一个困难是: 此时2个因子的交互作用不再像在2水平正交试验时那样与主效应的自由度相同(都为1), 而是等于主效应自由度的$q-1$倍; 从而用于分析单次2水平正交试验的方法不能被直接移植用于单次多水平正交试验的分析. 本文提出的Max$U_r^{(q)}$检验方法, 不仅解决了方差分析不能胜任的问题, 而且能够同时检验主效应和交互效应. 示例还表明, 在正交表存在空白列方差分析仍可实施的情况下, Max$U_r^{(q)}$检验比方差分析更灵敏, 其局部功效比方差分析大. 另外, Max$U_r^{(q)}$检验的计算类似于经典方差分析也很简单, 其不需要预先将各因子或交互作用的平方和分解成方差相等的正交对比, 从而给工程师和统计技术应用人员的使用带来方便.

《应用概率统计》版权所用

附　　录

$\mathrm{Max}U_r^{(3)}$检验临界值表$c_{13,r;\alpha}^{(3)}$, $q=3$, $m=13$

r	α			
	0.100	0.050	0.025	0.010
1	0.9921544	0.9961063	0.9981146	0.9992426
2	0.9977317	0.9990004	0.9995539	0.9998453
3	0.9990122	0.9995887	0.9998198	0.9999397
4	0.9994769	0.9997878	0.999919	0.999971
5	0.9996945	0.9998918	0.9999538	0.9999866
6	0.9997768	0.9999083	0.9999617	0.9999904
7	0.9998312	0.9999338	0.9999733	0.9999913
8	0.9998658	0.9999494	0.9999781	0.9999933
9	0.99987	0.9999557	0.9999815	0.9999946
10	0.9998813	0.9999559	0.9999829	0.9999952
11	0.9998893	0.9999569	0.9999829	0.9999958
12	0.9998974	0.9999628	0.9999847	0.9999959

$\mathrm{Max}U_r^{(3)}$检验临界值表$c_{4,r;\alpha}^{(3)}$, $q=3$, $m=4$

r	α			
	0.100	0.050	0.025	0.010
1	0.9753315	0.9879605	0.9939111	0.9975702
2	0.9867493	0.9938402	0.9969543	0.9989851
3	0.9902062	0.9954914	0.9979828	0.9992306

$\mathrm{Max}U_r^{(q)}$检验临界值表$c_{5,r;\alpha}^{(4)}$, $q=4$, $m=5$

r	α			
	0.100	0.050	0.025	0.010
1	0.979970128471530	0.989938628568160	0.995017848696130	0.998045415832520
2	0.990004925210940	0.995300224575720	0.997732529404840	0.999184818185770
3	0.993507971582500	0.997058819710050	0.998623438549930	0.999485800473380
4	0.994904249059320	0.997699436588660	0.998875286124940	0.999562592711210

$\mathrm{Max}U_r^{(q)}$检验临界值表$c_{6,r;\alpha}^{(5)}$, $q=5$, $m=6$

r	α			
	0.100	0.050	0.025	0.010
1	0.983467977603900	0.991967511997830	0.996280797871090	0.998494571551230
2	0.992796953807090	0.996756633278150	0.998441095673930	0.999314623510910
3	0.995440158862910	0.997921555577300	0.999050315463060	0.999648809161780
4	0.996447198512450	0.998528956159460	0.999329351167090	0.999766785267080
5	0.996737058770750	0.998553743910970	0.999360851262860	0.999729183825050

《应用概率统计》版权所用

参 考 文 献

[1] 茆诗松, 周纪芗, 陈颖, 试验设计, 中国统计出版社, 2004.

[2] Box, G.E.P. and Meyer, R.D., An analysis for unreplicated fractional factorials, *Technometrics*, **28**(1986), 11–18.

[3] 陈颖, 无重复试验的饱和析因设计的数据分析, 应用概率统计, **20(3)**(2004), 313–326.

[4] Chen, Y., Chan, C.K. and Leung, P.K.B., An analysis of three-level orthogonal saturated designs, *Computational Statistics and Data Analysis*, **54(8)**(2010), 1952–1961.

[5] Chen, Y. and Kunert, J., A new quantitative method for analysing unreplicated factorial designs, *Biometrical Journal*, **46(1)**(2004), 125–140.

[6] Hamada, M. and Balakrishnan, N., Analyzing unreplicated factorial experiments: A review with some new proposals, *Statistica Sinica*, **8(1)**(1998), 1–41.

A Significance Test in Multi-Level Orthogonal Designs with Only One Replicate

Chen Ying

(*School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, 200433*)

The ordinary ANOVA has been playing an important role in testing significant factors in experimental designs since it was developed. Especially for orthogonal designs, the ANOVA is almost the unique method of analyzing them. When the columns of an orthogonal array all are allocated by factors or interactions and it is impossible to do replicates, however, the ordinary ANOVA is no more applicable because there are no degrees of freedom left to estimate the error variance. This paper proposes a new test method for analyzing the orthogonal designs with only one replicate using complete orthogonal arrays. It is illustrated with some examples that the proposed method is practicable when the ANOVA is invalid. When there are a few empty columns of the orthogonal array used so that the ANOVA is applicable, the proposed method is more powerful than the ordinary ANOVA for some parameter configurations.

Keywords: Saturated design, multi-level, one replicate, complete orthogonal array.

AMS Subject Classification: 62K15.