二水平正规设计中主效应的混杂性质*

米乃瓦尔·亚森, 李智明†

(新疆大学 数学与系统科学学院, 新疆 乌鲁木齐 830017)

摘　要： 研究了二水平正规设计中主效应与其它各阶因子之间混杂指标集的计算公式, 讨论它与别名矩阵、别名效应数型之间的关系, 根据别名效应数型得到混杂指标集中重要元素的计算公式. 通过表格列出在不同条件下的最优设计并进行比较.

关键词： 别名矩阵; 混杂指标集; 别名效应数型; 二水平正规设计

DOI： 10.13568/j.cnki.651094.651316.2022.03.18.0001

中图分类号： O212.6　**文献标识码：** A　**文章编号：** 2096-7675(2023)02-0175-09

引文格式： 米乃瓦尔·亚森, 李智明. 二水平正规设计中主效应的混杂性质[J]. 新疆大学学报(自然科学版)(中英文), 2023, 40(2): 175-183.

英文引文格式： MINAIWAER Yasen, LI Zhiming. The confounding property of main effects in two-level regular designs[J]. Journal of Xinjiang University(Natural Science Edition in Chinese and English), 2023, 40(2): 175-183.

The Confounding Property of Main Effects in Two-Level Regular Designs

MINAIWAER Yasen, LI Zhiming

(*School of Mathematics and System Sciences, Xinjiang University, Urumqi Xinjiang 830017, China*)

Abstract： The paper aims to study the formulas of the confounding index pattern between main effects and other-order effects in two-level regular designs. The relationship between the alias matrix and aliased effect-number pattern is investigated with the index pattern. Through the aliased effect-number pattern, we provide the formula of the important elements in the confounding index pattern. The optimal designs are listed and compared under different conditions in tables.

Key words： alias matrix; confounding index pattern; aliased effect-number pattern; two-level regular designs

0　引 言

试验设计特别是二水平正规设计在科学研究的各个领域中备受关注. 文献[1]认为一个试验往往有多个设计方案, 试验者如何从中选择最优设计是研究中的难点问题. 研究者给出一些准则下的最优设计, 如最大分辨度准则[2]、最小低阶混杂准则[3]、纯净效应[4]等. 这些准则都基于效应排序原则: 低阶效应比高阶效应重要; 同阶效应同等重要, 它们从不同角度反映了一个设计中不同因子之间的混杂程度. 一个2^{n-m}正规设计D由n个因子$F_1,F_2,\cdots,F_n$组成, 共$N=2^{n-m}$次试验. 每个因子都是二个水平, 用+1和−1表示. 在因子$F_1,F_2,\cdots,F_n$中, 有$n-m$个独立列, m个额外列(称为定义字). 这m个定义字生成的子群称为定义对照子群G. 令A_i为定义对照子群G中字长为$i(i=1,2,\cdots,n)$的字母个数, 称序列$W=(A_1,A_2,A_3,\cdots,A_n)$为设计的字长型(WLP). 文献[3]通过顺序最小化WLP中元素选择最优设计. 文献[5]提出二水平正规设计的别名效应数型来描述所有因子之间的混杂信息, 得到字长型对应于别名效应数型中某些元素的函数, 并在别名效应数型的基础上提出了一般最小低阶混杂

* 收稿日期: 2022-03-18

基金项目: 国家自然科学基金“S水平正规及非正规设计的一般最小低阶混杂理论”(12061070); 新疆维吾尔自治区自然科学基金“因析设计的一般最小低阶混杂准则及性质”(2021D01E13),“多水平一般最小混杂设计的研究及应用”(2018Q011).

作者简介: 米乃瓦尔·亚森(1995-), 女, 硕士生, 从事数理统计的研究, E-mail: 1843072694@qq.com.

† 通讯作者: 李智明(1977-), 女, 博士, 教授, 主要从事数理统计的研究, E-mail: zmli@xju.edu.cn.

准则. 文献[6]计算了别名效应数型中的主要元素. 有关二水平一般最小低阶混杂准则的理论性质及设计构造可参阅文献[7-13].

在最优设计的选择上, 主效应和各阶交互效应的混杂情况往往更为重要. 为了更好地描述主效应与各阶因子交互效应混杂的情况, 文献[14]在某些二因子交互效应重要的先验情况下, 根据线性模型的最小二乘估计提出了描述主效应与各阶因子交互效应混杂的别名矩阵, 并引入混杂指标集来选择最优设计. 但是, 对于该指标集的性质及计算相关研究较少. 本文主要通过别名矩阵及别名效应数型来计算混杂指标集中的重要元素.

1 基本概念

令$q=n-m$及$N=2^q$. 对于一个2^{n-m}正规设计D, 若考虑主效应与其它高阶效应之间的混杂结构, 根据文献[15]建立一个线性模型:

$$Y=\beta_0 I+\sum_{i=1}^{n} X_i\beta_i+\varepsilon \tag{1}$$

其中: Y表示n个观测向量, β_0是均值向量, I表示所有元素为1的列向量, $\beta_r(r=1,2,\cdots,n)$是r阶因子交互效应的参数向量, X_1是设计D对应的$N\times n$矩阵, $X_r(r=2,\cdots,n)$是r阶交互效应的$N\times\binom{n}{r}$矩阵, ε是不相关随机误差向量且$E(\varepsilon)=0, D(\varepsilon)=\sigma^2(\sigma>0)$. 则模型(1)中$\beta_1$的最小二乘估计量$\hat{\beta}_1=(X_1'X_1)^{-1}X_1'Y$, 且

$$E(\hat{\beta}_1)=\beta_1+\sum_{i=2}^{n}(X_1'X_1)^{-1}X_1'X_i\beta_i \tag{2}$$

令$P_r=(X_1'X_1)^{-1}X_1'X_r=(a_{ijr})$, 称之为主效应与$r$阶因子交互的别名矩阵, 简称$r$阶别名矩阵. 因此, $\hat{\beta}_1$与β_1的偏差为

$$B(\hat{\beta}_1,\beta_1)=\sum_{i=2}^{n}(X_1'X_1)^{-1}X_1'X_i\beta_i=\sum_{i=2}^{n}P_i\beta_i \tag{3}$$

(3)式表明$\hat{\beta}_1$与β_1的偏差主要由β_r及$P_r(r=2,\cdots,n)$决定, 但β_r是未知的常数, 因此通过度量P_r来最小化$P_r\beta_r$. 文献[14]提出一种方法: $\|P_r\|^2=\mathrm{tr}(P_r'P_r)=\sum_{i,j}a_{ijr}^2$. 在效应排序原则下, 最小化$\hat{\beta}_1$与$\beta_1$的偏差等价于序列最小化$\|P_2\|^2,\cdots,\|P_n\|^2$. 记$N_r=\|P_r\|^2(r=2,\cdots,n)$, 称$N(D)=(N_2,\cdots,N_n)$为设计$D$的混杂指标集, 序列最小化$N(D)$设计为最优$N$设计.

注意到混杂指标集揭示了主效应与r阶因子交互效应之间的混杂信息. 文献[5]引入二水平设计的效应别名数型来反映所有因子的混杂信息, 包括主效应与各因子之间的混杂. 考虑一个2^{n-m}正规设计D, 令${}_i^{\#}C_j^{(l)}$为与l个j效应别名的i阶效应的个数, 称l为i阶效应与j阶效应的混杂程度, 则集合

$$\left\{{}_i^{\#}C_j^{(l)}, i,j=0,1,\cdots,n; l=0,1,\cdots,\binom{n}{j}\right\}$$

反映了二水平设计不同阶因子混杂的信息. 显然, l值越大, i阶效应与j阶效应混杂的个数越多. 记${}_i^{\#}C_j$=(${}_i^{\#}C_j^{(0)}$, ${}_i^{\#}C_j^{(1)}$, $\cdots$, ${}_i^{\#}C_j^{(K_j)}$), 其中$i,j=0,1,\cdots,n; K_j=\binom{n}{j}$. 根据效应排序原则, 对${}_i^{\#}C_j$按照以下规则排序: 若满足如下三个条件之一: $\max\{i_1,j_1\}<\max\{i_2,j_2\}$, $\max\{i_1,j_1\}=\max\{i_2,j_2\}$且$i_1<i_2$, 或$\max\{i_1,j_1\}=\max\{i_2,j_2\}$, $i_1=i_2$且$j_1<j_2$, 则${}_{i_1}^{\#}C_{j_1}$放到${}_{i_2}^{\#}C_{j_2}$的前面. 称

$${}^{\#}C(D)=({}_1^{\#}C_1,{}_0^{\#}C_2,{}_1^{\#}C_2,{}_2^{\#}C_1,{}_2^{\#}C_2,{}_0^{\#}C_3,{}_1^{\#}C_3,{}_2^{\#}C_3,{}_3^{\#}C_1,{}_3^{\#}C_2,{}_3^{\#}C_3,\cdots) \tag{4}$$

为设计D的别名效应数型(Aliased Effect-Number Pattern, 记AENP). 基于AENP, 顺序最大化(4)式得到设计称为一般最小低阶混杂(General Minimum Lower-Order Confounding)设计, 简记为GMC设计.

在AENP(4)中, ${}_1^{\#}C_j^{(l)}$表示与l个j阶效应混杂的主效应个数, 因此集合$\{{}_1^{\#}C_j^{(l)}, j=0,1,\cdots,n, l=0,1,\cdots, K_j\}$反映了主效应与其它效应之间所有的混杂信息, 其中$K_j=\binom{n}{j}$. 当$l$越大, 主效应与其它效应之间的混杂就越严重. 类似的, 根据混杂程度l得到向量

$${}_1^{\#}C_j=({}_1^{\#}C_j^{(0)},{}_1^{\#}C_j^{(1)},\cdots,{}_1^{\#}C_j^{(K_j)}) \tag{5}$$

其中$K_j=\binom{n}{j}$. 由于${}_1^{\#}C_0=(n)$, 可忽略不计. 按照效应排序原则对${}_1^{\#}C_j$进行排序, 若$j_1<j_2$, 则${}_1^{\#}C_{j_1}$置于${}_1^{\#}C_{j_2}$之前, 得到一个新的排序方式

$$M(D)=({}_1^{\#}C_2(D),{}_1^{\#}C_3(D),{}_1^{\#}C_4(D),{}_1^{\#}C_5(D),{}_1^{\#}C_6(D),\cdots) \tag{6}$$

称之为基于主效应的别名效应数型(Aliased Effect-Number Pattern Based on Main Effect, 记M-AENP). 顺序最大化式(6)得到设计称之为基于主效应的一般最小低阶混杂(General Minimum Lower-Order Confounding Based on Main Effects)设计, 简记为M-GMC设计.

例 1　考虑两个非同构的2^{12-7}设计

$$\begin{aligned}D_1&=\{1,2,3,4,5,345,234,235,245,123,124,134\},\\ D_2&=\{1,2,3,4,5,345,234,123,125,145,124,134\}.\end{aligned}$$

由(6)式可得

$$\begin{aligned}M(D_1)&=(12;0^{13},12;12;0^{50},12;12;0^{50},12;12;0^{13},12;12;0,12),\\ M(D_2)&=(12;0^{12},4,8;12;0^{50},8,0^3,4;12;0^{44},4,0^5,8;12;0^{13},8,0^3,4;12;4,8).\end{aligned}$$

根据混杂指标集定义, 可计算

$$N(D_1)=(0,156,0,600,0,600,0,156,0,12),\ N(D_2)=(0,152,0,616,0,576,0,172,0,8),$$

其中0^s指连续s个0. 设计$D_i(i=1,2)$中所有主效应都不与二阶交互效应别名, 由${}_1^{\#}C_3^{(12)}(D_2)>{}_1^{\#}C_3^{(12)}(D_1)$, $N_3(D_2)<N_3(D_1)$. 因此, 在M-GMC准则及混杂指标集下设计D_2优于D_1, 且D_2既是M-GMC设计、又是最优N设计. 但GMC准则下D_1优于D_2, 因此D_1为GMC设计.

由例1可知, M-GMC准则与GMC准则下的最优设计不一定是同一个设计.

2　主要结果

对二水平正规设计, 各因子效应之间的关系要么混杂, 要么正交, 因此模型(2)中r阶别名矩阵$P_r=(X_1'X_1)^{-1}X_1'X_r(r=2,3,\cdots,n)$中的元素只能取1或0. 令$a_{ijr}$为$r$阶别名矩阵$P_r$中第$i(i=1,2,\cdots,n)$行、第$j(j=1,2,\cdots,\binom{n}{r})$列的元素. 在表1中, 第$i$行表示第$i$个主效应, 第$j$列表示第$j$个$r$阶交互效应. 如果$a_{ijr}=1$, 意味着第$i$个主效应与第$j$个$r$阶交互效应别名, 否则正交, 即

$$a_{ijr}=\begin{cases}1, & \text{第}i\text{个主效应与第}j\text{个}r\text{阶交互效应混杂},\\ 0, & \text{第}i\text{个主效应与第}j\text{个}r\text{阶交互效应正交}.\end{cases}$$

令$a_{i.r}$为P_r第i行所有元素之和且$a_{i.r}=\sum_{j=1}^{K_r}a_{ijr}$, 其中$K_r=\binom{n}{r}$.

表 1　2^{n-m}正规设计的r阶别名矩阵P_r

P_r	1	2	$\cdots$	$j-1$	$\cdots$	K_r	$a_{i.r}$
1	a_{11r}	a_{12r}	$\cdots$	$a_{1(j-1)r}$	$\cdots$	a_{1K_rr}	$\sum_{j=1}^{K_r}a_{1jr}$
2	a_{21r}	a_{22r}	$\cdots$	$a_{2(j-1)r}$	$\cdots$	a_{2K_rr}	$\sum_{j=1}^{K_r}a_{2jr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	a_{n1r}	a_{n2r}	$\cdots$	$a_{n(j-1)r}$	$\cdots$	a_{nK_rr}	$\sum_{j=1}^{K_r}a_{nK_rr}$

由矩阵P_r可知, $a_{i.r}$可能取值分别为$0,1,2,\cdots,K_r$. 令m_{lr}表示$a_{i.r}$取某一数值对应的频数, 见表2.

表 2 $a_{i.r}$取值的频数分布

$a_{i.r}$	0	1	2	3	$\cdots$	j	$\cdots$	K_r
m_{lr}	m_{0r}	m_{1r}	m_{2r}	m_{3r}	$\cdots$	m_{jr}	$\cdots$	$m_{K_r r}$

因为$a_{ijr}^2=a_{ijr}(i=1,\cdots,n;j=1,\cdots,K_r;r=2,\cdots,n)$, 根据$N_r$及$m_{lr}$的定义可得,

$$N_r=\|P_r\|^2=\sum_{i=1}^{n}\sum_{j=1}^{K_r}a_{ijr}^2=\sum_{i=1}^{n}\sum_{j=1}^{K_r}a_{ijr}=\sum_{i=1}^{n}a_{i.r}=\sum_{l=0}^{K_r}lm_{lr} \tag{7}$$

根据(7)式直接可得以下等价性质.

定理 1 最小序列化$N=(N_2,\cdots,N_n)$等价于最小序列化

$$\left(\sum_{i=1}^{n}a_{i.2},\cdots,\sum_{i=1}^{n}a_{i.n}\right) \text{ 或 } \left(\sum_{l=0}^{K_2}lm_{l2},\cdots,\sum_{l=0}^{K_n}lm_{lK_n}\right).$$

当$a_{i.r}=l(l=0,1,\cdots,K_r)$对应的频数为$m_{lr}$时, 说明与$l$个$r$阶交互效应别名的主效应个数为$m_{lr}$. 根据${}_1^{\#}C_r^{(l)}$的定义, 得到

$${}_1^{\#}C_r^{(l)}(D)=m_{lr} \tag{8}$$

其中: $l=0,1,\cdots,K_r;r=2,\cdots,n$. 根据(8)式得到

$$N_r=\sum_{l=0}^{K_r}l{}_1^{\#}C_r^{(l)} \tag{9}$$

根据式(9), 可以得到最小序列化$N=(N_2,\cdots,N_n)$的另一个等价定义.

定理 2 最小序列化$N=(N_2,\cdots,N_n)$等价于最小序列化

$$\left(\sum_{l=0}^{K_2}l{}_1^{\#}C_2^{(l)},\cdots,\sum_{l=0}^{K_n}l{}_1^{\#}C_n^{(l)}\right).$$

定理1揭示了混杂指标集N与别名矩阵P_r之间的关系, 定理2反映了N与基于主效应的别名效应数型之间的关系.

例 2 考虑一个2^{6-2}设计$D=\{1,2,3,4,23,34\}$. 通过计算可得设计D的r阶别名矩阵$P_r(r=2,3,4,5)$. 以2阶别名矩阵P_2为例(见表3), 其它类似. 由表3可知, $a_{i.2}$的取值为0,1,2, 对应的频数m_{l2}分别为1,4,1. 用别名矩阵P_2计算设计D的N_2值: $N_2=a_{2.1}+a_{2.2}+a_{3.2}+a_{4.2}+a_{5.2}+a_{6.2}=0+1+2+1+1+1=6$. 类此的, 可以计算$N_3=4,N_4=6,N_5=2$. 因此, $N(D)=(6,4,6,2)$. 由式(8)可计算${}_1^{\#}C_2^{(0)}=m_{02}=1,{}_1^{\#}C_2^{(1)}=m_{12}=4,{}_1^{\#}C_2^{(2)}=m_{22}=1$. 所以, ${}_1^{\#}C_2(D)=(1,4,1,0,\cdots,0)$. 同样, 通过别名矩阵$P_3,P_4,P_5$可得${}_1^{\#}C_3(D)=(2,4,0,\cdots,0)$, ${}_1^{\#}C_4(D)=(1,4,1,0,\cdots,0)$及${}_1^{\#}C_5(D)=(4,2,0,\cdots,0)$.

表 3 二阶别名矩阵P_2

P_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$a_{i.2}$
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
3	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	2
4	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1

根据式(7), 利用别名矩阵P_r计算N_r比较直观及简洁, 但当因子个数增加时, 别名矩阵的计算过程比较繁杂. 在序列$N=(N_2,\cdots,N_n)$中, 元素N_2最重要. 下面以$r=2$为例, 通过${}_1^{\#}C_2^{(l)}$计算N_2.

定理 3 若2^{n-m}设计$D\subset F_{qq}$, 则$N_2(D)=0$.

证明 因为$D\subset F_{qq}$, 根据文献[7]知, ${}_1^{\#}C_2^{(0)}=n$和${}_1^{\#}C_2^{(l)}=0(l\geq 1)$. 由式(9)可得$N_2(D)=0$.

定理 4 设2^{n-m}设计$D=(D\setminus S_{qr})\cup S_{qr}$, 其中$S_{qr}=H_q\setminus H_r$. 则

$$N_2(D)=(n-N/2)(N-2^r)+\sum_{l\geq (N-2^r)/2} l_1^{\#}C_2^{(l-(N-2^r)/2)}(D\setminus S_{qr}).$$

证明 因为$D=\{D\setminus S_{qr},S_{qr}\}$, 根据文献[7]可知当$l=n-N/2$时, ${}_1^{\#}C_2^{(l)}(D)=N-2^r$. 当$l\geq (N-2^r)/2$时, ${}_1^{\#}C_2^{(l)}(D)={}_1^{\#}C_2^{(l-(N-2^r)/2)}(D\setminus S_{qr})$. 由式(9)可得结论.

例 3 考虑一个2^{10-6}设计$D=(D\setminus S_{43})\cup S_{43}=\{23,123\}\cup\{4,14,24,124,34,134,234,1234\}$.

显然, ${}_1^{\#}C_2(D\setminus S_{43})=(2)$, 其中$n-N/2=10-8=2,(N-2^r)/2=(16-2^3)/2=4$. 根据文献[7]可得${}_1^{\#}C_2(D)=(0,0,8,0,2)$. 因此, 由定理4知$N_2(D)=2\times(16-8)+4\times 2=24$.

下面利用AENP中的元素表达混杂指标集中N_3,N_4,N_5,N_6,N_7元素的计算公式.

定理 5 对分辨度$R\geq III$的2^{n-m}正规设计, 则

$$\begin{aligned}
N_3 &= \frac{2}{3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)},\\
N_4 &= \frac{1}{2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)}-\frac{1}{6}(n-3)\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)},\\
N_5 &= \frac{3}{5\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)}-\frac{2(n-4)}{5\times 3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)},\\
N_6 &= \frac{1}{5}\sum_{l=0}^{K_4} l_3^{\#}C_4^{(l)}-\frac{(n-5)}{5\times 2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)}-\frac{(n-3)}{5\times 2}\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)},\\
N_7 &= \frac{8}{10\times 7}\sum_{l=0}^{K_4} l_4^{\#}C_4^{(l)}-\frac{9(n-6)}{10\times 7\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)}+\frac{(n-4)(n-14)}{10\times 7\times 2}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)}.
\end{aligned}$$

证明 对2^{n-m}设计, 通过文献[10]给出的A_4,A_5,A_6和文献[16]给出的A_i, A_{i-2}, ${}_1^{\#}C_{i-1}$之间的关系, 可得

$$\sum_{l=0}^{K_3} l_1^{\#}C_3^{(l)} = \frac{2}{3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)} \tag{10}$$

$$\sum_{l=0}^{K_4} l_1^{\#}C_4^{(l)} = \frac{1}{2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)}-\frac{1}{6}(n-3)\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)} \tag{11}$$

$$\sum_{l=0}^{K_5} l_1^{\#}C_5^{(l)} = \frac{3}{5\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)}-\frac{2(n-4)}{5\times 3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)} \tag{12}$$

由式(9)和式(10~12)得

$$N_3=\frac{2}{3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)},\quad N_4=\frac{1}{2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)}-\frac{1}{6}(n-3)\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)},$$

$$N_5=\frac{3}{5\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)}-\frac{2(n-4)}{5\times 3}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)}.$$

由文献[17]可以给出

$$A_7=\frac{1}{7\times 5}\sum_{l=0}^{K_4} l_3^{\#}C_4^{(l)}-\frac{(n-5)}{7\times 5}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)}+\frac{(n-3)(n-6)}{7\times 5\times 2}\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)} \tag{13}$$

$$A_8 = \frac{1}{10\times 7}\sum_{l=0}^{K_4} l_4^{\#}C_4^{(l)} - \frac{(n-6)}{10\times 7}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)} + \frac{(n-4)(n-7)}{10\times 7\times 2}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)} \tag{14}$$

由文献[16]可得

$$A_7 = \frac{1}{7}\sum_{l=0}^{K_6} l_1^{\#}C_6^{(l)} - \frac{(n-5)}{7\times 5}\sum_{l=0}^{K_4} l_1^{\#}C_4^{(l)} + \frac{(n-3)(n-5)}{7\times 5\times 3}\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)} \tag{15}$$

$$A_8 = \frac{1}{8}\sum_{l=0}^{K_7} l_1^{\#}C_7^{(l)} - \frac{(n-6)}{8\times 6}\sum_{l=0}^{K_5} l_1^{\#}C_5^{(l)} + \frac{(n-4)(n-6)}{8\times 6\times 4}\sum_{l=0}^{K_3} l_1^{\#}C_3^{(l)} \tag{16}$$

由式(13), (15)知

$$\sum_{l=0}^{K_6} l_1^{\#}C_6^{(l)} = \frac{1}{5}\sum_{l=0}^{K_4} l_3^{\#}C_4^{(l)} - \frac{(n-5)}{5\times 2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)} - \frac{(n-3)}{5\times 2}\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)},$$

由式(14), (16)知

$$\sum_{l=0}^{K_7} l_1^{\#}C_7^{(l)} = \frac{8}{10\times 7}\sum_{l=0}^{K_4} l_4^{\#}C_4^{(l)} - \frac{9(n-6)}{10\times 7\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)} + \frac{(n-4)(n-14)}{10\times 7\times 2}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)},$$

由式(9)得

$$\begin{aligned} N_6 &= \frac{1}{5}\sum_{l=0}^{K_4} l_3^{\#}C_4^{(l)} - \frac{(n-5)}{5\times 2}\sum_{l=0}^{K_3} l_2^{\#}C_3^{(l)} - \frac{(n-3)}{5\times 2}\sum_{l=0}^{K_2} l_1^{\#}C_2^{(l)}, \\ N_7 &= \frac{8}{10\times 7}\sum_{l=0}^{K_4} l_4^{\#}C_4^{(l)} - \frac{9(n-6)}{10\times 7\times 2}\sum_{l=0}^{K_3} l_3^{\#}C_3^{(l)} + \frac{(n-4)(n-14)}{10\times 7\times 2}\sum_{l=0}^{K_2} l_2^{\#}C_2^{(l)}. \end{aligned}$$

3 部分最优设计表格

为比较混杂指标集下的最优设计与其它准则下最优设计的关系, 列出了部分二水平最优N设计、M-GMC设计和GMC设计. 每个2^{n-m}设计的q个独立列及m个额外列的取法可参阅文献[5]. 为方便起见, 表中第一列表示2^{n-m}最优N设计, 简记$n-m.i$; 第二列对应该设计的额外列; 第三列给出设计的混杂指标集$N_r(r=2,3,\cdots,n)$的值; 第四列给出M-AENP中${}_1^{\#}C_j$的元素; 第五、六列分别对比最优N设计, M-GMC设计、GMC设计. 记"Y"表示该设计为最优设计,"N"表示在该准则下不是最优设计.

由表4可知, 试验次数为16的2^{n-m}设计中, 在混杂指标集下序列最小化$(N_2,N_3,\cdots,N_n)$与序列最大化$({}_1^{\#}C_2, {}_1^{\#}C_3, {}_1^{\#}C_4,\cdots)$, $({}_1^{\#}C_2, {}_2^{\#}C_2, {}_1^{\#}C_3, {}_2^{\#}C_3,\cdots)$选的最优设计是相同的.

表 4 16次试验的二水平设计

最优N设计	额外列	N_2,N_3,N_4	${}_1^{\#}C_2;{}_1^{\#}C_3;{}_1^{\#}C_4$	M-GMC	GMC
6-2.1	14 7	0,12,0	6;0,0,6;6	Y	Y
7-3.1	14 7 11	0,28,0	7;0^4,7;7	Y	Y
8-4.1	14 7 11 13	0,56,0	8;0,8;8	Y	Y
9-5.1	14 7 11 13 3	12,56,64	0,8,0^2,1;1,0^6,8;0^7,8,1	Y	Y
10-6.1	14 7 11 13 3 6	24,72,136	0^2,8,0,2;0^4,2,0^3,8;0^{12},2,0,8	Y	Y
11-7.1	14 7 11 13 3 6 12	36,104,236	0^3,8,3;0^8,3,0,8;0^{20},3,0,8	Y	Y
12-8.1	14 7 11 13 3 6 12 9	48,156,384	0^4,12;0^{13},12;0^{32},12	Y	Y

由表5可知，当$n=7,8,10,13,\cdots,19$以及$n=22,23,\cdots,28$时，试验次数为32次的最优N设计既是M-GMC设计又是GMC设计．2^{12-7},2^{20-15},2^{21-16}最优N设计是M-GMC设计但不是GMC设计．2^{9-4},2^{11-6}最优N设计既不是M-GMC设计也不是GMC设计．

表 5　32次试验的二水平设计

最优N设计	额外列	N_2,N_3,N_4	${}^{\#}_{1}C_2;{}^{\#}_{1}C_3$	M-GMC	GMC
7-2.1	30 7	0,4,10	7;3,4	Y	Y
8-3.1	30 7 11	0,12,20	8;2,0,0,6	Y	Y
9-4.1	30 7 11 19	0,24,40	9;1,0,0,8	**N**	**N**
9-4.2	30 7 11 13	0,28,35	9;2,0^3,7	Y	Y
10-5.1	30 7 11 19 29	0,40,80	10;0^4,10	Y	Y
11-6.1	28 14 7 19 25 11	0,100,0	11;0^9,10,1	**N**	**N**
11-6.2	28 14 22 26 7 11	0,104,0	11;0^8,3,0,8	Y	Y
12-7.1	28 14 7 19 25 11 13	0,152,0	12;0^{12},4,8	Y	**N**
12-7.2	28 14 22 26 7 11 13	0,156,0	12;0^{13},12	**N**	Y
13-8.1	28 14 22 26 7 11 13 19	0,220,0	13;0^{16},1,12	Y	Y
14-9.1	28 14 22 26 7 11 13 19 21	0,308,0	14;0^{22},14	Y	Y
15-10.1	28 14 22 26 7 11 13 19 21 25	0,420,0	15;0^{28},15	Y	Y
16-11.1	28 14 22 26 7 11 13 19 21 25 31	0,560,0	16;0^{35},16	Y	Y
17-12.1	28 14 22 26 7 11 13 19 21 25 31 3	24,560,672	0,16,0^6,1;1,0^{34},16	Y	Y
18-13.1	28 14 22 26 7 11 13 19 21 25 31 3 6	48,592,1360	0,0,16,0^5,2;0^8,2,0^{27},16,	Y	Y
19-14.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12	72,656,2104	0^3,16,0^4,3;0^{16},3,0^{21},16	Y	Y
20-15.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 24	96,752,2944	0^4,16,0^3,4;0^{24},4,0^{16},16	Y	**N**
20-15.2	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9	96,756,2944	0^4,16,0^3,4;0^{25},4,0^{15},16	**N**	Y
21-16.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 24 17	120,880,3925	0^5,16,0^2,5;0^{32},5,0^{12},16	Y	**N**
21-16.2	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24	120,884,3920	0^5,16,0^2,5;0^{32},1,4,0^{11},16	**N**	Y
22-17.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18	144,1052,5072	0^6,16,0,6;0^{42},6,0^7,16	Y	Y
23-18.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23	168,1260,6440	0^7,16,7;0^{52},7,0^3,16	Y	Y
24-19.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23 29	192,1512,8064	0^8,24;0^{63},24	Y	Y
25-20.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23 29 5	228,1768,9952	0^9,24,0,0,1;0^{64},1,0^6,24	Y	Y
26-21.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23 29 5 10	264,2072,12184	0^{10},24,0,2;0^{76},2,0^3,24	Y	Y
27-22.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23 29 5 10 20	300,2424,14820	0^{11},24,3;0^{88},3,0,24	Y	Y
28-23.1	28 14 22 26 7 11 13 19 21 25 31 3 6 12 9 24 18 23 29 5 10 20 27	336,2828,17920	0^{12},28;0^{101},28	Y	Y

由表6可知，当$n=8,\cdots,12,19,20,29,30,31,32$时，试验次数为64的最优$N$设计既是M-GMC设计又是GMC设

计. 2^{23-17}, 2^{24-18}, 2^{28-22}最优N设计是M-GMC设计但不是GMC设计. 当$n=13,\cdots,18,21,22,25,26,27$时, 最优$N$设计既不是M-GMC设计也不是GMC设计.

表 6 64次试验的二水平设计

最优N设计	额外列	N_2,N_3,N_4	${}^{\#}_{1}C_2;{}^{\#}_{1}C_3$	M-GMC	GMC
8-2.1	60 15	0,0,10	8;8	Y	Y
9-3.1	60 15 22	0,4,20	9;5,4	Y	Y
10-4.1	60 15 22 39	0,8,40	10;2,8	Y	Y
11-5.1	60 15 22 39 21	0,16,70	11;1,4,6	Y	Y
12-6.1	60 15 22 39 21 59	0,24,120	12;0^2,12	Y	Y
13-7.1	60 15 22 39 19 41 26	0,56,140	13;0^3,5,0,4	**N**	**N**
14-8.1	60 15 22 35 26 37 19 46	0,88,200	14;0^4,2,0,4,8	**N**	**N**
15-9.1	60 15 22 35 26 37 19 46 59	0,120,300	15;0^8,15	**N**	**N**
16-10.1	60 15 22 35 26 37 19 46 59 29	0,172,405	16;0^{10},12,0,0,4	**N**	**N**
17-11.1	60 15 22 35 26 37 19 49 29 55 41	0,236,560	17;0^{12},9,0^3,8	**N**	**N**
18-12.1	60 15 22 35 26 37 19 49 29 55 41 50	0,312,720	18;0^{14},6,0^4,12	**N**	**N**
19-13.1	60 15 22 35 26 49 37 55 19 50 29 46 41	0,400,960	19;0^{16},3,0^5,16	Y	Y
20-14.1	60 15 22 35 26 49 37 55 19 50 29 46 41 59	0,500,1280	20;0^{25},20	Y	Y
21-15.1	56 11 22 37 7 59 28 42 14 49 19 38 21 41 26	0,816,0	21;0^{38},6,12,3	**N**	**N**
22-16.1	56 11 22 37 7 59 28 42 14 49 19 38 21 41 26 44	0,1000,0	22;0^{45},12,10	**N**	**N**
23-17.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35	0,1216,0	23;0^{51},1,1,21	Y	**N**
24-18.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38	0,1460,0	24;0^{60},4,20	Y	**N**
25-19.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52	0,1740,0	25;0^{69},10,15	**N**	**N**
26-20.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 25	0,2060,0	26;0^{79},20,6	**N**	**N**
27-21.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25	0,2420,0	27;0^{89},10,17	**N**	**N**
28-22.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25 31	0,2824,0	28;0^{100},4,24	Y	**N**
29-23.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25 31 44	0,3276,0	29;0^{112},1,28	Y	Y
30-24.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25 31 44 41	0,3780,0	30;0^{126},30	Y	Y
31-25.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25 31 44 41 62	0,4340,0	31;0^{140},31	Y	Y
32-26.1	56 11 22 37 7 59 28 42 14 49 13 26 47 50 19 21 35 38 52 55 25 31 44 41 62 61	0,4960,0	32;0^{155},32	Y	Y

4 结 论

本文主要研究2^{n-m}正规设计中主效应与其它各阶因子之间混杂指标集的计算公式. 首先通过线性回归模型的最小二乘估计的偏差得到了主效应与r阶因子交互效应之间的r阶别名矩阵P_r. 在低阶效应的基础上, 提出基于主效应的别名效应数型及一般最小低阶混杂准则. 分析了r阶别名矩阵P_r、混杂指标集以及别名效应数型之间的关系. 利用别名矩阵P_r计算N_r更直观及简洁, 但当因子个数增加时, 别名矩阵的计算过程比较繁杂. 在

序列$N=(N_2,\cdots,N_n)$中元素N_2最重要，因此通过M-AENP函数中${}^{\#}_{1}C_2^{(l)}$给出N_2的计算公式(定理3及定理4)．定理5给出了通过AENP函数计算混杂指标集中的元素．最后一节通过表格列出试验次数为16、32及64次试验的二水平设计来比较混杂指标集、M-GMC准则、GMC准则下的最优设计．

参考文献：

[1] MUKERJEE R, WU C F J. A modern theory of factorial designs[M]. New York: Spring Science Business Media Inc, 2006.

[2] BOX G E P, HUNTER J S. The 2^{k-p} fractional factorial designs part I[J]. Technometrics, 1961, 3(3): 311-351.

[3] FRIES A, HUNTER W G. Minimum aberration 2^{k-p} designs[J]. Technometrics, 1980, 22(4): 601-608.

[4] WU C F J, CHEN Y Y. A graph-aided method for planning two-level experiments when certain interactions are important[J]. Technometrics, 1992, 34(2): 162-175.

[5] ZHANG R C, LI P, ZHAO S L, et al. A general minimum lower-order confounding criterion for two-level regular designs[J]. Statistica Sinica, 2008, 18(4): 1689-1705.

[6] LI Z M, ZHANG R C. Results for two-level designs with general minimum lower-order confounding[J/OL]. The Scientific World Journal, 2015, 1-9[2021-12-18]. http://dx.doi.org/10.1155/2015/163234.

[7] LI P F, ZHAO S L, ZHANG R C. A theory on constructing designs with general minimum lower-order confounding[J]. Statistica Sinica, 2011, 21(4): 1571-1589.

[8] ZHANG R C, CHENG Y. General minimum lower order confounding designs: an overview and a construction theory[J]. Journal of Statistical Planning and Inference, 2010, 140(7): 1719-1730.

[9] HU J W, ZHANG R C. Some results on two-level regular designs with general minimum lower-order confounding[J]. Journal of Statistical Planning and Inference, 2011, 141(5): 1774-1782.

[10] CHEN J, LIU M Q. Some theory for constructing general minimum lower order confounding designs[J]. Statistica Sinica, 2011, 21(4): 1541-1555.

[11] GUO B, ZHOU Q, ZHANG R C. Some results on constructing general minimum lower-order confounding 2^{n-m} designs for $n\leq 2^{n-m-2}$[J]. Metrika, 2014, 77(6): 721-732.

[12] ZHOU Q, GUO B. A note on constructing clear compromise plans[J]. Statistics and Probability Letters, 2017, 130: 17-24.

[13] ZHANG R C, MUKERJEE R. Characterization of general minimum lower-order confounding criterion via complementary sets[J]. Statistica Sinica, 2009, 19(1): 363-375.

[14] KE W M, YAO R. Selection of non-regular fractional factorial designs when some two-factor interactions are important[J]. Journal of Modern Applied Statistical Methods, 2008, 7(1): 94-100.

[15] 胡锡健. 金融时间序列分析与建模[J]. 新疆大学学报(自然科学版), 2007, 24(2): 150-158.

[16] 李智, 李智明. 正规部分因析设计别名成分数型的一些性质[J]. 应用概率统计, 2021, 37(6): 585-597.

[17] ZHANG R C, PARK D K. Optimal blocking of two-level fractional factorial designs[J]. Journal of Statistical Planning and Inference, 2000, 91(1): 107-121.

责任编辑：赵新科