

2.2 2^k 因子设计及其部分实施（一）

王正明 易泰河

系统工程学院 军事建模与仿真系

2019 年 11 月 22 日

引言

2^k 因子设计:

- 包含 k 个二水平因子的试验, 全部处理有 2^k 个;
- 主要应用于:
 - 定性考察因子对响应的影响;
 - 筛选大量因子中有实质影响的因子的初级研究阶段.
- 优势: 试验次数可以控制在相对较少的范围内.
- 缺陷: 对于连续变化的定量因子, 不能发现非线性关系.

本节教学目标

- (1) 介绍 2^k 因子设计及其部分实施;
- (2) 掌握 $L_{2^k}(2^{2^k-1})$ 型正交表的构造与应用;
- (3) 理解正交表、别名、混杂、分辨度等基本概念.

2.2 2^k 因子设计及其部分实施

2.2.1 2^2 设计与正交表 $L_4(2^3)$

- (1) 固定效应模型
- (2) 主效应与交互效应
- (3) 正交表 $L_4(2^3)$

2.2.2 2^3 设计与正交表 $L_8(2^7)$

2.2.3 2^k 设计与正交表 $L_{2^k}(2^{2^k-1})$

2.2.4 2^k 因子试验的部分实施

- 两个二水平因子 A 和 B , 以 0、1 表示两个水平;
- 用 (1)、 a 、 b 、 ab 表示 4 个处理 (0, 0), (1, 0), (0, 1), (1, 1), 当代表因子的字母出现时, 该因子取水平 1, 否则取 0;
- 每个处理重复 m 次, y_{ijk} 表示处理 (i, j) 处第 k 次重复试验的响应值, 2^2 试验的固定效应模型为:

$$\begin{cases} y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, & \varepsilon_{ijk} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 0, 1, \quad j = 0, 1, \quad k = 1, \dots, m, \\ \tau_0 + \tau_1 = 0, \quad \beta_0 + \beta_1 = 0, \\ (\tau\beta)_{00} + (\tau\beta)_{01} = (\tau\beta)_{10} + (\tau\beta)_{11} = 0, \\ (\tau\beta)_{00} + (\tau\beta)_{10} = (\tau\beta)_{01} + (\tau\beta)_{11} = 0. \end{cases}$$

- 两个二水平因子 A 和 B , 以 0、1 表示两个水平;
- 用 (1)、 a 、 b 、 ab 表示 4 个处理 (0, 0), (1, 0), (0, 1), (1, 1), 当代表因子的字母出现时, 该因子取水平 1, 否则取 0;
- 每个处理重复 m 次, y_{ijk} 表示处理 (i, j) 处第 k 次重复试验的响应值, 2^2 试验的固定效应模型为:

$$\begin{cases} y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, & \varepsilon_{ijk} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 0, 1, \quad j = 0, 1, \quad k = 1, \dots, m, \\ \tau_0 + \tau_1 = 0, \quad \beta_0 + \beta_1 = 0, \\ (\tau\beta)_{00} + (\tau\beta)_{01} = (\tau\beta)_{10} + (\tau\beta)_{11} = 0, \\ (\tau\beta)_{00} + (\tau\beta)_{10} = (\tau\beta)_{01} + (\tau\beta)_{11} = 0. \end{cases}$$

Definition

称 $\tau := \tau_1 - \tau_0$ 为因子 A 的主效应, 表示 A 的水平变动对响应的影响; 称 $\beta := \beta_1 - \beta_0$ 为因子 B 的主效应, 表示 B 的水平变动对响应的影响; 称

$$\tau\beta := \frac{1}{2} [(\tau\beta)_{11} - (\tau\beta)_{01} - (\tau\beta)_{10} + (\tau\beta)_{00}]$$

为因子 A 与因子 B 的交互效应, 表示因子 B 固定在 1 水平时 A 的水平变动对响应的影响与因子 B 固定在 0 水平时 A 的变动对响应的影响之差的一半.

- 参数变换: $\{\tau_i, \beta_j, (\tau\beta)_{ij} : i, j = 0, 1\} \iff (\tau, \beta, \tau\beta).$

- 以符号 (1)、 a 、 b 、 ab 表示各处理上 m 个响应值的总和, 则各效应的估计为

$$\begin{aligned} A &:= \hat{\tau}_1 - \hat{\tau}_0 = \frac{y_{1..}}{2m} - \frac{y_{0..}}{2m} = \frac{1}{2m} [-(1) - b + a + ab], \\ B &:= \hat{\beta}_1 - \hat{\beta}_0 = \frac{y_{.1.}}{2m} - \frac{y_{.0.}}{2m} = \frac{1}{2m} [-(1) + b - a + ab], \\ AB &:= \frac{1}{2m} [+ (1) - b - a + ab] \end{aligned}$$

- 它们是互相正交的三个对照, 它们的平方和分别为

$$\begin{cases} SS_A = \frac{1}{4m} [-(1) - b + a + ab]^2; \\ SS_B = \frac{1}{4m} [-(1) + b - a + ab]^2; \\ SS_{AB} = \frac{1}{4m} [+ (1) - b - a + ab]^2. \end{cases}$$

- 将计算诸效应估计量的对照的系数符号列成表:

	A	B	AB
(1)	—	—	+
b	—	+	—
a	+	—	—
ab	+	+	+

- 第一列为**二分列**, 称第二列为**四分列**. 此外:
 - (1) 每列 “+” 号与 “—” 号的出现的次数相等;
 - (2) 任何两列组成四组不同的符号对, 其出现的次数相等.

Definition

称由一些符号组成的矩阵为**正交表**(orthogonal table), 如果任意两列中同行符号构成的若干符号的重复次数相等.

性质:

- 任意一列中不同符号出现的次数相等;
- 从一张正交表中挑选出部分列组成的子表依然是正交表.

- $L_4(2^3)$: L 表示正交表, 2 代表正交表中不同水平数, 4 代表表的行数, 3 代表表的列数.

	A	B	AB
(1)	—	—	+
b	—	+	—
a	+	—	—
ab	+	+	+

- 任意两列对应符号相乘得出另一列, 且任意一列均可由其余两列对应符号相乘得到.
- $L_4(2^3)$ 的任何两列的交互效应列是另一列, 表 $L_4(2^3)$ 是完备正交表, 在不改变行数的情况下不能增加列.

2.2 2^k 因子设计及其部分实施

2.2.1 2^2 设计与正交表 $L_4(2^3)$

2.2.2 2^3 设计与正交表 $L_8(2^7)$

- (1) 固定效应模型
- (2) 正交表 $L_8(2^7)$
- (3) 正交表的等价
- (4) 正交表 $L_8(2^7)$ 的构造
- (5) 正交表 $L_8(2^7)$ 的应用

- 设 3 个因子为 A 、 B 、 C , 以 “1” 和 “0” 分别表示因子的两个水平.
- 全部处理组合共 8 个:

$(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1),$

$(1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1).$

- 每个处理重复 m 次, 以 y_{ijkl} 表示处理 (i, j, k) 的第 l 次重复试验结果, 固定效应模型?

$$\left\{ \begin{array}{l} y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \varepsilon_{ijkl}; \\ \varepsilon_{ijkl} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2), \quad i = 0, 1, \quad j = 0, 1, \quad k = 0, 1, \quad l = 1, \dots, m; \\ \tau_0 + \tau_1 = \beta_0 + \beta_1 = \gamma_0 + \gamma_1 = 0; \\ (\tau\beta)_{00} + (\tau\beta)_{10} = (\tau\beta)_{01} + (\tau\beta)_{11} = (\tau\beta)_{00} + (\tau\beta)_{01} = (\tau\beta)_{10} + (\tau\beta)_{11} = 0; \\ (\tau\gamma)_{00} + (\tau\gamma)_{10} = (\tau\gamma)_{01} + (\tau\gamma)_{11} = (\tau\gamma)_{00} + (\tau\gamma)_{01} = (\tau\gamma)_{10} + (\tau\gamma)_{11} = 0; \\ (\beta\gamma)_{00} + (\beta\gamma)_{10} = (\beta\gamma)_{01} + (\beta\gamma)_{11} = (\beta\gamma)_{00} + (\beta\gamma)_{01} = (\beta\gamma)_{10} + (\beta\gamma)_{11} = 0; \\ (\tau\beta\gamma)_{0jk} + (\tau\beta\gamma)_{1jk} = 0, \quad j = 0, 1, \quad k = 0, 1; \\ (\tau\beta\gamma)_{i0k} + (\tau\beta\gamma)_{i1k} = 0, \quad i = 0, 1, \quad k = 0, 1; \\ (\tau\beta\gamma)_{ij0} + (\tau\beta\gamma)_{ij1} = 0, \quad i = 0, 1, \quad j = 0, 1. \end{array} \right.$$

- 称 $\tau := \tau_1 - \tau_0$ 为因子 A 的效应, 其估计量记作 A ; 称 $\beta := \beta_1 - \beta_0$ 为因子 B 的效应, 其估计量记作 B ; 称 $\gamma := \gamma_1 - \gamma_0$ 为因子 C 的效应, 其估计量记作 C .
- 称 $\tau\beta := \frac{1}{2}[(\tau\beta)_{11} - (\tau\beta)_{10} - (\tau\beta)_{01} + (\tau\beta)_{00}]$ 为因子 A 与 B 的交互效应, 其估计量记作 AB ; 称 $\tau\gamma := \frac{1}{2}[(\tau\gamma)_{11} - (\tau\gamma)_{10} - (\tau\gamma)_{01} + (\tau\gamma)_{00}]$ 为因子 A 与 C 的交互效应, 其估计量记作 AC ; 称 $\beta\gamma := \frac{1}{2}[(\beta\gamma)_{11} - (\beta\gamma)_{10} - (\beta\gamma)_{01} + (\beta\gamma)_{00}]$ 为因子 B 与 C 的交互效应, 其估计量记作 BC ;

● 称

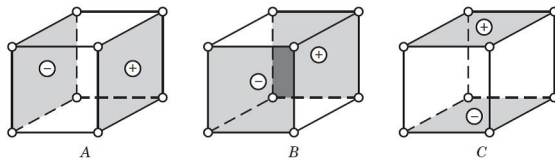
$$\frac{1}{2} \left\{ \frac{1}{2} [(\tau\beta\gamma)_{111} - (\tau\beta\gamma)_{011} - (\tau\beta\gamma)_{101} + (\tau\beta\gamma)_{001}] \right. \\ \left. - \frac{1}{2} [(\tau\beta\gamma)_{110} - (\tau\beta\gamma)_{010} - (\tau\beta\gamma)_{100} + (\tau\beta\gamma)_{000}] \right\}$$

为 3 个因子 A 、 B 、 C 的交互效应, 记作 $\tau\beta\gamma$, 表示因子 C 固定在水平 1 时 A 、 B 的交互效应与因子 C 固定在水平 0 时的 A 、 B 的交互效应之差的一半, 它的估计量记作 ABC .

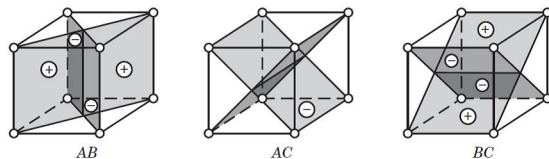
- 各试验点上 m 个观察值的总和 $y_{000\cdot}$ 、 $y_{100\cdot}$ 、 $y_{010\cdot}$ 、 $y_{110\cdot}$ 、 $y_{001\cdot}$ 、 $y_{101\cdot}$ 、 $y_{011\cdot}$ 、 $y_{111\cdot}$ 分别用 (1)、 a 、 b 、 ab 、 c 、 ac 、 bc 、 abc 表示, 则诸效应的无偏估计为

$$\left\{ \begin{array}{l} A = \frac{1}{4m} [- (1) - c - b - bc + a + ac + ab + abc]; \\ B = \frac{1}{4m} [- (1) - c + b + bc - a - ac + ab + abc]; \\ C = \frac{1}{4m} [- (1) + c - b + bc - a + ac - ab + abc]; \\ AB = \frac{1}{4m} [+ (1) + c - b - bc - a - ac + ab + abc]; \\ AC = \frac{1}{4m} [+ (1) - c + b - bc - a + ac - ab + abc]; \\ BC = \frac{1}{4m} [+ (1) - c - b + bc + a - ac - ab + abc]; \\ ABC = \frac{1}{4m} [- (1) + c + b - bc + a - ac - ab + abc]. \end{array} \right.$$

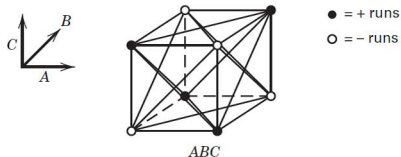
- 它们是七个互相正交的对照!



(a) Main effects



(b) Two-factor interaction



(c) Three-factor interaction

- 诸效应的对照系数的符号列称正交表 $L_8(2^7)$:

处理	A	B	AB	C	AC	BC	ABC
(1)	—	—	+	—	+	+	—
c	—	—	+	+	—	—	+
b	—	+	—	—	+	—	+
bc	—	+	—	+	—	+	—
a	+	—	—	—	—	+	+
ac	+	—	—	+	+	—	—
ab	+	+	+	—	—	—	—
abc	+	+	+	+	+	+	+

- A 列为**二分列**, B 列为**四分列**, C 列为**八分列**, 这三列即可用来安排试验, 又可用于分析试验结果;

- 正交性: 任何两列符号乘积之和为 0;
- 任意两列相乘, 得出表中的一列, 如 $AB \times C = ABC$ 等;
- 可由其它两列运算得到的列称为那两列的交互效应列, $L_8(2^7)$ 中的交互效应关系:

1	2	3	4	5	6	7	列 号
	3	2	5	4	7	6	1
		1	6	7	4	5	2
			7	6	5	4	3
				1	2	3	4
					3	2	5
						1	6

从试验设计的角度来看:

- **行置换不变性**: 任意两行之间可以互相置换, 即试验的次序可以自由选择;
- **列置换不变性**: 任意两列之间可以互相置换, 即试验因子可以自由地安排在各列上;
- **水平置换不变性**: 每一列的水平可以互相置换, 即因子的水平可以自由安排.

正交表不唯一!

Definition

称两张正交表**等价**, 如果对其中一张表进行适当的行置换和列置换可以得到另一张表; 称两张正交表**同构**, 如果对其中一张表进行适当的行置换、列置换和水平置换可以得到另一张表.

- 以水平记号“1”和“0”代替代替符号“+”和“-”，得到

No.	A	B	AB	C	AC	BC	ABC
1	0	0	1	0	1	1	0
2	0	0	1	1	0	0	1
3	0	1	0	0	1	0	1
4	0	1	0	1	0	1	0
5	1	0	0	0	0	1	1
6	1	0	0	1	1	0	0
7	1	1	1	0	0	0	0
8	1	1	1	1	1	1	1

- 把符号“+”和“-”互换, 然后分别以“1”和“0”替换“+”和“-”, 得到

No.	A	B	AB	C	AC	BC	ABC
1	0	0	0	0	0	0	0
2	0	0	0	1	1	1	1
3	0	1	1	0	0	1	1
4	0	1	1	1	1	0	0
5	1	0	1	0	1	0	1
6	1	0	1	1	0	1	0
7	1	1	0	0	1	1	0
8	1	1	0	1	0	0	1

- AB 列由 A 列和 B 列按照模 2 加法生成, 表示交互效应列.

- 利用诸效应的系数构造:

- 行按字母顺序排: A, B, AB, C, AC, BC, ABC ;
- 列按字母反序排: $(1), c, b, bc, a, ac, ab, abc$.

- 利用列名运算:

Step 1 构造二分列 A , 它的前四个元素为 0, 后四个元素为 1;

Step 2 构造四分列 B , 并利用对应元素的模 2 加法运算构造 B 与它前面的列 A 的交互效应列 AB ;

Step 3 构造八分列 C , 并利用对应元素的模 2 加法运算依此构造 C 与它前面的 A 列、 B 列和 AB 列的交互效应列 AC 列、 BC 列和 ABC 列.

- 利用诸效应的系数构造:

- 行按字母顺序排: A, B, AB, C, AC, BC, ABC ;
- 列按字母反序排: $(1), c, b, bc, a, ac, ab, abc$.

- 利用列名运算:

Step 1 构造二分列 A , 它的前四个元素为 0, 后四个元素为 1;

Step 2 构造四分列 B , 并利用对应元素的模 2 加法运算构造 B 与它前面的列 A 的交互效应列 AB ;

Step 3 构造八分列 C , 并利用对应元素的模 2 加法运算依此构造 C 与它前面的 A 列、 B 列和 AB 列的交互效应列 AC 列、 BC 列和 ABC 列.

小测试

- 利用“+”和“-”表示二水平因子的两个水平, 构造正交表 $L_8(2^7)$;
- 利用“0”和“1”表示二水平因子的两个水平, 构造正交表 $L_8(2^7)$.

Example

在梳棉机上纺粘锦混纺纱, 为了提高质量, 选了 3 个因子, 每个因子 2 个水平,

A 金属针布: $A_1 =$ 日本产, $A_2 =$ 青岛产;

B 产量水平: $B_1 = 6$ 公斤, $B_2 = 10$ 公斤;

C 锡林速度: $C_1 = 238$ 转/分, $C_2 = 320$ 转/分.

实践经验表明, 因子间可能有二因子交互效应.

Example (Cont.)

- 用正交表 $L_8(2^7)$ 来安排试验:

列号	1	2	3	4	5	6	7
因子	A	B	AB	C	AC	BC	ABC

试验号	$A(1)$	$B(2)$	$C(4)$
1	日本产	6	238
2	日本产	6	320
3	日本产	10	238
4	日本产	10	320
5	青岛产	6	238
6	青岛产	6	320
7	青岛产	10	238
8	青岛产	10	320

- 为什么因子 C 不能安排在第三列?

试验号	A(1)	B(2)	AB(3)	C(4)	AC(5)	BC(6)	ABC(7)	棉结粒数
1	0	0	0	0	0	0	0	(1) = 0.30
2	0	0	0	1	1	1	1	$c = 0.35$
3	0	1	1	0	0	1	1	$b = 0.20$
4	0	1	1	1	1	0	0	$bc = 0.30$
5	1	0	1	0	1	0	1	$a = 0.15$
6	1	0	1	1	0	1	0	$ac = 0.50$
7	1	1	0	0	1	1	0	$ab = 0.15$
8	1	1	0	1	0	0	1	$abc = 0.40$
T_0	1.15	1.30	1.20	0.80	1.40	1.15	1.25	$T = 2.35$
T_1	1.20	1.05	1.15	1.55	0.95	1.20	1.10	
m_0	0.2875	0.3250	0.3000	0.2000	0.3500	0.2875	0.3125	
m_1	0.3000	0.2625	0.2875	0.3875	0.2375	0.3000	0.2750	
R	0.0125	0.0625	0.0125	0.1875	0.1125	0.0125	0.0375	

● $R := \max\{m_0, m_1\} - \min\{m_0, m_1\}$ 为**极差**, 可用来衡量 3 个诸效应的主次关系, 极差越大表明相应的效应越大.

● 2^k 因子设计中, 极差就是效应的估计的绝对值!

两种方式计算各列的偏差平方和,

- 采用计算对照偏差平方和的公式:

$$\begin{aligned}SS_A &= \frac{1}{8} [abc + ab + ac + a - bc - b - c - (1)]^2 \\&= \frac{1}{8} [T_{A_1} - T_{A_0}]^2 = 0.0003125;\end{aligned}$$

- 正交表中常用的偏差平方和计算公式:

$$\begin{aligned}SS_A &= \frac{T_{A_0}^2}{4} + \frac{T_{A_1}^2}{4} - \frac{T^2}{8} = \frac{1.15^2 + 1.20^2}{4} - \frac{2.35^2}{8} \\&= 0.0003125.\end{aligned}$$

方差来源	平方和	自由度	均方	F 值	p 值
B	0.0078125	1	0.0078125	8.33	0.0447
C	0.0203125	1	0.0203125	75.00	0.0010
AC	0.0253125	1	0.0253125	27.00	0.0065
误差	0.0037500	4	0.0009375		
A	0.0003125	1			
AB	0.0003125	1			
BC	0.0003125	1			
ABC	0.0028125	1			
总和	0.1071825	7			

总结:

- 1 2^2 设计与正交表 $L_4(2^3)$;
- 2 2^3 设计与正交表 $L_8(2^7)$.

注意:

- 1 各种教材上符号有所不同!

习题:

- 1 写出 2^2 因子设计和 2^3 因子设计的方差分析表.

总结:

- 1 2^2 设计与正交表 $L_4(2^3)$;
- 2 2^3 设计与正交表 $L_8(2^7)$.

注意:

- 1 各种教材上符号有所不同!

习题:

- 1 写出 2^2 因子设计和 2^3 因子设计的方差分析表.

总结:

- 1 2^2 设计与正交表 $L_4(2^3)$;
- 2 2^3 设计与正交表 $L_8(2^7)$.

注意:

- 1 各种教材上符号有所不同!

习题:

- 1 写出 2^2 因子设计和 2^3 因子设计的方差分析表.