

2.4 正交设计的一般讨论

王正明 易泰河

系统工程学院 军事建模与仿真系

2019 年 11 月 13 日

知识回顾

- 2^k 设计与 $L_{2^k}(2^{\frac{2^k-1}{2-1}})$ 型正交表;
- 3^k 设计与 $L_{3^k}(3^{\frac{3^k-1}{3-1}})$ 型正交表;
- q^k 设计与 $L_{q^k}(q^{\frac{q^k-1}{q-1}})$ 型正交表;
- 部分实施: 混杂与别名, 定义关系, 字长, 分辨度.

2.4 正交设计的一般讨论

2.4.1 正交表的一般性质

- (1) 基本概念
- (2) 利用正交表安排试验
- (3) 表头设计的基本原则

2.4.2 等水平试验的正交设计

2.4.3 不等水平试验的正交设计

(1) 基本概念

- 交表 $L_n(q_1^{m_1} \times \cdots \times q_r^{m_r})$ 是一个 $n \times m$ 的矩阵, 其中 $m = m_1 + \cdots + m_r$, $q_i \geq 2$ 为正数, m_i 个列有 q_i 个不同的符号, 使得**正任意两列组成的符号对出现的次数相同**.

L : 表示正交表;

n : 正交表的行数, 也表示试验总数;

q_i : 列中不同符号的个数, 对应该列可安排因子的水平数;

m_i : 包含 q_i 个不同符号的列数, 表示最多能容纳的 q_i 水平因子的个数;

r : 表中不同水平数的数目.

(1) 基本概念

- 称所有的因子水平数相同的正交表为**对称正交表**或**等水平正交表**, 记作 $L_n(q^m)$,

- 完备正交表**

$$L_{q^k}\left(q^{\frac{q^k-1}{q-1}}\right), \quad k = 2, 3, \dots,$$

完备正交表在行数不变的情况下不能增加列.

- 当正交表 $L_n(q_1^{m_1} \times \dots \times q_r^{m_r})$ 中 $r > 1$ 时, 称为**非对称正交表**或**混合水平正交表**.
- 正交表查阅网址:

- <http://www.research.att.com/~njas/oaddir/>;
- <http://support.sas.com/techsup/technote/>

- 可通过**对称正交表的并列**得到混合正交表.
- 把正交表 $L_8(2^7)$ 的 A 列和 B 两列构成的四对水平组合替换成 4 个水平: $(0, 0) \rightarrow 0$, $(0, 1) \rightarrow 1$, $(1, 0) \rightarrow 2$, $(1, 1) \rightarrow 3$, 并删去它们的交互作用列 AB , 得到 $L_8(4 \times 2^4)$.

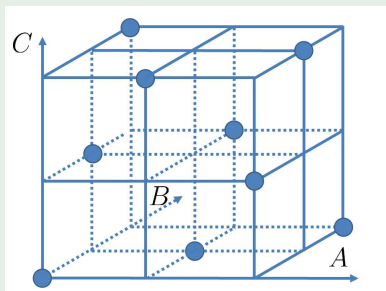
No.	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	1	1
3	1	0	0	1	1
4	1	1	1	0	0
5	2	0	1	0	1
6	2	1	0	1	0
7	3	0	1	1	0
8	3	1	0	0	1

(2) 利用正交表安排试验

- 利用正交表 $L_n(q_1^{m_1} \times \cdots \times q_r^{m_r})$ 设计试验时, 最多可安排 m_1 个 q_1 水平因子、 m_2 个 q_2 水平因子、 \cdots , 总共做 n 次试验.
- 正交表安排的试验方案具有**均衡分散**和**整齐可比**的特点.
 - 均衡分散: 挑选出来的处理在全部处理中的分布比较均匀
 - 整齐可比: 每一个因子的各水平间具有可比性

Example

利用正交表 $L_9(3^4)$ 的前三列安排 3^3 试验:



- 均衡分散: 任一平面内都包含 3 个试验点, 任一直线上都包含 1 个试验点.
- 整齐可比: 当比较 A 因子不同水平时, B 因子不同水平的效应相互抵消, C 因子不同水平的效应也相互抵消.

(3) 表头设计的基本原则

- **自由度原则:**

- (a) 因子和交互作用的自由度是指其偏差平方和作为统计量所服从的 χ^2 分布的自由度. 因子的自由度为该因子水平数减一, 交互作用的自由度为该交互作用中个因子的自由度的乘积;
- (b) 正交表的自由度为试验次数减一, 表中各列的自由度为该列的水平数减一;
- (c) 因子的自由度应等于所在列的自由度;
- (d) 交互作用的自由度应等于所在列的自由度或其之和;
- (e) 所有因子与交互作用的自由度之和不能超过所选正交表的自由度.

(3) 表头设计的基本原则

- 避免混杂原则:

- (a) 如果一列上出现的因子和交互作用不止一个, 当该列的效应显著时, 无法识别是哪个因子 (或交互作用) 显著, 称这种现象为**混杂现象**(confounded, aliased).
- (b) 表头设计时应尽量避免出现混杂现象.

Example

为了提高某化工产品的转化率, 试验者选择了影响转化率的 4 个主要因子

A: 催化剂种类, $A_1 = 1$, $A_2 = 2$;

B: 反应时间, $B_1 = 1.5\text{h}$, $B_2 = 2.5\text{h}$;

C: 反应温度, $C_1 = 80^\circ\text{C}$, $C_2 = 90^\circ\text{C}$;

D: 加减量, $D_1 = 5\%$, $D_2 = 7\%$.

这是一个 2^4 设计, 如果采用 $L_{16}(2^{15})$ 设计, 虽然可以估计出所有的效应, 但试验次数至少需要 16 次. 由于经费所限, 需要试验次数更少的试验方案.

如果经验告诉我们, 没有三因子和四因子交互作用, D 和其它三个因子没有交互作用, 而 A, B, C 之间可能有交互作用. 也就是说, 总共需要考虑 A, B, C, D 四个主效应和 AB, AC, BC 三个两因子交互效应, 共 7 个自由度. 我们知道正交表 $L_8(2^7)$ 的自由度恰为 7, 能否利用正交表 $L_8(2^7)$ 来规划这个试验呢?

Example (Cont.)

回顾 $L_8(2^7)$ 的交互作用表, 建立如下的列号与各因子及交互作用的对应关系:

列号	1	2	3	4	5	6	7	定义关系	分辨度
方案一	A	B	AB	C	AC	BC	D	$I = ABCD$	IV
			\updownarrow		\updownarrow	\updownarrow	\updownarrow		
			CD		BD	AD	ABC		

从表中可以看出, 交互效应 AB 与 CD 混杂、 AC 与 BD 混杂、 BC 与 AD 混杂、 D 与 ABC 混杂, 但经验告诉我们 CD 、 BD 、 AD 以及 ABC 不显著, 因而实际上并不存在混杂效应, 该方案是可行的。

(3) 表头设计的基本原则

- 混杂技术

- (a) 如果先验信息和试验资源都不足, 则混杂不可避免!
- (b) 基本原则: 忽略高阶交互效应、保证主效应和低价交互效应
 - 效应稀疏原则: 多因子试验中重要效应的个数不会太多;
 - 效应有序原则: 主效应比交互效应重要, 低阶交互效应比高阶交互效应重要, 同阶交互效应的重要性相同.
- (c) 首先保证估计主效应, 其次保证估计低阶交互效应, 让混杂发生在次要的交互作用之间, 这种办法称为**混杂技术**.

Example

继续讨论前一例中的 2^4 设计问题, 下面是另一种 2^{4-1} 方案:

列号	1	2	3	4	5	6	7	定义关系	分辨度
方案二	A	B	C	D	AD	BD	CD	$I = ABC$	III
	\updownarrow	\updownarrow	\updownarrow						
	BC	AC	AB						

若二阶交互作用可能存在, 方案一可以估计 4 个因子的主效应, 但所有的二阶交互效应互相混杂. 方案只能估计 D 的主效应和三个二阶交互效应. 按照效应系数原则和效应有序原则, 我们应该选择方案一, 优先保证主效应的估计.

小结

- ① 正交表的概念
- ② 正交表的并列
- ③ 表头设计的原则
- ④ 混杂技术

2.4 正交设计的一般讨论

2.4.1 正交表的一般性质

2.4.2 等水平试验的正交设计

- (1) 无交互作用情形
- (2) 有交互作用情形

2.4.3 不等水平试验的正交设计

Example (无交互作用情形)

为了提高某化工产品的转化率, 选择 3 个试验因子: 反应温度 A , $A_1 = 80^\circ\text{C}$, $A_2 = 85^\circ\text{C}$, $A_3 = 90^\circ\text{C}$; 反应时间 B , $B_1 = 90\text{min}$, $B_2 = 120\text{min}$, $B_3 = 150\text{min}$; 用碱量 C , $C_1 = 5\%$, $C_2 = 6\%$, $C_3 = 7\%$. 假设根据过去的经验, 所有交互效应都不显著.

- 共有 3 个三水平因子, 总的自由度等于 3 个三水平因子自由度的和, 即 _____;
- 根据自由度原则, 需要选择一张行数不小于 _____ 的正交表来安排试验.
- 不需考虑交互作用, 所选正交表的列数不小于 _____ 即可.
- 选什么表?

试验数据如下:

试验号	<i>A</i>	<i>B</i>	<i>C</i>	转化率 (<i>CP</i>)
1	0(80℃)	0(90min)	0(5%)	31%
2	0(80℃)	1(120min)	1(6%)	54%
3	0(80℃)	2(150min)	2(7%)	38%
4	1(85℃)	0(90min)	1(6%)	53%
5	1(85℃)	1(120min)	2(7%)	49%
6	1(85℃)	2(150min)	0(5%)	42%
7	2(90℃)	0(90min)	2(7%)	57%
8	2(90℃)	1(120min)	0(5%)	62%
9	2(90℃)	2(150min)	1(6%)	64%
T_1	123%	141%	135%	
T_2	144%	165%	171%	
T_3	183%	144%	144%	
m_1	41%	47%	45%	
m_2	48%	55%	57%	
m_3	61%	48%	48%	
R	20%	8%	12%	

(1) 验结果的直观分析

Step 1 计算诸因子在每个水平下的平均转化率和极差. 以第一列为例, “ T_1 ”行给出在反应温度 80°C 下三次试验转化率之和,

$$T_1 = 32 + 54 + 38 = 123,$$

其均值

$$m_1 = T_1/3 = 123/3 = 41,$$

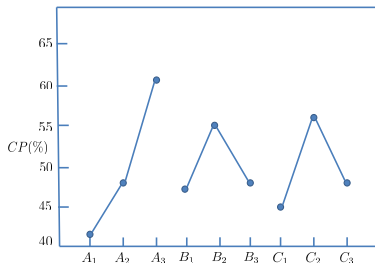
类似地, 在反应温度 85°C 和 90°C 下三次试验的平均转化率为 48 和 61, 极差

$$R = \max\{41, 48, 61\} - \min\{41, 46, 61\} = 20$$

列在表的最后一行.

(1) 验结果的直观分析

Step 2 画平均转化率图:



- 温度越高, 转化率越高, 以 90℃ 为最好;
- 反应时间以 120min 转化率最高;
- 用碱量以 6% 转化率最高.

综合起来以处理 (A_3, B_2, C_2) 最好.

(1) 验结果的直观分析

Step 3 将因子对响应的影响排序.

- 利用平均转化率图, 点散布范围越大的因子影响越大, 看出主次关系是 $A > C > B$.
- 利用极差 R : 极差越大的因子影响越大, A 、 B 、 C 三个因子的极差分布为 20、8、12, 故 $A > C > B$.

Step 4 追加试验. 9 次试验中不包含推断最佳水平组合为 (A_3, B_2, C_2) , 需要追加试验.

利用直观分析可以获得最佳或满意的处理, 还可以区分因子对响应影响的主次.

(2) 固定效应模型与参数估计

- 由于没有交互作用, 固定效应模型为:

$$\left\{ \begin{array}{l} y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \\ \varepsilon_{ijk} \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ \tau_1 + \tau_2 + \tau_3 = 0, \\ \beta_1 + \beta_2 + \beta_3 = 0, \\ \gamma_1 + \gamma_2 + \gamma_3 = 0. \end{array} \right.$$

τ_1, τ_2, τ_3 为因子 A 的 3 个水平的主效应; $\beta_1, \beta_2, \beta_3$ 为因子 B 的 3 个主效应; $\gamma_1, \gamma_2, \gamma_3$ 为因子 C 的 3 个主效应.

(2) 固定效应模型与参数估计

- 9 次试验数据可用矩阵表示为

$$\mathbf{y} = \begin{pmatrix} y_{111} \\ y_{122} \\ y_{133} \\ y_{212} \\ y_{223} \\ y_{212} \\ y_{313} \\ y_{321} \\ y_{332} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 & -1 & -1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & -1 & -1 \\ 1 & 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & -1 & -1 & 1 & 0 & -1 & -1 \\ 1 & -1 & -1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -1 & -1 & -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \gamma_1 \\ \gamma_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{111} \\ \varepsilon_{122} \\ \varepsilon_{133} \\ \varepsilon_{212} \\ \varepsilon_{223} \\ \varepsilon_{212} \\ \varepsilon_{313} \\ \varepsilon_{321} \\ \varepsilon_{332} \end{pmatrix}$$

- 由最小二乘法可得 $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (50, -9, -2, -3, 5, -5, 7)^T$;
- 利用这些估计值来预测最佳水平组合 (A_3, B_2, C_2) 的响应值:

$$\hat{y}_{322} = \hat{\mu} + \hat{\tau}_3 + \hat{\beta}_2 + \hat{\gamma}_2 = 50 + 11 + 5 + 7 = 73,$$

(2) 固定效应模型与参数估计

- 最优回归设计以未知参数的估计的精度来衡量试验设计的有效性.
- 由回归分析理论可知

$$\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 \begin{pmatrix} \frac{1}{9} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{9} & -\frac{1}{9} & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{9} & \frac{2}{9} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{9} & -\frac{1}{9} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{9} & \frac{2}{9} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{9} & -\frac{1}{9} \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{9} & \frac{2}{9} \end{pmatrix}$$

(2) 固定效应模型与参数估计

- $\text{Var}(\hat{\mu}) = \text{Var}(\hat{\tau}_i) = \text{Var}(\hat{\beta}_i) = \text{Var}(\hat{\gamma}_i) = \frac{2}{9}\sigma^2,$
 $i = 1, 2;$
- $\text{Cov}(\hat{\tau}_1, \hat{\tau}_2) = \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{Cov}(\hat{\gamma}_1, \hat{\gamma}_2) = -\frac{1}{9}\sigma^2;$
- $\text{Var}(\hat{\tau}_3) = \text{Var}(\hat{\tau}_1 + \hat{\tau}_2) =$
 $\text{Var}(\hat{\tau}_1) + \text{Var}(\hat{\tau}_2) + \text{Cov}(\hat{\tau}_1, \hat{\tau}_2) = \frac{2}{9}\sigma^2;$
- 类似地, $\text{Var}(\hat{\beta}_3) = \frac{2}{9}\sigma^2, \text{Var}(\hat{\gamma}_3) = \frac{2}{9}\sigma^2;$
- $\hat{\sigma}^2 = \mathbf{y}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{y} / (9 - 7) = 18/2 = 9.$

(2) 固定效应模型与参数估计

- 在 A_1 的平均转化率为 $\mu_{A_1} = \mu + \tau_1$, 它由 “ m_1 ” 行的值 $\hat{\mu}_{A_1} = \hat{\mu} + \hat{\tau}_1$ 估计,

$$\text{Var}(\hat{\mu}_{A_1}) = \text{Var}(\hat{\mu} + \hat{\tau}_1) = \frac{1}{3}\sigma^2,$$

- 类似地,

$$\text{Var}(\hat{\mu}_{A_1}) = \text{Var}(\hat{\mu}_{A_3}) = \cdots = \text{Var}(\hat{\mu}_{C_3}) = \frac{1}{3}\sigma^2.$$

- 虽然试验方案中每个处理只做了一次试验, 但参数估计的效果相当于每个处理都重复了 3 次!

(3) 方差分析

- 考虑反应温度 A 对转化率是否有显著的影响, 就是检验:

$$H_0^A : \tau_1 = \tau_2 = \tau_3 = 0, \quad H_1^A : \tau_1, \tau_2, \tau_3 \text{ 不全为零.}$$

- 利用方差分析和 F 验, 不难获得有关的 F 统计量, 由于正交表的特殊结构, 使得有关计算平方和公式变得十分容易和简单, 这里利用定义来计算偏差平方和.

(3) 方差分析

- 总平方和

$$\bar{y} = \frac{1}{9} \sum_{i=1}^9 y_i = 50, \quad SS_T = \sum_{i=1}^9 (y_i - \bar{y})^2 = 984.$$

- 因子 A 的平方和是它们的 3 个均值, m_1^A , m_2^A , m_3^A 的离差平方和乘以 3, 因为每个均值是由 3 次试验的结果平均而得, 即

$$\begin{aligned} SS_A &= 3[(m_1^A - \bar{y})^2 + (m_2^A - \bar{y})^2 + (m_3^A - \bar{y})^2] \\ &= 3[(41 - 50)^2 + (48 - 50)^2 + (61 - 50)^2] = 618. \end{aligned}$$

类似的, 由 B 和 C 的均值可算得 $SS_B = 114$, $SS_C = 234$.

- 误差平方和 SS_E 可通过平方和分解公式获得

$$SS_E = SS_T - SS_A - SS_B - SS_C = 984 - 618 - 114 - 234 = 18.$$

(3) 方差分析

方差来源	自由度	平方和	均方	F	p 值
A	2	618	309	34.33	0.0283
B	2	114	57	6.33	0.1364
C	2	234	117	13.00	0.0714
误差	2	18	9		
总和	8	984			

- 只有因子 A 在水平 $\alpha = 0.05$ 下显著. 因子 B 和因子 C 在 $\alpha = 0.05$ 下均不显著.

Example (有交互作用情形)

为了提高某化工产品的转化率, 试验者选择了影响转化率的 4 个主要因子

A: 催化剂种类, $A_1 = 1$, $A_2 = 2$;

B: 反应时间, $B_1 = 1.5\text{h}$, $B_2 = 2.5\text{h}$;

C: 反应温度, $C_1 = 80^\circ\text{C}$, $C_2 = 90^\circ\text{C}$;

D: 加减量, $D_1 = 5\%$, $D_2 = 7\%$.

是希望提高转化率, 其值越大越好. 假设根据经验, 认为可能存在交互作用 AB 和 AC .

- 如果全部交互效应都显著, 则需要采用正交表 $L_{16}(2^{15})$. 由于经验表明只存在两个二因子交互效应, 采用 2^{4-1} 实施即可. 如何设计表头?

列名	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>	定义关系	分辨度
方案一	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>D</i>	$I = ABCD$	IV
			↕		↕	↕	↕		
			<i>CD</i>		<i>BD</i>	<i>AD</i>	<i>ABC</i>		

- 如果全部交互效应都显著, 则需要采用正交表 $L_{16}(2^{15})$. 由于经验表明只存在两个二因子交互效应, 采用 2^{4-1} 实施即可. 如何设计表头?

列名	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>	定义关系	分辨度
方案一	<i>A</i>	<i>B</i>	<i>AB</i>	<i>C</i>	<i>AC</i>	<i>BC</i>	<i>D</i>	$I = ABCD$	IV
			↕		↕	↕	↕		
			<i>CD</i>		<i>BD</i>	<i>AD</i>	<i>ABC</i>		

试验号	$A(1)$	$B(2)$	$AB(3)$	$C(4)$	$AC(5)$	$D(7)$	转化率/%
1	1	1(1.5h)	1	1(80℃)	1	1	82
2	1	1(1.5h)	1	2(90℃)	2	2	78
3	1	2(2.5h)	2	1(80℃)	1	2	76
4	1	2(2.5h)	2	2(90℃)	2	1	85
5	2	1(1.5h)	2	1(80℃)	2	2	83
6	2	1(1.5h)	2	2(90℃)	1	1	86
7	2	2(2.5h)	1	1(80℃)	2	1	92
8	2	2(2.5h)	1	2(90℃)	1	2	79
m_1	80.25	82.25	82.75	83.25	80.75	86.25	
m_2	85.00	83.00	82.50	82.00	84.50	79.00	
R	4.75	0.75	0.25	1.25	3.75	7.25	

- m_1 和 m_2 的值对 A, B, C, D 所在的四列反映了四个因子分别在两个水平下的均值, 而 AB 和 AC 所在的两列的 m_1 和 m_2 没有统计意义.

(1) 试验结果的直观分析

- 但对于 2^k 设计而言, 由它们计算的极差 R 是有统计意义的, 就是效应的估计, 因此仍可以用 R 的值来衡量四个因子以及其交互作用的主次关系.
- 主次关系如下: $D > A > AC > C > B > AB$.
- 第 7 号试验的转化率最高, 该试验条件为 (A_2, B_2, C_1, D_1) ;
- 从各因子的平均转化率大小来看, (A_2, B_2, C_1, D_1) 也是最好的试验条件.

(2) 试验结果的方差分析

方差来源	平方和	自由度	均方	F 值	p 值
A	45.125	1	45.125	40.11	0.0997
B	1.125	1	1.125	1.00	0.5000
C	3.125	1	3.125	2.78	0.3440
D	105.125	1	105.125	93.44	0.0656
AB	0.125	1	0.125	0.11	0.7952
AC	28.125	1	28.125	25.00	0.1257
误差	1.125	1	1.125		
总和	183.875	7			

- 可以断定交互作用 AB 是不显著的, 但因子 B 和 C 的显著性还需要进一步考察, 方法剔除 AB 后再重新作方差分析.

(2) 试验结果的方差分析

- 新的方差分析表明因子 B 是不显著的, 需要剔除. 剔除 B 后得到的方差分析表

方差来源	平方和	自由度	均方	F 值	p 值
A	45.125	1	45.125	57.00	0.0048
C	3.125	1	3.125	3.95	0.1411
D	105.125	1	105.125	132.79	0.0014
AC	28.125	1	28.125	35.35	0.0094
误差	2.375	3	0.792		
{ B AB BC	0.125	1			
	1.125	1			
	1.125	1			
总和	183.875	7			

(3) 确定最佳处理组合

- 由于因子 D 和 A 最显著, 从它们两个水平的平均响应值可知, D 因子应取水平 D_1 , A 因子取水平 A_2 .
- 由于 AC 也比较显著, 根据 AC 来确定 C 的水平. A 和 C 共有四种搭配 (A_1, C_1) , (A_1, C_2) , (A_2, C_1) , (A_2, C_2) , 试验结果总结如下:

	A_1		A_2	
C_1	82	76	83	92
C_2	78	85	86	79

- A 和 C 的最佳水平搭配为 (A_2, C_1) , 其中 A 因子的最优水平与单独考虑 A 因子时的结果一致, 而 C 的最优水平与从所有数据中单独看 C 的最优水平也是一致的, 于是得到最优条件 (A_2, C_1, D_1) .
- 反应时间 B 的水平对转化率没有显著影响, 但还是可以从数据中选择平均转化率稍大的反应时间 B_2 , 最终得到最好的试验条件为 (A_2, B_2, C_1, D_1) , 它与直观分析的结果一致, 且恰好是第 7 号试验.

2.4 正交设计的一般讨论

2.4.1 正交表的一般性质

2.4.2 等水平试验的正交设计

2.4.3 不等水平试验的正交设计

- 如果各因子的水平个数不全相等, 如何利用正交表制定试验方案?
- 利用混合水平的正交表: $L_8(4 \times 2^4)$, $L_{16}(4 \times 2^{12})$, $L_{16}(4^4 \times 2^3)$, \dots ;
- 利用拟水平法.

Example (利用混合水平正交表)

为了探索缝纫机胶压板的制造工艺, 选了如下的因子和水平:

因子水平	压力 (A)	温度 (B)	时间 (C)
1	8kg	95℃	9min
2	10kg	90℃	12min
3	11kg		
4	12kg		

不考虑交互作用的前提下可以直接套用正交表 $L_8(4 \times 2^4)$,

列号	1	2	3	4	5
因子	A	B	C		

该试验由四位有经验的专家打分, 最高 6 分, 最低 1 分.

试验号	$A(\text{kg})$	$B(^{\circ}\text{C})$	$C(\text{min})$	打分				总分
1	8(0)	95(0)	9(0)	6	6	6	4	22
2	8(0)	90(1)	12(1)	6	5	4	4	19
3	10(1)	95(0)	9(0)	4	3	2	2	11
4	10(1)	90(1)	12(1)	4	4	3	2	13
5	11(2)	95(0)	9(0)	2	1	1	1	5
6	11(2)	95(0)	9(0)	4	4	4	2	14
7	12(3)	95(0)	12(1)	4	3	2	1	10
8	12(3)	90(1)	9(0)	6	5	4	2	17
T_1	41	48	64					
T_2	24	63	47					
T_3	19							
T_4	27							
m_1	5.1	3.0	4.0					
m_2	3.0	3.9	2.9					
m_3	2.4							
m_4	3.4							
R	2.7	0.9	1.1					

(1) 试验结果的直观分析

- 当水平数不同时, 即便两个因子对指标有等同影响, 水平多的因子极差应该也要大一些, 直接比较 R 是不符合直观.
- 统计学家们采用一个折算公式

$$R' = \sqrt{m} \times R \times \rho$$

来解决这一问题, 其中 m 表示在该因子在每一水平下的重复试验次数, ρ 为折算系数, 不同水平数因子的折算系数为:

水平数	2	3	4	5	6	7	8	9
折算系数	0.71	0.52	0.45	0.40	0.37	0.35	0.34	0.32

(1) 试验结果的直观分析

- 因子 A 是 4 水平, 其折算系数是 0.45, 折算结果

$$R'_A = \sqrt{m_A} \times R_A \times 0.45 = \sqrt{8} \times 2.7 \times 0.45 = 3.4,$$

- B 和 C 是二水平, 折算系数是 0.71, 折算结果为

$$R'_B = \sqrt{m_B} \times R_B \times 0.71 = \sqrt{16} \times 0.9 \times 0.71 = 2.6,$$

$$R'_C = \sqrt{m_C} \times R_C \times 0.71 = \sqrt{16} \times 1.1 \times 0.71 = 3.1,$$

最后用 R'_A , R'_B , R'_C 的大小来分主次关系, 得到主次关系 $A > B > C$.

- 最佳参数组合为压力 8kg(A_1), 温度 90℃(B_2), 时间 9min(C_1).

(2) 试验结果的方差分析

- 试验的方差分析结果如下, 表明 3 个因子对胶压板质量都有显著的影响.

方差来源	自由度	平方和	均方	F	p 值
压力	3	33.34375	11.11460	9.46	0.0002
温度	1	7.03125	7.03125	5.98	0.0215
时间	1	9.03125	9.03125	7.68	0.0102
误差	26	30.56250	1.17755		
总和	31	79.96875			

Example (拟水平法)

为了提高某化工产品的转化率, 选择 3 个试验因子: 反应温度 (A), 反应时间 (B), 用碱量 (C), 每个因子取 3 个水平: $A_1 = 80^\circ\text{C}$, $A_2 = 85^\circ\text{C}$, $A_3 = 90^\circ\text{C}$; $B_1 = 90\text{min}$, $B_2 = 120\text{min}$, $B_3 = 150\text{min}$; $C_1 = 5\%$, $C_2 = 6\%$, $C_3 = 7\%$.

- 如果还要考虑搅拌速度 D 这个因子, 而电动机只有快慢两挡, 能否用 $L_9(3^4)$?
- 解决的方法就是给搅拌速度凑足 3 个水平. 让搅拌速度快的 (或慢的) 一挡多重重复一次, 凑成第 3 个水平, 得到试验方案:

因子水平	温度 (A)	时间 (B)	加碱量 (C)	搅拌速度 (D)
1	80°C	90min	5%	快速
2	85°C	120min	6%	慢速
3	90°C	150min	7%	快速

这样一种凑足水平的方法叫做拟水平法.

Example (拟水平法)

为了提高某化工产品的转化率, 选择 3 个试验因子: 反应温度 (A), 反应时间 (B), 用碱量 (C), 每个因子取 3 个水平: $A_1 = 80^\circ\text{C}$, $A_2 = 85^\circ\text{C}$, $A_3 = 90^\circ\text{C}$; $B_1 = 90\text{min}$, $B_2 = 120\text{min}$, $B_3 = 150\text{min}$; $C_1 = 5\%$, $C_2 = 6\%$, $C_3 = 7\%$.

- 如果还要考虑搅拌速度 D 这个因子, 而电动机只有快慢两挡, 能否用 $L_9(3^4)$?
- 解决的方法就是给搅拌速度凑足 3 个水平. 让搅拌速度快的 (或慢的) 一挡多重重复一次, 凑成第 3 个水平, 得到试验方案:

因子水平	温度 (A)	时间 (B)	加碱量 (C)	搅拌速度 (D)
1	80°C	90min	5%	快速
2	85°C	120min	6%	慢速
3	90°C	150min	7%	快速

这样一种凑足水平的方法叫做**拟水平法**.

Example

玻璃绝缘子钢化试验选的因子水平如下:

因子水平	1	2	3	4	5
保温温度 (°C)	700	685	670	710	720
保温时间 (h)	5.5	4.5	3.5	2.5	1.5
上风压 (kPa)	130	80	110	160	180
下风压 (kPa)	240	300	340	380	440
风栅形状	I	II	III	IV	
主风嘴大小 (mm)	9	6	12		

- 其中前 4 个因子是 5 水平的, 后 2 个因子是 4 水平和 3 水平, 没有现成的正交表可以套用.
- 比较合适的办法是选正交表 $L_{25}(5^6)$, 将最后两个因子凑足 5 个水平.

- 在风栅形状这个因子中, 工程师估计形状 II 效果较好, 将它重复一次, 凑成第五个水平;
- 在主风嘴大小因子中, 估计 9 和 12 比较好, 将它们分别重复一次, 凑成五个水平, 于是因子、水平变成如下的分布:

因子水平	1	2	3	4	5
保温温度 ($^{\circ}\text{C}$)	700	685	670	710	720
保湿时间 (h)	5.5	4.5	3.5	2.5	1.5
上风压 (kPa)	130	80	110	160	180
下风压 (kPa)	240	300	340	380	440
风栅形状	I	II	III	IV	II
主风嘴大小 (mm)	9	6	12	9	12

- 拟水平法在实行时方法简单, 但处理组合出现的频率不平衡, 有时会给试验数据分析带来一些困难.

总结

- ① 正交表、表头设计的原则、混杂技术
- ② 等水平试验的正交设计
- ③ 不等水平试验的正交设计

有什么问题？