

案例 13

存在关联因子的裂区试验设计

来源	国家自然科学基金资助项目(70572044); 新世纪优秀人才支持计划(NCET — 04-0240)	类别	<input checked="" type="checkbox"/> 案例 <input type="checkbox"/> 方法
作者	何桢 梁昭磊	关键词	实验设计; 关联因子; 裂区设计; 嵌套效应建模
理论知识	关联因子; 裂区设计; 嵌套效应建模	适用层次	<input checked="" type="checkbox"/> 本科生 <input type="checkbox"/> 研究生 <input type="checkbox"/> MBA

1. 案例背景

本案例是一个裂区试验的案例,目的是优化塑料件的硬度,案例中烘烤时间和反应时间之间存在关联。拟利用嵌套因子建模的方法,尝试解决包括关联因子的裂区试验设计。

在工业的实际生产中,存在关联因子时一般被依赖的因子多为因子水平难以改变的因子(Hard To Change, HTC),比如高炉中的炉温、连续型生产中的原材料批次等,此时试验设计三原则之一的试验顺序随机化(Randomization)就很难实现或不可能实现,因为要实现炉温的随机化调节对于真实的工业生产来几乎是不现实的,这会使试验的周期和成本成倍提高。对于单纯的关于 HTC 因子的试验设计,目前已有一些关于裂区试验设计的方法可以解决。但对于裂区因子中存在关联因子的问题,在国内外的研究中还没有相关的介绍。

1.1 存在关联因子的裂区试验设计的试验空间及编码转换

当裂区试验设计中存在关联因子时,在不同裂区中副区的因子水平不统一,这是不符合正规裂区试验设计的要求的。目前,为了能使得试验设计人员可以应用试验设计理论进行分析,设计者不得不舍弃一些副区因子的水平区域,如在有关温度和压力的关联裂区试验设计中,在试验设备和技术允许的情况,试验者本可以在压力为低水平的时候,使温度的取值范围大于压力为高水平的情况,但为了满足试验设计的理论分析要求,必须使得各个裂区中的副区取值范围想同,这就导致试验者必须对一些副区的试验区域进行裁剪。虽然这可使得试验设计满足理论要求,但也给试验设计带来一定的风险,因为裁剪掉的区域都是满足实际试验条件的区域,是否这些区域中存在试验者期望的最优值就无从得知了。而且在某些情况下,如果不同试验裂区里副区因子的取值范围没有重合区域,就会使试验者无法进行试验区域的剪裁,从而无法安排试验设计。因此,为了能够使设计的试验水平组合更符合现实生产的条件,同时尽可能保存可能存在最优的试验区域,有必要对存在关联因子的裂区试验设计分析做相应的变化。

考虑存在两个相关联的因子 A 和 B 的关联因子裂区试验,其中 A 为 HTC 因子, B 为 ETC (Easy To Change) 因子,对应与 A 的每个水平, B 有 j 个因子水平,基本的裂区试验设计的构造如表 1 所示,此表是按照标准顺序排列,在试验中的需要将主区和裂区分别随机化后进行试验:

表 1 关联因子裂区试验构造表

A1			A2			...	Ai		
B11	...	B1j	B21	...	B2j	...	Bi1	...	Bij

我们注意到,虽然裂区中的试验点数都相同,但因为各副区中副区因子的取值已经不同,所以编码的转化方式也不同与一般的试验设计,此处的编码处理公式为式 3。

$$X_{ij} = \frac{Z_{ij} - Z_{i0}}{\Delta_i} \quad (3)$$

$$\text{其中, } \Delta_i = \frac{Z_{ij} - Z_{i1}}{2}$$

X_{ij} 表示在主区因素的第 i 个水平上, 关联因子 j 水平的编码; Z_{ij} 表示在主区因素的第 i 个水平上, 关联因子 j 水平的实际取值; Z_{i0} 表示在主区因素的第 i 个水平上, 关联因子实际值的中间值; Δ_i 表示主区因子的第 i 个水平上的变化半径。

由以上的编码处理可知, 经过编码处理后在各不同主区上, 副区因子的编码取值没有差别, 但此时已经不能简单的应用裂区试验设计的分析方法进行实验数据的处理, 因为不同的裂区中的相同的编码代表的意义已经不同。

为方便理解, 下面采用两因子两水平的情况给予说明。假定存在两因子 X_1 表示烘烤的温度, X_2 表示反应的时间, 两因子都分别为两水平。其因子的实际取值和编码处理后的取值如表 2 和表 3 所示

表 2 存在关联因子的两因子裂区试验因子水平对照表

		反应时间	
		-1	1
烘烤温度	-1	10	20
	1	5	12

表 3 存在关联因子的两因子裂区试验试验组合表

烘烤温度	反应时间	X_1	X_2
80	10	-1	-1
80	20	-1	1
200	5	1	-1
200	12	1	1

相应的试验区域也不再为标准的试验区域, 其试验空间如图 1 所示。

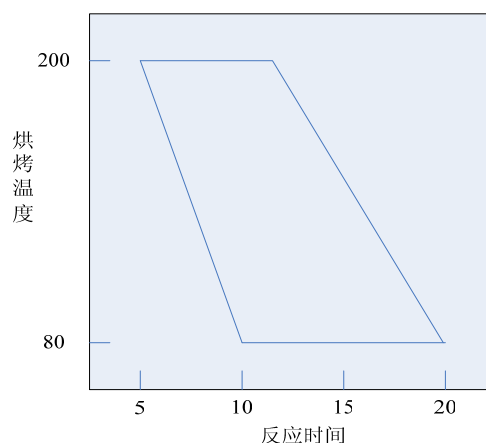


图 1 两因子关联裂区试验设计试验空间图

由图 1 可知, 由两因子的原始取值所画出的试验区域图不再是矩形, 而是一个梯形。当在此基础上通过如上所述的编码处理后, 可以看出, X_2 在反应时间为 5 和 10 时都被处理为编码为-1 的值, 按照经典试验设计的处理方式, 在此后的分析中可以认为这是相同的取值, 但由于关联因子的特殊性, 此时的-1 并不代表 X_2 相同的取值。因此在数据的分析中需要对关联因子进行特殊的处理。

1.2 存在关联因子的裂区试验设计的数据分析

因为嵌套在不同的 A 因子水平下的 B 因子的效应应该分别计算才能更真实的反应试验的本质。此时需要拟合的方程为式 3。

$$Y = \beta_0 + \beta_1 X_A + e_{wp} + \sum_{i=1}^n \beta_2 X_B(X_{Ai}) + e_{sp} \quad (3)$$

其中, Y 为响应值, $B_i (i = 0, 1, \dots, n)$ 为拟合方程的系数, $X_B(X_{Ai})$ 表示当 X_A 取 X_{Ai} 水平时 X_B 所取的水平。 e_{wp} 和 e_{sp} 分别表示主区和副区的拟合误差项。

所对应的方差分析表如表 4 所示:

表 4 关联因子裂区试验设计方差分析表

	变异来源	DF	SS
主区	A 因素	$a-1$	$mnb \sum_{i=1}^a (\overline{y_{i..}} - \overline{\overline{y_{...}}})^2$
	主区误差	$r-a$	$SS_M - SS_A$
	主区	$r-1$	$nb \sum_{k=1}^r (\overline{y_{..k}} - \overline{\overline{y_{...}}})^2$
副区	B 因素 (A 因素取水平 1 时)	$b-1$	$mn \sum_{j=1}^b (\overline{y_{1j.}} - \overline{\overline{y_{1..}}})^2$

	B 因素 (A 因素取水平 i 时)	$b-1$	$mn \sum_{j=1}^b (\overline{y_{ij.}} - \overline{\overline{y_{i..}}})^2$

	B 因素 (A 因素取水平 a 时)	$b-1$	$mn \sum_{j=1}^b (\overline{y_{aj.}} - \overline{\overline{y_{a..}}})^2$
	副区误差	$ab(r-1) + a - r$	$SS_T - SS_M - \sum_{i=1}^a SS_B$

	总变异	$rab-1$	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (y_{ijk} - \bar{\bar{y}})^2$
--	-----	---------	--

其中， m 、 n 分别代表因素 A 和因素 B 的重复次数， a 、 b 分别为因素 A 和因素 B 的水平数， r 为区组个数。

$$\bar{y}_{ij.} = \frac{\sum_{k=1}^r y_{ijk}}{r}; \quad \bar{y}_{i..} = \frac{\sum_{j=1}^b \sum_{k=1}^r y_{ijk}}{br}; \quad \bar{y}_{..k} = \frac{\sum_{i=1}^a \sum_{j=1}^b y_{ijk}}{ab}; \quad \bar{\bar{y}} = \frac{\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r y_{ijk}}{abr};$$

$$SS_M = nb \sum_{k=1}^r (\bar{y}_{..k} - \bar{\bar{y}})^2;$$

$$SS_A = mn \sum_{i=1}^a (\bar{y}_{i..} - \bar{\bar{y}})^2;$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (y_{ijk} - \bar{\bar{y}})^2;$$

$$SS_B = mn \sum_{j=1}^b (\bar{y}_{.j.} - \bar{\bar{y}})^2;$$

此外，因为因素 B 的编码在因素 A 的各不同水平中代表的实际意义是不同的，所以试验相对与因素 B 必须有重复，即满足条件： $m+n > 2$ 或在因素 A 的各不同水平中都设置重复的中心点，以便估计副区中的效应。

2. 案例描述

本案例中将关联因子定义为烘烤时间和反应时间之间存在关联。可能存在的影响因子为： A ：烘烤温度， B ：添加剂， C ：搅拌速度， D ：反应时间。其中 A 、 B 、 C 为二水平因子，高低水平的编码分别为 +1 和 -1， D 因子的取值依赖与 A 因子，当烘烤的时间较短，即 A 的水平为 -1 时，其取值范围较大，如 10-20 分钟；反之，当烘烤时间较长，即 A 的水平为 +1 时，其取值范围较小，如 5-15 分钟。为此，可设定 D 在 A 因素的高低水平的嵌入效应分别为 $D|A_{-1}$ 和 $D|A_{+1}$ ，两者的高低水平编码也都为 +1 和 -1，但此时编码代表的意义已经不同。因烘烤的温度调节起来耗时较多，为节约试验的运行时间，本试验采用了裂区试验设计，先将后三个因子按照全因子试验设计的原理进行随试验设计，两水平三因素的全因子设计需要 8 次试验，然后统一将这组试验进行烘烤，这样就可以显著的减少因调节烘烤温度所用的时间。在此试验设计中重复数为 2，因此，总的试验次数为 32 次。

由以上的试验说明可知，此试验为存在关联因子的裂区试验设计，其中，关联因子为因子 A 烘烤温度和因子 D 反应时间。此外，在此试验设计中，烘烤温度为 HTC 因子，其他三因子为 ETC 因子，所以裂区试验的主区因子为烘烤温度，其他三个因子为裂区因子。需要拟合的包含一阶主效应和两因子交互作用的方程为：

$$y = b_0 + b_1x_A + b_2x_B + b_3x_C + b_4x_{D|A_{+1}} + b_4x_{D|A_{-1}} + b_5x_Ax_B + b_5x_Ax_C + b_5x_Bx_C \quad (4)$$

其中，此试验设计的因子编码组合和采集的数据如表 5:

表 5 优化塑料件硬度试验设计编码及数据

主区	温度	添加剂	搅拌 速度	反应时间 温 度取高水平	反应时间 温 度取低水平	强度
1	1	-1	1	1	0	68.5
1	1	1	-1	1	0	66.8
1	1	-1	-1	-1	0	58.5
1	1	1	1	1	0	70.8
1	1	-1	1	-1	0	61.3
1	1	1	-1	-1	0	51.9
1	1	-1	-1	1	0	59.5
1	1	1	1	-1	0	66.2
2	-1	1	-1	0	-1	57.4
2	-1	1	-1	0	1	57.5
2	-1	-1	1	0	-1	56.5
2	-1	1	1	0	1	63.9
2	-1	-1	1	0	1	56.4
2	-1	1	1	0	-1	58.1
2	-1	-1	-1	0	1	53.2
2	-1	-1	-1	0	-1	59.5
3	-1	-1	-1	0	-1	66.6
3	-1	-1	-1	0	1	63.9
3	-1	1	1	0	-1	62.6
3	-1	1	1	0	1	63.2
3	-1	-1	1	0	-1	56.1
3	-1	1	-1	0	1	63.3
3	-1	-1	1	0	1	62.7
3	-1	1	-1	0	-1	65.0
4	1	-1	-1	-1	0	59.5
4	1	1	1	-1	0	64.0
4	1	-1	1	1	0	68.0
4	1	1	-1	-1	0	65.6
4	1	-1	1	-1	0	58.6
4	1	1	1	1	0	73.3
4	1	1	-1	1	0	61.5
4	1	-1	-1	1	0	64.2

根据以上的分析，我们可以利用 minitab 软件分析试验的数据，结果如表 6 所示:

表 6 塑料件硬度试验分析结果

来源	自由度	平方和	均方	F	P
温度	1	85.478	85.478	1.52	0.234
主区误差	2	112.391	56.196		
添加剂	1	45.363	45.363	4.23	0.055
搅拌速度	1	41.178	41.178	3.84	0.067
反应时间 高温	1	138.063	138.063	12.88	0.002
反应时间 低温	1	0.331	0.331	0.03	0.863
添加剂* 搅拌速度	1	27.938	27.938	2.61	0.125
温度* 添加剂	1	1.088	1.088	0.10	0.754
温度* 搅拌速度	1	78.438	78.438	7.32	0.015
添加剂* 反应时间 高温	1	0.360	0.360	0.03	0.857
添加剂* 反应时间 低温	1	3.331	3.331	0.31	0.585
搅拌速度* 反应时间 高温	1	12.250	12.250	1.14	0.300
搅拌速度* 反应时间 低温	1	34.516	34.516	3.22	0.091
误差	17	182.288	10.723		
合计	31	763.010			

由表 6 可知，在试验中有两个显著的因子，分别为在温度取高水平时反应时间，即 $D|A_{+1}$ ，温度和搅拌速度的交互作用，根据试验设计的要求，交互作用显著时，拟合的回归方程需要包含交互作用项中的主效应项，因此，此时参与拟合的共包含 5 个显著因素。在结果分析中，选用逐步回归的方法，与原分析中的结果比较如下表 7：

表 7 关联裂区分析结果与原结果的比较

关联裂区分析		原分析	
因素	R^2	因素	R^2
反应时间 高温	18.09	温度	11.20
温度	29.30	温度 * 搅拌速度	21.48
温度*搅拌速度	39.58	反应时间	31.44
搅拌速度	44.97	添加剂	37.38
		搅拌速度 * 反应时间	43.14

由表 7 可知，虽然由于数据的原因，两个模型的拟合结果都不能令人满意，但在关联因子的裂区试验设计中，我们仅用分析的 4 个显著因子就可以得到比原分析中 5 个显著因子还好的拟合效果。因此，可以认为关联因子的裂区分析比目前的分析方法更有效。

3. 结论与展望

针对相关联因子中存在 HTC 因子这类在工业试验中常见的约束试验设计，在裂区试验设计的基础上引入嵌套因子设计，利用裂区设计满足了存在 HTC 因子时试验顺序在不能完全随机化的约束条件，嵌套因子设计处理了因子间存在关联约束的问题，从而使裂区试验中存在关联因子的问题得以解决，扩展了试验设计的适用范围。通过实例的分析与对比，这种关联因子裂区试验设计的方法更能有效的解释试验因素的效应。

4. 参考文献

- [1] Montgomery D C. Design and Analysis of Experiments. Sixth. Hoboken, [M]. Wiley and Sons, 2005.
- [2] Box G E, Hunter J S, Hunter W G. Statistics for Experimenters: Design, Innovation, and Discovery. 2nd. [M]. John Wiley & Sons, Inc, 2005.
- [3] Taguchi G. System of experimental design[J]. White Plains, New York: UNIPUB/Krauss International. 1987.
- [4] Hamada M, Wu C F. The treatment of related experimental factors by sliding levels[J]. Journal of Quality Technology. 1995, 27: 45-55.
- [5] Cheng S W, Wu C F, Huwang L. Statistical modeling for experiments with sliding levels[J]. Time Series and Related Topics: In Memory of Ching-zong Wei. 2007.
- [6] Bingham D R, Sitter R R. Design issues in fractional factorial split-plot designs[J]. Journal of Quality Technology. 2001, 33: 2-15.
- [7] Vining G G, Kowalski S M. Exact inference for response surface designs within a split-plot structure[J]. Submitted for publication. 2006.
- [8] Goos P, Donev A N. Tailor-made split-plot designs for mixture and process variables[J]. J. Qual. Technol. 2007.