

均匀设计的方法与应用

张国权 吕小欢 罗志刚

(华南农业大学理学院, 广州 510642)

摘要 论述了均匀设计的思想和方法, 探讨其在农业试验中的应用.

关键词 试验设计; 均匀设计; 回归设计

中图分类号 O 212. 6

如何合理安排试验, 使通过尽量少次试验, 达到较好的试验效果, 是科学家们不断探索和研究的一个问题, 称为“试验设计”. 70 年代以来, 我国推广“正交设计”方法并取得丰硕的成果. 然而当试验需考察的因素较多, 且每个因素有较多的水平时, 运用“正交设计”方法所做的试验次数仍会较多, 以至难于安排试验. 例如 1978 年原七机部在进行某项产品的试验设计时(方开泰, 1994), 须考虑 5 因素 31 水平, 且要求试验次数不能超过 50 次. 5 因素 31 水平可能的试验次数多达 2 800 多万次, 为研究其数学模型曾试用国外的方法, 长时间得不到理想的结果, 而运用“正交设计”方法, 5 因素 31 水平的试验次数为 961. 为解决该难题, 我国著名的数理统计专家方开泰(1996)与数论专家王元合作, 将数论理论成功地应用于试验设计问题中, 创立了一种全新的试验设计方法——“均匀设计试验法”, 运用该方法于上述的 5 因素 31 水平的试验问题, 仅做 31 次试验, 其效果便接近于 2 800 多万次的试验, 成功地解决了该难题.

多年来, 我国数学界在数论的理论研究与应用研究两方面都卓有成效, “均匀设计”方法的创立就是其中一个例子. 10 多年来, “均匀设计”方法已广泛应用于国内的军工、化工、医药等领域, 并取得显著的成效. 在国际上, “均匀设计”也已得到承认和应用, 引起了国际数学界的重视.

本文介绍“均匀设计”方法及其应用, 希望引起国内农业科学界的重视.

1 均匀设计的思想

正交设计法是从全面试验中挑选部分试验点进行试验, 它在挑选试验点时有两个特点, 即“均匀分散, 整齐可比”. “均匀分散”使试验点具有代表性, “整齐可比”便于试验的数据分析. 然而, 为了照顾“整齐可比”, 试验点就不能充分地“均匀分散”, 且试验点的数目就会比较多(试验次数随水平数的平方而增加). “均匀设计”方法的思路是去掉“整齐可比”的要求, 通过提高试验点“均匀分散”的程度, 使试验点具有更好的代表性, 使得能用较少的试验获得较多的信息.

均匀设计沿用近 30 年来发展起来的“回归设计”方法, 运用控制论中的“黑箱”思想, 把整个过程看作一个“黑箱”, 把参与试验的因素 $x_1, x_2, \dots, x_n$, 通过运用均匀设计法安排试验, 作为系统的输入参数, 而把实验指标(结果) Y 作为输出参数, 如图 1.

在数学上可把输出参数 Y 与输入参数 $x_i(i=1, 2, \dots, n)$ 的关系用函数式表示为

$$Y=f(x_1, x_2, \cdots, x_n). \tag{1}$$

函数的模型对不同的系统可根据理论或凭经验进行假设, 然后根据试验结果运用回归分析方法确定模型中的系数, 具体计算可使用国内外现已广泛流行的统计软件 SAS、Minitab 等在计算机上进行.

不同系统的函数模型的形式及复杂程度可能相差很大. 线性模型是最简单也是较常用的一种, 但在现实中往往有其局限性, 尤其当输入参数取值范围较大时, 正如几何中曲线在局部可用直线段近似表示, 在较大范围内直线段表示就会有较大的偏差. 当线性模型假设失效时, 可以考虑多项式模型(2 次的或更高次的), 还可以考虑非多项式模型. 任何模型假设都必须通过试验进行检验和评价以确定取舍.

利用建立的回归模型, 可估计各因素的主效应和交互效应, 还可进行预测、预报等.



图 1 试验因素(输入)与试验指标(输出)系统

2 均匀设计表及其特点

2.1 均匀设计表及其使用表

均匀设计和正交设计相似, 也是精心构造一套试验表(方开泰, 1994), 利用试验表来安排试验. 表 1(a)、2(a)和 3(a)是均匀设计表(也称 U 表)的 3 个例子. 均匀设计表的代号 $U_n(q^s)$ 或 $U_n^*(q^s)$ 中的“U”表示均匀设计, “n”表示要做 n 次试验, “q”表示每个因素有 q 个水平, “s”是表中列的数目, 表示最多可安排 s 个因素, U 的右上角的“*”代表不同类型的均匀设计表. 例如 $U_6(6^4)$ 表示要做 6 次试验, 每个因素有 6 个水平, 该表有 4 列. 通常带“*”号的均匀设计表有更好的均匀性, 应优先选用.

表 1 均匀设计表 $U_6(6^4)$ 及其使用表

(a) $U_6(6^4)$				
试验号	1	2	3	4
1	1	2	3	6
2	2	4	6	5
3	3	6	2	4
4	4	1	5	3
5	5	3	1	2
6	6	5	4	1

(b) $U_6(6^4)$ 的使用表				
因素数	列 号			D
1	1	3		0.187 5
2	1	2	3	0.265 6
3	1	2	3 4	0.299 0

表 2 均匀设计表 $U_7(7^4)$ 及其使用表

(a) $U_7(7^4)$				
试验号	1	2	3	4
1	1	2	3	6
2	2	4	6	5
3	3	6	2	4
4	4	1	5	3
5	5	3	1	2
6	6	5	4	1
7	7	7	7	7

(b) $U_7(7^4)$ 的使用表				
因素数	列 号			D
2	1	3	3	0.239 8
3	1	2	3	0.372 1
4	1	2	3 4	0.476 0

每张均匀设计表都附有一张使用表, 它指示我们如何从均匀设计表中选用适当的列, 以及由这些列所组成的试验方案的均匀度. 表 1(b)、2(b)和 3(b)分别是表 1(a)、2(a)和 3(a)的附表. 表 1(b)是表 1(a), 即 $U_6(6^4)$ 表的使用表. 它告诉我们, 若有两个因素, 应选用 1, 3 两列来安排试验; 若有 3 个因素, 应选用 1, 2, 3 三列, ……最后 1 列 D 表示偏差值, 用于刻划均匀度的偏差 (discrepancy), 偏差值越小, 表示均匀度越好. 例如使用两个均匀设计表 $U_7(7^4)$ 和 $U_7^*(7^4)$ 及它们的使用表来安排含两个因素的试验, 若选用 $U_7(7^4)$ 的 1, 3 列, 其偏差 $D=0.239\ 8$; 若选用 $U_7^*(7^4)$ 的 1, 3 列, 相应偏差 $D=0.158\ 2$, 后者较小, 应优先择用.

另外, 均匀设计方法还构造了混合水平的均匀设计表.
均匀设计表及其使用表的构造均应用了数论的理论和方法(方开泰等, 1996).

2.2 均匀设计表的特点

均匀设计有其独特的布置试验点的方式, 其特点表现在以下几方面.

- (1) 每个因素的每个水平做一次且仅做一次试验.
- (2) 任两个因素的试验点描在平面的格子
上, 每行每列有且仅有一个试验点. 如表 $U_6(6^4)$ 的第 1 列和第 3 列点成图 2(a).

性质 1 和 2 反映了试验安排的均衡性, 即对各因素, 每个因素的各个水平给予同样的重视.

- (3) 均匀设计表任两列组成的试验方案一般是不平等的. 例如用 $U_6(6^4)$ 的 1, 3 和 1, 4 列中的点分别作图, 得图 2(a)和图 2(b). 可见, 图 2(a)的点分布均匀, 而图 2(b) 的点分布不那么均匀. 因此使用均匀设计表时我们还要挑选均匀性搭配比较好的列, 而每个均匀设计表的使用表的用途就是帮助我们进行挑选. 例如, 使用 $U_7^*(7^4)$ 表即表 3(a)时, 依据其使用表即表 3(b)可知, 当有 2 个因素时, 用 1, 3 列; 当有 3 个因素时, 用 2, 3, 4 列.

- (4) 运用“均匀设计法”时, 试验数随水平数的增加而增加. 如当水平数从 9 水平增加到 10 水平, 实验数也从 9 增加到 10. 而正交设计当水平增加时, 试验数按水平的平方而增加, 如当水平数从 9 到 10 时, 试验数将从 81 增加到 100. 实验数随水平数较缓慢的增加, 是均匀设计的一个突出优点.

3 均匀设计法的试验安排

运用均匀设计法安排试验, 其步骤和正交设计有相似之处, 通常有如下步骤.

表 3 均匀设计表 $U_7^*(7^4)$ 及其使用表
(a) $U_7^*(7^4)$

试验号	1	2	3	4
1	1	3	5	7
2	2	6	2	6
3	3	1	7	5
4	4	4	4	4
5	5	7	1	3
6	6	2	6	2
7	7	5	3	1

(b) $U_7^*(7^4)$ 的使用表

因素数	列 号				D
2	1	2	3		0.158 2
3	1	2	3	4	0.213 2

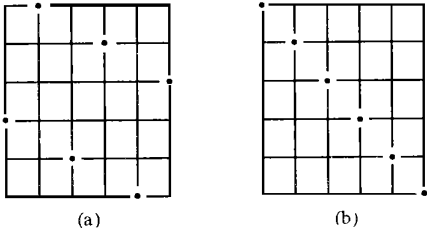


图 2 2 因素 6 水平试验的试验点分布

- (1) 确定试验目的. 必须首先明确试验的指标, 如果试验要考察的指标不只一个, 还要将各指标综合考虑.
- (2) 定因素. 挑选试验中可能影响指标的因素.
- (3) 选水平. 结合实验目的和以往经验对每个因素选择合适的水平.
- (4) 选均匀设计表. 视试验的因素个数、水平数等具体情况从有关文献提供的均匀设计表套中选择一张合适的均匀设计表.
- (5) 制定设计方案. 从均匀设计表的使用表中选出列号, 将因素分别安排到这些列号上, 并将这些因素的各水平值按所在列的编号分别对号填入, 编制出试验方案表.
- (6) 实施试验并把试验结果填写在实验方案表的最右一列上.
- 下面以我们近期作的一个试验为例说明均匀设计法的试验安排.

试验 1 我们用二甲酚橙分光光度法测定微量锆, 寻找最佳显色条件, 选取了显色剂用量(B)、酸度(c)、温度(θ)和络合物稳定时间(t)4 个因素, 它们的变化范围选择为: B : 0.1 ~ 1.3 mL, c : 0.1 ~ 1.3 mol $\cdot$ L $^{-1}$, θ : 20 ~ 80 $^{\circ}$ C, t : 0 ~ 24 h 各取了 13 个水平, 列于表 4.

表 4 试验 1 的因素水平表

水平列号	B/mL	$c_{\text{HClO}_4}/(\text{mol}\cdot\text{L}^{-1})$	$\theta/^{\circ}\text{C}$	t/h
1	0.1	0.1	20	0
2	0.2	0.2	25	2
3	0.3	0.3	30	4
4	0.4	0.4	35	6
5	0.5	0.5	40	8
6	0.6	0.6	45	10
7	0.7	0.7	50	12
8	0.8	0.8	55	14
9	0.9	0.9	60	16
10	1.0	1.0	65	18
11	1.1	1.1	70	20
12	1.2	1.2	75	22
13	1.3	1.3	80	24

表 5 均匀设计表 $U_{13}(13^4)$ 及其使用表

(a) $U_{13}^*(13^4)$				
试验号	1	2	3	4
1	1	5	9	11
2	2	10	4	8
3	3	1	13	5
4	4	6	8	2
5	5	11	3	13
6	6	2	12	10
7	7	7	7	7
8	8	12	2	4
9	9	3	11	1
10	10	8	6	12
11	11	13	1	9
12	12	4	10	6
13	13	9	5	3

(b) $U_{13}^*(13^4)$ 的使用表					
因素数	列 号				D
2	1	3			0.096 2
3	1	2	3		0.144 2
4	1	2	3	4	0.207 6

根据因素和水平, 我们选取均匀设计表 $U_{13}^*(13^4)$, 即表 5(a). 由其使用表 5(b)可查到, 当因素数 $s=4$ 时(此时偏差 $D=0.207 6$), 应选用表 $U_{13}^*(13^4)$ 的 1, 2, 3, 4 列的各水平编号分别对应因素 A, B, C, D 的各水平. 具体试验方案及试验结果(吸光度, 记为 A)列于表 6.

从表 6 可见, 第 10 号试验的结果是最好的, 其工艺条件为 $B=1.0$ mL, $c_{\text{HClO}_4}=0.8$ mol $\cdot$ L $^{-1}$, $\theta=45$ $^{\circ}$ C, $t=22$ h. 由于试验点具有较好的代表性, 因此可以猜想此工艺条件即使不是全面试验的最佳条件, 也是相对较优的条件. 下面将通过统计分析, 寻找更好的工艺条件.

表 6 试验 1 的试验方案 $U_{13}^*(13^4)$ 和结果¹⁾

试验号	1	2	3	4	5
	B/mL	$\text{CHClO}_4/(\text{mol}^*\text{L}^{-1})$	$\theta/^\circ\text{C}$	t/h	A
1	0.1(1)	0.5(5)	60(9)	20(11)	0.376
2	0.2(2)	1.0(10)	35(4)	14(8)	0.384
3	0.3(3)	0.1(1)	80(13)	8(5)	0.342
4	0.4(4)	0.6(6)	55(8)	2(2)	0.412
5	0.5(5)	1.1(11)	30(3)	24(13)	0.387
6	0.6(6)	0.2(2)	75(12)	18(10)	0.368
7	0.7(7)	0.7(7)	50(7)	12(7)	0.438
8	0.8(8)	1.2(12)	25(2)	6(4)	0.391
9	0.9(9)	0.3(3)	70(11)	0(1)	0.385
10	1.0(10)	0.8(8)	45(6)	22(12)	0.441
11	1.1(11)	1.3(13)	20(1)	16(9)	0.363
12	1.2(12)	0.4(4)	65(10)	10(6)	0.380
13	1.3(13)	0.9(9)	40(5)	4(3)	0.426

1) 表中圆括号内的数字是水平编号. 显色剂质量分数为 0.2%, 锆标准溶液用 3 mL, 铪标准溶液用 5 mL, 锆测定波长 550 nm

4 试验分析

试验设计的目的通常主要有 2 个, 一是揭示变量 Y 与各因素之间的定量、定性关系. 二是寻求最优工艺条件. 回归方程的建立可以同时达到这两个目的.

4.1 回归模型及其分析

现建立回归模型, 设吸光度 Y 、显色剂用量 x_1 、 HClO_4 浓度 x_2 、温度 x_3 和稳定时间 x_4 :

$$Y=f(x_1, x_2, x_3, x_4), \tag{2}$$

先考虑线性模型, 运用 Minitab 软件在计算机上进行回归计算, 发现因素 x_3 与其他因素高度相关而被自动剔除, 得回归方程:

$$Y=0.367\ 0+0.021\ 3x_1+0.017\ 1x_2-0.000\ 13x_4, \tag{3}$$

其显著性检验的统计量 $F=0.57$. 因为 $F_{3,9}(0.05)=3.86>F=0.57$, 故回归方程式 (3) 显然不可信. 在对回归系数的假设检验中, 发现最不显著的因素是 x_4 , 于是剔除 x_4 , 再进行回归计算, 得回归方程:

$$Y=0.365\ 0+0.022\ 1x_1+0.016\ 3x_2. \tag{4}$$

其显著性检验的统计量 $F=0.95$. 因为 $F_{2,10}(0.05)=4.10>F=0.95$, 故回归方程式 (4) 显然不可信. 由此猜测线性模型未必符合实际, 于是考虑二次回归模型. 经运用统计分析软件 Minitab 的最优子集回归法筛选变量, 发现在二次回归模型中因素 x_1 和 x_2 及其交互作用对响应值 Y 的影响最显著, 而因素 x_3 和 x_4 的影响都不显著, 应予剔除. 经二次回归计算, 得回归方程:

$$Y=0.282+0.101\ x_1+0.300\ x_2-0.091\ 5\ x_1^2-0.238\ x_2^2+0.070\ 5\ x_1x_2. \tag{5}$$

其显著性检验的统计量 $F=27.31$. 因为 $F_{5,7}(0.05)=3.97<F=27.31$, 故回归方程式 (5) 在显著水平 $\alpha=0.05$ 下是可信的.

4.2 寻求最优工艺条件

现以试验 1 为例来说明如何寻求最好的工艺条件. 从表 6 可见, 第 10 号试验的工艺条件是 13 次试验中最好的, 即 $B=1.0\text{ mL}$, $c_{\text{HClO}_4}=0.8\text{ mol}\cdot\text{L}^{-1}$, $\theta=45\text{ }^\circ\text{C}$, $t=22\text{ h}$. 这个工艺条件和最优工艺条件常常是很接近的.

上面得到的回归模型方程 (5) 一般仅在试验范围内成立, 即 $B=0.1\sim 1.3\text{ mL}$, $c_{\text{HClO}_4}=0.1\sim 1.3\text{ mol}\cdot\text{L}^{-1}$, $\theta=20\sim 80\text{ }^\circ\text{C}$ 和 $t=0\sim 24\text{ h}$. 寻求最优模型等价于在这个范围内求式 (5) 中 Y 的极大值. 对于已建立的回归方程, 我们可适当地选用一种优化算法来求最佳工艺条件.

对试验 1 中所得回归方程式 (5), 可以用简单的微积分求得极值. 经计算得, 当 $x_1=0.842\ 8$, $x_2=0.755\ 1$ 时, Y 取极大值.

为了验证上面求得的最优工艺条件的理论值, 我们进行了下述的试验 2.

试验 2 重复进行试验 1 中的第 7~10 号试验, 在该工艺条件下 (取 $x_1=0.84$, $x_2=0.76$) 进行试验, 具体试验方案及试验结果列于表 7.

从表 7 可见, 在理论最优工艺条件下得到最好的结果.

5 结语

农业试验往往具有周期长、季节性强、受天气等难预测的环境变化因素干扰等特点, 有些农业试验还涉及面积大、范围广, 试验次数通常难以太多; 面向推广的农业试验的结论的正确性、可推广性对农业生产具有重大后果, 必须严格谨慎.

“均匀设计法”是一种成熟的试验设计方法, 具有多种优点, 突出优点是适用多因素多水平而试验次数又比较少的设计, 实践已经证明该法适用于多个领域的科学试验. 鉴于农业试验的特点, 我们认为“均匀设计法”在农业科学试验上的应用, 是应该也值得推广的.

参 考 文 献

方开泰. 1994. 均匀设计与均匀设计表. 北京: 科学出版社, 1~30
方开泰, 王 元. 1996. 数论方法在统计中的应用. 北京: 科学出版社, 5~36

表 7 试验 2 的试验方案和结果¹⁾

试验号	B/mL	$c_{\text{HClO}_4}/\text{mol}\cdot\text{L}^{-1}$	A
7	0.7 (7)	0.7 (7)	0.442
8	0.8 (8)	1.2 (12)	0.392
9	0.9 (9)	0.3 (3)	0.378
10	1.0 (10)	0.8 (8)	0.444
理论最优值	0.84	0.76	0.451

1) 表中圆括号内的数字是水平编号, 取自表 6. 温度为室温, 时间在 1 h 内

METHOD OF UNIFORM DESIGN AND ITS APPLICATIONS

Zhang Guoquan Lu Xiaohuan Luo Zhigang

(College of Sciences South China Agric. Univ., Guangzhou 510642)

Abstract

The idea and method of uniform design were reviewed and its application to agricultural science was discussed.

Key words experimental design; uniform design; regression design