

回归通用旋转设计的几个问题*

卢恩双, 宋世德, 郭满才

(西北农林科技大学 生命科学学院, 陕西 杨陵 712100)

[摘要] 通过理论和实例着重论述了二次回归通用旋转设计中试验次数、设计的正交性、回归系数的计算及回归系数间的相关性等, 说明了在应用通用旋转设计与二次正交设计上的不同, 并指出了在进行回归系数显著性检验时需要注意的问题。

[关键词] 二次回归通用旋转设计; 正交性; 回归系数

[中图分类号] O 212.1

[文献标识码] A

[文章编号] 1671-9387(2002)05-0110-04

二次回归通用旋转设计因其能使回归预测值 $\hat{y}$ 在以试验区域的中心为球心, 半径为 ρ 的球内为一常数, 即设计的一致精度和需要试验次数较少等优点广泛应用于生产试验和科学试验中。但在不少应用论文中往往对它的一些基本特点和计算不清楚, 易与二次回归正交设计和二次回归正交旋转设计混淆, 出现计算和分析上的一些错误。现就二次回归通用旋转设计中的一些问题进行探讨, 以期为其的正确应用提供理论参考。

1 应用通用回归旋转设计的几个问题

1.1 试验次数和臂

二次回归通用旋转设计与二次回归正交设计和二次回归正交旋转设计一样都是组合设计。设计次数 n 由3部分组成: 即 $n = m_c + m_r + m_0$ 。 m_c 是析因试验次数, m_r 是在星号臂上进行的试验次数, m_0 是在试验区域的中心点进行的试验次数。在回归设计中, 确定了自变量数和试验是全实施还是半实施等, m_c 和 m_r 就已经确定。当设计选为二次回归正交设计时, m_0 可以自由选定, 当 m_0 取定后 n 就确定了; 在这个设计中, 根据试验要满足正交性来计算星号臂 r : $r^2 = \frac{(m_c + 2p + m_0)m_c - m_c}{2}$ 。但当设计选用正交旋转和通用旋转设计时, r 根据设计的旋转要求确定: $r = \sqrt[4]{m_c}$ 。在这2个设计中, n 都是事先确定的, m_0 不可以自由选定, 前者根据满足正交性条件而定, 后者根据满足通用性条件而定, 这里计算公式较繁, 不再给出, 但都有现成的表可查^[1]。

1.2 设计的正交性和回归系数计算及其显著性检验

在通用旋转设计中, 为达到精度的一致性是以牺牲部分的正交性为代价的, 因此回归系数的计算及其显著性检验以及分析变量的重要性等, 都与二次回归正交设计和二次回归正交旋转设计不同, 应特别注意。下面以 $p = 2$ 为例, 比较回归正交旋转设计与通用旋转设计的结构矩阵(分别为 X_1 和 X_2)。设响应变量

$$y = b_0 + b_1z_1 + b_2z_2 + b_{12}z_1z_2 + b_{11}z_1^2 + b_{22}z_2^2$$

式中, $b_0, b_1, \dots, b_{22}$, 为回归系数, z_1, z_2 分别是自变量 X_1, X_2 的规范变量。

	z_0	z_1	z_2	z_1z_2	z_1^2	z_2^2	
$X_1 =$	1	1	1	1	0.5	0.5	1
	1	1	-1	-1	0.5	0.5	2
	1	-1	1	-1	0.5	0.5	3
	1	-1	-1	1	0.5	0.5	4
	1	1.414	0	0	1.5	-0.5	5
	1	-1.414	0	0	1.5	-0.5	6
	1	0	1.414	0	-0.5	1.5	7
	1	0	-1.414	0	-0.5	1.5	8
	1	0	0	0	-0.5	-0.5	9
	1	0	0	0	-0.5	-0.5	10
	1	0	0	0	-0.5	-0.5	11
	1	0	0	0	-0.5	-0.5	12
	1	0	0	0	-0.5	-0.5	13
	1	0	0	0	-0.5	-0.5	14
	1	0	0	0	-0.5	-0.5	15
	1	0	0	0	-0.5	-0.5	16

* [收稿日期] 2002-04-05

[作者简介] 卢恩双(1952-), 男, 陕西南郑人, 副教授, 主要从事应用数学的研究。

$$X_2 = \begin{bmatrix} z_0 & z_1 & z_2 & z_1 z_2 & z_1^2 & z_2^2 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1.414 & 0 & 0 & 2 & 0 \\ 1 & -1.414 & 0 & 0 & 2 & 0 \\ 1 & 0 & 1.414 & 0 & 0 & 2 \\ 1 & 0 & -1.414 & 0 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

两矩阵外右边1列皆为试验号。 X_1 中 $z_{ij}' =$

$z_{ij}^2 - \frac{1}{n} \sum_{i=1}^n z_{ij}^2, i=1, 2, \dots, n; j=1, 2$, 是为了满足正交性所进行的二次项的中心化处理, 这时 $\hat{y} = b_0 + b_1 z_1 + b_2 z_2 + b_{12} z_1 z_2 + b_{11} z_1^2 + b_{22} z_2^2$

设系数(信息)矩阵 $A = X^T X$, 由最小二乘法得 $b = A^{-1} B$, $B = (B_0, B_1, B_2, B_{12}, B_{11}, B_{22})^T$, $B_0 = \sum_{i=1}^n y_i, B_j = \sum_{i=1}^n z_{ij} y_i, B_{kj} = \sum_{i=1}^n z_{ik} z_{ij} y_i, B_{jj} = \sum_{i=1}^n z_{ij}^2 y_i$ (对 X_1)或 $B_{jj} = \sum_{i=1}^n z_{ij}^2 y_i$ (对 X_2)。称 A^{-1} 为相关矩阵, 它正好是回归系数的相关阵。

$$A_1^{-1} = \text{diag}(1/16, 1/8, 1/8, 1/4, 1/6, 1/6)$$

$$A_2^{-1} = \begin{bmatrix} \frac{1}{5} & 0 & 0 & 0 & -\frac{1}{10} & -\frac{1}{10} \\ 0 & \frac{1}{8} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ -\frac{1}{16} & 0 & 0 & 0 & \frac{23}{160} & \frac{3}{160} \\ -\frac{1}{16} & 0 & 0 & 0 & \frac{3}{160} & \frac{23}{160} \end{bmatrix}$$

由 A_1^{-1} 明显可知, 所有回归系数间都是不相关的, 即设计具有正交性。而 A_2^{-1} 不是对角阵, 显然 b_0 与 b_{11} 和 b_{22} 间, b_{11} 和 b_{22} 间都具有相关性, 所以说, 通用回归旋转设计牺牲了部分正交性来达到精度的一致性。因此, 在分析通用回归旋转设计的试验结果时, 不可直接由 b_{ij} 的大小判断各二次项作用的大小, 也不能如正交设计一样, 任意去掉某个二次项,

而不改变其余回归系数。也正由于此, 在正交旋转设计中, 回归系数的计算与显著性的检验都很容易。如

回归系数的计算公式为 $b_0 = \frac{B_0}{16}, b_j = \frac{B_j}{8}, b_{ij} = \frac{B_{ij}}{4},$

$b_{jj} = \frac{B_{jj}}{6}, j=1, 2$ 。而在通用旋转设计中: $b_0 =$

$\frac{1}{5} B_0 - \frac{1}{10} (B_{11} + B_{22}), b_j = \frac{1}{8} B_j, b_{12} = \frac{1}{4} B_{12}, b_{11} =$

$-\frac{1}{10} B_0 + \frac{23}{160} B_{11} + \frac{3}{160} B_{22} = -\frac{1}{10} B_0 + \frac{20}{160} B_{11} +$

$\frac{23-3}{160} (B_{11} + B_{22}), b_{22} = -\frac{1}{10} B_0 + \frac{3}{160} B_{11} + \frac{23}{160} B_{22} =$

$-\frac{1}{10} B_0 + \frac{20}{160} B_{22} + \frac{23-3}{160} (B_{11} + B_{22})$ 。

除了回归系数计算不一样以外, 回归方程及回归系数的显著性检验也不同。在回归正交设计和正交旋转设计中都是直接先求回归平方和, 然后求剩余平方和($Q_{\text{剩}} = SS_T - U, SS_T$ 为总平方和), 而在通

用旋转设计中要先计算 $Q_{\text{剩}} = \sum_{i=1}^n y_i^2 - b_0 B_0 - \sum_{j=1}^p b_j B_j$

$- E \sum_{k=1}^{p-1} \sum_{j=1}^p b_{kj} B_{kj} - \sum_{j=1}^p b_{jj} B_{jj}$, 然后再计算 $U = SS_T$

$- Q_{\text{剩}}$, 回归系数的显著性检验也不同, 这里不再一一介绍。

2 实 例

以建立黄土高原春大豆综合农艺措施数学模型为例, 采用4因子(1/2实施)二次回归通用旋转组合设计, 对影响陕北黄土高原春大豆产量最主要的种植密度(x_1)、施氮量(x_2)、施磷量(x_3)、施有机肥量(x_4)等4项农艺措施进行田间试验, 以期建立黄土高原春大豆综合农艺措施的数学模型, 优选最佳农艺措施组合方案, 为提高大豆产量提供科学依据^[3]。

2.1 确定变量变化范围及编码

选4因子(1/2实施)二次回归通用旋转组合设计。 $p=4$ 时, $r=1.682, n=20, m_0=4, m_c=m_r=8$ 。因素水平编码如表1。

2.2 试验安排与试验结果记录

试验地设在陕西省榆林地区米脂县桥河岔乡远老山村的旱梯田上, 土壤为黄绵土, 前茬谷子, 地力均匀, 播前土壤0~20 cm土层的理化性质为: pH 8.1, 有机质 2.45 g/kg, 全氮 0.25 g/kg, 全磷 1.23 g/kg, 全钾 13.560 g/kg, 碱解氮 31.04 mg/kg, 速效磷 7.58 mg/kg, 速效钾 109.74 mg/kg, 供试品种为黄豆 8415。1991-05-01

播种, 小区面积为 13.34 m^2 , 锄地 4 次, 1991-10-10 收获, 单收单打, 风干称重, 大豆生育期降雨 390.3 mm , 前期降雨偏多, 后期旱象严重, 生育期积温为 $3\,349\text{ }^\circ\text{C}$, 试验安排及试验结果见参考文献 [3]。

表 1 因素水平编码表

Table 1 Codes of the factor levels				
z_j	$x_1/$ (株·hm ⁻²)	$x_2/$ (kg·hm ⁻²)	$x_3/$ (kg·hm ⁻²)	$x_4/$ (kg·hm ⁻²)
上星号臂水平(1.682) Level of upper astrisk arm	135 000	150	150	15 000
上水平(1) Upper level	120 000	120	120	12 000
零水平(0) Zero level	97 500	75	75	7 500
下水平(-1) Lower level	75 000	30	30	3 000
下星号臂水平(-1.682) Level of lower astrisk arm Δ_j	60 000	0	0	0
	22 500	45	45	4 500

2.3 回归方程与拟合度及回归方程显著性检验

由计算回归系数公式可得回归方程:

$$\hat{y} = 1\,056\,902 - 3\,097z_1 + 106\,190z_2 + 89\,665z_3 + 7\,840z_4 + 15\,000z_{12} - 26\,250z_{13} + 26\,250z_{14} + 26\,250z_{23} - 26\,250z_{24} - 15\,000z_{34} - 99\,645z_1^2 + 40\,875z_2^2 + 54\,120z_3^2 + 9\,060z_4^2$$

拟合度检验 F_{L_f} 不显著, 说明本试验无其他因素的显著影响, 模型是适合的, 拟合不足被否定。回归方程检验 F 在 0.05 水平上显著, 说明试验所选因子对黄土高原春大豆产量有显著影响。回归系数 t 检验结果表明, 除 $b_1, b_4, b_{12}, b_{34}, b_{44}$ 项外, 都对大豆产量有不同程度的影响, 其中施氮量 x_2 与施磷量 x_3 的一次项, 密度 x_1 的二次项都有显著的影响。为便于对模型进行分析讨论, 不显著的各项没有去掉。特别注意若要去掉 b_{44} 项, 会影响到 b_{11}, b_{22}, b_{33} 和 b_{00} 。因为在通用旋转设计中, b_0 与 b_{ij} 及 b_{jj} 之间还存在着相关性。

2.4 对模型中几个问题的分析

2.4.1 主效应分析 各一次项回归系数 b_j 之间,

各 b_j 与交互项、平方项的回归系数间都是不相关的, 因此, 可以由回归系数绝对值的大小来直接比较各因素一次项对大豆产量的影响。这里 $x_2 > x_3 > x_4 > x_1$, x_2 和 x_3 影响较大, 且都是正效应。在通用旋转设计中, 二次项回归系数间是相关的, 不能直接由它们的绝对值大小来比较二次项作用的大小。

2.4.2 因素间的交互效应分析 因素间的交互效应分析应在显著时进行分析其中最大的, 为丰富分析结果作参考, 本试验所建立模型中的 $x_1x_3, x_1x_4, x_2x_3, x_2x_4$ 的交互作用较 x_1x_2, x_3x_4 大, 但都在 0.1 水平上不显著。根据 1988~1990 年的试验, 氮肥和磷肥的交互作用最大, 因此, 这里仅分析氮磷的交互作用。将密度和有机肥 2 因素都固定在零水平上, 得如下方程:

$$\hat{y}_{z_2z_3} = 1\,056\,902 + 106\,190z_2 + 89\,665z_3 + 26\,250z_2z_3 + 40\,875z_2^2 + 54\,120z_3^2$$

将 z_2, z_3 的各个水平的编码值代入此方程, 计算各个 $\hat{y}$ 值, 结果如表 2 所示。

表 2 施氮和施磷肥对大豆产量的交互作用

Table 2 Interaction of phosphatic fertilizer and nitrogenous fertilizer on soybean yield									
z_2	z_3					统计参数			
	施 P 量 Phosphatic fertilizer					Statistic parameter			
	-1.682	-1	0	1	1.682	$\bar{x}$	s	CV/%	
N	-1.682	1 071 090	1 003 140	994 530	1 094 175	1 224 210	1 077.450	92 550	8.59
	-1	1 038 630	982 890	992 190	1 109 730	1 251 975	1 075.050	110 850	10.31
	0	1 059 780	1 021 950	1 057 500	1 201 290	1 361 430	1 140.450	141 450	12.40
	1	1 162 695	1 142 760	1 204 560	1 374 600	1 552 650	1 287.450	174 150	13.53
	1.682	1 279 770	1 272 045	1 351 740	1 539 690	1 729 950	1 434.600	197 250	13.75
$\bar{x}$		1 122 450	1 084 500	1 120 050	1 263 900	1 424 100			
s		99 900	121 950	155 550	190 200	214 200			
CV/%		8.91	11.24	13.89	15.05	15.04			

由表 2 可见, 当种植密度为 $97\,500$ 株/hm²、施有机肥 $7\,500$ kg/hm² (皆是 0 水平)、施氮肥和磷肥

各为 150 kg/hm² (都是 1.682 水平) 时, 大豆产量最高, 达 1 729.95 kg/hm²。由表 2 还可以看出, 在施 N 肥的后 4 个水平上, 施 P 肥在 -1 水平时, 无论是增加还是减少, 都导致大豆产量的提高。在施 P 肥的前 3 个水平上, 施 N 肥在 -1 水平时, 无论再增加还是减少, 都导致大豆产量的提高。在施 P 肥为 1 和 1.682 水平时, 随着 N 肥的增加, 产量都明显的增加。当施 N、P 肥皆为 -1 水平时, 大豆产量最低, 为 982.89 kg/hm²。

从表 2 可明显看出, 在施氮肥的不同水平下, 随着磷肥的增加, 产量改变的变异度不同, 且随着磷肥的增加而增加。在施磷肥的不同水平下, 产量改变的变异度有同样的变化趋势。

根据多元函数极值理论可得, 当 z_1, z_4 固定在 0 水平时, 将 $\hat{y}_{z_2 z_3}$ 分别对 z_2 和 z_3 求偏导, 并令其为零, 可解得: $z_2 = -1.1201$ (相当于每 hm² 施 N 肥 26.4 kg), $z_3 = -0.5568$ (相当于每 hm² 施 P 肥 50 kg), 此时的大豆产量为 973 kg/hm², 是极小值, 也是在

讨论的因子空间中的最小值。还可进行多种结果分析, 如边际产量效应等, 这里就不再一一进行了。

2.4.3 用统计选优方法确定各因素水平的最优组合 用统计选优方法, 每个因素取 5 个水平: $\pm 1.682, \pm 1$ 和 0。用计算机对 $5^4 = 625$ 个方案寻优, 可得大豆产量最高的农艺方案为 $z_1 = -1, z_2 = z_3 = 1.682, z_4 = -1.682$, 相当于种植密度为 7 500 株/hm², 施 N、P 肥各 150 kg/hm², 不施有机肥, 此时产量为 1 876.065 kg/hm²。

分析可得, 大豆产量高于 1 500 kg/hm² 的方案共有 33 个, 占全部方案的 5%, 频数分析如表 3 所示 (表 3 中第 9 行数据是把规范变量转换成自然变量)。由表 3 可知, 当农艺措施中种植密度为 81 840 ~ 94 890 株/hm², 施 N、P 和有机肥依次为 137.64 ~ 147.00, 148.56 ~ 149.16 和 3 769.5 ~ 7 387.5 kg/hm² 时, 大豆产量有 95% 的可能高于 1 500 kg/hm²。

表 3 $\hat{y} > 100$ kg 的变量数值与频率

Table 3 Variable value and frequency under $\hat{y} > 100$ kg				
编码 Codes	z_1 (密度)频数 Density frequency	z_2 (N 肥)频数 N itrogenous fertilizer frequency	z_3 (P 肥)频数 Phosphorescence frequency	z_4 (有机肥)频数 O rganic fertilizer frequency
- 1.682	5	0	0	10
- 1	10	0	0	8
0	13	0	0	7
1	5	9	6	4
1.682	0	24	27	4
平均值 $\bar{z}$ The average number	- 0.406	1.496	1.558	- 0.427
$S_{\bar{z}}$	0.148	0.053	0.046	0.205
95% 的置信区间 The confidence interval	- 0.696 ~ - 0.116	- 1.392 ~ 1.600	- 1.682	- 0.829 ~ - 0.025
农艺措施 (x_j) Cultivation measure	81 840 ~ 94 890	137.64 ~ 147.00	148.56 ~ 149.16	3 769.5 ~ 7 387.5

4 讨 论

1) 选定通用旋转设计后, 试验次数 n 和星号臂处试验点的指标是固定的。通常 n 和正交旋转设计的 n 是不一样的。回归系数的计算、回归方程及回归系数的显著性检验都与二次正交回归设计及二次正

交旋转设计不同。
2) 由通用旋转设计的试验数据得到的回归方程中, 常数项系数与二次项系数、二次项系数之间都具有相关性, 切不可随意去掉不显著的二次项系数。
3) 使用统计优选法是寻找最优组合的直观而实用的方法, 方便于实施。

[参考文献]

[1] 徐中儒. 回归分析与试验设计[M]. 北京: 中国农业出版社, 1998
[2] 茆诗松, 丁元, 周纪芑. 回归分析及其试验设计[M]. 上海: 华东师范大学出版社, 1981.
[3] 林关石, 郭培才. 陕北黄土高原春大豆综合农艺措施模型研究[J]. 大豆科学, 1993, (4): 51- 56

存所起的主导作用。利用这种方法,可以探讨自然条件下不同种群生存与死亡概率规律,它对预测不同种群在不同龄区的生存率与死亡率、科学编制种群

生命表、预报作物害虫的最佳防治时间、保护生态环境等提供了科学依据。

[参考文献]

- [1] 刘光祖 概率论与应用数理统计[M]. 北京: 高等教育出版社, 2000.
- [2] 袁志发 概率基础与数理统计[M]. 北京: 农业出版社, 1987.
- [3] 陈兰荪 生物数学引论[M]. 北京: 科学出版社, 1988.
- [4] 王金福 瓜螟生物学和生态学特性的研究[J]. 浙江大学学报, 1988, 14(2): 221- 225.
- [5] 王金福 印度野瓜螟增长及其密度制约[J]. 动物研究, 1989, (3): 232- 237.

A discussion on the survival and death regularities of the population and melon borer population

SHI Mei-ying, LIU Guang-zu, ZHENG Li-fei

(College of Life Sciences, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

Abstract: The mathematical methods were used to make research into the living populations lifetime distribution, population lifetime distribution model and the intrinsic death invariant of the population were observed, and population lifetime's mathematical expectation and variance were induced. Experiment data of melon borer's population was used to fit the second and third generation population distribution function, and testing result was quite good.

Key words: population; lifetime distribution; distribution function; intrinsic death invariant; mathematical expectation

(上接第 113 页)

Problems in second-order regression general rotation analysis

LU En-shuang, SONG Shi-de, GUO Man-cai

(College of Life Sciences, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

Abstract: The experiment number, the orthogonality of experimental design, the regression coefficient's calculation and its correlativity in second-order regression general rotational design were discussed by theoretic research and example analysis, the differences between general rotation design and second-order orthogonal design were showed, some problems that should be noticed in regression coefficient significance tests were also pointed out.

Key words: second-order regression general rotation designing; orthogonality; regression coefficient