

# 均匀设计 Design Expert 统计方法的优化

孟庆红<sup>1</sup> 李灿国<sup>2</sup> 魏 萌<sup>2</sup> 孔 玲<sup>2\*</sup> 陈志伟<sup>3\*</sup>

1. 蒙牛乳业泰安有限责任公司, 中国·山东 泰安 271000;

2. 山东理工大学生命与医药学院, 中国·山东 淄博 255000;

3. 山东理工大学食品与营养科学研究院, 中国·山东 淄博 255000.

**【摘 要】**均匀设计由于考虑试验点的“均匀分散性”可以大幅度减少试验次数, 因此在多领域得到广泛应用。但常用的均匀设计软件分析方法需要借助 SPSS 进行多元变量人工设置, 然后通过线性回归构建模型方程; 再借助 Excel 的规划求解功能计算方程最优解。本文介绍一种通过 Design Expert 软件有效简化均匀设计分析的方法, 具体使用 Historical Data 程序的 Analysis、Optimization 模块实施。通过对比 SPSS、Design Expert 软件在回归模型选择、最优解确定、试验结果预测方面的特性, 验证 Design Expert 软件在均匀设计数据分析方面的精确度和高效性。

**【关键词】**均匀设计; 试验设计; Design Expert; SPSS; 方法优化

## Optimization of Design Expert Statistical Methods for Uniform Design

Meng Qinghong<sup>1</sup>, Li Canguo<sup>2</sup>, Wei Meng<sup>2</sup>, Kong Ling<sup>2\*</sup>, Chen Zhiwei<sup>3\*</sup>

1. Mengniu Dairy Tai'an Co., Ltd., Tai'an, Shandong 271000;

2. School of Life and Medicine, Shandong University of Technology, Zibo 255000, Shandong;

3. Institute of Food and Nutrition Sciences, Shandong University of Technology, Zibo, Shandong 255000.

**[Abstract]** Uniform design is widely used in many fields because it can greatly reduce the number of tests by considering the "uniform dispersion" of test points. However, the commonly used uniform design software analysis method needs to use SPSS to manually set the multivariate variables, and then construct the model equation through linear regression; and then use the Solver function of Excel to calculate the optimal solution of the equation. This paper introduces a method to effectively simplify uniform design analysis through Design Expert software, which is implemented using the Analysis and Optimization modules of the Historical Data program. By comparing the characteristics of SPSS and Design Expert software in regression model selection, optimal solution determination, and test result prediction, the accuracy and efficiency of Design Expert software in the analysis of uniform design data are verified.

**[Key words]** Uniform design; Experimental design; Design Expert; SPSS; Method optimization

**【基金项目】**中国学位与研究生教育学会研究课题 (C-2015Y0501-050); 山东省研究生教育创新计划项目 (4052/115027); 山东理工大学混合式教学课程建设项目 (4003/119008)。

均匀设计是由我国学者方开泰、王元等于 1978 年提出, 将数论原理和多元统计相结合的多因素多水平的试验设计方法<sup>[1]</sup>。由于该方法主要考虑试验点“均匀分散性”, 不考虑“整齐可比性”, 从而大幅度减少试验次数, 已在国民经济各领域取得了显著的经济和社会效益<sup>[2-4]</sup>。“均匀试验设计的理论、方法及其应用”于 2008 年获得国家自然科学奖二等奖。

均匀设计常用 SPSS、Excel 等软件进行多元线性回归分析<sup>[5-6]</sup>, 但计算过程复杂, 计算量大, 使得降低分析效率下降。本文通过案例, 介绍一种应用 Design Expert 软件的 Historical Data 程序 (具有 Design、Analysis、Optimization、Post Analysis 四个模块) 进行高效分析的方法, 并比对不同软件在回归模型选择, 最优解案确定, 试验结果预测和验证方面的特性, 旨在为处理各种有关均匀设计的资料提供参考。

### 1 均匀设计表

均匀设计表 (均匀表) 是均匀设计的基础。等水平均匀表可用  $U_n(r^1)$  表示<sup>[7]</sup>, 其中  $U$  表示均匀表代号,  $n$  表示均匀表行数 (即需要做的试验次数),  $r$  是因素水平数, 与  $n$  相等, 1 是均匀表列数, 表示最多可以排列 1 个因素。均匀设计表需要

根据其对应使用表的推荐, 选取适当的列数来安排因素<sup>[8]</sup>。使用表的最后一列  $D$  (均匀度偏差值) 越小, 表示均匀分散性越好, 因此方案设计应选择  $D$  最小的排列方式<sup>[9-11]</sup>。通常,  $U$  右上角加有“\*”号比不加“\*”的均匀表具有更小的  $D$  值, 应优先选用。

### 2 SPSS 数据分析

选取优化甜茶叶中总黄酮提取工艺的文献数据为例<sup>[12]</sup>, 以单因素试验中优化出的 4 个单因素最佳值, 酶添加量 ( $X_1$ ) 3.5 mg、酶解液 pH ( $X_2$ ) 5.0、酶解温度 ( $X_3$ ) 40 °C、酶解时间 ( $X_4$ ) 65 min 为基础, 设计 4 因素 10 水平的均匀试验优化组合因素。根据试验结果, 应用 Design Expert 软件优化筛选总黄酮提取得率 ( $Y$ ) 最大时的工艺条件。

#### 2.1 设计试验方案

根据 4 因素 10 水平的均匀设计要求, 选取均匀设计表  $U*10(108)$ , 4 因素依据均匀设计使用表的推荐, 选取该表的 1、3、4、5 列进行放置。将单因素最佳值设置为中间水平 (本案例中设置为水平 6), 分别以相同的梯度编号向上向下设置其他水平值。本试验中,  $X_1$ 、 $X_2$  的梯度编号均设置为 0.2,  $X_3$  的

梯度编号设置为2, X4的梯度编号设置为3, 详见表1。根据试验方案实施试验的结果, 添加在表1的Y列。

表1 纤维素酶-乙醇提取法4因素 $U \cdot 10(108)$ 试验设计方案及试验结果

序号	X1/mg	X2	X3/℃	X4/min	Y/%
1	1(2.5)	3(4.4)	4(36)	5(62)	9.2
2	2(2.7)	6(5.0)	8(44)	10(77)	9.32
3	3(2.9)	9(5.6)	1(30)	4(59)	10.48
4	4(3.1)	1(4.0)	5(38)	9(74)	9.19
5	5(3.3)	4(4.6)	9(46)	3(56)	9.15
6	6(3.5)	7(5.2)	2(32)	8(71)	10.64
7	7(3.7)	10(5.8)	6(40)	2(53)	10.31
8	8(3.9)	2(4.2)	10(48)	7(68)	9.21
9	9(4.1)	5(4.8)	3(34)	1(50)	9.82
10	10(4.3)	8(5.4)	7(42)	6(65)	10.89

## 2.2 构建回归方程和确定最优解

参考刘宇锋等人的文献<sup>[13]</sup>, 应用SPSS对上述案例进行分析。调用Regression子菜单的Linear功能, 通过Transform菜单的Compute Variable逐个设置原数据中不存在的各交互变量, 结果见图1、图2。模型 $R^2 = 0.978$ , 调整 $R^2 = 0.960$ , 方程见公式1。

$$Y = 7.19138 + 0.00931 \times X1X4 + 0.01245 \times X2X4 - 0.00061 \times X3^2 - 0.00055 \times X4^2 \quad \text{公式1}$$

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.989 <sup>a</sup>	.978	.960	1.3929

a. Predictors: (Constant), X4X4, X3X3, X1X4, X2X4

ANOVA <sup>a</sup>					
Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	4.248	4	1.062	54.741	.000 <sup>a</sup>
Residual	.097	5	.019		
Total	4.345	9			

a. Predictors: (Constant), X4X4, X3X3, X1X4, X2X4  
b. Dependent Variable: Y

图1 SPSS计算的R值和ANOVA

Coefficients <sup>a</sup>					
Model	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	Std. Error	Beta	t	
1 (Constant)	7.19138	.378		19.025	.000
X1X4	0.00931	.001	.546	6.888	.001
X2X4	0.01245	.001	.867	9.221	.000
X3X3	-0.00061	.000	-.414	-5.491	.003
X4X4	-0.00055	.000	-.921	-9.681	.000

a. Dependent Variable: Y

图2 SPSS计算的方程系数

为寻找丁黄酮最大提取率, 需应用Excel进行规划求解。根据试验因素范围, 整理出约束条件:  $2.5 \leq X1 \leq 4.3$ ;  $4 \leq X2 \leq 5.8$ ;  $30 \leq X3 \leq 48$ ;  $50 \leq X4 \leq 77$ ; Y目标规划最大值, 公式3规划求解, 最佳条件是 $X1 = 4.3$ (水平10)、 $X2 = 5.8$ (水平10)、 $X3 = 30$ (水平1)、 $X4 = 77$ (水平10), 相应 $Y = 12.0241$ 。

该方法需要借助SPSS进行多元变量人工设置, 并通过线性回归构建模型方程, 再借助Excel的规划求解功能计算方程最优解, 导致计算过程复杂, 计算量大(尤其是对于变量过多的数据)。

## 3 Design Expert 数据分析

### 3.1 录入试验数据

通过Design Expert 10.0软件RSM(Response Surface Method)的Historical Data程序(Design Expert 12.0的Historical Data程序可通过Custom Designs模块)对上述案例进行均匀设计数据分析。Numeric factors根据因素水平设置, Rows根据实际数据规模输入。变量名称和范围的调整可以在数据表生成前设定, 也可以在生成后设定, 但后者的设定需在数据录入前, 通过Design模块Design Tool工具框的Column Info Sheet选项, 先输入High对应的最大值, 再输入Low对应的最小值, 否则会出现“Minimum value for factor/component X(变量) must be less than the maximum.”的提示, 之后返回Design Tool的Design Layout选项, 录入试验数据。

本案例中选择4个数值自变量10行、1个因变量。调用RSM的Historical Data程序, Numeric factors设置为4, Rows设置为10, 调整变量名称和范围, 录入表1的数值, 见图3。

Run	Factor 1 A:X1	Factor 2 B:X2	Factor 3 C:X3	Factor 4 D:X4	Response 1 Y
1	2.5	4.4	36	62	9.2
2	2.7	5	44	77	9.32
3	2.9	5.6	30	59	10.48
4	3.1	4	38	74	9.19
5	3.3	4.6	46	56	9.15
6	3.5	5.2	32	71	10.64
7	3.7	5.8	40	53	10.31
8	3.9	4.2	48	68	9.21
9	4.1	4.8	34	50	9.82
10	4.3	5.4	42	65	10.89

图3 Historical Data试验方案及试验结果

### 3.2 构建回归方程

在Historical Data程序的Analysis模块中, 分别调用Transform、Fit Summary、Model、ANOVA等子模块进行分析。调用Analysis模块, Fit Summary子模块给出Linear优选, 2FI备选建议, Model采用默认Process order选项Linear。ANOVA子模块分析结果见图4,  $F = 19.34$ ,  $R^2 = 0.9393$ , 调整 $R^2 = 0.8907$ 。

Model的Process order尝试选择2FI, 软件自动生成交互变量, ANOVA子模块分析结果中,  $R^2=0.9965$ , 大于Linear模型下的 $R^2=0.9393$ , 且展现出由于试验方案设计偏差而造成模型中出现多个Sum of Squares为0的变量, 去除该类变量后, F、R2、调整R2的数值均未改变, 说明含有交互变量的2FI模型可以更好地描述试验结果。(见图4)

在Model选项中依次从高到低去除p-value最高且不显著的

Use your mouse to right click on individual cells for definitions.

ANOVA for Response Surface Linear model					
Analysis of variance table [Partial sum of squares - Type III]					
Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	4.08	4	1.02	19.34	0.0030 significant
A-X1	0.83	1	0.83	15.76	0.0106
B-X2	1.63	1	1.63	30.82	0.0026
C-X3	0.56	1	0.56	10.64	0.0224
D-X4	0.29	1	0.29	5.48	0.0663
Residual	0.26	5	0.053		
Cor Total	4.35	9			

图4 ANOVA初步分析结果

变量, ANOVA 优化结果见图 5,  $F = 49.25$ ,  $R^2 = 0.9840$ , 调整  $R^2 = 0.9640$ , 实际水平方程见公式 2。

$$Y = 11.43503 - 1.48457 \times X1 + 0.79375 \times X2 - 0.046648 \times X3 - 0.087570 \times X4 + 0.032318 \times X1X4$$

公式 2

Use your mouse to right click on individual cells for definitions.

Response 1 Y

ANOVA for Response Surface Reduced 2FI model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Model	4.28	5	0.86	49.25	0.0011	significant
A-X1	0.83	1	0.83	47.89	0.0023	
B-X2	1.63	1	1.63	93.66	0.0006	
C-X3	0.56	1	0.56	32.35	0.0047	
D-X4	0.29	1	0.29	16.65	0.0151	
AD	0.19	1	0.19	11.19	0.0287	
Residual	0.069	4	0.017			
Cor Total	4.35	9				

图5 优化后ANOVA分析结果

### 3.3 确定最优解

Optimization可通过Solutions选项筛选出优化的工艺条件。为寻找丁黄酮最大提取率,调用Optimization模块Numerical的Criteria选项窗,设置Y的目标值(Goal)为maximize。运行Solution,选项窗共列出100个Desirability = 1.000的结果,其中当 $X1 = 4.230$ 、 $X2 = 5.759$ 、 $X3 = 32.443$ 、 $X4 = 71.166$ 时, $Y = 11.70$ 最大,暂时作为优化结果。

Graphs选项窗可以显示上述优化结果的所有Factors,见图6。在Graphs选项窗Factors Tool工具框中,鼠标移动各变量的红色指针进行数值改变时,发现界面左上角显示的Y值仍有增大的趋势,表明数据还有进一步挖掘的空间。这是由于目前因变量的取值范围是根据实际水平设定,而依据各因素水平筛选出的最优组合的预测值可能会超过该取值范围,所以可以扩大因变量目标值设定的上下限范围。本案例中,将因变量Upper的设定在默认值的基础上予以扩大(如扩大1倍),设置为21.78,此时Solution选项窗推荐的优化结果为 $X1 = 4.3$ 、 $X2 = 5.8$ 、 $X3 = 30$ 、 $X4 = 77$ , $Y = 12.21$ ,与原文献的优化结果基本一致。说明在合理优化Y值的取值范围后,可以获得更多的试验结果,预示着同类因素在不同组合下的潜在能力。(见图6)

### 4 SPSS、Design Expert 比对

对于总黄酮提取工艺案例,应用SPSS软件对Design Expert回归方程(公式2)的5个变量 $X1$ 、 $X2$ 、 $X3$ 、 $X4$ 、 $X1X4$ 进行线性

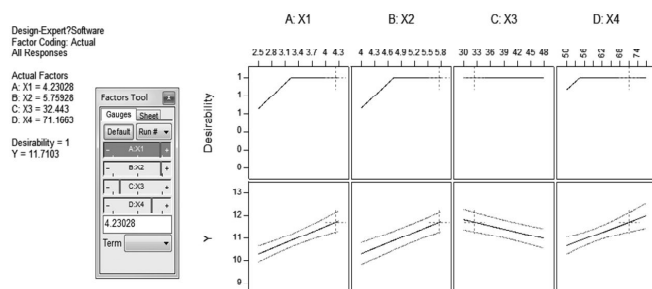


图6 All Factors

回归模型验证,结果见图7、图8。模型中 $X1$ 、 $X4$ 不显著,模型  $R^2 = 0.984$ , 调整  $R^2 = 0.964$ , 与Design Expert 方差一致,方程系数与公式2 四舍五入后一致。

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.992 <sup>a</sup>	.984	.964	.13178

a. Predictors: (Constant), X1X4, X2, X3, X4, X1

图7 SPSS 验证公式2的R值

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.435	2.222		5.146	.007
	X1	-1.485	.619	-.1294	-2.399	.074
	X2	.794	.082	.692	9.678	.001
	X3	-.047	.008	-.407	-5.687	.005
	X4	-.088	.033	-.1145	-2.630	.058
	X1X4	.032	.010	.1896	3.346	.029

a. Dependent Variable: Y

图8 SPSS 验证公式2的方程

依次去除不显著变量后,结果见图9、图10,  $R^2 = 0.975$ , 调整  $R^2 = 0.927$ , 方程系数均显著,数值与公式2 不一致。说明Design Expert 优化的模型方程未能通过SPSS验证。

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.975 <sup>a</sup>	.951	.927	.18752

a. Predictors: (Constant), X1X4, X2, X3

图9 SPSS 优化后验证公式2的R值

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.752	.809		7.113	.000
	X2	.832	.110	.725	7.586	.000
	X3	-.047	.012	-.407	-4.016	.007
	X1X4	.008	.002	.497	5.187	.002

a. Dependent Variable: Y

图10 SPSS优化后验证公式2的方程

应用Design Expert 软件验证SPSS模型中的4个变量 $X1X4$ 、 $X2X4$ 、 $X3X3$ 、 $X4X4$ ,结果见图11,  $R^2 = 0.0572$ , 调整  $R^2 = -0.6970$ , 方程系数均不显著,数值与公式2 不一致。说明SPSS优化的模型方程未能通过Design Expert 验证。(见图11)

为进一步比较统计软件在回归模型选择,最优解确定,试

Use your mouse to right click on individual cells for definitions.

Response 1 Y

ANOVA for Response Surface Reduced Quadratic model

Analysis of variance table [Partial sum of squares - Type III]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F
Model	0.25	4	0.062	0.076	0.9865 not significant
AD	3.706E-003	1	3.706E-003	4.523E-003	0.9490
BD	0.010	1	0.010	0.013	0.9144
C <sup>2</sup>	0.026	1	0.026	0.031	0.8662
D <sup>2</sup>	1.949E-003	1	1.949E-003	2.379E-003	0.9630
Residual	4.10	5	0.82		
Cor Total	4.35	9			

图11 Design Expert验证公式1



表2 试验结果与Design Expert (DE)、SPSS 预测结果比较

试验号	X1	X2	X3	X4	Y	公式1 (SPSS)		公式2 (DE)	
						预测结果	误差	预测结果	误差
1	2.5	4.4	36	62	9.2	9.126	-0.80%	9.1167	-0.91%
2	2.7	5	44	77	9.32	9.4783	1.70%	9.319	-0.01%
3	2.9	5.6	30	59	10.48	10.4343	-0.44%	10.5383	0.56%
4	3.1	4	38	74	9.19	9.1197	-0.77%	9.1688	-0.23%
5	3.3	4.6	46	56	9.15	9.1034	-0.51%	9.1098	-0.44%
6	3.5	5.2	32	71	10.64	10.7043	0.60%	10.6874	0.45%
7	3.7	5.8	40	53	10.31	10.3233	0.13%	10.3763	0.64%
8	3.9	4.2	48	68	9.21	9.2675	0.62%	9.3558	1.58%
9	4.1	4.8	34	50	9.82	10.0078	1.91%	9.819	-0.01%
10	4.3	5.4	42	65	10.89	10.7637	-1.16%	10.7192	-1.57%

验结果预测上的特性,在表1的基础上增加2个方程的预测结果及误差(误差=(预测结果-试验结果)/试验结果),见表2。

通过表2 预测结果比对10个试验点,检索公式1、公式2的预测误差都在 $\pm 2\%$ 范围内。通过Excel 规划求解进一步验证两个公式,已知公式1 规划求解的最佳条件是 $X1 = 4.3$ (水平10)、 $X2 = 5.8$ (水平10)、 $X3 = 30$ (水平1)、 $X4 = 77$ (水平10),相应 $Y = 12.0241$ ;公式2 规划求解,最佳条件是 $X1 = 4.3$ (水平10)、 $X2 = 5.8$ (水平10)、 $X3 = 30$ (水平1)、 $X4 = 77$ (水平10),相应 $Y = 12.2133$ 。2个公式规划求解的最佳水平相同、Y值相近。说明在本案例中,两种软件构建的模型均具有较好地可靠性和良好的回代效果,但对于最大产量Y的取值,由于两种软件选用的方程不同,预测值略有偏差。所以为了更好地描述SPSS、Design Expert 在均匀设计数据分析上的特性,还需更多案例探讨。

## 5 总结

通过案例,证实应用Design Expert 软件的Historical Data 程序可以合理、高效地进行均匀设计的数据分析。通过Analysis 模块自动计算交互变量,直接获取回归方程;利用Optimization 模块Graphs 的All Factors 配合Solution,可以避免Excel 规划求解,快速获取最优解。因此,应用Design Expert 可以有效简化分析步骤,提高均匀设计数据分析的效率。

通过比对Design Expert、SPSS 软件发现,两种软件构建的模型方程不同,Design Expert 部分可通过SPSS 相互验证;二者的方程与试验结果均具有较好的拟合度,但SPSS 仅在一定范围内有较好的回代效果。所以相比于SPSS,Design Expert 具有更高的精确度和分析效率。因此,应用Design Expert 进行均匀设计数据分析具有普及推广的价值。

## 参考文献:

- [1]张国秋,王文璇.均匀试验设计方法应用综述[J].数理统计与管理,2013,32(1):90-99.
- [2]Huihui Li,Lifeng Li,Wenpeng Wu,Liang Xu.Seismic fragility assessment framework for highway bridges based

on an improved uniform design-response surface model methodology[J].Bulletin of Earthquake Engineering,2020,8(5):2329-2353.

[3]Zhang Yongping,Wang Zuhua,Xu Jian,Yang Fangfang,Dai Chuanyang,Xie Weijie,Liang Zhu,Su Songbo.Optimization of the extraction and purification of Corydalis yanhusuo W.T.Wang based on the Q-marker uniform design

method[J].BMC chemistry,2020,14(1):9.

[4]张浩,刘秀玉,宗志芳.基于均匀设计和多元非线性回归法优化制备改性钠基蒙脱土[J].过程工程学报,2017,17(1):188-192.

[5]张学彬,陈孟涛,杨宇奇,庞永峰,孙丽,郑辉杰.均匀设计法优化超临界CO<sub>2</sub>萃取丁香精油工艺研究[J].食品研究与开发,2017,38(17):50-54+84.

[6]汪文俊,周永道,方开泰.在计算机模拟试验中不同试验设计的比较[J].数理统计与管理,2011,30(2):218-226.

[7]王元,方开泰.关于均匀分布与试验设计(数论方法)[J].科学通报,1981,26(2):65-70.

[8]王元,方开泰.Uniform Design of Experiments with Mixtures[J].Science in China (A),1996,44(3):345-350.

[9]Xueru Zhang,Minqian Liu,Yongdao Zhou.Orthogonal uniform composite designs[J].Journal of Statistical Planning and Inference,2020,206:100-110.

[10]刘华,万建平.均匀试验设计的方法与应用[J].阜阳师范学院学报(自然科学版),2003,20(1):12-16.

[11]范步高.均匀设计线性回归分析的Excel模型构建与应用[J].中国医药工业杂志,2013,44(5):515-522.

[12]王小明,陈碧,钟翠娟,覃逸明,廖政达,谢济运.均匀设计法优化甜茶叶中总黄酮提取工艺[J].食品研究与开发,2018,39(4):56-62.

[13]刘宇锋,苏天明,苏利荣,秦芳,李琴,梁芷姮.均匀设计试验多元回归分析方法在新型生物农药研制中的应用[J].农药,2018,57(2):105-110.

## 作者简介:

孟庆红(1975.9-),女,蒙古族,内蒙古通辽人,硕士,研究方向食品检验。

★通讯作者:孔玲(1974.09-),女,汉族,山东定陶人,博士,副教授,研究方向为生物分析化学。

★通讯作者:陈志伟(1972.06-),男,汉族,内蒙古呼伦贝尔人,博导,教授,研究方向为应用统计学。