应用SPSS配合Design Expert提高正交试验设计和数据挖掘的效率

魏　萌[1]　李灿国[1]　孔　玲[1*]　陈志伟[2*]
1.山东理工大学生命与医药学院，中国·山东　淄博　255000
2.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】正交设计可以通过较少试验次数达到全面试验等效的结果，因此在科研、工程诸多领域受到广泛关注。然而，常用软件通常不能够兼顾设计方案灵活性和试验数据深度挖掘高效的特性。通过案例，比对Design Expert的Taguchi OA程序与SPSS的Orthogonal design程序，展现SPSS在方案设计上的灵活性；比对SPSS的General Linear Model程序与Design Expert的Historical Data程序，展现Design Expert在结果分析和数据挖掘上的高效率。因此利用SPSS的Orthogonal design进行方案设计，通过Design Expert的Historical Data进行结果分析和数据挖掘，可以有效提高正交试验的灵活性和效率。

【关键词】正交设计；数据挖掘；Design Expert；SPSS；效率
【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；山东理工大学混合式教学课程建设项目；山东理工大学一流本科课程建设与培育项目（JX20200076）。

引言

正交设计（Orthogonal design）是研究多因素多水平的一种设计方法。可以根据因子设计的分式原理，采用由组合理论推导而成的正交表来安排设计试验，并对结果进行统计分析。1951年由日本统计学家田口玄一根据试验的优化规律提出的正交表，是正交试验设计的基本工具，使正交试验具备分散性和整齐可比性[1]。

正交设计相关应用软件数目较多，如SPSS、SAS、JMP等[2]。然而，常用软件通常不兼顾设计方案灵活性和试验数据深度挖掘高效的特性，因此根据各软件的特点，合理搭配、应用不同软件，共同进行正交试验，对提高正交试验效率有着重要意义。Design Expert软件结果分析方便、快捷，数据挖掘灵活、精确；SPSS软件方案设计灵活，对多因素不同水平（特别是未查到正交表）的方案设计尤为适用。为此，通过比对Design Expert、SPSS软件在正交方案设计、结果分析和数据挖掘上的特性，提出一种应用SPSS配合Design Expert的方式，以此提高正交试验设计和数据挖掘的灵活性和效率。

1　正交试验表

正交表是正交试验设计的基础，可以手动查找，也可以应用软件生成。Design Expert、SPSS等软件可自动生成正交表，有利于提高试验设计的效率。

正交表将各试验因素、各水平间的组合均匀搭配，合理安排，将试验因素各水平平均分布，实现了因素和水平的均匀分散性和整齐可比性，极大地减少了试验次数。因此，如何保证正交试验点的均匀分散性和整齐可比性成了正交设计的关键，这就涉及到正交表的设计和选取[3-6]。正交表常表示为Ln（tq）。L是Latin的第一个字母，n为试验次数，t为水平数，q为因素数。如L9（34）表示共需做9次试验，至多可以安排4个因素，每个因素为3水平，这是标准型正交表。如正交表中各列水平数不等，则称为混合型正交表，通常表示为Ln（t1q1×t2q2），如L8（4×24）表示第1列为4水平，其它均为2水平，共需做8次试验。

2　Design Expert正交试验设计及分析思路

2.1选择建立新的试验方案

Design Expert具有Taguchi OA程序，其中Design Designation下拉菜单可提供包括L4（23）、L8（27）、L9（34）.....L64（263）、L64（421）19种正交表组成方案，最多提供64组21因素4水平或63因素2水平的设计方案。界面底端的Runs可显示所选方案的运行次数[7]。选择方案后，设置自变量、因变量。

对于正交试验设计可以直接从19种正交表中选择方案。优先选择水平数，因素数量采用就近向上的原则选择，如3因素3水平的正交设计，在没有现成方案的情况下，可以选择4因素3水平的L9（34）。

2.2正交试验分析

Taguchi OA程序具有Design、Analysis、Optimization、Post Analysis四个模块，分析主要使用Analysis模块，主要包括Transform、Effect、ANOVA、Model Graphs等子模块。调用Analysis模块，Transform采用默认值。Effect子模块Selection Tool工具框有Pareto Chart和Numeric 2个选项，可以进行不同路径的结果分析[8]。一种是调用Numeric选项，通过Order选择Design Model或者选择最多交互4FI；另一种是调用Pareto Chart选项，Rank从高到低依次选择，根据t-Value值考察最大变量是否对模型有显著影响，也可以根据变量在Pareto Chart中的位置，对各变量作用初步排序。运行ANOVA模块，根据各变量p-value由高到低（Mean Square由低到高），依次回到Numeric或者Pareto Chart去掉相关变量，直到各变量的p-value达到显著水平，找出主要因素[9-14]。

对于模型回归方程、方程系数的显著水平及以及方程相关系数等可以在ANOVA子模块选项中查阅。对于编码变量的详细的方

程系数及系数显著水平也可以通过Post Analysis模块的Coefficients Table查阅。

Design Expert在Analysis模块Model Graphs子模块中通过Graphs Tool、Factors Tool工具框提供7种图形，可以展示变量水平对因变量的影响。All Factors可以展示全部因素在不同水平下对因变量的影响，使得正交分析最优水平组合的描述更加直观；3D Surface可以比较2个不同水平的自变量对因变量的直观影响。

3 单因变量的正交设计及分析

选取鳜鱼发酵工艺对感官评分影响的文献数据为例[15]，利用Design Expert进行单因变量的正交设计及分析。鳜鱼发酵工艺条件为发酵时间、温度、盐度，设计3因素4水平正交试验，以感官评分（评分与品质正相关）为衡量指标优选最优理论发酵条件。

3.1 试验方案设计、实施试验、实验结果汇总

Select	Std	Run	Factor 1 A:时间 d	Factor 2 B:温度 ℃	Factor 3 C:盐度 %	Factor 4 D:D	Factor 5 E:E	Response 1 感官评分
	1	9	Level 1 of A	Level 1 of B	Level 1 of C	Level 1 of D	Level 1 of E	69.84
	2	10	Level 1 of A	Level 2 of B	Level 2 of C	Level 2 of D	Level 2 of E	78.99
	3	2	Level 1 of A	Level 3 of B	Level 3 of C	Level 3 of D	Level 3 of E	82.54
	4	5	Level 1 of A	Level 4 of B	Level 4 of C	Level 4 of D	Level 4 of E	69.35
	5	12	Level 2 of A	Level 1 of B	Level 2 of C	Level 3 of D	Level 4 of E	87.26
	6	15	Level 2 of A	Level 2 of B	Level 1 of C	Level 4 of D	Level 3 of E	72.3
	7	16	Level 2 of A	Level 3 of B	Level 4 of C	Level 1 of D	Level 2 of E	80.94
	8	1	Level 2 of A	Level 4 of B	Level 3 of C	Level 2 of D	Level 1 of E	63.86
	9	6	Level 3 of A	Level 1 of B	Level 3 of C	Level 4 of D	Level 2 of E	79.79
	10	8	Level 3 of A	Level 2 of B	Level 4 of C	Level 3 of D	Level 1 of E	80.48
	11	7	Level 3 of A	Level 3 of B	Level 1 of C	Level 2 of D	Level 4 of E	73.45
	12	3	Level 3 of A	Level 4 of B	Level 2 of C	Level 1 of D	Level 3 of E	69.38
	13	14	Level 4 of A	Level 1 of B	Level 4 of C	Level 2 of D	Level 3 of E	73.74
	14	4	Level 4 of A	Level 2 of B	Level 3 of C	Level 1 of D	Level 4 of E	74.4
	15	11	Level 4 of A	Level 3 of B	Level 2 of C	Level 4 of D	Level 1 of E	78.93
	16	13	Level 4 of A	Level 4 of B	Level 1 of C	Level 3 of D	Level 2 of E	57.06

图1 Design Expert正交试验方案及试验结果

使用因子分析的Taguchi OA选择试验方案，由于Design Designation下拉菜单中没有3因素4水平的试验方案选择，所以按照向上相近的原则，选择5因素4水平的试验方案L16(45)，因变量1项，试验方案如图1。实施试验，填写试验结果。

3.2 试验结果方差分析及回归方程

调用Analysis模块，Transform采用默认值，通过Numeric选项对感官评分应用ANOVA子模块分析。

	Term	df	Sum of Squares	Mean Square	F Value	Prob > F
	Intercept					
M	A-时间	3	66.55	22.18		
M	B-温度	3	503.99	168.00		
M	C-盐度	3	241.48	80.49		
e	D-D	3	42.01	14.00		
e	E-E	3	16.78	5.59		
~	AB		Aliased			
~	AC		Aliased			
~	AD		Aliased			
~	AE		Aliased			
~	BC		Aliased			
~	BD		Aliased			
~	BE		Aliased			
~	CD		Aliased			
~	CE		Aliased			
~	DE		Aliased			
~	ABC		Aliased			

图2 Numeric Plot因子分布

在Numeric图中，Order选择Design Model或者选择最多交互4FI。运行ANOVA子模块，由于无法估计误差，返回Numeric选项去掉高阶项D，再次运行ANOVA子模块。根据各变量p-value由高到低，依次回到Numeric去掉相关变量，直到各变量的p-value达到显著水平，如图2。

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

Response 1 感官评分

ANOVA for selected factorial model

Analysis of variance table [Classical sum of squares - Type II]

Source	Sum of Squares	df	Mean Square	F Value	p-value Prob > F	
Model	812.02	9	90.22	9.21	0.0069	significant
A-时间	*66.55*	*3*	*22.18*	*2.26*	*0.1813*	
B-温度	*503.99*	*3*	*168.00*	*17.15*	*0.0024*	
C-盐度	*241.48*	*3*	*80.49*	*8.22*	*0.0152*	
Residual	58.78	6	9.80			
Cor Total	870.80	15				

图3 ANOVA初步分析结果

运行ANOVA子模块，结果如图3。根据p-value对3个主要因素排序B>C>A，其中A-时间的p-value=0.1813不显著。此时的R2=0.9325。

ANOVA子模块编码水平方程见公式(1)。更详细的参数可以通过Post Analysis模块的Coefficient Table子模块获得。用编码因子表示的方程可以用来预测每个因子在给定水平下的响应。默认情况下，高级别的因子编码为+1，低级别的因子编码为-1。该编码方程可用于通过比较因子系数来识别因子的相对影响。

感官评分=74.52+0.66×A[1]+1.57×A[2]+1.26×A[3]+3.14×B[1]+2.02×B[2]+4.45×B[3]-6.36×C[1]+4.12×C[2]+0.63×C[3]公式(1)

3.3 Model Graphs图形分析

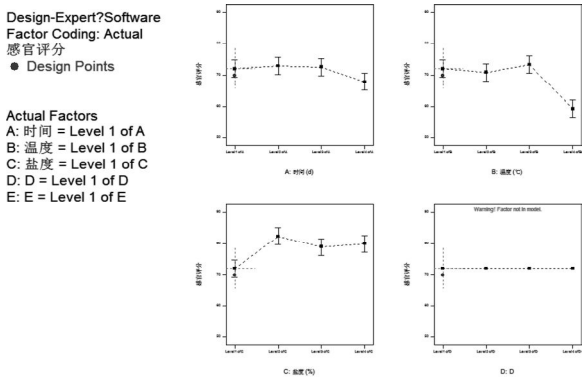

图4 All Factors图

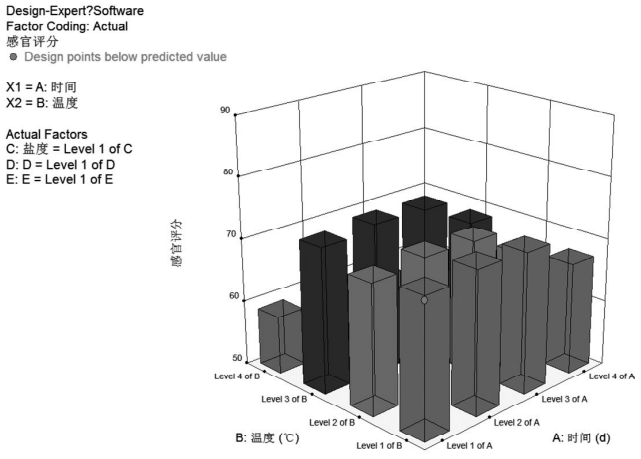

图5 自变量、因变量3D Surface

在Analysis模块Model Graphs子模块中可以通过Graphs Tool、Factors Tool工具框查看All Factors，如图4。图中展示全部因素在不同水平下对因变量的影响，可以清晰看出BC的最佳组合为B3C2，但对于A则不容易直接比较出A2、A3感官评分高低，表明还需进一步数据挖掘。自变量、因变量的3D Surface如图5，可以通过鼠标调整视角以便看到更多组合效果，也可以通过调整坐标比较出BC两个主要因素组合对感官评分的影响。

4 多因变量的正交设计及结果优化

选取研究影响透明果汁的产品质量的文献数据为例[16]，利用Design Expert进行多因变量的正交试验设计及分析。透明果汁产品质量受到填充液（X1）、加糖量（X2）、原果汁量（X3）、均质（X4）、增稠剂（X5）、冷却方法（X6）、灌装（X7）7个因素影响，每个因素2水平，选取香气和色泽两个重要指标进行考察。香气评价标准分为10个等级，最好的记为10，最差的记为1。本次试验共有7个因素，每个因素为2水平，选用L8（27）正交表来安排试验。

4.1 试验方案设计、实施试验、实验结果汇总

使用Design Expert因子分析模块的Taguchi OA设计试验方案，选择7因素2水平的试验方案L8（27），因变量2项，如图6。

4.2 试验结果方差分析及回归方程

调用Analysis模块，Transform采用默认值，通过Pareto Char选项分别对色泽和香气变量应用ANOVA子模块分析。

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

Response 1 色泽

ANOVA for selected factorial 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	1.21	7	0.17		
A-填充液	*0.14*	*1*	*0.14*		
B-加糖量	*0.14*	*1*	*0.14*		
C-原果汁量	*0.43*	*1*	*0.43*		
D-均质	*0.090*	*1*	*0.090*		
E-增稠剂	*0.17*	*1*	*0.17*		
F-冷却方法	*0.090*	*1*	*0.090*		
G-灌装	*0.17*	*1*	*0.17*		
Pure Error	0.000	0			
Cor Total	1.21	7			

图8 色泽ANOVA初步分析

Select	Std	Run	Factor 1 A:填充液	Factor 2 B:加糖量	Factor 3 C:原果汁量	Factor 4 D:均质	Factor 5 E:增稠剂	Factor 6 F:冷却方法	Factor 7 G:灌装	Response 1 色泽	Response 2 香气
	1	6	Level 1 of A	Level 1 of B	Level 1 of C	Level 1 of D	Level 1 of E	Level 1 of F	Level 1 of G	2.55	2
	2	7	Level 1 of A	Level 1 of B	Level 1 of C	Level 2 of D	Level 2 of E	Level 2 of F	Level 2 of G	2.7	4
	3	4	Level 1 of A	Level 2 of B	Level 2 of C	Level 1 of D	Level 1 of E	Level 2 of F	Level 2 of G	1.9	6
	4	8	Level 1 of A	Level 2 of B	Level 2 of C	Level 2 of D	Level 2 of E	Level 1 of F	Level 1 of G	1.9	8
	5	5	Level 2 of A	Level 1 of B	Level 2 of C	Level 1 of D	Level 2 of E	Level 1 of F	Level 2 of G	2.4	8
	6	1	Level 2 of A	Level 1 of B	Level 2 of C	Level 2 of D	Level 1 of E	Level 2 of F	Level 1 of G	1.4	6
	7	2	Level 2 of A	Level 2 of B	Level 1 of C	Level 1 of D	Level 2 of E	Level 2 of F	Level 1 of G	2.1	10
	8	3	Level 2 of A	Level 2 of B	Level 1 of C	Level 2 of D	Level 1 of E	Level 1 of F	Level 2 of G	2.1	10

图6 正交试验方案及试验结果

对于色泽变量，通过Effect子模块Pareto选项使用鼠标从左到右选择，如图7。选择之后显示变量名称，连续选择变量后最大值仍未超过t-Value值，说明最大变量未对模型有显著影响。通过连续选择可以看出各因素按Rank排序的结果是C>G>E>B>A>D>F。

因素全部选入模型后的ANOVA子模块结果如图8，获得Mean Square

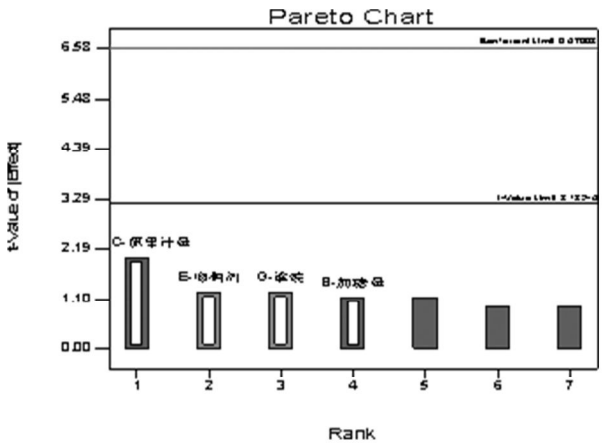

图7 色泽Pareto Chart因子Rank顺序

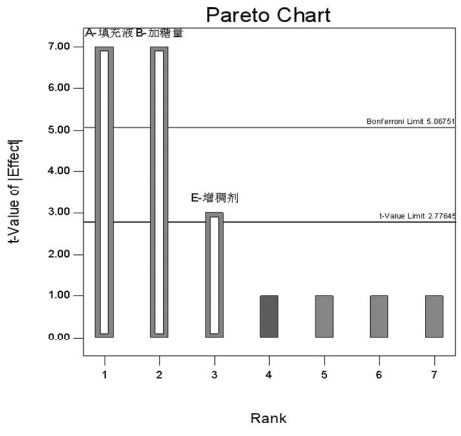

图9 香气Pareto Chart因子Rank顺序

的信息，而p-value空白。返回Effect子模块Pareto选项去掉影响最小的变量（Mean Square最小值），依次去掉每次ANOVA子模块运行p-value最大值，仍然没有找到显著因素。由于Effect子模块Pareto直观显示结果与ANOVA子模块运行结果相印证，所以各因素排序为C>G>E>B>A>D>F。

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

Response 2 香气

ANOVA for selected factorial 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	53.50	3	17.83	35.67	0.0024	significant
A-填充液	24.50	1	24.50	49.00	0.0022	
B-加糖量	24.50	1	24.50	49.00	0.0022	
E-增稠剂	4.50	1	4.50	9.00	0.0399	
Residual	2.00	4	0.50			
Cor Total	55.50	7				

图10 香气ANOVA初步分析结果

对于香气变量，通过Effect子模块Pareto选项使用鼠标从左到右选择，连续选择变量后最大值仍超过t-Value值，说明最大变量对模型有显著影响。通过连续选择可以看出各因素按Rank排序的结果是A>B>E>F>C>G>D。运行ANOVA子模块，返回Effect子模块Pareto选项依次去掉p-value最大值，获得3个主要因素A、B、E，其中A、B差异极显著，E差异显著，如图9和图10。编码因素方程见公式(2)，R2=0.9640。

香气=6.75+1.75×A+1.75×B+0.75×E 公式(2)

因此，影响香气的主要因素是ABE；色泽没有显著的影响因素，各因素排序为C>G>E>B>A>D>F。

5 小结

Design Expert的Taguchi OA程序可以实现对单因变量、多因变量的正交试验的设计和分析。根据实际情况选择合适水平、因素的试验方案，实施试验，填写试验结果；再通过Analysis模块的Effect子模块Numeric或Pareto程序对试验结果进行方差分析并得到回归方程。通过两例案例发现，Design Expert不仅方便试验方案的选择，且在结果分析上方式灵活多样，运行简单快捷，可以较好地提高正交试验的效率。

Design Expert的Taguchi OA程序虽然在一定程度上能够提高正交试验的效率，但仍有不足。一方面，Design Expert对于19种方案之外的试验方案尚不能设计，故如何提升方案设计的灵活性，将成为进一步提高正交试验效率的关键。另一方面，对于已有交叉项的正交试验结果的分析和挖掘，Taguchi OA程序如何解决，还需要进一步探讨。

参考文献:

[1]刘瑞江,张业旺,闻崇炜,等.正交试验设计和分析方法研究[J].实验技术与管理,2010,27(9):52-55.

[2]史周华.析因设计的SAS实现[J].数理医药学杂志,2005,18(6):604-605.

[3]杜家菊,陈志伟.使用SPSS线性回归实现通径分析的方法[J].生物学通报,2010,45(2):4-6.

[4]贾小波,李道强,涂庆.基于正交表的二次统计法在正交试验中的应用[J].统计与决策,2008(24):150-151.

[5]蔡火娣,邓林海.二次回归分析在正交设计中的应用[J].统计与决策,2007(5):25-26.

[6]程敬丽,郑敏,楼建晴.常见的试验优化设计方法对比[J].实验室研究与探索,2012,37(7):7-11.

[7]方开泰,马长兴.正交与均匀试验设计[M].北京:科学出版社,2001,40-56.

[8]葛宜元.试验设计方法与Design-Expert软件应用[M].哈尔滨:哈尔滨工业大学出版社,2015,106-111,120-127.

[9]储成顶,梁伟,赵国琴,等.建立分析测试方法的正交试验与电脑编程[J].实验室研究与探索,2012,31(9):36-39.

[10]曾尊祥.正交试验优化石墨炉原子吸收测定珠江水中Mn[J].实验室研究与探索,2011,30(11):224-226+305.

[11]张秀英,冯亚云,冯朝任,等.正交试验设计法在恒压过滤实验中的应用[J].实验技术与管理,1993,10(3):52-55.

[12]李建鹏,陶进转,陈冰.蔗糖酶水解蔗糖的正交试验与SPSS分析[J].化学研究与应用,2019,31(10):1807-1811.

[13]邓波,王莎,吴汉奇.正交试验设计在茶叶铁、锌、铜含量测定中的应用[J].食品研究与开发,2020,41(1):167-171.

[14]肖建昆,夏兆旺,卢仙兰.发动机排气噪声影响因素正交实验[J].实验室研究与探索,2014,33(4):29-32.

[15]杨召侠,刘洒洒,高宁,等.臭鳜鱼发酵工艺优化及挥发性风味物质分析[J].中国食品学报,2019,19(5):253-262.

[16]肖怀秋,刘洪波.试验数据处理与试验设计方法[M].北京:化学工业出版社,2013,70-89.

作者简介:

魏 萌（1999.02-），女，山东枣庄人，本科，研究方向：应用统计学。

李国灿（1995.03-），男，山东泰安人，硕士，研究方向：生物分析化学。

通讯作者:

*孔 玲（1974.09-），女，山东定陶人，博士，副教授，研究方向：生物分析化学、应用统计学。

*陈志伟（1972.06-），男，内蒙古呼伦贝尔人，博导，教授，研究方向：应用统计学。