

doi: 10.13241/j.cnki.pmb.2015.19.041

生物工程专业“试验设计与数据处理”教学中常用软件的探讨 *

何 环 黄冠华 冷云伟 刘海臣 朱红威

(中国矿业大学化工学院生物工程系 江苏 徐州 221116)

摘要:试验设计与数据处理在生物工程专业方向有着重要的应用,但是在本科教学中开设该课程的高校较少。本文结合生物工程中的常见案例,对该课程教学过程中常用的软件包括 Excel、正交试验设计助手、Data processing system (DPS)和 Design expert 在单因素、多因素、正交、均匀、P-B(Plackett-Burman)和响应面等数据处理方法进行了探讨。通过比较分析各种软件和数据处理方法之间特点,增加该课程教学的实践性和应用性,从而提高教学质量。

关键词:生物工程; 试验设计与数据处理; 正交设计; 均匀设计; 响应面设计

中图分类号:G642.0 **文献标识码:**A **文章编号:**1673-6273(2015)19-3752-03

The Discussion on Conventional Software during the Teaching of “Experiment design and data processing” in Bioengineering Major*

HE Huan, HUANG Guan-hua, LENG Yun-wei, LIU Hai-chen, ZHU Hong-wei

(Bioengineering department, School of Chemical Engineering and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, 221116, China)

ABSTRACT: The application of experiment design and data processing plays a very import role in bioengineering field. However, only a few schools open it in their undergraduate course. In present work, the conventional cases in bioengineering were involved to introduce the familiar software of experiment design (including Excel, Orthogonal experiment helper, DPS and Design expert) and the major data processing method (including single factor, multifactor, Orthogonal, uniformity, P-B, response surface et al.) were discussed. By comparison the difference among the features of these software and data processing method, we hope to improve the teaching quality by increasing its practice and appliance.

Key words: Bioengineering; Experiment design and data processing; Orthogonal design; Uniform distribution design; Response surface design

Chinese Library Classification(CLC): G642.0 **Document code:** A

Article ID: 1673-6273(2015)19-3752-03

前言

在生产和研究过程中经常碰到需要用较少的试验次数,在较短的时间达到预期目标的问题,这就必须对试验过程进行科学的设计。《试验设计与数据处理》是一门专门介绍工程技术和科研中常用的试验设计与数据处理方法的课程^[1]。生物工程领域中也会经常运用到该课程中的知识来优化控制过程的问题,如菌种选育条件、发酵工艺控制、目标产物分离等。目前全国开设生物工程专业的本科院校有 225 所,但是在专业课程中开设《试验设计与数据处理》的相对较少。随着生物工程学科发展的逐渐成熟,有必要推广和普及该课程的教学。由于该课程的理论性较强,只有将授课内容与实际操作结合,给学生提供动手解决实际问题机会,才能既充分调动学生的积极性,同时又能增加学生对课程内容的掌握程度^[2]。本文对生物工程领域常用的试验设计和数据处理软件进行介绍,希望为同行进行该科目教学时提供参考,也为将来从事该行业工作的人员提供帮助。

1 Excel 与单因素试验设计和方差分析

试验中用来表示试验结果特性的值称为指标,而将影响试验指标的条件称为因素,因素的不同状态称为水平^[3]。单因素试验是指只有一个因素(或仅考察一个因素)对实验指标构成影响的试验,该试验设计的前提是假设所考察的因素之间不存在交互作用,通过一次仅改变其中一个因素的水平而其它保持因素恒定,然后依次对其它因素进行类似分析^[4]。方差分析是一种非常实用、有效的统计检验方法,能用于检验试验过程中因素对实验结果影响的显著性^[5]。单因素试验的方差分析由于其操作起来简单容易,结果明了,因此一直都是生物工程领域常用试验设计和数据处理方法。但是,通过单因素试验方差分析仅能得到单个因素对试验结果影响的显著性,而且当考查的因素较多时,其工作量会相当大,并且通过该方法不能得到最优组合^[6]。因此常被用于初步筛选,得到对考察目标影响显著的因素后,再利用正交试验设计等方法对主要因素进一步分析,得

* 基金项目:中国矿业大学教改项目(2014YB22)

作者简介:何环(1981-),男,博士,副教授,主要研究方向:应用微生物学的教学与科研,E-mail:hehuan6819@cumt.edu.cn

(收稿日期:2014-12-10 接受日期:2014-12-30)

到其最优组合。

Excel 的统计功能因为被封装在一个外挂模块中,其受关注程度相对较少。其功能模块组合中包括了常用统计分析功能如:单因素方差分析、可重复双因素及无重复双因素方差分析等试验设计^[6]。Excel 提供的单因素方差分析结果通过表格 1 展示,其中三个数据大小可以判断出因素的重要性,F、F-crit 和 P

value,前者表示显著性水平为 0.05 时临界值,所以分析得到的 $F > F\text{-crit}$, 因素对试验指标有显著影响;P value 表示的是因素对试验结果无显著影响的概率,当 $P\text{ value} \leq 0.01$ 时说明因素对试验结果影响非常显著; $0.01 \leq P\text{ value} \leq 0.05$ 说明因素对试验结果影响一般显著。

表 1 方差分析结果^[3]

Table 1 Results of analysis of variance

Source of variance	SS	df	MS	F	P-value	F crit
Between group	303.6	4	75.9	15.18	0.000299	3.47805
Inside group	50	10	5			
Total	353.6	14				

*SS:sum of deviation square;df: degree of freedom;MS: mean square。

2 Design Expert 中 P-B 试验设计和方差分析

在多因素试验中,将所有因素的全部水平相互交叉组合进行试验,以考察各因素的主要效应与因素之间的相互效应称为析因设计,它又可分为全因子和部分因子设计^[7]。多因素全面组合设计由于实际操作起来较为繁琐,所以在实际应用中部分因子设计应用较多。Design- Expert 是由 State-East 公司开发的是一款专门面向实验设计以及相关分析的软件,P-B 试验设计是该软件因子分析功能菜单下一项独立的试验设计选项。P-B 试验可以通过做 n 次试验来从 n-1 个试验因素中筛选出最主要因素,其中试验设计中必需至少有一个空白因素,每个因素取两个水平,下水平为原始培养条件,上水平约取下水平的 1.25 倍,通过实验结果的回归分析方程可以对实验结果的可能优化值进行预测^[4,5]。当初步筛选出主要的影响因素后可以采用正交

设计等方法对试验做进一步优化。韩健等^[8]采用 P-B 和中心复合设计 (central composite design) 对影响地衣芽孢杆菌 TJ-101 发酵生产 B- 甘露聚糖酶粗发酵液絮凝操作的 8 个条件进行筛选优化,通过 11 次试验筛选出 3 个重要因子。黄丽金等^[9]采用 P-B 设计法对 12 种乳酸菌生长促进因子以及培养基的初始 pH 等 13 个因素对德氏乳杆菌保加利亚亚种 LB14 生长的影响进行评价,通过 20 次试验筛选出 2 个关键影响因子。表 2 为 P-B 试验结果的偏回归系数和影响因子及所建立模型,其模型的方差分析结果形式与表 1 相似。表中回归系数为所对应的回归方程因变量的系数,正负值分别对应着对试验结果影响是正相关还是负相关,标准误差表示对应回归系数的标准误差,E(x1)为每个变量对试验结果的影响值,直接决定影响的重要性,贡献值与因素的重要性密切相关。

表 2 P-B 试验设计中偏回归系数及影响因子的显著性分析^[8]

Table 2 The significance analysis of partial regression coefficient and impact factor in P-B experiment design

Factor	Coefficient estimate	Standard error	Stdized effect	Contribution	Importance
Intercept	0.44	0.032			
X ₁	0.10	0.032	0.210	20.89	Yes
X ₂	0.015	0.032	0.030	0.41	No
X ₃	0.049	0.032	0.098	4.60	No
X ₄	-0.022	0.032	-0.045	0.94	No
X ₅	-0.14	0.032	-0.27	35.04	Yes
X ₆	-0.12	0.032	-0.24	28.13	Yes
X ₇	0.051	0.032	0.01	5.03	No
X ₈	-0.0069	0.032	-0.014	0.091	No
X ₉	0.038	0.032	0.075	2.71	No
X ₁₀	0.033	0.032	0.066	2.06	No
X ₁₁	0.0066	0.032	0.013	0.084	No

3 正交试验设计助手与正交试验设计极差和方差分析

正交试验设计是利用正交表科学地安排与分析多因素试验的方法^[3]。正交试验结果可以进行直观分析和方差分析,直观分析可以得到因素的主次关系和最优化水平组合值,方差分析可以判断因素对实验结果影响的显著性。另外,根据因素水平

图还可以为后续优化试验提供参考。“正交试验设计助手”是由国内开发一款针对正交试验设计和数据处理的软件,软件不仅提供了大多数常见的等水平正交表,对于混合水平的正交表,该软件的专业版也提供了混合表编辑功能。运用正交试验设计助手设计试验时有几个细节问题需要注意,首先在设置因素水平表时,因素的水平应按照无序排列,尽量使设计的试验组合

是随机组合。其次,为了进行方差分析,应该选择较大的正交表,便于安排空白列。正交试验设计助手的数据表导入功能可以将直观分析和方差分析的表格保存成 CSV 格式,直接在 Word 中粘贴。根据实验结果是越大越好还是越小越好,直观分析结果可以直接得到最优化水平和因素的主次关系,而方差分析中根据 F 比和 F 临界值大小可以判断出因素对实验结果影响的显著性。

4 DPS 与均匀试验设计和回归分析

正交试验设计也存在其局限性,它只适于水平数不多的试验。若所考察的试验中每个因素的水平数较多(一般因素的水平数大于因素个数),均匀设计就非常实用。均匀设计法是由我国数学家方开泰和王元创造的一种多因素优化法,它只考虑试验点在试验范围内均匀分布,最大特点是试验次数可以等于最大水平数,而非试验因子数的平方,试验次数仅与需要考察的水平数有关,一般来说,试验次数选为试验因子个数的 3 倍左右为宜^[10]。与正交试验设计类似,均匀设计也有一套精心设计的均匀表,且每个设计表都有相应的使用表,实际使用时需要根据使用表安排。由于均匀设计表不具有正交表的整齐可比性的特点,只能进行直观分析和回归分析,而不能进行方差分析,近几年来均匀设计方法在微生物培养优化方面已取得了不少成功的例子^[11,12]。

DPS(data processing system)软件是国内集试验设计和统计分析功能的软件,它在均匀试验设计,混料试验设计方面具有国际领先水平^[10]。DPS 中试验设计就提供了有关均匀设计的所有方案,可以直接进入相应菜单完成操作。均匀设计的试验结果的直观分析可以从所进行的实验组中挑选最佳试验结果作为较优的工艺条件,由于均与设计的试验点分散均匀,用该方法找到的试验点一般距离最佳试验点也不会太远,所以该法也是一种非常有效的方法。借助与 DPS 菜单中多元分析的回归分析可以建立试验指标与影响因素之间的数学模型,确定因素的主次顺序和最优方案等,也可以通过 DPS 菜单中数值分析的方程求解得到最优方案。从回归统计中的相关系数 R 可以判断出所建立起的方程的回归效果。其所建立的回归方程分析结果中回归系数和标准误差与 P-B 试验中含义相同, t_{Stat} 及统计分析中 t 检验的结果,如果是多元回归分析则可直接根据 t_{Stat} 大小判断因素的主次顺序,P-value 值表示 t 检验偏回归系数不显著的概率,如果 P-value < 0.01,则可以认为该系数对应的变量对试验结果影响非常显著,如果其值介于 0.01 和 0.05 之间则认为影响一般显著。

5 Design expert 与响应面设计

响应面设计法(response surface methodology, RSM)是一种实验条件寻优的方法,适宜于解决非线性数据处理的相关问题,它囊括了试验设计、建模、检验模型的合适性、寻求最佳组合条件等众多试验和统计技术^[13]。该方法不仅可以减少试验次数,提高试验精确度,而且可以通过建立连续变量曲面模型同时对影响试验指标的各因子水平及其交互作用进行优化和评价。发酵培养基的优化中较常用的响应面设计方法有 Box-behnken 设计(BBD)和中心组合设计(central composite de-

sign, CCD)^[14]。Box-behnken 于 1960 年提出由 2 水平因子设计和不完全区组设计而成的一种组合设计方法,其中每个因素取 3 个水平,分别以(-1, 0, 1)编码,然后根据实验表进行设计,运用响应面法对试验后的数据进行分析^[15]。由 Box 和 Wilson 开发的 CCD 设计是国际上一种较为常用的响应面法。与 B-B 设计相比,CCD 在试验点设计上除了考虑最低水平,高水平外还增加了 α 水平,这样可以使得 CCD 设计的试验点具有旋转性质。整体上来看,CCD 较适用于 2-5 因素的优化试验。如果因素少于 4 个时,BBD 相较 CCD 更具有优势^[16]。

Design-Expert 功能虽然不如 SAS 强大,但可方便的进行 CCD 响应优化分析或 B-B 响应面优化分析,并且可以满足实验对数据处理的要求。其内嵌 Response surface 就包含了这两种常见的响应面方法,它同时包括试验方案的设计、数据分析和结果寻优三部分。从分析结果中因变量与自变量之间的对应关系可以判断所拟合的回归分析方程是否显著,另外其寻优处理部分也可以给出考察变量的最佳的理论组合值。

6 小结

许多研究工作表明,合理地使用试验设计和数据处理方法,会使研究工作起到事半功倍的效果。生物工程专业领域的蓬勃发展也带动了相应研究工作的发展,越来越多的研究工作者和产品生产者可以在更广泛的范围内考虑因素的组合,因素之间的交互作用,建立起因素与试验指标之间的数学模型,优化试验结果和预测试验值。除了本文介绍的一些常见的软件 and 数据处理方法外,SAS (Statistical analysis system)^[14], Minitab^[15] 和神经网络算法^[16]等也在该领域有一定应用。但是不管哪种设计软件和数据处理方法,它都有一定的局限性。因此事先必须进行调研,查询和充分的论证。本文希望通过介绍生物工程领域常用的试验设计与数据处理软件,能对该领域的教学和研究工作者利用这些软件时提供一定帮助。

参考文献(References)

- [1] 许彬,薛刚,李慧星,等.生物工程专业试验设计与数据处理课程教学对策[J].现代农业科技,2010,17: 31
Xu Bin, Xue Gang, Li Hui-xing, et al. The teaching method of experiment design and data processing in bioengineering [J]. Modern agricultural science and technology, 2010, 17: 31
- [2] 惠明,田青.生物工程专业《试验设计与数据处理》的教学探讨[J].河南科技学院学报,2009,37(3): 159-162
Hui Ming, Tian qing. The discussion of teching metod in experiment design and data processing of bioengineering [J]. Journal of school of science and technology He nan, 2009, 37(3): 159-162
- [3] 李云雁,胡传荣.试验设计与数据处理 [M].化学工业出版社,第二版,2008
Li Yun-yan, Hu Chuan-rong. Experiment design and data processing [M]. Press of chemical industry, second edition, 2008
- [4] 肖怀秋,李玉珍.微生物培养基优化方法研究进展 [J].酿酒科技,2010,1: 90-94
Xiao Huai-qiu, Li Yu-zhen. Research progress in the optimization of microbial culture medium[J]. Liquor science and technology, 2010, 1: 90-94

- Du Xiao-chun. On Medical College Teachers' Professional Development Issues in Transformation Period of Higher Education[J]. The Guide of Science & Education, 2014, 1(17): 56-57
- [13] 赵晶, 刘新平, 孟艳玲, 等. 生物化学物质代谢讨论课的教学体会[J]. 山西医科大学学报:基础医学教育版, 2010, 12(3): 252-253
- Zhao Jing, Liu Xin-ping, Meng Yan-ling, et al. Teaching Experience on Metabolism Discussion Course of Biochemistry [J]. Journal of Shanxi Medical University (Preclinical Medical Education Edition), 2010, 12(3): 252-253
- [14] 阳雨君, 梁雪清. 医学院校大学生自主学习能力的培养[J]. 华夏医学, 2006, 19(6): 1240-1241
- Yang Yu-jun, Liang Xue-qing. Cultivation of Students' Independent Learning Capability in Medical College [J]. Acta Medicine Sinica, 2006, 19(6): 1240-1241
- [15] 周梅. 医学院教育评估指南[J]. 医学教育, 2001, 1(2): 32-38
- Zhou Mei. Educational and Procedural Guidelines for Accreditation of Medical Schools[J]. Medical Education (China), 2001, 1(2): 32-38
- [16] 张贤斌, 王晨. 外国教育史[M]. 教育科学出版社, 2008
- Zhang Xian-bin, Wang Chen. Foreign History of Education [M]. Educational Science Publishing House, 2008
- [17] 巨瑛梅, 刘旭东. 当代国外教学理论[M]. 教育科学出版社, 2004
- Ju Ying-mei, Liu Xu-dong. Modern Teaching Theory Abroad [M]. Educational Science Publishing House, 2004
- [18] 周家伦. 创新型人才培养与大学生综合素质教育 [J]. 中国高等教育, 2006, 1(5): 1-2
- Zhou Jia-lun. Creative Talents Cultivation and Comprehensive Quality Education of College Students [J]. China Higher Education, 2006, 1(5): 1-2
- [19] 仇静莉, 许海元. 高校素质教育与创新型人才的培养[J]. 河北大学学报(哲学社会科学版), 2007, 32(2): 76-79
- Qiu Jing-li, Xu Hai-yuan. Quality Education and Cultivation of Innovative Intellectuals in University [J]. Journal of Hebei University (Philosophy and Social Science), 2007, 32(2): 76-79
- [20] 周志太. 讨论课培养学生的创新思维能力 [J]. 淮北煤炭师范学院学报(哲学社会科学版), 2009, 30(6): 140-144
- Zhou Zhi-tai. Discussion Course Cultivates the Students' Creative Thinking [J]. Journal of Huabei Coal Industry Teachers College (Philosophy and Social Sciences), 2009, 30(6): 140-144

(上接第 3754 页)

- [5] 代志凯, 张翠, 阮征. 试验设计和优化及其在发酵培养基优化中的应用[J]. 微生物学通报, 2010, 37(6): 894-903
- Dai Zhi-kai, Zhang Cui, Ruan Zheng. The application of experimental design and optimization techniques in optimization of microbial medium[J]. Microbiology China, 2010, 37(6): 894-903
- [6] 徐仲安. 计算机技术在正交试验设计数据分析中的应用[J]. 科技情报开发与经济, 2004, 14(5): 135-136
- Xu Zhong-an. Application of computer technique in the data analysis of the orthogonal experiment design [J]. Sci and tech information development and economy, 2004, 14(5): 135-136
- [7] 刘炯天, 樊民强. 试验研究方法[M]. 中国矿业大学出版社, 2006
- Liu Jiong-tian, Fan Ming-qiang. Experiment study method [M]. Press of China university of Mining technology, 2006
- [8] 韩健, 刘朝辉, 齐威, 等. β -甘露聚糖酶发酵液絮凝条件的统计学筛选与响应面优化[J]. 生物加工过程, 2007, 5(2): 29-34
- Han Jian, Liu Zhao-hui, Qi Wei, et al. Optimization of the flocculation conditions of β -mannanase broth using statistical screening and response surface method [J]. Chinese journal of bioprocess engineering, 2007, 5(2): 29-34
- [9] 黄丽金, 陆兆新, 袁勇军. 响应面法优化德氏乳杆菌保加利亚种增殖培养基[J]. 食品科学, 2005, 26(5): 103-107
- Huang Li-jin, Lu Zhao-xin, Yuan Yong-jun. Optimization of a culture medium for cell multiplication of Lactobacillus delbrueckii subsp bulgaricus LB14 using response surface methodology [J]. Food science, 2005, 26(5): 103-107
- [10] 唐启义. DPS 数据处理系统 - 实验设计、统计分析及数据挖掘[M]. 科学出版社, 第二版, 2010
- Tang Qi-yi. Data processing system, experiment design, statistical analysis and data excavation [M]. Press of science, second edition, 2010
- [11] 杨丽英, 马国顺. 均匀设计在微生物发酵条件优化试验中的应用[J]. 太原师范学院院报, 2009, 8(3): 23-25
- Yang Li-ying, Ma Guo-shun. The application of uniform design in experiment of microorganism fermentation [J]. Journal of Taiyuan normal university, 2009, 8(3): 23-25
- [12] 马懿, 周健, 左勇, 等. 均匀设计优化微生物絮凝剂产生菌的培养条件[J]. 工业水处理, 2013, 33(3): 65-67
- Ma Yi, Zhou Jian, Zuo Yong, et al. Optimizaiton cultural conditions of microbial flocculant-producing strain by homogeneous design[J]. Industrial water treatment, 2013, 33(3): 65-67
- [13] 郝学财, 余晓斌, 刘志钰, 等. 响应面方法在优化微生物培养基中的应用[J]. 食品研究与开发, 2006, 27(1): 38-40
- Hao Xue-cai, Yu Xiao-bin, Liu Zhi-yu, et al. The application of response surface methodoogy in optimization of microbial media[J]. Food study and development, 2006, 27(1): 38-40
- [14] 徐子钧, 李剑, 梁凤来, 等. 利用 SAS 软件优化 L-乳酸发酵培养基[J]. 微生物学通报, 2004, 31(3): 85-87
- Xu Zi-jun, Li Jian, Liang Feng-lai, et al. SAS application in studies on medium optimization lactic acid fermentation [J]. Bulletin of Microbiology, 2004, 31(3): 85-87
- [15] 赵丽坤, 张利平, 张秀敏, 等. 利用 Minitab 软件优化阿维菌素产生菌发酵培养基[J]. 河南农业科学, 2009, 6: 90-93
- Zhao Li-kun, Zhang Li-ping, Zhang Xiu-min, et al. Minitab applicationin Studies on Medium Optimization of avermectins fermentation [J]. He Nan Agricultural science, 2009, 6: 90-93
- [16] Jorjani E, Chehreh CS, Mesroghli S. Prediction of microbial desulfurization of coal using artificial neural net works [J]. Minerals Engineering, 2007, 20: 1285-1292