




全国高等农林院校“十一五”规划教材

试验设计与 统计分析

林学类专业用

洪伟 主编

 中国农业出版社

全国高等农林院校“十一五”规划教材

试验设计与统计分析

洪 伟 主编

林学类专业用

中国农业出版社

图书在版编目 (CIP) 数据

试验设计与统计分析/洪伟主编. —北京: 中国农业出版社, 2009. 1

全国高等农林院校“十一五”规划教材

ISBN 978-7-109-13193-4

I. 试… II. 洪… III. ①试验设计(数学)-高等学校-教材②统计分析(数学)-高等学校-教材 IV. 0212

中国版本图书馆 CIP 数据核字 (2008) 第 199630 号

中国农业出版社出版

(北京市朝阳区农展馆北路 2 号)

(邮政编码 100125)

责任编辑 戴碧霞 田彬彬

中国农业出版社印刷厂印刷 新华书店北京发行所发行

2009 年 6 月第 1 版 2009 年 6 月北京第 1 次印刷

开本: 820mm×1080mm 1/16 印张: 23.25

字数: 555 千字

定价: 35.00 元

(凡本版图书出现印刷、装订错误, 请向出版社发行部调换)

内 容 提 要

试验设计与统计分析是数理统计学的一个重要分支学科，随着试验设计技术与数据分析方法的不断丰富与发展，现已形成一门系统的专业基础学科，并在多学科广泛应用。本教材分为17章，包括试验设计、方差分析、回归分析、完全随机化和随机区组设计与分析、拉丁方与希腊拉丁方设计与分析、正交试验设计、平衡不完全区组设计与分析、裂区试验设计与分析、协方差分析、均匀设计与分析、回归正交试验设计与分析、回归旋转设计与分析、饱和 D -最优设计与分析、混料试验设计、三次设计、基于人工神经网络的试验设计分析模型、遗传算法在试验设计统计分析中的应用等内容。

本教材可作为高等农林院校林学类本科专业的教材和教学参考书，也可供农林科研院所科研人员以及生物学、园艺、作物、环境、食品科学等专业学生、教师、科研人员参考使用。

郑 重 声 明

中国农业出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 65005894, 59194974, 59194971

传 真：(010) 65005926

E - mail：wlxyaya@sohu.com

通信地址：北京市朝阳区农展馆北路2号中国农业出版社教材出版中心

邮 编：100125

购书请拨打电话：(010) 59194972, 59195117, 59195127

数码防伪说明：

本图书采用出版物数码防伪系统，用户购书后刮开封底防伪密码涂层，将16位防伪密码发送短信至106695881280，免费查询所购图书真伪，同时您将有机会参加鼓励使用正版图书的抽奖活动，赢取各类奖项，详情请查询中国扫黄打非网 (<http://www.shdf.gov.cn>)。

短信反盗版举报：编辑短信“JB, 图书名称, 出版社, 购买地点”发送至10669588128

短信防伪客服电话：(010) 58582300/58582301

主 编 洪 伟

副主编 吴承祯 杨俊霞

编 者 (按姓氏笔画排序)

邢世岩 (山东农业大学)

刘苑秋 (江西农业大学)

刘贤谦 (山西农业大学)

刘金福 (福建农林大学)

杨立峰 (河南科技学院)

杨俊霞 (河北农业大学)

吴承祯 (福建农林大学)

何东进 (福建农林大学)

洪 伟 (福建农林大学)

黄少伟 (华南农业大学)

前 言

试验设计与统计分析是运用概率论与数理统计的基础知识,研究如何在科研和生产中合理有效地获得试验数据,并对数据进行分析推断的一门学科。这门学科既有较高的理论要求,又有很强的实用性。试验设计方法不仅应用于农业、林业,还广泛适用于食品、电子、医学、航空、机械制造等学科,已形成一门较为完整的专业基础学科。它已渗透到各个学科,特别是目前电子计算机的应用已深入到各学科,因而试验设计的应用范围日益广泛。试验设计已成为制订科研方案和分析实验数据的重要手段,一个科学、正确的试验设计,可以最大限度地节约试验成本,提高试验成效,缩短试验周期,同时又能获得可知的结论。

随着科学技术的发展,试验设计与统计分析方法在不断发展与丰富,如混料试验设计试验方法的提出,遗传算法、神经网络等人工智能算法广泛应用于试验数据分析等。试验设计这一学科知识具有极强的应用性,是林学类专业的重要专业基础课,是开展林学专业科研与实践的重要手段。本教材共分17章,由单因素试验到多因素试验、简单试验到复杂试验、经典试验到现代生命科学研究中应用的最新试验设计方法,介绍了试验设计的基本理论和原理。通过试验的统计模型建立、统计量构建到假设的检验与模型参数的估计,介绍如何合理有效地获得数据,并对数据进行分析推断的原理与方法。

生物学科是一门实验性学科,很多生物学现象及其原理都是在实验中发现,并通过实验加以验证并得出结论的,因此试验设计的好坏直接影响结果的分析及其结论的正确、可靠与否。所以本教材把理论与实践相结合作为重要的教学手段,通过理论教学与实践相结合的方式,深化、扩展理论知识,培养学生科学研究的思维方式、认真踏实的学风、开拓创新的精神和科学严谨的科研作风。

考虑到不同专业对试验设计的要求不同、学时也可能有差异,因此,针对不同专业、不同学时教学内容应适当选择。

本书的编写分工如下：洪伟负责绪论及第十三章、第十六章，邢世岩负责第一章、第二章，刘贤谦负责第三章、第五章，黄少伟负责第四章，何东进负责第六章，杨俊霞负责第七章、第八章，刘苑秋负责第九章，杨立峰负责第十章、第十一章，刘金福负责第十二章、第十五章，吴承祯负责第十四章、第十七章。

鉴于编者水平有限，书中不妥之处在所难免，敬请广大读者及同行批评指正。

编者

2008年10月

目 录

前言

绪论	1
一、试验设计与统计分析概念	1
二、科学研究的方法与要求	2
三、试验研究的特点	3
四、本课程的作用与重要性	4
五、本课程的发展及其与其他学科的关系	4
思考题	5
第一章 试验设计	6
第一节 试验设计的几个基本概念	6
第二节 试验误差	7
第三节 试验设计基本原理	7
一、试验设计的“三要素”	7
二、试验设计的三个基本原则	8
思考题	9
第二章 方差分析	10
第一节 单向分组的方差分析	10
一、组内观测值数目相等的单向分组资料的方差分析	10
二、组内观测值数目不等的单向分组资料的方差分析	14
第二节 两向分组的方差分析	16
一、组合内无重复观测值的两向分组资料的方差分析	16
二、组合内有重复观测值的两向分组资料的方差分析	18
第三节 系统分组的方差分析	21
第四节 数据转换与漏失数据的弥补	25
一、数据转换	25
二、遗失数据的弥补	27
思考题	28

第三章 回归分析	30
第一节 一元线性回归	30
一、相关与回归的概念	30
二、直线相关	31
三、直线回归	33
第二节 曲线回归	40
一、曲线的类型与特点	40
二、曲线方程的配置实例	42
第三节 多元线性回归与相关分析	44
一、多元线性回归与相关分析概述	44
二、多元线性回归方程的建立	45
三、多元线性回归关系的假设检验	47
四、偏回归系数的假设检验	49
五、多元回归方程的区间估计	50
思考题	50
第四章 完全随机化和随机区组设计与分析	52
第一节 完全随机化设计与分析	52
一、单因素完全随机化设计与分析	52
二、双因素完全随机化设计与分析	55
第二节 随机区组设计与分析	57
一、单因素随机区组	57
二、双因素随机区组	61
三、三因素随机区组	64
思考题	68
第五章 拉丁方和希腊拉丁方设计与分析	71
第一节 拉丁方设计与分析	71
一、试验设计	71
二、单因素试验结果的分析	75
三、二因素试验结果的分析	79
第二节 希腊拉丁方设计与分析	83
思考题	86
第六章 正交试验设计	87
第一节 正交试验设计法	87

一、全面比较法	88
二、轮换比较法	88
三、正交试验设计	89
第二节 交互作用的试验设计与统计分析	95
第三节 有重复试验的正交试验	99
第四节 混合水平的正交试验设计	102
第五节 单纯形最优化方法的应用	105
思考题	110
第七章 平衡不完全区组设计与分析	113
第一节 平衡不完全区组设计与分析	113
一、平衡不完全区组设计	113
二、平衡不完全区组设计的统计分析	115
第二节 平衡格子设计与分析	119
一、平衡格子设计	119
二、平衡格子设计的统计分析	123
三、平衡格子设计的评价及使用时应注意的问题	128
第三节 部分平衡格子设计与分析	129
一、部分平衡格子设计	129
二、部分平衡格子设计的统计分析	131
思考题	137
第八章 裂区试验设计与分析	139
第一节 裂区试验设计与分析	139
一、裂区试验设计的原理与方法	139
二、裂区试验设计统计分析	142
三、再裂区试验设计与分析	148
四、裂区试验设计的评价与应用中的注意事项	155
第二节 林木多年多点试验设计与分析	156
一、林木多年多点试验方法	156
二、统计分析	156
思考题	160
第九章 协方差分析	161
第一节 协方差分析方法	161
第二节 单向分组的协方差分析	163
一、统计模型及其假定	163

二、协方差分析方法.....	164
第三节 两向分组的协方差分析	167
一、统计模型及其假设	167
二、协方差分析方法.....	167
第四节 多个协变量的协方差分析	172
思考题	177
第十章 均匀设计与分析.....	178
第一节 试验设计与均匀设计	178
一、试验设计	178
二、均匀试验设计方法	179
三、混合水平的均匀设计表	186
四、均匀试验设计与正交试验设计的比较	190
五、均匀试验设计的几个问题	191
第二节 配方均匀设计与分析	192
一、配方均匀设计	192
二、有约束的配方均匀设计	194
三、均匀设计在系统工程中的应用.....	196
思考题	197
第十一章 回归正交试验设计与分析	199
第一节 一次回归正交设计与结果分析	199
一、一次回归正交试验设计	199
二、一次回归正交设计试验结果的统计分析	202
三、一次回归正交设计试验结果的失拟性检验	204
四、回归正交设计的应用	205
五、快速登高法选择最佳处理组合.....	208
第二节 二次回归正交组合设计及其统计分析	210
一、二次回归正交试验的回归组合设计	211
二、二次回归正交组合设计试验结果的统计分析.....	218
三、四元二次回归正交组合设计示例	219
四、回归方程的应用.....	223
思考题	224
第十二章 回归旋转设计与分析	226
第一节 回归旋转设计的几个重要问题	226
一、旋转性和正交性条件	226

二、二次旋转组合设计的通用性	230
第二节 二次正交旋转组合设计的统计方法	231
第三节 通用旋转组合设计的统计方法	241
一、建立二次多项式回归方程	241
二、回归方程的显著性检查	243
三、回归系数的显著性检验	244
思考题	250
第十三章 饱和 D—最优设计与分析	253
第一节 D —优良性	253
一、回归模型扩展	253
二、 D —优良性	255
第二节 一次饱和 D —最优设计	256
第三节 二次饱和 D —最优设计	258
思考题	263
第十四章 混料试验设计	265
第一节 单形格子设计	265
一、基本概念	265
二、 $\{p, 1\}$ 设计	266
三、 $\{p, 2\}$ 设计	267
四、 $\{p, 3\}$ 设计	269
五、编码方法	274
六、控制点检验	275
七、单形格子设计的几点说明	276
第二节 单形重心设计	276
第三节 极端顶点设计	281
思考题	283
第十五章 三次设计	286
第一节 三次设计基本原理	286
第二节 三次设计在参数优化中的应用	287
思考题	289
第十六章 基于人工神经网络的试验设计分析模型	292
第一节 神经网络概述	292
第二节 神经网络理论及其统计分析	294

思考题	299
第十七章 遗传算法在试验设计统计分析中的应用	300
第一节 遗传算法及其数学基础	300
一、简单遗传算法 (GA)	300
二、数学基础	302
三、遗传算法的描述	303
四、遗传算法的特点	306
五、简单遗传算法的一些改进	306
第二节 遗传算法在参数优化设计中的应用	307
一、遗传算法在种群空间格局测定中的应用	307
二、遗传算法在杉木人工林自疏规律统计分析中的应用	309
思考题	312
附表 常用数理统计用表	314
附表 1 百分数的 $\sin^{-1}\sqrt{p}$ 变换	314
附表 2 t 分布的双侧分位数表	317
附表 3 F 检验的临界值表	319
附表 4 均匀设计表	327
附表 5 正交表	338
附表 6 χ^2 分布的上侧分位数表	343
附表 7 正交拉丁方表	344
附表 8 平衡不完全区组设计表	346
附表 9 多重比较中的 q 表	348
附表 10 多重比较中的 S 表	349
附表 11 邓肯 (Duncan) 多重比较 SSR 值临界值表	350
附表 12 检验相关系数 $\rho=0$ 的临界值表	352
附表 13 相关系数 r 与 R 的临界值表	353
主要参考文献	354

绪 论

提要 试验设计是自然科学研究方法论领域中的一个分支学科，是科学研究计划中关于研究方法步骤的一项内容。它是一项通用技术，是生物类、农学类专业基础课程，是当代科技和工程技术人员必须掌握的技术方法。本章主要阐述了试验设计与统计分析的概念、科学研究方法与要求、试验研究特点、本课程作用与重要性及其与其他学科的关系，目的是使学生了解试验设计与统计分析的发展、作用与地位及其与其他学科的关系。

一、试验设计与统计分析概念

“实验科学”一词在中世纪末期由英国进步思想家罗吉尔·培根（Roger Bacon）最先提出，他提倡面向自然、依靠实验、注重经验，反对盲目崇拜权威。他所说的实验科学是指运用实验方法的科学，但是实验科学方法论的总结和概括，则是由英国唯物主义哲学家弗朗西斯·培根完成的。弗朗西斯·培根作为近代实验科学的哲学代言人，在《新工具》一书中提出的科学方法主要包括两个方面的内容：一是要坚持扫扰人心的偏见，“使理智完全得到解放和刷新”，这样才能进入科学的大门；二是要改善人们在科学研究中的手段，采用试验归纳法，即在获取知识的方法上，主要靠观察和试验，在整理经验材料、发展知识的方法上，主要靠归纳。试验方法是从观察方法发展起来的，它以观察方法为基础，但又高于观察，具有观察方法所不具备的优点。从观察发展到试验，是科学认识方法和手段的一大飞跃。试验方法的确立、试验方法论原理的制定，对近代科学的发展起到了巨大的推动作用。

试验是人们从事最普遍的活动之一，它的应用范围十分广泛，从一般的家务安排，到工业现代化中的技术革新，如生命科学、电子科学、航空航天技术等都会用到试验。试验允许研究者通过有目的地改变一个输入变量的设置，寻找系统的响应或输出变化规律，进而用统计方法或通常意义的分析方法研究输入与输出变量之间的关系，达到对一个系统的输入变量是如何影响其性能的更深层次的理解，这就为选择最优输入设置提供了一个基本依据。

试验设计是一门研究如何进行科学试验的设计实施、数据收集、结果分析、结论推断的科学，是有关知识和技术的一个整体，它使研究人员能够找到好的试验、有效地进行数据分析并建立来自分析的结论和最初研究目标之间的联系。对从试验收集到的数据有必要进行分析，而且这种分析应与所用的设计相适应。这种分析包括模型拟合、通过残差分析评估模型的假定。统计分析的前提是具备大量的统计数据，没有充足的、系统的、准确的统计数据，统计分析就无从谈起，因此，对统计数据收集、整理，是进行统计分析的基础，也是关系统计分析质量的关键。所以，正确的试验设计是科学试验成功的必备条件，而优良的设计方法及分析技术可保证试验结果和推断的可靠性及科学性。

近代试验设计可以追溯到统计学家费舍尔 (R. A. Fisher) 于 20 世纪 20 年代在英国 Rothamsted 农业试验站的开创性工作, 他于 1925 年在《研究工作中的统计方法》一书中提出了“试验设计”的术语, 1935 年《试验设计》一书中首次系统地介绍了试验设计的原理和方法, 如试验设计的随机、重复、区域控制三个基本原理; 又如随机区组设计、拉丁方设计等基本的试验设计与分析方法。R. A. Fisher 的杰出工作以及 F. Yates 和 D. J. Finney 的卓越贡献都是受到农业和生物中的问题的激励。由于农业试验规划较大, 花费时间长, 而且必须妥善处理田间的差异, 这些因素导致了分区组、随机化、重复试验、正交性以及方差分析和部分因析设计等技术的发展。在林业生产和科学试验中, 要经常研究哪些生产措施和外界条件对林木生长有影响, 以寻求优质、高产、低耗的生产方案。培育林木育种、确定防治森林病虫害的措施等, 这些都离不开试验, 而试验与数据是分不开的。有经验的试验者为了使试验结果具有真实性、客观性, 总是在试验前就根据统计学的要求制定出一套完整的试验方案。试验过程就是方案的实施过程, 依靠合理的试验设计得出正确的判断和结论。在 20 世纪 40 年代, 英美两国在工业上采用试验设计方法进行工艺研究。在 20 世纪 50 年代, Cochran 和 Cox(1957) 及 Kempthorne(1952) 的两本经典著作系统、全面地介绍了试验设计的原理和方法。

科学试验作为一种研究性的实践活动, 体现了理论与经验研究的统一, 即要遵循试验研究的基本规范, 科学试验技术方法服务于三大要素。

(1) 假说的表达和论证。这是一个试验的灵魂或思想基础, 包含问题的提出、对问题解答的设想(假说或理论构想)及目标的拟订、对目标假说的论证。

(2) 条件控制和研究性操作。条件变量控制是一个试验取得科学可靠性的关键, 在科学试验中必须强调控制性、重复性与随机性, 这是获得试验数据进行统计分析的基础。

(3) 假说验证和效果评价。这是对一个试验的认识价值和实践价值的考核, 应贯穿全过程, 包含对事实数据进行定性和定量分析, 作出客观结论以验证假说, 过程中验证新假说, 对试验结果作出价值判断等。

二、科学研究的方法与要求

1. 科学研究的方法 科学研究方法, 就是在人们解决某些实际问题的研究中所采用的各种手段和步骤。科学技术研究必须要有正确的研究方法才能有效地进行。探索未知科学技术领域的问题, 没有已知的答案可以遵循, 需要科技工作者在黑暗中摸索前进, 因此, 需要有科学的研究方法作指导, 否则就会感到无所适从、无从下手, 找不到正确的前进方向。

生理学家伊凡·巴洛夫曾指出:“初期研究的障碍, 乃在于缺乏研究方法。无怪乎人们常说, 科学是随着研究方法所获得的成就而前进的。研究方法每前进一步, 我们就更提高一步, 随之在我们面前也就开拓了一个充满着种种新鲜事物的、更辽阔的远景。因此, 我们头等重要的任务是制定研究方法”。法国杰出的科学哲学家庞加莱认为, 科学创造“首先需要创造方法”。实证主义哲学的代表人物皮尔逊 (Karl Pearson, 英国数学家、哲学家) 提出:“科学方法是通向绝对知识或真理入口”和“唯一道路”, “整个科学的统一只在于其方法而不在于其材料”。

2. 试验的基本要求 为了使试验设计所得结果是可靠和正确的, 设计者必须注意以下基本要求。

(1) 试验条件要有代表性。试验条件要能代表将来准备采用这些试验结果的地区的自然条件,例如要代表本地区的土质、土壤肥力、海拔高度、气候和其他自然条件。只有这样,得出的试验结果才能在本地区或类似地区推广。

(2) 试验的重演性。在相同条件下再进行试验或实践时能重复获得与原试验相类似的结果,这对在农业生产中推广科研成果有重要意义。农业试验是受复杂的自然条件影响的,不同年份、不同地区进行相同试验往往结果不同。为了保证重演性,首先要保证准确性和代表性,在此基础上还要了解和掌握植物生长发育的生物学特性。

(3) 应当选择适当的试验指标,并有相应的数据分析方法。分析试验结果最基本的统计方法是方差分析,因此试验结果数据需满足方差分析的基本模型要求,如正态、独立、等方差等要求;若不满足则需采取相应措施,如数据转换等。

(4) 试验的结果要可靠。可靠性主要包括准确性和精确性。准确性是指试验中每一个处理的试验值或所研究特性的观察值与其真值的接近程度,愈是接近则试验愈准确,但在试验中真值往往未知,故准确性不易确定;精确性是指试验误差尽可能小,使处理间差异能精确地比较。在进行试验的全过程中,必须尽最大努力准确地执行各项试验技术,力求避免人为差错,特别要注意试验条件的一致性。除试验处理以外,其他条件应尽可能一致,这样才能提高试验结果的准确性和精确性。

三、试验研究的特点

农业和生物学领域试验的研究对象和材料是生物体本身,以农作物为主,也包括昆虫、病菌、土壤微生物、杂草等,从生物体本身生育过程的反应作试验指标,研究有关生长发育的规律、某些因素的作用、某些技术的效果等。由农作物或其他生物体本身的反应来直接检测试验的效果,这是田间试验的重要特点。由于自然界的生物体往往是一个具有多种遗传变异的群体,即使用纯系品种的种子也往往存在一定的变异性,因而试验材料本身便存在产生试验误差的多种因素。

试验大多是在开放的自然条件下进行的,试验环境包括土壤、气候,甚至病虫等生物条件,它们是多变的,再加上农作物试验周期长,尤其最后产品的测定要在田间自然条件下经历生长发育的全过程,因而田间试验的环境条件也存在导致试验产生试验误差,包括系统误差和随机误差的多种可能性。例如,试验的对象是生长于土壤中的作物品种,容易受到各种因素的干扰和制约,如风、雨、温度、湿度、光照、土壤肥力、排灌状况、人为的耕作施肥等。因此,在生产实践中,很难看到某一品种的产量都是 400 kg 或 500 kg,总是随着地点的不同而有高有低。可见,作物生产具有很强的地域性差异和年度间气候条件的差异等。其特点为:

(1) 试验周期长。作物完成一个生活世代的时间很长,短则 3~5 个月,长则十余个月,有的甚至几年;林木的生长周期更长,短轮伐期树种 5~6 年,一般 20~30 年,长则一百多年。因此,植物有的一年内可进行多次试验,如水稻;有的一年或几年只能进行一次试验,如茶树品种的区域试验;有的几年或几十年只能开展一次试验,如林木种源与区域试验。

(2) 非处理因素多。试验过程中,受非试验因素的干扰和影响较多,有自然条件的,如土壤肥力差异;也有人为操作管理的,如不同操作者技术水平差异客观存在;还有很多偶然因素的干