

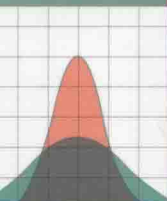


iCourse · 教材
高等农林院校基础课程系列



自主创新
方法先行

生物统计与试验设计



主编 徐辰武 章元明

高等教育出版社



iCourse · 教材
高等农林院校基础课程系列



自主创新
方法先行

生物统计与试验设计

主 编 徐辰武 章元明
副主编 张洪亮 陈志斌 徐向宏
编 委 (按姓氏拼音排序)

陈进卿 (福建农林大学)
宁海龙 (东北农业大学)
王春平 (河南科技大学)
徐辰武 (扬州大学)
杨泽峰 (扬州大学)
张洪亮 (中国农业大学)
朱永平 (云南农业大学)

陈志斌 (沈阳农业大学)
王 伟 (河南科技学院)
夏石头 (湖南农业大学)
徐向宏 (甘肃农业大学)
张恩盈 (青岛农业大学)
章元明 (华中农业大学)



高等教育出版社

内容简介

本书系统地介绍了生物统计与试验设计的原理和方法。全书共分9章,在扼要介绍生物统计学发展简史和主要功能的基础上,着重讨论了描述性统计、概率和理论分布、抽样分布、统计推断、方差分析、常用试验设计及其方差分析、一元和多元线性回归与相关分析。本书密切联系生物和农业试验实际,内容安排力求由浅入深、循序渐进。

本书可作为高等院校生命科学类、植物生产类和动物生产类相关专业本科生教材使用。

图书在版编目(CIP)数据

生物统计与试验设计 / 徐辰武, 章元明主编. -- 北京: 高等教育出版社, 2015.8

iCourse·教材. 高等农林院校基础课程系列

ISBN 978-7-04-042063-0

I. ①生… II. ①徐… ②章… III. ①生物统计-高等学校-教材②生物学-试验设计-高等学校-教材
IV. ①Q-33

中国版本图书馆CIP数据核字(2014)第097366号

项目策划 王瑜 李光跃 陈琪琳 李艳馥 吴雪梅

策划编辑 孟丽 责任编辑 单冉东 封面设计 张楠 责任印制 张泽业

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印刷 三河市华骏印务包装有限公司
开本 850mm×1168mm 1/16
印张 17.25
字数 460千字
购书热线 010-58581118

咨询电话 400-810-0598
网址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版次 2015年8月第1版
印次 2015年8月第1次印刷
定价 32.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换
版权所有 侵权必究
物料号 42063-00

iCourse · 数字课程 (基础版)

生物统计与试验设计

主编 徐辰武 章元明

<http://abook.hep.com.cn/42063>

登录方法:

1. 访问<http://abook.hep.com.cn/42063>, 点击页面右侧的“注册”。已注册的用户直接输入用户名和密码, 点击“进入课程”。
2. 点击页面右上方“充值”, 正确输入教材封底的明码和密码, 进行课程充值。
3. 已充值的数字课程会显示在“我的课程”列表中, 选择本课程并点击“进入课程”即可进行学习。

自充值之日起一年内为本数字课程的有效期
使用本数字课程如有任何问题
请发邮件至: lifescience@pub.hep.cn



iCourse · 教材
高等农林院校基础课程系列



自主创新
方法先行

生物统计与试验设计

主编 徐辰武 章元明

用户名 密码 验证码 进入课程

[注册](#)

[内容介绍](#)

[纸质教材](#)

[版权信息](#)

[联系方式](#)

[相关教材](#)

“生物统计与试验设计”数字课程与纸质教材一体化设计, 紧密配合。数字课程包括教学课件、拓展阅读、视频、深入学习、详细推导等多种资源类型, 丰富知识呈现形式, 以此提升课程教学效果, 辅助学生拓展知识, 引导学生自主学习。



农业生态学
林文雄 陈雨海



普通遗传学实验
张祖新 严长杰



动物生物化学实验技术教程
李庆章

高等教育出版社

数字资源 先睹为快



拓展阅读



详细推导



教学课件



视频

出版说明

“十二五”是继续深化高等教育教学改革、走以提高质量为核心的内涵式发展道路和农林教育综合改革深入推进的关键时期。教育教学改革的核心是课程建设,课程建设水平对教学质量和人才培养质量具有重要影响。2011年10月12日教育部发布了《教育部关于国家精品开放课程建设的实施意见》(教高[2011]8号),开启了信息技术和网络技术条件下校、省、国家三级精品开放课程建设的序幕。作为国家精品开放课程展示、运行和管理平台的“爱课程(iCourse)”网站也逐渐为高校师生和社会公众认知和使用。截至目前,已启动2911门精品资源共享课和696门精品视频公开课的立项建设,其中的1000多门精品资源共享课和600多门精品视频公开课已经在“爱课程(iCourse)”网站上线。

高等教育出版社承担着“‘十二五’本科教学工程”中国国家精品开放课程建设的组织实施和平台建设运营的重要任务,在与广大高校,特别是高等农林院校的调研和协作中,我们了解到当前高校的教与学发生了深刻变化,也真切感受到课程和教材建设所面临的挑战和机遇。如何建设支撑学生自主学习与校际共建共享的课程和新形态教材成为现实课题,结合我社2009年以来在数字课程建设上的探索和实践,我们提出了“高等农林院校基础课程精品资源共享课及系列教材”建设项目,并获批列入科技部“科学思维、科学方法在高等学校教学创新中的应用与实践”项目(项目编号:2009IM010400)。项目建设理念得到了众多农林高校的积极响应,并于2012年12月—2013年6月,分别在北京、扬州、武汉、哈尔滨、福建等地陆续召开了项目启动会议、研讨会和编写会议。2014年,项目成果“iCourse·教材:高等农林院校基础课程系列”陆续出版。

本系列教材涵盖数学、物理、化学化工、计算机、生物学等系列基础课程,在出版形式、编写理念、内容选取和体系编排上有不少独到之处,具体体现在以下几个方面:

1. 采用“纸质教材+数字课程”的出版形式。纸质教材与丰富的数字教学资源一体化设计,纸质教材内容精练适当,并以新颖的版式设计和内容编排,方便学生学习和使用;数字课程对纸质教材内容起到巩固、补充和拓展作用,形成以纸质教材为核心,数字教学资源配合的综合知识体系。

2. 创新教学理念,引导自主学习。通过适当的教学设计,鼓励学生拓展知识面和针对某些重要问题进行深入探讨,增强其独立获取知识的意识和能力,为满足学生自主学习和教师教学方法的创新提供支撑。

3. 强调基础课程内容与农林学科的紧密联系,始终抓住学生应用能力培养这一重要环节。教材和数字课程中精选了大量有实际应用背景的案例和习题,在概念引入和知识点讲授上也总是从实际问题出发,这不仅有助于提高学生学习基础课程的兴趣,也有助于加强他们的创新意识和创新能力。

4. 教材建设与资源共享课建设紧密结合。本系列教材是对各校精品资源共享课和教学改革成果的集成和升华,通过参与院校共建共享课程资源,更可支持各级精品资源共享课的持续建设。

建设切实满足高等农林教育教学需求、反映教改成果和学科发展、纸质出版与资源共享课紧密结合的新形

态教材和优质教学资源,实现“校际联合共建,课程协同共享”是我们的宗旨和目标。将课程建设及教材出版紧密结合,采用“纸质教材+数字课程”的出版形式,是一种行之有效的方法和创新,得到了高校师生的高度认可。尽管我们在出版本系列教材的工作中力求尽善尽美,但难免存在不足和遗憾,恳请广大专家、教师和学生提出宝贵意见与建议。

高等教育出版社

2014年7月

前言

“生物统计与试验设计”是研究如何收集、整理、分析和解释生物学相关研究中所得数据资料的一门科学。该门课程一直是高等院校生命科学类、植物生产类和动物生产类相关专业本科生的学科基础课程,在学生能力与综合素质培养以及毕业论文设计中发挥着十分重要的作用。近年来,随着我国高等教育教学改革的需要,很多课程的知识体系以及教育教学方法均发生了相应的变化,“生物统计与试验设计”课程教学也不例外。结合精品资源共享课的建设成果,在高等教育出版社“高等农林院校基础课程精品资源共享课及系列教材”建设项目的推动下,我们针对高等农林院校生命科学类、植物生产类和动物生产类相关专业人才培养的特点和要求,组织了12所高等农林院校的一线骨干教师,精心编写了本教材。

本教材在简要介绍生物统计学发展简史和主要功能的基础上,着重讨论了描述性统计,概率和理论分布,抽样分布,统计推断,方差分析,常用试验设计及其方差分析,一元和多元线性回归与相关分析。本教材内容密切联系生物学和农业试验实际,引入大量例题和习题,涉及生物科学、遗传育种、作物栽培、植物生理、植物保护、土壤肥料、畜牧和兽医科学、动物医学、蔬菜果树等学科方向。内容编排上,力求由浅入深,循序渐进。

本书采用“纸质教材+数字课程”的出版模式,在纸质教材知识架构基础上,利用数字课程辅以相关拓展知识,包括拓展阅读、深入学习、详细推导和视频等内容,可供学生进行课外自主学习。

本书共分9章,主要编委及编写分工为:徐辰武(第一章),夏石头(第二章),张恩盈(第三章),陈进卿(第四章),宁海龙(第五章第一节至第四节),徐向宏(第五章第五节至第七节),章元明(第六章第一节),朱永平(第六章第二节至第五节),杨泽峰(第七章第一节、第二节),王春平(第七章第三节至第八节),张洪亮(第八章),陈志斌(第九章);全书由徐辰武统稿、王伟协助校对,徐辰武、章元明、张洪亮审校样稿。在本书辑成过程中,特别要感谢高等教育出版社的孟丽老师,正是在她的推动和帮助下,本书得以如此顺利出版。在此,特别感谢各位参与人员,谢谢诸位的尽职尽责与辛勤付出!

由于水平所限,书中难免会有许多缺点和错误,敬请读者批评指正。

徐辰武 章元明

2015年4月

目 录

1 生物统计学概论	1	3.1.1 事件和概率的定义	34
1.1 生物统计学发展简史	2	3.1.2 事件的相互关系	35
1.2 生物统计学的主要功用	4	3.1.3 计算事件发生概率的法则	36
1.3 本书知识导图	5	3.2 二项分布	37
2 描述性统计	9	3.2.1 二项总体	37
2.1 变量与次数分布	10	3.2.2 二项分布	38
2.1.1 变量	10	3.2.3 二项成数(百分数)分布	40
2.1.2 次数分布	11	3.3 多项分布	41
2.2 次数分布表	12	3.4 泊松分布	42
2.2.1 间断性变量次数分布表	12	3.5 正态分布	43
2.2.2 连续性变量次数分布表	13	3.5.1 正态分布及其性质	43
2.3 次数分布图	15	3.5.2 利用正态分布计算概率的方法	46
2.3.1 间断性变量次数分布图	15	3.6 二项分布的正态近似	47
2.3.2 连续性变量次数分布图	16	习题	49
2.3.3 频率和累积频率分布图	18	4 抽样分布	51
2.4 集中性度量统计量	21	4.1 随机抽样和无偏估计	52
2.4.1 算术平均数	21	4.1.1 总体和样本	52
2.4.2 几何平均数	22	4.1.2 随机抽样	54
2.4.3 调和平均数	23	4.1.3 无偏估计	55
2.4.4 中位数	23	4.2 样本平均数的分布	57
2.4.5 众数	25	4.2.1 正态总体样本平均数的分布	57
2.5 离散性度量统计量	25	4.2.2 中心极限定理	58
2.5.1 极差	26	4.2.3 样本平均数分布的概率计算	59
2.5.2 分位数	26	4.3 样本平均数差数的分布	61
2.5.3 方差	27	4.3.1 样本平均数差数分布的理论推导	61
2.5.4 标准差	28	4.3.2 样本平均数差数分布的概率计算	63
2.5.5 变异系数	30	4.4 t 分布	64
2.6 分布偏度和峰度度量统计量	30	4.4.1 t 分布的特点及应用	64
2.6.1 偏度系数	31	4.4.2 t 分布概率计算	66
2.6.2 峰度系数	31	4.5 χ^2 分布	67
习题	31	4.5.1 χ^2 分布的特点及应用	67
3 概率和理论分布	33	4.5.2 χ^2 分布概率计算	67
3.1 事件与概率	34	4.6 F 分布	68

4.6.1 F 分布的特点及应用	68	6.3 双向分组资料的方差分析	128
4.6.2 F 分布概率计算	68	6.3.1 无重复观测值双向分组资料的 方差分析	128
习题	69	6.3.2 有重复观测值双向分组资料的 方差分析	132
5 统计推断	71	6.4 系统分组资料的方差分析	138
5.1 统计假设测验	72	6.4.1 二级系统分组资料的方差分析	138
5.1.1 统计假设测验的基本步骤	72	6.4.2 应用实例	140
5.1.2 假设测验的两类错误	75	6.5 变量转换	142
5.1.3 一尾测验和两尾测验	77	6.5.1 方差分析的基本假定	142
5.2 平均数的假设测验	78	6.5.2 常用变量转换方法	143
5.2.1 单个平均数的假设测验	78	习题	146
5.2.2 两个平均数成组比较的假设测验	79	7 常用试验设计及其方差分析	149
5.2.3 两个平均数成对比较的假设测验	82	7.1 试验设计概述	151
5.3 方差的假设测验	84	7.1.1 因素、水平和处理	151
5.3.1 单个方差的假设测验	84	7.1.2 效应与计算	152
5.3.2 两个方差的假设测验	85	7.1.3 试验误差	153
5.3.3 多个方差的假设测验	86	7.1.4 试验设计的基本原则	155
5.4 成数的假设测验	87	7.1.5 试验设计的小区技术	155
5.4.1 单个成数的假设测验	87	7.2 常用试验设计方法	157
5.4.2 两个成数的假设测验	89	7.2.1 完全随机化试验	158
5.5 χ^2 测验	92	7.2.2 随机区组试验	158
5.5.1 χ^2 测验的原理和方法	92	7.2.3 拉丁方试验	159
5.5.2 适合性测验	93	7.2.4 裂区试验	161
5.5.3 齐性测验	95	7.2.5 正交试验	161
5.5.4 独立性测验	97	7.3 完全随机试验设计的方差分析	162
5.6 参数估计	100	7.3.1 单因素完全随机试验结果的 方差分析	162
5.6.1 参数点估计与区间估计的原理	100	7.3.2 两因素完全随机试验结果的 方差分析	164
5.6.2 参数的点估计与区间估计	100	7.4 随机区组试验设计的方差分析	164
5.6.3 假设测验与区间估计的关系	106	7.4.1 单因素随机区组试验结果的 方差分析	164
5.7 样本容量的确定	107	7.4.2 两因素随机区组试验结果的 方差分析	167
5.7.1 样本容量估计的意义	107	7.5 拉丁方试验设计的方差分析	172
5.7.2 几个常用样本容量估计	107	7.5.1 数据模式	173
习题	111	7.5.2 变异分解	173
6 方差分析	115	7.5.3 应用实例	174
6.1 方差分析的基本原理	117	7.6 裂区试验设计的方差分析	177
6.1.1 数学模型	117	7.6.1 数据模式	177
6.1.2 平方和与自由度的分解	117		
6.1.3 F 测验	119		
6.1.4 多重比较	120		
6.2 单向分组资料的方差分析	123		
6.2.1 组内观测次数相等的方差分析	123		
6.2.2 组内观测次数不相等的方差分析	126		

7.6.2 变异分解	177	8.5.3 两个相关系数的假设测验	217
7.6.3 应用实例	178	8.5.4 相关系数的区间估计	218
7.7 正交试验设计的方差分析	182	8.6 线性回归和相关的内在关系及应用	
7.7.1 数据模式	182	注意事项	218
7.7.2 变异分解	183	8.6.1 线性回归和相关的内在关系	218
7.7.3 应用实例	184	8.6.2 线性回归和相关分析的注意事项	219
7.8 缺值估计	190	习题	220
7.8.1 缺值估计原理	190	9 多元线性回归和相关分析	221
7.8.2 缺值估计实例	191	9.1 多元线性回归分析	222
习题	192	9.1.1 多元线性回归模型	222
8 一元线性回归和相关分析	197	9.1.2 多元线性回归方程的求解和离回归	
8.1 线性回归和线性相关的概念	198	标准误的计算	224
8.1.1 变量间的函数关系与统计关系	198	9.1.3 多元线性回归的假设测验	229
8.1.2 散点图	199	9.1.4 自变量的统计选择与相对重要性	233
8.1.3 自变量与依变量	200	9.2 多元线性相关分析	233
8.1.4 回归分析和相关分析	200	9.2.1 多元相关分析	233
8.2 线性回归方程和离回归标准误	200	9.2.2 偏相关分析	235
8.2.1 线性回归方程及其参数估计	200	习题	238
8.2.2 线性回归中的变异分解与离回归		主要参考文献	241
标准误	205	附表1 标准正态分布累积函数表	242
8.3 线性回归方程的假设测验	206	附表2 标准正态分布的双侧百分位数	
8.3.1 单个线性回归方程的假设测验	206	$u_{\alpha/2}$ 值表	245
8.3.2 两个线性回归方程的假设测验	208	附表3 t 分布两尾临界值 $t_{\alpha/2, df}$ 表	246
8.4 线性回归的区间估计	210	附表4 F 分布右尾临界值 F_{α, df_1, df_2} 表	248
8.4.1 回归截距和回归系数的置信区间	210	附表5 多重比较的 q_{α} 值表	251
8.4.2 条件总体平均数 $\mu_{Y X}$ 的置信区间	212	附表6 多重比较的 SSR_{α} 值表	253
8.4.3 条件总体中个体观测值 $Y_{(p)}$ 的		附表7 χ^2 分布右尾临界值 $\chi^2_{\alpha, df}$ 表	255
预测区间	212	附表8 r 与 R 临界值表	257
8.4.4 条件总体平均数及单个观测值		附表9 常用正交表	259
预测区间的图示	213		
8.5 线性相关分析	213		
8.5.1 相关系数和决定系数	213		
8.5.2 相关系数的假设测验	216		

1

生物统计学概论

- 生物统计学发展简史
- 生物统计学的主要功用
- 本书知识导图

关键词

生物统计学 总体 样本 参数
统计数 试验误差

统计学是有关如何收集、整理、分析和解释反映客观现象总体特征的数据,以便给出正确认识的方法论科学。它是利用相对有限的样本数据,对特定的随机现象作出推断的学科。统计学可分数理统计学和应用统计学两个领域,前者更关注统计推断中新方法的发展,常需要较多的数学知识;而后者则是如何将数理统计方法应用到特定的领域,如生物学、医学和农学。生物统计学是应用统计学的一个分支,即将统计方法应用到生物学及农学领域。本章首先回顾生物统计学的发展历史;然后概括生物统计学的主要功用,并以导图形式展示生物统计学各章的主要内容及其相互关系,以便读者了解本书概貌。

1.1 生物统计学发展简史

生物统计学的研究开始于 19 世纪末。1870 年,英国遗传学家弗朗西斯·高尔顿(Francis Galton, 1822—1911)应用统计方法研究人种特性,分析父母与子女的变异,探索其遗传规律。高尔顿自 1882 年起开设“人体测量实验室”,大量收集了有关身高、体重、胸围等人体特征方面的资料,并深入钻研这些资料中隐藏着的内在联系,最终得出“祖先遗传法则”。他努力探索那些能把大量数据加以描述与比较的方法和途径,引入了中位数、百分位数、四分位数、四分位差以及分布、相关、回归等重要的统计学概念与方法,因此,被后人推崇为生物统计学的创始人。

卡尔·皮尔逊(Karl Pearson, 1857—1936)全面继承和发展了高尔顿的思想,将回归与相关的理论进一步推广和一般化,并建立了相应的数学基础。1891 年之后,皮尔逊潜心研究区分物种时所用数据的分布理论,提出了“概率”和“相关”的概念。接着,又提出标准差、正态曲线、平均变差、均方根误差等一系列数理统计基本术语。皮尔逊致力于大样本理论的研究,他发现不少生物方面的数据有显著的偏态,不适合用正态分布去刻画,为此他提出了后来以他的名字命名的分布族,为估计这个分布族中的参数,他提出了“矩法”。为考察实际数据与族分布的拟合优劣问题,他引进了著名的“ χ^2 测验法”,并在理论上研究了其性质。这个测验法是假设测验最早、最典型的方法,他在理论分布完全给定的情况下求出了测验统计量的极限分布。皮尔逊、高尔顿与威尔登(Raphael Weldon, 1860—1906)为了推广统计在生物学上的应用,于 1901 年创立统计学期刊 *Biometrika*,使得生物统计有了自己的学术阵地。

人类历史进入 20 世纪后,无论社会领域还是自然领域都向统计学提出更多的要求。各种事物与现象之间繁杂的数量关系以及一系列未知的数量变化,单靠记录或描述的统计方法已难以奏效。因此,相继产生“推断”的方法来研究事物总体的真正联系以及预测未来的发展。从描述统计学到推断统计学,这是统计学发展过程中的一大飞跃。统计学发展中的这场深刻变革是在农业田间试验领域中完成的。对现代推断统计的建立贡献最大的是英国统计学家威廉·戈塞特(William Sealy Gosset, 1876—1937)和罗纳德·费希尔(Ronald Aylmer Fisher, 1890—1962)。

戈塞特在生产实践中对样本标准差进行了大量研究。1908 年,戈塞特首次以“Student”(学生)为笔名在该年的 *Biometrika* 上发表了论文“平均数的概率误差”,创立了小样本测验代替大样本测验的理论,即 t 分布(也称为学生氏 t 分布)和 t 测验法。 t 测验已成为当代生物统计工作的基本工具之一,为小样本理论的形成和应用奠定了基础。为

● 拓展阅读 1-1
弗朗西斯·高尔顿
生平简介

● 拓展阅读 1-2
卡尔·皮尔逊生平
简介

● 拓展阅读 1-3
威廉·戈塞特生平
简介

● 拓展阅读 1-4
罗纳德·费希尔生
平简介

此,许多统计学家把1908年看作是统计推断理论发展史上的里程碑。后来,戈塞特又连续发表了“相关系数的概率误差”(1909)、“非随机抽样的样本平均数分布”(1909)和“从无限总体随机抽样平均数的概率估算表”(1917)等论文。这些论文的完成,为“小样本理论”奠定了基础,同时也为以后的样本资料的统计分析与解释开创了一条崭新的道路。由于戈塞特开创的理论使统计学开始由大样本向小样本、由描述向推断发展,因此,有人把戈塞特推崇为推断统计学的先驱者。

20世纪20—50年代间,费希尔对当时被广泛使用的统计方法进行了一系列理论研究,内容涉及估计理论、假设测验、方差分析和试验设计等重要领域。

在对统计量及抽样分布理论的研究方面,费希尔系统地发展了正态总体下各种统计量的抽样分布,并于1915年在*Biometrika*杂志上发表论文“无限总体样本相关系数的频率分布”,该文被称为现代推断统计学的第一篇论文。费希尔还利用多重积分方法,给出了 t 分布的严密推导,使研究小样本函数的精确理论分布中一系列重要结论有了新的开端。在对参数估计的研究中,费希尔在1912年提出了一种重要而普遍的点估计法——极大似然估计,并在之后工作中又加以发展,从而建立了以极大似然估计为中心的点估计理论,在推断总体参数中应用这个方法,不需要有关事前概率的信息,这是数理统计史上的一大突破。这种方法直到目前为止仍是构造估计量的最重要的一种方法。在假设测验的研究方面,费希尔于1923年发展了显著性测验及估计理论,提出了 F 分布和 F 测验。“方差”和“方差分析”两个概念是费希尔于1918年在“孟德尔遗传试验设计间的相对关系”一文中首创。方差分析也称变异分析,其系统研究开始于1923年费希尔与麦凯基合写的“对收获量变化的研究”一文中。在生物统计中,方差分析有着广泛的应用,特别是在费希尔发表了《供研究人员用的统计方法》专著后,对推动和促进农业科学、生物学和遗传学的研究与发展,起到了奠基作用。在试验设计方面,自1923年起,费希尔陆续发表了关于在农业试验中控制试验误差的论文,1925年提出随机区组和正交拉丁方试验设计,并在卢桑姆斯坦德(Rothamsted)农业试验站得到检验与应用。费希尔在创建试验设计理论的过程中,还提出了十分重要的“随机化”原则,他认为这是保证取得无偏估计的有效措施,也是进行可靠的显著性测验的必要基础,1938年费希尔和耶茨(Frank Yates, 1902—1994)合编了随机数字表,利用随机数字表保证总体中每一元素有同等被抽取的机会。此外,费希尔在1930年代初期引进了一种构造区间估计的方法,即“信念推断法”,为某些困难的统计问题提供了简单可行的解法。

美国学者奈曼(Jerzy Neyman, 1894—1981)和伊贡·皮尔逊(Egon Pearson, 1895—1980)进行了统计理论研究,分别于1936和1938年提出一种统计假设测验学说,假设测验作为数学上的最优化问题,对促进统计理论研究和对试验作出正确结论具有非常实用的价值。此外,马瑟(Kenneth Mather, 1911—1990)对生统遗传学,耶茨和犹勒(Udny Yule, 1871—1951)对田间试验设计等都作出了杰出贡献。

国内对生物统计学的研究始于1930年代。1935年王绶编著的《实用生物统计法》是我国出版最早的生物统计专著之一。1942年范福仁编著了《田间试验技术》,这些对推动我国农业生物统计和田间试验方法的应用都产生了很大影响。新中国成立后,统计学的研究与应用出现了崭新的局面,许多生物学研究工作者积极从事统计学理论和实践的应用研究,使生物统计学在农业科学、医学、遗传学和生态学等学科领域发挥了重要作用。应用试验设计方法和统计分析理论,进行农作物品种产量比较试验、病虫害的预测预报、动物饲养试验、饲料配方、毒理试验、动植物资源的调查与分析、动植物育种中遗传资源和亲

子代遗传分析等都取得了较好成果。

近年来,随着后基因组时代分子生物学高通量测序等技术的飞速发展,各种生物学数据呈爆炸性指数增长,如何从大规模、结构复杂的数据中提炼有用的科学信息,对以数据分析为主要内容的生物统计学提出了前所未有的挑战和机遇。面对大数据时代的挑战,统计学家们发展了一系列处理高维数据的参数及非参数的统计学方法,例如最小绝对缩减和变量选择算子(LASSO)、随机搜索变量选择法(SSVS)、岭回归分析、惩罚极大似然及贝叶斯方法等。统计学已成为现代生物学和医学研究中不可缺少的重要部分。未来的生物学工作者,必须掌握足够的生物统计学基础知识与分析技能。

1.2 生物统计学的主要功用

生物统计学的主要功用可以概括为以下几个方面。

(1) 描述试验数据的特征

试验数据通常都是有变异的,若不整理往往庞杂零乱,又不能说明任何问题。统计方法提供了整理资料、化繁为简的科学程序,以及由众多观测值归纳出几个能描述该变量特征的数值计算方法,使试验者从少数的特征数或一些简单的图表中了解到数据中所蕴藏的信息,便于进一步的计算和分析,这类统计处理统称为**描述性统计**,主要内容包括:次数分布的制作和平均数、变异数计算等。

(2) 由样本推论总体

统计上的总体(population),是指在同一组条件下所有成员的某种性状变量的集合,或者说是某一变数的全部可能值的集合。描述总体的特征数称为参数(parameter),它是一个常数,但是只有在总体的全部个体皆属已知时才能计算出来。样本(sample)是指从总体中抽出的一部分。为了使样本能代表总体,并进而可用概率论的方法处理,必须使总体中的每一个个体都有同等的机会被取为样本,这样的样本称为随机样本。描述样本的特征数则称为统计数(statistic)。统计数是一个变数,它会随样本的不同而不同。

生物统计学的一个重要任务是要知道所研究总体的参数,但由于总体中的个体数极为庞大,即使是个体数有限(有限总体),通常亦难以得到其参数。但不管是何种类型的总体,我们总是可以通过随机抽样的方法获得该总体的随机样本,然后由样本的实际结果推断相应总体的特征。这部分内容包含在**统计推断**一章,分为统计假设测验和参数估计两部分。但是,为了正确进行统计推断,首先需要学习一些概率和理论分布的知识以及关于从总体中抽出的样本的统计数的分布问题,这些内容分别包含在**概率与理论分布**一章和**抽样分布**一章。

(3) 分析试验数据的变异

科学试验中一般要求除了被研究的因素具有控制的不同水平(这种不同水平称为处理)外,其余因素作为试验背景对不同处理都应保持一致,这样才能精确地测定处理的效应。但实际上,无论试验条件控制的如何严格,两个处理所处的试验背景决不可能完全相同。因而观察到的效应总是混杂着试验背景这样或那样的不一致而产生的偶然效应。这种偶然效应称为试验误差(experimental error)。因此,试验数据的变异是有不同来源的。最简单的是有两种来源:一是处理不同;二是试验误差。试验观察到的效应究竟是处理的不同造成、还是误差造成,尚有待“测验”,才能使我们的判断建立在科学的基础上。统计

学提供了相应于各种试验设计的分析变异的方法,它能够把各种变异来源的变异一一分解开来,从而帮助试验者发现起主导作用的变异来源。这类分析变异的统计方法称为方差分析。这部分内容包含在**方差分析**一章以及**常用试验设计及其方差分析**一章。

(4) 分析试验数据变量间的相关关系

在农学和生物学的研究中,相关关系是普遍存在的。例如,研究种植密度和作物产量的关系,即使一切可控因素(如品种、播期、施肥、管理等)都能严格控制一致,一定密度下的产量仍将是具有一定概率分布的随机变数,而不是常数,所以我们必须采用适当的统计方法来处理具有相关关系的数据资料。这些内容分别包含在**一元线性回归和相关分析**一章以及**多元线性回归和相关分析**一章。

(5) 提供进行科学试验设计的一些重要原则

为了以较少的人、物、财力取得最多的试验信息和精确地估计处理效应,试验必须有科学的设计。而科学的设计又是和统计方法密切相关的。这是由于:①设计本身的科学性需要应用统计的方法加以研究。例如在田间试验上,抽样方法、处理设置、重复次数以及试验小区的大小、形状和排列方法等,都是要有一定的统计学依据的。②根据试验结果的误差分析,又可为进一步试验或抽样时如何改进试验方法,以减少误差和鉴别出较小的处理效应等,提供理论依据和实践指导。常可看到一些试验资料,由于设计上的某些差错而丧失了大量试验信息,究其原因,多半由于缺乏一定的统计知识。这部分内容主要包含在**常用试验设计及其方差分析**一章。

1.3 本书知识导图

统计方法的应用贯穿于试验研究的始终,对于农业科学和生物学的研究,起到了重要的推动作用。为了更好的掌握有关统计方法,现将生物统计学各章内容及相互关系以导图形式展示如下。

