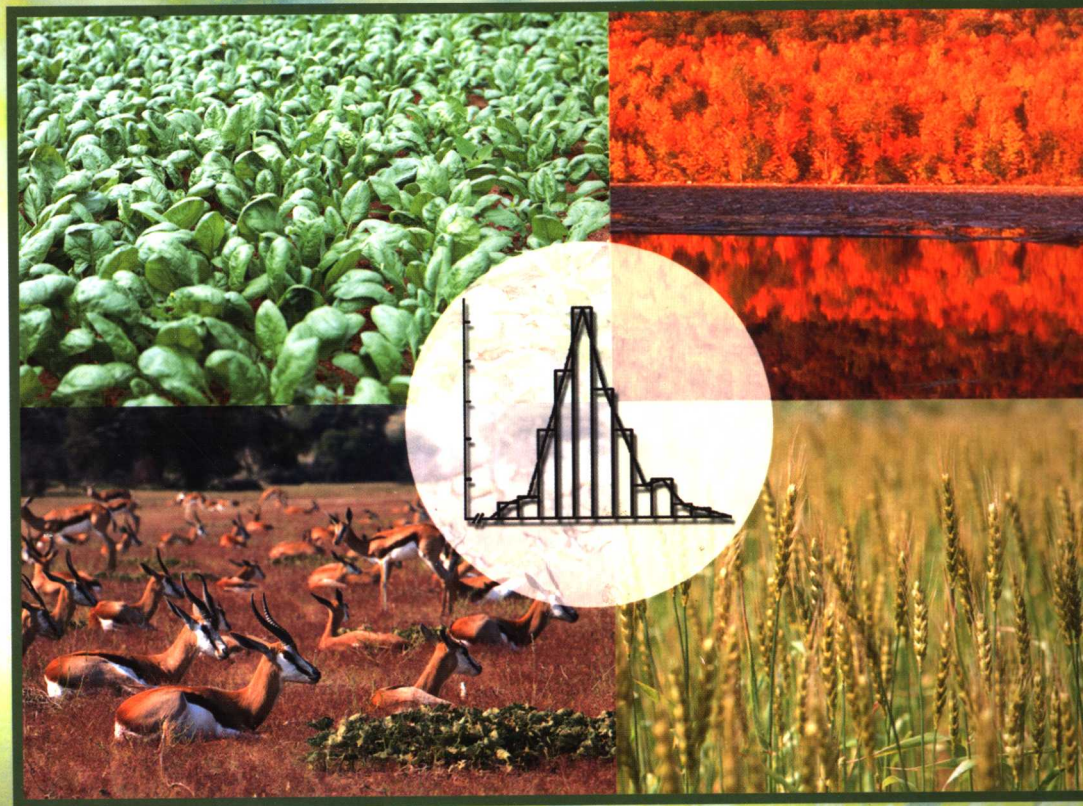


21世纪高等院校教材 · 生物科学系列

生物统计学 (第三版)

■ 李春喜 姜丽娜 邵云 王文林 编著



科学出版社
www.sciencep.com

21 世纪高等院校教材——生物科学系列

生物统计学

(第三版)

李春喜 姜丽娜 邵云 王文林 编著

科学出版社

北京

内 容 简 介

本书较为系统地介绍了生物统计学的基本原理和方法,在简要叙述了生物统计学的产生、发展及其研究对象与作用、生物学研究中试验资料的整理、特征数的计算、概率和概率分布、抽样分布基础上,着重介绍了平均数的统计推断、 χ^2 检验、方差分析、直线回归与相关分析、可直线化的非线性回归分析、协方差分析、多元回归与多元相关分析、逐步回归与通径分析和多项式回归分析,同时对抽样原理和方法、试验设计原理及对比设计、随机区组设计、平衡不完全区组设计、裂区设计、拉丁方设计、正交设计等常用试验设计及其统计分析也进行了详细叙述。在上述内容的基础上,对聚类分析、判断分析、主成分分析、因子分析、典型相关、时间序列分析等多元分析也作了简要介绍。

本书可供综合性大学、师范院校生物类及其相关专业的本科生作为教材使用,也可作为从事生命科学、农业科学、林业科学、医学、畜牧兽医、水产科学等专业的科研工作者、教师和研究生的参考书。

图书在版编目(CIP)数据

生物统计学/李春喜等编著.—3版.—北京:科学出版社,2005.7

(21世纪高等院校教材——生物科学系列)

ISBN 7-03-015155-0

I. 生… II. ①李…②姜…③邵…④王… III. 生物统计-高等学校-教材 IV. Q-332

中国版本图书馆 CIP 数据核字(2005)第 035171 号

责任编辑:周 辉 张晚春/责任校对:鲁 素

责任印制:安春生/封面设计:陈 敬

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

双 音 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

1997年8月第一版 开本:850×1168 1/16

2000年11月第二版 印张:23 1/2

2005年7月第三版 字数:488 000

2005年8月第二次印刷 印数:4 001—8 000

定价:28.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

第三版前言

进入 21 世纪,生命科学已成为现代科学技术发展中最活跃、最具有活力和最富有挑战性的学科领域之一。生命科学从古到今,从对生物体整体的观察到种群、群落、生态圈、生物多样性的研究和利用显微技术对细胞结构的认识、利用分子生物学技术对生命活动机理的探索无不都伴随着数学方法的运用。随着生命科学研究的深入发展和相关学科的融合与渗透,数学方法运用于生物科学的范围将会越来越广泛,越来越深入。数学在生命科学上的应用,总的说来主要包括三个方面:一是应用数学知识来解决生命科学中的实际问题,如应用数学模型研究生物生长发育的过程,利用计算机模拟对其生育过程进行预测与控制等;二是数学方法与生命科学融合交叉产生新的学科,如数量生态学、数量遗传学、生物信息学等;三是从生命科学中提炼出数学问题进行研究,进而发展新的数学理论,如基因如何从时间、空间上来精确发育过程,多个核苷酸是如何构成功能基因的等。可见,生物数学对生命科学的发展是必不可少的。生物统计学作为生物数学的重要分支,在生命科学研究与探索的过程中发挥了巨大的推动作用。因此,各高等学校生物类专业都将生物统计学作为一门重要的基础课程纳入到教学体系中,促进了新型综合性生物学高级人才的培养。

为适应 21 世纪生命科学发展和生物学人才培养的要求,在本书 1997 年第一版、2000 年第二版的基础上,对全书内容重新进行了编排和审核,增加了部分内容,修订和改正了原书中存在的一些错误。与前两版相比,本书突出了以下几个特点:①内容更加丰富,增加了平衡不完全区组设计、倒数函数曲线、通径分析等内容;②编排更科学,将全书分解为 16 章,各章节的安排更加注重了内容的循序渐进;③针对性更加明确,内容突出了本书主要作为生物类及相关专业教材这个重点,更换了第二版中部分针对性不强的例子,对各章思考练习题及其答案重新进行了核对,对相关名词与术语增加了英文标注,并重新编排了中英文对照索引,以便于学习和检索。

在本书第三版的修订和出版过程中,得到了科学出版社和河南师范大学的大力支持,周辉先生在书稿的编审方面做了大量工作,周其源、鲁旭阳、苗永平、李丹丹、侯小丽、张蓓蓓、肖长新、冯淑利、刘玲 9 位研究生承担了书稿的部分核对工作和习题答案的核对工作,在此一并感谢。

本书能够第三次出版,是广大读者支持和厚爱的结果。希望各位读者在阅读和使用过程中对本书的谬误和不妥之处给予批评指正。

李春喜
于河南师范大学
2005 年 1 月

第二版前言

近代生物学的发展有两个显著的特点：一个是向微观方向的发展，通过显微技术对生物的细胞和细胞结构进行深入研究；另一个是向宏观方向的发展，从生物体的器官、整体到种群、群落、生态圈进行研究。这两个发展方向共同趋势都是需要运用数学方法对生物体、生物器官、细胞及分子结构所观察和实验的结果进行综合分析，研究各种因素间的相互作用，通过建立数学模型，并对模型进行数学推理，来发现和解释新的生命现象。随着科学的发展，数学方法在生物学研究中的应用会越来越广泛，其作用也将会越来越重要。因此，一门新兴的边缘性学科——生物数学也就应运而生了。在生物数学领域中，生物统计学是应用最早也最广泛的一门学科，起先是应用生物学科，后来是纯生物学科，它们都对生物统计学的应用有一定的深度和广度，特别是信息科学的迅猛发展和计算机的迅速普及，为在生物学研究中运用生物统计学原理和方法提供了更为广阔的空间。生物统计学作为基础性工具课程，越来越为高校生物类专业所重视。

本书第一版出版以来，在部分高校生物类专业作为教材使用，一些科学研究单位也作为工具书应用，总的说来反映是良好的。有不少读者对本书进行了仔细研究，提出了不少修改意见，对本书第一版中出现的错误诚恳地提出了批评。根据读者的意见和生物统计学应用的需要，这次修订，对第一版各章节作了较大幅度的调整，将全书分为十四章，补充了拉丁方设计和裂区设计两种试验设计方法，将抽样原理和方法、常用试验设计及其统计分析放在了可直线化的非线性回归分析之后进行介绍，使章节编排体系更符合读者学习的要求。书中还增加了对全文关键词汇和术语的索引，并在书后附上了各章部分思考练习题答案。同时，对本书第一版中的不妥和错误之处进行了订正，更换了部分引用例题，以使这些例题更能反映本章内容和便于读者的学习和理解。

在本书第二版的修订和出版过程中，得到了河南省科委和科学出版社的大力支持，尚玉磊、邱宗波、董媛、史小琴同志承担了部分书稿的校对工作，在此一并表示感谢。

由于作者水平有限，谬误和不当之处，敬请读者批评指正！

李春喜

于河南师范大学

2000年4月

第一版前言

生物统计学是运用数理统计的原理和方法来分析和解释生物界各种现象和试验调查资料的一门科学。随着生物学的不断发展,对生物体的研究和观察已不再局限于定性的描述,而是需要从大量调查和测定数据中,应用统计学方法,分析和解释其数量上的变化,以正确制订试验计划,科学地对试验结果进行分析,从而作出符合科学实际的推断。目前,生物统计学在农学、林学、畜牧、医药、卫生、生态、环保等领域已有广泛应用,但在纯生物学研究方面的应用,不管是在深度上还是广度上都不及上述领域。有鉴于此,在生物学研究中,迫切需要加强生物统计学的应用,对高校生物类专业,它也是一门应被十分重视的“工具”课程。本书正是为了满足这些需要而编写的。

本书的写作是在作者多年从事生物统计学教学和应用研究的基础上完成的。书中的内容主要侧重于各种统计方法的应用,在统计原理方面,一般只作概念上的介绍和公式的简单推导,对有些较复杂的统计公式则只给出公式,其目的主要是为了让读者不但对统计学原理有较全面的了解,更重要的是结合实例了解和掌握各种常用统计方法。在本书的安排上,全书共分十二章,概括起来主要有五个方面:第一章至第三章介绍统计和概率的基础知识,包括生物统计学的概念和内容、数据的搜集与整理、平均数和变异数的计算、概率和概率分布等;第四章、第五章介绍统计推断,包括样本平均数的检验、样本频率的检验、方差同质性检验、非参数检验和 χ^2 检验;第六章至第九章介绍统计分析方法,主要内容有方差分析、直线回归与相关分析、可直线化的曲线回归分析、多元回归与相关分析、逐步回归分析、多项式回归、协方差分析;第十章、第十一章介绍抽样与试验设计,主要包括抽样误差估计、抽样方法、抽样方案制订及常见的试验设计,如对比设计、随机区组设计、正交设计及其相应的统计分析方法;第十二章对近年来应用越来越多的多元统计分析进行了简单介绍。每章都附有一定数量的思考练习题,供读者参考。

本书中的例子主要有两个来源,一个是近年来有关生物学、农学、林学、医学、畜牧、水产、环保等领域或学科的实际研究资料,另一个是有关著作中的一些例题。崔党群教授在百忙中通审了全书,并提出了富有建设性的建议。贾玉书同志承担了本书的大部分绘图工作。姜丽娜同志在本书的录排中做了大量工作。在本书的出版过程中,得到了科学出版社的大力支持,特别是张晓春同志在书稿的编审和发行方面做了大量工作,在此一并表示谢意。

本书通俗易懂,具有一定的深度和广度,适合生物学、农学、医学、畜牧、水产、环保等领域或学科的科学工作者阅读,也可供本、专科院校生物类专业作为教材使用。

由于作者水平的限制和资料占有的局限性,本书难免会有错误和不妥之处,敬请读者批评指正,以便日后修订完善。

李春喜 王文林

1997年3月

目 录

第三版前言

第二版前言

第一版前言

| | |
|---------------------------------|------|
| 第一章 概论 | (1) |
| 第一节 生物统计学的概念 | (1) |
| 第二节 生物统计学的内容与作用 | (2) |
| 第三节 生物统计学发展概况 | (3) |
| 一、古典记录统计学 | (3) |
| 二、近代描述统计学 | (3) |
| 三、现代推断统计学 | (4) |
| 第四节 常用统计学术语 | (5) |
| 一、总体、个体与样本 | (5) |
| 二、变量与常数 | (5) |
| 三、参数与统计数 | (6) |
| 四、效应与互作 | (6) |
| 五、误差与错误 | (6) |
| 六、准确性与精确性 | (7) |
| 思考练习题 | (7) |
| 第二章 试验资料的整理与特征数的计算 | (8) |
| 第一节 试验资料的搜集与整理 | (8) |
| 一、试验资料的类型 | (8) |
| 二、试验资料的搜集 | (9) |
| 三、试验资料的整理 | (10) |
| 第二节 试验资料特征数的计算 | (16) |
| 一、平均数 | (16) |
| 二、变数 | (19) |
| 思考练习题 | (22) |
| 第三章 概率与概率分布 | (24) |
| 第一节 概率基础知识 | (24) |
| 一、概率的概念 | (24) |
| 二、概率的计算 | (25) |
| 三、概率分布 | (27) |
| 四、大数定律 | (29) |
| 第二节 几种常见的理论分布 | (30) |
| 一、二项分布 | (30) |
| 二、泊松分布 | (34) |
| 三、正态分布 | (35) |

| | |
|---|-------------|
| 第三节 统计数的分布 | (40) |
| 一、抽样试验与无偏估计 | (41) |
| 二、样本平均数的分布 | (42) |
| 三、样本平均数差数的分布 | (43) |
| 四、 t 分布 | (45) |
| 五、 χ^2 分布 | (46) |
| 六、 F 分布 | (47) |
| 思考练习题 | (48) |
| 第四章 统计推断 | (49) |
| 第一节 假设检验的原理与方法 | (49) |
| 一、假设检验的概念 | (49) |
| 二、假设检验的步骤 | (50) |
| 三、双尾检验与单尾检验 | (52) |
| 四、假设检验中的两类错误 | (53) |
| 第二节 样本平均数的假设检验 | (54) |
| 一、大样本平均数的假设检验—— u 检验 | (54) |
| 二、小样本平均数的假设检验—— t 检验 | (57) |
| 第三节 样本频率的假设检验 | (62) |
| 一、一个样本频率的假设检验 | (63) |
| 二、两个样本频率的假设检验 | (64) |
| 第四节 参数的区间估计与点估计 | (66) |
| 一、参数区间估计与点估计的原理 | (66) |
| 二、一个总体平均数 μ 的区间估计与点估计 | (67) |
| 三、两个总体平均数差数 $\mu_1 - \mu_2$ 的区间估计与点估计 | (68) |
| 四、一个总体频率 p 的区间估计与点估计 | (69) |
| 五、两个总体频率差数 $p_1 - p_2$ 的区间估计与点估计 | (70) |
| 第五节 方差的同质性检验 | (71) |
| 一、一个样本方差的同质性检验 | (71) |
| 二、两个样本方差的同质性检验 | (72) |
| 三、多个样本方差的同质性检验 | (72) |
| 第六节 非参数检验 | (73) |
| 一、符号检验法 | (74) |
| 二、秩和检验法 | (75) |
| 思考练习题 | (78) |
| 第五章 χ^2 检验 | (80) |
| 第一节 χ^2 检验的原理与方法 | (80) |
| 第二节 适合性检验 | (82) |
| 第三节 独立性检验 | (85) |
| 一、 2×2 列联表的独立性检验 | (85) |
| 二、 $2 \times c$ 列联表的独立性检验 | (87) |
| 三、 $r \times c$ 列联表的独立性检验 | (88) |
| 思考练习题 | (89) |

| | |
|---------------------------------|-------|
| 第六章 方差分析 | (91) |
| 第一节 方差分析的基本原理 | (92) |
| 一、相关术语 | (92) |
| 二、方差分析的基本原理 | (93) |
| 三、数学模型 | (94) |
| 四、平方和与自由度的分解 | (95) |
| 五、统计假设的显著性检验—— F 检验 | (98) |
| 六、多重比较 | (99) |
| 第二节 单因素方差分析 | (104) |
| 一、组内观测次数相等的方差分析 | (104) |
| 二、组内观测次数不相等的方差分析 | (106) |
| 第三节 二因素方差分析 | (108) |
| 一、无重复观测值的二因素方差分析 | (108) |
| 二、具有重复观测值的二因素方差分析 | (111) |
| 第四节 多因素方差分析 | (118) |
| 第五节 方差分析缺失数据的估计 | (122) |
| 一、缺失一个数据的估计方法 | (123) |
| 二、缺失两个数据的估计方法 | (123) |
| 第六节 方差分析的基本假定和数据转换 | (124) |
| 一、方差分析的基本假定 | (124) |
| 二、数据转换 | (125) |
| 思考练习题 | (128) |
| 第七章 抽样原理与方法 | (131) |
| 第一节 抽样误差的估计 | (131) |
| 一、样本平均数的标准误和置信区间 | (132) |
| 二、样本频率的标准误和置信区间 | (132) |
| 第二节 样本容量的确定 | (133) |
| 一、平均数资料样本容量的确定 | (133) |
| 二、频率资料样本容量的确定 | (134) |
| 三、成对资料和非成对资料样本容量的确定 | (134) |
| 第三节 抽样的基本方法 | (136) |
| 一、随机抽样 | (136) |
| 二、顺序抽样 | (139) |
| 三、典型抽样 | (139) |
| 第四节 抽样方案的制定 | (139) |
| 一、抽样调查的目的和指标要求 | (140) |
| 二、确定调查对象 | (140) |
| 三、确定抽样调查的方法 | (140) |
| 四、确定样本容量和抽样分数 | (140) |
| 五、总体单位编号 | (141) |
| 六、编制抽样调查表 | (141) |
| 七、抽样调查的组织工作 | (141) |
| 思考练习题 | (141) |

| | |
|--------------------------------|-------|
| 第八章 试验设计及其统计分析(一) | (143) |
| 第一节 试验设计的基本原理 | (143) |
| 一、试验设计的意义 | (143) |
| 二、生物学试验的基本要求 | (144) |
| 三、试验设计的基本要素 | (145) |
| 四、试验误差及其控制途径 | (145) |
| 五、试验设计的基本原则 | (147) |
| 第二节 对比设计及其统计分析 | (148) |
| 一、对比设计 | (148) |
| 二、对比设计试验结果的统计分析 | (150) |
| 第三节 随机区组设计及其统计分析 | (151) |
| 一、随机区组设计 | (151) |
| 二、随机区组设计试验结果的统计分析 | (151) |
| 第四节 平衡不完全区组设计及其统计分析 | (158) |
| 一、平衡不完全区组设计 | (158) |
| 二、平衡不完全区组设计试验结果的统计分析 | (160) |
| 思考练习题 | (164) |
| 第九章 试验设计及其统计分析(二) | (166) |
| 第一节 裂区设计及其统计分析 | (166) |
| 一、裂区设计 | (166) |
| 二、裂区设计试验结果的统计分析 | (167) |
| 第二节 拉丁方设计及其统计分析 | (173) |
| 一、拉丁方设计 | (173) |
| 二、拉丁方设计试验结果的统计分析 | (174) |
| 第三节 正交设计及其统计分析 | (177) |
| 一、正交表及其特点 | (177) |
| 二、正交试验的基本方法 | (179) |
| 三、正交设计试验结果的统计分析 | (181) |
| 思考练习题 | (184) |
| 第十章 直线回归与相关分析 | (186) |
| 第一节 回归和相关的概念 | (187) |
| 第二节 直线回归 | (188) |
| 一、直线回归方程的建立 | (188) |
| 二、直线回归的数学模型和基本假定 | (191) |
| 三、直线回归的假设检验 | (191) |
| 四、直线回归的区间估计 | (194) |
| 五、直线回归的应用及注意事项 | (197) |
| 第三节 直线相关 | (198) |
| 一、相关系数和决定系数 | (198) |
| 二、相关系数的假设检验 | (200) |
| 三、相关系数的区间估计 | (201) |
| 四、应用直线相关的注意事项 | (202) |

| | |
|--|--------------|
| 第四节 直线回归和直线相关的关系 | (202) |
| 一、区别 | (202) |
| 二、联系 | (203) |
| 思考练习题 | (204) |
| 第十一章 可直线化的非线性回归分析 | (205) |
| 第一节 非线性回归的直线化 | (206) |
| 一、曲线类型的确定 | (206) |
| 二、数据变换的方法 | (206) |
| 第二节 倒数函数曲线 | (207) |
| 第三节 指数函数曲线 | (210) |
| 第四节 对数函数曲线 | (213) |
| 第五节 幂函数曲线 | (215) |
| 第六节 Logistic 生长曲线 | (218) |
| 一、Logistic 生长曲线的由来和基本特征 | (218) |
| 二、Logistic 生长曲线方程的配合 | (219) |
| 思考练习题 | (221) |
| 第十二章 协方差分析 | (222) |
| 第一节 协方差分析的作用 | (223) |
| 一、降低试验误差,实现统计控制 | (223) |
| 二、分析不同变异来源的相关关系 | (223) |
| 三、估计缺失数据 | (224) |
| 第二节 单向分组资料的协方差分析 | (224) |
| 一、计算各项变异的平方和、乘积和与自由度 | (226) |
| 二、检验 x 和 y 是否存在直线回归关系 | (227) |
| 三、检验矫正平均数 $\bar{y}_{i(x=\bar{x})}$ 间的差异显著性 | (227) |
| 四、矫正平均数 $\bar{y}_{i(x=\bar{x})}$ 间的多重比较 | (229) |
| 第三节 两向分组资料的协方差分析 | (230) |
| 一、乘积和与自由度的分解 | (231) |
| 二、检验 x 和 y 是否在线性回归关系 | (233) |
| 三、检验矫正平均数 $\bar{y}_{i(x=\bar{x})}$ 间的差异显著性 | (233) |
| 第四节 协方差分析的数学模型和基本假定 | (234) |
| 一、协方差分析的数学模型 | (234) |
| 二、协方差分析的基本假定 | (234) |
| 思考练习题 | (235) |
| 第十三章 多元线性回归与多元相关分析 | (236) |
| 第一节 多元线性回归分析 | (236) |
| 一、多元线性回归模型 | (236) |
| 二、多元线性回归方程的建立 | (237) |
| 三、多元线性回归的假设检验和置信区间 | (242) |
| 第二节 多元相关分析 | (247) |
| 一、多元相关分析 | (247) |
| 二、偏相关 | (248) |

| | |
|-----------------------|-------|
| 思考练习题 | (251) |
| 第十四章 逐步回归与通径分析 | (253) |
| 第一节 逐步回归分析 | (253) |
| 一、逐个淘汰不显著自变量的回归方法 | (254) |
| 二、逐个选入显著自变量的回归方法 | (259) |
| 第二节 通径分析 | (263) |
| 一、通径与通径系数的概念 | (263) |
| 二、通径系数的求解方法 | (264) |
| 三、通径分析的假设检验 | (266) |
| 思考练习题 | (269) |
| 第十五章 多项式回归分析 | (271) |
| 第一节 多项式回归的数学模型 | (271) |
| 第二节 多项式回归方程的建立 | (272) |
| 一、多项式回归方程的建立与求解 | (272) |
| 二、多项式回归方程的图示 | (275) |
| 第三节 多项式回归方程的假设检验 | (275) |
| 第四节 相关指数 | (277) |
| 第五节 正交多项式回归分析 | (277) |
| 一、正交多项式回归分析原理 | (277) |
| 二、正交多项式回归分析示例 | (279) |
| 思考练习题 | (281) |
| 第十六章 多元统计分析简介 | (282) |
| 第一节 数据矩阵与相似系数 | (282) |
| 一、数据矩阵 | (282) |
| 二、相似系数 | (283) |
| 三、距离系数 | (286) |
| 第二节 聚类分析 | (287) |
| 一、类与类之间的距离 | (288) |
| 二、系统聚类的分类过程 | (289) |
| 三、系统聚类法的统一模型和方法评价 | (290) |
| 第三节 判别分析 | (291) |
| 第四节 主成分分析 | (294) |
| 第五节 因子分析 | (299) |
| 一、因子分析的数学模型 | (299) |
| 二、因子分析的计算过程 | (300) |
| 第六节 典型相关分析 | (304) |
| 一、典型相关分析的数学模型 | (305) |
| 二、典型相关系数的检验 | (306) |
| 三、典型相关分析的计算过程 | (306) |
| 第七节 时间序列分析 | (308) |
| 一、平稳时间序列的线性外推法 | (309) |
| 二、显著性相关函数值预报法 | (312) |

| | |
|--------------------------|-------|
| 思考练习题 | (313) |
| 附表 | (315) |
| 附表 1 正态分布的累积函数 $F(u)$ 值表 | (315) |
| 附表 2 正态离差 (u) 值表(双尾) | (317) |
| 附表 3 t 值表(双尾) | (317) |
| 附表 4 χ^2 值表(右尾) | (318) |
| 附表 5 F 值表(右尾) | (319) |
| 附表 6 符号检验表 | (323) |
| 附表 7 秩和检验表 | (323) |
| 附表 8 新复极差检验 SSR 值表 | (324) |
| 附表 9 q 值表(双尾) | (325) |
| 附表 10 平衡不完全区组设计参数表 | (326) |
| 附表 11 平衡不完全区组设计表 | (327) |
| 附表 12 正交拉丁方表 | (329) |
| 附表 13 常用正交表 | (330) |
| 附表 14 r 与 R 的临界值表 | (339) |
| 附表 15 正交多项式系数表 | (340) |
| 思考练习题答案 | (344) |
| 索引 | (348) |
| 主要参考文献 | (355) |

第一章

概 论

本章 提要

生物统计学是数理统计的原理和方法在生命科学领域的具体应用,它是运用数理统计原理和方法对生物有机体开展调查和试验,目的是以样本的特征来估计总体的特征,对所研究总体进行合理的推论,得到对客观事物本质和规律性的认识。生物统计学的主要研究内容包括试验设计和统计分析两大部分,其作用主要有四个方面:提供整理、描述数据资料的科学方法并确定其数量特征,判断试验结果的可靠性,提供由样本推断总体的方法,提供试验设计的原则。本章还介绍了生物统计学的发展概况和六组统计学常用术语。

第一节 生物统计学的概念

生物统计学(biostatistics)是数理统计(mathematical statistics)在生物学研究中的应用,它是用数理统计的原理和方法来分析和解释生物界各种现象和试验调查资料的科学,属于应用统计学的一个分支。随着生物学研究的不断发展,运用统计学方法来认识、推断和解释生命过程中的各种现象,也越来越广泛。尽管生物统计在应用过程中曾经受到过一些批评,但绝大多数生物学家、农学家、园艺学家、育种学家、畜牧学家、医学工作者以及人口学家还是在自己的研究领域越来越普遍地应用生物统计分析方法,并把它作为学科自身发展的需要。

生物学研究的对象是生物有机体,与非生物相比,它具有更加特殊的复杂性。生物有机体的生长发育、生理活动、生化变化及有机体受外界各种随机因素的影响等,都使生物学研究的试验结果有较大的差异性,这种差异性往往会掩盖生物体本身的特殊规律。在生物学研究中,大量试验资料内在的规律性,也容易被杂乱无章的数据所迷惑,从而被人们忽视。因此,在生物学研究中,应用生物统计学就显得特别重要。生物学研究的实践证明,只有正确地应用生物统计原理和分析方法对生物学试验进行合理设计,对数据进行客观分析,才能得出科学的结论。

在对事物的研究过程中,人们往往是通过某事物的一部分(样本),来估计事物全部(总体)的特征,目的是为了以样本的特征对未知总体进行推断,从特殊推导一般,对所研

究的总体作出合乎逻辑的推论,得到对客观事物的本质和规律性的认识。在生物学研究中,我们所期望的是总体,而不是样本。但是在具体的试验过程中,我们所得到的却是样本而不是总体。因此,从某种意义上讲,生物统计学是研究生命过程中以样本来推断总体的一门学科。

生物统计学是在生物学研究过程中,逐渐与数学的发展相结合所形成的,它是应用数学的一个分支,属于生物数学的范畴。生物统计学的过程就是把数学的语言引入具体的生命科学领域,把具体生命科学领域中亟待研究的问题抽象为数学问题的过程。它是数学的概率论和数理统计为基础,这其中要涉及到数列、排列、组合、矩阵、微积分等知识。作为一门工具课,生物统计学一般不过多讨论数学原理,而主要偏重于统计原理的介绍和具体分析方法的应用。

第二节 生物统计学的内容与作用

生物统计学的基本内容,概括起来主要包括试验设计(experimental design)和统计分析(statistical analysis)两大部分。在试验设计中,主要介绍试验设计的有关概念、试验设计的基本原则、试验设计方案的制定、常用试验设计方法,其中主要的有对比设计、随机区组设计、平衡不完全区组设计、拉丁方设计、裂区设计以及正交设计等。在统计分析中,主要包括数据资料的搜集和整理、数据特征数的计算、统计推断、方差分析、回归和相关分析、协方差分析、主成分分析、聚类分析等。

从生物统计学的基本作用上来讲,其任务可以概括为以下几个方面:

(1) 提供整理和描述数据资料的科学方法,确定某些性状和特性的数量特征。一批试验或数据资料,若不整理则杂乱无章,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,它可以从众多的数据资料中,归纳出几个特征数或绘制出一定形式的图表,使研究者从少数的特征数或一些简单的图表中了解大批资料所蕴藏的信息。

(2) 判断试验结果的可靠性。一般在试验中要求除试验因素以外,其他条件都应控制一致,但在实践中无论试验条件控制得如何严格,其试验结果总是受试验因素和其他偶然因素的影响。偶然因素的影响就是造成试验误差的重要原因。要正确判断一个试验结果是由试验因素造成的还是由试验误差造成的,就必须应用统计分析方法。

(3) 提供由样本推断总体的方法。试验的目的在于认识总体规律,但由于总体庞大,一般无法实施,在研究过程中都是抽取总体中的部分作为样本,用统计方法以样本来推断总体的规律性。这是生物统计学的精髓所在,也是其成为一门科学的缘由。在这种推断中,统计学原理和方法起到了理论上的保证作用。

(4) 提供试验设计的一些重要原则。为了以较少的人力、物力和财力取得较多的试验信息和较好的试验结果,在一些生物学研究中,就需要科学地进行试验设计,如对样本容量的确定、抽样方法、处理设置、重复次数的确定以及试验的安排等,都必须以统计学原理为依据。从统计分析和试验设计的关系来看,统计学原理可以为试验设计提供合理的依据,而试验设计又是统计分析方法的进一步运用。以统计学原理为指导,进行科学合理的试验设计,可以在较少人力、物力、时间和条件下,得出可靠而准确的数据和信息。以往

有一些试验资料,由于设计不当而丧失了大量的试验信息,究其原因多是由于缺乏科学的统计学知识,从而使试验的效率大大降低。尽管统计学原理和分析方法对试验设计有着积极的指导意义,但它绝对不可能代替试验设计。如果试验目的、要求不明确,设计不合理,试验条件不合适,统计数据不准确,这种试验绝对不会成功,统计学原理和分析方法也不可能挽救试验的这种失败。

第三节 生物统计学发展概况

现代统计学起源于17世纪,它主要有两个来源,一是政治科学的需要,二是当时贵族阶层对机率数学理论很感兴趣而发展起来的。另外,研究天文学的需要也促进了统计学的发展。统计学的发展过程大体上可以分为古典记录统计学、近代描述统计学和现代推断统计学三个阶段。

一、古典记录统计学

古典记录统计学(record statistics)的形成是在17世纪中叶至19世纪中叶。在最初兴起时,通过用文字或数字如实记录与分析国家社会经济状况,初步建立了统计研究的方法和规则。到概率论被引进之后,才逐渐成为一项成熟的方法。

瑞士数学家 J. Bernoulli (1654~1705) 系统论证了大数定律。后来, J. Bernoulli 的后代 D. Bernoulli (1700~1782) 将概率论的理论应用到医学和人类保险。

法国天文学家、数学家 P. S. Laplace (1749~1827) 发展了概率论的研究,建立了严密的概率数学理论,并在天文学、物理学的研究中推广应用了概率论。他研究了最小二乘法,提出了“拉普拉斯定理”(中心极限定理的一部分),初步建立了大样本推断的理论基础,为后人开创了抽样调查的方法。

正态分布理论对研究生物统计的理论是十分重要的,它最早是由 De Moivre 于 1733 年发现的。德国天文学和数学家 G. F. Gauss (1777~1855) 在研究观察误差理论时,也独立推导出测量误差的概率分布方程,提出了“误差分布曲线”。这条分布曲线称为 Gauss 分布曲线,也就是正态分布曲线。Gauss 对统计学的另一重要贡献是首先提出了统计上非常重要的最小二乘法。

二、近代描述统计学

近代描述统计学(description statistics)的形成是在19世纪中叶至20世纪上半叶,这个时期也是统计学用于生物学研究的开始和发展时期。

1870年,英国遗传学家 F. Galton (1822~1911) 在19世纪末应用统计方法研究人种特性和遗传,分析父母与子女的变异,探索其遗传规律,提出了相关与回归的概念,开辟了生物学研究的新领域。尽管他的研究当时并未成功,但由于他开创性地将统计方法应用于生物学研究,后人推崇他为生物统计学的创始人。

在此之后, Galton 和他的继承人 K. Pearson(1857~1936)经过共同努力于 1895 年成立了伦敦大学生物统计实验室, 1889 年发表了《自然界的遗传》一文, 并于 1901 年创办了 *Biometrika*(生物统计学报或称为生物计量学报)权威杂志。在该杂志的创刊词中, Galton 和 Pearson 首次为他们所运用的统计方法论明确提出了“生物统计”(biometry)一词, Galton 解释为: 所谓生物统计学, 就是应用于生物学科中的现代统计方法。在《自然界的遗传》一文中, K. Pearson 首先提出了相关与回归分析问题, 并给出了计算简单相关系数和复相关系数的计算公式。K. Pearson 在研究样本误差效应时, 提出了测量实际值与理论值之间偏离度的指标卡方(χ^2)的检验问题, 它在属性统计分析中有着广泛的应用。例如, 遗传学孟德尔豌豆杂交试验中, 高豌豆品种与低豌豆品种杂交后, 它的后代理论比率应该是高 3: 低 1, 但实际后代数是否符合 3: 1, 需用 χ^2 进行检验。

三、现代推断统计学

现代推断统计学(inference statistics)的形成是在 20 世纪初至 20 世纪中叶。随着社会科学和自然科学领域研究的不断深入, 各种事物与现象之间的表面关系及未知的一些数量变化, 要求采用推断的方法来掌握事物之间的真正联系并对事物进行预测。从描述统计学到推断统计学, 这是统计学发展过程中的一个巨大飞跃。

K. Pearson 的学生 W. S. Gosset(1876~1937)对样本标准差进行了大量研究, 于 1908 年以笔名“Student”在 *Biometrika* 上发表论文《平均数的概率误差》, 创立了小样本检验代替大样本检验的理论和方法, 即 t 分布和 t 检验法。 t 检验已成为当代生物统计工作的基本工具之一, 它也为多元分析的理论形成和应用奠定了基础。因此, 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。

英国统计学家 R. A. Fisher(1890~1962)于 1923 年发展了显著性检验及估计理论, 提出了 F 分布和 F 检验, 创立了方差和方差分析。在从事农业试验及数据分析研究时, 他提出了随机区组法、拉丁方法和正交试验设计。1925 年, Fisher 发表了《试验研究工作中的统计方法》, 对方差分析及协方差分析进一步作了完整的解释, 从而推动和促进了农业科学、生物学和遗传学的研究与发展。自 20 年代 Fisher 的方差分析问世以来, 各种数理统计方法不但在实验室中成为研究人员的析因工具, 而且在田间试验、饲养试验、临床试验等农学、医学和生物学领域也得到了广泛应用。

Newman(1894~1981)和 S. Pearson 进行了统计理论的研究工作, 分别于 1936 年和 1938 年提出了一种统计假设检验学说。假设检验和区间估计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。

另外, P. C. Mabeilinrobis 对作物抽样调查、A. Waecl 对序贯抽样、Finney 对毒理统计、K. Mather 对生统遗传学、F. Yates 对田间试验设计等都做出了杰出的贡献。

我国对生物统计学的应用始于 1913 年顾澄教授翻译的统计名著《统计学之理论》。这是英国统计学家在 1911 年出版的关于描述统计学的著作, 也是英美数理统计学传入中国的开始。中华人民共和国成立以后, 许多生物科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、医学科学、生物学、遗传学、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