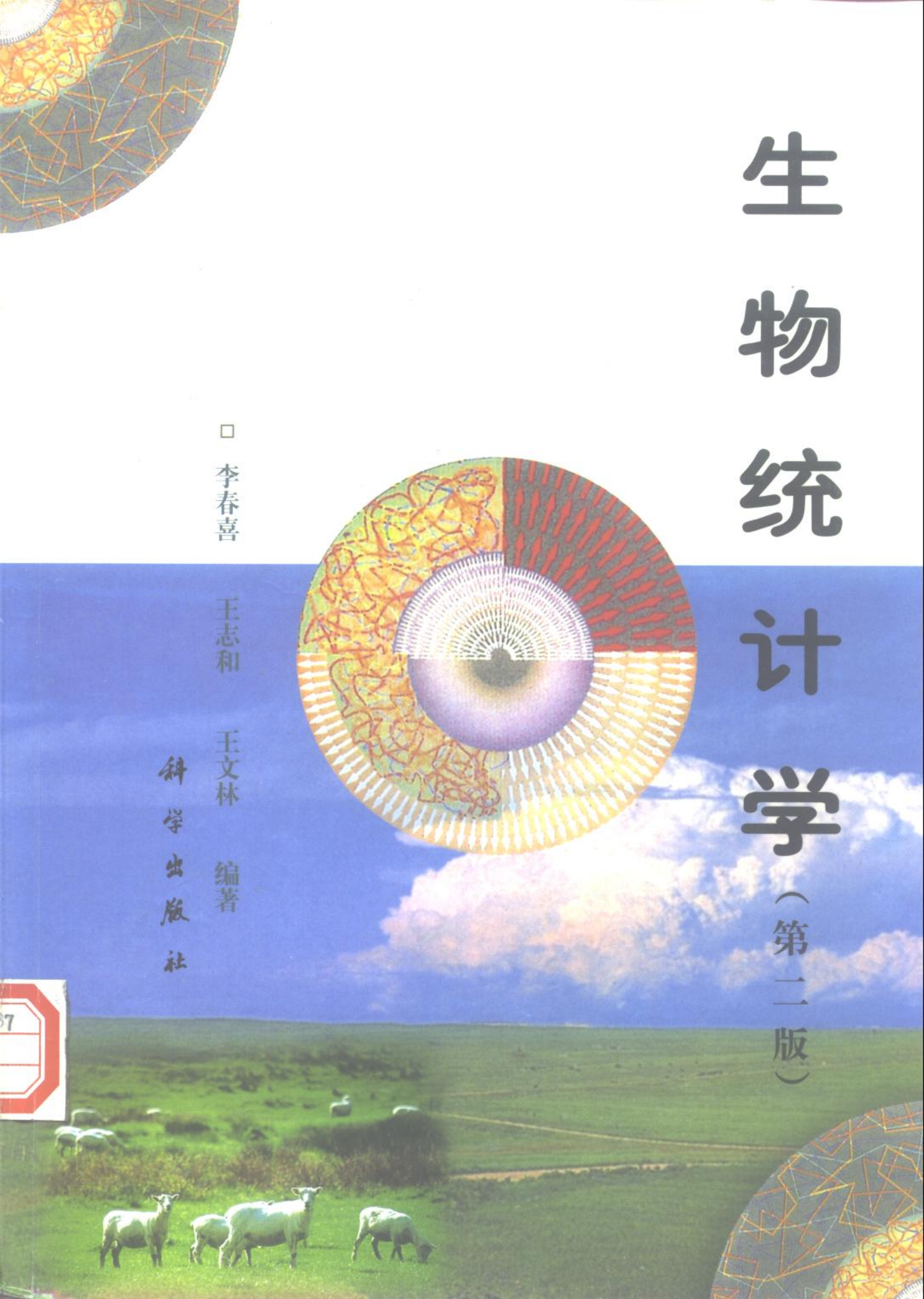
生物统计学（第二版）

李春喜
王志和
王文林
编著

科学出版社



生 物 统 计 学

（第 二 版）

李春喜　王志和　王文林　编著

科 学 出 版 社

2000

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和多项式回归分析，同时对抽样原理和方法、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。

本书通俗易懂，具有一定的深度和广度，适合从事生命科学、农业科学、医学工作者阅读，也可供本、专科院校生物类专业作为教材使用。

图书在版编目(CIP)数据

生物统计学(第二版)/李春喜等编著. -北京：科学出版社，2000.6
ISBN7-03-006069-5

Ⅰ.生… Ⅱ.李… Ⅲ.生物统计 Ⅳ.Q-724

中国版本图书馆 CIP 数据核字(97)第 07412 号

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
北京双青印刷厂 印刷
新华书店北京发行所发行 各地新华书店经售
*
1997 年 8 月第 一 版 开本：787×1092 1/16
2000 年 6 月第 二 版 印张：19 1/4
2000 年 6 月第三次印刷 字数：439 000
印数：5 551—12 550
定价：28.00 元
(如有印装质量问题，我社负责调换〈环伟〉)

第二版前言

近代生物科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。这两个发展方向的共同趋势，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各专业所重视。

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研究的工具书应用，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，以正确制定试验计划，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作了大量工作，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

第一章

概　论

第一节　生物统计学的概念

生物统计学是数理统计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。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，并把它变为学科自身发展的需要。

生物学研究的对象是复杂的生物有机体，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环境因素的影响等，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，容易被人们所忽视。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）的特征的，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，得到对客观事物本质的和规律性的认识。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。生物统计学以数学的概率论为基础，也涉及到数列、排列、组合、矩阵、微积分等知识。生物统计学作为一门重要的工具课，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　生物统计学的主要内容

生物统计学的基本内容，概括起来主要包括试验设计和统计分析两大部分。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、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出一定形式的图表，使试验研究者能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。一个试验结果，是由试验因素造成的还是试验误差造成的，要正确判断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，在这种推断中，统计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时，可以使在较少人力、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，究其原因多半是由于缺乏一定的统计知识，使试验的效率大大降低。当然，统计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，这种试验也绝对不会成功，统计原理和分析方法都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。瑞士数学家 J. Bernouli(1654～1705)系统论证了大数定律。后来，J. Bernouli 的后代 D. Bernouli(1700～1782)将概率论的理论应用到医学和人类保险。

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 Moiver 于 1733

年发现的，后来德国天文学和数学家 Gauss(1777～1855)在研究观察误差理论时，也独立发现了正态分布的理论方程，因此，常有人将正态分布称为 Gauss 分布。

统计学用于生物学的研究，开始于 19 世纪末。1870 年，英国遗传学家 Galton(1822～1911)在 19 世纪末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，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. Plarson(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，于 1889 年发表了《自然的遗传》一书。在该书中，K. Plarson 首先提出了回归分析问题，并给出了计算简单相关系数和复相关系数的计算公式。K. Plarson 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数卡方（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，在遗传上孟德尔豌豆杂交试验，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。

K. Plarson 的学生 Gosset(1876～1937)对样本标准差进行了大量研究，于 1908 年以笔名“Student”在该年的生物统计学报(Biometrika)上发表论文，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。

英国统计学家 Fisher 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。他在从事农业试验及数据分析研究时，创立了正交试验设计和方差分析。在生物统计中，方差分析有着广泛的应用，特别是在他发表了《试验研究工作中的统计方法》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。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。

Neyman(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说检验学说。假说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。

国内对生物统计学的应用始于 30 年代。中华人民共和国成立以后，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的分析等都取得了较好成果。

近年来，生物统计学发展迅速，从中又分支出生统遗传学(群体遗传学)、生态统计学、生物分类统计学、毒理统计学等。由于数学与生物学和农学的应用，使生物数学成为一门新的学科，生物统计学只是它的一个分支学科。1974 年，联合国教科文组织在编制学科分类目录时，第一次把生物数学作为一门独立的学科列入生命科学类中。随着计算机的普及和生物学研究的不断深入，生物统计学的研究和应用必将越来越广泛，越来越深入。

第四节　常用统计学术语

一、总体与样本

具有相同性质的个体所组成的集合称为总体，它是指研究对象的全体，而组成总体的基本单元称为个体。总体按总体单位的数目可分为有限总体和无限总体。个体极多或无限多的总体称为无限总体。例如，某一地区棉田棉铃虫的头数，可以认为是无限总体。另外，也可从抽象意义上来理解无限总体，比如通过临床试验来推断某一种药品比另一种药品的治愈率高，这里无限总体指的是一个理论性总体。个体有限的总体称为有限总体，如对某一班学生身高进行调查，这时总体是指这一班中每一名学生的身高。

要研究总体的性质，一般情况下我们无法一一对总体中的个体全部取出进行调查或研究。因为在实际研究过程中，我们常常会遇到两种难以克服的困难：一是总体的个体数目较多，甚至无限多；二是有时总体的数目虽然不多，但试验具有破坏性，或者试验费用很高，不允许做更多的试验，因而只能采取抽样的方法，从总体中抽取一部分个体进行研究，作为统计的依据。从总体中抽出的若干个个体所构成的集合称为样本，构成样本的每个个体称为样本单位，样本个体数目的大小称为样本容量。通过从样本计算出来的统计数，如平均数、标准差等，来对该总体在一定可靠程度上进行推断。样本的作用在于估计总体。例如可以调查某一地区棉田 100 株棉花上的棉铃虫头数，来推断该地区棉铃虫的发生状况，以采取相应的对策。一般在生物学研究中，样本容量在 30 个以下称为小样本，30 个以上称为大样本。在一些计算和分析检验方法上，大样本和小样本是不同的。

二、变量与常数

相同性质的事物间表现差异性或差异特征的数据称为变量或变数，它是表示在一个界限内变动着的性状的数值。自然界同类事物中，都存在着一定的变异，如人的身高、体重，棉花的株高、分枝数、衣分，同窝动物的身长及生理指标等都会存在一定的差异。所有这些差异均可用量来表示，通常记为 x，如 10 个人的身高在 155～180cm 之间，共有 158，167，173，155，180，165，175，178，170，162cm10 个变量值，记作 $x_i(i = 1,2,\cdots,10)$，表示 x_1 到 x_{10} 之间任一数值，亦称 x_i 为随机变量。

变量按其性质可分为连续变量和非连续变量。连续变量表示在变量范围内可抽出某一范围的所有值，这种变量之间是连续的、无限的。如小麦的株高在 80～90cm，在此范围内可以取得无数个变量。非连续变量，也称为离散变量，表示在变量数列中，仅能取得固定数值。如菌落中的菌数、单位面积水稻的茎数、小白鼠每胎产仔数等。

变量可以是定性的，也可以是定量的。定性的变量往往表示某个体属于几种互不相容的类型中的一种，如果蝇的翅有长翅与残翅，人的血型有 A、B、AB 和 O 型，豌豆花的颜色有白色、红色和紫色，等等。定量的变量是指可测量的，如出栏时猪的重量、花生的百仁重、电泳酶谱上的带数等。

常数表示能代表事物特征和性质的数值，通常由变量计算而来，在一定过程中是不变的。如某样本平均数、标准差、变异系数等。

三、参数与统计数

参数也称参量，是对一个总体特征的度量。如总体平均数、总体标准差等均为参数。因为总体一般都很大，有的甚至不可能取得，所以总体参数一般不可能计算出来。可以通过对总体抽取样本，计算样本的统计数，来估计总体参数。从样本中计算所得的数值称为统计数，它是总体参数的估计值。

四、效应与互作

引起试验差异的作用称为效应，如不同饲料使动物的体重增加表现出差异，不同品种的玉米产量不同等。互作，也称连应，是指两个或两个以上处理因素间的相互作用产生的效应。如氮、磷肥共施会对作物产量产生互作效应。互作有正效应，也有负效应，如果氮、磷共施的产量效应大于氮、磷单施效应之和，说明氮磷互作为正效应，如果氮、磷共施的产量效应小于氮、磷单施效应之和，说明氮磷互作为负效应。

五、机误与错误

机误，也叫试验误差，是指试验中由于无法控制的随机因素所引起的差异。如在抽样中，会出现较大或较小的数据，这是由于总体中的个体间存在一定的差异，它是不可避免的，试验中只能设法减小，而不能完全消灭。增加抽样或试验次数，可以降低机误的数值。

错误是指在试验过程中，人为的作用所引起的差错。如试验人员粗心大意，使仪器校正不准、药品配制比例不当、称量不准确、将数据抄错、计算出现错误等都是由于人为因素造成的，在试验中是完全可以避免的。

六、准确性与精确性

统计工作是用样本的统计数来推断总体参数的。我们用统计数接近参数真值的程度，来衡量统计数准确性的高低，用样本中的各个变量间变异程度的大小，来衡量该样本精确性的高低。因此，准确性不等于精确性。准确性是说明测定值对真值符合程度的大小，而精确性则是多次测定值的变异程度。

思考练习题

1.1 生物统计学的主要内容和作用是什么？

1.2 举例说明什么叫总体？什么叫样本？什么是参数？什么是统计数？

1.3 什么是机误？什么是错误？如何避免试验错误的发生？

第二章

试验资料的整理与特征数的计算

在生物学试验及调查中，能够获得大量的原始数据，这是在一定条件下，对某种具体事物或现象观察的结果，我们称之为资料。这些资料在未整理之前，一般是分散的、零星的和孤立的，是一堆无序的数字。统计分析就是要依靠这些资料，通过整理分析进行归类，使其系统化，列成统计表，绘出统计图，计算出平均数、变异数等特征数。

第一节　试验资料的搜集与整理

一、试验资料的类型

对试验资料进行分类是统计归纳的基础，若不进行分类，大量的原始资料就不能系统化、规范化。对试验资料进行分类整理时，必须坚持“同质”的原则。只有“同质”的试验数据，才能根据科学原理来分类，使试验资料正确反映事物的本质和规律。

对于生物学试验及调查所得的资料，由于使用方法和研究的性状特性不同，其资料性质也不相同。根据生物的性状特性，大致可分为数量性状和质量性状两大类，因而，我们所得到的资料有时是定量的，有时则是定性的，所以这些资料可以分为数量性状资料和质量性状资料。

(一)数量性状资料

数量性状资料一般是由计数和测量或度量得到的。由计数法得到的数据称为计数资料，也称作为连续变量资料，如鱼的尾数、玉米果穗上籽粒行数、种群内的个体数、人的白细胞计数等。计数资料的变量值以正整数出现，不可能带有小数。如鱼的尾数只可能是1，2，…，n，绝对不会出现2.5，4.8等这样的数据。

由测量或度量所得的数据称为计量资料，也称为连续变量资料，数据通常用长度、重量、体积等单位表示，如人的身高、玉米的果穗重量、仔猪的体重、奶牛的产奶量等。计量资料不一定是整数，在相邻值之间有微小差异的数值存在。如小麦的株高为80～95cm，可以是85cm，也可以是86cm，甚至可以是86.5cm或86.54cm等变量值，随小数位数的增加，可以出现无限个变量值。至于小数位数的多少，要依试验的要求和测量仪器或工具的精度