

生物统计学入门

责任编辑：林德琼
封面设计：李德华

生物统计学入门

周季维编著

*

云南人民出版社出版

(昆明市书林街100号)

云南新华印刷厂印装 云南省新华书店发行

*

开本：787×1092 1/32 印张：8.125 字数：188,000

1985年3月第一版 1985年3月第一次印刷

印数：1—8,000

统一书号：13116·96 定价：1.05元

前 言

数理统计学是上一世纪发展起来的一门学科，它对很多科学研究的理论、方法提供了巨大的帮助和推动。在生物学、农学、医学以及社会科学的领域中，由于数理统计方法的应用，获得了重大的成果。尤其在生物学的一些分支，如遗传学、生态学、生理学以至古生物学中，基本上没有人否认它所带来的好处。

近年来在数理统计的方法和应用上又出现了许多新的进展。同时在生物科学研究方面应用数理统计分析的要求也与日俱增。然而由于传统常常把生物学与数学看成是互不相关的两门科学，因此造成生物学工作者不熟悉数学，数学工作者也不善于将数学应用到生物学领域中去。一些生物学工作者，特别是农业科学技术工作者不习惯或缺乏数理统计的应用和训练，往往辛勤而反复地做了不少试验与观察，但所获得的资料却未能充分利用，大量可贵的数据由于缺乏系统科学的整理和分析，使一些具有规律性的真知卓识，失去了有价值的线索，或者由于对数据的处理方法不当或轻率片面地误用，反而得出相反或偏倚的结论。

本书为两类读者提供应用生物统计学的基础知识。一是对生物学方面的初学者，着重入门基本知识和概念的介绍，使读者能够较快地进入自己熟悉的领域，并且用数理统计的方法来描述和评价观察试验的结果。一是对农业科学研究的工作者在

实际工作中作为参考的工具书，可对常用的试验设计类型查出分析计算的步骤和方法，检验其试验观察的可靠性与置信程度和应用数学模型评价其结果。因此，并不打算编写得完整和包罗万象。全书分为两个部分。为了第一类读者，各章开头的一些部分说得比较简单和通俗，例子也很浅显。主要说明一般性的概念和数理统计是怎样用于这个领域的。后一部分针对实验研究工作者，选择了一些普遍常用的统计分析例子，逐步加以说明其运算方法以及分析测验的步骤并举例示范。只要具有一些基本的数学知识就能应用和理解其方法与主要内容。同类或相似的试验设计只要把编号和处理换一下，就可以代入或举一反三地加以比拟应用。

目前数理统计在学术上有从显著性的测验逐渐向点和区间估计量转移的趋势。虽然估计量往往要比显著性的测验能提供更多的资料，但实际应用中涉及了理论方面和专门数学的推导。此外，还有各式各样非参量性的测验法。这些都超出了本书的范围。由于繁冗的计算有把人们的注意力从数理统计的目的性转移出来的倾向，因此本书作为生物统计基础入门与应用，力图把计算的负担减轻，把注意力集中在生物学的目的性上。同时着重将方差分析的方法尽快地介绍给读者，这是现实应用中较广泛而实用的。均数、标准差等的概念和基本计算都是方差分析的基础，介绍得详尽一些。概率以及理论分布作为一般概念，只作简要叙述，因为它是数理统计的理论基础，同时也是进一步深入点和区间估计量的依据。在实际应用中可跳过这些章节。如在计算中要弄清有些统计量的意义，可再回转来读一读，将是有益的。

为了读者能更容易地阅读国内外有关生物学、遗传学等的文献期刊，有必要将某些统计符号和名词加以统一。本书内所

用符号以“广播电视大学教材”《概率统计讲义》人民教育出版社1980年第1版,《田间试验和统计方法》农业出版社1980年2版及英国《生物学统计方法》(“Statistical Methods in Biology”, Hodder & Stoughton, Second edition, 1981)为主。书后附“符号”及“公式”集注,以便对照查阅。

本书编写过程中,参考了一些生物统计书籍和教材,特别是在例题和试验结果分析方面,吸取并引用了它们中的不少材料,并得到云南省农业科学院程侃声研究员的帮助,在此致以谢忱。

由于编著者水平有限,书中的缺点和错误一定不少,请读者批评指正。

编 者

一九八四年一月

目 录

第一章 生物统计学的发展历史和内容..... (1)

1 生物统计学的历史和发展..... (2)

2 生物统计的主要内容..... (5)

一、资料的整理与基本分析..... (7)

二、显著性测验..... (7)

三、相关与回归的度量..... (8)

3 生物统计中常用的几个术语概念..... (10)

一、总体与样本..... (10)

二、随机抽样..... (10)

三、变数与变异数列..... (10)

四、参数与统计量..... (11)

五、机误与错误..... (11)

六、准确性与精确性..... (11)

第二章 几个主要统计值的意义和计算方法 (12)

1 样本集中性的度量——平均数…………… (12)

一、平均数的意义…………… (12)

二、平均数的计算方法…………… (13)

2 样本变异性(离中性)的度量——标准差
…………… (23)

一、标准差的意义…………… (23)

二、标准差的计算方法…………… (24)

3 变异系数…………… (28)

一、变异系数的意义…………… (28)

二、变异系数的计算方法…………… (29)

4 平均数标准差(标准误)…………… (31)

一、平均数标准差的意义…………… (31)

二、平均数标准差的计算方法…………… (32)

第三章 数据资料的整理和分析…………… (34)

1 数据整理的方法…………… (34)

一、数量资料的分类…………… (37)

二、原始数据的检查与核对…………… (40)

三、数据的分组…………… (41)

2 数据表示的方法.....(48)

一、次数分布表.....(48)

二、次数分布图.....(48)

第四章 概率与理论分布.....(55)

1 什么叫概率.....(55)

一、事件.....(56)

二、概率.....(58)

三、概率的计算.....(62)

2 什么是理论分布.....(68)

一、正态分布.....(70)

二、二项分布.....(84)

三、泊松 (Poisson) 分布.....(91)

第五章 统计推断与分析.....(96)

1 显著性测验.....(98)

一、平均数差异显著性测验——t 测验.....(98)

二、百分数显著性测验.....(105)

2 参数的区间估计.....(110)

3 方差 (变量) 分析.....(112)

一、方差分析的基本方法.....(113)

二、方差分析的多重比较.....(119)

第六章 实验结果的统计分析.....(123)

1 单因素试验结果的统计分析.....(124)

一、简单对比试验和调查的分析.....(124)

二、对标比排列试验.....(135)

三、间比法试验.....(142)

四、随机区组排列试验.....(148)

五、拉丁方试验的统计分析.....(168)

2 多因素试验结果的统计分析.....(179)

3 多年多点试验或调查结果的统计分析.....(189)

第七章 生物统计中相关回归分析及其应用.....(198)

1 相关分析、相关系数及其显著性测验.....(198)

一、相关系数的计算.....(204)

二、相关系数的显著性测验.....(209)

2 回归分析及其计算.....(210)

| | | |
|---------------|--------------------------------|-------|
| 一、 | 直线回归方程和回归系数的计算····· | (211) |
| 二、 | 应用举例····· | (213) |
| 附录 I | 主要统计符号和注解····· | (218) |
| 附录 II | 生物统计常用公式集注····· | (222) |
| 附录 III | 生物统计基本概念和统计分析步骤表解····· | (228) |
| 附表 I | t 值表 (两尾)····· | (237) |
| 附表 II | 5 % F 值表 (一尾)····· | (239) |
| | 1 % F 值表 (一尾)····· | (241) |
| 附表 III | 邓肯新复极差测验 5 % 和 1 % SSR 值表····· | (243) |
| 附表 IV | 相关系数 r 值显著性界限表····· | (247) |

第一章

生物统计学的发展历史和内容

生物生长、发育、遗传、变异的发生发展，千百年来成为人类的一项重要研究对象。动物、植物、微生物以及人类自己，这一课题越来越成为探索其规律的综合科学。

生物统计学是处理大量偶然现象中的规律性的方法。它使我们能从浩繁的事实中抽出意义深远的结论，使估计上的误差和人力物力上的浪费减至最低度。因而它有助于作出行之有效的决定，成为对生物科学与技术的设计和改进行起关键作用的一环。

生物统计学是数理统计在生物科学中的应用。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数量资料的科学。人们在从事生物科学研究时，往往是通过某事物的一部份（样本），来估计事物全体（总体）的特征特性的。其目的在于由样本推断总体，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。

在生物科学研究中，我们所期望知道的是总体，而不是样本，可是在具体工作中，我们所得到的却是样本而不是总体（由于在生物科学研究中不可能也做不到对总体全面进行研

究)，因此生物统计在某种意义上来说，它就是研究生物学中以样本来推断总体的一门学科。

1

生物统计学的历史和发展

统计是个老名词，它很可能是作为算术开始的。奴隶社会或封建社会的统治者为了抽取人民的税收或者发动一场战争，需要知道所统治的人民的财富和名数。所有那些有意把历史记载下来的文化也会无意的记下统计的数字。公元前27—14年罗马皇帝凯撒·奥古斯都曾下过一道命令，要全世界都向他纳税，于是他要每个人都向就近的统计师登记，那时候的统计师就是收税人。英国的威廉大帝下令测量英国的土地，其目的是为了征税和征兵役，这种测量的记录称为“定罪日记录”（Domesday Book）。我国史籍上和各地方志书中的田赋、徭役、丁（男）口（女）记载也是这类的统计。这些都已成为历史的陈迹。

十四世纪，在殖民主义开始向外扩张侵略年代里的船运保险中，经验概率得到了应用，比利时和荷兰的法兰德斯人（Flemish）船运业中，逐渐发展成为在统计中一种被人重视的形式，称为保险。

以机遇博奕的方式出现的赌博，大约在十七世纪中叶引起了由巴斯噶和斐马（Pascal Fermat）所创始的概率论，这是由于他们对谢利埃·得·梅雷（Chevalier de Me're）伯爵的赌博经验发生了兴趣。对统计师以及实验科学家来说，在这概率论中包含着不少对数据处理有实际用途的方法。

在数理统计的发展中，正态曲线或者误差的正态曲线是非

常重要的。在1733年得·穆瓦弗雷 (De—Moivre) 首先发表了这一曲线的方程，但是并未想到要把它的结果应用在实验观测数据上，因而他的论文没有得到注意，一直到1924年才被卡尔·皮尔逊 (Karl Pearson) 在一家图书馆中发现。不过同样的结果后来曾由两个数学天文学家拉普拉斯(Laplace1749—1827)和高斯 (Gauss 1777—1855) 各自独立地推导出，而后来由高斯完成，因而称为高斯分布。

一个重要的统计理论在十九世纪由查理士·莱耶尔 (Charles Lyell) 应用于地质学问题，在1830年到1833年期间，出版了莱耶尔著述的三卷《地质学原理》他确定了第三纪岩石地层的次序和年代，并加以命名，他同法国的贝类学家迪希耶 (Deshayes) 一起把各个地层中化石的种加以辨认和记录，并确定其中迄今在海洋中还活着的种的百分率，他根据这些百分率制定了更新世、上新世、中新世以及始新世的地层名称。莱耶尔的论据实质上是属于数理统计的同一类型。这些地层名称的确定被学术界所承认并至今沿用。

生物统计是一门新兴的学科，实际上在十九世纪末叶才开始建立起来。1870年英国遗传学家哥尔通 (F. Galton) 通过研究人类体高的遗传，认为子女的体高与父母的体高有着直接关系，发现有回归的趋向，这是在数理统计中，回归这个术语的由来，因而后人推崇他为生物统计的创始人。

生物学家达尔文 (1809—1882) 在搭乘贝格尔 (Beagle) 号船到南美洲考察的旅程中带了莱耶尔 (Lyell) 新著的《地质学原理》第二卷，达尔文的工作在本质上主要是应用生物统计和数理统计的，从而提出了进化论。1899年皮尔逊 (K. Pearson) 提出了一个测量实际数与预计数 (或理论数) 之间的偏离度的指数卡方 (X^2)，它在属性统计分析上有着广泛的

应用。例如，在遗传上孟德尔豌豆杂交试验，高豌豆品种与矮豌豆品种杂交之后子一代均为高豌豆，高株这个性状属于显性，而矮株属于隐性，当子一代自交时，它的后代数预计是应该有高 3 : 矮 1，但实际的后代数是否符合 3 : 1 需用 X^2 进行测验。这种方法至今在遗传学方面仍广泛应用。

K·皮尔逊 (Karl Pearson 1857—1936) 原是一位数学物理学家，由于达尔文的进化论引起了他在生物学上的热衷，他将数学用于生物学几乎费了半个世纪的时间，他从事认真的数理统计研究，并创立了《生物统计学报》(Biometrika) 这份杂志以及一个数理统计学校。

在皮尔逊致力于大样本研究的时候，大样本理论对一些只能用小样本的实验工作者来说多少被证实是不妥当的。他们之中有个名叫戈斯特 (W. S. Gosset 1876—1937) 的在一家酿酒厂工作的科学家，是皮尔逊的学生。戈斯特发现大样本的理论不能用以求得他所关注的一些统计量，诸如样本标准离差、样本平均数对样本离差之比以及相关系数等的恰当分布的目的，后来他采用从弄乱的卡片中抽样，计算并积累经验的频率分布，所得结果以《t 测验》论文发表在1908年的《生物统计学报》上署名为“学生氏”(Student)。现今 t 测验已成为生物统计工作的一种基本工具应用非常广泛。

1923年英国的费雪 (R. A. Fisher 1890—1962) 第一个把变异来源不同的均方 (S^2) 比值称为 F 值，当 F 值大于理论上 5 % 概率水准的 F 值时，该项变异来源的必然性效应就从偶然性变量中分析了出来。这个分析方法，被称为方差分析 (也称变量分析)。在生物统计中，方差分析应用极为广泛。特别是在他发表了《试验研究工作中的统计方法》的专著后，对推动和促进农业科学、生物学和遗传学的研究和发展起了一

定的奠基作用。

纳耶曼 (J. Neyman) 和 E. S. 皮尔逊 (E. S. Pearson) 分别在1936年提出一种统计假设检验学说, 对促进理论研究以及对试验研究作出结论具有实用价值, 此外数理统计学家, 阿布拉罕·韦尔德 (Abraham Wald 1902—1950) 的序贯分析和《统计判决函数》被认为是数理统计学上的很大成就。

我国早在三十年代, 数理统计已成为农业科学、生物学、遗传学的一项研究工具, 并起到积极作用, 以后扩大应用于医学和工业上。近年来又提出了优选法和安排与分析多因素试验的正交试验法, 这个方法在工农业方面的试验研究中, 也相继获得了很大的成就, 最近, 回归分析方法在工农业生产和科学实验中大量应用, 如在进行数据处理时, 寻求经验公式, 探索新工艺、新配方, 以及某些性状的预测预报都积累了不少新经验, 获得了许多新成果。各种类型电子计算机的发明, 使运算技术出现新的跃进, 变得日益精确和迅速。在生物科学现代化的进程中, 应用统计方法以及先进的试验设计, 来进行分析、研究, 在今后工作中将越来越显得重要。

2

生物统计的主要内容

现代科学发展的重要趋势是日益走向量化的方向。在生物科学领域内的许多量的问题, 不再以过去那种定性地描述各种生物现象为满足, 而在生物学的应用方面, 单靠定性的指导, 在许多场合下已经不能应付实际的需要, 只有通过定量的科学实验和调查研究加上定量的数据分析和总结, 才能更为有效地解决实际问题, 如果不进行这种量的关系的研究, 就几乎

不能得到任何关于这类事物规律性的认识。

生物统计所研究的主要内容就是这种量的变化关系，即生物的变异量。通常我们所说的“这个男人的体重超过60公斤”；“这株水稻高达90公分”；“那个鸡蛋的重量是50克”等等都只是涉及到一些随个体而变异的特征，仅只是定性的描述。至于是否所有（总体）男人的体重都超过60公斤，或者平均一个地区（或者一个学校、一个家庭）成年男子的平均体重都是这样呢？或者某一些“来航”鸡蛋的个体平均重量是50克？这就得用生物统计的方法来加以研究。个体表现出变异性或变差的特征称为变量（也叫机遇变量、随机变量或变数）。对于变量的值在统计上还要有相对频率这种概念。例如，人们对一只鸡重达50公斤的传闻是不大会相信的；一个鸡蛋重一市两则被认为是正常的。这样我们就把一个变量的值跟一种度量联系起来，这就是一种事物的正常性或缺乏正常性的度量范围，在数理统计中就称这种变量具有一个分布，或者一个概率或频率分布。生物统计中所研究的变异量，都是服从于一定的理论分布。以鸡蛋重量来讲，它是服从于两头小，中间大，即最大或最小的蛋均较少，中等重量的蛋较多的一种正态分布（也叫常态分布）。正态分布的特点之一，是呈现连续性变异，例如动物的体重身高；农作物的穗长、粒重等都属此类，是定量的。但也有一些性状是属于非连续性的变异，在个体属于几种互不相容的类别中的一种（它不能同时属于两种类别）时，呈现离散性变异，所作的观测是定性的，而且一般是非数字的，如家畜家禽的性别比例，非公即母；植物种子的颜色；药物试验的治愈或无效等。这种非连续性的变异是服从于数学上二项式分布的，具有离散型变量的概率分布。而研究这些分布都要应用统计分析的原理和方法。本书作为科学研究方面一般生物统计的

入门基础，着重借助数理统计方法的概念与计算，来科学地分析问题，对理论及数学原理仅简要说明其意义，不作深入探讨和推导。因此从统计方法的主要内容及其应用来讲，大致可以分为以下三个方面。

一、资料的整理与基本分析

生物统计是以样本推断总体。以一定数量的样本通过统计资料初步整理之后，再从资料中计算出三个主要的统计量，即平均数、标准差及标准误，从而推断总体的特征。即一、资料的集中性，以平均数来表示；二、资料的离中性，以标准差来表示；三、衡量平均数的可靠性，以标准误的大小来表示。

二、显著性测验

由于所获得的资料只不过是一个样本，它是从总体中取出的一部分，在样本与总体之间必然有一定的差异，因此，必须探究样本平均数的可靠性，看它是否能代表总体，即可靠的程度，在生物统计上称为显著性（可靠性）测验。

测验显著性的方法较多，根据试验研究和抽取样本方法的不同，以及假设与适合性的检验标准不同，近年在数理统计上，又有不少新的理论和发展，但常用和基础的是以下三类。

（一）平均数间差异的比较 当遇到两组或两组以上数据的平均数时，要在它们之间进行比较，就有一个有无显著差异的问题，或者说孰优孰劣的问题。平均数之间的比较，要通过一定的统计方法，并且要注意抽样是否合理否则就没有比较的基础。例如有两批孵化日期不同但生长期同样为两个月的小鸡的体重；甲组（3月9日出壳）共61支，平均体重为269.3克；乙组（4月15日出壳）共73支，平均体重为275.6克；试问