

SHENGWU TONGJIXUE

生物统计学

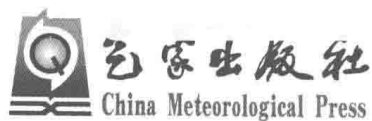
吴荣军 李琪 戴庆龙 编著

非外借



生物统计学

吴荣军 李 琪 戴庆龙 编著



内 容 简 介

本书较系统地介绍了生物统计学的基本原理与方法及其应用。在简要叙述生物统计学的概念、发展、试验资料的整理及资料特征数的计算、概率及概率分布和抽样分布的基础上,重点介绍了假设检验和区间估计、卡方(χ^2)检验、方差分析、直线回归与相关分析,并基于 SPSS 统计软件介绍了生物统计学在相关问题中的应用。

本书适合从事生命科学、生态学、农业气象学、农业资源、环境科学和医学工作者阅读,也可供相关专业的本科生和研究生作为教材使用。

图书在版编目(CIP)数据

生物统计学 / 吴荣军, 李琪, 戴庆龙编著. — 北京:
气象出版社, 2019. 1

ISBN 978-7-5029-6934-9

I. ①生… II. ①吴… ②李… ③戴… III. ①生物统
计-教材 IV. ①Q-332

中国版本图书馆 CIP 数据核字(2019)第 026886 号

SHENGWU TONGJIXUE

生物统计学

出版发行: 气象出版社

地 址: 北京市海淀区中关村南大街 46 号

邮政编码: 100081

电 话: 010-68407112(总编室) 010-68408042(发行部)

网 址: <http://www.qxcbs.com>

E-mail: qxcbs@cma.gov.cn

责任编辑: 黄红丽

终 审: 吴晓鹏

封面设计: 楠竹文化

责任技编: 赵相宁

印 刷: 三河市百盛印装有限公司

开 本: 720 mm×960 mm 1/16

印 张: 13.5

字 数: 247 千字

版 次: 2019 年 1 月第 1 版

印 次: 2019 年 1 月第 1 次印刷

定 价: 48.00 元

本书如存在文字不清、漏印以及缺页、倒页、脱页等,请与本社发行部联系调换。

前 言

生物统计学是利用数理统计的原理和方法来分析和解释生物领域各种现象和试验、调查及观测资料的一门科学。随着学科不断发展,生物统计学也在与生物学相关或者交叉学科领域有着广泛的应用。不管是纯生物学领域,还是农学、林学、医学、生态学、环境科学等领域的相关工作都需要用到生物统计学的原理和方法。从一定数量的数据资料中,应用统计分析方法,分析和解释数据的数量变化及其差异性特征,从而解释或判断相关试验、调查和观测指标的变化特征和规律,得出合理的科学结论。因此,统计分析方法是很多科研工作的重要环节。有了良好的统计学基础,数据处理的很多结果就能知其然,且知其所以然,有助于科研工作的进一步深入。

本书是作者在十几年来在从事生物统计学教学和相关应用研究的基础上编写完成的。书中的内容主要围绕样本推断总体这条主线,针对统计方法的基本原理仅作基本介绍和公式的简单推导,而更侧重于统计方法的应用,结合实例让读者了解和掌握各种统计方法,并能够针对不同问题采用适合的统计推断方法。本书概括起来主要有四个方面的内容:第一,生物统计学的介绍和数据的初步识别,涉及第 1 章和第 2 章,主要介绍生物统计学的概念和发展、数据收集和整理、数据特征值的计算;第二,在概率论的基础上介绍总体分布和抽样分布,涉及第 3 章,主要介绍概率的概念和计算,以及针对总体参数的二项分布、泊松分布和正态分布,并在中心极限定理的基础上,引出了样本统计数的抽样分布;第三,在抽样分布的基础上,介绍统计推断方法,涉及第 4 章到第 7 章,主要介绍样本平均数和标准差的假设检验、区间估计、 χ^2 检验,以及特殊的统计推断方法方差分析和回归分析;第四,介绍不同统计方法在 SPSS 软件的应用,涉及第 8 章,SPSS 软件包容量大、功能全面,并受到了广泛应用,为便于读者在了解统计方法基本原理的基础上,熟练运用统计推断方法开展应用,本章结合实例介绍了 SPSS 软件在数据的描述性统计、均值比较和方差分析、 χ^2 检验

和相关分析和回归等方面的应用,并给出了详细的步骤和分析结果的解读。本书还在每章最后都附有一些习题,供读者参考,本书所附习题很多是从书后参考文献中引用的,在这里对原著作者深表感谢。

本书在编写过程中,得到了南京信息工程大学教材建设基金的资助,得到了教务处和应用气象学院领导和相关同仁给予的大力支持,教师发展与评估中心的陈娟副研究员通阅了全书,并提出了许多宝贵的建设性意见。吴荣军同志承担了本书第1、3、4、5章的编写,李琪同志承担了本书第2、6章的编写,戴庆龙同志承担了本书第7、8章的编写。徐家婧、周宁、边晨昱、田丰、孙茂森等同学在本书的录排工作中做了大量工作。在本书的出版过程中,得到了气象出版社的大力支持,特别是黄红丽副编审在书稿的编审方面所做的大量工作,在此一并表示感谢。各方经过2年时间的努力,终于将《生物统计学》付诸出版。

本书的编写力求简明实用、通俗易懂、由浅入深、层次分明,在保证科学性、系统性的基础上,强调针对性,注重实践性,突出应用性,适合相关学科的科学工作者阅读,也可供本科院校生态学、应用气象学、生物学类专业作为教材使用。同时,由于作者水平的限制,本书难免会有错误和不妥之处,敬请读者批评指正,以便今后修订完善再版。

作者

2018年11月

目 录

前 言	
第 1 章 绪论	(1)
1.1 生物统计学的概念	(1)
1.2 生物统计学的发展	(2)
1.3 生物统计学的内容与作用	(4)
1.4 生物统计学的特点及其学习方法	(6)
1.5 常用统计学术语	(7)
思考题	(11)
第 2 章 数据的收集与整理	(12)
2.1 试验资料的收集与整理	(12)
2.2 试验资料的特征数计算	(20)
思考题	(29)
第 3 章 概率分布与抽样分布	(30)
3.1 概率的基础知识	(30)
3.2 概率的概念	(35)
3.3 概率的计算	(41)
3.4 几种常见的概率分布	(51)
3.5 中心极限定理	(64)
3.6 抽样分布	(66)
思考题	(73)
第 4 章 统计推断	(77)
4.1 假设检验的原理与方法	(77)
4.2 单个样本的假设检验	(84)
4.3 两个样本的差异显著性检验	(88)
4.4 总体参数的区间估计	(94)

思考题	(99)
第 5 章 χ^2 检验	(102)
5.1 χ^2 检验的原理与方法	(102)
5.2 适合性检验	(104)
5.3 独立性检验	(107)
思考题	(111)
第 6 章 方差分析	(112)
6.1 方差分析的基本原理	(112)
6.2 两因素方差分析	(125)
6.3 方差分析的基本假定和数据转换	(135)
思考题	(138)
第 7 章 一元回归与简单相关分析	(140)
7.1 回归和相关的概念	(140)
7.2 回归分析	(142)
7.3 相关分析	(156)
思考题	(158)
第 8 章 基于 SPSS 软件的统计分析	(160)
8.1 数据管理及数据资料的基本统计分析	(160)
8.2 均值比较	(167)
8.3 方差分析	(174)
8.4 χ^2 检验	(182)
8.5 回归和相关分析	(189)
思考题	(194)
参考文献	(195)
附表	(196)
附表 A 正态分布的累积函数 $F(u)$ 值表	(196)
附表 B 标准正态分布分位数表	(198)
附表 C t 值表	(199)
附表 D χ^2 值表(右尾)	(200)
附表 E F 值表(右尾)	(201)
附表 F 新复极差检验 SSR 值表	(209)
附表 G 相关系数临界值表	(210)

第 1 章 绪 论

1.1 生物统计学的概念

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,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各种现象和试验、观测和调查资料的科学,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。

生物研究的对象是生物有机体的生长发育、生理生化特征及变化,以及生物与环境的相互作用,研究过程中获得的各类生物及环境因子等观测、试验和调查数据具有较大的差异性,这种差异性往往伴随着一定的误差成分。例如,小麦纯系是经过多代自交得到的,遗传上已经纯合化,个体间遗传成分可以认为是均一的。将自交系的单株后代种植在生长条件都相同的环境中,具有相同的肥力、水分、光照、温度和通风条件等,即使这样,个体间的株高、穗长、穗重、干物重等生物指标仍存在变异,这就是典型的随机现象。如何识别并揭示这种差异性特征,就必须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。

由于个体间的变异,要识别环境和气候因素对小麦生长和产量的影响差异,最理想的办法是对所有小麦植株的生长和生理指标进行测量,但事实上根本不可能像人口普查那样对全部植株进行测量,因为这要消耗大量的人力、物力、财力和时间成本。因此,测量全部植株的所有指标既不现实也不可能。我们只能从全部研究对象中抽出一部分个体(样本)来,通过分析这一部分个体的相关指标来推断全体(总体)

的差异性特征。一般来说,在生物科学研究中,我们所期望的是总体,而不是样本,但在具体的试验过程中,我们所得到的却是样本而不是总体。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。因此,用样本推断总体是生物统计学的主线,生物统计学的所有内容都是围绕样本推断总体这一目标进行的。

生物统计学作为一门重要的应用型专业工具课,一般不过多的推导其数学公式,也不详细地讨论其数学原理,而主要偏重于统计原理的介绍及具体统计分析方法在生物学、生态学及其相关学科中的应用。

1.2 生物统计学的发展

正态分布是生物统计学科发展最重要的基础理论,它最早是由法国数学家棣莫弗(A. De Moivre, 1667—1754)于1733年发现,后来德国物理学家、天文学家和数学家,近代数学奠基人之一高斯(J. C. F. Gauss, 1777—1855)在进行天文观察和研究土地测量误差理论时,也独立发现了正态分布(又称常态分布)的理论方程,提出“误差分布曲线”。因此,后人为了纪念他,将正态分布称为 Gauss 分布。贝努里(J. Bernoulli, 1654—1705)系统论证了“大数定律”,即样本容量越大,样本统计数与总体参数之差越小。

统计学用于生物学的研究,开始于19世纪末。在生物学的研究中,测量误差是重要问题,必须应用统计处理分析其变异。最早有意识地将数理统计学引入到生物学领域的先驱者是比利时数学家阿道夫·凯特勒(L. A. J. Quetelet, 1796—1874),提出了生物统计的思想。1866年,孟德尔的豌豆杂交试验是最早运用数理统计于生物实验的一个成功范例。

19世纪末,英国著名的人类学家和生物统计学家高尔登(F. Galton, 1822—1911)的工作巩固了生物测量学和优生学的基础,于1884年创建了世界上第一个人类测量实验室。应用统计方法研究人种特性、处理遗传学中的变异,分析父母与子女的变异,探索其遗传规律,指出了子代的身高不仅与亲代的身高相关,而且有向平均值“回归”的趋势,即高个子父母的子女,其身高有低于其父母身高的趋势,而矮个子父母的子女,其身高有高于其父母的趋势,创立了平均数离差法则,提出了“回归”

和“相关”的概念和算法,奠定了生物统计的基础。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,并和他的学生皮尔逊创办了 *Biometrika*(《生物统计学报》),首次明确“Biometry(生物统计)”一词。后人推崇他为生物统计学的创始人。

数学家皮尔逊(K. Pearson, 1857—1936)继承了高尔登的研究工作,经过共同的努力于 1895 年成立了伦敦大学生物统计实验室,进行了回归和相关特别是复相关、泊松型分布数、频率累加法、 χ^2 检验等数理统计学的研究,并制成了很多统计数值表,于 1889 年发表了《自然的遗传》一书。在该书中,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,发展了统计回归和相关的理论,阐明了测量实际值与理论值之间偏离度的指数卡方(χ^2)的检验方法。例如,在遗传上孟德尔豌豆杂交实验,高豌豆品种与矮豌豆品种杂交后,它的后代理论比率是否符合 3 : 1,需用进行 χ^2 检验。

皮尔逊的学生戈塞特(W. S. Gosset, 1876—1937)在生产实践中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》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,也称为学生式分布。 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,为此,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。

英国统计学家费歇尔(R. A. Fisher, 1890—1962)指出,只注意事后的数据分析是不够的,他使实验设计成为生物统计的一个分支,并创立了正交试验设计方法。同时,他于 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,创立了数量遗传学和方差分析。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。

此后,美国统计学家内曼(J. Neyman, 1894—1981)和 K. 皮尔逊(K. Pearson)的儿子 E. S. 皮尔逊(E. S. Pearson, 1895—1980)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,合著了《统计假设试验理论》,发

展了假设检验的数学理论,其要旨是把假设检验问题作为一个最优化问题来处理。他们把所有可能的总体分布看作一个集合,其中考虑了一个与零假设相对应的备择假设,引进了检验功效函数的概念,以此作为判断检验程序好坏的标准。这种思想使统计推断理论变得非常明确。内曼还从数学上定义可信区间,提出了置信区间的概念,建立置信区间估计理论。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。

国内对生物统计学的应用始于 20 世纪 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,基因组学和功能基因组学、遗传流行病学、群体遗传学和生态学、计算神经科学等现代生物学领域对生物统计提出了新的要求,生物统计学发展迅速从中又分支出统计遗传学(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。随着计算机技术和信息技术的大量引入,以及生物学数据的爆炸性增长,使得生物统计与生物信息的结合,生物统计学的研究和应用必将越来越广泛,越来越深入。

1.3 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和统计分析。试验设计是指应用数理统计的原理与方法,制定试验方案,选择试验材料,合理分组,降低试验误差,使我们可以利用较少的人力、物力和时间,获得丰富而可靠的数据资料。试验设计主要介绍试验设计的有关概念、试验设计的基本原则、设计方案的制定、常用设计方法,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。统计分析是指应用数理统计的原理与方法对数据资料进行分析与推断,认识

客观事物的本质和规律性及事物间的相互关系,得出合理的结论。统计分析中,主要包括数据资料的收集和整理、数据特征的计算、统计推断和方差分析、回归和相关分析、协方差分析、主成分分析、聚类分析等。统计分析与试验设计是不可分割的两部分。试验设计须以统计分析的原理和方法为基础,而正确设计的试验又为统计分析提供了丰富、可靠的数据资料信息,两者紧密结合推断出合理的结论,不断地推动生物统计学的应用及应用生物科学研究的发展。

1964年,英国著名统计学家叶茨(F. Yates, 1902—1994)和希利(M. J. R. Healy, 1924—)曾说过:“非常痛心地看着,因为数据分析的缺陷和错误,那么多好的生物研究工作面临被葬送的危险。”事实上,迄今为止,统计方法应用的缺陷和错误从未停止。2014年《自然》(*Nature*)杂志发表了美国华盛顿加劳德特大学 Regina Nuzzo 的观察文章 *Scientific method: Statistical errors* 讨论了大量发表的论文中 P 值(统计学指发生 I 型错误的概率)的滥用问题。同时,50%的医学论文存在统计学瑕疵(Altman et al., 1991)。足以看出正确使用生物统计方法对于生命科学研究是何等重要,生物统计学的作用主要体现在以下几个方面。

(1) 提供数据和描述数据资料的科学方法。

一些实验或数据资料,若不整理则杂乱无章,无规律可循。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图表,从而可以了解大批资料所蕴藏的内在规律性信息。

(2) 运用显著性检验,判断试验结果的可靠性。

一般在实验中要求除实验因素以外,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。例如,在比较转基因与非转基因水稻产量时,可以采用对比设计、分两个区块进行试验,尽管施肥、管理、浇水等耕作技术都一致,但由于不同区块土壤肥力必然存在一定的差异、每株水稻的生理生化特性的差异,以及其他一些不可控的因素对水稻产量均存在一定的影响。因而,每株水稻的产量必然包括转基因的真实效应以及偶然因素两个效应。因此,我们在分析结果时,只有正确区分这两种效应,判断它们各自效应的大小,才能对试验结果作出科学可靠的结论。根据不同的资料类型,生物统计学提供了推断试

验结果可靠性的不同方法。

(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)提供科学的试验设计方法和原则。

科学的试验设计可用较少的人力、物力和时间取得丰富可靠的试验资料,尽量降低试验误差。在一些生物科学研究中,就需要科学地进行试验设计,如对样本容量的确定、抽样方法的挑选、处理水平的选择、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,进行科学的试验设计。

1.4 生物统计学的特点及其学习方法

生物统计学的主要特点包括概率性、二元性和归纳性。概率性也即明确概率论是生物统计学的理论基础。生物统计学的一个重要特点就是通过样本来推断和估计总体,这样得出的结论有很大的可靠性,但有一定的错误的概率,因此,从概率的角度来思考问题和分析科学实验非常重要。二元性指的是理论结合实际数据。归纳性,是指从现实资料中归纳出一般的原理并应用于较大的范围,从特殊到一般,从样本到总体,属于归纳推理的范畴。在生物统计学的教学中,一些学生反映生物统计学比较难懂,这主要是因为生物统计学是生物科学与数理统计的交叉学科,涉及的数学公式、数学概念、数学符号及数学用表等较多,需要付出较多的时间和精力来掌握这些基本概念、基本原理和基本方法,记忆一些常用的公式和符号,理解并推导一些基本的数学表达式。事实上,生物统计学是一门逻辑性和应用性较强的学科,作为一个并非专门从事生物统计学原理与方法研究的人员,其主要目的是应用生物统计学方法解决生产和研究中遇到的实际问题,简化数学公式的推导。学好生物统计学,可以为其他课程和今后的学习与研究奠定基础。如数量遗传学就是应用生物统计方法研究数量性状遗传与变异规律的一门学科,如果不懂生物统计学,也就无法掌握遗传学。此外,阅读中外科技文献也常常会碰到统计分析的问题,也必须有生物统计的基础知识。因此,生物科学工作者必须学习和掌握统计方法,才能正确

认识客观事物存在的规律性。

生物统计学是一种很有用的工具,正确使用这一工具可以使科学研究更加有效,主要的学习方法主要有:

(1)必须理论联系实际,结合专业知识,不可忽视专业知识的重要性。生物统计对试验数据提供收集、整理、分析的方法,并作出结论。因此,生物统计学对试验结果作出科学的推断时,必须结合专业知识进行全面综合分析才能获得客观、科学的结论。例如,某研究者测得某地区 200 名正常成年男性转氨酶和身高的数据,用简单相关与回归对这两个变量间的关系进行分析,并开展了正确的显著性检验,得出的结论是:可用身高去预测转氨酶的含量。显然这个结论与实际不符,缺乏专业理论知识的支撑。

(2)要理解生物统计学的基本理论和基本公式。要正确理解每一个公式和生物统计分析方法的实际含义和应用条件。例如,进行 t 检验时,要求原始资料服从 t 分布,否则可能会得出错误结论;进行方差分析时必须满足正态性、可加性和方差齐性等三个条件。如果不考虑应用条件,生搬硬套,会得出与事实不符,甚至截然相反的结果。

(3)掌握各种符号及其意义,定期整理所学过的公式与符号,不必深究其数学推导过程。因此,平时需多留意国内外杂志、书籍中的图表、数据及其分析方法和对结果的解释,从而熟悉表达方法及其应用。

(4)要及时认真地做好课后习题。结合常用的生物统计分析软件加深对统计的基本理论和基本方法的理解和掌握,达到能熟练运用生物统计方法分析实际问题的目的,提高生物统计分析的效率,及时认真地复习是保证学习效果的关键。

(5)应注重培养科学的统计思维方法。生物统计意味着一种新的思考方法——从不确定性或概率角度来思考问题和分析科学试验的结果。生物统计学不仅仅是试验设计和统计方法的教学,更重要的还是有助于大学生思维方式的培养。

1.5 常用统计学术语

1.5.1 总体与样本

研究的目的是要了解总体,然而能观测到的却是样本,通过样本来推断总体是

统计分析的基本特点。因此,总体与样本是生物统计中的两个最基本概念。

具有相同性质或属性的个体所组成的集合称为总体,组成总体的基本单元称为个体,总体是我们研究的全部对象,总体既分为无限总体和有限总体。例如,我们要研究在某种生长条件下小麦的株高,因为无法估计生长在这种条件下的小麦的数量,可以设想这一总体是无限的。或者研究新生儿的体重,因为新生儿是不断增加的,所以这一总体也可以设想是无限的。如果我们要调查一所学校今年新生的身高,这一总体则是有限的。生物统计学中所遇到的总体多数是无限总体。

从总体中抽出若干个体所构成的集合称为样本,样本内包含的个体数目称为样本容量,常记为 n 。

1.5.2 变量与常数

变量(变数)是指相同性质的事物间表现差异性 or 差异特征的数据,对变量的表现进行观察或测量所获得的数值称为观测值,这些数值有时也被称为变数。变量可分为定量变量和定性变量,而定量变量又可分为连续变量和非连续变量。变量表现在不同个体间或不同组间存在变异性,例如身高、体重、产仔数、性别、毛色、血型等,显然有的变量的表现可以用数值来表示(如体重),而有的变量的表现则要用文字来描述(如毛色),但有时为了计算分析方便,我们也可将这种用文字描述的变量转换成数值,例如可将红、白、黑三种毛色用 1, 2, 3 来表示。

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。

1.5.3 参数与统计数

为了表示总体和样本的数量特征,需要计算出几个特征数,包括平均数和变异数(极差、方差、标准差等)。

描述总体特征的数量称为参数,也称参量,常用希腊字母表示参数,例如用 μ 表示总体平均数,用 σ 表示总体标准差。

描述样本特征的数量称为统计数,也称统计量,常用拉丁字母表示统计数,例如用 $\bar{x}$ 表示样本平均数,用 s 表示样本标准差。统计量可以由样本观测值计算得到。

一般来说,每一个总体参数都有一个对应的样本统计量。我们前面所说的由样

本推断总体也可以理解为由样本统计量推断总体参数。

1.5.4 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。

1.5.5 准确性与精确性

准确性是指观测值或估计值与真值的接近程度,精确性是对同一物体的重复观察值或估计值彼此之间的接近程度。这两个概念可用图 1.1 来说明:图中显示了三个射手打靶结果的弹点分布情况,第一种情况(图 1.1a)弹点分布较散且大部分偏离靶心,说明该射手的射击精确性和准确性都较差;第二种情况(图 1.1b)所有弹点都偏离了靶心,但彼此间很集中,说明该射手射击的准确性差,但精确性高;第三种情况(图 1.1c)则是精确性高准确性也高。

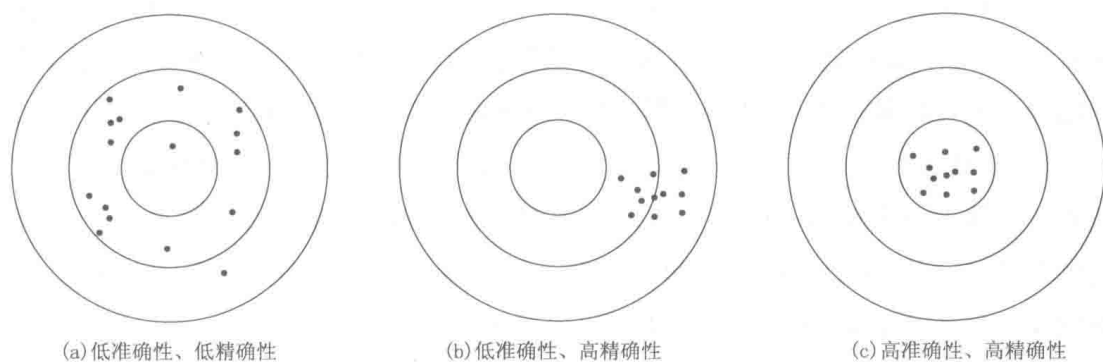


图 1.1 准确性与精确性示意图

需要注意的是,精确性还与数据的有效位数有关。有的重复观测值在有效数较少时(例如仅取整数)精确性很高,但如增加有效位数,差别就显示出来了。例如,对一头仔猪体重的两次度量值分别是 15.6 kg 和 16.4 kg,二者间的差是 0.8 kg,但如

果只取整数(四舍五入),二者就没有差别。有效位数的多少除了人为的取舍外,还与测量仪器的精度有关。例如一个普通天平,只能称得以克为单位的重量,克以下就不敏感,而一个电子天平可以称得 0.001 g 或者 0.0001 g 的重量。对测量仪器精度的要求和对有效位数的取舍取决于我们所研究问题的要求。在生物学的研究中,有一个常用的简单规则,即 30—300 规则,即以数据间差异的最小单位计算,最小值与最大值间的差异应在 30—300 个单位。例如有一群仔猪,如果对它们的体重以千克为单位,则体重最小的为 12 kg,最大的为 20 kg,差 8 kg,显然这样的体重数据的精确性太低,因为一个单位的度量误差就会造成比较大的差别。如果以 0.1 kg 为单位,则体重最小的为 12.3 kg,最大的为 19.8 kg,差 7.5 kg,即差了 75 个单位,这样的精确性对于大多数统计分析来说已经足够了,但如果以 0.01 kg 为单位,精确性就显得太高,记录这样的数据不仅对称重设备的要求较高,也会增加许多工作量。

1.5.6 误差与错误

误差也叫试验误差,是指试验中不可控因素所引起的观测值偏离真值的差异。试验中出现的误差可以分为两类:随机误差和系统误差。随机误差也称为抽样误差、偶然误差,它是由于试验中许多无法控制的偶然因素所造成的试验结果与真实结果之间产生的误差,是不可避免的。我们可以通过试验设计和精心管理设法减小随机误差,而不能完全消除。随机误差影响试验的精确性。统计上的试验误差就是指随机误差。增加抽样或试验次数,可以降低随机误差。系统误差也称为片面误差,是由于试验处理以外的其他条件明显不一致所产生的带有倾向性的或定向性的偏差。系统误差主要由一些相对固定的因素引起,例如仪器调校的差异,各批药品间的差异,不同操作者操作习惯的差异等。系统误差在某种程序上是可控制的,只要试验工作做得精细,在试验过程中是可以克服的。

错误是指在试验过程中,人为的作用所导致的差错。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、计算出现错误等由于人为因素造成的,在试验中是完全可以避免的。这类错误原则上是不允许产生的。