

生物统计学

主 编 杨运清 岳永生 于长江
主 审 缪尧源 涂万里



哈尔滨工程大学出版社

A-332

425565

Y30

生物统计学

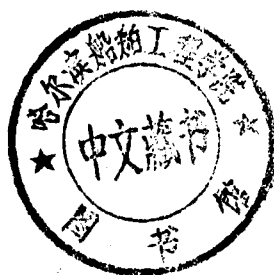
主 编 杨运清 岳永生 于长江

副主编 陈德全 张廷荣 许晓曦

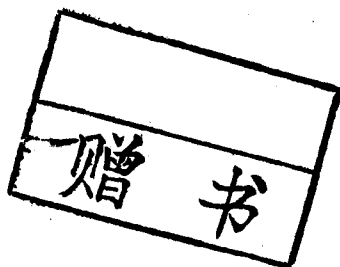
李树怀 李 瑛 张 宏

王 辉 李 馨

主 审 缪尧源 涂万里



00425585



哈尔滨工程大学出版社

内 容 简 介

本书围绕着由样本推断总体的基本思路,分四大部分系统地介绍了有关统计比较与相关回归分析方法。主要内容有资料的搜集与整理;概率与正态分布;总体特征数测定;显著性检验的基础理论;两均数差异的统计比较;多个均数差异的统计比较;试验设计资料的方差分析;计数资料的统计比较;简单直线相关与回归分析;多元线性相关与回归分析;非线性回归分析;协方差分析;统计比较与方差分析的统一等。该书可作为动物科学类包括畜牧、兽医、公共卫生、生物工程和食品加工等专业的专科生、本科生教材;也可作为相关领域现场与科研工作者的参考书。

生 物 统 计 学

主 编 杨运清 岳永生 于长江
责任编辑 徐达山

哈尔滨工程大学出版社出版发行
新华书店经销
哈尔滨市华茂印刷厂印刷

开本 787×1092 1/16 印张 17.4 字数 400 千字
1996年5月第1版 1996年5月第1次印刷
印数: 1—3 000 册

ISBN 7-81007-677-9/Q·5

定价: 16.90 元

序

生物界广泛存在着变异，掌握变异的规律是人类认识生物界的有效途径。《生物统计学》为人们掌握变异的规律性提供了科学的手段和方法，并由此使人类对生物界的认识产生了重大飞跃。

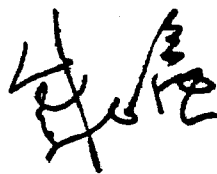
虽然《生物统计学》被引入动物科学领域并且正式成为一门课程的历史并不很悠久，但它对推动动物科学研究的发展发挥了重要作用。为使动物科学工作者更好地掌握这门科学，加强《生物统计学》的教材建设势在必行。

由东北农业大学杨运清倡导并组织山东农业大学、黑龙江八一农垦大学、吉林农业大学、莱阳农学院、熊岳农业专科学校和北安农业专科学校等六所农业院校的部分教师，历经两年时间编写了这本《生物统计学》。该书具有两个突出的特点：

一是系统性强。该书按着教学大纲的要求，围绕着由样本推断总体这一主题，将《生物统计学》从总体上分为四个部分，这些部分既相互联系又彼此独立成篇，同一部分内章与章之间、节与节之间衔接有序。所阐述的问题明朗、透彻。

二是实用性强。该书针对动物科学的特点而选择介绍有关的统计理论和方法，力图做到形象直观，理论联系实际，例题选用恰当，习题丰富实用。

总之，这本《生物统计学》内容丰富，选材精当，而且言简意赅，通俗易懂，是农业院校动物科学专业学生的一本良好教材，也可作为广大动物科学工作者的参考书。



一九九六年五月一日

前 言

历时两年，这本《生物统计学》教材终于面世了。

该书可作为畜牧、兽医、公共卫生、生物工程、食品加工等专业本、专科生教材，也可作为同类科研和生产工作者的参考书。

经常听到有一些读者反映《生物统计学》晦涩难懂，是一块“难啃的硬骨头”，认真分析一下其中的缘由，其关键所在，可能是它的特点——理论性强，比较抽象。这一方面表现为与形象直观的其它专业基础课相比，学习方法将由原来的“少理解多记忆”转变为“多理解少记忆”，一时有点不太适应；另一方面表现为预备知识的准备可能不够充分，尤其是必要的、足够的应用数学知识。所以如何将抽象的理论方法通过形象直观的途径传播，应是《生物统计学》编写与教学工作首先需要解决的问题。为此，在编写这本《生物统计学》教材过程中我们重点作了如下努力：

首先，根据动物科学研究的基本特点，将《生物统计学》从总体上划成四大板块，明确各板块所解决的主要问题，加强版块内容间的联系；缩减与概率论和数理统计相同基础理论的重复讲解；将所引用的概率论与数理统计方法有机地与动物生产和科研问题结合在一起，不要求读者一定要先学会概率论与数理统计即可读懂。

其次，尽管还是按照原统编教材的思路组织内容，即先说明方法步骤，再举例演示，但本教材在对每个统计理论方法的直观解释上下了很大功夫，使理论与应用衔接得尽量自然一些；可以肯定，书中基本上杜绝了“后面方法前面引用”问题的出现。

本书的参编人除了主编与副主编外，还有李淑玲、单玉明和王明文同志，他（她）们在提供习题，计算机验算以及部分章节的编写上作了一定的工作。

特别感谢盛志廉教授、杨汝康教授对该书编写工作的支持与鼓励。

《生物统计学》是在多年教学手稿的基础上加以整理并集合作者的教学经验而成。有的学术观点可能属于一家之言，偏颇之处在所难免，恭请读者不吝赐教，批评指正，以便再版时加以更正，使本书更臻完善准确。

杨 运 清
1996.5.1

目 录

序	(1)
前 言	(1)
绪 论	(1)
一、生物统计学概念	(1)
二、本学科的研究内容	(1)
三、学习本学科应注意的几个问题	(4)
四、本学科发展概况	(5)
五、常用统计术语	(6)
第一部分 资料的基本特征分析	(8)
第一章 资料的获取与整理	(9)
第一节 有效数字	(9)
第二节 资料的检查与核对	(10)
第三节 资料的分类	(11)
第四节 资料的整理方法	(12)
第二章 概率与正态分布	(19)
第一节 事件与概率	(19)
第二节 正态分布	(20)
第三章 总体特征数测定	(26)
第一节 反映总体集中趋势的统计量	(26)
第二节 反映总体离散程度的统计量	(32)
第二部分 统计比较	(39)
第四章 显著性检验的基础理论	(40)
第一节 显著性检验的基本原理	(40)
第二节 显著性检验的基本步骤	(41)
第三节 单尾检验与双尾检验	(42)
第四节 两类统计错误	(42)
第五章 两均数差异的统计比较	(45)
第一节 t 分布与 t 检验	(45)
第二节 与已知总体均数差异的显著性检验	(46)
第三节 两样本所在总体均数间差异的显著性检验	(48)
第四节 总体均数的置信区间	(53)
第五节 样本含量的估计	(55)
第六章 多个均数差异的统计比较	(59)
第一节 方差分析的基本原理	(59)

第二节	F 分布与 F 检验	(61)
第三节	多重比较	(64)
第四节	处理均数的独立分组比较	(70)
第五节	方差分析的基本假定及数据转换	(74)
第七章	试验设计资料的方差分析	(80)
第一节	试验设计简介	(80)
第二节	单因素试验设计资料的方差分析	(82)
第三节	两因素试验设计资料的方差分析	(98)
第四节	正交设计资料的方差分析	(109)
第八章	计数资料的统计比较	(117)
第一节	相对数分析	(117)
第二节	二项分布	(119)
第三节	两属性率差异的显著性检验	(121)
第四节	χ^2 检验	(126)
第三部分	相关与回归分析	(141)
第九章	简单直线相关与回归分析	(143)
第一节	简单直线相关分析	(143)
第二节	简单直线回归分析	(150)
第三节	简单直线相关与回归的关系	(159)
第四节	简单相关和简单回归的应用	(160)
第十章	多元线性回归与相关分析	(162)
第一节	多元线性回归分析	(162)
第二节	多元线性相关分析	(175)
第三节	自变量作用程度分析	(179)
第十一章	曲线回归分析	(183)
第一节	可直线化曲线方程的拟合	(183)
第二节	多项式回归方程的拟合	(189)
第三节	曲线拟合效果的评价	(191)
第四部分	统计比较与回归分析的结合与统一	(194)
第十二章	协方差分析	(195)
第一节	一元协方差分析	(195)
第二节	二元协方差分析	(207)
第十三章	回归分析与方差分析的统一	(213)
第一节	有关线性模型的基本知识	(213)
第二节	方差分析模型是一种特殊的线性模型	(214)
第三节	效应的最小二乘分析	(218)
第四节	方差组分估计	(223)
思考与习题		(227)

《生物统计学》名词术语汉英对照表	(241)
附表 1. 标准正态分布密度函数表	(245)
附表 2. 标准正态分布函数表	(246)
附表 3. 正态离差 ν_α 值表(两尾)	(248)
附表 4. 理论 t 值表(两尾)	(249)
附表 5. 5%(上)和 1%(下)理论 F 值表(一尾)	(250)
附表 6. 5%(上)和 1%(下)理论 SSR 值表(两尾)	(256)
附表 7. 5%理论 q 值表(两尾)	(258)
附表 8. 1%理论 q 值表(两尾)	(259)
附表 9. 理论 χ^2 值表(一尾)	(260)
附表 10. r 与 R 5%和 1%的理论值表	(261)
附表 11. 常用正交设计表	(262)
参考文献	(267)

绪 论

一、生物统计学概念

生物统计学是应用概率论和数理统计原理来研究生物界数量变异规律的一门学科。根据这些原理和方法，能正确设计科学试验，正确处理试验结果，从而推导出较为客观的结论。半个世纪以来，由于生物统计学的发展，现已广泛应用于田间试验、数量遗传、动物育种、繁殖、饲养、兽医、公共卫生和食品等科研生产工作中，它对加速农业、动物产业的科学研究和农牧业生产的发展起着重要的促进作用。同时农牧业生产的发展和农牧业科学研究的不断深入又促进了生物统计学的进一步发展与提高，使本学科已成为当代在农业科学研究和生产上必不可少的工具了。

人们在从事试验研究时，往往是通过总体的一部分（即样本），来估计总体的特性特征。其目的是为了由样本来推断总体，以便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。例如，畜牧业生产中，研究中国荷斯坦（原名黑白花）奶牛 305 天产奶量时，可以发现其资料是相当庞大的，几乎不可能全部测出，而只能采用抽样的方法。从研究对象所在总体中抽出一部分来加以研究，即用样本的统计量，来估计总体的参数。因此，生物统计学从研究思路上看，它是以样本来推断总体的一门学科。

根据生物统计学基本概念与研究思路，将不难概括出该学科所包含的特点：

1. 概率性 它不仅表现在生物统计学的研究手段是概率论，以及建立在概率论基础上的数理统计方法，更主要的是利用生物统计方法解决生物科学问题时所得出的每一个结论是不确切的，是在一定误差概率的允许下，作出的可靠推断。
2. 归纳性 从辩证法角度看，样本是研究对象的特殊规律，生物统计学由样本推断总体的研究思路乃是由特殊到一般的归纳过程。
3. 实践性 生物统计学在运用归纳法研究生物界数量变异规律的同时，还非常重视将其所归纳的结论应用于生物科学实践，检验统计结论的准确性或演绎出更为具体的理论方法。

二、本学科的研究内容

如果将所学的课程结构比作“金字塔”，那么位于塔底的是基础课，塔中间的是专业基础课，塔顶就是专业课。生物统计学应处于“金字塔”三层结构的哪个位置呢？大部分院校都将其列为专业基础课。是数理统计课不能完全替代的，因为这门课与数理统计课相比至少更注重于实用。

那么，生物统计学是如何实现由样本来推断总体的呢？图 1 给出了较为形象的解

释。据此将本书的内容归纳为四大部分：

第一部分 资料的基本特征分析

本部分包括全书前三章：第一章为资料的获取与整理；第二章为概率与正态分布；第三章为总体特征数的测定。内容梗概：

分析资料的基本特征是生物统计学的首要任务，是实现由样本推断总体的基础工作。

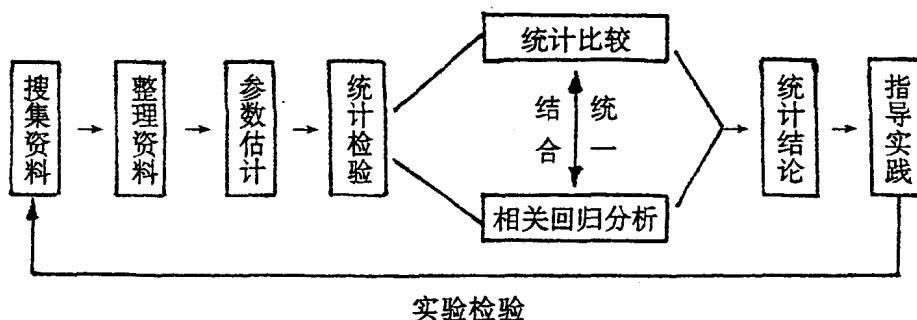


图 1 生物统计学研究思路示意

大量的数据资料将来自于动物生产与科学试验中，为保证数据资料的完善性、真实性和准确性，在获取数据资料过程中要特别注意有效数字的舍取与计算，获得资料后，还应对资料进行检查与核对，去粗取精，去伪存真。

由于原始资料起初只是一些零星的、孤立的现象，所以欲从中找出它们的内部联系和规律性，就必须进行资料的整理和分析，资料整理是指按照一定的标志，把性质相同或相似的资料进行分类或合并，通过整理资料，可以直观地了解到资料各观测值的分布情况。

多数资料的观测值都呈正态分布或近似正态分布，正态分布在各种生物统计分析方法中处于举足轻重的地位。

不论是资料还是资料所在总体，观察资料整理后的分布情况，便会发现其分布取决于两个基本特征，那就是集中趋势与离散程度。如果能够将资料（样本）的两个基本特征用数值表示出来，那么就实现了对总体相应特征数的估计。

第二部分 统计比较

本部分占全书内容的最大比重，包括从第四章至第八章共五章内容：第四章为显著性检验的基础理论；第五章为两均数差异的统计比较；第六章为多个均数差异的统计比较；第七章为试验设计资料的方差分析；第八章为计数资料的统计比较。

在动物生产与科研中，经常遇到两个或两个以上研究对象的比较问题，借此区别和选择不同的处理、手段、方法或状态。由于研究对象对应的总体往往是无限的，因此，在人力、物力所不及的条件下，只有通过样本间的差异来推断总体间的差异，将这种运用统计方法比较不同处理、手段、方法或状态间差异的过程称为统计比较。

那么究竟比较什么？又如何进行比较呢？为回答这些问题，需要引入数。有关“显著性检验”的理论与方法。所谓的显著性检验，也称假设检验，是指运用信息判定有关总体特征（总体分布、参数、相关性等）每一个论断的过程。根据所判定的总体分布是否已知，又将显著性检验分成参数显著性检验与非参数显著性检验两种。本部分统计比较所依赖的理论基础乃为参数显著性检验，即在研究对象总体分布已知的条件下，利用样本信息对总体参数差异的显著性检验，并把统计比较的重点放在各种试验设计情况下的计量和计数资料上。

第三部分 回归与相关分析

本部分是该课程的第二大重点内容，包括从第九章至第十一章的共三章内容：第九章为简单直线相关与回归分析；第十章为多元线性相关与回归分析及第十一章曲线回归分析。

前几章所讨论的都是一种事物或一个现象，即一个变量的特征及其变化的规律。但是在自然界中，任何一种事物或现象的形成，都是受着另外一些事物或现象的影响，亦即各种事物或现象都不会是彼此孤立的存在，而是有着相互依赖和相互制约的关系。生物现象之间的关系亦不例外，生物的各性状之间以及性状与环境条件之间都有着一定的相互关系，改变某一性状，就会引起另一性状也发生变化。这种联系反映在数量方面就表现为变量与变量的相互联系和相互制约。

变量间的关系可分为两种类型：确定性的关系和非确定性的关系。

在统计学中，相关分析是研究变量间相互依赖非确定性的关系的，通过分析确定变量间相互联系的性质和紧密程度；而回归分析则是研究一个变量对另一个变量单向依存的非确定性的关系，即研究一个变量是如何随着另一个变量的变化而变化的，通常前一个变量称为依变量或应变量，后一个变量称为自变量。在相关与回归分析过程中，根据被研究变量的数目及其相互关系所对应的几何意义，可将变量间的关系划分为简单直线相关与回归和多元线性相关与回归，以及一元或多元曲线相关与回归几种类型。

本部分根据这些划分，着重讲述有关线性相关和回归分析内容。

第四部分 统计比较与回归分析的结合与统一

统计比较是用来推断各种资料形式下不同处理效果差异显著性的，而回归分析则是用来通过计算回归系数和建立回归方程，明确不同变量间相互关系的。虽然统计比较与回归分析的研究目的有所不同，但在一些情况下，两者还必须结合起来，才能有效地实现对生物界数量变异规律的发掘和认识。如：

- 1) 在比较不同处理效果的差异过程中，对指标的影响，除了试验因素外，还有其它不可控因素；
 - 2) 多因素交叉分组试验设计资料，当次级样本含量不等时，总平方和将失去可加性；
 - 3) 多因素试验的目的不是为了统计比较，而是研究各因素作用的变异幅度。
- 诸如此类问题，是单凭统计比较或回归分析一种方法所不能圆满解决的。其实数理

统计学早已证明，方差分析与回归分析的理论基础相同，都源于线性模型理论。

根据线性模型的基本思想，回归分析与方差分析的有机结合，就产生了协方差分析法；最小二乘分析对解决次级样本含量不等资料的统计比较问题起了不可估量的作用；如果线性模型为随机或混合模型，那么方差组分的估计可以用来研究各因素的变异幅度。这样，作为线性模型理论的初步，本部分便将统计比较与回归分析结合和统一起来了。

三、学习本学科应注意的几个问题

利用生物统计学可对取得的实验资料进行综合与分析，它是结果判断的有力工具。但试验者必须善于运用这个工具，才能使研究工作收到事半功倍的效果。这里应注意以下几点：

1. 试验者运用生物统计应注意统计方法与演算技术来源于数理统计的概念与原理，任何公式都在一定条件下成立，因此对概念与原理要有一定程度的了解，否则就易犯机械套用公式的毛病，使最终结论引入歧途。例如常用的 t 检验方法，要求原始资料服从正态或近似正态分布，一堆严重偏离正态的数据，应用 t 检验可能会得出错误结论；又如方差分析资料必须满足三个基本假定才有效。如果不考虑应具备的条件，生搬硬套地运用公式，会导致荒谬的结果。

2. 应用生物统计的同时，不可忘记专业知识的重要性。生物统计是对试验提供如何收集资料、分析资料以及作出结论的方法，它确能使试验收到经济、有效、省时、省力的效果，对试验的最终结果作出科学的判断，但必须结合专业知识进行全面分析才能获得实事求是的结论。例如在检定药物毒性时，有将试验的药物用恒速静滴注于麻醉动物，以观察动物心跳停止时所用药物总量。一个在体内代谢慢速的滴注所得的中毒量往往由于作用的叠加要比快速滴注来得小。这个道理显然不能用生物统计而只能用药理学知识来解释；又例如一组服安眠药的病人比另一组未服安眠药的病人平均睡眠时间长半小时，经统计分析，两组有显著差别。但任何一个有经验的医生都知道，这个差别并无实际意义。因此利用统计显著性检验作出的结论还必须结合专业知识考虑其是否有实质显著性。

3. 生物统计从来不仅仅是在实验最终结束后对取得的实验资料进行分析，而应该在实验开始前，就在生物统计的理论指导下，对实验作出精心的安排，以消耗最少的人力、物力和时间换得最多最有效的资料。在实验过程中，还必须对产生误差的因素或可能的意外进行估计，并随时改进和控制实验条件，以减轻这些额外因素的干扰。只有这样，才能使生物统计发挥充分效果。反之，一堆未经设计的实验数据，实验本身安排不合理，资料不正确，则任凭用什么“高超”的统计方法也决不能从中找出正确的答案来。

4. 生物统计有时需要冗长繁复的计算，近年来由于电子计算机的应用，这个问题已获得逐步解决。但仍有大量的计算工作要借助于袖珍计算器，对复杂的统计分析仍花时较多。为此，试验者常追求一些简单且粗糙的方法代替，甚至为了减轻计算工作，不

惜放弃一些有用的资料。须知应用生物统计方法对最终的实验数据所耗的演算劳动和用于整个实验的劳动相比，毕竟只占很小部分。因此，试验者应把生物统计的计算工作看作是整个过程的一个重要环节，必须充分利用已取得的所有资料，不厌其烦地将资料分析演算工作完成得正确无误。

5. 应用统计学总是在一定的显著水准上作出结论，这是因为对存在变异的现象，要作出完全正确的判断是不可能的，其中还包含了结论错误的可能性，尽管这种错误的可能性很少。因此试验者必须对结论的性质做到“胸中有数”，并可根据具体情况，对规定的显著水准结合专业知识进行斟酌，必要时可提高或降低显著水准，使最终结论分寸得当。

总之，试验者既不可视生物统计为累赘，又不可将生物统计看作万能，惧怕生物统计中某些艰涩难懂的抽象理论以及烦琐冗长的数学运算而加以回避，或者认为生物统计既根据于严密的数学理论，一切结论均由统计通过，就深信勿疑，这两种极端想法都是错误的。

四、本学科发展概况

统计一词起源很早，远在奴隶、封建社会时，由于战争和税收的需要，就开始有统计技术了。但是把概率论与数理统计的原理应用到生物学中来，使生物统计成为一门独立的学科，是在十九世纪的末期。1870年英国遗传学家高尔顿(F. Galton, 1822~1911)通过研究人类身高的遗传，发现儿子的平均成长高度，介于父亲高度与一般高度(种族高度)之间。父亲矮的，其儿子的平均高度则较父亲高，但比一般高度矮；父亲高的，其儿子的平均高度则较父亲矮，但比一般高度高。儿子的高度，有趋向于种族平均高度的趋势，亦即回归于一般平均高度。这就是“回归”一词在遗传上的含义。1889年他发表了回归分析方法在遗传学上应用的论文，因而后人推崇他为生物统计的创始人。

正态分布对研究生物统计的理论是十分重要的，它早在1733年就被迪—摩弗来(De-Moivre)发现，而被后来的高斯(C. F. Gauss, 1775~1855)所完成，因此也有人称之为高斯分布。

1899年K.皮尔逊(K. Pearson, 1857~1936)提出了一个测量实际数与理论数之间偏离度的指数卡方，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。高株这之性状属于显性，而矮株属于隐性，当子一代自交时，它的后代数预计是高与矮之比为三比一，但皮尔逊是一位数学物理学家，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》这份杂志和一个数理统计学校，对促进生物统计学科的发展作出了贡献。

古斯特(W. S. Gosset, 1876~1937)是K.皮尔逊的学生。他发现大样本的理论不能用以求得他所关注的一些统计量，为此提出了 t 统计量，就是估计平均数距总体平均数以标准误为单位而计量的离差数。后来他从事小样本理论的研究，于1908年以“学

生”为笔名，在《生物统计学报》上发表了《 t 检验》的论文。这一检验方法已成为当今生物统计工作者的一种基本工具，应用非常广泛。

1923年英国费歇(R.A.Fisher, 1890~1962)第一个把变异来源不同的均方与误差均方的比值称为 F 值，当 F 值大于理论上5%显著水准的 F 值时，该项变异来源的处理效应就从偶然性变量中分析了出来，这个方法被称为方差分析。他在首次发表了《试验研究工作中的统计方法》的专著后，提出了田间试验的基本原则和主要设计方法，成为试验设计的经典著作。同时他还使统计学与群体遗传学相结合形成数量遗传学，对推动和促进农业科学、生物学以及遗传学的研究和发展，起了奠基的作用，被后人公认为当代生物统计学的奠基人。

J.纳耶曼(J.Neyman)和E.S.皮尔逊(E.S.Pearson)分别在1936~1938年提出统计假设检验学说，对促进理论研究以及对试验研究作出科学结论具有重大实际意义。

我国早在30年代，生物统计与田间试验就成为农学专业的必修课程。最早出版的书籍有王绶编著的《实用生物统计法》(1935年出版，1953年再版)。在动物科学专业方面应用则比较迟。1945年吴仲贤教授在成都开设的高级遗传学并同时讲授生物统计，作为研究变异的统计方法。在饲养学中则进行饲养试验设计与分析。尔后在兽医药检方面也开始应用检验。进入70年代以来，由新的教学计划设置的《生物统计附试验设计》已成为畜牧、兽医两专业的必修课程。在杨纪珂教授的倡导下，受农业部委托，由盛志廉教授主持为畜牧界开办了多期生物统计培训班，从而使生物统计与试验设计在畜牧业上的应用日益广泛起来。近十年来，我国数量遗传学取得了不少成果，例如，遗传参数的估测，进一步推动了我国畜牧界生物统计的发展。目前多元统计分析、最小二乘分析等新技术的出现，各类电子计算机的推广和应用，使计算技术出现了新的跃进。随着电子计算机的应用发展，生物统计将成为今后科学研究工作中，越来越重要的工具。

五、常用统计术语

1. 总体与个体 根据研究目的确定的、符合指定条件的全部观测结果称为总体。构成总体的每一个观测值，称为该总体的个体。例如，研究中国荷斯坦(黑白花)奶牛在正常饲养管理条件下305天的产奶量，那么，凡是在这种相同条件下饲养的黑白花奶牛，其305天的产奶量构成一个总体，而每一个产奶量值则是一个体。根据所含个体的多少，总体可以分为有限总体与无限总体两种。个体数目有限的总体称为有限总体，个体数目无限的总体称为无限总体。生物统计学的研究对象主要是无限总体，尽管有限总体可以全部被观测，但受限于人力、物力，尤其是有的观测还具有破坏性，故相对样本来说，统计学上也将其假定为无限总体。

2. 抽样与样本 在研究总体特征时，因为总体的无限性或测量方法具有破坏性，不允许对总体中的每一个个体一一进行观测，而采用抽样的方法，从总体中抽取一定数量的个体进行研究。这部分个体叫做样本。样本中所含个体的数目叫做样本含量。由于样本含量不同，一般又将样本分大样本(样本含量 $n>30$)和小样本(样本含量 $n<30$)。从

总体获取样本过程中，必须严格贯彻随机抽样的原则，以保证所得样本的代表性，只有样本能够代表总体，才能比较准确地实现生物统计学由样本推断总体的目的。所谓随机抽样是指抽样时，不掺杂人为的主观愿望，总体中每个个体被抽取的机会均等。因样本由随机抽样而得，故也将样本称随机样本。

3. 参数与统计量 由总体计算的数称为参数。参数是一真值，是统计学的推断对象。由样本计算的数称为统计量，它是参数的估计值，受抽样变动的影响。为了区别，统计上规定凡是参数均用希腊字母表示。如总体平均数用 μ 符号，总体标准差用 σ 。凡是统计量均用拉丁字母表示。如样本平均数用 $\bar{x}$ 符号，样本标准差用 S 。

4. 系统误差与随机误差 误差按来源的不同可分为系统误差和随机误差，前者是由于试验条件所带来的差异，如测量工具不准确，试验过程中少考虑了某些因素的固定作用等，这类误差的表现具有定向性，是可以克服的；而随机误差则是由偶然因素引起的差异，该误差表现为随机波动、不定向，无法完全克服，只能尽量减少或降低，如在相同条件下重复测量一物体长度，每次测量结果却不尽相同，这是因为环境温度、温度、风速的微小变化都可能影响测量结果，而通常这些偶然条件是难以控制的，因此所造成的随机误差也无法克服。

5. 准确性与精确性 观测值与真值接近的程度称为准确性；同一性状重复观测各观测值彼此接近的程度称为精确性。以打靶为例，有三名射手进行比赛，每个各射五次。情况是：

第一人每次都打中靶心，可以说这是个准确性高精确性也高的好射手；

第二人每次都没有打中靶心，而且很分散，是一个准确性和精确性都很差的射手；

第三人每次虽没有打中靶心，但在中心外某一点上很集中，可以说这是一个精确性高但准确性差的射手。

第一部分 资料的基本特征分析

分析资料的基本特征是生物统计学的首要任务，是实现由样本推断总体的基础工作。

大量的数据资料将来自于动物生产与科学试验，为保证数据资料的完善性、真实性和准确性，在获取数据资料过程中要特别注意有效数字的舍取与计算，获得资料后，还应对资料进行检查与核对，去粗取精，去伪存真。

由于原始资料起初只是一些零星的、孤立的现象，所以欲从中找出它们的内部联系和规律性，就必须进行资料的整理和分析，资料整理是指按照一定的标志，把性质相同或相似的资料进行分类或合并，通过整理资料，可以直观地了解到资料各观测值的分布情况。

多数资料的观测值都呈正态分布或近似正态分布，正态分布在各种生物统计分析方法中处于举足轻重的地位。

不论是资料还是资料所在总体，观察资料整理后的分布情况，便会发现其分布取决于两个基本特征，那就是集中趋势与离散趋势，如果能够将资料（样本）的两个基本特征用数值表示出来，那么就实现对总体相应特征数的估计了。

第一章 资料的获取与整理

数据资料的获取主要有三个途径：生产记录、科学试验和调查研究。而获取这些资料的方法既有利用度量衡测量的，又有计数的，究竟采用哪种方法要根据资料的性质而定。关于获取资料的具体过程不将作过多讲解，这里仅就与资料准确性有关的“有效数字”回顾一下。

第一节 有效数字

一个观测值中决定数量大小的数字称有效数字，用以正确地表达观测与计算结果。由于受观测工具或仪器的限制，不可能使观测值有绝对的正确性。即使最精密的长度测量仪，也不可能对某些长度作出绝对正确的测量。最简单的例子是当直角三角形的两个直角边皆为 1 单位时，其斜边长度是 $\sqrt{2}$ 单位，这条斜边就无法用长度测量仪测得绝对正确。如果测量以 m 为单位，最细刻度是 cm，则测量可精密到小数点后 2 位，即可量出 1.41m，只能对小数的第 3 位进行估计，如果这个估计约为 4mm，则可记录这个测量长度为 1.414m 这个数字也可成 141.4cm 或者 1414mm。但不管小数点的位置如何，前面 3 个数是准确测量的，最后一个数是估计的。这个测量值共有 4 位有效数字。

有效数字究竟是几位要视观测仪器究竟能达到多少精密度而定。例如用分析天平称药物，若灵敏度为 0.1mg，则称得的药物重例如为 5.0032g，这里可有 5 位有效数字；若灵敏度为 1mg，则可称得 5.003g，有 4 位有效数字；若灵敏度为 10mg 的台秤，则可称得 5.00g，有 3 位有效数字；从这里可以看出有效数字包括小数点后的 0，当然也包括数字中间的 0，并与小数点的位置无关。如果将这个重量用灵敏度为 1g 的磅秤称量时，只可能得出 5g，或者是 0.005kg。由于 5 是一位有效数字，0.005 当然也是一位有效数字，于是可以认为有效数字不包括数字前的 0。

对有效数字位数的决定可归纳为以下三点：

- (1) 位于数字中或数字后的 0 都是有效数字；
- (2) 有效数字与小数点位置无关；
- (3) 位于数字前的 0 不作有效数字。

对数值进行运算时，其算得的结果到底是几位有效数字与原数据有几位有效数字有关，这里举出加减法与乘除法的规则如下：

由于在有效数字中，最后一位是估计数，因此若有两个数相加或相减后得出的数中，有两个或两个以上的估计数，估计数仅需保留最前一位。例如

$$15.273 + 0.2359 + 7.23 = 22.7386$$

以上相加的三个数中，各数有底线的皆为估计数，因此相加的数只能是最前一位估计数 3，3 以后的 8，6 两个数就无意义，于是正确的答数应是 22.74。

若有几个数字相乘除时，其答数的有效数字应与原数字中有效数字最少的一个相