21世纪高等院校教材

# 生物统计学

主编 吴占福

科学出版社
www.sciencep.com

21世纪高等院校教材

# 生物统计学

主　编　吴占福

科学出版社

北　京

## 内 容 简 介

生物统计学是高等农业院校动物科学、动物医学等专业的主要专业基础课之一。为进一步适应21世纪市场经济和我国高等农业教育的发展，为适应高等农业教育改革的深化，我们编写了本书。本书共10章，包括概论、资料的整理与特征数、概率与理论分布、均数差异显著性检验、方差分析、次数资料分析——$\chi^2$检验、直线相关与回归、多元线性回归与相关、协方差分析、试验设计，并附有习题、实验实习指导及常用统计学用表。在编写中特别注意突出重点，阐述基本概念与基本方法，便于读者举一反三。对内容的阐述力求重点突出、深入浅出，语言力求简洁通俗，便于自学。

本书适用于高等农业院校动物科学、动物医学、兽药生产、饲料加工、卫生检验等专业，也适合于农牧业科技工作者阅读。

**图书在版编目(CIP)数据**

生物统计学 / 吴占福主编. —北京：科学出版社，2005.4
(21世纪高等院校教材)
ISBN 7-03-015211-5

Ⅰ. 生… Ⅱ. 吴… 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022779号

责任编辑：[illegible] 吴朝杰 / 责任校对：钟 洋
责任印制：刘士平 封面设计：卢秋红



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
双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2005年4月第 一 版 开本：787×1092 1/16
2005年4月第一次印刷 印张：17
印数：1—4 000 字数：392 000

**定价：29.80元**

(如有印装质量问题，我社负责调换〈双青〉)

# 《生物统计学》编者名单


**主　　编**　吴占福

**副 主 编**　马旭平　王成杰　曹春梅　高志花　吴秀品　杨国忠

**编　　者**　（按姓氏笔画为序）

马旭平　王　净　王成杰　白　升　刘海斌　李文海

杨国忠　吴占福　吴秀品　吴淑琴　高志花　耿光瑞

曹春梅　韩英强　薛　荣　穆秀明

# 前　言

《生物统计学》是高等农业院校动物科学、动物医学等专业的主要专业基础课之一。

为进一步适应21世纪市场经济和我国高等农业教育的发展，为适应高等农业教育改革的深化，以历年《生物统计学》的教学讲义为基础，参照国家教育部组织编写的统编教材及国内外有关专著，在编写《实用生物统计学》基础上，我们编写了《生物统计学》。在编写过程中，根据高等农业院校动物科学、动物医学等专业培养目标和相应教学计划的要求，教材内容在坚持科学性、系统性基础上，突出应用性，加强实践性，力求编写出具有农林院校特色的教学用书，使本书既保持本学科的系统性，又遵循难易适度循序渐进的教学规律；既有广泛的适用性，又具有时代特征。

本书共10章，包括概论、数据的整理与特征数、概率与理论分布、均数差异显著性检验、方差分析、次数资料分析——$\chi^2$检验、直线相关与回归、多元线性回归与相关、协方差分析、试验设计，并附有习题、实验实习指导和常用统计学用表。第一至三章是学习生物统计学的基础知识，第四至九章是统计分析方法，第十章是试验设计的基础知识。为适应高等教育的不断革新，编入了抽样方法、计算器的使用、统计软件SPSS应用的实验实习内容。本书必学和选学两部分内容共50～65学时。在编写中特别注意突出重点，阐述基本概念与基本方法，便于读者举一反三。对内容的阐述力求重点突出、深入浅出，语言力求简洁通俗，便于自学。

本书适用于高等农业院校动物科学、动物医学、兽药生产、饲料加工、卫生检验等专业，也适合于农牧业科技工作者阅读。

由于编写时间紧迫，编写人员水平有限，欠妥和错误之处在所难免，恳请读者批评与指正。

编　者

2004年11月

# 目　录



* 符号为选学内容。

# 第一章 概 论

## 第一节 生物统计与试验设计的概念

生物统计(biometry)是数理统计在生物科学中的应用,是用数理统计的原理和方法分析和解释生物界各种现象与数量资料的一门学科。即借助于数理统计的理论和方法,对生产和试验调查所得到的有变异的数据做出正确的判断,找出内在的客观规律,再以生物学观点加以解释。随着生产和科学技术的发展,生物统计的应用日益广泛,日益成为处理数据的必需手段。生物统计是研究如何科学地搜集、整理和分析数据的统计分析方法,具有很强的实用性。

人们从事动物生产和科学研究的对象总是预研究事物的一部分(样本),例如:测定畜禽的生产性能不可能全部测定,而只能抽取部分个体;药物的疗效观察也只能用少数畜禽做试验,而我们希望了解的是事物全体(总体)的特征特性。运用生物统计方法就能由部分推断全体,由个别推断一般,这种研究方法称为统计推断,是生物统计的重要内容之一。生物统计处理的数据要求具有质的共同性,不同质的事物或现象混在一起统计,会得出荒谬的结果;但同质的研究对象也不可避免地存在数量方面的差异,例如:同品种的奶牛,即使被观察牛的性别、年龄相同,但它们之间在体重、体尺等方面也各不相同。生物统计的任务就是要认识有变异的事物或现象。自然界的事物或现象间总是互相联系的,生物统计就是要研究它们间的相互关系,揭示其客观规律,从一事物或现象的观察预测另一事物或现象的发生,为充实理论和生产服务。

试验设计(experimental design)是指试验研究工作进行前应用生物统计原理,制订试验方案,选择试验动物,合理分组,使我们可以利用较少的人力、物力和时间,获得多而可靠的信息资料,得出科学的结论。生物统计与试验设计是不可分割的两部分。试验设计需要以统计的原理和方法为基础,而正确设计试验又为统计方法提供了丰富可靠的信息,两者紧密结合推断出较为客观的结论,不断地推动动物科学、动物医学、水产业和科学研究的发展。

广义的试验设计是指试验研究课题的整体设计,也就是指整个试验计划的拟定,包含课题名称、试验目的、研究依据、内容及预期达到的效果、试验方案、供试单位的选取、重复数的确定、试验单位的分组、试验的记录项目和要求、试验结果的分析方法、经济效益或社会效益的估计、已具备的条件、需要购置的仪器设备、参加研究人员的分工、试验时间、地点、进度安排和经费预算、成果鉴定、学术论文撰写等内容。狭义的试验设计主要是指试验单位(如试验的畜、禽)的选取、重复数目的确定及试验单位的分组。生物统计中的试验设计主要指狭义的试验设计。合理的试验设计能控制和降低试验误差,提高试验的精确性,为统计分析获得试验处理效应和试验误差的无偏估计提供必要的数据。

## 第二节 本课程的主要内容

本课程是动物科学、动物医学等专业的主要专业基础课之一。其主要目的是培养学生具有

动物科学试验设计的能力和对试验资料进行统计分析处理的能力，是为学习数量遗传学、动物育种学和动物饲养学、动物医学等打好基础。

本课程的内容主要分为以下几个方面。

## 一、资料的整理及统计分析

由生产或试验所得的数据资料(data)是很多的，需要加以整理和分析。整理的内容主要是检查原始数据的完整性、正确性，做次数表和统计图；并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，用这些统计量来估计总体的参数，分析资料的集中性(以平均数来表示)、离中性(以标准差来表示)以及平均数的可靠性(以标准误大小来表示)，做初步统计分析。

## 二、显著性检验

由于所获得的资料仅是一个样本，与总体间必然有一定差异，因此，必须检验样本统计量的可靠性，看它是否能代表总体，以及样本均数之间的差异主要由处理效应(treatment effect)引起的，还是主要由试验误差(experimental error)所造成。这在统计上称为显著性检验(test of significance)。

显著性检验的方法很多，常用的有以下三类：

### (一) 平均数间差异显著性检验

1908 年，英国统计学家 Cosset(1876～1937)首次以“学生”(student)为笔名，在《生物计量学》杂志上发表了《平均数的概率误差》，又连续发表了《相关系数的概率误差》(1909)、《非随机抽样的样本平均数分布》(1909)、《从无限总体随机抽样平均数的概率估算表》(1917)等论文。这些论文的完成，为“小样本理论”奠定了基础；同时，也为以后的样本资料的统计分析与解释开创了一条崭新的路子。由于 Cosset 开创的理论使统计学开始由大样本向小样本、由描述向推断发展，提出了著名的 $t$ 检验的方法和理论。因此，有人把 Cosset 推崇为推断统计学的先驱者。

在分析资料时，经常要对两组平均数进行比较，比较平均数之间是否存在着显著性的差异，即检验差异由偶然性引起的可能性有多大？如是小样本则常用 $t$ 检验法，如是大样本则用 $u$ 检验法。检验时要注意惟一差异性，即注意是否具有可比的基础。

### (二) $\chi^2$检验

1900 年，Pearson 独立地提出 $\chi^2$ 检验，又重新发现了 $\chi^2$ 分布，并提出了有名的“卡方检验法”。Pearson 获得了统计量 $\chi^2 = \sum(\text{实际次数} - \text{理论次数})^2 / \text{理论次数}$，并证明了当观察次数充分大时，$\chi^2$ 总是近似地服从自由度为 $(k-1)$ 的 $\chi^2$ 分布，其中 $k$ 是表示所划分的组数。在自然现象的范围内，$\chi^2$ 检验法运用得很广泛，经英国统计学家 Fisher 补充，成为了小样本推断统计的早期方法之一。$\chi^2$ 检验是属性资料的统计检验方法。有许多性状不能用直接测量的方法加以衡量，一般称之为属性性状。例如，在猪的杂交试验中，子代毛色的黑与白，性别中的公与母以及药物试验中的治愈或无效，均可以应用属性统计检验方法。通过对具有相同属性的计数来分析、检验它实际的观测值是否与理论值相符。

### (三) 方差分析——*F* 检验法

方差分析(analysis of variance,简称 AOV)又名变量分析,是由英国统计学家 Fisher(1890～1962)于 1923 年提出的。方差分析在科研工作中极为重要,特别是在多因素试验中,可以帮助我们分析出起主导作用的变异来源。方差分析从数学模型上看有固定模型(fixed model)和随机模型(random model)以及混合模型(mixed model)三种。进一步列出方差分析的各自的期望均方(expected mean squares,EMS),从而估计出各种效应值,是近代生物统计学科的发展。

## 三、相关与回归

统计相关法是由 Galton 创造的。关于相关研究的起因,最早是他因度量甜豌豆的大小,觉察到子代在遗传后有"返于中亲"的现象。1877 年,他搜集大量人体身长数据后,计算分析高个子父母、矮个子父母以及一高一矮父母的后代各有多少个高个子和矮个子子女,从而把父母高的后代高个子比较多、父母矮的后代高个子比较少这一定性认识具体化为父母与子女之间在身长方面的定量关系。1888 年,Galton 在《相关及其主要来自人体的度量》一文中,充分论述了"相关"的统计意义,并提出了高尔登相关函数(即现在常用的相关系数)的计算公式。研究两个变量之间相互关系的密切程度,称为相关(correlation),以相关系数来表示。例如,黄牛胸围与体重存在着一定程度的相关。胸围越大,其体重也可能越大。

1870 年,Galton 在研究人类身长的遗传时发现高个子父母的子女,其身长有低于他们父母身长的趋势;相反,矮个子父母的子女,其身长却往往有高于他们父母身长的趋势。从人口全局来看,高个子的人"回归"于一般人身长的期望值,而矮个子的人则作相反的"回归"。这是统计学上"回归"的最初含义。1886 年,Galton 在论文《在遗传的身长中向中等身长的回归》中,正式提出了"回归"概念。回归(regression)是指两个或两个以上的变量存在着从属关系,即一个变量变化时,引起另一个变量的相应变化。它们的关系可以从量的方面加以估算,这就是回归分析。

变量之间的关系可以是线性的也可以是非线性的,可以是一元的也可以是多元的。所谓多元相关与回归是指多个变量对某一个变量的影响,这在研究一些复杂的问题时很有用处。例如,猪的瘦肉率受很多因素影响,为了估测瘦肉率,可以利用多因素间的相关关系来建立多元回归方程式,从而估计出某一头猪的瘦肉率。

## 四、试验设计

自 1923 年起,Fisher(1890～1962)陆续发表了关于在农业试验中控制试验误差的论文。1925 年,他提出随机区组法和拉丁方法,到 1926 年 Fisher 发表了试验设计方法的梗概。这些方法在 1935 年进一步得到完善,并首先在卢桑姆斯坦德农业试验站中得到检验与应用,后来又被他的学生推广到许多其他科学领域。

Fisher 在创建试验设计理论的过程中,提出了十分重要的"随机化"原则。他认为这是保证取得无偏估计的有效措施,也是进行可靠的显著性检验的必要基础。所以,他把随机化原则放在极重要的地位,"要扫除可能扰乱资料的无数原因,除了随机化方法外,别无他法"。1938 年,他和耶特斯合作编制了有名的 Fisher Yates 随机数字表。利用随机数字表保证总体中每一元素有同等被抽取的机会。这样,Fisher 就把随机化原则以最明确、最具体化的形式引入统计工作

与统计研究中。

Fisher 在统计发展史上的地位是显赫的。这位多产作家的研究成果特别适用于农业与生物学领域，但它的影响已经渗透到一切应用统计学，由此所提炼出来的推断统计学已越来越被广大领域所接受。因此，美国统计学家 Kohnson 于 1959 年出版的《现代统计方法：描述和推断》一书中指出："从 1920 年起一直到今天的这段时期，称之为统计学的 Fisher 时代是恰当的"。是对多个平均数进研比较的一种统计检验方法，它的基本特点是把试验的总变异剖分为各个不同变异来源的变量，然后把处理均方与误差均方进行比较计算 $F$ 值，使处理的真实效应显示出来。

本书第十章重点讨论试验设计原理、任务和要求以及试验计划和方案的拟定，还分别介绍了一些常用的试验设计方法，如完全随机的设计、配对设计、随机单位组设计、交叉设计、拉丁方设计、正交设计。

## 五、SAS 与 SPSS 统计软件系统

SAS 是"Statistic Analysis System"的缩写，是一个用来管理分析数据和编写报告的组合软件系统，其基本部分是 SAS / BASE 软件。1966 年，美国 North Carolina 州立大学开始开发 SAS 统计软件包，1976 年该系统完成，同时成立 SAS 研究所。当初该系统只能运行于大型计算机系统，1985 年出现了当今我们广泛使用的 SAS 微机版本。SAS 系统具有统计分析方法丰富、信息储存简单、语言编程能力强、能对数据连续处理、使用简单等特点。SAS 是一个出色的统计分析系统，它汇集了大量的统计分析方法，从简单的描述统计到复杂得多的变量分析，编制了大量的使用简便的统计分析过程。SAS 系统运行的几个重要前提条件：①SAS 系统运行时要同时打开的文件较多，因此在微型计算机的系统配置文件 CONFIG.SYS 中应指定 FILES = 50 或以上。②SAS系统软件有时间租期限制，因此只有机器时间(DATE)在软件有效期内才能运行。时间租期取决于 SAS 出售版本日期，即所谓的 SAS 诞生日(BIRTHDAY)。③SAS 系统应全部安装到硬盘的 SAS 子目录下，硬盘应至少有 10M 空间。

SPSS(Statistic Package for the Social Science，社会科学统计软件包)是世界著名的统计分析软件之一。1968 年，三位美国斯坦福大学的学生开发了最早的 SPSS 统计软件系统，并基于这一系统于 1975 年在芝加哥合伙成立了 SPSS 公司。20 世纪 80 年代以前，SPSS 统计软件主要应用于企事业单位。1984 年，SPSS 总部推出了世界第一个统计分析软件微机版本 SPSS/PC$^+$，开创了 SPSS 微机系列产品的开发方向，从而确立了该软件在个人用户市场第一的地位。迄今为止，SPSS 软件已有 30 余年的成长历史，拥有全球约 25 万的产品用户，它们分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业，是世界上应用最广泛的专业统计软件。SPSS 使用 Windows 的窗口方式展示各种管理和分析数据的方法，使用对话框展示出各种功能选择项，只要掌握一定的 Windows 操作技能，并了解统计分析原理，就可以使用该软件为科研工作服务。SPSS 的基本功能包括数据管理、统计分析、图表分析、输出管理等。其过程包括描述性统计、均值比较、一般线性模型、相关分析、回归分析、对数线性模型、聚类分析、数据简化、生存分析、时间序列分析、多重响应等大类，每类中又分好几个统计过程。如回归分析中又分线性回归分析、曲线估计、logistic 回归等几个统计过程，并且每个过程中又允许用户选择不同的方法及参数。SPSS 中还有专门的绘图系统，可以根据数据绘制各种图形。

最近，伴随着 SPSS 产品服务领域的扩大和服务深度的增加，SPSS 公司已决定将它的英文全称更改为 Statistical Product and Service Solutions，意为“统计产品与服务解决方案”。到目前为止，SPSS 已具有适合于 DOS、Windows、UNIX、Macintosh、OS/2 等多种操作系统使用的产品，国内常用的是其适用于 DOS 和 Windows 的版本。SPSS for DOS 通常称为 SPSS/PC$^+$，现已较少使用。SPSS for Windows 界面友好，功能强大，使用者越来越多。SPSS for Windows 的主要版本有 SPSS V7.0、SPSS V7.5、SPSS V8.0、SPSS V9.0、SPSS V10.0、SPSS V11.0 等。SPSS V10.0 以上版本有两种结构：一种是服务器(Server)/客户机(Client)结构，由 SPSS Server 和 SPSS for Windows 两部分组成；另一种结构是单机版本，即 SPSS for Windows 标准版。

## 第三节 常用术语

### 一、总体与样本

总体(population) 是指根据研究目的确定的、符合指定条件的研究对象的全体。它是由相同性质的(个体)成员所构成的集团。构成总体中的个体是有限的称为有限总体，构成总体中的个体是无限的称为无限总体。总体所含个体(单元)的多少称为总体容量，以 $N$ 表示。例如，研究北京白鸡的产蛋量，凡是按饲养管理手册饲养的各地的北京白鸡个体的产蛋量构成一个总体。可见，总体是由相对同质的个体(单元)构成。样本 (sample) 是指从总体中抽取一定数量的个体所组成的集合。样本所含个体(单元)的多少称为样本容量，以 $n$ 表示。通常以样本容量 30 为界，$n>30$ 的称为大样本，$n<30$ 的称为小样本。生物统计要求由总体抽取一部分个体应遵循“随机抽样”的原则，随机就是总体中的所有个体均有同等机会被抽到。随机抽到的那部分个体就称为随机样本，这个抽样过程称为随机抽样。

### 二、参数与统计量

参数(parameter) 是指由总体计算的用来描述总体的特征性数值。它是一个真值，通常用希腊字母表示。如总体平均数以 $\mu$ 表示，总体标准差以 $\sigma$ 表示。统计量(statistics) 是指由样本计算的用来描述样本的特征性数值。它受抽样波动的影响，是总体参数的估计值，常以英文字母表示，如样本平均数为 $\bar{x}$，样本标准差为 $s$。

### 三、误差与错误

误差(error)是指试验中由无法控制的非试验因素所引起的差异。它是不可避免的，试验中只能设法减少，而不能消除。误差按来源可分为两类：一类为随机误差(random error)，也叫偶然误差，这是由于许多无法控制的内在和外在的偶然因素所引起的差异。它具有随机性质，在试验中即使小心管理也难以完全消除。它影响试验的精确性。只有用增加试验重复、合理分组等措施来减少随机误差。另一类为系统误差(systematic error)，也叫片面误差(lopsided error)，这是由试验条件不一致造成的。如试验动物的初始条件，年龄、初始重、性别、健康状况等相差较大，饲料种类、品质、数量、饲养条件未控制相同所引起。系统误差影响试验的准确性。这两类误差可统称为试验误差，是难以避免的非本质性的差异，不同于处理条件所造成的实质性差异。试验误差的来源包括三个方面：一是试验材料的固有差异；二是饲养管理和操作技术上的

不一致所引起的差异；三是外界条件的差异。例如，在同一处理组内，试验动物尽管做到在性别、年龄、体重等方面的一致，但试验结果同组不同个体在指标效应上仍有差异，这就是上述两类误差造成的。虽然难以避免，如果设计更为合理，可以减少试验误差，提高精确性与准确性。

错误(mistake)是指试验过程中人为的作用所引起的差错，在试验中完全可以避免，是指试验或调查观察时由于过失造成的差异，如读错刻度、记错数据和仪器不准确等。只要责任心强、细心操作、遵循设计的标准和方法，错误是可以避免的，也是应该消除的。

### 四、精确性与准确性

精确性(precision)是指试验或调查中同一试验指标或性状的重复观察值彼此的接近程度。若观测值彼此接近，即任意两个观测值 $x_i$、$x_j$ 相差的绝对值 $|x_i - x_j|$ 小，则观测值精确性高；反之则低。好比打靶，弹点密集靶上某处叫射击的精确性高；弹点散开，叫精确性低。试验观察中各次测量值接近即试验误差小，其精确性高，重复测量值间不接近即试验误差大，其精确性低。准确性(accuracy) 是指试验或调查中某一试验指标或性状的观察值(统计量)与真值(或总体参数)之间的接近程度。设某一试验指标或性状的真值为 $\mu$，观测值为 $x$，若 $x$ 与 $\mu$ 相差的绝对值 $|x - \mu|$ 小，则观测值 $x$ 的准确性高；反之则低。好比打靶，每发子弹(每次观察值)都命中或接近靶心(真值)，叫准确性高；否则，准确性低。准确性与精确性不是相等的概念，统计上以统计量接近参数的程度衡量准确性，以统计量的变异程度衡量精确性。

## 第四节 生物统计的功用

生物统计已经广泛地应用于动物生产和科学研究中，其主要功用可以归纳为下列几个方面。

### 一、提供整理与描述数据的科学方法

由试验或调查所得到的数据不仅是大量的，而且也是杂乱的，很难从中提取有用的信息。生物统计能够提供科学的整理方法，将数据化繁为简，使之系统化、条理化，同时提供具体的描述方法，或用图表，或用简单的数值，或用公式来阐述数据资料的内在规律、事物的本质及其彼此的关系。例如，面对大量的某品种奶牛产奶量的原始数据，很难从这些杂乱的数据中看到什么规律，只有通过科学的整理与描述，才能了解该品种奶牛各胎产奶量的一般情况及各胎产奶量高低的变化规律，以及胎次产奶量的变异特征，是指导奶牛生产有用的信息。

### 二、提供由样本推断总体的科学方法

调查或试验研究所取得的数据都属样本资料。例如，调查某品种奶牛的产奶量，可能取 50 头或 100 头奶牛产奶量的数据。这些数据只是该品种奶牛总体的一部分。试验研究某配套肉鸡饲喂发酵血粉的增重效果，可能取 100 只或 500 只肉鸡供试，试验的结果也仅是该肉鸡总体的一部分鸡的饲喂效果。这都是样本的数据，在一定程度上可以反映总体，但用样本数据确实地反映总体特征还有偏差，即存在抽样误差。由于样本与总体间有内在联系，所以由样本推论总体是可能的。生物统计就是用这种内在联系，提供了由样本推断总体的科学方法。

## 三、提供鉴定试验处理效应的科学方法

研究两种商品肉猪饲料配方的饲喂试验，供试猪100头，在品种、性别、年龄、初始重等条件一致情况下随机分为两组，每组50头，各喂一种配方饲料，试验结束后观察其增重与饲料消耗。由于前述“试验误差”的存在，该试验结果不同程度地混有试验误差，那么，两组试验猪在试验指标上的差异，是试验误差造成的非本质差异，还是确因饲料配方不同所造成的本质性差异？生物统计可以提供鉴别试验处理效应的科学方法，将试验结果的两类差异予以区分，从而获得可靠的结论。

## 四、提供相关和回归分析的科学方法

客观事物是普遍联系的，我们在日常生活和科学研究中，经常可以看到有些事物间存在着一定的关系。例如，动物体重与日龄的关系，日龄越大，其体重也可能越大。动物生产中为了方便饲养管理，常常把幼畜成批集中断奶。由于出生日期不同，同一天断奶的幼畜生长日龄往往不同，这样它们的体重就不好比较，需要矫正到同一日龄的体重。根据体重与日龄的关系，我们先求出体重依日龄的回归方程，并用回归方程算出各日龄体重估计值，而后以标准断奶日龄的体重估计值与各日龄体重估计值的比值作为矫正系数。生物统计学提出了相关和回归分析科学方法。

## 五、提供调查或科学试验设计的原则

无论调查研究或科学试验，事先应有周密的计划和合理的试验设计，否则就不可能得到正确可靠的结果。例如，调查某品种奶牛的产奶量，仅取少数几头奶牛为调查对象，显然是不够的；但头数过多，人力、物力又造成浪费。即使头数不少，若集中于某一地域取样，其代表性就差。科学试验也涉及试验分组，试验动物的选择，组内试验动物多少等若干问题。生物统计可以从理论上提供如何设计试验和如何实施试验的原则，可以做到尽量降低试验误差，使试验结果能代表总体，而且从试验所得的数据中能够无偏地估计出试验处理效应和误差效应的估值，能够最大限度保证以较少的投入获得多而正确的结论。

## 六、提供制订规划和进行决策需要的依据

动物生产中制订育种计划时，常常要规划出经若干年的选育，某个经济性状提高到何等水平，这就必须将大量资料经过统计学处理，得到某些表型参数与遗传参数，在此基础上估测经选育后该性状改进的程度；在市场经济体制下，动物养殖场既要考虑利润，也要考虑畜产品的滞销以及资金周转能力，就必须首先充分地做好市场调查，经过统计分析取得依据，以正确决定饲养动物的规模及其畜产品的数量。上述两例说明，在养殖业中生产的决策与规划的制订离不开生物统计分析。因此，生物统计学是一门实用性强的科学。我们应当深入地、正确地理解其概念和熟练掌握其方法，应当树立起统计思想。

## 习 题

1. 什么叫生物统计？生物统计在动物科学、动物医学、水产试验或调查中有哪些功用？

2. 解释下列术语：①总体与样本；②参数与统计量；③变数与变量；④错误与误差。

3. 生物统计工作中的准确性与精确性有什么区别？

4. 简述 SAS 与 SPSS 统计软件在生物科学上的应用。

# 第二章　数据的整理与特征数

## 第一节　资料的收集

收集资料(collection data)是统计工作的第一步,也是整个分析工作的基础。要使收集的原始资料(raw data)具有应用价值,则必须根据各类资料的要求,力求准确观察度量。如果收集的资料不正确或不完整,则用任何统计分析方法也无法弥补。

### 一、资料的来源

原始资料一般来源于以下三个方面。

#### (一) 经常性记录

在日常工作中,将工作情况记录下来,经过一段时间,可以从中总结出许多规律性的东西,还可以从中发现一些新的、异常的东西。如兽医院门诊、住院、临床化验,畜禽场的日常记录表格,统计报表等。

#### (二) 试验研究记录

科学研究是资料的重要来源之一。为了推动畜牧业的发展,常常需要进行科学研究。例如畜禽、水产新品种的选育、新的饲养管理技术研究、兽药疗效研究、畜禽生理指标正常值范围的制定等等。在这些研究中,一般要通过设置各种类型的试验来获取样本资料,其规模不一定很大,以便于控制条件。所做的记录,便是随机样本的资料。

#### (三) 调查记录

调查是获取资料的又一重要途径。调查方法有两种:一种是普查;一种是抽样调查。普查是指对研究对象的每个个体都进行观测的一种调查,比如人口普查、牲畜存栏头数普查等。普查一般要求在一定的时间或范围内进行,主要是摸清研究对象的全部情况。在生物学研究中,普查仅在极少数情况下应用,多数情况还是抽样调查,而生物统计所涉及的主要内容就正好是有关试验和抽样调查的原理和方法。

抽样调查是一种非全面调查,它是根据一定的原则对研究对象抽取一部分个体进行观测,把得到的数据资料作为样本进行统计处理,然后利用样本特征数对总体做出推断和估计。要使样本能无偏地估计总体,除了使样本容量尽可能大外,重要的是采用科学的抽样方法,抽取有代表性的样本。实践证明,正确的抽样调查不仅能节约人力、物力和财力,而且与相应的统计分析方法相结合,可以做出比较准确的估计和推断。

从有限总体中做抽样调查,一般有随机抽样、机械抽样和典型抽样三种。

1. 随机抽样　生物统计学是以概率论和数理统计的原理和方法为依据,它要求用来推断