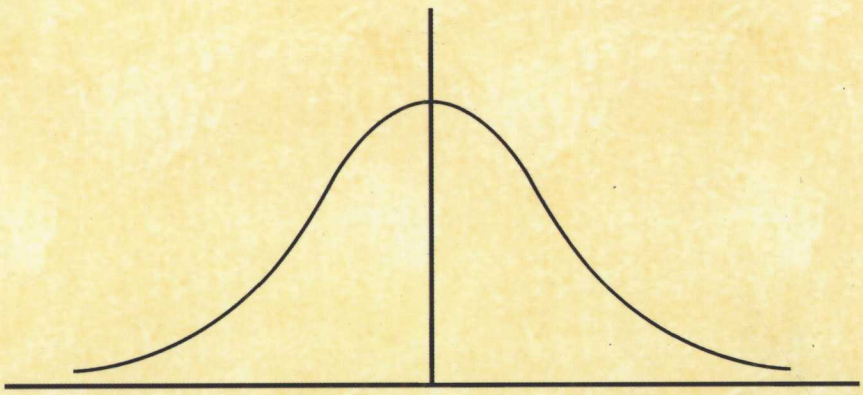
普通高等教育“十一五”国家级规划教材

《生物统计学》（第四版）立体化教材

生物统计学学习指导

李春喜　姜丽娜　邵　云　编著



科学出版社
www.sciencep.com

普通高等教育“十一五”国家级规划教材

《生物统计学》（第四版）立体化教材

生物统计学学习指导

李春喜　姜丽娜　邵　云　编著

科学出版社

北京

内 容 简 介

本书是普通高等教育“十一五”国家级规划教材《生物统计学》（第四版）立体化教材项目之一。本书旨在为《生物统计学》的学习提供概要性总结、资料扩充、难点解析，通过增加具体实例和对习题的解答，帮助学生进一步理解和掌握基本概念、基本内容和基本方法。其内容编排与教材各章内容相对应，共14章。内容包括目的要求、内容提要、难点评析、例题解析、习题解答和自我测验6部分。书后附有自我测验答案。

本书可作为综合性大学、师范院校生物类及其相关专业的本科生学习《生物统计学》的配套学习辅导书，也可作为从事生命科学、生物工程、农业科学、林业科学、医学、畜牧兽医、水产科学等专业的科研工作者、教师和研究生的参考书。

图书在版编目(CIP)数据

生物统计学学习指导/李春喜，姜丽娜，邵云编著. —北京：科学出版社，2008

（普通高等教育“十一五”国家级规划教材《生物统计学》（第四版）立体化教材）

ISBN 978-7-03-021949-7

Ⅰ. 生… Ⅱ. ①李…②姜…③邵… Ⅲ. 生物统计-高等学校-教学参考资料 Ⅳ. Q-332

中国版本图书馆CIP数据核字（2008）第067888号

责任编辑：甄文全/责任校对：刘小梅

责任印制：张克忠/封面设计：耕者设计工作室

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

铭浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

*

2008年7月第 一 版 开本：787×1092 1/16

2008年7月第一次印刷 印张：14 1/4

印数：1—4 000 字数：325 000

定价：25.00元

（如有印装质量问题，我社负责调换〈长虹〉）

前　言

生物统计学是一门实用性很强的工具性课程。学习生物统计学需要举一反三，既要对生物统计学的基本概念、基本内容有一定的理解和掌握，也要通过例题学习来了解不同统计问题的解题思路和解题方法，更要通过习题练习来熟练掌握这些方法。因此，编写一本与《生物统计学》教材配套的学习指导书就显得十分必要。多年来，作者编写的《生物统计学》得到了广大读者的厚爱，其使用范围也在不断扩大。但由于在教材编写时受课程教学的限制，内容体系只能涉及基本的统计问题和部分扩展性知识，用于介绍和解析各种统计方法的例题也只能选择少部分经典实例，这就不可避免地会使一些问题得不到细致分析，部分内容的叙述和公式推导也不够深入。对习题部分，也只是给出简单的参考答案，而没有细致的解题过程。因此，不断有读者提出上述问题，询问其相关解决方法。为了有效解决上述问题，更好地配合生物统计学教学，我们结合《生物统计学》（第四版），组织编写了这本《生物统计学学习指导》。

本书在内容体系安排上与《生物统计学》（第四版）保持一致，共14章。内容包括目的要求、内容提要、难点评析、例题解析、习题解答和自我测验等6部分。书后附有自我测验答案。目的要求部分提出了本章要达到的基本要求；内容提要部分概要地介绍了本章的主要知识点和难点、关键点；难点评析部分是对本章的疑难问题进行较细致的剖析，适当扩充了部分内容，对重要问题的解题思路、解题方法以及注意事项作了介绍；例题解析部分是在教材例题的基础上，重点选取部分代表性的例子对其解题过程进行了系统分析、计算和评述；习题解答部分对教材每章后所附思考练习题一一进行了详细解答。自我测验部分则是结合《生物统计学》各类考题形式，设计了部分题目，主要包括填空、判断、名词解释、单项选择和计算等5种类型，供读者练习。书后附有自我测验答案，供参考。

本书在编写和出版过程中，得到了科学出版社甄文全先生、周辉先生、王国栋先生和河南师范大学教务处、生命科学学院的大力支持，在此一并表示感谢。

由于作者在编写学习指导书时缺乏经验，掌握的素材也不够多，在内容设计与编排、解题方法和技巧等方面会存在许多不足之处。殷切希望广大读者对书中的疏漏和不妥之处及时给予批评指正，以便本书再版时进一步完善。

李春喜　姜丽娜　邵　云
2008年3月于河南师范大学

目　录

第1章 概 论

1.1 目的要求

(1) 掌握生物统计学的基本概念，了解生物统计学的主要内容和基本作用。

(2) 了解生物统计学的发展概况。

(3) 掌握统计学常用的术语。

1.2 内容提要

生物统计学是数学语言在生命科学领域的具体应用，它是运用数理统计的原理和方法对生物有机体开展调查和试验，目的是以样本的特征来估计总体的特征，对所研究的总体进行合理的推论，得到对客观事物本质和规律性的认识。生物统计学的研究包括试验设计和统计分析两大部分，其作用主要有4个方面：提供整理、描述数据资料的科学方法并确定其数量特征，判断试验结果的可靠性，提供由样本推断总体的方法以及试验设计的原则。统计学的发展经历了古典记录统计学、近代描述统计学和现代推断统计学3个阶段。在统计学中要掌握总体、个体与样本，变量与常数，参数与统计数，主效与互作，误差与错误，准确性与精确性这几组常用基本术语的概念和区别。

生物统计学是以样本的统计数估计和推断总体的参数，对总体和样本的理解是本章的重点内容。总体是指研究对象的全体，从总体中抽出的若干个体即组成了样本。由此可以看出，生物统计学的研究包含了两个过程，一个是从总体抽取样本的过程，即抽样过程；另一个是从样本的统计数到总体参数的过程，即统计推断的过程。

在生物学研究中，由于不可控制因素所引起的观测值偏离真值的差异称为误差，误差分为随机误差和系统误差两类。正确理解两类误差产生的原因，才能在生物学试验中尽量减少误差。

1.3 难点评析

1. 正确理解统计学是应用数学的一个分支

统计学是研究数据资料的搜集、整理、分析和解释的科学。搜集资料是取得数据资料的过程。例如，通过抽样调查或科学试验获取资料。正确的结论只能来自高质量的资料。整理资料是对数据资料进行初步归纳分析，找出数据资料的基本特征，并以图、表等适当的形式表示这些数据资料，以便对数据的基本特征有清晰、直观的了解。分析资料是针对要研究的问题，通过对数据的深入分析，从数据资料中获取所需有关信息的过程。解释资料是在分析结果的基础上对所研究的问题作出统计推断。综上所述，统计学是与数据密切相关的科学，因而可将统计学看成是应用数学的一个分支。

2. 生物统计学是生物科学研究的基本工具

生物统计学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一门学科。生物统计学是生物科学研究和应用必不可少的基本工具，这是由生物现象的基本特征所决定的，生物现象有如下基本特征：

（1）变异性。遗传和变异是生物的两大基本现象，在任何两个生物个体之间均存在差异。

（2）不确定性（随机性）。生物个体之间的差异往往由很多偶然因素造成，因而是随机的，不能准确预测。

（3）复杂性。造成生物变异性的因素通常有很多，既有遗传方面的，也有环境方面的。从遗传上来说，生物个体所携带的基因系统是复杂的，从上代传递给下代的方式也是多种多样的，人们至今对生物的遗传机制也还没有一个完全清楚地认识；从环境上来说，生物的很多性状都受到各种环境因素的影响，其中有的环境因素是可以人为控制的，但更多的是不可控制的、对不同个体所产生的随机性影响的环境因素。

生物现象的这些特点决定了不能通过描述性的定性科学或决定性的数量来解决生物学领域中的众多问题。只能通过一定数量的数据资料进行分析才能得到答案，而只有统计学才能告诉我们如何通过科学的调查或试验获得高质量的数据，如何对所获得的数据进行分析，如何根据分析的结果做出尽可能可靠的结论。

3. 正确理解总体、个体和样本的概念

总体是研究对象的全体，是具有相同性质的个体所组成的集合；个体是组成总体的基本单元；样本是由总体中抽出的若干个体所构成的集合。

总体指的是“统计总体”，是指一个统计问题研究对象的全体，它是具有某种（或某些）共同特征的元素的集合。总体的定义是根据所要研究的问题而定的，如果研究某一个班级学生的身高，则总体由该班全体学生的身高组成；如果研究该班学生的体重，则总体由该班全体学生的体重组成。由此可以看出，总体是由要研究的随机变量的所有可能取值构成的。

总体中的每一个研究对象称为个体。组成总体的个体，至少要具有某一共同的属性或特征，这是区分不同统计总体和辨别总体真假的基本依据。有时，关于总体的确定还需要事先制订某种规则或标准，对总体和它包括的边界作出详细的说明。在具体的研究中，往往只关心研究对象的某个指标（即变量）。因此，总体又可以是指变量或指标取值的全体。这样，对同一个研究对象，如果要调查几个方面的情况，就可以把它们分成几个总体分别进行研究。例如，上面所提到的，对某个班级的学生，既要了解他们的身高情况，又要研究他们的体重情况，就有身高总体和体重总体两个方面。

总体范围的大小与统计研究的目的和任务有关，也与总体存在的规模有关。按总体中所含个体的数目是否有限，可将总体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，个体有限的总体称为有限总体。也可以从抽象意义上来理解无限总体。例如，研究某种药物对某种疾病的治疗效果是有效的还是无效的，将利用一些发病个体进行药效试验。这部分个体可以看成是来自一个假想总体的样本，这个假想总体由

所有发病个体对此药物的治疗效果构成，个体数目可能是无限的。但这个总体并不现实存在，因为并未对所有发病个体用药，但从理论上可以对所有发病个体用药。

如果按照总体的构成是否随时间的变化而变化来划分，可将总体分为动态总体和固定总体；按总体是否由现实存在的研究对象所构成的来划分，可将总体分为现实总体和假想总体。例如，研究河南省2007年出生婴儿的体重，则总体由河南省2007年所有出生婴儿的体重构成。这是一个有限总体、固定总体和现实总体。如果在这个总体的定义中取消时间的限制，则这个总体就是一个无限总体（因为时间可以无限地延续下去）和动态总体。

统计分析的目的就是要对总体的特征、不同总体间的差异等作出推断。但由于总体往往很大，在实际的统计分析中，遇到的总体大多是无限总体、动态总体和假想总体，不可能得到总体中全部个体的数据资料。通常的做法是从总体中按一定方法抽取部分具有代表性的个体，这部分抽自总体的个体所构成的集合即是样本。由此也可以看出，统计分析的基本任务就是通过对样本的分析来推断总体。

应该指出的是，在一次调查中，总体是唯一确定的，而样本却是随机变化的。从一个总体中可以抽取多个样本，这取决于样本容量、抽样方式、推断精确度的要求等诸多因素。

4. 正确理解变量的概念

变量是相同性质的事物间表现差异性的某种特征。总体内部的各个体，除了具有某种或某些共同属性外，在调查项目具体表现上存在的差异称为变异。例如，研究某个班级学生的体重情况，体重是调查项目，而每个学生的体重是不完全一样的，这就是变异现象。从某种意义上说，变异是统计存在的前提，若是群体中的各个个体无论哪方面的表现都一样，显然就没有进行统计工作的必要了。统计工作的重要任务，就是通过对存在变异的大量现象的调查研究，剔除偶然性因素的影响，最终得到对总体必然性的认识。因此，变异是相对于调查结果而言的，而变量则是相对于调查项目来说的。从广义上讲，凡是能够取不同的值或取值多于一个以上的量均可以称为变量。例如，身高、体重、温度、酶的活性、叶片叶绿素的含量等都是变量。

5. 正确理解准确性和精确性的概念与区别

准确性指在调查或试验中某一试验指标或性状的观测值与真值的接近程度；精确性是指调查或试验中同一试验指标或性状的重复观测值彼此接近程度的大小。准确性不等于精确性。准确性反映了观测值或估计值与真值的接近程度，而精确性则是反映多次重复测定值的变异程度。通常样本所属总体的真值是未知的，只能由样本的统计数来估计和代替，因此准确性无法衡量；而精确性则可以通过样本的统计数据进行衡量。此外，精确性与数据的有效位数有关。有的重复观测值在有效位数较小时（如仅取整数）精确性很高，但如果增加有效位数后，差别就显示出来了。例如，对一个植株个体高度的两次测量值分别为85.8cm和86.3cm，二者之间的差为0.5cm，但如果只取整数，二者就没有差别了。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g为单位的质量，g以下则不敏感，而一个电子天平可以称得

以 0.001g 或 0.0001g 为单位的质量。如果要取得高的精确性，需要测量仪器精度较高，同时还要求保留一定的有效位数，这无疑在一定程度上加大了工作量。

1.4 习题解答

习题 1.1 什么是生物统计学？生物统计学的主要内容和作用是什么？

答 生物统计学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，是研究生命过程中以样本来推断总体的一门学科。

生物统计学主要包括试验设计和统计分析两大部分的内容。其基本作用表现在以下 4 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；②判断试验结果的可靠性；③提供由样本推断总体的方法；④提供试验设计的一些重要原则。

习题 1.2 解释以下概念：总体、个体、样本、样本容量、变量、参数、统计数、效应、互作、试验误差。

答 总体是具有相同性质的个体所组成的集合，是指研究对象的全体。

个体是组成总体的基本单元。

样本是从总体中抽出的若干个个体所构成的集合。

样本容量是指样本个体的数目。

变量是相同性质的事物间表现差异性的某种特征。

参数是描述总体特征的数量。

统计数是描述样本特征的数量。

效应是由因素而引起试验差异的作用。

互作是指两个或两个以上处理因素间的相互作用产生的效应。

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，可以分为随机误差和系统误差。

习题 1.3 随机误差与系统误差有何区别？

答 随机误差也称为抽样误差或偶然误差，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。随机误差可以通过试验设计和精心管理设法减小，而不能完全消除。

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，在某种程度上是可控制的。

习题 1.4 准确性与精确性有何区别？

答 准确性也称为准确度，指在调查或试验中某一试验指标或性状的观测值与其真值接近的程度。精确性也称精确度，指调查或试验中同一试验指标或性状的重复观测值彼此接近程度的大小。

准确性是说明测定值对真值符合程度的大小，用统计数接近参数真值的程度来衡量。精确性是反映多次测定值的变异程度，用样本中的各个变量间变异程度的大小来衡量。

1.5 自我测验

一、填空

1. 变量按其性质可以分为_____变量和_____变量。
2. 样本统计数是总体_____的估计值。
3. 生物统计学是研究生命过程中以样本来推断_____的一门学科。
4. 生物统计学的基本内容包括_____、_____两大部分。
5. 统计学的发展过程经历了_____、_____、_____3个阶段。
6. 生物学研究中，一般将样本容量_____称为大样本。
7. 试验误差可以分为_____、_____两类。

二、判断

(　　) 1. 对于有限总体不必用统计推断方法。
(　　) 2. 资料的精确性高，其准确性也一定高。
(　　) 3. 在试验设计中，随机误差只能减小，而不可能完全消除。
(　　) 4. 统计学上的试验误差，通常指随机误差。

三、名词解释

样本　　总体　　连续变量　　非连续变量　　准确性　　精确性

第 2 章　试验资料的整理与特征数的计算

2.1　目 的 要 求

（1）熟悉不同类型资料的整理和相关统计图表的制法。

（2）掌握常用几种平均数和变异数的基本概念及计算方法。

2.2　内 容 提 要

试验资料的搜集与整理是数据资料处理的首要环节。根据生物的性状特征，试验资料可分为数量性状资料和质量性状资料两类。数量性状资料是由计数或测量的方法得到的，又分为计数资料（非连续变量资料）和计量资料（连续变量资料）。质量性状资料（属性资料）常常经过数量化再进行统计分析。试验资料搜集的常用方法有调查和试验，其中，调查又可分为普查和抽样调查两种方式。资料的整理一般需经过对原始资料的检查、核对，确保资料正确无误后即可制作次数（频率）分布表和次数（频率）分布图。作次数（频率）分布表时，根据资料分类不同，计数资料可用单项式分组法、计量资料用组距式分组法分组，再统计各组的次数，计算其频率和累积频率，最终制成表格。次数（频率）分布图主要有适合于计数资料和属性资料的条形图、饼图，适合于计量资料的直方图、多边形图，反应变量间相关性及变化趋势的散点图等。通过制作统计图表可以定性地反映资料的特征，但要定量描述其特征，还要进一步计算资料的特征数。

试验资料均具有集中性和离散性两种基本特征。平均数是反映集中性的特征数，变异数是反映离散性的特征数。常用平均数包括算术平均数、中位数、众数和几何平均数等。算术平均数具有离均差之和等于零和离均差平方和为最小等基本性质，可以用直接计算法、减去（加上）常数法和加权计算法来计算。常用变异数包括极差、方差、标准差和变异系数等，极差是资料中最大值和最小值之差，计算简单，但只能反映数据的最大波动范围；方差等于观测值离均差的平方和除以其自由度，可以反映出资料中每一个观测值的变异；标准差是方差的平方根，其单位和变异程度与平均数相符，是表示资料变异程度的一项重要指标；用标准差再除以其平均数即为变异系数，变异系数是变量的相对变异量，可以进行平均数相差悬殊或单位不同的资料间变异程度的比较。

2.3　难 点 评 析

1. 关于自由度

在方差计算公式中，样本的方差不以样本容量 n 作为除数，而是以自由度 $n-1$ 作除数。这是因为通常所掌握的资料，大多不知其总体平均数 μ 的数值，常用样本平均数

$\bar{x}$ 来估计 μ。由于 $\bar{x}$ 与 μ 总有差异，根据平均数的基本性质——离均差平方和为最小，如把 μ 看成 a，则比 $\sum(x-\mu)^2$ 为小。因此由公式 $\dfrac{\sum(x-\bar{x})^2}{n}$ 算出来的方差对总体所作的估计是偏小的，用 $n-1$ 来代替 n，去除离均差的平方和，可以避免偏小估计的问题，从而实现样本方差对总体方差的无偏估计。在统计上，自由度（$df=n-1$）是指样本内独立而能自由变动的观测值的个数。在计算 n 个观测值的样本方差时，每个 x 与 $\bar{x}$ 比较，虽有 n 个离均差，但只有 $n-1$ 个是自由变动的，最后一个离均差由于受 $\sum(x-\bar{x})=0$ 的限制不能自由变动。例如，5 个观测值的样本，如果 4 个离均差为 2，3，1，−2，则第 5 个离均差一定为 −4；如果 4 个离均差为 1，−3，4，−4，则第 5 个离均差一定为 2，这样才能使 $\sum(x-\bar{x})=0$ 成立。由于能自由变动的离均差是 4 个，故自由度为 4，即自由度为 $n-1$。在计算其他统计数时，如果受到 k 个条件的限制，则其自由度应为 $n-k$。

2. 关于标准差

在标准差的特性中，“如果对各观测值加上或减去一个常数 a，其标准差不变；如果给各观测值乘以或除以一个常数 a，则所得到的标准差扩大或缩小了 a 倍”可以通过以下公式推导得到。

已知一组观测值，其标准差 $s=\sqrt{\dfrac{\sum x^2-(\sum x)^2/n}{n-1}}$。若将每一个观测值 x_i 加上常数 a，把 x_i+a 代入标准差公式中，则有

$$s'=\sqrt{\frac{\sum(x+a)^2-\left[\sum(x+a)\right]^2/n}{n-1}}$$

$$=\sqrt{\frac{\sum x^2+2\sum ax+\sum a^2-(\sum x+\sum a)^2/n}{n-1}}$$

$$=\sqrt{\frac{\sum x^2+2a\sum x+na^2-\left[(\sum x)^2+2na\sum x+n^2a^2\right]/n}{n-1}}$$

$$=\sqrt{\frac{\sum x^2-(\sum x)^2/n}{n-1}}=s$$

s' 与原标准差 s 相等。若每个观测值减去常数 a，可同理论证。

若将每一个观测值 x_i 乘上常数 a，将 ax_i 代入标准差公式中，则有

$$s''=\sqrt{\frac{\sum(ax)^2-\sum(ax)^2/n}{n-1}}=\sqrt{\frac{a^2\sum x^2-a^2\sum x^2/n}{n-1}}$$

$$=\sqrt{\frac{a^2\left[\sum x^2-(\sum x)^2/n\right]}{n-1}}=a\sqrt{\frac{\sum x^2-(\sum x)^2/n}{n-1}}=as$$

s'' 为原标准差 s 的 a 倍。若每个观测值除以常数 a，可同理论证。

2.4 例题解析

例题 2.1 现有 126 头基础母羊的体重（单位：kg）资料见表 2-1。

表 2-1

53.0	50.0	51.0	57.0	56.0	51.0	48.0	46.0	62.0	51.0	61.0	56.0	62.0	58.0	46.5
48.0	46.0	50.0	54.5	56.0	40.0	53.0	51.0	57.0	54.0	59.0	52.0	47.0	57.0	59.0
54.0	50.0	52.0	54.0	62.5	50.0	50.0	53.0	51.0	54.0	56.0	50.0	52.0	50.0	52.0
43.0	53.0	48.0	50.0	60.0	58.0	52.0	64.0	50.0	47.0	37.0	52.0	46.0	45.0	42.0
53.0	58.0	47.0	50.0	50.0	45.0	55.0	62.0	51.0	50.0	43.0	53.0	42.0	56.0	54.5
45.0	56.0	54.0	65.0	61.0	47.0	52.0	49.0	49.0	51.0	45.0	52.0	54.0	48.0	57.0
45.0	53.0	54.0	57.0	54.0	54.0	45.0	44.0	52.0	50.0	52.0	52.0	55.0	50.0	54.0
43.0	57.0	56.0	54.0	49.0	55.0	50.0	48.0	46.0	56.0	45.0	45.0	51.0	46.0	49.0
48.5	49.0	55.0	52.0	58.0	54.5									

试将该资料整理成次数分布表，并绘制次数分布图。

解 本题为计量资料的整理，故应采用组距式分组法，制作次数分布表，绘制直方图或多边形图，实际应用时只绘制其中一种统计图即可。

（1）制次数分布表（表 2-2）。

表 2-2

组 限/kg	组中值/kg	次数(f)	频率	累积频率
36.0 ～	37.5	1	0.0079	0.0079
39.0 ～	40.5	1	0.0079	0.0158
42.0 ～	43.5	6	0.0476	0.0634
45.0 ～	46.5	18	0.1429	0.2063
48.0 ～	49.5	26	0.2063	0.4162
51.0 ～	52.5	27	0.2143	0.6269
54.0 ～	55.5	26	0.2063	0.8332
57.0 ～	58.5	12	0.0953	0.9285
60.0 ～	61.5	7	0.0556	0.9841
63.0 ～	64.5	2	0.0159	1.0000
合 计		126	1.0000	

全距为：

$$R = 65.0 - 37.0 = 28.0(\text{kg})$$

确定组数和组距、根据样本容量与分组数关系表，初步取 10 组，则组距＝全距/组数＝28.0/10＝2.8≈3(kg)。

本题资料中最小值为 37.0，第一组的组中值取 37.5，因组距已确定为 3，则第一组的下限＝37.5－(1/2)×3＝36.0，之后每 3kg 分为一组，则可分组为 36.0～，39.0～，…，63.0～。将 126 个数据资料归组，统计各组次数，并计算各组频率和累积频率，列入表 2-2 中。

根据表 2-2 中的数据，绘制直方图和多边形图，如图 2-1 与图 2-2 所示。

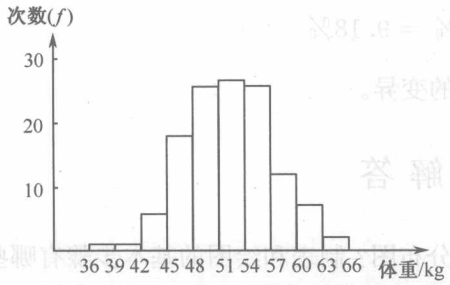


图 2-1　126 头基础母羊体重的次数分布直方图

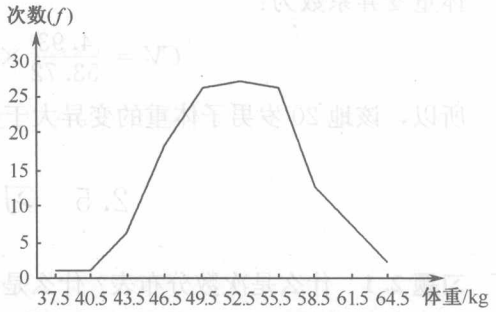


图 2-2　126 头基础母羊体重的次数分布多边形图

从次数分布表和次数分布图中可看出，126 头基础母羊体重资料分布的一般趋势为，体重的变异范围为 37.0～65.0kg，各组分布数据不均，大部分母羊的体重为 48.0～57.0kg，占所有观测值的 62.69%，即表格或图形的中部，并向两侧逐渐减少。

例题 2.2　测得 1960～1972 年越冬三代棉红铃虫在江苏东台的羽化高峰期（单位：日）依次为（以 6 月 30 日为 0）8，6，10，5，6，6，10，－1，12，11，9，1，8。试求其平均数、极差、标准差和变异系数。

解　本题为 13 个数据特征数求解的问题，带入平均数、极差、标准差和变异系数的计算公式即可。

平均数为：

$$\bar{x}=\frac{1}{n}\sum_i x_i$$

$$=\frac{8\times2+6\times3+10\times2+5\times1-1+1+12+11+9}{13}=7(\text{日})$$

极差为：

$$R=12-(-1)=13(\text{日})$$

标准差为：

$$s=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}=\sqrt{\frac{(8-7)^2+(6-7)^2+\cdots+(8-7)^2}{13-1}}=3.79(\text{日})$$

变异系数为：

$$CV=\frac{s}{\bar{x}}\times100\%=\frac{3.78}{7}\times100\%=54.1\%$$

例题 2.3　某地 20 岁男子 100 人，其身高平均数为 166.06cm，标准差为 4.95cm；其体重平均数为 53.72kg，标准差为 4.93kg。比较身高与体重的变异情况。

解　本题为两个资料变异程度比较的问题。若仅从标准差绝对量大小上看，身高的变异大于体重，但因两个资料的度量单位不同，而极差、方差、标准差都是绝对变异量，此时不适用，故应使用表示相对变异量的变异系数来比较，最终结果体重的变异大于身高。

身高变异系数为：

$$CV=\frac{4.95}{166.06}\times100\%=2.98\%$$

体重变异系数为：

$$CV = \frac{4.93}{53.72} \times 100\% = 9.18\%$$

所以，该地 20 岁男子体重的变异大于身高的变异。

2.5 习题解答

习题 2.1 什么是次数分布表？什么是次数分布图？制表和绘图的基本步骤有哪些？

答 将数据资料分组，统计各组次数，并计算其频率和累积频率，制成表格，即为次数分布表。若将分组作横坐标，次数作纵坐标，将各组及其次数用柱形、线段、点等形状表示在坐标系中，即为次数分布图。

作次数分布表时，首先根据资料的性质分组：计数资料使用单项式分组法，直接用样本变量一个或几个自然值分组；计量资料使用组距式分组法，计算全距，确定组数和组距，确定组限，再进行分组。然后将资料归在各组内，统计各组的次数，计算其频率和累积频率，制成次数分布表。

作次数分布图时，资料分组方法同次数分布表，将分组作横坐标，次数作纵坐标，将分组及其次数用柱形、线段、点等形状表示在坐标系中，即绘成次数分布图。其中计数资料可做成条形图、饼图等，计量资料可做成直方图、多边形图等，研究变量间相关性及变化趋势可做成散点图等。

习题 2.2 算术平均数与加权平均数形式上有何不同？为什么说它们的实质是一致的？

答 算术平均数计算公式为 $\bar{x} = \dfrac{\sum x}{n}$。

加权平均数计算公式为 $\bar{x} = \dfrac{\sum fx}{\sum f}$。

加权平均数的分母为 $\sum f$，即为 n，分子为 $\sum fx$，即为 $\sum x$，所以算术平均数与加权平均数的实质是一样的。加权平均数的计算更适用于分组资料求解平均数。

习题 2.3 平均数与标准差在统计分析中有什么用处？它们各有哪些特性？

答 平均数的用处：①平均数指出了一组数据资料内变量的中心位置，标志着资料所代表性状的数量水平和质量水平；②作为样本或资料的代表数据与其他资料进行比较。

平均数的特性：①离均差之和等于零；②离均差平方和为最小。

标准差的用处：①标准差的大小，受试验或调查资料中多个观测值的影响，如果观测值与观测值间差异较大，其离均差也大，因而标准差也大，反之则小；②在计算标准差时，如果对各观测值加上或减去一个常数 a，标准差不变；如果给各观测值乘以或除以一个常数 a，则所得的标准差扩大或缩小了 a 倍；③在正态分布中，一个样本变量的分布可作如下估计：$\bar{x} \pm s$ 内的观测值个数约占观测值总个数的 68.26%，$\bar{x} \pm 2s$ 内的观测值个数约占观测值总个数的 95.49%，$\bar{x} \pm 3s$ 内的观测值个数约占观测值总个数