

SHENGWU TONGJIFA SHENGWU TONGJIFA

生物统计法

童一中编著 湖南科学技术出版社

生物统计法

童一中 编著

责任编辑：彭少富 熊穆葛

*

湖南科学技术出版社出版

(长沙市展览馆路8号)

湖南省新华书店发行 湖南省新华印刷二厂印刷

*

1987年10月第1版第1次印刷

开本：787×1092毫米 1/32 印张：16.25 字数：373,000

印数：1—3,600

ISBN 7—5357—0077—2/Q·5

统一书号：13204·156 定价：3.90元

湘目87—31

前 言

具有广泛研究领域的生物科学，既有定性描述的学科，也有定量描述的学科；即便原来是定性描述的学科，目前，也正在逐步趋向量化的研究。因此，作为一门研究生物界数量现象的科学方法——生物统计法，在广大生物学工作者的心目中，愈来愈受到重视。近年来，为适应教学、科研的需要，各地院校生物学系，纷纷开设生物统计课，或生物统计专业。生物学系学生对生物统计课程的学习，主要在于应用；然而，也有不少学生，对一些统计公式的来源，颇感兴趣。鉴此，作者在本书中作了两方面的努力：一是对生物统计公式的应用，重点介绍其如何运算方面，并结合实例进行叙述，介绍时力求具体、详细，以便读者举一反三，触类旁通；二是对书中所出现的公式，尽可能地作了数学推导，以利读者对公式的理解和记忆。另外，每章都附有相当数量的习题，以便读者通过解题掌握书中要点。

由于编者水平有限，书中难免有错误和不当之处，深望专家、学者及广大读者批评指正。

编者

1986年8月于上海师范大学

目 录

第一章 绪论	(1)
一、生物统计法的性质和作用.....	(1)
二、学习生物统计法的预备知识.....	(5)
习题一.....	(15)
第二章 次数分布	(16)
一、次数分布的意义.....	(16)
二、次数分布表.....	(16)
三、次数分布图.....	(24)
四、累积次数表和累积次数分布图.....	(28)
习题二.....	(32)
第三章 平均数和变异数	(33)
一、集中趋势和离散趋势的度量.....	(33)
二、平均数.....	(34)
三、变异数.....	(49)
习题三.....	(61)
第四章 概率、排列和组合	(63)
一、概率的概念.....	(63)
二、期望值.....	(65)
三、概率的运算.....	(66)
四、大数定律.....	(77)
五、排列与组合.....	(79)

习题四	(83)
第五章 二项分布和普阿松分布	(85)
一、二项分布	(85)
二、普阿松分布	(112)
习题五	(119)
第六章 正态分布	(120)
一、正态分布曲线	(120)
二、正态分布的概率计算	(127)
三、正态分布的应用	(137)
习题六	(146)
第七章 抽样和抽样分布	(148)
一、总体参数的无偏估计	(148)
二、样本平均数的分布	(152)
三、两个样本平均数差的分布	(161)
四、二项总体抽样	(166)
五、抽样误差和样本大小	(170)
习题七	(175)
第八章 显著性测验	(176)
一、显著性测验的基本原理和方法	(176)
二、平均数的显著性测验	(192)
三、百分数(成数)的显著性测验	(203)
四、普阿松分布的显著性测验	(211)
习题八	(214)
第九章 χ^2 测验	(216)
一、 χ^2 测验的基本原理和方法	(216)
二、适合性测验	(220)
三、独立性测验	(231)

四、齐性测验·····	(243)
五、 χ^2 测验的计算方法问题·····	(248)
习题九·····	(252)
第十章 非参数统计 ·····	(255)
一、非参数法的意义·····	(255)
二、符号测验法·····	(256)
三、添号秩次测验法·····	(261)
四、秩和测验法·····	(265)
习题十·····	(269)
第十一章 方差分析 ·····	(271)
一、方差分析的基本原理和方法·····	(271)
二、单向分组资料的方差分析·····	(294)
三、两向分组资料的方差分析·····	(307)
四、方差分析的基本假定和数据转换·····	(327)
习题十一·····	(335)
第十二章 直线回归和相关 ·····	(339)
一、直线回归和相关的意义·····	(339)
二、直线回归方程·····	(342)
三、直线回归的显著性测验·····	(363)
四、相关系数的性质及其与回归系数、决定系数的关 系·····	(377)
五、相关系数的计算·····	(384)
六、相关系数的显著性测验·····	(388)
习题十二·····	(396)
第十三章 多元回归和相关 ·····	(398)
一、多元回归和偏回归·····	(398)
二、多元回归的显著性测验·····	(413)

三、多元相关和偏相关·····	(417)
习题十三·····	(425)
第十四章 协方差分析 ·····	(427)
一、协方差分析的意义和应用·····	(427)
二、单向分组资料的协方差分析法·····	(428)
三、两向分组资料的协方差分析法·····	(442)
习题十四·····	(449)
第十五章 正交试验 ·····	(450)
一、多因素的试验问题·····	(450)
二、正交试验的基本方法·····	(451)
三、有交互作用的正交试验·····	(462)
四、水平数不等的正交试验·····	(472)
习题十五·····	(477)
附表 ·····	(479)
参考文献 ·····	(510)

第一章 绪 论

一、生物统计法的性质和作用

(一) 什么是生物统计法

生物统计法，也可称为生物统计学，它是运用统计学原理和方法，来研究生物界数量现象的科学方法。或者说，它是一门应用统计方法或数学逻辑来搜集、整理、分析和解释生物界数量现象的科学方法。因而，是一门统计学与生物科学相结合的应用边缘科学。

生物界各种数量现象具有普遍的变异性。引起变异的因素是多方面的，有主要因素，也有次要因素，而这些因素对总体中各个个体的影响是随机的，也是不确定的。因此，统计研究的各种现象的表现结果是一种随机事件，这种随机事件，只要在一定数量重复条件下，它就会呈现出规律性。然而，对于客观事物数量方面的规律性，我们不可能也没有必要对事物的全部（总体）一一进行研究后再来认识，一般只可能通过某事物的一部分（样本）所传递的信息来进行估计。由样本来推断总体，这在统计学上便叫统计推断，它的性质与一般归纳法极相类似。我们并不排斥演绎法在统计学中的地位，然而，我们还是可以说，统计学的主要精神是归纳性的。因此，从这个角度来看，把生物统计法看作是一门以样本为依据，运用数学

模型来推断总体的科学方法，也是可以的。

由于各种现象的表现结果是一种随机事件，因此，从样本来推断总体，其结论的可靠程度就应该是概率性的。

因而，生物统计法就不同于一般只是进行描述，而不进行计算的某些生物学学科。也就是说，它要的是总体中个体间的度量值，而不是个体间的分类；它要的是数据，而不是修饰词。另外，它又不同于某些确定性的数量学科，因为它的结果只能是以某种概率出现。例如， $s = \pi r^2$ ，在这里，圆面积的大小，完全可由半径的大小来确定，这是一种严密的函数关系，而这种严密的函数关系，在生物试验中，是极为罕见的。比如利用相同性质的种子，在相同条件下进行发芽试验，我们可能得到发芽率（概率），而不能作出一定能有几粒种子发芽的决定。

总之，统计学是概率性的，而它的主要精神又是归纳性的，这样，概率性和归纳性便构成为统计学的两个非常重要的属性。我们开始学习生物统计法时，就应充分理解，并给以肯定。

（二）为什么要学习生物统计法

1. 生物统计法能为我们提供数据整理分析的方法

例如，假设要研究地理隔离对生物的影响，其中有一部分工作，是要了解某海岛上田鼠的一些有关性状。如果将调查所得的每一性状的每一观察值都列在一起，往往会显得庞杂零乱，且又不能说明任何问题。而生物统计可为我们提供数据整理分析的方法，可将所得的大量数据加以整理、简缩，制成统计表，绘成统计图，或就这些数据的分布特征，如集中趋势、离中趋势、相关程度等等，计算、凝炼成为几个具有概括性的统计数字（如平均数、标准差、相关系数等）作为统计指标。而后通过这些概括性的数字，我们就有可能从杂乱无章的数据中了解到蕴藏的信息。

2. 生物统计法能为我们提供由样本推论总体的方法

仍以在某海岛上调查田鼠的一些有关性状为例。我们永远也不能奢望去捕捉岛上的全部田鼠，我们捕捉到的也只能是一部分。这一部分究竟该要多少？我们调查的仅是一部分田鼠，而我们所要了解的是整个岛上全部田鼠的有关性状。如何通过一个或几个样本的调查，就能从所获得的有限数据中，对所研究的问题作出全面的论述，并得出具有科学意义的结论呢？有关这类问题，生物统计法都可为我们提供切实可行的办法。

3. 生物统计法能为我们分析变异因素提供一系列决策和方法

导致某一性状的变异因素是多方面的，因此，在阐明客观存在的规律性时，就有必要对变异的原因进行追踪。比如，在遗传学中，表现型（指性状）是基因型（遗传）和环境相互作用的结果，这就是说，表现型是受遗传和环境两类因素制约的。所以，经常遇到这样的情况：基因型改变，表现型随着改变；环境改变，表现型也随着改变。因此，如要研究某一数量性状的遗传，如不将遗传变异和环境变异从总变异中分裂开来，则研究就成为一句空话。

在通常的比较试验中，也同样存在着这类情况。比如，对甲、乙两品种进行一次产量比较试验，试验结果见表1—1。

表1—1

甲、乙两品种产量比较试验

（单位：公斤/亩）

生产单位 品 种	I	II	III	IV	V	平均
甲	410	382.5	416	392.5	430	406.2
乙	330	385.5	403	382.5	365	373.2

从表面数据来看：品种甲的平均亩产比品种乙要高33公斤。能否就因此得出结论说品种甲的产量潜力就一定胜过品种乙呢？回答是否定的。因为一个品种的产量高低，除与品种的种性（基因型）有关外，还包括环境因素。也就是说，在判别一个品种的产量潜力时，也有必要将变异因素进行分裂后才能识别。

生物统计法中的方差分析，能帮助我们将各种变异逐步分裂出来，这样，我们就可追查变异的原因，可以从表现为偶然的数据中，找出其必然性规律。

4. 生物统计法可帮助我们分析现象之间的关系

统计法不是孤立地研究各种现象，而是通过一定数量的观察，从这些现象里研究事物间的相互关系，阐明事物客观存在的规律。我们可通过变数间的变化趋势，从一个（或一个以上）变数的数量变化，来预测另一个变数的数量变化。比如，一些夏季病虫盛发期的迟早，与春季温度的高低有关。如果我们记录了历年春季旬平均温度的累积值和病虫盛发期，就可求得一个经验公式。再将当年的春季旬平均温度的累积值代入经验公式，就可预测某一种病虫在当年的盛发期。这样，对防治病虫害是非常有利的。

5. 生物统计法对试验设计也有着指导作用

在试验中，我们只能注意到控制进行的试验因素（即处理因素），而对那些非试验因素（偶然因素）往往难以控制，这样，在整个试验中，处理因素由于受到偶然因素的干扰，致使真象被蒙蔽，实验也就缺乏敏感性。因此，如果希望试验能获得正确的结果，对这类偶然因素，就必须严加控制。而对偶然因素的控制，需要借助统计学中的随机原理和重复原理，来增加试验的敏感性。

我们之所以要学习生物统计法，是因为它对科学实验有着积极的指导意义，此外，通过生物统计法的学习，还可锻炼我们的科学思维，学会科学的推理与思考方法，养成实事求是地对待一切事物的唯物主义态度。但也必须注意到，统计程序的本身，并不能保证试验能避免产生错误、误差，或不正确的结论。再说，生物本身的生长、发育是受生物规律而不是统计规律所制约的。因此，对生物统计法的运用，必须建立在对生物本身充分认识的基础上，而后应用正确的资料和相应的方法。也只有在这个时候，生物统计法才能充分发挥作用。

二、学习生物统计法的预备知识

(一) 统计常用术语

1. 总体和样本

医生根据病人一小滴血液的检验，来诊断病人的全部血液，这便是个抽样问题。而要了解抽样问题，首先就得要弄清什么是总体，什么是样本，以及总体与样本间的关系。

总体，简单地说，就是我们所研究的全部个体。它可以是无限的，也可以是有限的。如果把小麦品种“扬麦一号”作为研究对象，这个品种的全部个体，便是个无限总体。如果我们以一个小区的“扬麦一号”作为研究对象，就这个小区的全部“扬麦一号”来说，却是一个有限的总体。生物统计学所研究的对象，总是以无限总体为基础的。

样本，就是从总体中随机抽取的若干个体，这若干个体便构成一个样本。样本取多少个个体，在统计学上称为样本容量。从“扬麦一号”中，随机抽取100个植株，这100个植株便是一个样本；如从中随机抽取100张剑叶或30个花药，这100张

剑叶是个样本，这30个花药也是一个样本。根据样本容量的多少，可分为大样本和小样本。在生物统计上，一般习惯把样本容量超过30的，划为大样本；等于或小于30的，划为小样本。把样本划分成大、小样本是有必要的，因为在一些计算方法和分析检验上，大样本和小样本是有区别的。

很清楚，一个样本是总体的一部分，但在有些情况下，一个样本也可能包括总体的全部。统计学上所要研究的总体，是有其一定的特点；总体中各个体的量或特征的表现是有变异的，并且是不确定的，是随机的。要从样本作出关于总体的推断，这个样本必须是总体的代表。也就是说，也并不是每个样本都能作为总体的代表。比如，假定某一实验取样的目的，是要测定在一群小鸡的总体内，喂以一种新饲料所产生的生长速率。在实验中，从这群小鸡中抽取10只（作为一个样本）放在笼内。可是，这个笼子被放在一个冷而吹风的地方，或是放在一个阴暗的角落里；或笼内的小鸡间传播着一种尚未被觉察出的传染病。如果这样，这些小鸡的生长速率，就不能作为正常小鸡生长速率的总体代表。或者，你尽量挑选一些身强体壮的小鸡来做上述实验，即使把笼子放在正常的环境里，这些小鸡的生长速率，也同样不能代表总体。总之，要取得一个有代表性的样本，除了要注意实验条件的正常性外，很重要的一点，那就是抽样必须是随机性的。所以，通常指的样本就是指随机样本。因为只有根据随机样本，才有可能计算出其推断的误差概率，从而得知用它来估计总体的准确程度。

2. 变数和变量

身长就是一个变数。概括地讲，凡具有变异的某种性状，而且又可以进行度量或称量的，这在统计学上，就称它为变数。比如，测量了30个人的身高，他们之间是有差异的，这30个人

的身高数据就成为一个变异数列，这个变异数列，可简称为变数，通常以 x 表示；而组成变异数列的每一具体数值，就叫变量（variate），或叫观察值，以 $x_1, x_2, \dots, x_n$ 来表示。

医学中的血压、血球计数、心电图波段的时间间隔等等，都是变数。每个变数都有一个总体，就血压来说，45岁以上男子的血压是个变数，则全部45岁以上男子的血压数据就构成一个总体，而作为样本的某个45岁以上男子的血压数据，就是变量。

变数又可分为离散型变数和连续型变数。

分蘖这个性状也存在着变异，也可以度量，无疑，分蘖也是个变数。调查分蘖数是采用计数方式，每个观察值都是以整数来表示的，在两个相邻的整数间，不容许有带小数的数值存在。由于在两个相邻整数间是不连续的，所以，象分蘖这类变数就称为离散型（间断性或非连续性）变数，对所记载的观察值，就叫离散型变量。

种子的粒重也是个变数，它是以量来表示的。在称量时，各个观察值并不限于整数，在两个相邻的数值间，可以有微量差异的其他数值存在。例如测定小麦的颗粒重，在35克与36克两个观测值之间，可以有35.5克，也可以有35.54克等等数值存在，其小数位数的多少，可随度量的精确度而定。象这类数据，在理论上可取无限个连续变异的数值，所以，就称为连续型变量；而就粒重这个性状来说，它便是连续型变数。

3. 参数和统计量

参数，是指由总体的全部观察值演算而得的总体特征数。如总体平均数、标准差等。总体特征数虽是真值，但不易直接求得，只能以样本的统计量来估计。所谓样本的统计量，是指从样本中求得的样本平均数、标准差等特征数。由于总体内

存在着变异，而样本仅是总体中的一小部分个体，统计量与参数并不完全一致，但应该是近似的，近似的程度取决于抽取样本的科学性。尽管统计量只能在一定程度上反映总体，但把样本的统计量作为总体相应参数的估计值还是可行的。

4. 误差、偏差和错误

由于试验过程中，经常会受到各种无法控制的随机因素——自然因素、人为因素、偶然因素等的影响，尽管我们在抽样时很随机，在度量时很准确，在计算时很细心，而结果仍有一定的误差和偏差，这类误差称之为试验误差。试验误差是随机性的，有时大，有时小，有时正，有时负；误差的大小以及正负误差的出现，完全由概率决定。随着度量次数的增加，试验误差的算术平均值将逐渐趋近于零，而多次度量结果的算术平均值将更接近于真值。由于试验误差的发生是随机的，受概率所支配，因而，也可把试验误差，简称为机误。

误差与偏差是有区别的。误差是指观察值与真值之差；偏差是指观察值与平均观察值之差。实际上，真值往往为人不知，只能由平均观察值来估计。因此，人们在习惯上，常将两者混用。

错误，不同于误差，错误可以避免，误差是免不了的，它是属于一种过失性误差，是由于粗心大意所引起。比如将试验材料弄错，或将不同材料相混合，或由于记录、度量、计算不正确。对这类错误，生物统计学是无法进行研究的，也不能通过增加重复次数来消除，在这里，多次重复也不过重复错误而已。

5. 准确度和精确度

在日常生活中，准确度和精确度是同义语，但在统计学中，它们是两个不同的概念。准确度是指试验中某一性状的观察值与真值符合的程度；两者越接近，则试验数据越准确；但在—

般试验中，真值为未知数，所以对准确度不易确定。精确度是指试验中同一性状的重复观察值彼此间的接近程度，一般是指抽样误差。抽样误差越小，则处理间的比较越精确。

(二) 有效数字和计算法则

1. 有效数字

实验结果的准确度应该反映在记录这些结果的数字上面。有效数字就是实验结果最简单的概括，它不仅表示结果的大小，而且也反映结果的准确程度。

有人认为，直接观测值的小数点后面的位数取得越多，测量就越准确；在计算时，计算结果中保留的位数越多，计算就越精确。殊不知这种想法和做法是错误的。

第一，观测值的准确度不能超出仪器的准确度所允许的范围。在直接度量中，有效数字正是反映了所用仪器的准确程度，度量值的准确度决不能超越仪器的准确度所允许的范围，读数只能读到一定的位数。多读了没有意义，少读了不能如实地反映度量结果的准确度。正确的原则是：直接测得值的位数，除最末一位数是欠准确或可疑的以外，其余各位数字都应是准确可靠的。例如用米尺(准确度为1毫米)测量长度时，读数在14.4厘米和14.5厘米之间，目测估计为14.46厘米。这里，“14.4”是准确的，而最末一位“6”是估计的，欠准确的。尽管“6”是可疑的，但对于测量结果来说，读出它比不读出要准确得多，因此，我们称此时所读的数字为有效数字。有效数字的位数称为有效数位。14.46厘米有四个有效数位，就是有四位有效数字。如仅读出14.4厘米或14.5厘米，则只有三位有效数字，这样，就不能正确反映所用仪器的准确度，从而也降低了观测值的准确度。如读成14.467厘米，则成了五位有效数字，而这一准确度是米尺所不可能达到的，因而写出“7”这一位就毫无意义。

从这里可看出，有效数字与测量误差是密切相关的。一般认为测量误差是仪器最小刻度的二分之一，如米尺的测量误差为0.05厘米。发生误差的数位称作可疑数位，这位数字就叫可疑数字。因此，对于一个测得值，从左边第一个非零数字算起，到第一个发生误差的数字为止，都算作有效数字。

在有效数字中，由于只允许保留一位可疑数字，因此，关于小数点后面的0，它可以是有效数字，也可以是非有效数字。比如对准确度为1毫米的米尺而言：8.70厘米或8.00厘米，这里小数点后面的0就都是有效数字；而8.710厘米，这个小数点后面的0是没有意义的，是非有效数字。

第二，小数点的位置不是决定准确度的标准，小数点的位置只是与所用单位的大小有关。比如，测量长度若采用厘米或米作单位时，3.00厘米与0.0300米是相同的测量值，都是三位有效数字，它们的准确度也完全相同。

对那些为了标明小数点而加的0，应不作为有效数字计。例如，0.01206和0.001206都只含有四位有效数字。可是，在诸如12060和1 20600之类的数字中，就看不清后面的0是有效数字还是非有效数字。为了避免混淆，此时最好用指数表示，可把它们写成为 1.206×10^4 和 1.206×10^5 ，以表明它们都只有四位有效数字。

最后，我们可将有效数字的判别原则归纳为下列四条：

(1) 经简化后，非零的位数均为有效数字。

(2) 零居于两个有效数字之间，或零居于有效数字之后，但含小数点时，恒为有效数字。

(3) 以零开头的数值，其居于第一个非零数字之前的零不为有效数字。

(4) 零居于最后一个有效数字之后，则有效数字的位数不