

内蒙古大学教材丛书

生物统计学

杨 持 编著

内蒙古大学出版社

生物统计学

——

——

——

内蒙古大学教材丛书

生物统计学

杨 持 编著

内蒙古大学出版社

内蒙古大学出版社基金委员会

主 任：旭日干

副主任：梁希侠（常务） 呼格吉勒图

委 员：旭日干 梁希侠 呼格吉勒图

曹之江 包 祥 王 璋

刘 成 刘树堂 陈羽云

生物统计学

杨 持 编著

内蒙古大学出版社出版发行

（呼和浩特市大学西路1号）

内蒙古自治区新华书店经销

内蒙古大学印刷厂印刷

开本：850×1168/32 印张：9 字数：216千

1996年5月第1版 1996年5月第1次印刷

印数：1—1500册

ISBN7-81015-617-9/Q·19

定价：9.00元

内 容 提 要

本书是作者在多年教学实践中写成的。内容包括两章,第一章讲统计分析的基础知识;第二章讲数据整理、统计推断、统计假设检验(t 测验, χ^2 测验, 方差分析), 回归分析等。本书可供综合性大学和师范院校生物系以及农林医院校有关专业师生使用, 也可供有关科研工作者参考

序

生物科学是一门实验性学科,需要用数理统计的原理和方法去分析和解释生物界的各种现象和数据资料,从而从大量、普遍存在的偶然现象中揭示出真实的规律性。又由于现代科学技术的发展,特别是电子计算机在生物学研究中的应用,大大地促进了生物科学研究的定量化。近代生物科学发展的两个显著特点就是定量化和动态化,这都要求有数学的渗透,特别是生物统计学的应用。

本书是作者在 10 多年教学讲义的基础上修改而成。内容侧重于对统计学基本原理的了解和对各种统计方法的应用,不强调各种公式的严格推导。全书分两章,共 16 节。第一章有 6 节,主要讲统计分析的基础知识,即概率论与数理统计的基本概念及运算;第二章有 10 节,主要讲数据整理,统计推断,统计假设检验(t 测验, χ^2 测验,方差分析),回归分析,协方差分析等。本书还给出了学生课外作业题 106 题,平均每次课留课外作业 3—4 个题。

本教材需 72 学时讲完,其中讲授 52 学时,习题课及计算机统计软件应用练习 20 个学时。随着电子计算机科学的发展,目前已有许多不同来源的统计软件可供使用,即使没有学习过“生物统计学”的人也可以利用计算机作统计分析,那为什么还要强调学生学习“生物统计学”呢?这是因为,只有真正理解统计学的原理,才能应用好那些软件,才能对各种类型的问题给出科学的解释。另一方面,也要紧跟科学的发展,利用现代化手段,所以在学时安排上,压缩课堂讲授,加强了计算机软件应用的练习。

书中的例题及习题,有相当多的部分选自其他作者编写的书刊,在此深表谢意。

书中的插图请张海燕同志绘制,也在此深表谢意。

由于作者的专业及学识所限,错误在所难免,希望使用本教材的教师、学生和生物学工作者提出宝贵意见。

编 者

1996. 3

目 录

第一章 统计分析的基础知识.....	(1)
第一节 几个基本概念.....	(2)
1 必然与偶然	
2 随机现象与随机试验	
3 随机事件、基本事件、复合事件	
4 样本点与样本空间	
5 随机变量	
6 随机样本与统计总体	
7 事件之间的关系与事件的运算	
第二节 概率的计算.....	(7)
一 概率的基本概念	
1 概率的统计定义	
2 概率的古典定义	
二 概率的运算	
1 概率加法法则	
2 概率乘法法则(包括条件概率与独立事件的概率)	
3 全概率公式	
4 逆概率公式	
第三节 随机变量与分布函数	(17)
一 随机变量	
二 分布函数	
第四节 离散分布	(20)
一 离散型随机变量及其概率函数	
二 离散型随机变量的数学期望与方差	
三 几个常见的离散分布	
1 伯努里(Bernoulli)分布	
2 二项分布	

3	普阿松分布	
4	超几何分布	
5	几种分布之间的关系	
第五节	连续分布	(35)
一	连续型随机变量及其密度函数	
二	连续型随机变量的数学期望及方差	
三	几个重要的连续分布	
1	均匀分布	
2	正态分布	
第六节	中心极限定理	(47)
1	同分布的中心极限定理	
2	二项分布的正态近似	
	习题	
第二章	数据资料的统计分析	(56)
第一节	数据资料的获得与整理	(56)
一	抽样技术	
1	总体类型与抽样类型	
2	抽样方法	
3	取样数目	
二	试验设计	
1	试验处理设计	
2	试验方法设计	
三	试验数据的整理	
第二节	样本平均值与标准差	(73)
1	样本平均值	
2	样本标准差	
3	变异系数	
4	偏斜度	

5 峭度

第三节 总体参量估计 (78)

一 点估计

二 μ 的区间估计

1 σ 已知

2 σ 未知

3 $\bar{x}$ 的控制图

第四节 单个样本的统计假设检验 (84)

一 一般原理

1 假设

2 小概率原理

3 单侧检验与双侧检验

4 两种类型的错误

二 单个样本显著性检验

1 σ 已知情况下, 单个平均数的显著性检验—— u 检验

2 σ 未知时的单个平均数的显著性检验—— t 检验

3 变异性的显著性检验—— χ^2 检验

第五节 两个样本的差异显著性检验 (89)

一 两个方差的检验—— F 检验

二 两个平均数差异的检验—— t 检验

1 配对数据均数间的比较

2 标准差(σ_1)已知时, 两个均数间的差异显著性检验

3 标准差(σ_1)未知, 但方差具齐性时, 两均数间的差异显著性检验

4 标准差(σ_1)未知, 且方差不具齐性时, 两均数间的差异显著性检验

三 两个百分率的比较

第六节 非参量检验法 (101)

- 一 划级法
 - 1 D 值法(适用于配对数据)
 - 2 S 值法(适用于成组数据比较)
- 二 符号检验(适用于配对数据)
- 三 秩和检验(适用于成组数据比较)
- 四 秩相关检验
- 五 级和分布法
- 六 两组以上数据的比较
- 七 游程检验

习题

第七节 计数资料的显著性测验— χ^2 测验 (120)

- 一 四格表 χ^2 测验
- 二 多次重复的四格表测验
- 三 多格表 χ^2 测验
- 四 2×2 列联表的精确检验法
- 五 柯尔莫哥洛夫—斯米尔诺夫检验

习题

第八节 变量分析 (139)

- 一 单因素多组群的方差分析
 - 1 重复数相等的几个均数的比较
 - 2 重复数不等的几个均数的比较
- 二 两因素多组群的方差分析
 - 1 随机区组试验资料的方差分析(无重复)
 - 2 两因素分组资料的方差分析(有重复)
- 三 拉丁方资料的方差分析
- 四 正交拉丁方试验资料的方差分析

习题

第九节 回归分析 (168)

一	一元线性回归	
1	回归直线的求法	
2	相关系数及其显著性检验	
3	回归直线方程效果的检验	
4	两条回归直线的比较	
二	一元非线性回归	
三	多元线性回归分析	
四	逐步回归分析	
第十节	协方差分析	(209)
	习题	
附表 1.	标准正态分布表	(226)
附表 2.	正态分布上侧分位数(u_α)表	(230)
附表 3.	t 分布的分位数表	(231)
附表 4.	样本标准差分布的矩系数 C_4 、 C_5	(232)
附表 5.	χ^2 分布的上侧分位数(χ_α^2)表	(233)
附表 6.	F 检验的临界值(F_α)表	(237)
附表 7.	多重比较中的 Q 值表	(241)
附表 8.	相关系数检验表	(245)
附表 9.	z、r 值转换表	(246)
附表 10.	D 值表	(247)
附表 11.	B 值表	(247)
附表 12.	S 值表	(248)
附表 13.	柯尔莫哥洛夫检验临界值表	(250)
附表 14.	斯米尔诺夫检验临界值表	(252)
附表 15.	游程总数检验表	(254)
附表 16.	游程长度检验表	(256)
附表 17.	符号检验表	(257)
附表 18.	秩和检验表	(258)

附表 19. 秩相关系数检验表	(259)
附表 20. 正交拉丁方表	(259)
附表 21. 平衡不完全组设计	(262)
附表 22. 正交表	(267)
参考书目	(276)

第一章 统计分析的基础知识

在一切事物和现象的发展变化中,有两类变化同我们所要讨论的问题有着十分密切的关系,一类是指事物和现象在一定条件下所发生的必然性变化和偶然性变化;另一类是指事物和现象在其发展历程中所发生的质变和量变。

首先看事物和现象的必然性变化和偶然性变化:一粒有生命力的种子,只要给予适量的水分、空气和热量等条件,必然生根发芽;两头具有正常生育能力的纯种种畜,其繁殖的后代也必然为纯种仔畜。然而伴随着必然性变化一并而来的是偶然性变化。关于这一点,恩格斯曾经指出过:“偶然性是必然性的表现形式,是必然性的补充”。如采自同一母株的一批种子,由于各自的内在遗传性不可能完全相同,加上外界发芽条件又难绝对保持一致,因此各粒种子萌发的迟早、子叶和胚根的长度与干重等,都事先难于精确预料;出自同胎的一批仔畜,在性别、初生重以及其它经济性状方面,也不完全一样,很难事先断言。诸如此类的偶然性变化,在各种科学实验中大量地和普遍地存在,它掩盖住,有时甚至严重地蒙蔽着人们对事物真象的认识。科学研究的一个重要任务就是要设法了解这些偶然性变化的规律,对于这类随机现象,我们无法利用“因果关系”加以严格控制,或准确预测;也不能用一些简单的物理定律加以概括,而须从大量观测中去综合分析,归纳出一些“大量现象”的规律来,从而透过偶然现象去把握住事物本质及其必然性变化。正如恩格斯所指出:“被断定为必然的东西,是由种种纯碎的偶然所构成的,而被认为是偶然的東西,则是一种有必然性隐藏在里面的形式”,“偶然性本来也具有必然性”。这些论断,乃是数理统计最根本的认识论的依据。

再来看事物和现象的另一类变化,即质变和量变。事物的质是指一事物区别于其它事物的内部规定性,但是,要研究事物的质,是断然离不开同一事物的量的规定性,种子的千粒重、大小、成分含量、发芽率等等,都是标志具体事物存在规模和发展程度的数量属性,反映着事物的本质。在农业及生物科学研究中,经常要同大量的实验数据或者观察值打交道,通过量的研究,区别差异,揭示本质,阐明变化,探索规律,以便为促进农业生产服务。因此,注意事物的量变,深入研究量与量之间的关系,是从事任何一项农业及生物科学实验所不应忽视的一个重要方面。

事物的必然性变化与偶然性变化,质变与量变,虽然统一于事物运动发展过程之中,彼此不可分割,但是,科学研究的方法论并不排斥把一个复杂事物的各种变化分解开来,侧重研究某方面的变化,便于加深认识。人们可以在必然性与偶然性统一的前提下,侧重于研究事物的偶然性变化;当然也可以在质与量统一的前提下,侧重于研究事物的数量方面。

概率论和数理统计就是研究自然界偶然现象或者叫做随机现象的数量特征和规律性的科学,生物统计则是数理统计的原理和方法在农业及生物科学实验中的应用。因此,要讨论生物统计的概念和方法,就离不开概率论和数理统计最一般的知识;此外,在描述和讨论一个生物统计问题时,也离不开数学上所特有的方法,即数学符号的逻辑推理方法。在第一章里主要是介绍概率论和数理统计的最一般概念。

第一节 几个基本概念

1. 必然与偶然

必然现象是一种确定性现象,例如,将小白鼠放在充满一氧化

碳的罐子中,它必然死亡。要了解这种确定性现象的规律,只需一次性试验就可得到。微积分学和线性代数是研究这类现象的数学工具。

偶然现象是一种非确定性现象或称为随机现象,例如,考察动物的出生重量,可以发现个体之间存在很大差异,初看起来,这些随机现象似乎是无规律的,但当你多次重复观察之后,就会发现动物出生重的差异是有规律的。这种在大量重复观察和试验中所呈现出的固有规律性,就是经常提到的统计规律性。概率论与数理统计是研究和揭示随机现象的统计规律性的一门数学学科。

2. 随机现象与随机试验

在个别观察试验中呈现出不确定性,在大量重复试验中又具有统计规律性的现象或变化,称为随机现象或变化。

随机现象可以通过随机试验来研究,所谓随机试验就是在同一组综合条件下,在每一次观测前,并不能准确得知出现的结果。这里明确地给出随机试验的三个特性:

- ①可以在同一组综合条件下重复地进行;
- ②每次试验的可能结果不止一个,并且能事先明确试验的所有可能结果;
- ③进行一次试验之前不能确定哪一个结果会出现。

3. 随机事件、基本事件,复合事件

把随机试验的结果叫做事件,因为事件的出现是随机的,因此把一次试验可能出现也可能不出现,而在大量重复试验中却具有某种规律性的事情,称为此随机试验的随机事件。

在一随机试验中,它的每一个可能出现的结果都是一个随机事件,如掷一个骰子,1、2、3、4、5、6点都是可能出现的结果,每个结果都是一个随机事件,而且这些结果都是这个随机试验的最简单的,不可再分的随机事件,称为基本事件。

由几个基本事件组成的事件称为复合事件,如结果是偶数的

随机事件为 2、4、6 三个点，它包括了三个基本事件。

4. 样本点与样本空间

每一个基本事件都可以用一个点来描述，它是对随机试验结果最细的描述。

为了便于研究随机试验，将随机试验的所有基本事件组成的集合叫做样本空间。例如，抛三枚硬币的样本空间为 $2^3=8$ ；从一副扑克牌中抽两张黑桃的样本空间为 $C_{52}^2=1236$ 。

5. 随机变量

给随机事件加以数量化以后，也就是用一个数量来表示随机试验的结果，这个数量就叫做随机变量。随机变量随着试验的结果而取不同的值，因而，在试验之前只知它可能取值的范围，而不能预知它取什么值。根据随机变量可能取得的值，将随机变量分为离散型随机变量和连续型随机变量。如果随机变量可能取得的数值为有限个，或可数无穷个孤立的数值，则称为离散型随机变量，例如每 10 只新生动物中，雄性动物的只数。如果随机变量可取某一区间内的任何数值，则称为连续型随机变量，例如我国男青年身高即为一连续型随机变量。

6. 随机样本与统计总体

科学研究工作中，研究的问题是各种各样的，要观测记录的项目也各不相同，有株高、穗长、千粒重、产量、初生重等等，这些不同的项目是区别不同研究问题的标志。每观测一次，就相当于作了一个随机试验，于是就得到一个一定标志下的标志值，同一随机试验所有的标志值的集合，就叫做统计总体。

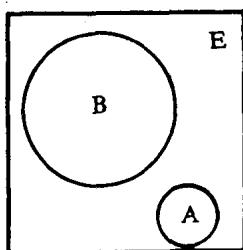
由于一个研究对象所包含个体的数目往往很大，甚至无穷，因此，实际上进行观察的只是总体中的一小部分，从总体中抽出的这个有限部分就构成了样本。因为从统计学的要求，总体中每个元素被抽出作为样本的机会都必需是均等的，那就要保证随机抽样，由随机抽样得到的样本就叫做随机样本。以 r 来表示样本所包含的

元素数目,这叫做样本的容量,根据样本容量的大小,又分大样本与小样本,一般以 30 为界,容量大于 30 的为大样本,小于 30 的为小样本。

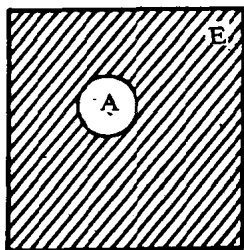
7. 事件之间的关系与事件的运算

为了便于说明事件之间的关系,可以先介绍集合论中的一些基本概念,看下面的对照表。

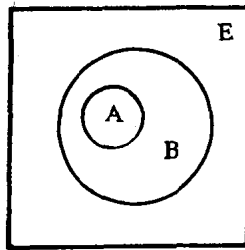
符号	概率论	集合论
S	样本空间 必然事件	全集(包含了全部元素的集合)
Φ	不可能事件	空集(不包含任何元素的集合)
e	基本事件 样本点	元素
A	事件	子集
$\bar{A}$	A 的对立事件	A 的余集
$A \subset B$	事件 A 发生导致事件 B 发生	A 是 B 的子集(包含)
$A = B$	事件 A 与事件 B 相等	A 与 B 相等
$A \cup B$	事件 A 与事件 B 至少有一个发生	A 与 B 的并集
$A \cap B$	事件 A 与事件 B 同时发生	A 与 B 的交集
$A - B$	事件 A 发生而事件 B 不发生	A 与 B 的差集
$AB = \emptyset$	事件 A 与事件 B 互不相容	A 与 B 没有相同元素



A, B 不相容



$\bar{A}$



$A \subset B$