



世纪高等院校“十一五规划”教材

# 生物统计学 (第3版)

内蒙古自治区生命科学教材编委会 组编  
主编 杨 持 执行主编 王铁娟  
王铁娟 张慧荣 张颖娟 张彩琴 编著



内蒙古大学出版社  
INNER MONGOLIA UNIVERSITY PRESS

21

世纪高等院校“十一五规划”

# 生物统计学 (第3版)

内蒙古自治区生命科学教材编委会 组编  
主编 杨 持                      执行主编 王铁娟  
王铁娟 张慧荣 张颖娟 张彩琴 编著



内蒙古大学出版社  
INNER MONGOLIA UNIVERSITY PRESS

## 内 容 提 要

本书是作者在多年教学实践中写成的。内容包括四章,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( $t$ 检验, $\chi^2$ 检验,方差分析)、回归分析等;第三章讲多元分析方法;第四章介绍了SPSS软件的应用。本书可供综合性大学和师范院校生物类专业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。

### 图书在版编目(CIP)数据

生物统计学/王铁娟等编著. —呼和浩特:内蒙古大学出版社,2014.3  
ISBN 978-7-5665-0324-4

I. ①生… II. ①王… III. ①生物统计 IV. ①Q-332  
中国版本图书馆CIP数据核字(2014)第033710号

## 生物统计学(第3版)

主编 杨 持 执行主编 王铁娟

王铁娟 张慧荣 张颖娟 张彩琴 编著

内蒙古大学出版社出版、发行

(呼和浩特市昭乌达路88号 邮政编码010010)

责任编辑 张 志 封面设计 张燕红

内蒙古军区印刷厂印刷

开本 850毫米×1168毫米 1/32 印张 11.625 字数 292 000

1996年5月第1版

2005年8月第2版

2014年2月第3版 2014年2月第3次印刷

印数 30 001—32 000

ISBN 978-7-5665-0324-4

定价:18.60元

## 内蒙古自治区生命科学教材编委会

主任:李东升 杨 持(常务)

副主任:王黎元 布日额 齐宝瑛 李春林 杨 劼

张秋良 郑根昌 徐振军 韩国栋 蔡 禄

委员:于 卓 马德慧 王迎春 王纯洁 王思珍

王黎元 布日额 田 迅 白恒勤 齐宝瑛

刘玲玉 刘晓光 李东升 李青丰 李春林

杨 劼 杨 持 杨保收 张 宁 张昕原

张秋良 张淑艳 郑根昌 周欢敏 周延林

庞保平 郝永清 赵萌莉 哈斯巴根

栗淑媛 徐振军 郭 成 唐立红 韩国栋

蔡 禄 德永军

## 第三版前言

《生物统计学》自1996年第一版,2005年第二版,到2013年已经使用了将近20年,发行量超过3万多册,为生物学、生态学、农业院校和师范院校的教学提供了适用的教材。

从第二版到现在也已经8年多了,从适应学科发展的角度看,对第二版重新修订已是十分必要。

修订工作由王铁娟等四位博士来完成,一方面,她们从事生物统计学的教学工作多的10年左右,少的也有3—4年;另一方面,她们在教学工作中对生物统计学有了比较深入的理解,对第二版中存在的不足和缺失有了新的思考,由她们来完成第三版的修订编写,一定会使教材的质量有大幅度的提高。我作为第一版和第二版的主编,对她们在教材建设上所付出的辛苦是很赞赏的。

本次修订的重点是:修改错误,补充内容,特别是对第四章的修订改动较大,在SPSS软件的应用上做了更为详细的阐述,使学生操作起来更加容易和理解。

编写分工如下:第一章由内蒙古大学张慧荣修订;第二章1—6节由内蒙古师范大学张颖娟修订;第二章第7节和第四章由内蒙古师范大学王铁娟修订;第二章8—9节、第三章由内蒙古农业大学张彩琴修订。

细心和认真不等于十全十美,错误在所难免,希望使用本教材的教师和学生,以及其他生物学工作者们提出宝贵意见。

杨 持

2013年11月20日

# 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)
第二节 概率的计算 .....	(7)
一、概率的基本概念 .....	(7)
二、概率的运算 .....	(9)
第三节 随机变量与分布函数 .....	(17)
一、随机变量 .....	(17)
二、分布函数 .....	(19)
第四节 离散分布 .....	(21)
一、离散型随机变量及其概率函数 .....	(21)
二、离散型随机变量的数学期望与方差 .....	(22)
三、几个常见的离散分布 .....	(27)
第五节 连续分布 .....	(36)
一、连续型随机变量及其密度函数 .....	(36)
二、连续型随机变量的数学期望及方差 .....	(40)
三、两个重要的连续分布 .....	(41)
第六节 中心极限定理 .....	(47)
一、同分布的中心极限定理 .....	(47)
二、二项分布的正态近似(棣莫佛—拉普拉斯定理) .....	(49)
习题 .....	(53)

第二章 数据资料的统计分析 .....	(58)
第一节 数据资料的获得与整理 .....	(58)
一、抽样技术 .....	(58)
二、试验设计 .....	(64)
三、试验数据的整理 .....	(73)
第二节 样本平均值与标准差 .....	(76)
第三节 总体参量估计 .....	(82)
一、点估计 .....	(82)
二、 $\mu$ 的区间估计 .....	(83)
第四节 单个样本的统计假设检验 .....	(87)
一、一般原理 .....	(88)
二、单个样本显著性检验 .....	(91)
第五节 两个样本的差异显著性检验 .....	(94)
一、两个方差的检验—— $F$ 检验 .....	(95)
二、两个平均数差异的显著性检验 .....	(98)
三、两个百分率的比较 .....	(107)
第六节 计数资料的显著性测验—— $\chi^2$ 测验 .....	(108)
一、适合性检验 .....	(111)
二、独立性检验 .....	(115)
第七节 方差分析 .....	(125)
一、单因素多组群的方差分析 .....	(126)
二、两因素多组群的方差分析 .....	(138)
三、拉丁方资料的方差分析 .....	(144)
四、正交拉丁方试验资料的方差分析 .....	(146)
五、方差分析的基本假定和数据转换 .....	(151)

第八节 回归分析 .....	(154)
一、一元线性回归分析 .....	(155)
二、一元非线性回归分析 .....	(172)
三、多元线性回归分析 .....	(180)
四、逐步回归分析 .....	(187)
第九节 协方差分析 .....	(196)
习题 .....	(203)
第三章 多元分析 .....	(226)
第一节 数据的处理 .....	(226)
一、数据的简缩、转换及标准化 .....	(226)
二、相似系数 .....	(227)
第二节 分类方法 .....	(232)
一、等级聚合分类 .....	(233)
二、等级分划分类 .....	(233)
第三节 排序方法 .....	(234)
第四章 SPSS 软件的应用 .....	(239)
第一节 SPSS 的使用简介 .....	(239)
一、SPSS for Windows 的主界面 .....	(239)
二、定义变量 .....	(240)
第二节 SPSS 的应用 .....	(245)
一、置信区间的估计 .....	(245)
二、单个样本 $T$ 检验 .....	(247)
三、两个样本 $T$ 检验 .....	(248)



四、方差分析 .....	(254)
五、回归分析 .....	(287)
六、协方差分析 .....	(304)
七、列联表的 $\chi^2$ 检验 .....	(308)
八、因子分析 .....	(312)
附表 1 标准正态分布表 .....	(327)
附表 2 正态分布上侧分位数 ( $u_\alpha$ ) 表 .....	(331)
附表 3 $t$ 分布的分位数表 .....	(332)
附表 4 $\chi^2$ 分布的上侧分位数 ( $\chi_\alpha^2$ ) 表 .....	(333)
附表 5 $F$ 检验的临界值 ( $F_\alpha$ ) 表 .....	(337)
附表 6 多重比较中的 $Q$ 值表 .....	(341)
附表 7 相关系数检验表 .....	(345)
附表 8 $r$ 与 $z$ 的转换表 .....	(346)
附表 9 正交拉丁方表 .....	(347)
附表 10 正交表 .....	(350)
附表 11 从百分率坐标变换为概率坐标表 .....	(359)
参考书目 .....	(360)

# 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、初生重以及其他经济性状方面,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普遍存在,它掩盖住、有时甚至严重地蒙蔽着人们对事物真相的认识.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,而需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,是由种种纯粹的偶然所构成的,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他事物的内部规定性.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.在生物科学研究中,经常要同大量的试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.因此,注意事物的量变,深入研究量与量之间的关系,是从事任何一项生物科学试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,生物统计学则是数理统计学的原理和方法在生物科学研究中的应用.因此,要讨论生物统计学的概念和方法,就离不开概率论和数理统计学最一般的知识.此外,在描述和讨论一个生物统计学问题时,也离不开数学上所特有的方法,即逻辑推理方法.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。

偶然现象是一种非确定性现象或称为随机现象,例如,考察动物的初生重,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,就会发现动物初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.概率论与数理统计学是研究和揭示随机现象的统计规律性的一门科学.

## 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件,称为随机事件.通常用大写英文字母 A, B, C …表示.

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,如事件为掷出偶数点,其结果为2,4,6,它包括了三个基本事件.

#### 4. 样本点与样本空间

现代集合论为表述随机试验提供了一个方便的工具.

每一个基本事件都可以用一个点来描述,称为样本点.

为了便于研究随机试验,将随机试验的所有样本点组成的集合叫做样本空间.例如,抛三枚硬币的样本空间为  $2^3 = 8$ ;从一副不含大小王的扑克牌中抽两张牌的样本空间为  $C_{52}^2 = 1326$ .

#### 5. 随机变量

给随机事件加以数量化,即可用一个变量来表示随机试验的结果,这个变量就叫做随机变量.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,这些不同的观测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.

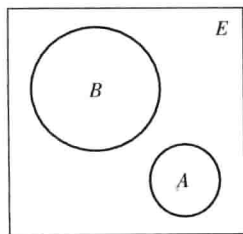
由于一个研究对象所包含个体的数目往往很大,甚至无穷,因此,实际上进行观察的只是总体中的一小部分,从总体中抽出的这个有限部分就构成了样本.因为从统计学的要求,总体中每个元素被抽出作为样本的机会都必须均是均等的,这就要保证随机抽样.由随机抽样得到的样本就叫做随机样本.以  $r$  来表示样本所包含的元素数目,这叫做样本的容量.根据样本容量的大小,又分大样本

与小样本.一般以 30 为界,容量大于 30 的为小样本,小于、等于 30 的为小样本.

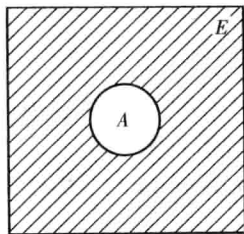
### 7. 事件之间的关系与事件的运算

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.

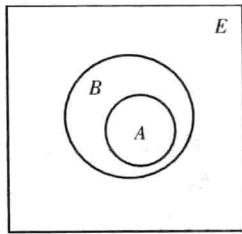
符号	概率论	集合论
$S$	样本空间 必然事件	全集(包含了全部元素的集合)
$\Phi$	不可能事件	空集(不包含任何元素的集合)
$E$	基本事件 样本点	元素
$A$	事件	子集
$\bar{A}$	$A$ 的对立事件	$A$ 的余集
$A \subset B$	事件 $A$ 发生导致事件 $B$ 发生	$A$ 是 $B$ 的子集(包含)
$A = B$	事件 $A$ 与事件 $B$ 相等	$A$ 与 $B$ 相等
$A \cup B$	事件 $A$ 与事件 $B$ 至少有一个发生,也可表示为 $A + B$	$A$ 与 $B$ 的并集
$A \cap B$	事件 $A$ 与事件 $B$ 都发生,也可表示为 $AB$	$A$ 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(互斥)	$A$ 与 $B$ 没有相同元素



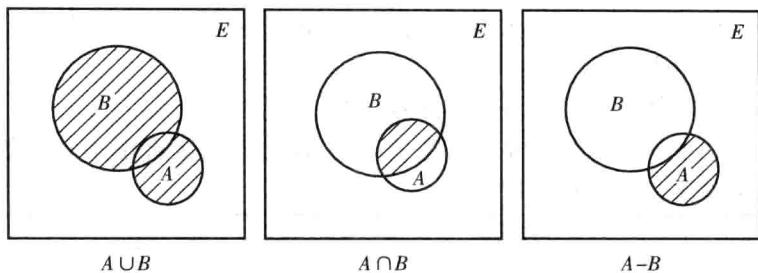
$A, B$  不相容



$\bar{A}$



$A \subset B$



概率论中事件之间的关系与运算和集合论中集合之间的关系与运算是一致的.

例 1.1 如果  $A \subset B$ , 则  $A \cup B = \underline{\hspace{2cm}}$ ,  $A \cap B = \underline{\hspace{2cm}}$ .

解: 设事件  $A$  以图中的小圆表示, 事件  $B$  以大圆表示, 则有  $A \subset B$ , 于是

$$A \cup B = B \quad A \cap B = A$$

例 1.2 证明 (1)  $\overline{A \cup B} = \bar{A} \cap \bar{B}$

$$(2) \overline{A \cap B} = \bar{A} \cup \bar{B}$$

证: (1) 事件  $A \cup B$  表示  $A$  与  $B$  中至少有一事件发生, 对立事件就是  $A$  与  $B$  都不发生, 即  $\bar{A} \cap \bar{B}$ , 所以  $\overline{A \cup B} = \bar{A} \cap \bar{B}$

(2) 设  $\bar{A} = C$ ,  $\bar{B} = D$ , 则  $\bar{C} = A$ ,  $\bar{D} = B$ , 根据已证明的(1), 可知

$$\overline{C \cup D} = \bar{C} \cap \bar{D}$$

即

$$\overline{\bar{C} \cap \bar{D}} = C \cup D \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$$

例 1.3 化简:  $(A \cup B)(A \cup C)$

解:  $(A \cup B)(A \cup C) = AA \cup AB \cup AC \cup BC$

因为

$$AA = A, \quad AB \cup AC \subset A$$

所以

$$AA \cup AB \cup AC = A$$

因此

$$(A \cup B)(A \cup C) = A \cup BC$$

## 第二节 概率的计算

### 一、概率的基本概念

#### 1. 概率的统计定义

设随机事件  $A$  在  $n$  次试验中发生了  $m$  次, 则比值  $m/n$  叫做随机事件  $A$  的频率, 记作:

$$W(A) = \frac{m}{n}$$

在不同的试验序列中, 当试验次数  $n$  充分大时, 随机事件的频率将围绕某一确定的常数  $p$  作平均幅度愈来愈小的摆动, 这种性质称为随机事件频率的稳定性. 其中的  $p$  即为事件  $A$  的概率, 记作  $P(A)$ .

在实际中, 当概率不易求出时, 人们常取试验次数很大时事件的频率作为概率的估计值, 称此概率为统计概率.

例 1.4 在每一个试验序列中, 抛掷钱币 500 次, 记录正面向上(设为事件  $A$ ) 的次数, 并计算正面向上的频率, 进而估算出概率值.

序列	1	2	3	4	5	6	7	8	9	10
频数 ( $m$ )	251	249	256	253	251	246	244	258	262	247
频率 $W(A)$	0.502	0.498	0.512	0.506	0.502	0.492	0.488	0.516	0.524	0.494

解:



可以认为正面向上的概率  $P(A) = 0.5$ .

概率是事件在试验结果中出现可能性大小的定量计量. 概率有以下明显的性质:

(1) 任何事件( $A$ )的概率均满足

$$0 \leq P(A) \leq 1$$

(2) 必然事件的概率为 1

$$P(S) = 1$$

(3) 不可能事件的概率为 0

$$P(\phi) = 0$$

## 2. 概率的古典定义

概率的统计定义是在大量试验中,以频率的稳定性为基础提出来的,它适用于一般的随机现象. 对于某些随机现象,可以用概率的古典定义来计算事件的概率. 这类随机现象的特点是:(1) 随机试验的全部可能的结果(基本事件数)是有限的;(2) 各基本事件间是互不相容且等可能的. 对于这类随机现象的概率类型习惯上称为古典概率. 在满足上述条件下,事件  $A$  的概率是  $A$  中所包含的基本事件数(样本点数)  $m$  与基本事件总数(样本空间)  $n$  的比值:

$$P(A) = \frac{\text{事件包含的样本点数}}{\text{样本空间}} = \frac{m}{n}$$

例 1.5 一组 20 只小白鼠,其中有 8 只雄鼠,12 只雌鼠,从中任取 5 只,问其中有 2 只雄鼠,3 只雌鼠的概率是多少?

解:从 20 只小白鼠中,任取 5 只的每一种取法都是等可能的. 从 20 只中任取 5 只的每一个结果是一个基本事件,基本事件的总数为  $C_{20}^5$ ,事件  $A$  是 5 只中有 2 只雄鼠,3 只雌鼠的事件,因此,事件  $A$  中的基本事件数为  $C_8^2 C_{12}^3$ ,根据概率的古典定义,事件  $A$  的概率为: