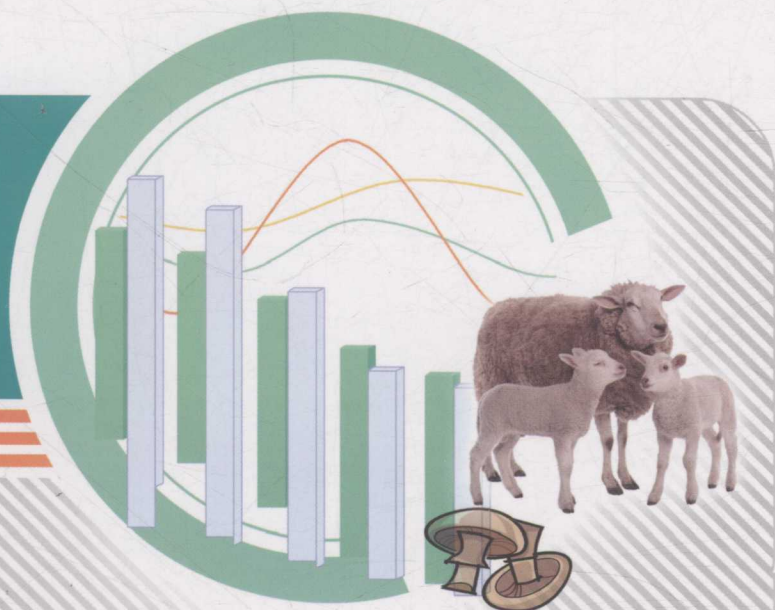




高等学校规划教材 | 畜牧兽医类

生物 统计学



主编 ● 刘安芳 伍莲

SHENGWU

TONGJIXUE

案例版



西南师范大学出版社
国家一级出版社 全国百佳图书出版单位



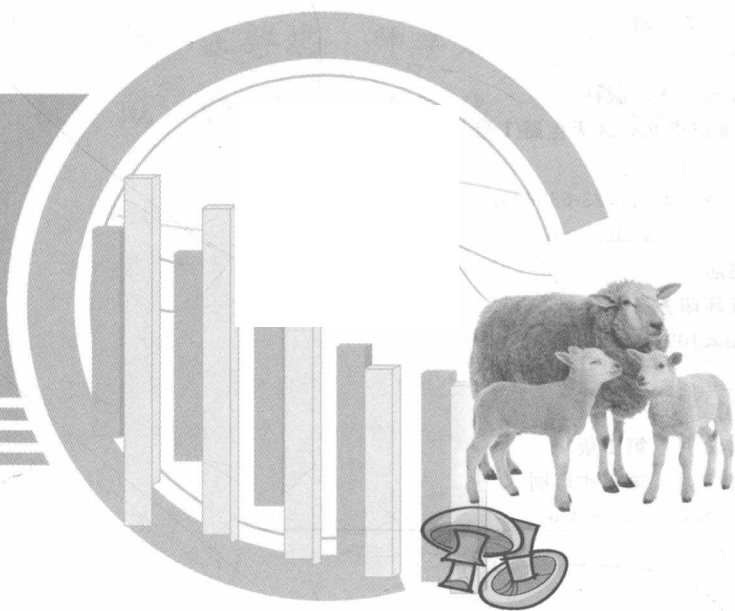
高等学校规划教材 | 畜牧兽医类

生物 统计学

主编 ● 刘安芳 伍莲

SHENGWU

TONGJIXUE



西南师范大学出版社
国家一级出版社 全国百佳图书出版单位

图书在版编目(CIP)数据

生物统计学 / 刘安芳, 伍莲主编. —重庆: 西南师范大学出版社, 2013. 6

ISBN 978-7-5621-6304-6

I. ①生… II. ①刘… ②伍… III. ①生物统计
IV. ①Q-332

中国版本图书馆 CIP 数据核字(2013)第 139458 号

生物统计学

主 编 刘安芳 伍 莲
副主编 孙桂荣 虞德兵 袁树楷

责任编辑: 杜珍辉

责任校对: 莫 琼 荣 霞

封面设计: 魏显锋

出版发行: 西南师范大学出版社

地址: 重庆市北碚区天生路 1 号

邮编: 400715

市场营销部电话: 023-68868624

<http://www.xscbs.com>

经 销: 新华书店

印 刷: 重庆五环印务有限公司

开 本: 787mm×1092mm 1/16

印 张: 25.5

字 数: 650 千字

版 次: 2013 年 8 月 第 1 版

印 次: 2013 年 8 月 第 1 次印刷

书 号: ISBN 978-7-5621-6304-6

定 价: 42.00 元

衷心感谢被收入本书的图文资料的原作者, 由于条件限制, 暂时无法和部分原作者取得联系。恳请这些原作者与我们联系, 以便付酬并奉送样书。

并有印装质量问题, 请联系出版社调换。

版权所有 翻印必究

高等学校规划教材·畜牧兽医类

总编委会 / ZONG BIAN WEI HUI

总主编：王永才 刘娟

编委(排名不分先后)：

刘娟	黄庆洲	伍莉	朱兆荣
罗献梅	甘玲	谢和芳	刘安芳
兰云贤	曾兵	杨远新	黄琳凯
陈超	王鲜忠	帅学宏	黎德斌
段彪	伍莲	陈红伟	

《生物统计学》

编委会 / BIAN WEI HUI

主 编：刘安芳 伍 莲

副主编：孙桂荣 虞德兵 袁树楷

顾 问：谢和芳(西南大学)

参 编(按姓名拼音顺序排列)：

刘安芳(西南大学)

李菊娣(辽宁石油化工大学)

李 辉(贵州大学)

罗宗刚(西南大学)

康 丽(山东农业大学)

孙桂荣(河南农业大学)

施晓丽(贵州大学)

伍 莲(西南大学)

虞德兵(南京农业大学)

袁树楷(西南大学)

主 审：朱汉春(西南大学)

前 言

《生物统计学》是动物科学、动物医学、水产养殖、食品科学、草业科学、生物技术等相关专业学生既重要又实用的专业基础课,随着计算机应用技术的快速发展,生物统计学的理论及实验教学方法也发生了较大变化。结合当今教学的实际情况,在保持本学科的系统性和科学性的前提下,在教材编写中注重引入相关学科发展的新知识、新成果及统计分析 with 计算机科学的结合,注重拓宽学生的知识面和 提高实践能力,力求体现强基础、重应用和素质教育、创新教育的教学目标,做到科学性 with 实用性、先进性与针对性相统一。

在内容方面着重于基本理论、基本技能和基本方法的讲授,力求由浅入深、循序渐进、深入浅出、简明易懂,注重培养学生的统计思维方法和正确地分析问题和解决实际问题的能力。每一种分析方法都安排有步骤完整、过程详细的案例或案例分析予以说明;在每章后安排有思考与练习题,作为学生课内课外作业练习之用;附录中增加了实验指导,内容侧重于对所学方法的训练。

本书由国内 6 所高等院校的 10 多位教学经验丰富、在教学第一线的教师联合编写而成,西南大学谢和芳教授担任顾问。他们中间既有多年从事生物统计学教学实践与改革的专家、教授,也有充实到生物统计学教师队伍中的年青教师。具体分工如下:第一章,刘安芳;第二章,康丽;第三章,李辉;第四章,刘安芳;第五章,虞德兵;第六章,伍莲;第七章,李菊娣;第八章,施晓丽;第九、十、十一章,孙桂荣;实验指导,伍莲、袁树楷、刘安芳;附录、附表,罗宗刚。全书由刘安芳统稿,朱汉春审稿。

本教材主要面向高等院校动科、动医、水产、食品等相关专业的本科学生,也可作为综合性大学、师范院校生物专业本科生、研究生教学用书,还可作为农林牧渔业、生物技术科技工作者进行科学研究的参考书。

本教材在编写过程中参考了有关生物统计学教材、相关的中外文献和专著,在此,对这些教材、专著、文献的作者表示衷心的感谢!

本书在全体编者的大力协作及共同努力下完成,各位编者认真负责、尽心尽力,为保证教材的质量付出了辛勤劳动,在此谨向各位参编者表示深切的谢意!

限于编者的知识水平和编写能力,书中错误、缺点在所难免,敬请生物统计学专家和广大读者批评指正,以便修改。

编 者

2013 年 5 月

目 录

第一章 绪论

第一节	生物统计学概述	1
第二节	生物统计学的主要内容	5
第三节	生物统计在生物科学研究中的作用	6
第四节	生物统计学的发展概况	7
第五节	生物统计的常用术语	11

第二章 资料的整理与描述

第一节	资料的分类	15
第二节	资料的来源	17
第三节	资料的整理	17
第四节	常用统计表与统计图	22
第五节	平均数	27
第六节	变异数	31

第三章 概率及其理论分布

第一节	事件与概率	41
第二节	随机变量及其概率分布	47
第三节	正态分布	49
第四节	二项分布	55
第五节	波松分布	58
第六节	样本平均数的抽样分布	61
第七节	t 分布	64

第四章 均数差异显著性检验—— t 检验

第一节	显著性检验的意义	69
第二节	显著性检验的原理	72
第三节	样本平均数与总体平均数的差异显著性检验	80
第四节	两样本均数差异的显著性检验	84
第五节	百分数资料的 t 检验	95
第六节	总体参数的区间估计	100

第五章 次数资料的假设检验——χ^2 检验	
第一节 χ^2 检验的意义和基本原理	109
第二节 适合性检验	112
第三节 独立性检验	118
第六章 方差分析	
第一节 方差分析的基本原理与步骤	131
第二节 单因素试验资料方差分析	143
第三节 两因素试验资料的方差分析	146
第四节 数据的转换	160
第五节 方差分析的数学模型与期望均方	162
第七章 简单回归与相关	
第一节 概述	177
第二节 直线相关	179
第三节 直线回归	186
第四节 曲线回归	200
第八章 多元线性回归与相关	
第一节 多元线性回归	214
第二节 复相关分析和偏相关分析	221
第三节 多项式回归	225
第九章 协方差分析	
第一节 协方差分析的原理与意义	233
第二节 单因素资料的协方差分析	234
第三节 两因素资料的协方差分析	242
第十章 非参数检验	
第一节 符号检验	251
第二节 秩和检验	254
第三节 非参数的等级相关分析	261
第十一章 半致死量	
第一节 半数致死量的概念	265
第二节 半数致死量的实验设计方法	265
第三节 半数致死量的测定方法	267
附录一 常用统计分析软件简介	
一 SAS 统计软件应用方法简介	273

二 SPSS 统计软件应用方法简介	279
-------------------------	-----

附录二 生物统计实验指导

实验一 Excel 在资料整理中的运用	287
实验二 Excel 在资料描述中的运用	295
实验三 Excel 在二项分布和正态分布概率计算中的运用	296
实验四 Excel 在 t 检验中的运用	300
实验五 Excel 在方差分析中的运用	306
实验六 Excel 在 χ^2 检验中的运用	313
实验七 Excel 在一元线性回归与相关中的运用	316
实验八 Excel 在多元回归分析与相关中的运用	320
实验九 Excel 在协方差分析中的运用	327
实验十 Excel 在试验设计中的运用	332

附表

附表 1 标准正态分布表	345
附表 2 正态分布的双侧分位数 U_{α} 值表	348
附表 3 t 值表(两尾)	349
附表 4 χ^2 值表(一尾)	351
附表 5 F 值表(方差分析用)	353
附表 6 q 值表(方差分析用)	356
附表 7 Duncan's 新复极差检验 SSR 值表	359
附表 8 r 与 R 的显著数值表	363
附表 9 符号检验用 K 临界值表	365
附表 10(1) 符号秩和检验用 T 临界值表	365
附表 10(2) 秩和检验用 T 临界值表	366
附表 10(3) 秩和检验用 H 临界值表	367
附表 11 等级相关系数 r_s 临界值表	368
附表 12 随机数字表	369
附表 13 百分数反正弦($\sin^{-1}\sqrt{x}$)转换表	375
生物统计学汉英文名词对照表	379
Excel 统计分析函数	395

参考文献

第一章 绪论

【本章导读】

生物统计是数理统计的原理和方法在生物科学研究中的应用,它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数量资料的学科。其目的是通过样本的特征来估计总体,对所研究的总体进行合理的推断,得到对客观事物本质和规律性的认识。本章主要介绍生物统计学的特点、常用的统计术语,学习生物统计学的目的及学习方法,阐述生物统计学的主要内容、在生物科学研究中的作用及生物统计学的发展概况。

【本章重点】

生物统计学的作用及常用的统计术语。

【本章难点】

对统计术语的理解,建立统计思维方法。

第一节 生物统计学概述

一、生物统计学的概念及特点

(一)生物现象的基本特征

1. 变异性

由于影响生命活动的各种因素错综复杂,从而导致生命现象呈现出多样性特征,表现为有性繁殖的种群中没有两个完全相同的个体。例如:同一组小鼠,即使品种、性别、年龄相同,但它们的体重、尾长和各项生理指标的数值都会有所不同;同一栏猪,由于吃了发霉的饲料而发生肠炎,病情的轻重也有所不同,这些都是普遍存在变异的实例。所以,通过调查和试验所获得的资料常常是一组存在变异的数据。

如测量 800 头猪三月龄的体重,所获得的 800 个数据,彼此不完全相同,表现出一定程度的变异。产生这种变异的原因,有的已被人们所了解,例如品种、性别、年龄、初始重、健康状况、饲养条件等不同,使得所测猪的三月龄体重表现出差异,另外还有许多内在和外在的因素未被人们所认识,由于这些人们已了解的因素和人们尚未认识而无法控制的因素的作用,使得通过调查或试验得来的数据普遍具有变异性。

2. 不确定性(随机性)

生命现象是物质运动的表现,呈现千差万别、多种多样的特性,有的在一定条件下必然发生或不发生,如鱼儿离不开水、动物缺氧一定会死亡、未受精的蛋孵化不出小鸡等,这类现象称为确定性现象;有的在一定条件下不一定发生,如接受同一药物处理的动物,不一定产生相



同的疗效,母猪的妊娠期不一定是 114 天,这类现象称为不确定性现象或随机现象,随机现象是生物科学中大量存在、经常遇到的一类现象,生物个体之间的差异许多是由随机因素造成的。正是由于生物之间存在这种不确定性,使得生物界的事物不能被准确预测。

3. 复杂性

生物的性状受许多因素的影响,有环境、遗传以及环境和遗传的互作等,虽然目前已能从分子水平认识生物的遗传现象,但至今对它的遗传机制还没有完全清楚的认识,有的环境因素是可以人为控制的,但更多的是不可控制的,这些就造成了生物现象的复杂性。

(二) 生物统计学的概念

生物统计学是数理统计的原理和方法在生物科学研究中的应用,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数量资料的学科。在生物科学研究的过程中,逐渐与数学的发展结合起来形成了本门学科,因此,它是应用数学的一个分支,属于生物数学的范畴。

(三) 生物统计学的特点

1. 由样本推断总体

生物统计在某种意义上来说,就是研究生物科学中以样本来推断总体的一门学科,由样本来推断总体是它的第一个特点。人们在从事科学研究时,往往是通过某事物的一部分即样本,来估计事物全体即总体的特征,目的是为了由样本推断总体,从特殊推导一般,从而对所研究的总体作出合乎逻辑的推论,得到对客观事物本质和规律性的认识。

在生物科学研究中,我们所期望知道的是总体,而不是样本,可是在实践中,我们所得到的却是样本而不是总体。例如,研究中国黑白花乳牛 305 天产乳量,所有中国黑白花乳牛 305 天产乳量资料太庞大,几乎不可能全部记载,只能采用抽样的方法,计算样本的统计量,用样本的统计量来推断总体参数。我们可以随机抽测 200 头中国黑白花乳牛 305 天的产乳量作为样本来推断中国黑白花乳牛 305 天的产乳量。

2. 具有概率性

生物统计学中的“由部分推断全部”的统计结论在一定概率保证下是正确的,具有概率性是它的第二个特点。

生物统计学的推理思想与其他自然学科不同,例如“血是红的,天鹅是白的,乌鸦是黑的”,这是通过简单枚举产生的结论,然而,在南极洲却发现了白色血液的鱼,在澳洲发现了黑天鹅,在日本发现了白乌鸦。如果根据其他自然学科推理方法,出现反例就否定原假设,这就要否定“血是红的,天鹅是白的,乌鸦是黑的”的结论。如果要概括上述结果,可以这样说,“有多大的可能性动物的血是红色的,有多大的可能性天鹅是白色的,有多大的可能性乌鸦是黑色的”。

3. 理论和实践密切联系

生物统计学不是一门纯理论的学科,而是一门理论和实践并重,理论和实践密切联系的学科,这是它的第三个特点。生物统计学有其自身的理论体系,但同时必须面对大量来源于实践的数据资料,否则就失去了它存在和发展的价值,处理和分析数据资料是生物统计学的首要任务。

(四) 统计分析的基本特点

1. 通过样本来推断总体

研究的目的是要了解总体,然而能观测到的却是样本,通过样本来推断总体是统计分析的基本特点。总体有无限总体和有限总体,有限总体也存在着包含的个体数目太多的情况,要获得全部观测值需花费大量人力、物力和时间,或者观测值的获得带有破坏性,例如,猪瘦肉率的测定,要求将猪屠宰、分割后才能测定和计算出瘦肉率,破坏性大,不允许也没有必要对每一头猪一一进行屠宰测定,常常随机选择一定数量的猪,用屠宰测定后的样本数据来对总体作出推断。

2. 有很大的可靠性但有一定的错误率

为了更可靠地从样本来推断总体,要求样本具有一定的容量和代表性,但样本毕竟只是总体的一部分,通过样本来推断总体不可能是百分之百的正确,结论有很大的可靠性但也有一定的错误率这是统计分析的又一特点。

二、为什么要学习生物统计学

生物统计是分析生物现象客观规律的基本工具,作为从事动物科学相关研究与生产管理的实践者,学习生物统计学具有重要的意义。

(一) 采用统计学方法,可发现不确定现象背后隐藏的规律

变异(*Variation*)是生物科学中的普遍现象,变异使得试验或观察的结果具有不确定性。生物现象的特点决定了不能仅通过描述性的定性科学或决定性的数量科学来解决生物学领域中的众多问题。例如,新引入品种是否对本地品种有改良效果,对白鲢鱼增长速度有影响的因素有哪些,遗传与环境哪个更重要,如何选择遗传上最优良的种畜……对于诸如此类的问题,只有通过生物的数据资料进行统计分析才能得到答案,只有生物统计学才能告知如何通过科学的调查或试验获得高质量的数据资料,怎样对所获得的数据资料进行无偏差地分析,如何根据分析的结果做出尽可能可靠的结论,并发现现象背后隐藏的规律。

(二) 用统计学思维方式考虑有关生物研究中的问题

通过学习生物统计学,可以培养用统计学思维方式考虑有关生物和农业研究中的问题,例如,仔猪初生重与母猪的体重是否有关,两种中草药促生长的配方哪种更好,两个肉鸡品种在相同饲养条件下哪个品种生长更快,假设甲品种肉鸡42日龄平均体重为2.5 kg,乙品种的平均体重为2.3 kg,需要对观察到的现象用统计方法分析其产生的误差大小,才能作出统计推断,是否甲品种的生长速度高于乙品种。

(三) 正确撰写研究论文,获得循证生物学证据的主要手段

通过学习生物统计学,能保证科研工作者的研究论文能通过统计学审查,对不同生物学杂志发表论文的调查显示,有统计错误的论文所占比例高达72%,最低也有20%,国内外权威生物学杂志对来稿都有统计学要求或统计学指导。

“良好愿望的生物学”(Well-meaning biology)转入“以证据为基础的生物学”(Evidence-based biology, EBB)需要有统计学方法的支持,全世界每年发表的生物学术论文,能作为可靠“证据”的论文并不多,而学习生物统计学,可让发表的生物学术论文作为可靠“证据”的论文。



(四)为学习相关学科打下基础

要学好动物科学、动物医学、水产养殖等生物学科,必须学好生物统计学,例如,数量遗传学是应用生物统计方法研究数量性状遗传与变异规律的学科,如果没有生物统计的知识,就无法学好数量遗传学,生物科学工作者也必须学习和掌握统计方法,才能正确认识客观事物存在的规律性,提高工作质量。

总之,生物统计是一种很有用的工具,正确使用这一工具可以使生物科学研究更加有效,它是每位从事生物科学研究的科研工作者都必须掌握的基本工具。

三、学习生物统计学的方法和要求

生物统计学是数学与生物学相结合的一门学科,与生物学的其他学科具有很大的不同,在性质上属于生物学领域的应用数学,所包含的公式很多。

(一)弄懂统计的基本原理和基本公式

在生物统计的学习中首先要弄懂统计的基本原理和基本公式,要理解每一公式的含义和应用条件,但一般可不必深究其数学推导、证明和数学原理。

(二)认真做好习题作业

作为一门工具课,必须认真做好习题作业,加深对公式及统计步骤的理解。只有通过一定数量的实践和练习,熟练使用函数型电子计算器,正确使用统计软件(Excel, SPSS 和 SAS),才能达到熟练应用生物统计方法的目的。

(三)注意培养科学的统计思维方法

生物统计意味着一种全新的思考方法,从概率的角度来思考问题和分析科学试验的结果,避免绝对肯定或绝对否定的武断结论,及单凭感觉不做检验的简单判断。

(四)理论联系实际

在日常生产实践中,无论是科学研究的成果、畜禽生产数据的总结,还是学生考试成绩的分析、公司的业绩年报、国家经济指标的公布等,都是统计分析后的结果,可见统计分析在我们的生产生活中无处不在。因此,如何在学好生物统计学理论知识的基础上,更好的在生产实践中应用它具有重要的意义,即要做到理论联系实际。

要做到理论联系实际,需要我们首先能利用所学的生物统计知识读懂教材案例、科研论文、报表等中的统计分析方法、统计结果及表达方式,分析它们为什么采用这些统计分析方法及如何对数据分析的结果进行解释等;学会根据实际情况,针对毕业课题、生产数据、调查或试验结果等选择合适的统计分析方法、分析出正确的结果、得出有意义的统计结论。

第二节 生物统计学的主要内容

生物统计的基本内容,概括起来主要包括试验设计和统计分析两大部分,从统计分析方法来讲大致可分为描述性统计、显著性检验和相关与回归分析。

一、描述性统计

描述性统计就是对原始数据进行整理并做基本分析,由生产或试验所得的数据资料(Data)很多,需要加以整理和分析。整理的主要内容就是检查原始数据的完整性、正确性,作次数分布表和统计图,通过统计表、图可以大致看到所得资料的集中、离散情况,并从资料中计算出几个主要的统计量,即平均数、标准差及标准误等,以表示该资料的数量特征,用这些统计量估计总体参数,分析资料的集中趋势、变异程度以及平均数的可靠性,作出初步统计分析。

二、显著性检验

显著性检验又称假设检验,常用的有 t 检验、方差分析、 χ^2 检验等。

(一) t 检验

t 检验主要用于检验两个处理平均数差异是否显著,或单个样本平均数与总体平均数是否有显著差异。在生物科学研究中,经常需要比较两组平均数之间的差异是否显著,即检验差异由偶然因素引起的可能性有多大。例如,调查两个不同品种雏鹅的 70 日龄体重,甲品种 80 只,其平均体重为 3.56 kg,乙品种 85 只,其平均体重为 3.36 kg,如果要检验这两个鹅品种 70 日龄平均体重之间的差异是由于品种不同造成的,还是由其他偶然因素所引起的,这就可采用 t 检验的统计分析方法,做出比较可靠的判断。

(二) 方差分析

方差分析又名变异量分析,是对多个处理平均数间差异是否显著进行比较的一种统计检验方法,其原理是应用数学方法把试验的总变异剖分为各个不同变异原因所形成的各种变异,进行显著性检验,并在显著性检验的基础上再进行多重比较。

(三) χ^2 检验

χ^2 检验是最常用的属性资料的统计检验方法,主要用于由质量性状得来的次数资料的显著性检验等。

生物学领域中有许多性状不能直接用测量的方法来加以衡量,一般称之为属性性状。例如,猪的毛色、性别中的雌和雄、药物试验的治愈或无效等均可以应用属性统计的方法,通过对具有相同属性的计数来分析、检验其实际观测值与理论值是否相符。

三、相关与回归

研究呈平行关系变量之间的密切程度,称为相关(Correlation),两者间的相关程度可用相关系数来表示,例如:人的身高与体重存在着一定程度的相关,一般身高越高,其体重越大。



回归(*Regression*)是指两个或两个以上的变量存在着从属关系,一个变量(x)变化时,引起另一变量(y)相应变化的估计,它们的关系可以从量的方面加以估算,这就是回归分析。

通过对资料进行相关、回归分析,可以揭示出试验指标或性状间的内在联系。

四、试验设计

所谓试验设计,主要是指如何选择试验材料或动物,进行合理的分组和安排实验等,其目的是为了尽量减少和控制试验误差,并对试验误差做出无偏的估计,主要的试验设计方法有完全随机设计、配对设计、随机单位组设计、拉丁方设计、正交设计等。为了使试验结果成为有用而可靠的科学资料,在开始试验前,认真地进行试验设计是非常必要的。

第三节 生物统计在生物科学研究中的作用

生物统计在畜禽、水产等科学研究中具有十分重要的作用。

一、提供整理、分析资料的科学方法

生物科学研究收集到的资料,往往是比较零乱的,看不出任何的规律。必须对资料进行整理,并在整理的基础上进行合理的统计分析,从中归纳出事物的规律性,用以指导实践。

例如,调查某品种鸡的产蛋数,可以得到不同季节、不同年龄、不同个体母鸡产蛋数的大量原始数据,从这些杂乱的原始数据难以看出规律性。若运用生物统计方法对这些数据进行加工整理,使之条理化,即可大体了解不同季节、不同年龄、不同个体母鸡产蛋数的一般情况及变异特征,得到母鸡在什么季节产蛋数最高,什么年龄产蛋数开始下降,以及母鸡适宜的利用年限等许多有用信息。

二、判断试验结果的可靠性与有效性

一般在试验中要求除试验因素以外,其他条件都应尽可能控制一致,但在实践中无论试验条件控制得如何严格,其试验结果总会受到试验因素和其他偶然因素的影响。由于存在试验误差,从试验所得的数据资料必须借助于统计分析方法才能获得可靠的结论。

例如,某试验场要研究甲乙两种饲料对肉用仔鸡增重及饲料报酬的影响,选择同品种及体重接近的500只肉用雏鸡,随机分为两个组,一组肉鸡喂甲种饲料,另一组肉鸡喂乙种饲料,8周龄后称其体重并计算饲料消耗量。分析比较这些资料,就应该运用统计方法进行分析,分析两组鸡体重与饲料消耗的差异,是由饲料不同造成的,还是由其他偶然因素所引起的,统计分析之后才能得出结论。

三、确定事物之间的相互关系

科学试验的目的,不仅要研究事物的特征特性,同时还要研究事物间相互关系的联系形式,研究事物之间的联系形式以及相关程度的方法是生物统计的一个重要部分。

例如,测定某群奶牛第一胎的产乳量和它以后几胎的产乳量之间的相互关系,就可以根据第一胎产乳量的高低来推断它终生的产乳量,这可为早期选择和淘汰低产乳牛提供科学根据。



四、提供试验设计的一些原则

一个好的试验设计,可以用较少的人力、物力和时间,最大限度地获得丰富而可靠的资料,尽量降低试验误差,从试验所得的数据中能够无偏地估计处理效应和试验误差,并得出正确的结论。相反,试验设计不周,不仅不能得到正确的试验结果,而且还可能带来经济上和其他方面的损失。为了以较少的人力、物力和财力取得较多的试验信息和较好的试验结果,在一些生物学研究中,需要科学地进行试验设计,包括样本容量的确定、抽样方法、处理设置、重复次数的确定以及试验的安排等,这必须以生物统计原理为依据。

生物统计对于进行生物科学研究极其重要,它是每一个生物科技工作者必须掌握的基本工具,随着生物统计方法的普及、计算工具的改进、统计计算程序的编制,已有越来越多的科技工作者掌握并在实际研究工作中应用了生物统计,取得了显著成效。

第四节 生物统计学的发展概况

统计一词起源很早,远在奴隶、封建社会,由于战争和税收的需要,就开始有统计技术了,因此,统计发展史可以追溯到远古的原始社会。而能使人类的统计实践上升到理论,开始成为一门系统的学科,距今只有 300 多年的历史。从统计学的产生和发展过程来看,大致可划分为古典记录统计学、近代描述统计学和现代推断统计学三个时期。

一、古典记录统计学

古典记录统计学形成大致在 17 世纪中叶至 19 世纪中叶,在这个兴起阶段,统计学还是一门意义和范围不太明确的学问,在它用文字或数字如实记录与分析国家社会经济状况的过程中,初步建立了统计研究的方法和规则,直到概率论被引进之后,才逐渐成为一项较成熟的方法。

(一)拉普拉斯的主要贡献

最初卓有成效地把古典概率论引进统计学的是法国天文学家、数学家、统计学家拉普拉斯(*P. S. Laplace*, 1749~1827),后来比利时的大统计学家凯特勒指出,统计学应从拉普拉斯开始,拉普拉斯的主要贡献有:

1. 发展了概率论的研究

拉普拉斯最早系统地把数学分析方法运用到概率论研究中去,建立了严密的概率数学理论,成为古典概率论的集大成者。

2. 推广了概率论在统计中的应用

拉普拉斯相当明确地指出,概率论能在广泛范围中应用、解决一系列的实际问题,1809~1812年,他结合概率分布模型和中心极限思想来研究最小二乘法,首次为后来统计学中这项最常用的手段奠定了理论基础。

3. 明确了统计学的大数法则

拉普拉斯发现在观察天体运动现象中,当次数足够多时,能使个体的特征趋于消失,而呈现出某种同一现象,他指出这其中一定存在着某些原因,而非出于偶然。



4. 进行了大样本推断的尝试

在统计发展史上,他利用样本来推断总体的思想方法,为后人开创了一条抽样调查的新途径。

(二)高斯的主要贡献

对概率论与统计学的结合研究上作出贡献的另一位大数学家是德国的高斯(C. F. Gauss, 1777~1855),其主要贡献有:

1. 建立最小二乘法

在学生时代,高斯就开始了最小二乘法的研究,1798年完成最小二乘法的整个思考结构,1809年正式发表。

2. 发现高斯分布

高斯以他丰富的天文观察和在1821~1825年间土地测量的经验,发现观察值 x 与真正值 μ 的误差变异服从正态分布,并推导出测量误差的概率分布公式和提出误差分布曲线,也称为高斯分布曲线,即正态分布曲线。高斯所发现的一般误差概率分布曲线以及据此来测定天文观察误差的方法,不仅在理论上,而且在应用上都具有极其重要的意义。

二、近代描述统计学

近代描述统计学大致在19世纪中叶至20世纪上半叶期间形成,生物统计学派的创始人是英国的高尔顿(F. Galton, 1822~1911),主将是高尔顿的学生皮尔森(K. Pearson, 1857~1936)。

(一)高尔顿的主要贡献

1. 初创生物统计学

1901年,高尔顿及其学生皮尔森首次为他们所运用的统计方法论明确提出了生物统计(Biometry)一词。高尔顿认为,生物统计学是应用于生物学科中的现代统计方法,从高尔顿及后续者的研究实践来看,他们把生物统计学看作是一种应用统计学。

2. 对统计学的贡献

(1)关于“变异” 变异是进化论中的重要概念,1889年,高尔顿把总体的定量测定法引入遗传研究中,通过总体测量发现,对动物或植物的每一个种别都可以决定一个平均类型,在一个种别中,所有个体都围绕着这个平均类型,并把它当作轴心向多方面变异,这就是他在提出的“平均数离差法则”。

(2)关于“相关” 统计相关法是由高尔顿创造的,1877年他搜集大量人体身高数据后,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女,从而把父母高的其后代高个子比较多、父母矮的其后代高个子比较少这一特性认识具体化为父母与子女之间在身高方面的定量关系;1888年,高尔顿在“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,并提出了现在常用的相关系数的计算公式。

(3)关于“回归” 1870年,高尔顿在研究人类身高的遗传时发现,高个子父母的子女,其身高有低于他们父母身高的趋势;相反,矮个子父母的子女,其身高却往往有高于他们父母身高的趋势;从人口全局来看,高个子的人“回归”于一般人身高的期望值,而矮个子的人则作相反的“回归”。1886年,高尔顿在论文“在遗传的身高中向中等身高的回归”中,正式提出了“回归”概念。