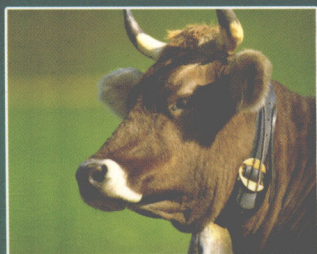


全国高等学校农林规划教材

Statistics for
Veterinary Science

兽医统计学

谢庄 贾青 主编



高等教育出版社
Higher Education Press

全国高等学校农林规划教材

Statistics for Veterinary Science

兽医统计学

谢 庄 贾 青 主编



高等教育出版社

内容提要

本书内容丰富,全书共 13 章,全面系统地介绍了兽医统计学的基本原理和方法;针对性强,除了介绍常用的几种基本统计方法以外,还介绍了兽医学科中需要使用的统计分析方法,如非参数检验、序贯分析、判别分析和半数致死量等;突出应用性和实践性,每一章都安排了大量的例题和复习思考题;内容新颖,大部分例题和复习思考题都引自最新的各种期刊,并介绍了 Excel 电子表格的统计学功能和统计软件 SPSS 的基本使用方法。

本书除作为兽医专业本科生的必修教材外,还可作为兽医科研工作者、兽医临床工作者、兽医教学工作者及兽医学科研究生的参考用书。

图书在版编目(CIP)数据

兽医统计学/谢庄,贾青主编. —北京:高等教育出版社,2005.8

ISBN 7-04-017365-4

I. 兽... II. ①谢...②贾... III. 家畜卫生 - 卫生统计 - 高等学校 - 教材 IV. S851.67

中国版本图书馆 CIP 数据核字(2005)第 062542 号

策划编辑	潘 超	责任编辑	舒敬江	封面设计	于文燕	责任绘图	黄建英
版式设计	范晓红	责任校对	朱惠芳	责任印制	韩 刚		

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100011
总 机 010-58581000

经 销 北京蓝色畅想图书发行有限公司
印 刷 北京鑫丰华彩印有限公司

开 本 787×1092 1/16
印 张 17.75
字 数 430 000

购书热线 010-58581118
免费咨询 800-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>

版 次 2005 年 8 月第 1 版
印 次 2005 年 8 月第 1 次印刷
定 价 20.70 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 17365-00

兽医统计学

主 编	谢 庄	贾 青	
副主编	张 勤	戴国俊	徐宁迎
	王钦德	周鸿飞	
主 审	盛志廉		
编 者	王钦德	李齐发	李学斌
	余 梅	陈 静	张 勤
	张廷荣	周鸿飞	赵宗胜
	贾 青	徐宁迎	景桂英
	谢 庄	戴国俊	

前 言

中国是个畜牧业大国。在中国畜牧业迅猛发展的过程中,兽医学科的保障作用是至关重要的。而随着兽医科学自身的发展,其科研、临床实践、教学等都需要有一个专门处理其发展过程中所产生的数据资料的这样一个有力工具,这个工具就是统计学。

随着科学事业的发展,统计学已大踏步地跨进了一个又一个学科领域,其原因是每个学科领域在其自身发展的过程中都会产生并还将产生大量的数据资料,而统计学就是为了处理这些数据资料而产生并发展着的一门学科。兽医学也是频频使用统计学、并越来越离不开它的一门学科。越来越多的兽医科研工作者、临床实践工作者、兽医教学工作者已越来越清醒地认识到了统计学的重要性。统计学能使人们从纷繁复杂的数据资料中发现事物的潜在规律。很难想象,自然科学发展到今天,如果没有统计学这一有力工具,我们将如何来面对和处理日益庞大的数据资料群,特别是和生物学有关的各类学科。

到目前为止,专门给兽医专业的学生所使用的统计学教材和参考书还很少。本书即是兽医专业的本科生在已学习并初步掌握了概率论和线性代数等基本知识的基础上,进一步系统学习统计学方法而编写的一本入门教材。本书并不想十分系统地全面介绍统计学的理论和概念,而是在一定的基本理论和基本概念的保障下,让同学们掌握统计学的基本方法,同时为同学们在今后的深造、工作实践和科学研究中继续深入学习和掌握更深的统计学原理和方法打下一个牢固的基础。

本书在体例和内容的安排上,力求坚持学以致用原则;在坚持科学性、系统性的前提下,突出应用性和实践性。因此,每一章都安排了大量的例题和习题,其目的就是希望同学们通过这些例题和习题的练习能掌握在各种场合下该如何使用每一种统计学方法。

本书除介绍了几种常用的基本统计方法(t 检验、 χ^2 检验、方差分析、相关与回归分析)外,还安排了几章介绍了兽医学科中需要使用的其他统计分析方法,如非参数检验、序贯分析、判别分析及半数致死量等。这些统计学方法对于临床资料的处理、药物研究、疾病诊断是很有用的。

在今天,学习统计学已离不开现代化的计算工具。许多实际资料的处理已不大可能单纯依靠手工计算和计算器,因而在学习统计学的同时,学习一个或几个统计学软件是必需的。故而本书末专门准备了一个附录,介绍 Excel 电子表格的统计学功能和专用的统计软件 SPSS。统计软件的使用可以大大地减轻我们的工作强度,提高运算速度和运算准确性。但如果以为只要统计软件而不需要学习统计学的基本概念和基本统计方法,那就大错而特错了。因为只有学习并掌握了统计学的基本知识,才能更好地使用统计软件,解读和理解统计软件处理数据资料后得出的结果,也才能对之进行更合理地解释,并更好地对科研结果、临床实践进行科学的指导。

本书既可作为兽医专业本科生的必修教材,也可作为兽医科研工作者、兽医临床工作者、兽

医教学工作者及兽医学科研究生的参考用书。

在本书的编写过程中,得到了许多关心本教材的人士的大力支持。大量的例题和习题除引自书后所列参考书外,还有许多来自各种杂志,或为了便于教学而改编自各类杂志,由于所引的参考量十分庞大,因此本书书末除必要的参考书目外,不再一一列出相关的参考文献,在此一并致以深深的谢意。

本书的编写,来自各位编写者的集体智慧。在编写过程中,召开了两次编写会,充分讨论了本书的编写意图、编写指导思想及体例和内容的安排,最后主编和部分副主编又对初稿进行了详细的审议,在主审盛志廉教授精心审稿的基础上最后定稿。对此,我们再一次对盛志廉教授 80 高龄却为我们认真审稿表示由衷的敬佩和衷心的感谢。即便如此,由于我们水平有限,所有编写者虽都常年工作在兽医统计学教学第一线,但都不是专业统计学研究者,难免存在错误、疏漏和不足,因此恳请读者和使用者在学习和使用本书的过程中不吝赐正。

谢 庄

2005 年 2 月于南京

郑 重 声 明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任，构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人给予严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话：(010) 58581897/58581896/58581879

传 真：(010) 82086060

E - mail：dd@hep.com.cn

通信地址：北京市西城区德外大街 4 号

高等教育出版社打击盗版办公室

邮 编：100011

购书请拨打电话：(010)58581118

目 录

第一章 绪论	1	第二节 完全随机设计资料的方差分析	65
第一节 兽医统计学的概念与功能	1	第三节 随机区组设计资料的方差分析	76
第二节 兽医统计学的特点	2	第四节 析因设计资料的方差分析	81
第三节 常用统计术语	3	第五节 系统分组设计资料的方差分析	86
复习思考题	5	复习思考题	91
第二章 资料的整理与基本分析	6	第六章 χ^2 检验	94
第一节 资料的采集与核对	6	第一节 χ^2 检验的原理	94
第二节 资料的整理	8	第二节 适合性检验	96
第三节 常用统计图表	12	第三节 独立性检验	101
第四节 平均数	16	复习思考题	105
第五节 变异数	20	第七章 非参数检验	107
第六节 正常值范围的确定	22	第一节 成对数据的显著性检验	107
复习思考题	24	第二节 两组资料的秩和检验	110
第三章 分布与统计推断	25	第三节 多个样本资料的秩和检验	110
第一节 正态分布	25	第四节 等级相关	115
第二节 二项分布与泊松分布	29	第五节 Ridit 分析	116
第三节 样本平均数的抽样分布	32	复习思考题	121
第四节 χ^2 分布、 t 分布和 F 分布	34	第八章 序贯检验	123
第五节 统计推断的意义与原理	36	第一节 单向检验	123
第六节 参数估计	44	第二节 双向检验	126
复习思考题	45	第三节 闭锁型序贯检验	130
第四章 差异显著性检验	47	复习思考题	133
第一节 单个平均数的假设检验	47	第九章 相关与回归分析	135
第二节 成组资料的两个平均数的假设 检验	48	第一节 简单相关	135
第三节 配对资料两平均值检验	52	第二节 直线回归	138
第四节 率的假设检验	54	第三节 曲线回归	143
复习思考题	59	第四节 多元线性回归分析	147
第五章 方差分析	62	复习思考题	154
第一节 基本假定和数据转换	63	第十章 判别分析	156
		第一节 基本概念	156

第二节 两类判别函数	157	附表	233
第三节 多类判别	163	附表 1 标准正态分布的分布函数表	233
复习思考题	167	附表 2 标准正态分布的双侧分位数表	235
第十一章 半数致死量	169	附表 3 χ^2 分布的右侧分位数表	236
第一节 基本概念	169	附表 4 t 分布的双侧分位数表	238
第二节 半数致死量的测定方法	170	附表 5 Dunnett t 表(单侧检验)	240
第三节 半数致死量的试验设计与应用	174	附表 6 F 分布的右侧分位数表	242
复习思考题	176	附表 7 q 值表(SNK 法)	247
第十二章 抽样调查	177	附表 8 Duncan's 多重极差检验的 5% 和 1% SSR 值表	248
第一节 抽样调查的基本概念和意义	177	附表 9 符号检验表	250
第二节 抽样方案的制订	178	附表 10 符号秩和检验表(双尾)	251
第三节 抽样调查的方法	180	附表 11 成组资料秩和检验表	252
第四节 抽样规模的确定	183	附表 12 Spearman 秩相关系数检验临界 值表	254
复习思考题	186	附表 13 计数型资料闭锁型序贯检验边界 及中界线坐标	255
第十三章 试验设计	187	附表 14 计量资料序贯检验闭锁线坐标 n' , y'_n 数值	256
第一节 试验设计的原则	188	附表 15 r 和 R 的 5% 和 1% 显著值	258
第二节 单因素试验设计	193	附表 16 百分率与概率单位对照表	260
第三节 多因素试验设计	203	附表 17 正态性 D 检验界值表	261
复习思考题	214	附表 18 随机数字表	263
附录 兽医统计学实验	217	附表 19 常用正交表	265
实验一 数据整理与基本分析	217	参考书目	272
实验二 差异显著性检验	223		
实验三 方差分析	225		
实验四 χ^2 检验	229		
实验五 回归与相关分析	230		

第一章 绪 论

本章主要介绍兽医统计学的概念、特点和基本内容,介绍总体、样本、变异、变数、变量、参数、统计量、系统误差、随机误差、准确度、精确度等兽医统计学中的常用术语。

第一节 兽医统计学的概念与功能

人类对生物现象的认识是通过对生物体复杂多样的特征性状的认识实现的。对个体而言,这些特征性状表现为描述性或可测量的具体数字,尤其是具有这些特征性状的群体表现有很大的随机性。如何处理、分析这些看似杂乱无章的数据资料,并发现其内在的规律,一般的数学方法显得力不从心,甚至无能为力。将概率论和数理统计学原理应用于生物现象中带有随机性的数量变化规律的研究,就形成了**生物统计学**(biostatistics)。兽医统计学是生物统计学的一个分支,具体说就是应用概率论和数理统计学的原理和方法研究如何用有效的方法收集、整理、分析兽医学科学研究与临床实践中产生的带有随机性的数据,对所研究的问题做出统计推断,提供决策依据的这样一门学科。

目前,生物统计学已广泛应用于生物学科各个领域,兽医学科也不例外。越来越多的兽医工作者已认识到生物统计学的重要性,越来越多的兽医科研工作者使用生物统计学的知识来设计科学试验,处理试验数据和调查结果,从而得出合理、客观、正确的结论。可以看出,生物统计学对兽医学科的科学研究、疾病防治、临床诊断正起着越来越重要的促进作用。

兽医统计学的基本功能大致有以下几方面内容:

对资料进行整理和描述。一般来说,从兽医实践和兽医科研中得来的原始资料往往都是杂乱无章的,不经过整理看不出其中的规律,也说明不了任何问题,因而有必要对这些资料进行科学、合理地整理,并用经过整理的资料来对我们的研究对象进行描述和说明。

用局部数据来推断和估计总体研究对象的特征。我们总希望针对全部研究对象进行科学研究,但在实践中这是不现实的,也是不可能的。因而我们只能抽取少部分有代表性的个体进行研究和观测,用由此得出的结论在一定的概率保证下来估计和推断全部研究对象的特征,从而得出带有普遍意义的一般规律。

通过显著性检验来鉴定试验效应。兽医科研和兽医实践中,一般都是比较性试验,即首先使可能影响试验的外部因素保持一致,然后把被研究因素根据试验要求划分成若干个等级(水平),根据设计要求进行试验,对各个水平的效应通过误差分析进行比较,用比较结果来得出具有一般意义的结论。

寻找因素间的相互关系。世间许多事物都不是孤立发生的,而是存在着某种平行或消长关

系,兽医学研究的对象也不例外.两个或多个变量间到底存在着何种关系,一个或一些变量对另一个或另一些变量起着什么样的作用,它们的变化规律如何,……,我们都可以对通过试验或从医学实践中得到的资料和数据进行分析研究,从而对这些变化规律进行定量地描述.

提供试验设计的一般原则.用尽可能少的人力、物力、财力和时间获取尽可能多的试验信息,并能精确地估计处理效应和试验误差,是每一个兽医科研工作者的愿望.因此,试验必须进行科学合理的设计,而科学合理的试验设计依赖于统计学为其提供科学的设计原则.这些原则保证了试验的合理性、科学性、公正性、客观性,从而使得由试验得到的数据是正确的,结论是可靠的,效果是可信的.

第二节 兽医统计学的特点

统计学(statistics)的推理思维与其他自然学科不同.在自然界中,我们可以总结出许多普遍规律,但也总能找到一些例外.例如,血是红的,这是普遍规律.但有人却发现了白色血液的鱼.按某些自然学科的推理方法,出现反例就应当否定原假设,即否定“血是红的”这一结论.用统计学的语言我们可以这样说,“至少99%的动物的血液是红的”,这就是事物的概率性.在自然界中,概率性是普遍存在的.概率性的特点是所做结论并不是100%正确的,而是在一定概率保证下是正确的.因此,概率性是统计学的第一个特点.兽医统计学也不例外,也以概率性为其第一特点.

任何一门独立的学科都有其自身的理论体系,兽医统计学也一样.但兽医统计学又必须同时面对大量来源于实践的数据资料.如果没有这些数据资料,兽医统计学就失去了其存在和发展的必要.收集、整理、分析来自兽医科研、临床、防治第一线的数据资料是兽医统计学的主要任务.因此,兽医统计学不是一门纯理论的学科,而是理论和实践并重、理论和实践密切结合的学科,这是兽医统计学的第二个特点.

理论上,我们总希望能获得并处理具有同一性质的所有资料,然而在实践中这往往是不可能的,我们总是只能获得其中具有代表性的一部分资料,对这一小部分资料进行分析和处理,从而得出一个结论,并用这一结论在一定的概率保证下进行统计推断.因此,统计学的第三个特点就是归纳性,即对部分资料进行整理分析,得出一个结论,在一定的概率保证下推断总体资料的带有普遍意义的规律,即从特殊推断一般,从局部推断总体;即用样本的数量特征值、数量关系和数量变化规律来推断总体相应的数量规律.但这一推断过程的前提是样本必须随机取得(即随机样本),且具有代表性.当然,统计学并不排斥演绎性.

一般来说,兽医统计学所要分析的资料来源于以下两个方面:科学试验与调查.这就涉及抽样和试验设计.正确地确定抽样方案,正确地对将要进行的试验进行科学设计是统计工作的基础.所谓试验设计(experimental design),就是指在试验工作进行之前,应用统计学原理,制订出合理的试验方案,如最适样本大小、最佳样本配置、正确的试验动物种类、试验整个过程的安排等等,以使我们可以使用最少的人力、物力、财力和时间,以获得尽可能多的、可靠的信息和资料进行统计分析,得到可信的科学结论.

从兽医科研实践中所得到的数据资料具有变异性、随机性和复杂性,而数据资料往往又是最能表明事物变化规律的证据.因此,学习兽医统计学的目的就是要使用统计学的原理和方法来定

量地处理和分析生物数据的这些变异性、不确定性和复杂性,从而得出最令人信服的结论,以阐明事物发展的规律。

兽医统计方法是兽医学科中的一个重要工具,它能帮助兽医工作者发现隐藏在纷繁复杂的表面现象下面的客观规律。因此,学习统计学首先要确立统计学的思维方式,要学会用统计学的思想来武装自己的头脑,用统计学的思考方式来观察世界;其次,在兽医科研、临床诊断、疾病防治等方面要用好用活统计学。除了学好统计学,掌握统计学的基本原理、计算公式、数学概念和含义,具有一定的电脑知识和操作技能外,还必须有扎实的兽医专业方面的知识及丰富的兽医实践经验;最后,用兽医统计学处理和分析每一批资料和数据,都必须有充分的生物学意义和兽医学意义,而所做的试验也必须有兽医学科的理论意义和实践意义,否则,计算结果再正确、再精确,也毫无实际意义。因此,兽医统计学的学习及统计学方法的应用不能孤立地、单独地进行,它必须紧密结合兽医学科实践,以取得具有指导意义的结果。

第三节 常用统计术语

一、总体和样本

所谓**总体** (population), 是指具有相同性质的所有观测值所组成的**集合** (set)。由于每一生物体具有许多个**性状** (trait), 因而相似的生物体所组成的集合, 如同一物种, 同一类群, 就不是兽医统计学意义上的总体, 只有相似生物体所具有的某一相同性状所表现出来的值的集合才能作为统计学中的总体。例如, 成年母牛的血压、仔猪的血糖等。总体可以是无限的, 也可以是有限的。无限的总体既有时间上的含义, 又有空间(地域)上的含义。当我们把某一总体限定于某一时间、某一地域时, 总体就成了有限的。如上述总体(成年母牛的血压)就是无限的, 但如果要研究 2005 年某一奶牛场中成年母牛的血压, 就成了有限总体了。总体还有虚、实之分。在试验某一新药时, 我们总假想这一部分被试动物就来自于已施这一新药的动物总体, 而这一总体其实在新药推广之前还不存在, 是一虚总体, 即只有当这一新药试验成功并加以推广后这一总体才存在, 才是实总体。总体往往是无限的, 即使是有限总体, 其量往往也很大, 因此在实际工作中不可能对总体中所有的观测值一一加以考察, 而只能对其中具有代表性的一小部分进行研究。为了能对总体有一个很好的了解和认识, 被研究的这一小部分观测值必须来自于这一总体, 并具有很好的代表性。这样的一批观测值的集合就称为**样本** (sample)。从总体中得到样本的过程称为**抽样** (sampling)。

一个样本内观测值即变量的个数, 称为**样本容量**, 用 n 表示(相对应的, 有限总体的大小用 N 表示), 根据样本容量的大小, 可以把样本分为大样本和小样本。大小样本之间实际上并没有严格的界限。习惯上, 可以用 $n = 30$ 作为大小样本的分界线, 但在很多情况下, 这一界限是不够合理的。

二、变异和变量

在实践中, 无论是总体还是样本, 无论是调查还是试验, 所得到的数值都是有差别的, 这种差别在统计学中称为统计数据的**变异** (variation)。例如, 同一年龄的成年母牛, 其血压、体重、体格

大小都不会相同. 这种具有变异性质的数量在生物统计学中就称为**变量**(variate), 变量在某一个体具体表现出来的数值又称为**变数**或**观测值**(observation value). 变量是和常量相对应的一个概念.

三、参数和统计量

用来描述总体特征的数值称为**参数**(parameter). 例如, 总体平均值反映了总体的集中程度和一般水平, 因此总体平均值就是参数. 参数用希腊字母表示. 相应的, 由样本观测值计算得到的描述样本特征的数值称为**统计量**(statistic). 例如, 样本平均值反应了样本变量的集中程度和一般水平, 因此样本平均数是统计量. 统计量用拉丁字母表示. 参数一般为常量, 由于总体很大, 且往往是无限的, 因此在绝大多数情况下, 参数很难直接由计算得到, 而只能通过样本的统计量来进行估计(estimation). 从同一总体中抽取不同的样本所计算得到的同一性质的统计量是不会相同的, 但这些统计量都可以用来估计相应的参数.

四、误差

在科学试验中, 除了对希望所要研究或讨论的某一个或几个试验条件人为地加以区别外, 其余外部及内部条件都应当保持一致, 以使试验所得到的结果符合真值. 然而, 在生物学科中, 人们几乎无法把非试验条件绝对地控制在同一水平上, 同时试验对象也是错综复杂的生物体, 因此, 很难使所得到的试验结果完全符合真值. 试验结果和真值之间的这种差异和偏离, 就是**误差**(error). 误差按其来源和性质可分为系统误差和随机误差.

系统误差(systematic error)是指由于某些特定的非试验条件所造成的使试验结果朝某一个方向发生有规律的偏移. 造成系统误差的原因有以下几种: 度量工具的不精确或未经校正, 试验仪器及其读数器发生偏差或未经校正, 外界试验条件发生了很大的变化, 观测时间及顺序的影响, 试验人员操作及观测时的偏爱和习惯, 试验动物分组时发生的偏差等. 这些因素都会使得试验结果有规律地偏离真值. 由于系统误差影响了试验的准确性, 因此应当在试验前就加以预防和克服. 一般来说, 系统误差是能被消除的.

随机误差(random error)是指由种种偶然因素引起的、无法加以预测和控制的无规律的偏差. 随机误差又称为偶然误差. 随机误差的大小、方向都无法确定. 例如, 不管试验条件控制得如何严格, 试验仪器多么精密, 观测手段多么完善, 所得到的试验结果总会发生大小不等、方向不定的偏差. 消除系统误差以后, 试验过程中主要的误差来源就是随机误差. 在不发生歧义的情况下, 随机误差简称为误差. 可以发现, 如果观测次数足够多的话, 随机误差有统计学上的意义. 每一次观测所产生的随机误差都是独立发生的, 且服从一定的规律. 通过各种手段可以把随机误差降到最低的程度, 但我们却无法消灭它. 实际上, 随机误差是进行假设检验的基础. 降低随机误差, 可以提高试验的精确性, 可以更好地区别误差效应和处理效应, 使得试验结果更准确, 对试验处理间的差异所做出的评定更准确、更可靠.

由于工作人员的粗心大意或不负责任(如仪器使用不当, 错读数据, 记录不准, 任意涂改, 凭空杜撰等)所产生的测定值与真值的偏差, 称为**错误**. 错误不是统计学的研究内容. 在试验和调查中, 错误应当、同时也可以得以避免.

五、准确度和精确度

准确度和精确度是和两类误差密切相关的。

准确度 (accuracy) 是指观测值与真值接近的程度。当发生系统误差时, 观测值都会有规律地向某一个方向偏离真值, 因而降低了试验的准确度。**精确度** (precise) 是指在同一处理条件下, 同一批观测值间相互接近的程度。当随机误差较大时, 数据较离散, 精确度较低。

相比之下, 准确度是比精确度更重要的一个概念。由于在很多情况下, 准确度和精确度两者往往不可兼得, 因此在制订试验设计或做试验时, 应当很好地加以权衡。原则上, 可以适当放弃一些精确度以保证足够的准确度, 即首先应当将系统误差降至为零或降至最小程度, 或将系统误差化为随机误差, 以保证有足够的准确度。

复习思考题

1. 生物学数据的本质特性是什么?
2. 兽医统计学的概念和特点是什么?
3. 试举例说明总体和样本、参数和统计量的概念及相互关系。
4. 什么是误差? 如何控制误差? 误差与准确度、精确度的关系如何?

第二章 资料的整理与基本分析

本章主要介绍数据资料的分类、采集、检查与核对;不同类型资料的整理与分组;常用统计图、表;资料的平均数、方差、标准差和变异系数的计算。

采集资料是进行科学研究的一个重要内容,也是进行统计分析工作的第一步和全部统计推断工作的基础。如果不能正确地收集原始资料,无论分析方法多么正确,都难以获得反映事物本质规律的正确结论。

原始资料进行检查核对后,仍然是一堆杂乱无序的数据,往往不能立即从中看出问题。在对它们进行分析之前,必须加以整理和分组。整理首先是按资料的类型、性质或时间不同等进行分类,将性质相同的资料归纳到一起,使资料系统化,这样才能反映事物的本质。因此在资料整理时必须坚持“同质”原则,才能显示资料内部的规律性,得出正确的结论。资料经过整理以后,根据观测值的多少确定是否分组。观测值不多时不必分组,直接进行统计分析,或按数值从小到大(或从大到小)进行排列,以看出资料的变化情况。当观测值较多时,宜将观测值分成若干组,以便进行统计分析。将观测值分组后,制成次数分布表,即可看出资料的集中和变异情况。不同类型的资料其整理的方法略有不同。

第一节 资料的采集与核对

一、资料的种类

不同性质的资料其统计分析方法也不同。按性质的不同,资料一般可以分为两大类。

(一) 连续性资料

连续性资料(continuous data),是指对每个观测单位使用仪器或试剂来测定其某项指标的数值大小而得到的资料。其数值特点是各个观测值不一定是整数,两个相邻的整数间可以有带小数的任何数值出现,其小数的位数随测量仪器或工具的精确性而变化,它们之间的变异是连续性的,因而称为连续性资料。常见的连续性资料有动物的各种生理、生化指标,药动力学指标等,如血液中血红蛋白含量。计量资料一般都为连续性资料。

(二) 间断性资料

间断性资料(discrete data),是指在一定范围内只取有限种可能值的数据资料。间断性资料又可进一步分为计数资料和分类资料两种。

计数资料(counting data),是指用计数方式得到的数据资料。在兽医学研究和工作中有些指

标不能用仪器或工具直接测量,而只能用分类计数的方式获得.这类数据资料全都是整数,各观测值之间是不连续的.常见的资料有畜禽疫病感染个数、发病数、死亡数、呼吸次数、单位容积内细胞数、细菌数等.

分类资料(categorical data),是指可自然或人为地分为两个或多个不同类别的资料.有些只能观察到而不能直接测量的性状指标,如充血程度、精神状态、生死、发病类型等.这种资料可以转化为数值表示,如药物治疗畜禽某种疾病的疗效分为治愈、显效、好转和无效,可用1、2、3、4分别表示.分类资料各类别用数字表示时,不同数字仅代表不同的类别,不代表大小.要获得这类性状的数据资料,须对观察结果按性状的类别统计次数,以获得性状指标的数据.如中毒的潜伏期按不同的时间点分别计数,用药物治疗畜禽某种疾病的疗效分为治愈、显效、好转和无效等4种情况分别计数等,其数据资料既有程度差别,又有量的不同,兽医学上把这类资料也称为次数资料,或称半定量资料.

在兽医学中,分类资料常用相对数(率)或构成比的形式表示.

率是一种频率指标,是在一定条件下,某种现象实际发生的次数在总次数中的比例,用以说明该类现象发生的频率或强度.率以百、千等为基数,称为百分率、千分率,以百分率(percentage)为常见.百分率的计算公式为:

$$\text{百分率} = \frac{\text{某一现象实际发生的次数}}{\text{总次数}} \times 100\%. \quad (2-1)$$

在兽医临床实践中常见的率有:发病率、死亡率、感染率、治愈率、免疫率、阳性率等.如:

$$\text{发病率} = \frac{\text{一定时期内发病的家畜数}}{\text{同一时期同类家畜饲养总数}} \times 100\%.$$

构成比是某类事物可以分成两个以上不同的部分,每一部分在总的数据中所占的比率.例如调查一个鸡场鸡死亡的原因,我们可以发现,鸡死亡的原因很多,每一类原因引起的鸡死亡的比例是不同的.假设有500羽鸡死亡,其中由于鸡白痢死亡的有160羽、鸡法氏囊病死亡的有87羽、啄肛死亡的有158羽、由于拥挤死亡的有38羽,其余原因死亡的有57羽,则各种死亡原因引起的死亡率分别为32%、17.4%、31.6%、7.6%、11.4%,这就是构成比.

可以这样理解,在分类资料中,如果强调各种分类中某一类在总类别中的比率,通常称为百分率,而各种类别在总类别中的比率常叫做构成比.例如上面所提及的各类死亡原因中,如果我们仅统计该鸡场的病死率,假设该鸡场共饲养10 000羽鸡,总的死亡率应为 $p = \frac{500}{10\,000} = 5.0\%$,

而病死率则显然为 $p = \frac{160 + 87}{10\,000} = 2.47\%$.这里,总死亡率和病死率均为率,病死率在总死亡率中所占的比重就是构成比.

二、资料的采集与核对

(一) 资料的采集

在搜集资料之前,必须根据研究的目的认真考虑所需资料的来源、要求、项目、内容及搜集的方法,尽可能用最经济的人力和物力采集到所需资料.一般通过两个途径采集数据资料,一是通过调查搜集资料,二是通过科学试验获得资料.

1. 调查

根据调查方法、对象、目的的不同,调查又可分为:

(1) 对历史资料的调查. 如查阅各种兽医卫生工作情况的年报表、畜禽疫情月报表等. 对以往历史资料的调查分析, 可以获得兽医工作的基本情况和疫情规律, 制订兽医工作计划和措施, 指导兽医卫生防疫的生产实践和科学研究.

(2) 临床兽医工作记录的调查. 如临床化验报告、门诊病例等, 也是研究畜禽疫病病情和疗效的基本资料. 通过对这些资料的研究, 可以分析疫病发生的规律及治疗措施的有效性. 为使这些资料具有科学研究价值, 平时要认真填写, 并妥善保存.

(3) 现场专题调查. 一是普查, 二是抽样调查. 普查是一种全面调查, 即对研究对象的全部个体进行观测, 如牛结核病、乳房炎、布氏杆菌病等的普查. 普查能全面了解群体的基本情况, 甚为重要. 抽样调查是指根据预定的抽样方法从研究总体中抽取一部分个体作为样本进行观测, 对观测结果进行统计分析, 并对总体的情况做出估计. 抽样调查如使用得当, 能在较少的人力和物力条件下, 对总体做出相当准确的推断.

2. 试验

在兽医实践和科学研究工作中, 很多问题的解决必须通过试验来完成. 例如药物试验、临床疗效试验、畜禽生理生化指标测定及病毒的致病机理研究等. 一般试验研究都是有计划地通过随机样本进行观测, 所得到的资料都是随机样本的结果.

(二) 资料的检查与核对

检查与核对原始资料的目的在于确保原始资料的完整性和正确性. 所谓完整性, 是指原始资料没有遗缺或重复; 所谓正确性, 是指原始资料的测量和记录无差错, 没有不合理的合并. 检查中尤其要注意特别大、特别小和异常的数据, 要结合专业知识做出判断. 对于有重复、异常或遗缺的资料, 经核实确认予以删除; 对于有错误、相互矛盾的资料经检查核对后进行更正, 必要时进行复查或重新试验. 资料的检查与核对工作虽然简单, 但在统计分析工作中却是非常重要的一项工作, 因为只有完整、正确的数据资料, 才能真实地反映出调查或试验的客观规律, 经过统计分析得出正确的结论.

第二节 资料的整理

一、间断性资料的整理

间断性资料的整理常采用单项式分组法. 它的特点是用样本的观测值直接进行分组, 每组均用一个观测值表示. 分组时, 将资料中的每个观测值归入相应的组内, 然后画线记数, 制成次数分布表.

【例 2-1】 200 只鸡新城疫血球凝集抑制滴度检测资料, 按不同凝集抑制滴度分组整理, 制成次数分布表, 如表 2-1.

从表中可以看出鸡的新城疫血球凝集抑制滴度在 $1:10 \sim 1:640$ 范围内变动, 有 7 个不同的观测值, 主要集中在 $1:40 \sim 1:160$ 范围内, 以 $1:80$ 为最多.

有些计数资料观测值较多, 变异范围较大, 若以每一个观测值为一组, 则组数太多, 而每组内包含的观测值又太少, 资料的规律性显示不出来. 对于这样的资料, 可扩大为将几个相邻观测值