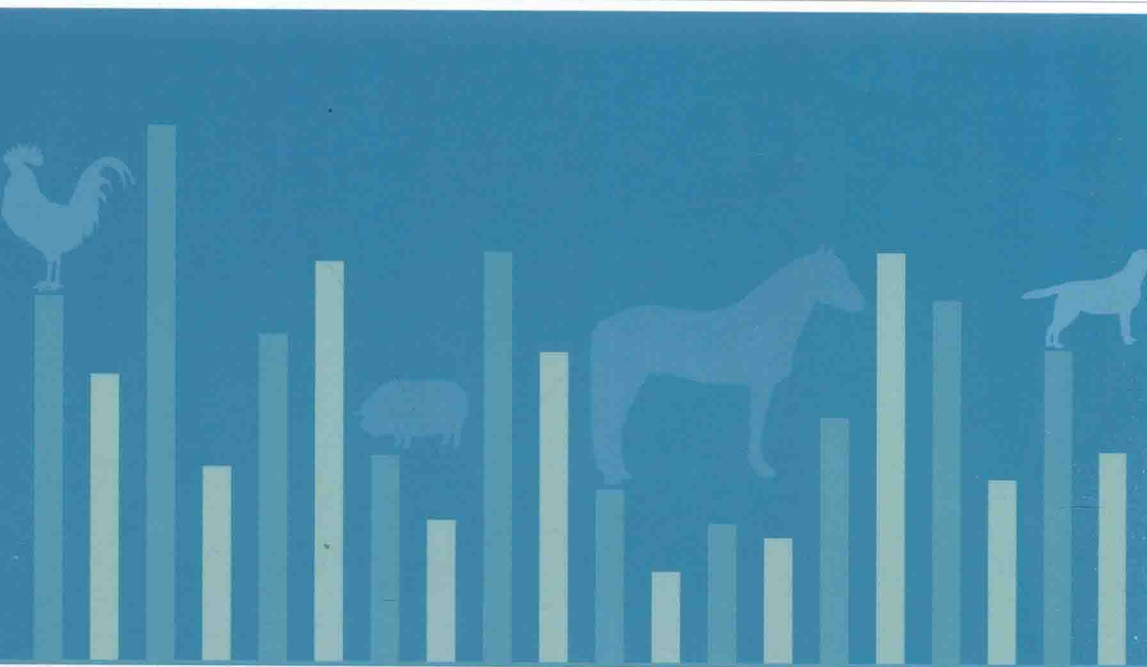


兽医统计学

SHOUYI TONGJI XUE

孙向东 王幼明 主编

宇传华 黄保续 主审



兽医统计学

孙向东 王幼明 主编
宇传华 黄保续 主审

中国农业出版社

图书在版编目 (CIP) 数据

兽医统计学 / 孙向东, 王幼明主编. —北京: 中国农业出版社, 2015.7

ISBN 978-7-109-20694-6

I. ①兽… II. ①孙…②王… III. ①家畜卫生-卫生统计 IV. ①S851.67

中国版本图书馆CIP数据核字 (2015) 第163892号

中国农业出版社出版

(北京市朝阳区麦子店街18号楼)

(邮政编码100125)

责任编辑 邱利伟 神翠翠

文字编辑 赵淑新

北京中兴印刷有限公司印刷 新华书店北京发行所发行

2016年3月第1版 2016年3月北京第1次印刷

开本: 720mm × 960mm 1/16 印张: 22

字数: 360千字

定价: 65.00元

(凡本版图书出现印刷、装订错误, 请向出版社发行部调换)

编 委 会

主 编：孙向东 王幼明

主 审：宇传华 黄保续

参编人员（以姓名笔画为序）：

于丽萍	王震坤	韦欣捷	刘爱玲	刘丽荣	李 妍	李金花	李 印
沈朝建	杨吉星	张 毅	贾智宁	洪希成	姚雪梅	赵 雯	徐全刚
康京丽	崔芳芳	章亚男	曾倩鲍	俊 哲	廖 祺	潘 涛	

前 言

本书是对应用于兽医实验室数据分析和兽医流行病学中统计学方法的概述。统计学广泛应用于多个领域，是进行数据分析的重要工具。然而，我国当前兽医工作者，特别是工作在动物疫病防控领域的学者，对于统计学方法的运用水平还有待提高。为了提高工作在动物疫病防控第一线的广大技术人员统计学运用水平，根据农业部兽医局领导的安排，针对目前统计学应用的现状编写本书。

本书的目标读者群是工作在动物疫病检测实验室的技术人员和进行流行病学调查与分析的工作人员。撰写本书前，编者进行了需求调研，在内容的设计上，充分考虑了当前兽医工作者对于统计学工具和方法的实际需求。对难度较大的和不常用的统计学方法，本书并未进行阐述，读者如果希望进行更深入的学习和研究，可以参照其他统计学书籍。

为了便于读者学习相关统计学知识，编者通过实际例子引出问题，并逐步展开阐述，将原理性内容和有关公式放在后面。这么做的目的，一是希望能够使读者以实际问题为导向进行理解和学习；二是希望能够由浅入深，引导读者逐步领会统计学方法；三是希望读者不要刚开始阅读本书就产生为难情绪，能够坚持将本书读完。

本书共包含18章内容，包括统计学术语总体介绍（第一章）、基本统计学方法（第二章到第九章）和高级统计学方法（第十章到第十五章）、流行病学调查研究设计技术（第十六章、第十七章）和诊断试验评价（第十八章）。本书中所用的例子，除个别是为了说明问题而人为设计数据外，绝大部分都是实际工作中的实例。坚持用实际例子说明问题是编者设计本书的初衷。

本书在编写过程中得到很多人的支持。在需求调研阶段和初稿完成后很多人提出了宝贵意见，特别是陈继明博士和韩雪博士提出了很好的意见，再次表示诚挚的感谢！

本书的编写由孙向东和王幼明总体设计，字传华和黄保续总体审校。由于作者水平有限，书中难免有不妥之处，敬请读者给予批评指正。

孙向东

2015年5月



前 言

第一章 Chapter 1

绪论	1
第一节 兽医统计学与学习要求	3
第二节 兽医统计学的常用术语	4
第三节 兽医统计学的基本步骤	9
第四节 统计分析方法概述	12

第二章 Chapter 2

资料的统计描述	17
第一节 资料的采集	19
第二节 定量资料描述	30
第三节 定性资料描述	34
第四节 参考值范围的确定	41

第三章 Chapter 3

统计表与统计图	45
第一节 常用统计表	48
第二节 常用统计图	53

第四章 Chapter 4

变量的分布及应用	61
第一节 正态分布	63
第二节 二项分布	66
第三节 Poisson分布	70
第四节 负二项分布	73

第五章 Chapter 5

总体均数的估计及假设检验 77

第一节	抽样误差与标准误	79
第二节	t 分布	80
第三节	总体均数的估计	82
第四节	假设检验概述	86

第六章 Chapter 6

t 检验 93

第一节	t 检验的应用条件	95
第二节	单样本 t 检验	95
第三节	两独立样本资料的 t 检验	96
第四节	配对资料的 t 检验	99

第七章 Chapter 7

方差分析 103

第一节	方差分析的基本思想及应用条件	105
第二节	两样本方差齐性的 F 检验	108
第三节	完全随机设计资料的方差分析	109
第四节	随机区组设计资料的方差分析	110
第五节	析因设计资料的方差分析	113
第六节	重复测量资料的方差分析	116
第七节	多样本均数间的比较	119
第八节	多样本方差比较	120

第八章 Chapter 8

卡方检验 123

第一节	卡方检验的原理	125
第二节	四格表资料的卡方检验	127
第三节	$R \times C$ 列联表的卡方检验	134
第四节	Fisher确切概率检验	137

第九章
Chapter 9

非参数检验 139

- 第一节 配对样本的Wilcoxon符号秩检验 141
- 第二节 两个独立样本的Wilcoxon秩和检验 144
- 第三节 完全随机设计多个样本比较的
Kruskal-Wallis H检验 147

第十章
Chapter 10

简单相关与回归 151

- 第一节 直线相关 153
- 第二节 直线回归 157
- 第三节 Spearman等级相关 166
- 第四节 曲线回归 170

第十一章
Chapter 11

多重回归 177

- 第一节 多重线性回归 179
- 第二节 自变量的选择 183
- 第三节 多重线性回归的应用与注意事项 185
- 第四节 Logistic回归 190
- 第五节 Logistic回归的应用与注意事项 196

第十二章
Chapter 12

判别分析 199

- 第一节 判别分析的基本思想 201
- 第二节 Fisher判别 203
- 第三节 Bayes判别 207
- 第四节 逐步判别 210
- 第五节 判别分析中应注意的问题 213

第十三章 Chapter 13

聚类分析	215
第一节 聚类分析基本思想	216
第二节 相似系数	217
第三节 系统聚类	219
第四节 注意事项	232

第十四章 Chapter 14

主成分与因子分析	233
第一节 主成分分析	235
第二节 主成分分析的数学模型	236
第三节 主成分分析的求法及性质	236
第四节 主成分分析的应用	241
第五节 因子分析的基本思想	241
第六节 因子分析数学模型	242
第七节 因子分析的性质	243
第八节 因子载荷的求解	244
第九节 因子旋转	248
第十节 因子分析注意事项	250

第十五章 Chapter 15

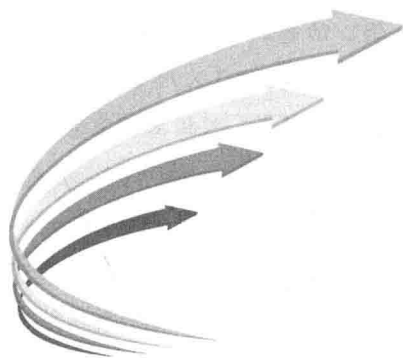
时间序列分析	253
第一节 时间序列分析概述	255
第二节 指数平滑法	256
第三节 ARIMA模型	261

第十六章 Chapter 16

抽样调查	273
第一节 抽样调查概述	275
第二节 抽样方法	277
第三节 样本量估算	284
第四节 案例分析和示范	288

第十七章 Chapter 17	研究设计	297
	第一节 观察性研究设计	298
	第二节 实验研究设计	307
第十八章 Chapter 18	诊断试验评价与ROC分析	315
	第一节 常用的诊断试验评价指标	316
	第二节 ROC曲线	326
名词术语		332
参考文献		335

第一章 绪论



【例1-1】

为研究中药对奶牛乳房炎的治疗作用，从中随机抽取65头临床型乳房炎患牛进行实验。将抽到的牛随机分为2组，试验组奶牛用中药复方治疗，对照组用青霉素和链霉素治疗，观察临床效果。结果中药复方治疗患牛32例，治愈27例，治愈率为84.3%；对照组治疗33例，治愈26例，治愈率为78.8%。据此认为用中药治疗奶牛乳房炎的疗效高于西药；将相关资料及结论撰写成文，投稿至某杂志编辑部，但被拒，理由是该文未经统计学处理。

【问题1-1】

- (1) 上表中的资料属于何种类型？
- (2) 该试验属于何种设计方案？
- (3) 为何杂志编辑部编辑要求用统计学处理该资料？
- (4) 该资料需要用何种统计学方法处理？

【分析】

- (1) 研究的结果分为治愈和未治愈两种，属于二分类定性资料；
- (2) 实验采用随机抽样和随机分组，属于完全随机实验设计；
- (3) 研究总体是所有临床型乳房炎患牛，而研究对象是65头临床型乳房炎患牛，使研究结果不可避免地存在抽样误差，故需要作统计学处理；
- (4) 根据资料的类型及其设计方案，应采用四格表 χ^2 检验。

该资料在统计描述的基础上，运用SPSS软件进行四格表 χ^2 检验。结果 $p>0.05$ ，差异无统计学意义，说明用中药复方治疗奶牛乳房炎与常用抗生素的治疗效果基本相同。虽然试验结果显示差异无统计学意义，但是由于中药成本较低、副作用较小且使奶牛乳汁无药物残留等优点，其仍具有很大的推广价值；经统计分析后的论文很快就被录用。如果没有统计学的帮忙，不仅会使精心撰写的论文失去发表、交流的机会，还可能导致辛苦的研究工作得不到相应的价值体现。

第一节 兽医统计学与学习要求

一、兽医统计学

统计学 (statistics) 是在相对有限的样本数据上, 对特定的随机现象作推断的学科。统计学的范围可分为两个主要的领域: 数理统计学和应用统计学。数理统计学更关注统计推断中新方法的发展, 同时需要运用多种数学工具。应用统计学则把数理统计学方法应用到特定的领域, 比如经济学、心理学、公共卫生及动物卫生等。

兽医学 (veterinary medicine) 是一门研究预防和治疗畜禽疫病的科学。家畜、家禽、伴随动物 (如犬、猫等)、经济野生动物、实验动物、观赏动物、经济昆虫 (如蜜蜂、蚕等) 和鱼类的保健和疫病防治工作均属兽医学范畴。

通过上述统计学和兽医学的概念, 可界定出兽医统计学的基本定义: 兽医统计学 (veterinary statistics) 作为应用统计学的分支, 是一门运用统计学的原理和方法, 进行兽医学科学研究与临床实践中有关数据的搜集、整理、分析和推断的应用科学。它通过对偶然性 (不确定性) 剖析、研究, 发现事物的必然性 (确定性), 并用以指导兽医学的理论和实践。

兽医统计学作为一门科学, 不论从执行统计方法的角度, 还是进行科学研究的角度, 必须如实地反映现状。科学研究的目的在于得到真实、可靠的结果, 这也是兽医统计学的求实性。兽医统计学必须在统计理论的指导下, 正确运用统计学思维, 针对数据特点, 巧妙地选用恰当、高效的统计分析方法, 从而得到可靠的结果和科学的结论。

二、学习要求

统计学方法很多，对每一种方法的学习要求可以概括为如下三个要点：

（1）这个统计学方法是什么？即需要掌握该法的基本思想、原理和有关概念；

（2）这个统计学方法是怎么来的？即需要初步了解该法的具体证明方法和推导过程；

（3）这个统计学方法能够用来干什么？即需要掌握该法的相关用途和适用条件。

数学基础较为薄弱的基层兽医工作者不大可能也没有必要对每一种统计学方法的证明推导、来龙去脉都一一深入学习。所以学习重点应放在1和3上。对于绝大多数学习兽医统计学的有关人员，在经过一定的统计学培训后，能知晓统计学的术语、懂得运用各类统计学方法为自身服务，并能看懂、理解和正确解释统计学的结果，那就达到了本书的学习要求。

第二节 兽医统计学的常用术语

一、总体与样本

任何统计研究都必须先确定**观察单位**（observed unit），亦称**研究单位**（study unit）。观察单位是统计研究中最基本的单位，可以是一个人，一头动物，也可以是某一特定的动物群体；还可以是一个器官、一个细胞、一个采样点等。

总体（population）是根据研究目的而确定的同质观察单位的全体，也就

是性质相同的研究对象中所有观察单位某种变量值的集合。例如研究某奶牛场2013年成年母牛的血压值,则研究对象是该奶牛场该年的所有成年母牛,观察单位是母牛个体,变量是血压,变量值是测得的血压值,该奶牛场2013年全部成年母牛的血压值构成一个总体。它的同质基础是同一地区,同一年份,同为成年母牛。这里的总体只包括有限个观察单位,称为**有限总体**。有时总体是设想的,如研究某新型抗生素对奶牛乳房炎的治疗效果,这里总体同质基础是同为患乳房炎的病牛,同用某新型抗生素治疗,包括设想用该抗生素治疗的所有患乳房炎的病牛,其观察单位数显然是不确定的、无限的,称为**无限总体**。

正常情形下,为节省人力、物力、财力和时间,许多医学研究都采取从总体中抽取样本 (sample),根据样本信息来推断总体特征的方法,即运用抽样研究 (sampling research) 的方法来实现,这种从总体中抽取部分观察单位的过程称为抽样 (sampling)。为保证样本的代表性,抽样时,必须遵循随机 (randomization) 原则,因而从总体中随机抽得的部分观察单位,其实测值的集合,就称为**样本 (sample)**。该样本中所包含的观察单位数称为该样本的**样本量 (sample sizes)**。如上例,可从某奶牛场2013年成年母牛中,随机抽取100头成年母牛,逐个进行血压测量,得到100头成年母牛的血压测量值,组成样本;也可从就诊的患乳房炎的病牛中,随机抽取100名头病牛,并观察该新型抗生素治疗前和治疗一段时期后的病情变化。例如,实验室诊断的体细胞计数变化,组成反映治疗结果的样本。应当强调,获取样本仅仅是手段,而通过样本信息来推断总体特征才是研究的目的。

二、变量与资料

确定总体之后,应对每个观察单位的某项特征进行测量和观察,这种特征,能表现观察单位的变异性,称为**变量 (variable)**。对变量的测得值称为**变量值 (value of variable)**或**观察值 (observed value)**,由变量值构成**资料 (data)**。例如,以人为观察单位调查某牧区某年7岁正常儿童的生长发育状况,性别、身高、体重等都可视为变量,性别有男有女,身高可高可矮,体重可轻可重,不

同个体不尽相同,这种个体间的差异称为变异。这些变异来源于一些已知的或未知的,甚至是某些不可控制的因素所导致的随机误差。变量的观察结果可以是定量的,例如身高(厘米);也可以是定性的,例如性别(男或女)。按变量的值是定量的、还是定性的,可将变量分为以下几种类型:

1. 定量资料(quantitative data) 又称定量资料(measurement data)或数值变量资料(numerical variable)。是为测定每个观察单位某项指标的大小而获得的资料。其变量值是定量的,表现为数值大小,一般有度量衡单位。如调查某养殖场某年成年肉牛的生长发育状况,以个体为观察单位,肉牛的胸围(cm)、体重(kg)、血压(kPa)等资料均属此类资料。

2. 定性资料(qualitative data) 又称计数资料或无序分类资料。为将观察单位按某种属性或类别分组计数,分组汇总各组观察单位数后而得到的资料。其变量值是定性的,表现为互不相容的属性或类别,如试验结果的阳性、阴性,某症状的有、无等。分两种情形:

(1) 二分类:如检查某养殖场猪仔粪便中的蛔虫,以每只猪仔为观察单位,结果可报告为猪蛔虫卵阴性或阳性两类;如观察某药治疗某病畜的疗效,以每头病畜为观察单位,结果可归纳为治愈与未愈两类。两类间相互对立,互不相容。

(2) 多分类:如口蹄疫病毒目前有O、A、C、SAT1、SAT2、SAT3(即南非1、2、3型)和Asia1(亚洲1型)7个血清型。

3. 等级资料(ranked data) 又称半定量资料或有序分类资料。为将观察单位按某种属性的不同程度分成等级后分组计数,分类汇总各组观察单位数后而得到的资料。其变量值具有半定量性质,表现为等级大小或属性程度。如观察禽类某血清反应,以每只禽类为观察单位,根据反应强度,结果可分-、±、+、++、+++、++++六级;又如观察用某药治疗某病畜的疗效,以每头病畜为观察单位,结果可分为治愈、显效、好转、无效四级等。

统计分析方法的选用与资料类型密切联系。在资料分析过程中,在有关专业理论指导下,各类资料间可以互相转化,以满足不同统计分析方法的要求。例如,以人为观察单位观察某人群脉搏数(次/min),属定量资料;若根据医学专业理论,定义脉搏数在60~100次/min为正常,<60次/min或>100次/min为异

常,按“正常”与“异常”两种属性进行分类,此时资料就转化为定性资料;若进一步定义脉搏数 <60 次/min为缓脉, >100 次/min为速脉,按“缓脉”、“正常”与“速脉”三个等级进行分类,此时的资料就是等级资料。以上的例子是先获取定量资料,再将定量资料转化为定性资料或等级资料。在专业理论的支持下,确定不同属性或不同等级的数量界限,这种转化是不难实现的。在研究设计中,对于能测量的指标,应尽可能设计为定量指标,这将为分析中的资料转化带来了方便。相反,定性资料是不能逆转化成等级资料或定量资料的。所以我们在收集资料时能定量应尽可能定量,定量资料所含有的信息量大于等级资料,等级资料所含有的信息量大于定性资料。

三、误差

误差(error)泛指实测值与真值之差。按其产生原因和性质可粗分为随机误差(random error)与非随机误差(nonrandom error)两大类,后者又可分为系统误差(systematic error)与非系统误差(nonsystematic error)两类。

1. 随机误差 是一类由多种无法控制的因素引起的、不恒定的、随机变化的误差。例如,在实验过程中,在同一条件下对同一对象反复进行测量,虽极力控制或消除系统误差后,每次测量结果仍会出现一些随机变化,即出现了随机测量误差(random error of measurement)。在抽样过程中由于抽样的偶然性而出现的抽样误差(sampling error)也是随机误差。

随机误差是不可避免的,在大量重复测量中,它可出现或大或小、或正或负,呈一定规律性的变化。由于造成随机误差的影响因素太多太复杂,其具体规律难以掌握,但随机误差呈正态分布,可用统计学的方法进行分析。

2. 系统误差 是实验过程中产生的误差,它的值恒定不变,或遵循一定的变化规律,其产生原因往往是可知的或可能掌握的。这种误差一是可能来自于受试者抽样不均匀,分配不随机;二是可能来自于不同实验者个人感觉或操作上的差异;三是可能来自于不标准的仪器;四是可能来自于外环境非实验因素的不平衡等。在实际工作中应尽可能设法预见到各种系统误差的具体来源,力