



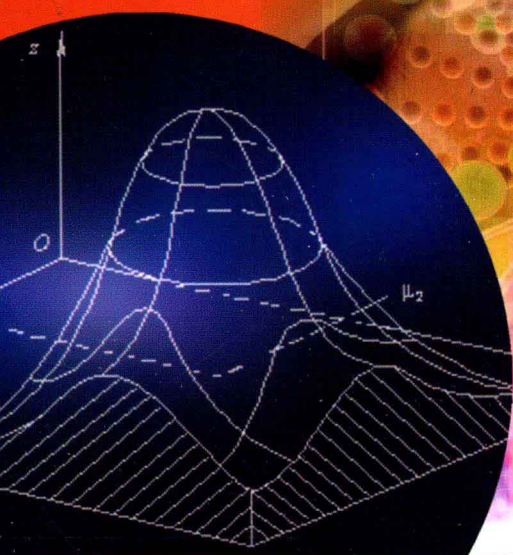
中国科学院教材建设专家委员会规划教材

全国高等医药院校规划教材

医药应用概率统计

第2版

高祖新 韩可勤 主编



科学出版社

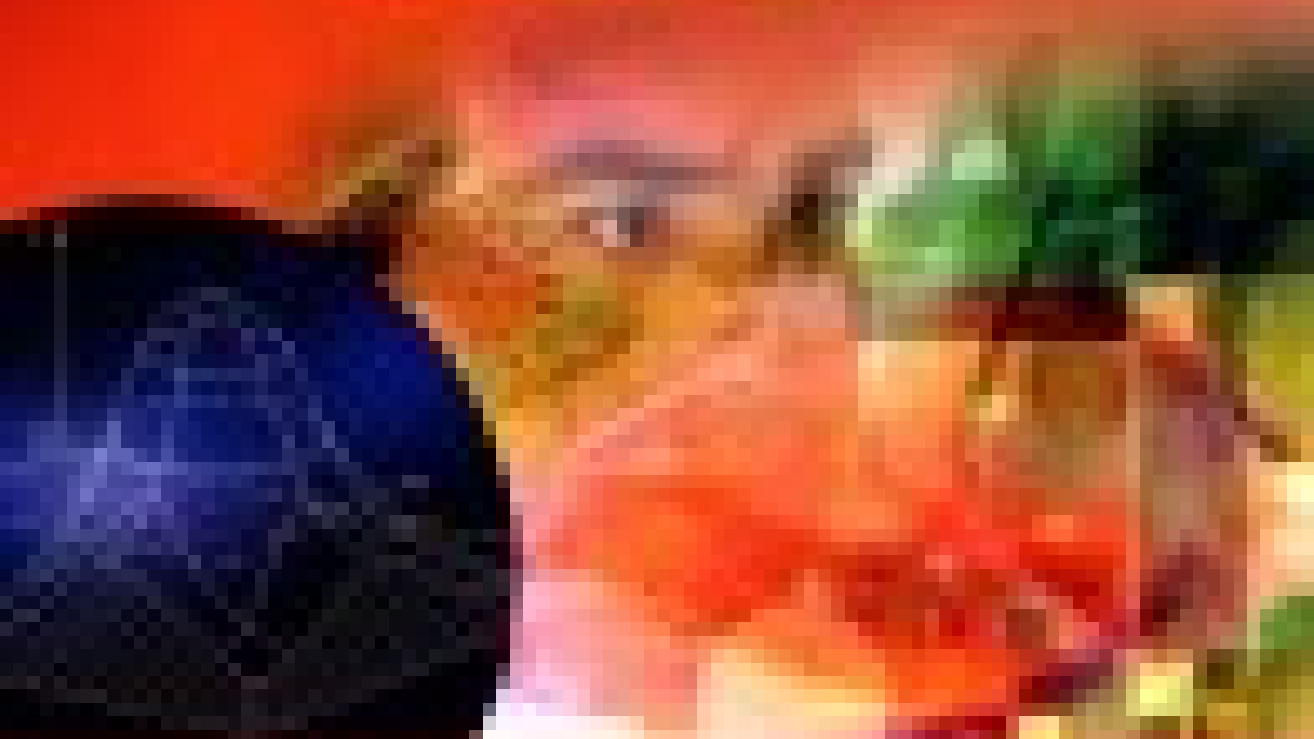


中华人民共和国国家卫生健康委员会
全国高等医药院校规划教材

医药应用概率统计

第三版

主编 陈永祥 副主编 王 浩



人民卫生出版社

中国科学院教材建设专家委员会规划教材
全国高等医药院校规划教材

医药应用概率统计

第2版

主 编 高祖新 韩可勤
副主编 言方荣
编 者 (按姓氏笔画排序)
王 菲 李雪玲 言方荣
周 森 徐 畅 高祖新
盛海林 韩可勤

科学出版社

北京

· 版权所有 侵权必究 ·

举报电话:010-64030229;010-64034315;13501151303(打假办)

内 容 简 介

本书是编者在所进行的“统计及应用课程的教改研究和实践”研究成果荣获国家级教学成果二等奖的基础上所编写的,是经过全面修订完善的颇具特色的医药统计课程教材,内容更加切合教学实际,结构体系也更为合理完善。

教材全面介绍了数据的统计概括处理与图表呈现,简明系统的概率论理论基础、数理统计的基本概念和基本知识,常用统计推断和统计分析方法,试验设计的统计分析,统计软件(Excel 数据分析模块)的操作应用指导等内容,并已包含全国硕士研究生入学统一考试数学考试大纲所要求的概率统计知识。各章正文以医药应用案例贯穿全程,内容系统全面,阐述深入浅出,例题经典实用;并配有兼具基础性和技巧性解题指导的综合例题(含考研真题)、具体翔实的统计软件操作指导、高度精练的各章内容提要简表、相关统计历史和大师简介的知识链接、题型多样丰富的思考与练习和习题、统计软件应用的上机实训题及参考答案等,从而方便教师教学,全面提升学生医药统计应用的学习和实践创新能力。

本书可作为医药类各专业学生学习医药统计或数理统计等基础课程的教材或教学参考书,也可供各类专业人员特别是医药卫生工作者学习参考。

图书在版编目(CIP)数据

医药应用概率统计 / 高祖新, 韩可勤主编. — 2 版. — 北京: 科学出版社, 2013. 1

中国科学院教材建设专家委员会规划教材·全国高等医药院校规划教材
ISBN 978-7-03-036551-4

I. 医… II. ①高… ②韩… III. 医用数学-概率统计-医学院校-教材
IV. R311

中国版本图书馆 CIP 数据核字(2013)第 017620 号

责任编辑:王 颖 胡治国 / 责任校对:钟 洋
责任印制:肖 兴 / 封面设计:范璧合

版权所有, 违者必究。未经本社许可, 数字图书馆不得使用

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京市农林印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2005 年 8 月第 一 版 开本: 787×1092 1/16

2013 年 1 月第 二 版 印张: 23

2013 年 1 月第九次印刷 字数: 529 000

定价: 54.80 元

(如有印装质量问题, 我社负责调换)

第 2 版前言

《医药应用概率统计》是编者在所进行的“统计及应用课程的教改研究和实践”研究成果荣获国家级教学成果二等奖的基础上,所编写的颇具特色的医药统计课程教材,自 2005 年 8 月出版至今已经印刷 8 次,并在 2009 年被评为江苏省高等学校精品教材。

为适应新形势下我国药学高等教育改革和发展的需要,结合我们多年的教学实践和教材编写经验以及国内外优秀统计学教材的成果,我们对第一版的内容体系进行了全面的修订和完善。本次再版的主要特色有:

一、本次再版在保持第一版特色和优势的基础上,本着“夯实概率统计基础,强化医药应用背景,增强统计软件训练,提升学习实践能力”的编写指导方针,在保持概率统计学科系统性前提下,进一步加强统计理论与医药实际的结合,突出医药专业学生的针对性和统计软件应用的实用性,特别是注重学生的实践能力和自主学习能力的培养,更好地体现“学以致用”的教学目的性,同时也兼顾更高层次(如硕士研究生入学考试)的教学要求。

二、本教材内容涵盖数据的统计处理与图表呈现,系统简明的概率论理论基础、数理统计的基本概念和基本知识,常用统计推断和统计分析方法,统计软件(Excel 数据分析模块)的实际操作应用等,并已包含全国硕士研究生入学统一考试数学考试大纲中要求的概率统计知识。本次修订增加了绪论部分,各章正文均采用医药实际案例引导并贯穿全程的新内容结构,并加以全面的修订完善,精简了方差分析、回归分析和试验设计等章节的部分内容,使修订再版的教学内容更加符合医药人才培养目标的要求,更切合目前的教学实际,结构体系也更为合理完善。

三、为真正体现以学生为中心的教材编写理念,全面提升学生的学习和实践创新能力,本次再版的各章正文以医药应用的案例引导贯穿教学全程,内容系统全面,阐述深入浅出,例题经典实用;每章附有相关统计历史、统计学家简介等的知识链接,以开阔学生视野和知识面;各章的最后两节分别为用以拓展提高的综合例题(包含典型的考研真题解析)和 Excel 软件统计应用的上机操作指导部分;并配有以简表形式高度概括的各章内容提要;适当的各种题型的思考与练习、习题及参考答案;统计软件应用的上机实训题等,从而有效帮助学生消化并巩固所学知识内容,不仅提高其学习效率和知识拓展的能力,同时也培养其运用 Excel 软件的统计分析工具去解决实际问题的操作技能,全面提升其医药统计应用的学习和实践创新能力,同时也方便教师的教学及相关医药工作者的学习使用。

本教材共分 11 章,由高祖新、韩可勤、言方荣等共同负责全书的修订编著和统稿纂定。本书编著时注重博采众长,参考了国内外多种教材和参考文献,同时还得到科学出版社、有关专家及广大读者的大力支持、帮助和指正,在此一并表示衷心的感谢。

本书虽经认真修订,但由于编写时间和水平有限,这一版中仍会有疏漏和不妥之处,恳请各位专家、读者批评指正,以便今后修正完善。

编者

2012 年 7 月

第 1 版前言

随着我国医药高等教育改革和发展的不断深入,高校医药人才的培养模式和要求有了极大的变化,同时也对我国医药类相关专业的统计及应用课程的教学提出了更高更新的要求。为此,我们进行了“统计课程结构、教学手段及面向专业的教学改革”重点教改课题的专题研究,对国内外概率统计及应用课程的教材和内容等进行深入的分析研究,同时结合学科发展动态、医药领域统计应用的特点要求和多年教改实践经验,对课程结构、内容、教学模式和教材等各方面进行了有益的改革探索,所取得的教改成果“统计及应用课程教改研究和实践”于 2001 年荣获国家级教学成果二等奖。在此基础上我们编写了这本颇具特色的《医药应用概率统计》课程教材。

概率统计是研究随机现象统计规律性的学科。本教材的编写既考虑到概率统计学科知识结构的科学性和系统性,又结合了医药领域对统计应用的具体要求和特点,同时针对医药本科学生的基顾同和培养要求,适当选取教材内容的深度和广度,并反映学科发展的时代特征,内容系统而全面,例题典型而实用,编写力求简明易懂,深入浅出。其主要特点是:

一、作为医药类相关专业基础课教材,在尽量保持概率统计学科的科学性和系统性的前提下,以“夯实学科基础,掌握概念方法,强化专业应用,培养实用技能”为重点,不片面追求理论的推导和证明,而强调理论与实际的结合,体现了学以致用目的,并充分考虑更高层次(如硕士研究生入学考试)的教学要求。

二、所选内容涵盖概率论基础、医药应用领域数据处理和统计分析的基本原理、基本知识和常用统计方法,在系统而简明地介绍概率论知识的基础上,以统计数据的处理和分析为核心,注重统计方法思想和实际医药应用的阐述,结合数据和医药专业应用实例说明统计方法的特点、应用条件和场合等,从而形成以概率基础、统计原理、统计方法及统计软件应用为主体并面向医药领域实际应用的内容体系。

三、强化以计算机应用为基础的统计技能的培养。现代医药领域数据处理和统计分析离不开计算机统计软件的应用,根据医药院校本科学生所用软件应满足普及、实用的要求,本教材选用了最为常用的 Microsoft Office 系统的 Excel 软件统计模块来进行统计软件应用的教学,操作指导具体翔实,便于自学,并配有上机训练题,从而能真正提高读者运用统计工具分析和解决实际问题的实际操作能力,达到学以致用的目的。

四、目前已有日趋增多的医药相关专业(诸如医药经济类、医药管理类和医药工程类等)研究生录取都需要参加全国“数学”硕士研究生入学考试,为适应这些更高层次的教学要求,本教材内容已完全包含参加全国“数学”硕士研究生入学考试教学大纲中关于概率统计知识要求的内容,并在每章配有内容提要与综合举例,对核心内容以简表形式进行高度概括,并精选综合性典型例题和考研真题进行详解和分析,为准备考研等更高学习要求的读者提供极为有效的、兼具基础性和技巧性的解题指导。

五、由于本书内容系统全面,阐述深入浅出,用例经典实用,概括高度精练,并兼顾本科基

基础课教学和更高教学要求不同层次的需求,使之不仅适用于医药类各相关专业概率统计、数理统计、应用统计等基础课教材或参考书,而且还可用于农林经管等各类非理工科专业同类课程的参考书,并可作为其他教学,如考研辅导及数学建模等的参考书。而书中极具实用性的医药统计的 Excel 软件操作应用,使之成为从事医药研究和工作的相关人员不可多得的统计应用参考书。

教材的主要内容有统计数据的描述和概括、随机事件和概率、随机变量及其分布、随机向量及其分布、大数定律与中心极限定理、抽样分布、参数估计、假设检验、方差分析、相关分析与回归分析、正交设计与均匀设计等章节,并在每章的最后一节给出 Excel 软件对应概率统计功能的操作应用,同时辅之以相应的典型实例、适当的思考与练习题和应用统计软件的上机训练题,以帮助学生消化、巩固所学内容,真正掌握统计应用的原理和方法。而在每章最后配有的内容提要与综合举例,不仅为读者重点掌握各章的核心内容提供便利,同时也为有更高要求的读者在综合性和技巧性解题方面的提高给予有效的帮助指导,使本教材能适应不同层次教学要求的需要。

本教材供一学期教学使用,也可根据课时和教学要求的不同选用部分内容。其中每章的最后一节——Excel 软件的统计应用,则可根据课时、计算机设备等客观条件来灵活取舍或安排自学。

本教材共分 11 章,其中第 1~6 章和各章 Excel 软件应用、内容提要等部分由高祖新负责编写,第 7~11 章由韩可勤负责编写,最后共同统稿纂定。

本教材的编者,得到了有关专家的关心和帮助,并参考了大量的教材和文献,在此表示衷心的感谢。由于时间和水平有限,书中疏漏和不妥之处在所难免,恳请各位读者批评指正。

编者
2005 年 7 月

目 录

绪论	(1)	第五节 概率的性质与运算法则	(42)
第一章 数据的统计描述和概括	(4)	一、概率公理化定义	(42)
第一节 数据的类型和统计整理	(4)	二、概率的重要性质	(43)
一、数据的分类	(4)	第六节 条件概率和事件的独立性	(45)
二、数据的统计整理和图示	(6)	一、条件概率	(45)
第二节 数据分布特征的统计概括	(10)	二、乘法公式	(46)
一、数据分布集中趋势的测度	(10)	三、事件的独立性	(47)
二、数据分布离散程度的测度	(12)	第七节 全概率公式和贝叶斯公式	(49)
三、数据分布形状的测度	(17)	一、全概率公式	(49)
第三节 数据的直观描述:统计图表	(18)	二、逆概率公式(贝叶斯公式)	(50)
一、统计图	(18)	第八节 独立试验与贝努里概型	(51)
二、统计表	(22)	第九节 综合例题	(53)
第四节 综合例题	(23)	本章内容提要	(56)
第五节 数据整理与统计作图的 Excel 应用	(23)	思考与练习二	(57)
一、Excel 数据分析程序的安装	(24)	习题二	(58)
二、用 Excel 进行统计作图	(24)	第三章 随机变量及其分布	(60)
三、用 Excel 生成频数分布表与直方图	(26)	第一节 随机变量及其概率分布	(61)
四、用 Excel 计算常用统计量	(28)	一、离散型随机变量的分布	(61)
本章内容提要	(30)	二、随机变量的分布函数	(62)
思考与练习一	(32)	三、连续型随机变量的分布	(63)
习题一	(32)	第二节 随机变量的数字特征	(66)
上机训练题	(33)	一、数学期望	(66)
第二章 随机事件和概率	(34)	二、方差与标准差	(70)
第一节 随机事件及其运算	(34)	三、矩	(73)
一、随机试验和随机事件	(34)	第三节 常用离散型随机变量分布	(74)
二、随机事件的关系和运算	(35)	一、二项分布	(74)
第二节 古典概率	(38)	二、泊松分布	(76)
第三节 几何概率	(40)	三、超几何分布	(78)
第四节 统计概率	(41)		

第四节 常用连续型随机变量分布	(78)	第六节 二维随机向量的数字特征	(119)
一、均匀分布	(78)	一、二维随机向量的数学期望和方差	(119)
二、指数分布	(79)	二、协方差与相关系数	(120)
三、正态分布	(80)	第七节 综合例题	(123)
第五节 随机变量函数的分布	(86)	本章内容提要	(127)
一、离散型随机变量函数的分布	(86)	思考与练习四	(129)
二、连续型随机变量函数的分布	(88)	习题四	(130)
第六节 综合例题	(90)	第五章 大数定律与中心极限定理	(133)
第七节 用 Excel 进行常用分布的概率计算	(93)	第一节 大数定律	(133)
一、二项分布	(93)	一、切比雪夫不等式	(133)
二、泊松分布	(95)	二、大数定律	(134)
三、正态分布	(95)	第二节 中心极限定理	(136)
四、指数分布	(98)	第三节 综合例题	(138)
本章内容提要	(99)	本章内容提要	(142)
思考与练习三	(100)	思考与练习五	(142)
习题三	(101)	习题五	(143)
上机训练题	(104)	第六章 抽样分布	(144)
第四章 随机向量及其分布	(105)	第一节 总体、样本和统计量	(144)
第一节 二维随机向量及其分布函数	(105)	一、总体与样本	(144)
第二节 二维离散型随机向量	(107)	二、统计量	(145)
第三节 二维连续型随机向量	(110)	第二节 抽样分布	(147)
一、二维连续型随机向量	(110)	一、样本均值的分布	(147)
二、常用二维连续型分布	(113)	二、 χ^2 分布与样本方差的分布	(148)
第四节 条件分布	(114)	三、 t 分布与 t 统计量	(150)
一、离散型随机变量的条件分布律	(114)	四、 F 分布与两个样本方差比的分布	(152)
二、连续型随机变量的条件密度	(116)	第三节 综合例题	(153)
第五节 二维随机向量函数的分布	(117)	第四节 用 Excel 进行 χ^2 、 t 、 F 分布的计算	(155)
一、二维离散型随机向量函数的分布	(117)	一、 χ^2 分布	(155)
二、二维连续型随机向量函数的分布	(118)	二、 t 分布	(156)
		三、 F 分布	(158)
		本章内容提要	(160)
		思考与练习六	(161)
		习题六	(162)
		上机训练题	(162)

第七章 参数估计	(163)	三、配对比较总体均值的检验	(195)
第一节 参数的点估计	(163)	四、正态总体方差的检验	(196)
一、矩估计法	(163)	第三节 两个正态总体参数的假设	(197)
二、最大似然估计法	(165)	检验	(197)
三、估计量的优良性	(168)	一、两个正态总体方差比较的	(197)
第二节 正态总体参数的区间估计	(170)	检验	(197)
一、区间估计的概念	(170)	二、两个总体均值比较的检验	(199)
二、正态总体均值的区间估计	(171)	(成组比较)	(199)
三、正态总体方差的区间估计	(173)	第四节 非正态总体参数的假设	(202)
第三节 二项分布和泊松分布	(174)	检验	(202)
参数的区间估计	(174)	一、总体均值的假设检验	(202)
一、大样本正态近似法	(174)	二、总体率的假设检验	(203)
二、小样本精确估计法(查表法)	(177)	第五节 拟合优度检验	(207)
第四节 综合例题	(178)	一、 χ^2 拟合优度检验	(207)
第五节 用 Excel 求总体参数的	(180)	二、列联表的独立性检验	(209)
置信区间	(180)	第六节 秩和检验	(213)
一、用 Excel 求正态总体均值的	(180)	一、配对比较的符号秩和检验	(213)
置信区间	(180)	二、两总体比较的秩和检验	(214)
二、用 Excel 求正态总体方差的	(182)	第七节 综合例题	(215)
置信区间	(182)	第八节 用 Excel 进行假设检验	(217)
本章内容提要	(184)	一、用 Excel 进行单个正态总体的	(217)
思考与练习七	(185)	参数检验	(217)
习题七	(185)	二、用 Excel 进行两个正态总体的	(221)
上机训练题	(186)	参数检验	(221)
第八章 假设检验	(188)	三、用 Excel 进行列联表的 χ^2	(224)
第一节 假设检验的基本概念	(188)	独立性检验	(224)
一、假设检验的基本原理	(188)	本章内容提要	(225)
二、假设检验的一般步骤	(190)	思考与练习八	(227)
三、两类错误	(190)	习题八	(228)
第二节 单个正态总体参数的假设	(191)	上机训练题	(231)
检验	(191)	第九章 方差分析	(232)
一、方差已知时正态总体均值的	(191)	第一节 单因素方差分析	(232)
检验	(191)	一、方差分析的原理和方法	(232)
二、方差未知时正态总体均值的	(193)	二、方差分析的步骤与实例	(235)
检验	(193)	第二节 多重比较	(238)
		一、Tukey 法	(239)

二、Scheffé 法	(240)	三、用 Excel 进行一元线性回归	
第三节 两因素方差分析	(241)	分析	(276)
无重复试验的两因素方差分析		本章内容提要	(278)
.....	(241)	思考与练习十	(280)
第四节 综合例题	(245)	习题十	(280)
第五节 用 Excel 进行方差分析		上机训练题	(281)
.....	(247)	第十一章 正交设计与均匀设计 ..	(283)
一、用 Excel 进行单因素方差		第一节 正交表与正交设计	(284)
分析	(247)	一、正交表	(284)
二、用 Excel 进行两因素方差		二、正交设计的基本步骤	(285)
分析	(248)	第二节 正交试验的直观分析 ..	(286)
本章内容提要	(250)	一、正交试验的表头设计	(286)
思考与练习九	(251)	二、直观分析法的分析步骤	(286)
习题九	(252)	三、考虑交互作用的正交试验	
上机训练题	(253)	设计	(289)
第十章 相关分析与回归分析	(254)	第三节 正交试验的方差分析法	
第一节 相关分析	(255)	(291)
一、散点图	(255)	一、正交试验的方差分析法原理	
二、相关关系与样本相关系数		(291)
.....	(256)	二、用正交试验的方差分析法	
三、等级相关分析	(258)	求解	(292)
第二节 一元线性回归分析	(260)	第四节 均匀试验设计	(294)
一、一元线性回归的统计模型		一、均匀设计表和均匀设计	(294)
.....	(260)	二、均匀设计试验结果分析	(296)
二、一元线性回归方程的建立		三、均匀设计应用举例	(296)
.....	(261)	第五节 综合例题	(298)
三、一元线性回归方程的显著性		本章内容提要	(302)
检验	(262)	思考与练习十一	(304)
四、用回归方程进行预测和控制		习题十一	(304)
.....	(266)	参考文献	(306)
五、相关与回归分析的注意事项		附录一 常用统计表	(307)
.....	(268)	附表 1 二项分布表	(307)
六、一元拟线性回归分析		附表 2 泊松分布表	(310)
.....	(269)	附表 3 标准正态分布表	(312)
第三节 综合例题	(271)	附表 4 标准正态分布双侧临界	
第四节 用 Excel 进行相关与回归		值表	(313)
分析	(273)	附表 5 χ^2 分布表	(314)
一、用 Excel 制作散点图	(274)	附表 6 t 分布表	(315)
二、用 Excel 计算相关系数	(275)	附表 7 F 分布表	(316)

附表 8	二项分布参数 p 的置信 区间表····· (320)	附表 13	多重比较中的 q 表····· (330)
附表 9	泊松分布参数 λ 的置信 区间表····· (324)	附表 14	多重比较中的 S 表····· (332)
附表 10	$\varphi=2\arcsin\sqrt{p}$ 数值表 ····· (325)	附表 15	检验相关系数 $\rho=0$ 的临界 值表····· (334)
附表 11	配对比较符号秩和检验用 T 界值表····· (328)	附表 16	等级相关系数的临界值表 ····· (335)
附表 12	两总体比较秩和检验用 T 界值表····· (329)	附表 17	正交表····· (336)
		附表 18	均匀设计表与使用表··· (343)
		附录二	习题参考答案····· (346)

绪 论

概率论与数理统计(probability and mathematical statistics)是从数量侧面来研究随机现象统计规律性的数学学科。而医药应用概率统计(或医药应用数理统计 mathematical statistics in medicine)则是应用概率论与数理统计的原理和方法,对医药、生物等相关领域研究对象的数据资料信息进行搜集、整理、分析和解释,以显示其总体特征和统计规律性的应用学科。

统计作为一种社会实践活动由来已久,其含义也较丰富。它可以指统计数据的搜集活动,即统计工作;也可以指统计活动的结果,即统计数据;还可以指分析统计数据的方法和技术,即统计学。

统计学(statistics)是关于研究对象的数据资料进行搜集、整理、分析和解释,以显示其总体特征和统计规律性的科学。可见,统计学是与统计数据密不可分的,其主要方面包括:①数据搜集,也就是取得统计数据,这是进行统计分析的基础;②数据整理,即用图表等形式来展示数据特征,使数据更加系统化、条理化,从而便于统计分析;③数据分析,是利用描述统计和推断统计等统计方法来研究数据,是统计学的核心;④数据解释,就是对统计分析结果进行说明和应用。

虽然人类的统计实践活动可以追溯到远古的原始社会,但是将统计实践上升到理论加以总结和概括,使之成为一门系统的学科——统计学,距今只有 300 多年的历史。最初的统计方法是随着社会政治和经济的需要而逐步得到发展的,直到 18 世纪概率论被引进之后,统计才逐渐成为一门成熟的学科。

17 世纪中叶,法国数学家帕斯卡(B. Pascal, 1623—1662)和费马(P. Fermat, 1601—1665)等对赌徒 Méré 提出的赌局问题的解决,开创了概率论研究的新纪元,并于 1657 年出版了《论赌博中的计算》,书中提出了数学期望、概率的加法定理与乘法定理等基本概念。瑞士数学家贝努里(Jakob Bernoulli, 1655—1705)创立了最早的大数定理——贝努里定理,建立了描述独立重复试验序列的“贝努里概型”。1662 年,英国统计学家格朗特(J. Graunt)基于伦敦死亡人数资料的研究所进行的死亡率推算,是历史上最早出现的统计推断。1763 年,英国统计学家贝叶斯(T. Bayes, 1702—1761)发表《论机会学说问题的求解》,给出了“贝叶斯定理”,可视为最早的数学化统计推断。而最先将古典概率论引进统计学领域的是法国天文学家、数学家拉普拉斯(P. S. Laplace, 1749—1827),他提出了研究随机现象的分析方法,完善了古典概率论的结构,并阐明了统计学大数法则。比利时统计学家凯特勒(A. Quetelet, 1796—1874)发现了大量随机现象的统计规律性,并开创性地应用许多统计方法完成了统计学和概率论的结合,被认为是数理统计学的创始人。此后,以概率论为基础的统计理论和方法被称为数理统计。

从 19 世纪中叶到 20 世纪中叶,数理统计得到蓬勃发展并逐渐达到成熟。法国医生路易斯(P. C. A. Louis, 1787—1872)1835 年提出了医学观察中的抽样误差和混杂概念、临床疗效对比的前瞻性原则和疗效比较的“数量化”方法,被誉为“临床统计之父”;他的学生盖

瓦勒特(J. Gavarret, 1808—1890)1840年在巴黎出版了世界上首部医学统计学教科书。德国的大地测量学者赫尔梅特(F. Helmert, 1843—1917)在1876年研究正态总体的样本方差时,发现了 χ^2 分布(卡方分布);英国生物学家高尔顿(F. Galton, 1822—1911)在生物遗传学中提出了回归、相关概念,创立了回归分析法。英国数学家、统计学家皮尔逊(K. Pearson, 1857—1936)进一步发展了回归与相关的理论,提出了总体、标准差、正态曲线等重要术语和矩估计法、 χ^2 拟合优度检验法,并创建了生物统计学,为数理统计和生物统计学的发展奠定了基础。英国统计学家戈塞特(W. S. Gosset, 1876—1937)于1908年导出了 t 统计量的精确分布—— t 分布,建立了“小样本理论”;而英国统计学派的代表人物费希尔(R. Fisher, 1890—1962)系统地发展了抽样分布理论,建立了以最大似然估计法为中心的估计理论,首创了试验设计法并提出方差分析法,对现代数理统计的形成和发展作出了极大的贡献。1933年前苏联的著名数学家柯尔莫哥洛夫(A. N. Kolmogorov, 1903—1987)出版了经典名著《概率论基础》,首次以测度论为基础建立了概率的公理化定义,从而使概率论建立在完全严格的数学基础之上,奠定了现代概率论的理论基础。其后波兰统计学家奈曼(J. Neyman, 1894—1981)等在20世纪30年代提出了似然比检验,并建立了置信区间理论;美国统计学家沃尔德(A. Wald, 1902—1950)所建立的序贯分析和统计决策理论,开创了数理统计学的新局面。1946年瑞典数学家克拉默(H. Cramer, 1893—1985)发表《统计学的数学方法》,运用测度论方法总结数理统计的成果,使现代数理统计趋于成熟。

随着自然科学和社会经济的进步和发展,数理统计在理论上不断完善与成熟,应用上日益广泛和深入。特别是随着电子计算机的应用和普及,统计数据的处理能力大为增强,以往受计算能力限制的数理统计有关理论和方法,其处理实际问题的能力也得到了空前提高。数理统计已成为研究自然现象和社会经济现象数量特征的极为有力的工具,并逐步渗透到各个学科领域,形成了许多边缘学科,如生物统计、医药统计、社会统计、水文统计、计量经济学、计量心理学、信息论、决策论、排队论、博弈论、可靠性理论、自动控制、质量管理、临床试验等,成为现代科学发展的一个重要标志。

目前,在医药研究和生产中,无论是疾病防治、药物研发、临床试验、公共卫生等领域,还是新药研制、药物鉴定、药理分析、试验设计、药政管理、处方筛选、医药信息等医药领域的各个方面,都需要进行大量的数据资料的整理和分析,医药统计作为利用相关数据资料进行医药科学研究的重要前提和手段,其相关知识、方法和必要的统计软件应用技能训练,已成为每个医药科技工作者必不可少的专门知识和技能,学习和掌握医药应用概率统计,对于正确有效地利用数据资料进行医药领域的研究和实践具有极为重要的意义。

知识链接

“统计”名词的来历

统计语源最早出现于中世纪拉丁语的“status”,意思指各种现象的状态和状况。由这一语根组成意大利语“stato”,表示国家结构和国情知识的意思。德国政治学教授亨瓦尔(G. Achenwall)在1749年所著的《近代欧洲各国国家学纲要》绪言中首次将

“statistika”(统计)作为国家学名使用,原意是指“国家显著事项的比较和记述”。此后,各国相继沿用“统计”这个词,并把这个词译成各国的文字。法国译为“statistique”,意大利译为“statistica”,英国译为“statistics”,日本最初译为“政表”、“政算”、“国势”、“形势”等,直到1880年在太政官中设立了统计院,才确定以“统计”二字正名。

1903年由钮永建等翻译了4本日本横山雅南所著的《统计讲义录》,把“统计”这个词从日本传到我国。1907年彭祖植编写的《统计学》是我国最早的一本“统计学”书籍。“统计”一词就成了记述国家和社会状况的数量关系的总称。

第一章 数据的统计描述和概括

统计学是对研究对象的数据资料进行搜集、整理、分析和研究的学科。在英文中,“statistics”以单数名词出现时表示统计学,而以复数名词出现时则表示统计数据或资料。可见,统计学与统计数据是密不可分的。

案例 1.1(受教育程度资料)

根据《中国统计年鉴 2010》(国家统计局编)提供的 2009 年全国人口变动情况抽样调查样本数据资料(抽样比为 0.873‰),人口的受教育程度分为未上过学、小学、初中、高中、大专及以上共 5 类,在抽样调查的我国 6 岁以上共计 109.18 万抽样人口中,7.77 万人的受教育程度是未上过学;32.89 万人的受教育程度是小学;45.50 万人的受教育程度是初中;15.06 万人的受教育程度是高中;7.96 万人的受教育程度是大专及以上。

问题:

如何对上述受教育程度资料进行统计整理,并用统计图表显示?

本章就讨论如上述案例所示的有关数据资料的统计整理、图表显示和统计概括等问题。

第一节 数据的类型和统计整理

一、数据的分类

数据(data)是对客观事物计量的结果。例如,对药品质量的计量可得到药品是正品或次品的数据;对药物在试验对象血液中含量的计量可得到血液浓度数据等。统计数据是利用统计方法进行分析的基础。

(一) 数据的类型

统计数据根据观察或实验结果的表现形式是否能用数值表示大体上分为两大类:定性数据和定量数据。

1. 定性数据 定性数据(qualitative data)也称品质数据,是观察或实验结果不可以用数值大小表示只能用文字描述的数据资料,一般不带有度量衡单位。这类数据资料说明的是事物的品质特征,它的特点是每个观察结果或实验结果之间没有量的大小区别,表现为互不相容的类别或属性。根据观察结果是否有等级或顺序,定性数据又可进一步分为定类数据和定序数据两类。

(1) 定类数据(categorical data 或名义数据 nominal data、计数数据 count data):是对事物按照其属性进行分类或分组的计量结果,其数据表现为文字型的无序类别,可以进行每

一类别出现频数的计算,但不能进行排序和加减乘除的数学运算。例如,人的性别分为男、女两类;人的血型分为O型、A型、B型和AB型四类,这些均属于定类数据。

(2) 定序数据(ordinal data 或等级数据 rank data):是对事物之间等级或顺序差别的计量结果,其数据表现为有序类别,可以进行类别的频数计算和排序,但不能进行加减乘除的数学运算。例如,某种药物的疗效可分为无效、有效、显效、痊愈等;新药的等级可分为一类、二类、三类、四类、五类等,均属于定序数据。案例 1.1 中的我国抽样人口的受教育程度数据也属于定序数据。

2. 定量数据 定量数据(quantitative data),也称数值数据(numerical data)或计量数据(measurement data),是观察或实验结果可以用数值大小表示的数据资料,一般带有度量衡单位。这类数据资料是用自然或度量衡单位对事物进行计量的结果,其特点是每个观察值或实验值之间有量的大小的区别,既可进行频数计算和排序,又可进行加减乘除的数学运算。例如,百分制的考试成绩(分)、人的体重(kg)、血压(kPa)、红细胞数(个/L)等,均为定量数据。

实际问题中绝大多数数据资料是定量数据,本书所介绍的统计方法也主要用于定量数据的分析处理,只有非参数方法等可用于定性数据的研究。虽然只有定量数据可转化为定性数据,但也可以通过每类赋值(即编码)的方法,使定量数据的统计分析方法应用于定性数据。

(二) 变量及其类型

在统计中,将说明现象的某种属性或标志称为变量(variable),对变量进行测量或观察的值称为观察值(observation)或变量值(variable value)。统计数据就是统计变量的观察值。根据变量的记录形式分别为定类数据、定序数据和数值数据,相应地变量可以分为定类变量(categorical variable 或名义变量 nominal variable)、定序变量(ordinal variable 或等级变量 rank variable)和数值变量(numerical variable 或 metric variable)。

数值变量中,如果变量可以取有限个值,或可以一一列举,称为离散变量(discrete variable),如制药公司数、仪器个数等。如果数值变量可以取无穷多个值,其取值是连续不断的,不能一一列举,就称为连续变量(continuous variable),如时间、温度、血药浓度等。实际应用时,当离散变量的取值很多时,也可以当做连续变量来处理。

由于在实际中,应用最多的是数值变量,大多数统计方法所处理的也都是数值变量,故一般将数值变量简称为变量,即通常所说的变量主要是数值变量。

区分数据的类型非常重要,如表 1-1 所示,对不同类型的数据必须采用不同的统计方法来进行处理和分析。

表 1-1 不同数据类型之比较

数据类型	定性数据(品质数据)		定量数据
	定类数据(计数数据)	定序数据(等级数据)	数值数据(计量数据)
表现形式	类别(无序)	类别(有序)	数值(+ - × ÷)
对应变量	定类变量	定序变量	数值变量(离散变量、连续变量)
主要统计方法	计算各组频数,进行列联表分析、 χ^2 检验等非参数方法		计算各种统计量,进行参数估计和检验、回归分析、方差分析等参数方法