

全国高等医药院校药学类规划教材

医药统计学

Yiyao Tongji Xue

梁维君 主编



中国医药科技出版社

全国高等医药院校药理学类规划教材

医药统计学

主 编 梁维君

副主编 刘艳杰 张丕德

编 委 (以姓氏笔画为序)

让慰清 (南华大学)

宇传华 (华中科技大学)

吕 媛 (湖南师范大学)

刘桃成 (湖南师范大学)

刘启贵 (大连医科大学)

刘艳杰 (沈阳药科大学)

胡平成 (中南大学)

张丕德 (广东药学院)

李君荣 (江苏大学)

袁秀琴 (南华大学)

宾晓农 (广州医学院)

梁维君 (湖南师范大学)

谭红专 (中南大学)

主编助理 吕 媛

中国医药科技出版社

内 容 提 要

本书为全国高等医药院校药理学类规划教材。在阐述统计学的基本理论、基本知识和基本技能的基础上突出培养学生应用统计学方法的能力,同时增加了医药学科研统计设计和论文中的统计表达等知识,加强了针对性。本书适合医药院校本科生,成人教育、自学教育使用。

图书在版编目(CIP)数据

医药统计学/梁维君主编. —北京:中国医药科技出版社, 2006.7

全国高等医药院校药理学类规划教材

ISBN 7-5067-3465-6

I. 医... II. 梁... III. 医药统计-统计学-医学院校-教材 IV. R311

中国版本图书馆 CIP 数据核字 (2006) 第 081696 号

美术编辑 陈君杞

责任校对 张学军

版式设计 郭小平

出版 中国医药科技出版社

地址 北京市海淀区文慧园北路甲 22 号

邮编 100088

电话 010-62244206

网址 www.cspyp.cn www.mpsky.com.cn

规格 787×1092mm $\frac{1}{16}$

印张 20 $\frac{3}{4}$

字数 470 千字

印数 1—5000

版次 2006 年 8 月第 1 版

印次 2006 年 8 月第 1 次印刷

印刷 北京昌平百善印刷厂印刷

经销 全国各地新华书店

书号 ISBN 7-5067-3465-6/G·0506

定价 32.00 元

本社图书如存在印装质量问题请与本社联系调换

全国高等医药院校药理学类规划教材编委会

名誉主任委员	吴阶平	蒋正华	卢嘉锡
名誉副主任委员	邵明立	林蕙青	
主任委员	吴晓明	(中国药科大学)	
副主任委员	吴春福	(沈阳药科大学)	
	王温正	(中国医药科技出版社)	
	黄泰康	(国家食品药品监督管理局)	
	彭师奇	(首都医科大学药学院)	
	叶德泳	(复旦大学药学院)	
	张志荣	(四川大学华西药学院)	
秘书长	姚文兵	(中国药科大学)	
	朱家勇	(广东药学院)	
委员	(按姓氏笔画排列)		
	丁安伟	(南京中医药大学中药学院)	
	丁红	(山西医科大学药学院)	
	刁国旺	(扬州大学化学化工学院)	
	马毅	(山东轻工业学院化学工程系)	
	元英进	(天津大学化工学院)	
	王广基	(中国药科大学)	
	王月欣	(河北工业大学制药工程系)	
	王地	(首都医科大学中医药学院)	
	王存文	(武汉工程大学)	
	王志坚	(西南师范大学生命科学学院)	
	王岳峰	(西南交通大学药学院)	
	王玮	(河南大学药学院)	
	王恩思	(吉林大学药学院)	
	王康才	(南京农业大学园艺学院)	
	韦玉先	(桂林医学院药学院)	
	冯怡	(上海中医药大学中药学院)	
	史录文	(北京大学医学部)	
	叶永忠	(河南农业大学农学院)	
	白钢	(南开大学生命科学学院)	

乔延江 (北京中医药大学中药学院)
乔海灵 (郑州大学药学院)
全 易 (江苏工业学院化学工程系)
刘 文 (南开大学医学院)
刘巨源 (新乡医学院药学系)
刘永琼 (武汉工程大学)
刘红宁 (江西中医学院)
刘 羽 (武汉工程大学)
刘克辛 (大连医科大学药学院)
刘利萍 (浙江绍兴文理学院化学系)
刘志华 (湖南怀化医学高等专科学校药学系)
刘明生 (海南医学院药学系)
刘杰书 (湖北民族学院医学院)
刘 珂 (山东省天然药物工程技术研究中心)
刘俊义 (北京大学药学院)
匡海学 (黑龙江中医药大学)
印晓星 (徐州医学院药学系)
吉 民 (东南大学化学化工系)
孙秀云 (吉林化学学院制药与应用化学系)
曲有乐 (佳木斯大学药学院)
朱大岭 (哈尔滨医科大学药学院)
朱景申 (华中科技大学同济药学院)
朴虎日 (延边大学药学院)
毕开顺 (沈阳药科大学)
纪丽莲 (淮阴工学院生物工程与化学工程系)
齐香君 (陕西科技大学生命科学与工程学院)
吴 勇 (四川大学华西药学院)
吴继洲 (华中科技大学同济药学院)
吴基良 (咸宁学院)
吴清和 (广州中医药大学中药学院)
吴满平 (复旦大学药学院)
吴 翠 (徐州师范大学化学系)
张大方 (长春中医学院药学院)

张丹参 (河北北方学院基础医学部)
张树杰 (安徽技术师范学院动物科学系)
张振中 (郑州大学药学院)
张晓丹 (哈尔滨商业大学药学院)
张崇禧 (吉林农业大学中药材学院)
李元建 (中南大学药学院)
李永吉 (黑龙江中医药大学药学院)
李青山 (山西医科大学药学院)
李春来 (莆田学院药学系)
李勤耕 (重庆医科大学药学系)
杨世民 (西安交通大学药学院)
杨宝峰 (哈尔滨医科大学)
杨得坡 (中山大学药学院)
沈永嘉 (华东理工大学化学与制药学院)
肖顺汉 (泸州医学院药学院)
辛 宁 (广西中医学院药学院)
邱祖民 (南昌大学化学工程系)
陈建伟 (南京中医药大学中药学院)
周孝瑞 (浙江科技学院生化系)
林 宁 (湖北中医学院药学院)
林 强 (北京联合大学生物化学工程学院)
欧珠罗布 (西藏大学医学院)
罗向红 (沈阳药科大学)
罗焕敏 (暨南大学药学院)
郁建平 (贵州大学化生学院)
郑国华 (湖北中医学院药学院)
郑葵阳 (徐州医学院药学系)
姚日生 (合肥工业大学化工学院)
姜远英 (第二军医大学药学院)
娄红祥 (山东大学药学院)
娄建石 (天津医科大学药学院)
胡永洲 (浙江大学药学院)
胡 刚 (南京医科大学药学院)

胡先明 (武汉大学药学院)
 倪京满 (兰州医学院药学院)
 唐春光 (锦州医学院药学院)
 徐文方 (山东大学药学院)
 徐晓媛 (中国药科大学)
 柴逸峰 (第二军医大学药学院)
 殷 明 (上海交通大学药学院)
 涂自良 (郧阳医学院药学系)
 秦雪梅 (山西大学化学化工学院药学系)
 贾天柱 (辽宁中医学院药学院)
 郭华春 (云南农业大学农学与生物技术学院)
 郭 姣 (广东药学院)
 钱子刚 (云南中医学院中药学院)
 高允生 (泰山医学院药学院)
 崔炯漠 (延边大学医学院)
 曹德英 (河北医科大学药学院)
 梁 仁 (广东药学院)
 傅 强 (西安交通大学药学院)
 曾 苏 (浙江大学药学院)
 程牛亮 (山西医科大学)
 董小萍 (成都中医药大学药学院)
 虞心红 (华东理工大学化学与制药工程学院制
药工程系)
 裴妙荣 (山西中医学院中药系)
 谭桂山 (中南大学药学院)
 潘建春 (温州医学院药学院)
 魏运洋 (南京理工大学化工学院)

全国高等医药院校药学类规划教材编写办公室

主 副	主	任	姚文兵 (中国药科大学)
		任	罗向红 (沈阳药科大学)
			郭 姣 (广东药学院)
			王应泉 (中国医药科技出版社)

编写说明

经教育部和全国高等医学教育学会批准，全国高等医学教育学会药学教育研究会于2004年4月正式成立，全国高等医药院校药学类规划教材编委会归属于药学教育研究会。为适应我国高等医药教育的改革和发展、满足市场竞争和医药管理体制对药学教育的要求，教材编委会组织编写了“全国高等医药院校药学类规划教材”。

本系列教材是在充分向各医药院校调研、总结归纳当前药学教育迫切需要补充一些教学内容的基础上提出编写宗旨的。本系列教材的编写宗旨是：药学特色鲜明、具有前瞻性、能体现现代医药科技水平的高质量的药学教材。也希望通过教材的编写帮助各院校培养和推出一批优秀的中青年业务骨干，促进药学院校之间的校际间的业务交流。

参加本系列教材的编写单位有：中国药科大学、沈阳药科大学、北京大学药学院、广东药学院、四川大学华西药学院、山西医科大学、华中科技大学同济药学院、复旦大学药学院、西安交通大学药学院、山东大学药学院、浙江大学药学院、北京中医药大学等几十所药学院校。

教材的编写尚存在一些不足，请各院校师生提出指正。

全国高等医药院校药学类
规划教材编写办公室
2004年4月16日

前 言

本书是全国高等医药院校药学类规划教材之一。主要教学对象为全国高等医药院校药学专业、五年制本科临床医学专业学生,亦可作为医学检验、护理学、医学影像等专业、成人教育及自学教材。

医药统计学是国内外医药学专业的一门重要的基础课程,它是医药学工作者从事医药学工作和科学研究工作必须掌握的武器。本教材根据概率论与统计学的基本任务,结合编著者多年从事统计学教学和科研的成果与经验,以及国际统计学科的发展趋势,为适应当今我国医药领域对具备相关统计学知识药学人才的要求,对本门课程内容作了必要的调整。全书内容紧紧围绕统计应用这条主线,强调应用实例,阐明统计方法的基本原理和思想,以提高学生学习统计学的兴趣和应用统计方法分析解决实际问题的能力,为学生毕业后从事医药学科研奠定必要的基础。

本书绪论介绍了医药统计学的性质、任务以及统计学的若干术语。第二章介绍了概率基础知识和常用概率分布;第三章为统计描述;第四章至第五章介绍统计推断的逻辑思维和基本方法;第六章至第九章为常用的统计方法;第十章至第十二章为现代统计方法内容;第十三章至第十四章介绍了科研设计和医药科研论文的统计表达;第十五章介绍了统计软件 SPSS 的应用。

本教材的主要特点一是重点阐述统计学的基本理论、基本知识和基本技能;二是突出统计思维和分析思路,培养学生应用统计方法的能力;三是增加了医药学科研统计设计和论文中的统计表达等知识,加强了针对性。

当今计算机、统计软件、多媒体教学课件已在我国高等医药院校普遍使用,课堂讲授的信息量将会大幅度增加。本课程的教学方法强调适应这一形势。讲授本课程教学基本要求的全部内容(主要为第一章至第九章)一般需要 50 学时左右。现代统计方法和医药学科研统计设计等内容供各校根据课时情况选择参考。

本书所用的课程教学体系是著者与全国九所医药院校的教授商讨后拟定的,凝结着各位编委的智慧和心血。本教材的编写得到中国医药科技出版社、湖南师范大学领导和有关部门的大力支持。我尊敬的老师,中南大学的孙振球教授对编写大纲提出了指导性的意见并始终关心着本教材的编写。我的老师黄镇南教授审阅了全部书稿。湖南师范大学医学院以及卫生统计学教研室在本教材编写过程中给予了不少关心与支持。易尚辉老师协助了第四章的编

写，钟贵良副教授、邹亨玉老师为本书做了大量的案头工作。在此，著者向他们致以诚挚的谢意。

由于编者水平有限，教材中一定存在许多不足和纰漏，恳请同行和广大师生不吝赐教。

梁维君

2006年元月于长沙

目 录

第一章 绪论	(1)
第一节 统计与医药统计学	(1)
第二节 变量与数据类型	(3)
第三节 统计学的若干概念	(4)
第四节 医、药学研究的统计过程	(6)
第五节 如何学习医药统计学	(7)
第二章 概率与常用概率分布	(9)
第一节 随机变量的概率	(9)
第二节 概率的基本计算	(11)
第三节 随机变量的概率分布特征	(17)
第四节 随机变量的数字特征	(21)
第五节 常见的两种离散型随机变量的分布	(24)
第六节 正态分布	(28)
思考与练习	(33)
第三章 数据特征与统计描述	(36)
第一节 频数分布	(36)
第二节 描述集中趋势的特征数	(38)
第三节 描述离散趋势的特征数	(42)
第四节 正态分布及其应用	(46)
第五节 医学参考值范围的制定	(48)
第六节 分类资料的统计描述	(50)
第七节 统计图表	(52)
思考与练习	(59)
第四章 抽样分布与参数估计	(63)
第一节 抽样分布与标准误	(63)
第二节 常见的几种抽样分布	(68)
第三节 总体参数的估计	(72)
思考与练习	(79)
第五章 假设检验	(81)
第一节 假设检验的概念与原理	(81)

第二节 单个总体均数的假设检验	(84)
第三节 两个总体均数的假设检验	(87)
第四节 单个总体率的假设检验	(94)
第五节 两个总体率的假设检验	(95)
第六节 等效性检验	(96)
第七节 检验效能及其影响因素	(97)
第八节 区间估计与假设检验	(100)
思考与练习	(100)
第六章 方差分析	(103)
第一节 方差分析的基本思想和应用条件	(103)
第二节 完全随机设计资料的方差分析	(105)
第三节 随机区组设计资料的方差分析	(107)
第四节 多个样本均数间的多重比较	(109)
第五节 方差齐性检验	(113)
第六节 重复测量资料的方差分析	(114)
思考与练习	(118)
第七章 χ^2 检验	(121)
第一节 四格表资料的 χ^2 检验	(121)
第二节 行 $\times$ 列表资料的 χ^2 检验	(126)
第三节 多个样本率间的多重比较	(127)
第四节 分类资料的关联性检验	(129)
第五节 χ^2 检验的注意事项	(130)
第六节 四格表资料的 Fisher 确切概率法	(131)
第七节 频数分布拟合优度的 χ^2 检验	(133)
思考与练习	(135)
第八章 秩转换的非参数检验	(138)
第一节 配对样本比较的 Wilcoxon 符号秩检验	(138)
第二节 两个独立样本比较的 Wilcoxon 秩和检验	(140)
第三节 完全随机设计多个样本比较的 Kruskal - Wallis H 检验	(143)
第四节 随机区组设计多个样本比较的 Friedman M 检验	(147)
思考与练习	(150)
第九章 直线相关与回归分析	(154)
第一节 直线相关	(154)
第二节 等级相关	(159)
第三节 直线回归	(160)
第四节 过原点的直线回归	(166)
思考与练习	(168)
第十章 多重线性回归	(171)

第一节	多重线性回归模型	(171)
第二节	回归方程的检验	(175)
第三节	回归方程的评价	(177)
第四节	自变量的筛选方法	(178)
第五节	多元线性相关	(180)
第六节	多重线性回归分析的应用	(183)
	思考与练习	(188)
第十一章	logistic 回归分析	(191)
第一节	logistic 回归模型	(191)
第二节	logistic 回归模型的应用举例	(197)
第三节	应用 logistic 回归时应注意的几个问题	(204)
	思考与练习	(205)
第十二章	meta 分析	(209)
第一节	meta 分析的基本步骤	(209)
第二节	meta 分析的用途与特点	(211)
第三节	meta 分析方法	(211)
第四节	meta 分析的注意事项	(218)
	思考与练习	(219)
第十三章	科研设计	(222)
第一节	实验研究设计	(222)
第二节	观察性研究设计	(239)
	思考与练习	(243)
第十四章	医药科研论文中的统计表达	(246)
第一节	摘要撰写中的统计学要求	(246)
第二节	材料与方法中的统计方法描述	(246)
第三节	结果的统计表达	(249)
第四节	讨论中的统计表达	(251)
第五节	医药科研论文中的英文表达	(253)
第六节	统计表达中的常见错误	(257)
	思考与练习	(259)
第十五章	SPSS 统计软件	(260)
第一节	SPSS 基础知识	(260)
第二节	SPSS 基本统计方法	(264)
第三节	SPSS 多元统计分析方法	(271)
附录一	统计用表	(274)
附录二	英汉医药统计学词汇	(301)
参考文献		(314)

第一章

绪 论

第一节 统计与医药统计学

英国齐麦曼 (E. A. Zimmeman) 博士于 1787 年首次将德语 “statistik” 译成英语 “statistics” (统计学)。后者与 “state” (国家) 一词来自同一词源。可见早期的 “统计” (数据) 主要指官方 (国家) 收集的信息。

今天, 不同文化背景的人对 “统计” 一词有不同的理解, 但大多把统计数据的收集活动 (统计工作), 统计活动的结果 (统计数据), 以及分析统计数据的方法和技术 (统计学) 统称为 “统计”。

当今, 人类社会已进入信息时代, 统计学已应用到多个学科领域, 它已与其他学科交叉融合形成许多分支学科。给统计学下一个准确的定义确非易事。罗伯特·约翰逊认为 “统计学是收集、描述和解释数据的科学”。本教材给统计学的解释为: 统计学是一门研究数据的收集、整理、分析、表达和解释的方法科学; 目的是探索数据的内在数量规律性, 以发现事物的必然性 (确定性)。

医药统计学 (medical and pharmaceutical statistics): 是运用统计学原理和方法研究生物、医药学问题, 以揭示生物、医药学客观总体的内在数量规律的应用科学, 见图 1-1。

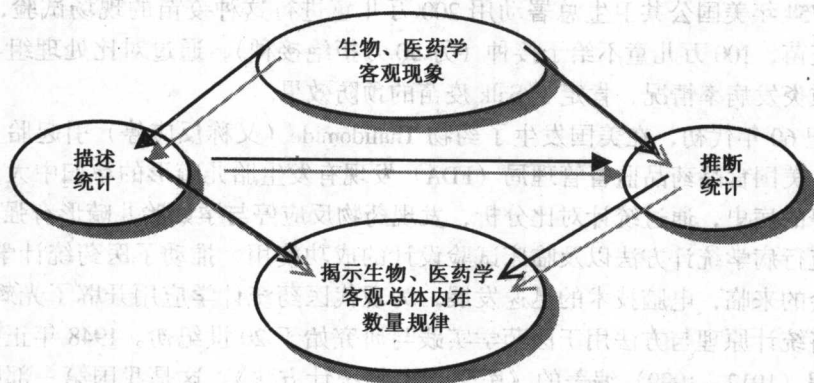


图 1-1 统计学探索生物、医药学数量规律过程

医药统计学的理论方法已广泛渗透到医药卫生研究与实践的各个方面。在阅读文献、研究设计、收集资料、数据记录、观察结果、整理分析、撰写论文等环节无不应用

到统计学。在临床实验、新药研发、生物医学实验室研究、流行病学调查、公共卫生管理、突发卫生事件防制、食品卫生安全等无不寻求统计学工作者的参与。

人类与疾病斗争的历史, 统计学与流行病学写下了不朽的篇章。英国政治算术学派的代表人物之一 John Graunt (1620 ~ 1674) 利用政府定期公布有关人口出生和死亡的数字, 研究后发表了《关于死亡表的自然观察和政治观察》(1662 年)。首次通过大量观察, 发现男婴多于女婴, 约为 14:13; 男性的死亡率高于女性; 一般疾病死亡率较稳定, 而传染病的死亡率波动较大。

早期用观察对比方法用于霍乱研究的是 Farr 和 Snow。他们分析 1853 ~ 1854 年伦敦霍乱流行的死亡资料时, 发现 Lambeth 自来水公司在污染轻的泰晤士河上游取水, Southwark 和 Vauxhall 自来水公司在污染重的泰晤士河下游取水。观察结果在下游取水的死亡率 315.4 (1/万户) 明显高于在上游取水的死亡率 37.5 (1/万户), 支持了 Snow 的水污染假说。1964 年 Doll 和 Hill 采用前瞻性流行病学调查方法, 将 4 万名英国注册医生分为吸烟与不吸烟两组, 观察数年后, 发现吸烟组和不吸烟组肺癌的平均发病率分别为 1.66% 和 0.07%, 相对危险度 $RR = 1.66/0.07 = 23.7$, 强烈提示吸烟的致癌作用。

1747 年英国的 Lind 医生, 将 12 名主要症状是牙龈溃烂、皮肤出血斑点、两脚无力的患者带到一艘船上, 分为 6 组进行了干预实验, 发现每天食 2 个橘子、1 个柠檬的患者, 病情好转最快。用现代医学知识解释, Lind 医生是对坏血病患者 (维生素 C 缺乏) 进行了试验。Lind 把病人放在一艘船上, 便于监督服药和观察病情; 设立对照组, 用 6 种干预相互比较, 这些都运用了统计学知识。

19 世纪, 法国医生 Louis 研究了当时流行的用“放血”疗法治疗伤寒和肺炎效果, 提出了抽样误差和混杂 (Confounding) 的概念以及临床疗效对比的前瞻性原则, 对医学研究的方法学做出了贡献。

英国临床医学研究理事会第一次采用生物统计方法进行临床干预试验, 评价了链霉素治疗肺结核的疗效。

1916 年脊髓灰质炎流行袭击了美国。此后美国人 Jonas Salk 培养的一种疫苗似乎最有希望。1754 年美国公共卫生总署动用 200 万儿童进行这种疫苗的现场试验, 其中 50 万接种了疫苗, 100 万儿童不给予接种 (另 50 万拒绝接种), 通过对比处理组与对照组的脊髓灰质炎发病率情况, 肯定了 Salk 疫苗的预防效果。

20 世纪 60 年代初, 在美国发生了药物 Thalidomide (又称反应停) 引起胎儿畸形的暴发流行。美国食品药品监督管理局 (FDA) 发现有发生胎儿畸形的孕妇中大多有服用药物反应停的病史, 通过统计对比分析, 发现药物反应停与孕妇胎儿畸形有强关联。

上述流行病学统计方法以及临床试验设计的成功使用, 推动了医药统计学的发展。信息化社会的来临, 电脑技术的迅速发展, 为现代医药统计学应用开辟了光辉的道路。在我国, 将统计原理与方法用于医药学实践与研究始于 20 世纪初。1948 年正中书局出版了郭祖超 (1912 ~ 1999) 编著的《医学与生物统计方法》, 这是我国第一部医学统计学教材。新中国成立后, 全国不少高等医药院校相继开设了数理统计、医学统计或卫生统计学课程。特别是改革开放后, 医学统计学迅速普及和发展, 在我国已形成一支老、中、青结合的医学统计学专家队伍, 为疾病防治、药物评价、临床试验、公共卫生等领

域的科学研究提供了有效手段,统计学方法已广泛应用于医药学研究的各个领域。

第二节 变量与数据类型

医药学科研中,常对观察单位(个体)的某项或某些特征(指标)进行观测,统计学把表达这种特征的指标称为变量(variable),变量的取值称为变量值(value of variable)或观察值(observed value)。统计学中所指的变量通常是指随机变量(random variable),它是指取值不能事先确定的观察结果。随机变量统计上用大写的拉丁字母表示。例如调查某地某年女中学生的身高可记为 X ,某一女中学生的身高值记为 X_i 。以人为观察单位调查某地某年女中学生的生长发育状况,如年龄、身高、体重等都可视为变量,相同年龄女中学生的身高、体重、性别不尽相同,这种个体间的差异为变异。这些变异可能来源于一些已知或未知的因素,亦可能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,也可以是定性的。

一、统计数据的分类

由变量值构成数据或资料(data),统计数据可分为如下三类:

1. 计量资料

计量资料(measurement data)又称数值变量(numerical variable)资料。是用仪器、工具或其他定量方法获得的结果,一般带有度量衡单位,如某种患病儿童的身高(cm)、体重(kg)、血压(kPa)、脉搏(次/min),红细胞计数($10^{12}/L$)等均属计量资料。

2. 计数资料

计数资料(count data)又称无序分类变量(unordered categorical variable)资料,亦称名义变量(nominal variable)资料。是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后而得到的资料。可分为二项分类变量和多项分类变量资料。二分类:如观察某药治疗某病患者的疗效,结果可归纳为有效、无效两类。两类间相互对立,互不相容。多分类:如调查某人群的血型分布,其结果可分为A、B、AB、O四种互不相容的血型。

3. 等级资料

等级资料(ranked data)又称有序分类变量(ordinal categorical variable)资料。是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获得的资料。是半定量的观察结果。例如临床检验中常以-、±、+、++、+++等表示若干等级;观察某药治疗某病的疗效,结果常分为治愈、有效、无效、恶化四个等级。

二、资料类型的转换

在统计分析中,为满足不同统计方法的要求,可将一种类型的资料转化为另一种类型。变量可由高级向低级转化,数值变量可转化为有序分类变量或二分类变量。如白细

胞计数属计量(数值变量)资料,若按 $4000 \sim 10000$ (个/ m^3) 属正常,不在此范围属异常,则形成二分类变量资料;若按 ≤ 4000 (个/ m^3), ≥ 10000 (个/ m^3), 可将此分为低,正常,高 3 个类别,此时成为等级资料(有序分类变量资料)。

第三节 统计学的若干概念

一、总体与样本

统计学中把被研究观察单位的全体叫做总体(population),确切地说,总体是根据研究目的确定的,同质的全部观察单位(observed unit)的某个观察值(即某个随机变量 X 可能取的值)的全体。组成总体的每个单元(或元素)叫做个体(individual),个体或称观察单位可以是 1 个基因、1 个蛋白质、1 个细胞;可以是 1ml 血液、1 只眼睛、1 个人;还可以是 1 个自然村(组)、1 个社区等。总体分为有限总体(finite population)和无限总体(infinite population)。有限总体是指在确定的时间和空间范围内总体的个体数可以确定。例如,欲调查某市 2005 年的正常女中学生的血红蛋白值,观察单位(个体)是每一个女学生,观察值(变量值)是每个女生的血红蛋白值,某市 2005 年全部正常女中学生的血红蛋白值就构成一个有限总体。同为某市、同是 2005 年、同为正常女中学生是总体的同质基础。若总体中的个体数量无限,没有时间、空间的限定,只是设想在理论上存在则称为无限总体。例如,欲研究用某药治疗 II 型糖尿病患者的疗效,总体应设想为用该药治疗的所有 II 型糖尿病患者的治疗结果,它没有时间和空间范围的限制,为无限总体。

在医药学的科学研究中,多数情况是采用从总体中抽取样本(sample),根据样本信息推断总体特征的方法,即抽样研究(sampling research)的方法来实现。从总体中抽取一部分个体,按照数理统计思维,就是对随机变量 X 进行若干次试验(观测)。这种从总体中抽取部分个体(观察单位)的过程称为抽样(sampling)。为使样本具有充分的代表性,从总体中抽取样本,必须满足随机性和独立性两个条件。随机性要求抽样必须遵循随机化(randomization)原则,即应使总体中的每个个体都有同等的机会被抽取;独立性指各次抽样必须相互独立,即每次抽样的结果既不影响其他各次抽样的结果,也不受其他各次抽样结果的影响。满足了这两个条件,把从总体中随机抽取的部分个体或确切地说 X 的一组试验数据(观测值),称为随机样本(random sample),样本中所包含的个体的数量(观察单位数)叫做样本含量(sample size),亦称样本容量。如可从某市 2005 年正常女中学生中,随机抽取 120 名,逐个进行血红蛋白的测定,得到 120 名女中学生的血红蛋白值,组成样本。在总体中抽取样本,目的是为了用样本信息推断(inference)总体特征。

二、统计量与参数

根据样本信息计算得出的量称为样本统计量(statistic)。如计算样本 120 名女中学生的血红蛋白平均值就是一个统计量。计量资料常用的统计量有样本均数、样本方差或