



研究生规划教材

全国高等医药院校教材·全国高等医药教材建设研究会规划教材

医学统计学

供 研 究 生 用

Graduate
Student
Graduate
Student

主 编 孙振球
副主编 徐勇勇



人民卫生出版社

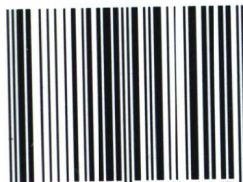


研究生规划教材

责任编辑 杨晋 / 封面设计 赵京津 / 版式设计 韩红梅

Graduate
Student
Graduate
Student

ISBN 7-117-04842-5



9 787117 048422 >

定价(含光盘): 57.00 元

全国高等医药院校教材

供研究生用

医学统计学

主 编 孙振球

副主编 徐勇勇

编者(按章节顺序)

孙振球 (中南大学湘雅医学院)	徐勇勇 (第四军医大学)
潘晓平 (四川大学华西公共卫生学院)	冯学山 (复旦大学公共卫生学院)
颜 虹 (西安交通大学医学院)	李 康 (哈尔滨医科大学公共卫生学院)
康晓平 (北京大学公共卫生学院)	陈冠民 (武汉大学公共卫生学院)
尹 平 (华中科技大学同济医学院)	夏结来 (第四军医大学)
王洁贞 (山东大学公共卫生学院)	田考聪 (重庆医科大学公共卫生学院)
黄镇南 (中南大学湘雅医学院)	易 东 (第三军医大学)
何大卫 (山西医科大学公共卫生学院)	陈平雁 (第一军医大学)
柳 青 (中山大学公共卫生学院)	贺 佳 (第二军医大学)

学术秘书

颜 艳 (中南大学湘雅医学院)

王乐三 (中南大学湘雅医学院)

人 民 卫 生 出 版 社

图书在版编目(CIP)数据

医学统计学/孙振球主编. —北京:

人民卫生出版社, 2002

ISBN 7-117-04842-5

I. 医… II. 孙… III. 医学统计-研究生-教材
IV. R195.1

中国版本图书馆 CIP 数据核字 (2002) 第 049642 号

医学统计学

主 编: 孙 振 球

出版发行: 人民卫生出版社 (中继线 67616688)

地 址: (100078) 北京市丰台区方庄芳群园 3 区 3 号楼

网 址: <http://www.pmph.com>

E - mail: pmph@pmph.com

印 刷: 北京市增富印刷有限责任公司 (尚艺)

经 销: 新华书店

开 本: 850×1168 1/16 印张: 36

字 数: 870 千字

版 次: 2002 年 8 月第 1 版 2003 年 3 月第 1 版第 3 次印刷

标准书号: ISBN 7-117-04842-5/R·4843

定价 (含光盘): 57.00 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究

(凡属质量问题请与本社发行部联系退换)

全国高等医药院校研究生规划教材出版说明

《中国医学教育改革和发展纲要》明确指出，在今后的5~15年我国医学教育要加速发展研究生教育，到2005年，本专科教育（含高等职业技术教育）和研究生教育年招生总量占总体的比例要达到60%以上，到2015年增长到70%以上。为适应这一要求，经全国高等医药教材建设研究会和卫生部教材办公室研究决定，自2001年8月起组织编写一套供研究生使用的规划教材。此套教材较五年制和七年制教材要体现“更高”、“更新”、“更深”的特点；在教材的“三基”（基础理论、基本知识、基本技能）、“五性”（思想性、科学性、先进性、启发性、适用性）方面要更强调启发性，以培养善于思考、勇于探索、敢于创新的临床型和科研型人才。与以课程教育为主的本科学历教育不同，研究生学历教育是课题教育，研究生可根据自己的课题方向选择性地研修相关课程。这就要求我们除了考虑整套教材的一定系统性和交叉内容外，还要指出每种课题中有争论的问题，以及其前沿和发展的方向，以启发研究生在学习中的兴趣，甚至产生科学灵感。

这次编写的19种为第一批研究生规划教材，今后将陆续编辑出版，以供广大读者使用。

KA159-01

前 言

本教材是在教育部的直接领导下，按照全国高等医药教材建设研究会关于研究生公共基础课程教材规划，为适应新世纪高等医药院校研究生教育发展趋势和教学改革的需要，而组织编写的第一本高等医药院校研究生《医学统计学》规划教材。

本教材作为医学统计学教学内容和教学方法的知识载体，作为医学统计学教学的基本工具，在编写风格上有如下特点：

1. 具有继承性和创新性，在继承中有所创新。既继承了国内外同类教材注重统计学教学的共性，又根据研究生教学与科研实践的需求，增补了诸如生存质量分析、统计预测预报、多指标综合评价等多类新方法，并试图用通俗而简练的语言，深入浅出地介绍各类方法的统计学基本原理，同时注重各类统计方法的医学应用及其与常用统计软件（如 SPSS、SAS）间的衔接，扩充了讲授的信息量，有利于启发学生的统计思维，诱导学生的推理乐趣，提高学生运用统计学的基本原理和方法解决医学科研中实际问题的能力。

2. 具有针对性和实用性，在针对中注重实用。全书分为三篇，共 28 章。第一篇为基本统计方法，包括基本统计描述方法和统计推断方法及其应用（第一章至第十章）；第二篇为高级统计方法，包括多元分析方法、生存分析方法、生存质量分析方法、多因素统计预测方法和多指标综合评价方法等多类较为新颖、较为复杂、较为常用的统计方法的基本原理及其应用（第十一章至第二十四章）；第三篇为数据处理的一般方法和统计软件的应用，包括数据处理和评价的一般原则和方法、统计结果的分析与表达、统计软件 SPSS 和 SAS 的特点、运行环境及其在统计分析中的应用（第二十五章至第二十八章），并附有运用这些软件处理和分析本书各章例题的光盘。这种编排格局可适应各高等医药院校研究生学位课程改革和医学统计学教学内容改革的需要，不同学历背景、不同专业背景、不同统计学知识水平、不同类型研究生可根据本专业要求及个人背景有针对性地分篇必修或选修有关内容。除作为教材外，本书还可以供医学专业的博士研究生及医学科研工作者作为参考书或工具书使用。

3. 具有互补性，本教材与同时出版的另外两本研究生规划教材《医学科研方法学》和《临床流行病学》，在教学内容与方法上存在互补性。这里尤其需要说明的是作为医学统计学重要内容的“统计学研究设计”部分，已在《医学科研方法学》有关章节中介绍，本书从略。

本书在编写过程中，得到中南大学、第四军医大学、西安交通大学和有关部门的高度重视和大力支持；来自全国和全军十六所知名高等医药院校的全体编委和编写人员为本教材的诞生付出了心血和劳动；第四军医大学卫生统计学教研室、西安交通大学卫生统计学教研室和中南大学卫生统计学教研室在本书编写工作会议上提供了优质的后勤保障；中南大学的罗建清、虞仁和两位老师，胡国清、史静琤、许林勇、傅建国四位在读硕士研究生为本书例题的复核做了大量细致的工作；本教材学术秘书颜艳副教授、王乐三副教授在联络各编委、筹

备会议以及教材定稿编排等方面付出了辛勤的劳动，谨致以衷心的感谢！

受本人能力及水平限制，加之编写时间不尽宽裕，教材中存在的不足在所难免，企望广大师生和医学界同仁批评指正。

孙振球

2002年4月于长沙

第一批研究生教材目录

1. 医学科学技术哲学	主 编	冯显威	
2. 医学计算机实用教程	主 审	王行言	
	主 编	童隆正	
3. 医学统计学	主 编	孙振球	
4. 临床流行病学	主 审	李立明	
	主 编	黄悦勤	
5. 医学科研方法学	主 编	梁万年	
6. 医学分子生物学	主 审	刘德培	
	主 编	查锡良	
7. 医学分子生物学实验技术	主 编	药立波	
8. 医学细胞分子生物学	主 编	宋今丹	
9. 组织和细胞培养技术	主 编	章静波	
10. 分子病理学	主 编	李玉林	
11. 组织病理技术	主 审	王伯沅	
	主 编	李甘地	
12. 医学遗传学	主 编	夏家辉	
13. 神经生物学	主 编	鞠 躬	
14. 分子病毒学	主 编	黄文林	
15. 基础与临床药理学	主 编	姚明辉	
16. 实验核医学	主 编	张永学	
17. 肿瘤学（第二版）	主 编	曾益新	
18. 外科学——前沿与争论	主 编	邹声泉	龚建平
19. 外科常用实验方法及动物模型的建立	主 编	陈孝平	

目 录

第一篇 基本统计方法

第一章 绪论	1
第一节 统计学的几个基本概念.....	1
第二节 医学统计工作的基本步骤.....	4
第三节 统计学发展简史.....	6
第二章 计量资料的统计描述	9
第一节 频数分布.....	9
第二节 集中趋势的描述.....	12
第三节 离散趋势的描述.....	15
第四节 正态分布.....	17
第五节 医学参考值范围的制定.....	19
练习题.....	21
第三章 总体均数的估计与假设检验	24
第一节 均数的抽样误差与标准误.....	24
第二节 t 分布.....	26
第三节 总体均数的估计.....	27
第四节 t 检验和 u 检验.....	30
第五节 两均数的等效检验.....	36
第六节 假设检验的基本步骤及其注意事项.....	38
第七节 正态性检验和两样本方差齐性检验.....	43
练习题.....	47
第四章 多个样本均数比较的方差分析	51
第一节 方差分析的基本思想和应用条件.....	51
第二节 完全随机设计资料的方差分析.....	53
第三节 随机区组设计资料的方差分析.....	56
第四节 拉丁方设计资料的方差分析.....	59
第五节 两阶段交叉设计资料的方差分析.....	62
第六节 多个样本均数间的多重比较.....	64

第七节 多样本方差齐性检验	67
附 常用基本拉丁方表	69
练习题	70
第五章 计数资料的统计描述	74
第一节 常用相对数	74
第二节 应用相对数的注意事项	76
第三节 率的标准法	77
第四节 动态数列及其分析指标	81
练习题	83
第六章 几种离散型变量的分布及其应用	87
第一节 二项分布	87
第二节 Poisson 分布	92
第三节 负二项分布	96
练习题	101
第七章 χ^2 检验	103
第一节 四格表资料的 χ^2 检验	103
第二节 配对四格表资料的 χ^2 检验	106
第三节 四格表资料的 Fisher 确切概率法	107
第四节 行 \times 列表资料的 χ^2 检验	110
第五节 多个样本率比较的 χ^2 分割法	113
第六节 有序分组资料的线性趋势检验	115
第七节 频数分布拟合优度的 χ^2 检验	118
练习题	119
第八章 秩转换的非参数检验	123
第一节 配对样本比较的 Wilcoxon 符号秩检验	123
第二节 两个独立样本比较的 Wilcoxon 秩和检验	126
第三节 完全随机设计多个样本比较的 Kruskal-Wallis H 检验	130
第四节 随机区组设计多个样本比较的 Friedman M 检验	133
练习题	136
第九章 双变量回归与相关	140
第一节 直线回归	140
第二节 直线相关	146
第三节 秩相关	150
第四节 秩回归	152

第五节	加权直线回归	154
第六节	两条回归直线的比较	156
第七节	曲线拟合	161
练习题		165
第十章	统计表与统计图	169
第一节	统计表	169
第二节	统计图	172
练习题		177
第二篇 高级统计方法		
第十一章	多因素试验的方差分析	181
第一节	析因设计的方差分析	181
第二节	正交设计与方差分析	188
第三节	嵌套设计资料的方差分析	194
第四节	裂区设计资料的方差分析	195
练习题		198
第十二章	重复测量设计的方差分析	200
第一节	重复测量资料的数据特征	200
第二节	重复测量数据的两因素两水平分析	203
第三节	重复测量数据的两因素多水平分析	206
第四节	重复测量数据统计分析常见的误用情况	209
练习题		210
第十三章	协方差分析	212
第一节	协方差分析的基本思想和步骤	212
第二节	完全随机设计资料的协方差分析	215
第三节	随机区组设计资料的协方差分析	217
练习题		220
第十四章	多变量数据的统计描述与统计推断	222
第一节	描述统计量	222
第二节	组间差别比较	225
第三节	重复测量资料的多变量分析	232
第四节	轮廓分析	234
练习题		237
第十五章	多元线性回归分析	241

第一节	多元线性回归	241
第二节	自变量选择方法	247
第三节	多元线性回归的应用及其注意事项	251
练习题		254
第十六章	logistic 回归分析	257
第一节	logistic 回归	257
第二节	条件 logistic 回归	264
第三节	logistic 回归的应用及其注意事项	267
练习题		270
第十七章	生存分析	272
第一节	生存分析中的基本概念	272
第二节	生存率的估计与生存曲线	275
第三节	生存曲线的 log-rank 检验	280
第四节	Cox 比例风险回归模型	282
第五节	寿命表	290
练习题		293
第十八章	判别分析	295
第一节	Fisher 判别	295
第二节	最大似然判别法	298
第三节	Bayes 公式判别法	299
第四节	Bayes 判别	300
第五节	逐步判别	303
第六节	判别分析中应注意的问题	305
练习题		305
第十九章	聚类分析	307
第一节	相似系数	307
第二节	系统聚类	308
第三节	动态样品聚类	314
第四节	有序样品聚类	315
第五节	应用注意事项	318
练习题		318
第二十章	主成分分析与因子分析	321
第一节	主成分分析	321
第二节	因子分析	330

练习题	340
第二十一章 典型相关分析	342
第一节 典型相关分析的基本思想	342
第二节 典型相关分析的数学模型	342
第三节 典型相关变量及典型相关系数的求法	343
第四节 典型相关系数的假设检验	345
第五节 应用实例	346
练习题	348
第二十二章 常用统计预测方法	351
第一节 统计预测概述	351
第二节 指数平滑法	353
第三节 ARIMA 预测方法	358
练习题	371
第二十三章 常用综合评价方法	373
第一节 综合评价的基本概念	373
第二节 综合评分法	381
第三节 综合指数法	383
第四节 层次分析法	386
第五节 Topsis 法	390
第六节 秩和比法	393
练习题	396
第二十四章 生存质量评价	397
第一节 概述	397
第二节 生存质量评价量表	398
第三节 生存质量测定量表的制定与评价	401
第四节 生存质量评价资料的统计方法	407
附：世界卫生组织生存质量测定量表(WHOQOL-100)	413
练习题	425

第三篇 数据处理与统计软件应用

第二十五章 数据处理的一般原则与方法	427
第一节 原始数据的录入	427
第二节 数据处理中的几个基本问题	428
第三节 统计方法选择的基本思路	429

第二十六章 SPSS 统计软件	434
第一节 SPSS 基础知识	434
第二节 SPSS 基本统计方法	438
第三节 SPSS 多元统计分析方法	447
第二十七章 SAS 统计软件	456
第一节 SAS 的基础知识	456
第二节 SAS 数据集的创建	459
第三节 SAS 基本统计方法	463
第四节 多元统计分析方法	477
第二十八章 数据处理的其它统计方法	485
第一节 效度与信度评价	485
第二节 统计结果在论文中的正确表达	495
第三节 医学文献的系统综述方法与 meta 分析	499
练习题	506
附录一 医学人口统计与疾病统计常用指标	510
附录二 统计用表	520
附表 1 标准正态分布曲线下的面积, $\Phi(-u)$ 值	520
附表 2 t 界值表	521
附表 3 F 界值表	522
附表 4 q 界值表 (Newman-Keuls 法用)	526
附表 5 Dunnett- t 检验临界值表	527
附表 6 百分率的可信区间	529
附表 7 Poisson 分布 λ 的可信区间	532
附表 8 χ^2 界值表	533
附表 9 T 界值表 (配对比较的符号秩和检验用)	534
附表 10 T 界值表 (两样本比较的秩和检验用)	535
附表 11 H 界值表 (三样本比较的秩和检验用)	536
附表 12 M 界值表 (随机区组比较的秩和检验用)	537
附表 13 r 界值表	538
附表 14 r_s 界值表	539
附表 15 随机数字表	540
附录三 英汉医学统计学词汇	541
附录四 参考文献	555

附录五 各章例题 (见光盘)

附录六 各章例题 SPSS 数据文件 (见光盘)

附录七 各章例题 SPSS 计算步骤 (见光盘)

附录八 各章例题 SPSS 计算结果 (见光盘)

附录九 各章例题 SAS 计算程序 (见光盘)

附录十 各章例题 SAS 计算结果 (见光盘)

第一篇 基本统计方法

第一章 绪 论

随着医学的发展,作为医学科学研究方法学的医学统计学(statistics of medicine)已逐渐为广大医务工作者和医学科学工作者所认识、所接受,并广为应用。如果说,统计学(statistics)是一门研究数据的搜集、整理、分析的科学,那么,医学统计学就是一门运用统计学的原理和方法,研究医学科研中有关数据的搜集、整理、分析的科学。医学统计学与生物统计学(biostatistics)和卫生统计学(health statistics)既有联系,也有区别,生物统计学应用于整个生物学范畴的有关研究,似比医学统计学范围更广;卫生统计学用于医学与卫生学领域的有关研究,似更侧重于医学与卫生学的社会方面,例如居民健康状况统计和卫生服务统计等。

David Freedman 提出“统计学是对令人困惑费解的问题做出数学设想的艺术”。我们似可把它引伸为“医学统计学是处理医学资料中的同质性(homogeneity)和变异性(variation)的科学与艺术”。首先,体现在资料的数字化上,体现在以数量反映质量的特点上。其次,医学及其相关数据必然蕴藏着丰富的信息,因此,医学统计学的全过程又可看成是一个搜集信息、处理信息、分析信息,从而提炼新的信息的过程。正是在这个过程中,完成了它的使命。即从事物同质性与变异性的数量表现出发,通过一定数量的观察、对比、分析,揭示那些困惑费解的医学问题的规律性,亦即由偶然性(不确定性)的剖析中,发现事物的必然性(确定性),并指导医学的理论和实践。

医学统计学作为一门科学,必须如实地反映现状,不论从实施统计法的角度,还是进行科学研究的角度,我们的目的都在于得到真实的数据,这就是医学统计学的求实性。医学统计学同时作为一门艺术,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,巧妙地选用恰当的高效的统计分析方法,从而得到可靠的结果和科学的结论。

第一节 统计学的几个基本概念

一、总体与样本

观察单位(observed unit),亦称个体(individual),是统计研究中的最基本单位。它可以是一个人,一头动物,也可以是特指的一群人(例如一个家庭、一个幼儿园、一个自然村等);可以是一个器官、甚至一个细胞、一个采样点等。根据研究目的而确定的同质观察单位的全体称为总体(population),更确切地说,它是同质的所有观察单位某种观察值的集合。例如调查某地

2000年7岁正常男童的身高,则观察对象是该地2000年全体正常7岁男童,观察单位是每个男童,观察值(变量值)是测得的身高值,该地2000年全体7岁正常男童的身高值就构成一个总体。它的同质基础是同一地区、同一年份、同一年龄的正常男童。这里的总体明确规定了空间、时间、人群范围内有限个观察单位,称为有限总体(finite population)。在另一些情形下,总体的概念是设想的或抽象的,如研究保列治治疗慢性前列腺增生症的疗效,这里总体的同质基础是慢性前列腺增生症患者,该总体应包括用该药治疗的所有前列腺增生症患者的治疗结果,没有时间和空间范围的限制,其观察单位的全体数只是理论上存在,因而可视为“无限”,称为无限总体(infinite population)。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,必须遵循随机化(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,该样本中所包含的观察单位数称为该样本的样本含量(sample size)。如上例,可从某地2000年7岁正常男童中,随机抽取110名男童,逐个进行身高测量,得到110名男童的身高测量值,组成样本;也可从就诊的前列腺增生症患者中,随机抽取100名患者,并观察保列治治疗前和治疗一段时期后的病情变化,例如,前列腺容量的变化,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,研究者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地某年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。变量的观察结果也可以是定量的,例如身高的厘米数;也可以是定性的,例如新生儿属男属女。按变量属定量亦或定性,可将资料分为以下几种类型:

1. 计量资料(measurement data) 又称定量资料(quantitative data)或数值变量(numerical variable)资料。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。如调查某地某年7岁正常男童的身体发育状况,以男童为观察单位,男童的身高(cm)、体重(kg)、血压(kPa)、坐高指数(坐高/身高,%)等资料均属此类资料。

2. 计数资料(enumeration count data) 又称定性资料(qualitative data)或无序分类变量(unordered categorical variable)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,家族史的有无等等。分两种情形:

- (1)二分类:如检查某小学学生大便中的蛔虫,以每个学生为观察单位,结果可报告为蛔虫卵阴性或阳性两类;如观察某药治疗某病患者的疗效,以每个患者为观察单位,结果可归纳为