

研究生教学用书

教育部学位管理与研究生教育司推荐

获首届全国高等学校医药教材**优秀奖**

# 医学统计学

(第二版)

*Medical Statistics*  
(Second Edition)

主编 孙振球 副主编 徐勇勇



人民卫生出版社

研究生教学用书

教育部学位管理与研究生教育司推荐

# 医学统计学

(第二版)

## Medical Statistics

(Second Edition)

主编 孙振球 副主编 徐勇勇

## 图书在版编目(CIP)数据

医学统计学/孙振球主编. —2版. —北京:  
人民卫生出版社, 2005. 7  
ISBN 978-7-117-06928-1

I. 医… II. 孙… III. 医学统计—医学  
院校—教材 IV. R195. 1

中国版本图书馆 CIP 数据核字(2005)第 064419 号

## 医学统计学 第二版

主 编: 孙振球  
出版发行: 人民卫生出版社(中继线 010-67616688)  
地 址: 北京市丰台区方庄芳群园 3 区 3 号楼  
邮 编: 100078  
网 址: <http://www.pmph.com>  
E - mail: [pmph@pmph.com](mailto:pmph@pmph.com)  
购书热线: 010-67605754 010-65264830  
印 刷: 北京铭成印刷有限公司  
经 销: 新华书店  
开 本: 787×960 1/16 印张: 56.75  
字 数: 982 千字  
版 次: 2002 年 9 月第 1 版—2007 年 4 月第 2 版第 12 次印刷  
标准书号: ISBN 978-7-117-06928-1/R·6929  
定 价(含光盘): 88.00 元  
版权所有, 侵权必究, 打击盗版举报电话: 010-87613394  
(凡属印装质量问题请与本社销售部联系退换)

# 编 者

(按章节顺序)

- 孙振球 (中南大学湘雅医学院)  
颜 艳 (中南大学湘雅医学院)  
潘晓平 (四川大学华西公共卫生学院)  
颜 虹 (西安交通大学医学院)  
康晓平 (北京大学医学部)  
尹 平 (华中科技大学同济医学院)  
王洁贞 (山东大学公共卫生学院)  
王乐三 (中南大学湘雅医学院)  
王 彤 (山西医科大学)  
柳 青 (中山大学公共卫生学院)  
徐勇勇 (第四军医大学)  
冯学山 (复旦大学公共卫生学院)  
李 康 (哈尔滨医科大学)  
陈冠民 (武汉大学公共卫生学院)  
夏结来 (第四军医大学)  
田考聪 (重庆医科大学)  
易 东 (第三军医大学)  
陈 峰 (南京医科大学)  
孙 高 (中国医科大学)  
杨士保 (中南大学湘雅医学院)  
姚 晨 (北京大学医学部)  
陈平雁 (南方医科大学)  
贺 佳 (第二军医大学)

## 学术秘书

- 史静琚 (中南大学湘雅医学院)  
胡国清 (中南大学湘雅医学院)

## 第2版前言

由卫生部教材办公室和全国高等医药教材建设研究会规划的研究生教材《医学统计学》于2002年8月出版以来,为全国大多数高等医药院校选用,反映良好。2004年5月,经教育部遴选、教育部学位管理与研究生教育司推荐本教材为全国“研究生教学用书”,并指令于2005年8月1日前再版。卫生部教材办公室和全国高等医药教材建设研究会按照教育部的要求,组织了全书的修订工作,保证了本书按期再版。

作为当代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,在征求广大医药院校对本书修订意见,并谋求进一步完善和优化本书知识结构的基础上,第2版对第1版的原内容作了更精细的加工,并主要作了如下修订:

1. 由第1版的三篇二十八章扩充为第2版的四篇三十三章:增加了“医学科学研究设计”篇(含四章)和“Stata统计软件”章;将原“生存质量评价”章拓展为“量表研制与量表资料的统计分析方法”章;在第二十二章“常用统计预测方法”中,增加了“人工神经网络在统计预测中的应用”节;在第二十三章“常用综合评价方法”中,增加了“决策树法在综合评价中的应用”节。

修订后的全书结构为:第一章“绪论”,包括医学统计学的定义,重要的基本概念等;第一篇“基本统计方法”,包括基本统计描述和统计推断方法及其应用(第二章至第十章);第二篇“高级统计方法”,包括设计较为复杂的方差分析方法,协方差分析方法,多元统计分析方法,生存分析方法,统计预测方法,综合评价方法和量表的研制与考评方法等一系列较为新颖、复杂而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“医学科学研究设计”,包括观察性研究设计,实验研究设计和临床试验研究设计的基本原理、基本类型、基本方法、基本原则及其应用(第二十五章至第二十八章);第四篇“数据处理与统计软件应用”,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,SPSS、SAS和Stata统计软件的特点,运行环

境及其在医学资料分析中的应用(第二十九章至第三十三章),并附有运用这些软件处理和分析本书各章例题的光盘。

这种编排格局可进一步适应各高等医药院校研究生学位课程设置和医学统计学教学内容改革的需要:按不同学历背景,不同专业背景,不同统计学基础水平,不同类型研究生可根据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。

本书除作为教材外,还可供医药院校博士研究生和医学科研工作者等作为参考书或工具书使用。

2. 作为医学统计学教材建设改革的尝试,在卫生部教材办公室和人民卫生出版社的支持下,本教材编委会拟探索编辑“医学统计学系列教材包”。与本书同时出版的还有两本配套教材,即《医学统计学习题解答》和《SPSS13.0统计软件应用教程》,并计划出版另外两本配套教材,即SAS统计软件和Stata统计软件最新版本的应用教程。

本书再版过程中,得到中南大学有关领导和有关部门的高度重视和大力支持;来自全国和全军十八所知名高等医药院校的全体编委和编写人员为本教材的修订付出了辛勤劳动;哈尔滨医科大学公共卫生学院和卫生统计学教研室,中南大学公共卫生学院和卫生统计学教研室为本书修订会议提供了优质服务和后勤保障;中南大学流行病学与卫生统计学系黄镇南教授,胡平成、曾小敏副教授,虞仁和、王一任讲师,在读博士研究生许林勇,在读硕士研究生郭燕芳、张磊、韦玉华、周绍英、夏强、郑磊等为书稿复核、校对做了大量深入细致的工作;本书第2版学术秘书史静琤、胡国清讲师在联络各位编委、筹备和组织会议以及本书的编排定稿等方面付出了忘我劳动。谨在此一并致以崇高敬意和衷心谢意。

受本人能力及水平所限,教材中不足之处在所难免,企望广大师生和医学界同仁拨冗指正。

孙振球

2005年4月于长沙

# 目 录

第一章 绪论	1
第一节 统计学的几个基本概念	2
第二节 医学统计工作的基本步骤	6
第三节 统计学发展简史	8

## 第一篇 基本统计方法

第二章 计量资料的统计描述	13
第一节 频数分布	13
第二节 集中趋势的描述	17
第三节 离散趋势的描述	21
第四节 正态分布	24
第五节 医学参考值范围的制定	28
练习题	30

第三章 总体均数的估计与假设检验	33
第一节 均数的抽样误差与标准误	33
第二节 $t$ 分布	38
第三节 总体均数的估计	40
第四节 $t$ 检验	44
第五节 假设检验的注意事项	54
第六节 正态性检验和两样本方差比较的 $F$ 检验	57
练习题	64

第四章 多个样本均数比较的方差分析	69
第一节 方差分析的基本思想及其应用条件	69
第二节 完全随机设计资料的方差分析	72
第三节 随机区组设计资料的方差分析	75

第四节	拉丁方设计资料的方差分析 .....	79
第五节	两阶段交叉设计资料的方差分析 .....	83
第六节	多个样本均数间的多重比较 .....	85
第七节	多样本方差比较的 Bartlett 检验和 Levene 检验 .....	89
附	常用基本拉丁方表 .....	91
练习题	.....	93
<b>第五章</b>	<b>计数资料的统计描述 .....</b>	<b>98</b>
第一节	常用相对数 .....	98
第二节	应用相对数的注意事项 .....	100
第三节	率的标准法 .....	103
第四节	动态数列及其分析指标 .....	108
练习题	.....	110
<b>第六章</b>	<b>几种离散型变量的分布及其应用 .....</b>	<b>114</b>
第一节	二项分布 .....	114
第二节	Poisson 分布 .....	122
第三节	负二项分布 .....	128
练习题	.....	134
<b>第七章</b>	<b><math>\chi^2</math> 检验 .....</b>	<b>136</b>
第一节	四格表资料的 $\chi^2$ 检验 .....	136
第二节	配对四格表资料的 $\chi^2$ 检验 .....	141
第三节	四格表资料的 Fisher 确切概率法 .....	142
第四节	行 $\times$ 列表资料的 $\chi^2$ 检验 .....	146
第五节	多个样本率间的多重比较 .....	149
第六节	有序分组资料的线性趋势检验 .....	152
第七节	频数分布拟合优度的 $\chi^2$ 检验 .....	155
练习题	.....	156
<b>第八章</b>	<b>秩转换的非参数检验 .....</b>	<b>161</b>
第一节	配对样本比较的 Wilcoxon 符号秩检验 .....	162
第二节	两个独立样本比较的 Wilcoxon 秩和检验 .....	165
第三节	完全随机设计多个样本比较的 .....	



Kruskal-Wallis $H$ 检验	170
第四节 随机区组设计多个样本比较的 Friedman $M$ 检验	174
练习题	178
<b>第九章 双变量回归与相关</b>	182
第一节 直线回归	182
第二节 直线相关	190
第三节 秩相关	196
第四节 加权直线回归	199
第五节 两条回归直线的比较	202
第六节 曲线拟合	209
练习题	215
<b>第十章 统计表与统计图</b>	219
第一节 统计表	219
第二节 统计图	222
练习题	231
<b>第二篇 高级统计方法</b>	
<b>第十一章 多因素试验资料的方差分析</b>	235
第一节 析因设计资料的方差分析	236
第二节 正交设计资料的方差分析	245
第三节 嵌套设计资料的方差分析	253
第四节 裂区设计资料的方差分析	255
练习题	258
<b>第十二章 重复测量设计资料的方差分析</b>	261
第一节 重复测量资料的数据特征	261
第二节 重复测量数据的两因素两水平分析	266
第三节 重复测量数据的两因素多水平分析	269
第四节 重复测量数据统计分析常见的误用情况	274
练习题	275
<b>第十三章 协方差分析</b>	277

## 4 / 目 录

第一节	协方差分析的基本思想和步骤	277
第二节	完全随机设计资料的协方差分析	281
第三节	随机区组设计资料的协方差分析	284
练习题		287
<b>第十四章</b>	<b>多变量数据的统计描述与统计推断</b>	289
第一节	描述统计量	289
第二节	组间差别比较	292
第三节	重复测量资料的多变量分析	302
第四节	轮廓分析	305
练习题		309
<b>第十五章</b>	<b>多元线性回归分析</b>	313
第一节	多元线性回归	313
第二节	自变量选择方法	321
第三节	多元线性回归的应用及其注意事项	326
练习题		330
<b>第十六章</b>	<b>logistic 回归分析</b>	333
第一节	logistic 回归	333
第二节	条件 logistic 回归	342
第三节	logistic 回归的应用及其注意事项	346
练习题		350
<b>第十七章</b>	<b>生存分析</b>	354
第一节	生存分析中的基本概念	354
第二节	生存率的估计与生存曲线	358
第三节	生存曲线的 log-rank 检验	364
第四节	Cox 比例风险回归模型	367
第五节	寿命表	377
练习题		382
<b>第十八章</b>	<b>判别分析</b>	384
第一节	Fisher 判别	384

第二节	最大似然判别法	388
第三节	Bayes 公式判别法	390
第四节	Bayes 判别	391
第五节	逐步判别	394
第六节	判别分析中应注意的问题	396
练习题		397
<b>第十九章</b>	<b>聚类分析</b>	399
第一节	相似系数	399
第二节	系统聚类	400
第三节	动态样品聚类	408
第四节	有序样品聚类	409
第五节	应用注意事项	412
练习题		413
<b>第二十章</b>	<b>主成分分析与因子分析</b>	415
第一节	主成分分析	415
第二节	因子分析	425
练习题		438
<b>第二十一章</b>	<b>典型相关分析</b>	440
第一节	典型相关分析的基本思想	440
第二节	典型相关分析的数学模型	441
第三节	典型相关变量及典型相关系数的求法	441
第四节	典型相关系数的假设检验	444
第五节	应用实例	445
练习题		448
<b>第二十二章</b>	<b>常用统计预测方法</b>	451
第一节	统计预测概述	451
第二节	指数平滑法	454
第三节	ARIMA 预测方法	461
第四节	人工神经网络在统计预测中的应用	477
练习题		492

第二十三章 常用综合评价方法	493
第一节 综合评价的基本概念	493
第二节 综合评分法	503
第三节 综合指数法	505
第四节 层次分析法	510
第五节 TOPSIS 法	514
第六节 秩和比法	518
第七节 决策树法在综合评价中的应用	522
练习题	527
第二十四章 量表研制与量表资料的统计分析方法	529
第一节 概述	529
第二节 量表的研制方法	531
第三节 量表的考评方法	535
第四节 量表资料的统计分析方法	541
练习题	547
<b>第三篇 医学科学研究设计</b>	
第二十五章 医学科学研究设计概述	549
第一节 医学科学研究的基本任务与特点	550
第二节 医学科学的分类与研究方法	552
第三节 医学科学研究的基本程序与步骤	555
练习题	560
第二十六章 观察性研究设计	561
第一节 调查研究概况	561
第二节 调查设计	562
第三节 抽样调查的样本含量估计	579
第四节 横断面研究	583
第五节 病例对照研究	587
第六节 队列研究	596
练习题	610
第二十七章 实验研究设计	614

第一节	实验设计的基本要素	614
第二节	实验设计的基本原则	617
第三节	样本含量的估计	623
第四节	常用实验设计方法	627
	练习题	634
<b>第二十八章</b>	<b>临床试验研究设计</b>	637
第一节	临床试验的特点和基本要求	637
第二节	临床试验设计的基本要素	638
第三节	临床试验中的偏倚及其控制方法	641
第四节	临床试验的对照组选择	648
第五节	临床试验设计的常见类型	651
第六节	临床有效性的统计学评价	654
第七节	临床诊断试验与评价	659
	练习题	669
<b>第四篇 数据处理与统计软件应用</b>		
<b>第二十九章</b>	<b>数据处理的一般原则与方法</b>	671
第一节	原始数据的录入	671
第二节	数据处理中的几个基本问题	673
第三节	统计方法选择的基本思路	674
<b>第三十章</b>	<b>SPSS 统计软件</b>	679
第一节	SPSS 基础知识	679
第二节	SPSS 基本统计方法	687
第三节	SPSS 多元统计分析方法	695
<b>第三十一章</b>	<b>SAS 统计软件</b>	702
第一节	SAS 的基础知识	702
第二节	SAS 数据集的创建	706
第三节	SAS 基本统计方法	712
第四节	SAS 多元统计分析方法	729
<b>第三十二章</b>	<b>Stata 统计软件</b>	740

第一节 Stata 基础知识 .....	740
第二节 Stata 基本统计方法 .....	745
第三节 Stata 多元统计分析方法 .....	750
<b>第三十三章 数据处理的其他统计方法</b> .....	<b>756</b>
第一节 效度和信度评价 .....	756
第二节 统计结果在论文中的正确表达 .....	770
第三节 医学文献的系统综述方法与 meta 分析 .....	775
练习题 .....	784
<b>附录一 医学人口统计与疾病统计常用指标</b> .....	<b>789</b>
<b>附录二 统计用表</b> .....	<b>803</b>
附表 1 标准正态分布曲线下的面积, $\Phi(u)$ 值 ( $u \leq 0$ ) .....	803
附表 2 $t$ 界值表 .....	804
附表 3 $F$ 界值表 .....	806
附表 4 $q$ 界值表 .....	814
附表 5(1) Dunnett- $t$ 界值表(单侧) .....	815
附表 5(2) Dunnett- $t$ 界值表(双侧) .....	816
附表 6 百分率的可信区间 .....	817
附表 7 Poisson 分布 $\lambda$ 的可信区间 .....	822
附表 8 $\chi^2$ 界值表 .....	823
附表 9 $T$ 界值表(配对比较的符号秩和检验用) .....	824
附表 10 $T$ 界值表(两样本比较的秩和检验用) .....	825
附表 11 $H$ 界值表(三样本比较的秩和检验用) .....	826
附表 12 $M$ 界值表(随机区组比较的秩和检验用) .....	827
附表 13 $r$ 界值表 .....	828
附表 14 $r_s$ 界值表 .....	830
附表 15 随机数字表 .....	832
附表 16 样本均数与总体均数比较(或配对比较)所需 样本含量 .....	834
附表 17 两样本均数比较所需样本含量 .....	835
附表 18 $\psi$ 值表(多个样本均数比较所需样本含量的 估计用) .....	836

附表 19(1) 两样本率比较所需样本含量(单侧) .....	838
附表 19(2) 两样本率比较所需样本含量(双侧) .....	839
附表 20 $\lambda$ 界值表 (多个样本率比较所需样本含量的 估计用) .....	840
附录三 英汉名词对照 .....	843
附录四 汉英名词对照 .....	863
附录五 参考文献 .....	883
附录六 各章例题 (见光盘)	
附录七 各章例题 SPSS 数据文件 (见光盘)	
附录八 各章例题 SPSS 计算步骤 (见光盘)	
附录九 各章例题 SPSS 计算结果 (见光盘)	
附录十 各章例题 SAS 计算程序 (见光盘)	
附录十一 各章例题 SAS 计算结果 (见光盘)	
附录十二 各章例题 Stata 数据文件 (见光盘)	
附录十三 各章例题 Stata 命令与结果 (见光盘)	

# 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(medical statistics)已逐渐为广大医务工作者和医学科学工作者认识和接受,并广为应用。如果说,统计学(statistics)是一门“研究数据的收集、整理和分析的科学”,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收集、整理和分析的应用科学”。医学统计学与生物统计学(biostatistics)和卫生统计学(health statistics)既有联系,又有区别:生物统计学应用于整个生物学范畴的有关研究,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,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”,可把它引申为“医学统计学是处理医学资料中的同质性(homogeneity)和变异性(variation)的科学和艺术”。首先,体现在以数字说明问题,以数量反映质量的特点上。例如,艾滋病患者前五年病死率为50%,这个数据,就充分展示了“高”的水平。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收集信息、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),并用以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,不论从实施统计法学的角度,还是进行科学研究的角度,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),是统计研究中的基本单位。它可以是一个人,一只动物,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。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 2004 年 7 岁正常女童的身高,则观察对象是该地 2004 年全体正常 7 岁女童,观察单位是每个女童,观察值(变量值)是测得的身高值,该地 2004 年全体 7 岁正常女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、同一年龄的正常女童。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的,因而可视为“无限”,称为无限总体(infinite population)。在医学研究中,为节省人力、物力、财力和时间,一般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。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 2004 年 7 岁正常女童中,随机抽取 110 名女童,逐个进行身高测量,得到 110 名女童的身高测量值,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观察或测