

中华人民共和国教育部推荐
医学研究生教学用书



附光盘

主编 方积乾

医学统计学与电脑实验

Medical Statistics and Computer Experiments

with CD-ROM, Fourth Edition

第四版



SEU 2628420

中国医药出版社
CHINA MEDICAL PUBLISHING HOUSE



ISBN 7-5067-3000-0

医学统计学与电脑实验

Medical Statistics and Computer Experiment

李春涛 主编



2628420

R195.1
3/-4

中华人民
医学研

医学统计学与电脑实验(附光盘)

Medical Statistics and
Computer Experiments

(第四版)

主编 方积乾



此书附光盘

上海科学技术出版社

0518588

图书在版编目(CIP)数据

医学统计学与电脑实验:附光盘 / 方积乾主编.
—4版. —上海:上海科学技术出版社,2012.6

(中华人民共和国教育部推荐医学研究生教学用书)

ISBN 978-7-5478-1192-4

I. ①医… II. ①方… III. ①医学统计—统计分析—
应用软件,SAS IV. ①R195.1-39

中国版本图书馆 CIP 数据核字(2012)第 022511 号

上海世纪出版股份有限公司 出版、发行
上海科学技术出版社

(上海钦州南路71号 邮政编码 200235)

新华书店上海发行所经销

常熟市兴达印刷有限公司印刷

开本 787×1092 1/16 印张 39.75

字数:900千字

1997年1月第1版

2001年7月第2版

2006年8月第3版

2012年6月第4版 2012年6月第5次印刷

ISBN 978-7-5478-1192-4/R·383

定价:88.00元

本书如有缺页、错装或损坏等严重质量问题,
请向承印厂联系调换

内 容 提 要

本书共三篇,第一篇为统计学基本概念,第二篇为多元统计方法,第三篇为医学研究的设计与分析,共 33 章。以医学统计学基本概念和常用的设计与分析为主干,辅以统计软件系统 SAS 简介和关于重要统计现象的 100 项电脑实验。每章设有思考、练习与实验题。另附光盘一片,内含习题参考答案、全部电脑实验和部分例题的 SAS 源程序,以及 12 项医学研究课题的数据文件、输入格式和指标的含义等。本书系教育部研究生办公室推荐的医学各科研究生医学统计学课程的教材,第一篇(11 章)可用于必修课,约 100 学时;第二篇(10 章)和第三篇(12 章)可用于选修课,各 50 学时左右。本书也可作为生命科学和农林专业有关课程的教学参考书。

医学统计学与电脑实验

(第四版)

主 编 方积乾

副 主 编 徐勇勇 余松林 郝元涛 张晋昕

编委会成员(以院笔画为序)

上海交通大学医学院	苏炳华	何清波		
山西医科大学	王 彤			
中山大学公共卫生学院	方积乾	柳 青	骆福添	郝元涛
	张晋昕	凌 莉	林爱华	许宗利
	倪涛洋	吴少敏	曾芳芳	朱淑明
	顾 菁	郝 春		
中山大学数学与计算科学学院	李彩霞			
中国医科大学	孙 高			
中南大学湘雅医学院	熊国强			
北京大学公共卫生学院	康晓平			
北京大学第一附属医院	姚 晨			
四川大学华西医学院	李晓松			
兰州大学医学院	丁建生			
华中科技大学同济医学院	余松林	王增珍	尹 平	董时富
武汉大学公共卫生学院	宇传华			
复旦大学公共卫生学院	赵耐青			
南京医科大学公共卫生学院	陈 峰			
南方医科大学公共卫生学院	陈平雁			
第三军医大学	易 东			
第四军医大学	徐勇勇	陈长生	刘丹红	万 毅

前 言

(第四版)

作为教育部推荐的研究生教学用书,《医学统计学与电脑实验》中文第一、二、三版分别于1997、2001和2006年由上海科学技术出版社出版;英文第一版由新加坡 World Scientific Publishing Co. 出版。后者的内容同中文第三版。

这次再版的动议是新加坡 World Scientific Publishing Co. 首先提出的。2010年5月,该公司的编辑来信称,“It is a quite good book and has garnered good reviews. Would you be able to update it to a new edition?”其中,“good reviews”是指各种“书评”。过去几年,该出版社十分关注国际上关于本书英文版的评论文章,并及时转发给我,读之每每受到鼓舞。鉴于中文第三版递增的市场需求,上海科学技术出版社也希望早日再版。于是,我们决定中英文版同时修订,先出中文第四版。

按照国际惯例,再版至少要有30%左右的更新。我们力求在“基本”与“更新”之间取得平衡。首先,原有各章均系“医学统计学”必修课、选修课和医药卫生科学研究所必需的基本内容,在保证不予削弱的前提下,全部章节皆删繁就简、吐故纳新。例如,直方图是必需的,但两组资料的直方图大部分重叠时,看不清孰优孰劣,我们推荐了累积频率图,两组资料表现为上下两条阶梯型曲线,差别一目了然。又如,方差分析的核心是就各种变异来源对总离均差平方和进行分解,计算的便利性则处于第二位;我们放弃了以前各种教科书惯用的一套平方和分解的计算公式,转而强调各种离均差平方和的定义式,凸显不同设计下总离均差平方和不同分解方式的本质;并且,根据误差平方和的定义式,统一用“残差分析”考察各种实验设计资料方差分析的前提条件是否满足。第二,我们根据医学研究发展的需要,以新的概念和知识加强了原有的“基本”内容。例如,适应大规模临床试验的发展,我们增加了人们日益重视的非劣检验(non-inferiority test)、等效检验(equivalence test)和期中分析(interim analysis);针对非正态数据的普遍存在,我们增加了逐渐成为常规武器的随机置换检验(permutation test)。第三,现代临床医学研究一方面宏观地强调“真实世界”临床效果的头-

对-头比较试验(head-to-head comparative trials),另一方面又微观地关注基因和分子水平的海量信息。于是,我们除了加强有关量表研究的统计方法和遗传统计两章外,又新增两章,分别介绍比较效果研究(comparative effectiveness research, CER)和生物信息学(bioinformatics)研究的设计与分析。

自本书的第一版提倡“统计学电脑实验”这一崭新的教学方式以来,国内外同行普遍接受,开设电脑实验课的学校越来越多。这一版,我们除了充实提高原有的电脑实验外,又与课文同步新增了一些实验内容。如随机置换检验、倾向性匹配(propensity matching)和量表条目的功能差异性(differential item functioning, DIF)等,我们提供了 SAS 程序,对海量基因数据、大规模观察性资料以及多中心量表资料的分析与整合十分有用。和以前一样,所有电脑实验的 SAS 程序和部分 SPSS 程序均存在所附光盘内,前者还同时印在书上。

第四版的修订工作主要是在 2011 年暑假期间进行的。余松林教授在“火炉”武汉冒着酷暑勤奋工作,有时竟在深夜回复我发去的电子邮件,讨论本书的修改;徐勇勇教授就方差分析和序贯分析两章大动手术,使之彻底改观;郝元涛教授就光盘内容,制订了“统一要求”,并付诸实施,提供“模板”;张晋昕主任和顾菁老师带领中山大学公共卫生学院的医学统计与流行病学研究生,担当起第四版的第一批读者,分别对纸质版和光盘“横挑鼻子竖挑眼”;易东教授以其深厚的基本功,从大量新文献萃取精华,为我们撰写了反映面貌全新的生物信息学书稿。陈长生、刘丹红、万毅、郝春等老师,创造性地参与了不少章节繁重的修订工作;吴少敏、曾芳芳和朱淑明老师一如既往,在绘图、打字、制盘等技术方面提供了有益的帮助。许多同事曾对第一、二、三版做出过重要贡献,他们的名字均一一记录在第四版的内封上和每一章的后面。本书的责任编辑以极高的效率积极推动中文第四版的问世。在此,谨向以上提到的和未及一一列举的朋友们一并致以衷心的感谢。

经主编和作者们的通力合作,这一版在内容和形式上虽然较前三版有所进步,但是,限于能力和时间;疏漏之处在所难免。欢迎广大师生通过各种方式提出批评建议,我们一定虚心接受,将通过日后的重印及时改正,并尽量体现于随后英文版的修订。

方积乾

2011 年 10 月

前 言

(第三版)

《医学统计学与电脑实验》是教育部推荐的研究生教学用书。前两版分别于 1997 年和 2001 年出版,现在奉献给读者的是第三版(中文),均由上海科学技术出版社出版。一年前,新加坡 World Scientific Publishing Co. 出版了第三版(英文)。第三版中英文两个版本的内容基本相同。

在过去的九年中,许多医学院校使用了本书,2 万册一售而空;通过会议、来信和网络我们不断感受到来自各方面的鼓励和鞭策。广大师生普遍认为本书的内容是医学研究者所必需的,而每一章的电脑实验十分有助于理解统计学概念和完成分析计算。许多国外的同行发现本书集中反映了中国大陆医学统计学老师们特有的教学经验和教学方式,正是在他们热情催促下我们先出了英文版,尽快进行国际交流。

第三版不是简单的再版,而是前两版的进一步调整和改善。首先,根据英文版专家评审的建议,章和篇的顺序有所调整,使内容的线条更加清晰。第一篇仍是统计学基本概念,第二篇是多元统计方法,第三篇是医学研究的设计与分析。我们把紧密联系的主题更好地串起来,如“简单相关-简单回归-多重回归-非线性回归-logistic 回归-生存分析-对数线性模型”等,旨在消除重复和脱节。在每一章内,根据读者的反馈,我们对某些内容作了认真的修改。例如,关于离散变量的统计描述,以前曾把相对数划分为“比、比率和速率”。然而,学生乃至老师常常被比率和频率、比率和构成比、比率和速率搞糊涂,而发病率和死亡率之类却既不是比率又不象速率,“无家可归”。为此,在这一版我们把相对数明确划分为“比、频率和强度”;过去所说的比率和构成比无非是频率和频率分布,而发病率和死亡率之类则属于强度;强度是流行病学熟知的概念,分子是人数,分母是人-时,可以理解为“单位时间内的频率”,它反映单位时间内某事件发生的机会。又如,关于条件 logistic 回归的本质,这一版明确指出,其基础仍然是 logistic 模型,只是配对的数据使得我们只能通过对于内部的条件概率来估计回归系数,常数项却估计不了,因而得不到完整的 logistic 模型,也就不能预测事件

发生的概率。

九年来,统计学电脑实验这一崭新的教学方式已被同行普遍接受,许多学校开设了电脑实验课,深受欢迎,本科教学也在逐渐推广。全书原有的 79 个电脑实验全部保留,大致可分为两类:一类供实现常用的统计学设计与分析,读者只需输入自己的数据,改变若干参数就可以自行计算;另一类供实现模拟实验,读者可以利用我们提供的程序反复“产生数据”、“分析数据”,观察统计现象及其规律。有关的 SAS 程序和 SPSS 程序均做了必要的修改,存在所附光盘内,前者还同时印在书上。

许多朋友参与了这一版的工作,王彤、陈长生、刘丹红、薛富波、万毅等老师协助修改了部分章节。张晋昕老师带领杜建、王兵等 15 名研究生修改了统计用表(附录二)和习题答案(附录三),调试了所有的程序(附录六、七)。Ying Lu、Lin Fei、Kai W Ng、Sin Kai Luo、Min Yang、Boguang Zeng 和 Zongli Xu 等身在海外的朋友们曾创造性地参与了第三版(英文)的工作,他们对英文版的贡献也在中文版都有所反映。许多同事曾对第一和第二版做出重要贡献,他们的名字均一一铭记在第三版每一章的后面。吴少敏、曾芳芳和朱淑明老师在绘图、打字等技术方面提供了有益的帮助。本书的责任编辑罗安妮女士和出版社总编辑应小雄先生更是以极高的效率积极推动第三版(中文)的尽早面世。在此,谨向以上提到的和未及一一列举的朋友们一并致以衷心的感谢。

经作者、同事和朋友们的共同努力,本书的第三版(中文)终于在质量上较前两版进一步提高和成熟了,但是,一旦印出,肯定又会发现不少新的不足和缺点,再次感叹新的遗憾。我们将以此为鞭策,不断努力,精益求精,为完善我国的医学统计学教科书奋斗不已。

方积乾

中山大学公共卫生学院
卫生统计与流行病学系
2006 年春节,广州

前 言

(第二版)

接中华人民共和国教育部教研办[2000]12号文件,得悉经专家评审《医学统计学与电脑实验》(上海科学技术出版社,1997)一书被列为教育部推荐的研究生教学用书。教育部研究生工作办公室建议第二版适当补充新内容和参编单位。于是,我们于2000年6月24~26日在广州中山医科大学召开了修编会议。参加会议的有20多名统计学专家,分别来自上海医科大学、上海第二医科大学、华西医科大学、第四军医大学、同济医科大学以及中山医科大学等14所高等医药院校。经三天认真的讨论,确定了这次修编的原则是:尽量维持原书结构;适当增加新内容;强化概念,突出应用,淡化计算。经过四个月的共同努力,第二版脱稿了。

从全貌来看,第二版较之第一版在可读性方面有较大进步。第一,每章都是从医学问题的挑战开始,围绕实例讨论思路与对策、介绍统计学概念与方法、最终落实于问题的解决和结果的解释。第二,除正面介绍现代医学研究中常用的统计学概念、设计与方法外,本书还不时通过展示典型的不妥与错误和医学实例点评等从反面提醒读者,加深印象。第三,电脑实验更容易实施,除原有的SAS程序外,还用SPSS和EXCEL平行地编写了全部程序,读者可从本书所附的磁盘中调用。

第二版的实用性也较前鲜明。第一,完整地新增了两章:连续型重复测量资料的分析(现第22章)和遗传资料统计分析方法(现第31章);前者是当前应用最广和出错最多的内容,后者是人类基因组学推动之下必定越来越广泛采用的内容。第二,大部分章节都适当增加了现代医学研究中十分重要的内容,如临床试验中的统计学原则(第11章)、ROC曲线(第18章)、通径分析(原第22章、现第21章)、量表的效度、信度与反应度(原第20章,现第13章)、确证性因子分析(第26章)、对应分析(第27章)、多分类Logistic回归模型(第28章)、Poisson回归(第30章)。第三,附录扩充成书面和磁盘两部分,前者包括SAS简介、统计用表、参考答案、参考文献和汉英、英汉两种统计词汇对照与索引;后者包括电脑实验所需的SAS和SPSS可调用程序以及医学科研数据实例。

第二版仍然分三篇。第一篇统计学基本概念,包含十章;第二篇医学研究的设计与分析,包含九章;第三篇现代统计方法,包含十一章。第一和第二篇可用于必修课,理论课和实习课各 50 学时左右;第三篇可用于选修课,理论课约 30 学时,实习课约 20 学时。若全书内容一章不漏地用于教学则需 170~180 学时,各校可根据自己的教学计划和学时数酌情取舍;有些章节标有星号,可在课程进行期间组织学生课外阅读或作读书报告用。

这次修编首先得力于许多兄弟院校工作在教学第一线的优秀中青年教师们,通过参与修编他们无私地贡献了各自的智慧和经验;中山医科大学医学统计学教研室的老师们虚心学习,认真总结,努力将本校和兄弟院校多年来使用第一版进行教学的成功经验融入新版之中;该教研室的技术员吴少敏、曾芳芳等和研究生王梅华、王奇玲、陈妍和刘国辉等全力支持,既为老师们尽心,也确实得到了锻炼。谨代表编委会对参与此次修编的全体人员一并致谢。

在短短四个月内,有这么多学校、这么多专家参与修编,千头万绪难免疏漏;何况,智者千虑必有一失,也许旧错误改掉了,新毛病又出现了。敬请读者和同行继续呵护这本书,严格要求,不吝指正,以便不断修编,一版胜过一版。

方积乾

2000 年 10 月,广州

前 言

(第一版)

我国医学院校和研究机构现行的研究生教学计划大多将医学统计学列为必修课,约 100 学时;有些还另设选修课,约 50 学时。可见医学统计学在医学科研中地位之重要。国内外虽不乏医学统计学佳作,但取之作为上述研究生课程的教材却有诸多不便。为此,许多同行和学生鼓励我们尽快编写一本与现行教学计划基本相吻合的教材。我们这个编写组于 1993 年成立。往日的友谊使我们很容易切磋交流,统一意向。我们将书名定为《医学统计学与电脑实验》,全书以医学统计学基本概念和常用的设计与分析为主干,辅以统计软件包 SAS 的操作和关于重要统计现象的电脑实验。我们商定共同努力营造如下特色:以中学初等数学为起点深入浅出地解释统计学概念与思维逻辑;复盖经典的和现代的重要医学统计方法;将医学研究的设计与分析有机结合;精心设计每章的电脑实验使抽象理论变得具体生动;每章设思考、练习与实验帮助学有所用,用有所思。我们于 1995 年 6 月完成初稿,经互相审阅后一并交主编协调、润色,求得内容、形式和风格的统一。付印之前,主要章节又在作者们各自的学校试教,并征求同行意见。本书共 30 章,每 10 章为一篇。第一篇为统计学基本概念,第二篇为医学研究的设计与分析,第三篇为现代多元统计方法。连同理论课与实习课,每篇各需 50 学时。第一、二两篇合起来可作医药卫生领域研究生必修课医学统计学的教材;第三篇可作医药卫生领域研究生选修课的教材。实际上,第一篇也可作医疗系本科生医学统计学课程的教材,第一、二两篇也可作预防医学系本科生必修课《卫生统计学》的教学参考书。附录中有统计软件包 SAS 的入门简介。每章之后各设一节指导读者利用 SAS 进行电脑实验。电脑设备较多的院校可将现行的实习课扩展为电脑实习课,在教员指导下利用书中提供的 SAS 程序,学员可主动地观察随机现象的统计规律。电脑设备尚感不足的院校可由教员演示电脑实验过程或讲解实验所见。由于电脑实验一节相对独立,完全不具备示教条件的单位不妨跳过这一节,仍可进行其他各节的教学。每章之后的思考、练习与实验可供实习讨论课选用,部分也可作为课外作业和巩固复习内容。多数题目来自实践或文献,某些甚至没有

最佳而只有较佳答案,颇有思考、讨论、实验和回味的余地。本书的附录二提供了若干医学实例和数据,读者可借以练习解决实际问题的能力,附录三是汉英名词对照,汇集了全书散在的主要术语,方便读者查阅。

我们衷心地感谢本书的二位顾问第四军医大学郭祖超教授和中山医科大学胡孟璇教授的指导,感谢中山医科大学研究生处的支持,也感谢上海科学技术出版社唐仲华编审的帮助。还要感谢作者们所在单位的同事和研究生们,在搜集资料、调试程序、打字、校对以及其他事务性工作方面他们默默地伸出援助之手,他们的名字不胜枚举,一一记在作者们的心中。

限于我们的学识和精力,这本书还有许多不足之处,我们将虚心吸取广大读者的批评与建议,争取在再版中给予弥补。

方积乾

1995年9月,广州

目 录

绪论	1
第一篇 统计学基本概念	5
1 描述性统计	7
2 概率与分布	32
3 样本均数的抽样误差与置信区间	50
4 连续型变量的假设检验	62
5 分类变量资料的 χ^2 检验	84
6 假设检验的功效与样本量	109
7 单因素方差分析	124
8 基于秩次的非参数检验	148
9 线性相关	166
10 线性回归	175
11 实验设计的统计学基本原则	196
第二篇 多元统计方法	213
12 多重回归与相关	215
13 多元统计量与多元方差分析	237
14 判别分析	250
15 Logistic 回归分析	266
16 聚类分析	287
17 主成分分析	296
18 因子分析	307
19 典则相关与对应分析	320
20 生存分析	336
21 对数线性模型与 Poisson 回归	347
第三篇 医学研究的设计与分析	361
22 多因素方差分析	363
23 连续型重复测量资料的分析	381
24 横断面研究的设计与分析	392

25	追踪研究的设计与分析	409
26	病例-对照研究的设计与分析	428
27	诊断和筛查试验研究的设计与分析	444
28	序贯试验设计与分析	459
29	系统综述与 Meta 分析	470
30	比较效果研究	485
31	量表研制的统计方法	505
32	遗传流行病学资料的统计方法	523
33	生物信息学资料的统计方法	537
附录一	统计软件 SAS 简介	555
附录二	统计用表	565
附录三	汉英名词对照	600
附录四	英汉名词对照	609
附录五	SAS 程序和数据	见光盘
附录六	习题答案	见光盘
附录七	医学科研数据实例	见光盘

绪 论

什么是统计学

统计学(statistics)和统计数字(statistics)在英语中共用一个单词,作为复数名词,意指统计数字,作为单数名词,表示统计学。这个词来源于 state,可见早期的统计数字是指官方所要求的信息。现在仍然如此,但不限于此,各行各业都有大量的统计数字,其中蕴含着丰富的信息。Webster 国际大辞典(第三版)中说,统计学是“a branch of mathematics dealing with the collection, analysis, interpretation and presentation of masses of numerical data”。Last JM 主编的《流行病学辞典(第四版)》中说,统计学是“The science and art of collecting, summarizing, and analyzing data that are subject to random variation. The term is also applied to the data themselves, and to summarizations of the data”。其他文献也有类似表述。由此看来,首先,统计学是从统计数字中挖掘知识、特别是对付资料中变异性的科学和艺术。将“科学”和“艺术”两项桂冠同时授予一门学问,这是很不寻常的,可能是资料中的变异性太普遍,太棘手的缘故;第二,统计学的目的在于取得可靠的结果,其求实性毫不含糊,既不为装点门面,也不为自欺欺人;第三,统计学是在收集、归类、分析和解释大量数据的过程中完成其使命的,这一点,统计学的顾客们并非都很了解,许多人到了分析数据阶段才想起统计学,不免发生“悔之晚矣”的憾事。

总体与样本

统计学中称试图了解和研究的全部个体(individual)为总体(population)。这里的个体可以指人,可以指动物,或学校、工厂等任何观察单位,也可以指由观察单位所得的数值。在多数场合,总体是无限的,人们不可能掌握所有个体;对于较大的有限总体,即使有可能作全数观察,也不应提倡,因为人力物力投入很大,观察的质量却不易控制。对于有限总体,有时也不允许作全数观察,例如,打碎安瓿作注射剂的品质检查,查一支毁一支,逐一查毕,全部报废。一般说,人们总是从总体中抽取一部分个体,构成样本(sample),通过对样本的分析来推断总体的规律性。

统计学好比总体与样本间的桥梁,它帮助人们设计与实施从总体中科学地抽取样本的过程,使样本中的个体不多也不少,代表性好;帮助人们提取信息(information)、采集数据(data)、推断关于总体的知识(knowledge);帮助人们确切地描述样本中的现象,恰当地解释总体中可能存在的规律。