

研究生教学用书
教育部研究生工作办公室推荐

方积乾 主编

医学统计学与电脑实验 (附光盘)

Medical Statistics and Computer Experiments

With CD-ROM, Third Edition

(第三版)

上海科学技术出版社

研究生教学用书

教育部研究生工作办公室推荐

医学统计学与电脑实验 (附光盘)

Medical Statistics and
Computer Experiments
With CD-ROM, Third Edition

(第三版)

方积乾 主编

上海科学技术出版社

图书在版编目 (C I P) 数据

医学统计学与电脑实验 / 方积乾主编. —3 版. —上海: 上海科学技术出版社, 2006.8

ISBN 7 - 5323 - 8415 - 2

I . 医... II . 方... III . ①医学统计②医学统计—
统计分析—应用软件, SAS IV . R195.1

中国版本图书馆 CIP 数据核字 (2006) 第 026777 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

上海新华印刷有限公司印刷

开本 787 × 1092 1/16 印张 35.5

字数 844 000

1997 年 5 月第 1 版

2006 年 8 月第 3 版

2006 年 8 月第 3 次印刷

定价 68.00 元

本书如有缺页、错装或坏损等严重质量问题,
请向工厂联系调换

内 容 提 要

本书共 3 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 31 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 的操作和关于重要统计现象的 81 项电脑实验。每章设有思考、练习与实验题,书后提供参考答案。另附光盘 1 张,内含统计软件 SAS 简介、全部电脑实验的 SAS 源程序和大部分电脑实验的 SPSS 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。第一篇和第三篇可用于必修课,约 100 学时;第二篇可用于选修课,约 50 学时。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。

主 编	方积乾	中山大学公共卫生学院
副 主 编	徐勇勇	第四军医大学
	余松林	华中科技大学同济医学院
	苏炳华	上海交通大学医学院
	柳 青	中山大学肿瘤医院
	骆福添	中山大学公共卫生学院

编委会成员 (以校名笔画为序)

何清波	上海交通大学医学院
王 彤	山西医科大学
张晋昕	中山大学公共卫生学院
许宗利	中山大学公共卫生学院
林爱华	中山大学公共卫生学院
郝元涛	中山大学公共卫生学院
凌 莉	中山大学公共卫生学院
李彩霞	中山大学公共卫生学院
倪涛洋	中山大学公共卫生学院
朱淑明	中山大学公共卫生学院
吴少敏	中山大学公共卫生学院
曾芳芳	中山大学公共卫生学院
孙 高	中国医科大学
熊国强	中南大学湘雅医学院
李晓松	四川大学华西医学院
康晓平	北京大学公共卫生学院
姚 晨	北京大学第一医院
丁建生	兰州医学院
王增珍	华中科技大学同济医学院
尹 平	华中科技大学同济医学院
董时富	华中科技大学同济医学院
宇传华	华中科技大学同济医学院
陈 峰	南京医科大学
陈平雁	南方医科大学
赵耐青	复旦大学医学院
陈长生	第四军医大学

前 言

《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。1 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 9 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,以尽快进行国际交流。

第三版不是简单的再版,而是对前两版内容的进一步调整和完善。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像速率,“无家可归”。为此,在这一版中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各个对子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。

9 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,深受广大师生欢迎,大学本科教学也在逐渐推广。全书原有的 81 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。上海科学技术出版社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 年春节于广州

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;覆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两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辑的帮助。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月于广州

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,分别来自复旦大学医学院、上海第二医科大学、华西医科大学、第四军医大学、华中科技大学同济医学院以及中山医科大学等14所高等医学院校。经3日认真地讨论,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4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。

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(现第二十二章)和遗传资料统计分析方法(现第三十一章);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(第十一章)、ROC曲线(第十八章)、通径分析(原第二十二章、现第二十一章)、量表的效度、信度与反应度(原第二十章,现第十三章)、确证性因子分析(第二十六章)、对应分析(第二十七章)、多分类logistic回归模型(第二十八章)和Poisson回归(第三十章)。第三,附录扩充成书面和光盘两部分,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,包含十二章。第一篇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、刘国辉和郑迎东等全力支援,既为老师们尽心服务,又确实得到了锻炼。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 4 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 月于广州

目 录

绪论	1
第一篇 统计学基本概念	5
第一章 描述性统计	7
第一节 变量与数据	7
第二节 频数表与直方图	8
第三节 样本平均水平的度量	12
第四节 样本变异性的度量	15
第五节 相对数与标准化方法	17
第六节 电脑实验	20
思考、练习与实验	22
第二章 概率与分布	25
第一节 概率的意义及相关的一些概念	25
第二节 随机变量的分布特征	27
第三节 二项分布	29
第四节 Poisson 分布	32
第五节 正态分布	35
*第六节 概率的基本运算	39
第七节 电脑实验	43
思考、练习与实验	44
第三章 样本均数的抽样误差与置信区间	46
第一节 样本均数的分布	46
第二节 t 分布	50
第三节 正态分布总体均数的置信区间	51
第四节 两正态总体均数之差的置信区间	53
第五节 二项分布总体概率以及概率之差的置信区间	55
第六节 估计置信区间所需的样本量	57
第七节 电脑实验	58

思考、练习与实验	59
第四章 连续型资料的假设检验	62
第一节 假设检验的独特逻辑	62
第二节 单组完全随机化设计资料均数的 t 检验	64
第三节 随机化配对设计资料均数的 t 检验	65
第四节 两组完全随机化设计资料均数的 t 检验	66
第五节 两组完全随机化设计资料方差齐性的 F 检验	69
第六节 二项分布和 Poisson 分布大样本资料参数的 Z 检验	70
第七节 实例点评	76
第八节 电脑实验	76
思考、练习与实验	78
第五章 假设检验的功效与样本量	81
第一节 两类错误与功效	81
第二节 影响功效的四要素	82
第三节 功效与四要素的定量关系	84
第四节 常用统计检验的样本量估算	88
第五节 实例点评	89
第六节 电脑实验	90
思考、练习与实验	91
第六章 离散型分类计数资料的 χ^2 检验	92
第一节 χ^2 分布和 Pearson 拟合优度检验	92
第二节 比较两独立样本频率的 χ^2 检验	93
第三节 2×2 交叉分类资料的 χ^2 检验	96
第四节 $R \times C$ 列联表资料的 χ^2 检验	100
第五节 检验频数分布的 χ^2 检验	103
第六节 两个标准化率相等的假设检验	104
*第七节 四格表精确概率检验法	107
第八节 电脑实验	108
思考、练习与实验	112
第七章 线性相关	114
第一节 线性相关概念	114
第二节 相关系数	115
第三节 相关系数的统计推断	116

第四节 等级相关	118
第五节 线性相关分析的注意事项	119
第六节 电脑实验	121
思考、练习与讨论	121
第八章 线性回归	123
第一节 线性回归的统计描述	123
第二节 回归系数的统计推断	124
第三节 线性回归的应用及其前提条件	128
第四节 电脑实验	132
思考、练习与讨论	133
第九章 非线性回归	135
第一节 利用线性回归拟合曲线	135
第二节 非线性回归的最小二乘拟合	138
第三节 电脑实验	140
思考、练习与实验	141
第十章 实验设计的统计学基本原则	143
第一节 实验设计的基本概念	143
第二节 临床试验中的统计学原则	145
第三节 随机化技术	150
第四节 医学实例点评	152
第五节 电脑实验	153
思考、练习与实验	155
第十一章 方差分析	157
第一节 完全随机设计与结果的方差分析	157
第二节 随机区组设计与结果的方差分析	164
第三节 拉丁方设计与结果的方差分析	166
第四节 析因设计与结果的方差分析	169
* 第五节 裂区设计与结果的方差分析	172
* 第六节 二阶段交叉设计与结果的方差分析	177
第七节 电脑实验	179
思考、练习与实验	182
第十二章 基于秩次的非参数检验	186
第一节 配对样本的 Wilcoxon 符号秩检验	186

第二节	两独立样本分布位置相同的假设检验	189
第三节	多个样本分布位置相同的假设检验	192
第四节	电脑实验	198
	思考、练习与实验	200
第二篇	多元统计方法	203
第十三章	多重回归与相关	205
第一节	多重线性回归模型及基本分析步骤	205
第二节	多重线性相关	209
第三节	回归分析中自变量的选择	211
第四节	多重线性回归应用中的若干技巧	214
* 第五节	通径分析	219
第六节	电脑实验	222
	思考、练习与实验	224
第十四章	多元统计量与多元方差分析	226
第一节	常用多元统计量	226
第二节	两个均数向量的比较——Hotelling T^2 检验	229
第三节	多个均数向量的比较——多元方差分析	232
第四节	电脑实验	235
	思考、练习与实验	237
第十五章	判别分析	240
第一节	判别分析的基本思想	240
第二节	Fisher 准则下的判别分析	241
第三节	Bayes 准则下的判别分析	242
第四节	逐步判别分析	245
第五节	回顾性考核和前瞻性考核	246
第六节	应用中的注意事项	247
第七节	电脑实验	248
	思考、练习与实验	249
第十六章	logistic 回归分析	250
第一节	logistic 回归模型	250
第二节	条件 logistic 回归模型	257
* 第三节	多分类 logistic 回归模型	260

第四节	logistic 模型的应用	263
第五节	电脑实验	265
	思考、练习与实验	267
第十七章	聚类分析	269
第一节	聚类分析的意义	269
第二节	系统聚类法	270
第三节	快速聚类法	273
第四节	对变量的聚类	274
第五节	电脑实验	275
	思考、练习与实验	276
第十八章	主成分分析	278
第一节	主成分的基本概念	278
第二节	主成分的计算与解释	280
第三节	主成分回归	282
第四节	电脑实验	285
	思考、练习与实验	286
第十九章	因子分析	289
第一节	因子模型	289
第二节	初始因子的提取	290
第三节	因子图和因子的旋转	292
第四节	因子得分与因子模型的应用	295
* 第五节	证实性因子分析	296
第六节	电脑实验	300
	思考、练习与实验	301
第二十章	典则相关与对应分析	302
第一节	典则相关	302
* 第二节	对应分析	308
第三节	典则判别	313
第四节	电脑实验	314
	思考、练习与实验	315
第二十一章	生存分析	317
第一节	生存分析的基本概念	317
第二节	生存率估计的乘积极限法	318

第三节	生存曲线比较的 log-rank 检验和 Breslow 检验	320
第四节	Cox 回归分析	322
第五节	电脑实验	325
	思考、练习与实验	326
第二十二章	对数线性模型与 Poisson 回归	328
第一节	列联表的对数线性模型	328
第二节	Poisson 回归	336
第三节	电脑实验	338
	思考、练习与实验	340
第三篇	医学研究的设计与分析	343
第二十三章	连续型重复测量资料的分析	345
第一节	重复测量资料举例	345
第二节	分析中常见的不妥及其原因	346
第三节	综合指标的分析	347
第四节	重复测量资料的一般线性模型	348
第五节	电脑实验	351
	思考、练习与实验	352
第二十四章	序贯试验设计与分析	354
第一节	基本概念	354
第二节	质反应序贯设计与分析	354
第三节	量反应序贯设计与分析	355
第四节	成组序贯试验	356
第五节	电脑实验	360
	思考、练习与实验	361
第二十五章	横断面研究的设计与分析	363
第一节	研究设计	363
第二节	估计总体参数的抽样方法与参数估计	364
第三节	样本含量的估计	368
第四节	现时寿命表	371
第五节	电脑实验	375
	思考、练习与实验	376