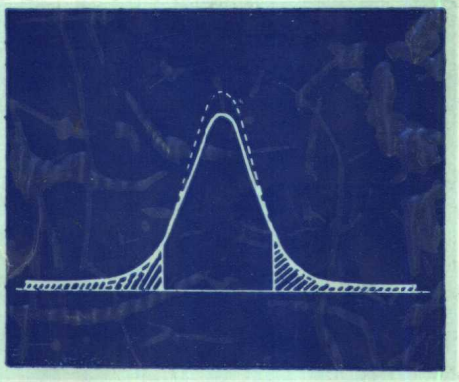
方积乾　徐勇勇　余松林　苏炳华　吴启宏　等编著

医学统计学与电脑实验



上海科学技术出版社

医学统计学与电脑实验

主　编

方积乾（中山医科大学）

副 主 编

徐勇勇（第四军医大学）
余松林（同济医科大学）
苏炳华（上海第二医科大学）
吴启宏（香港大学）

编写组成员

柳　青（中山医科大学）
骆福添（中山医科大学）
何清波（上海第二医科大学）
宇传华（第四军医大学）
董时富（同济医科大学）
王增珍（同济医科大学）

上 海 科 学 技 术 出 版 社

责任编辑 唐仲华

医学统计学与电脑实验

主 编 方积乾

上海科学技术出版社出版、发行

（上海瑞金二路 450 号）

新华书店上海发行所经销 常熟文化照相制版彩印厂印刷

开本 787×1092 1/16 印张 28.5 字数 678,000

1997 年 4 月第 1 版 1997 年 4 月第 1 次印刷

印数 1—10,000

ISBN 7-5323-4033-3/R・1136

定价：28.00 元

前　言

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，使读者学有所用，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，汇集了全书中的主要术语，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，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. 什么是统计学　统计学和统计数字在英语中共用 statistics 一词。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。统计学是和这类信息有关的一门学问。Webster 国际大辞典（第三版）中说，统计学是“a science dealing with the collection, analysis, interpretation and presentation of masses of numerical data”。Last JM 主编的一本流行病学辞典中说，统计学是“the science and art of dealing with variation in data through collection, classification and analysis in such a way as to obtain reliable results”。其他文献也有类似表述。由此看来，首先，统计学是处理资料中变异性的科学和艺术。将“科学”和“艺术”两顶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，统计学是在搜集、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。

2. 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，也可以指动物或学校、工厂等任何观察单位。在多数场合，总体是无限的，人们不可能对所有个体进行全数观察。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。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。一般说，人们总是从总体中抽取一部分个体，构成样本；通过对样本的观察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，信息丰富，代表性好；帮助人们挖掘样本中的信息，推断总体的规律性；帮助人们确切地描述样本中观察到的现象，恰当地解释总体中可能存在的规律。

3. 若无变异无需统计　在统计学术语中，变异就是多样性、不确定性，其反面就是千篇一律、千人一面。总体中若无变异，所有个体一模一样，只需观察任一个体，总体便清清楚楚。

事实上，客观世界充满了变异，生物医学领域更是如此。同类生命体的结构与功能千差万别，对内外环境的变迁反应各异，即使同一个体的各种生理量也在时间与空间中悄悄地变异不休。

哪里有变异，哪里就需要统计学。例如任何测量结果既含真值，又含变异，那么，已知测量值，其真值大约是多少？再如任何两份样本之间总有差别，也许出自同一总体，差别乃变异所致；也许出自不同总体，真值的差别甚于变异，那么，光看样本怎知总体同与不同？又如，两个量似有某种联系，由于变异作祟，必然性与偶然性难解难分，那么，如何拨雾见青天，

揭示其必然性?

4. 医学科研中统计学的作用 现在我们列举两个实例,帮助读者感受医学研究中统计学是如何发挥作用的。

(1) 麻醉剂氟烷的一场风波。美国于1958年开始在外科手术中采用麻醉剂氟烷(Halothane)。以其不易燃、不易爆、副作用小等优点,到1962年时大约普及到了一半手术。不料,突然有报告称数例病人术后恢复的同时突然恶化、发烧、死亡,尸解显示肝脏大片坏死,一时间疑云四起。氟烷是否损害肝脏?是否应禁止用于手术?为回答这些问题,自然需要与其他常用麻醉剂比较。其他麻醉剂对肝脏损害如何?是否氟烷有其特殊的副作用?查书、专家咨询均不得要领。急性动物实验随时可做,但与病人的外科手术相去甚远。为此,卫生当局决定进行调查。

当时使用氟烷的手术病例已逾千万,不可能也不必全数调查。于是,决定在有较完整病案资料的34家医院中抽取1960~1964年间的85万例手术病例,其中记有所需个体信息,如性别、年龄、使用何种麻醉剂,是否死于术后6周内、术前状况、手术方式等等。不分死因,其中共有17000例死于术后6周内,粗死亡率为2%。按所用麻醉剂区分,与氟烷、喷妥撒(pentothal)、环丙烷(cyclopropane)、乙醚(ether)和其他麻醉剂相应的粗死亡率分别为1.7%,1.7%,3.4%,1.9%和3.0%。是否由此可以认为氟烷的死亡威胁小于其他麻醉剂的死亡威胁呢?

显然,死亡威胁与术前状况有关,病情轻者死亡率仅0.25%,重者可达30%;也与手术本身的危险性有关,各种手术死亡率低者仅0.25%,高者达14%;还与年龄有关,病情和手术相同时青少年死亡率低,老年死亡率高;此外,死亡危险还与性别、医院等有关。各种麻醉剂使用对象不同,直接比较上述粗死亡率便毫无意义。这时,必须借助各种统计方法加以校正。假定各种麻醉剂的使用对象具有相同的术前状况,采用相同的手术、同年龄、同性别和同一所医院,相应的死亡率会如何?

经过校正,与氟烷、喷妥撒、环丙烷、乙醚和其他麻醉剂相应的死亡率分别为2.1%,2.0%,2.6%,2.0%,2.5%。从而说明:所有的麻醉剂都可能导致一部分病人术后6周内死亡,不同麻醉剂相应的死亡率水平不甚悬殊,氟烷的死亡威胁并无特殊。

统计学对这场风波的平息是有贡献的,其中抽样调查、粗死亡率概念和统计学校正等使得不确实的印象得到科学的澄清,也使人们对各种麻醉剂的死亡威胁增进了认识。

(2) 吸烟危害健康的论证。吸烟危害健康,如今已成国际共识。回顾历史,从科学角度论证这一命题浸透了统计学家的心血。人们不可能像药物研究那样,随机地安排一部分人去吸烟,安排另一部分人不吸烟,追踪观察其结局,从而需要特殊的统计学设计和分析。

Muller (1939)采用病例-对照设计搜集一组肺癌患者,并配置一组其他特点与之相仿的非肺癌患者作为对照组,逐一询问吸烟否、吸烟量和时间,发现肺癌组吸烟者比率高,对照组吸烟者比率低。Pearl (1938)调查了数百个家庭,形成了吸烟多、吸烟少和不吸烟三个组,根据各人的寿命,编制了三份寿命表,相应地绘制了三条生存曲线,反映各组随年龄增长减员的过程。发现吸烟多的一组几乎以直线下降的趋势减员,不吸烟的一组以先凸后凹的曲线趋势缓慢减员,而吸烟少的一组减员趋势则介于以上两组之间。

较之这类回顾性研究更进一步的是 Doll and Hill(1964) 的一项出色的前瞻性研究。他们向60000名英国医生发出关于吸烟的问卷,其中40000名应答,据此将他们分成吸烟组和

不吸烟组。借助英国良好的死亡登记系统追踪他们的结局，发现吸烟组肺癌的年发病率为1.66‰，心脏病 年发作率为5.99‰，而不吸烟组肺癌的年发病率为0.07‰，心脏病年发作率为4.22‰。

类似于以上的研究还有很多，但相反的意见仍很强盛。烟草公司和相关的行业以及政府财税部门反对禁烟，不足为怪。岂料，有两位在别的问题上常有分歧的统计学权威 Sir Ronald Fisher 和 Prof. Joseph Berkson 对吸烟有害的研究却一致提出挑战。

Berkson 认为，据称吸烟能提高许多种死因下的死亡率，这一点不可理解。除非有证据说明吸烟对整体健康有害或加速老化，不然，可认为类似上述的研究中对照组的选择以及吸烟状况的资料搜集方式等的偏倚导致了吸烟提高多种死亡率的假象。

Fisher 认为，已有的大量研究不能排除遗传因素。也许具有某种遗传因素者既爱吸烟，又易得癌，而无此因素者既厌吸烟，又难得癌。若果真如此，戒烟并不能摆脱癌症。

两位统计学权威的挑战促进了吸烟有害研究的深入。瑞典国家双生研究即为一例。他们调查了一方吸烟、另一方不吸烟的双生对，其中，同卵双生男274对，女264对，异卵双生男733对，女 653对，发现咳嗽的患病率在同卵双生的吸烟者中男女分别为14.6%和13.6%，而不吸烟者中男女分别为7.7%和7.6%；在异卵双生的吸烟者中男女分别为12.3%和14.5%，而不吸烟者中男女分别为5.5%和5.7%。吸烟与不吸烟相比，咳嗽的相对危险度约为1.8～2.5倍。

此外，还有更多的间接证据。例如，狗的模拟吸烟导致与人类肺癌相似的结局、吸烟有损动脉血管、吸烟与死亡率有剂量-反应关系、戒烟时间长短对死亡率有不同影响等等。鉴于吸烟与禁烟已非单纯的生物医学问题，有的统计学家建议当局利用统计决策理论权衡公众行动起来禁烟的代价和由此给社会带来的得与失，从而作出合理的决策。

目前，许多发达国家已经断然采取了种种限制吸烟的措施，此中融有统计学家的贡献。他们在吸烟有害问题的研究中所创造的方法学也为其他公害的研究提供了借鉴。

5. 学点统计，迎接挑战　诚然，如上所述，欲从事医药方面的科学研究必须从设计、分析到解释全面借助统计学。然而多数医药卫生工作者并非以研究为业，与统计学有何相干？

有人调查过 New England J. Medicine，British Medical J. 和 Lancet 等著名医学杂志上发表的文章，其中大约70%应用了统计学。中国国内较为优秀的医学杂志也如此，多数文章均经统计分析。又据国内外报道，在医学论文所应用的统计学知识中约70%是最基本的概念和经典的统计方法，其余则是较为复杂的、近代发展起来的统计理论和技术，而出现错误最多的却偏偏是前一部分。这一事实表明，研究人员和编辑固然需要提高统计学素养，普通的医药卫生工作人员作为读者也需提高识别统计学错误的本领，否则就不能正确对待出版物中的结果与结论，人云亦云，贻误自身的工作。

其实，由于繁忙而并非经常阅读医学出版物的医药卫生工作者在日常工作中也时常受到挑战。例如，如何正确理解与运用医学指标的正常参考范围？如何总结自己的治疗经验？如何证实自己提出的诊断方法不亚于现有方法？如何考证民间验方的优劣？如何进行社区调查？如何进行达标验收？医药工作者需要学点统计，迎接挑战。学统计学基本概念，学统计学独特的思维方式，学常用统计方法和电脑统计软件的使用。至于复杂的统计学理论与技术，专业性较强，不必人人掌握，需要时可通过协作解决问题。

第一篇　统计学基本概念

第一章　描述性统计

统计分析的目的是由样本推断总体，故统计学的主体是统计推断(statistical inference)。然而，描述性统计(descriptive statistics)也不可忽视，它是数据处理的必经之途。通常，总是利用统计表和统计图或利用某些简单统计量来描述资料的某些特征，为严格的统计推断奠定基础。本章介绍常用描述性统计方法，除频率分布外，还包括反映计量资料平均水平与变异性的指标和描述计数资料的不同性质的相对数指标。

1.1　变量与数据

1. 变量的类型　在总体中，个体的许多属性存在变异性，统计学上把反映这类属性的指标称为变量，如年龄、性别等。针对不同类型的属性，需采用不同类型的变量，因而产生不同类型的资料。

(1) 连续型变量与计量资料。诸如个体身高、体重、血压、脉搏和血细胞计数等变量均可经测量取得数值。限于测量精度，身高、体重之类并不能取任意位小数；如脉搏、血细胞计数之类，测量值只能是正整数。尽管如此，为便于理论研究，人们将这类变量视为连续型变量(continuous variable)，近似地将其取值范围当作实数轴上的一个连续区间。有的书上将这类变量的观察值构成的资料称为计量资料(measurement data)。

(2) 离散型变量与计数资料。某些属性只能定性地划分成少数几个互相排斥的类型，如性别之男女，职业之各种行档，药物反应之阴性和阳性等。描述性别这个属性的变量取"值"范围只是男和女两个类别，描述职业这个属性的变量取"值"范围只是工、农、商、学、兵等少数几个类别。这类变量称为类别变量或名义变量。在有些场合，如果给各类别适当赋值，它们便成为"假变量"(dummy variable)，可以和其他变量一起进行运算。

例 1.1　性别变量 X 可定义为一个二值变量(binary variable)，

$$X=\begin{cases}0 & \text{女性}\\ 1 & \text{男性}\end{cases}$$

例 1.2　职业变量。若职业分工农商学兵 5 个行档，可用 4 个二值变量来描述。令

$$X_1=\begin{cases}0 & \text{非工}\\ 1 & \text{工}\end{cases},\quad X_2=\begin{cases}0 & \text{非农}\\ 1 & \text{农}\end{cases},\quad X_3=\begin{cases}0 & \text{非商}\\ 1 & \text{商}\end{cases},\quad X_4=\begin{cases}0 & \text{非学}\\ 1 & \text{学}\end{cases}$$

于是可有

职业类别	变量			
	X_1	X_2	X_3	X_4
工	1	0	0	0
农	0	1	0	0
商	0	0	1	0
学	0	0	0	1
兵	0	0	0	0

上述只能在孤立的几个数中取值的变量称离散型变量(discrete variable)。二值变量是最简单的离散型变量。

人们时常在一批对象中清点某属性各类别出现的次数,称为频数(frequency)。

例 1.3 一批 108 名病人构成的样本中,按性别划分男性 63 人,女性 45 人;职业按工农商学兵划分各有 28 人、23 人、24 人、18 人和 15 人;药物反应呈阳性者 28 人,呈阴性者 80 人。

类似上述离散型变量的频数资料有的书上称为计数资料(count data)。

按前述性别变量的赋值,108 名病人各有一个 X 的数值,或 0 或 1。这些个体值之和就是 108 名病人中男性的数目。据例 2 中职业变量的赋值,108 名病人各有一个 X_1 的数值,或 0 或 1;这些个体值之和就是 108 名病人中工人的数目。类似地,108 名病人的 X_2 值相加,得到农民的数目。由此可见,一批样本中关于某个类别的计数相当于该样本中相应 0～1 二值变量的个体值之和。

(3) 有序变量与等级资料。某些测量手段只能提供半定量结果。例如,临床中常以－,±,+,++,+++等表示若干等级。另外,有些属性的各个类别存在自然的等级。例如,药物疗效这一属性常可分为治愈、有效、无效和恶化四个等级鲜明的级别。个体的这类属性也可用一个变量来描述,但变量的取值并不反映该个体的确切定量值,只反映类别的等级或秩次(rank)。这样的变量称为有序变量(ordinal variable)。

与计数资料类似,实践中人们也时常清点样本中各等级出现的频数,有的书上称这类频数资料为等级资料(ranked data)。

2. 数据的结构与特点 任何试验和观察的结果必须转变为数据后才能进行统计分析。医学研究中的绝大多数研究结果可用一种统一的数据结构表达,即共有 N 个基本观察单位,每个基本观察单位共记录了 p 个项目。这种数据结构可写成一个 N 行 p 列的方阵,也叫作数据矩阵,SAS,SPSS 和 BMDP 等统计软件都以此作为数据录入的基本格式,如程序 12.2中有一个 18×2 的数据矩阵,即有 18 个基本观察单位,每个单位有处理和试验结果 2 个记录。程序 12.4 中有一个 12×3 的数据矩阵,即有 12 个基本观察单位,每个单位有批次、测量条件和试验结果 3 个记录。

(1) 基本观察单位。基本观察单位是按研究需要确定的采集数据的基本单位。观察对象本身可以是一个基本观察单位,也可以同时具有若干个基本观察单位。以高血压临床治疗的临床试验为例,如果以患者治疗四周后的收缩压和舒张压作为研究数据,则每个患者是一个基本观察单位;如果将患者治疗后一周、二周、四周的收缩压和舒张压作为研究数据,由于采集数据的条件有了变化,每个患者具有三个基本观察单位。

(2) 记录项目。用于统计分析的记录项目通常由分组因素、反应变量和协变量三部分组成。如表 1.1 为一个 100×7 的数据矩阵,在 7 个记录项目中,治疗方法为分组因素,收缩压、

舒张压、心电图、疗效判定为反应变量，年龄、性别为协变量。

(3)“硬”数据与“软”数据。“硬”数据('hard' data)指用现代化工具、仪器或实验方法测得的结果，这种数据受主观因素影响较小，并且大多是数字化的定量结果，因此具有较好的准确性和可重复性。“软”数据('soft' data)则指不能用客观方法准确测量的数据，如患者的主诉、疗效评价等，这种数据容易受主观因素的影响。在临床试验中，“软”数据的重要性并不亚于“硬”数据，甚至比“硬”数据更有价值。例如临床上评价恶性肿瘤的治疗效果，患者的疼痛程度、心情状态和生活自主能力对患者本人和家属来说，往往比肿瘤体积、甚至比生存时间还要重要。因此，在临床试验中已日益重视“软”数据，如生存质量评价，并且以此作为疗效判定的一个重要方面。

表 1.1 100 名高血压患者治疗后的临床记录

患者编号	年龄(岁)	性别	治疗分组	收缩压(kPa)	舒张压(kPa)	心电图	疗效评定
1	37	男	A 药	18.67	11.47	正常	显效
2	45	女	对照	20.00	12.53	正常	有效
3	43	男	B 药	17.33	10.93	正常	有效
4	59	女	对照	22.67	14.67	异常	无效
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	54	女	B 药	16.80	11.73	正常	有效

“软”数据经过“硬”处理后也可用常规统计方法进行定量分析。常用的“硬”处理方法是建立一些相对客观的准则对“软”数据作定量化处理，以提高数据的准确性和可重复性，如用烧伤指数判断烧伤的严重程度、用特制的问卷评价受试者的生存质量等。

1.2 频数表与直方图

频数表(frequency table)与直方图(histogram)不仅是最常用的综合描述样本资料的方法，而且孕育出统计学中十分重要的关于概率分布的概念。

1. 频数表 在一批样本中，相同情形出现的次数就是该情形的频数。将互相排斥的情形的频数无遗漏地列在一起便是频数表。

对于离散型变量，上述“所有互相排斥的情形”就是某属性的各种类别。由例 1.3 中的资料可列出两个频数表(表 1.2 和表 1.3)。其中，频率等于频数与合计数之商，频率之和等于 100%；累积频率是将频率依次累加的结果，例如，

表 1.2 108 名病人中性别频数表

类别	频数	频率(%)	累积频数	累积频率(%)
女	45	41.7	45	41.7
男	63	58.3	108	100.0
合计	108	100.0		

表 1.3 累积频率栏中，与“工”相对应的数值与频率栏的数值相同，与“农”相对应的等于频率栏中“工”、“农”所对应频率之和。