

YIYONG TONGJI FANGFA
JIQI SPSS RUANJIAN SHIXIAN

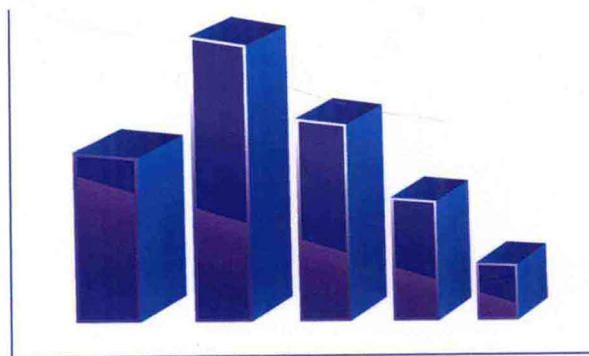
全国高等学校医学规划教材

医用统计方法 及其SPSS软件实现

主 编 潘发明

副主编 郝元涛 姚应水 吴学森 朱继民

主 审 叶冬青 夏结来



中国科学技术大学出版社



中国医药出版社
CHINA MEDICAL SCIENCE PUBLISHING HOUSE

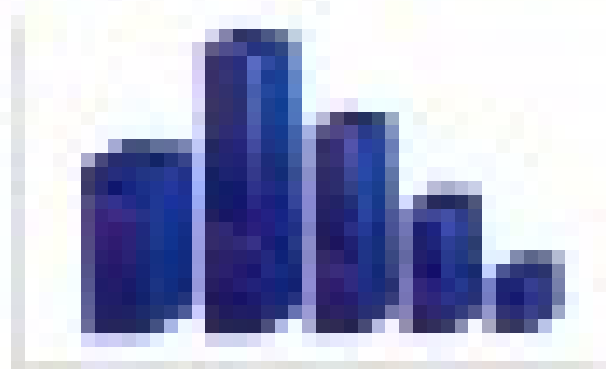
中国医药出版社 医药统计学系列教材

医用统计方法 及其SPSS软件实现

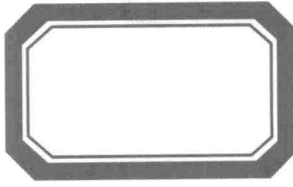
主编 李树刚

副主编 李树刚 李树刚 李树刚 李树刚 李树刚

主审 李树刚 李树刚



ISBN 978-7-5168-1000-0



国高等学校医学规划教材

医用统计方法及其 SPSS 软件实现

主 编 潘发明
副主编 郝元涛 姚应水
 吴学森 朱继民
主 审 叶冬青 夏结来

中国科学技术大学出版社

· 合 肥 ·

内 容 简 介

本书是为适应医学类专业“统计学”教学的需要而编写的,主要介绍医用统计方法及 SPSS 统计软件在医学工作中的应用。依据医学类专业人才教育培养目标的要求,从统计学基本原理和方法出发,在不打破传统的卫生统计学学科体系(适当调整)的基础上,将医用统计方法与 SPSS 软件操作结合起来,注重理论联系实际,弥补了目前高校医学统计学教材的缺憾。SPSS 软件实现部分以 SPSS 16.0 及以上版本为基础,不仅简要介绍了 SPSS 统计软件的数据管理及常用菜单功能,而且较详尽地介绍了 SPSS 统计软件的基本操作和常用技巧,重点在于如何利用 SPSS 统计软件分析实际的医学问题。为了便于读者复习巩固和加深理解,本书各章均有学习小结和适量的习题,并附有参考答案,方便读者自学与提高。

全书共 16 章,内容主要涉及医学领域的统计分析,既可作为高等医学院校“统计学”课程教材,也可供医学统计研究人员学习参考。

图书在版编目(CIP)数据

医用统计方法及其 SPSS 软件实现/潘发明主编. —合肥:中国科学技术大学出版社,2012.9
全国高等学校医学规划教材
ISBN 978-7-312-03040-6

I. 医… II. 潘… III. ①医学统计—方法—医学院校—教材 ②医学统计—统计分析—软件包——医学院校—教材 IV. R195.1

中国版本图书馆 CIP 数据核字(2012)第 124034 号

责任编辑:张善金

出版者:中国科学技术大学出版社

地 址:合肥市金寨路 96 号 邮编:230026

网 址:<http://www.press.ustc.edu.cn>

电 话:发行部 0551-3606086-8808

印刷者:安徽省瑞隆印务有限公司

发 行 者:中国科学技术大学出版社

经 销 者:全国新华书店

开 本:787mm×1092mm 1/16

印 张:29.25

字 数:730 千

版 次:2012 年 9 月第 1 版

印 次:2012 年 9 月第 1 次印刷

印 数:1—5000 册

定 价:45.00 元

《医用统计方法及其 SPSS 软件实现》

编 审 委 员 会

主 编 潘发明

副主编 郝元涛 姚应水 吴学森 朱继民

主 审 叶冬青 夏结来

编 委(排名不分先后,按汉语拼音首字母排列)

艾 东

皖南医学院

段振华

安徽医科大学

范引光

安徽医科大学

贾贤杰

蚌埠医学院

郝元涛

中山大学

刘 思

安徽医科大学

倪进东

广东医学院

潘发明

安徽医科大学

唐晓武

合肥职业技术学院

王 斌

安徽医科大学

王 静

安徽医科大学

吴学森

蚌埠医学院

武 松

安徽中医学院

夏结来

第四军医大学

张承业

安徽医科大学

曾 臻

安徽医科大学

叶冬青

安徽医科大学

姚应水

皖南医学院

朱继民

安徽中医学院

邹延峰

安徽医科大学

秘 书 范引光 刘 思 安徽医科大学



序 言

《医用统计方法及其 SPSS 软件实现》是为面向在校研究生开设的必修课“医用统计方法”课程而编写的教材,主要目的在于提高在校研究生利用统计软件分析和处理数据的能力,同时也为相关医学科研人员提供数据管理与处理分析的入门教材和参考资料。

以往出版的卫生统计学教材,偏重于抽象的概念和理论教学者居多,而实际的可操作性较强的很少,目前国内出版的研究生卫生统计学教科书中,能够把上述两个方面都兼顾起来的教材更少。然而,研究生教学的重点是提高研究生的科研设计能力、计算分析能力及结果解释能力。因此,编写一本适应高技术发展要求,理实交融,适合研究生层次教学需要的医用统计学教材十分必要和迫切。本书以人民卫生出版社出版的《卫生统计学》教材为主线,在不打破传统理论教学框架的情况下,把 SPSS 软件实践教学内容纳入到研究生的教学中,一方面要求掌握基本的统计知识,理解各种计算步骤和统计指标的含义,另一方面要求使用者能够做到理论结合实际,重点提高研究生科研设计和统计处理分析的能力,为在校研究生更好地学习卫生统计学提供一本有价值的教材,为从事科学研究的科研工作者提供一个可以参考的学习指南。

该书由安徽医科大学公共卫生学院流行病与卫生统计学系牵头,组织安徽医科大学、皖南医学院、蚌埠医学院、安徽中医学院、广东医学院等兄弟医学院校具有丰富研究生教学经验的卫生统计学专家参加编写,同时邀请了中国人民解放军第四军医大学、中山大学等院校著名卫生统计学专家编写相关章节,并对教材的目录及章节内容提出了许多宝贵意见。专家们利用寒暑假和其他休息时间,精心编写,字斟句酌,将自己平时积累的宝贵教学、实践经验和心血倾注到书稿编写中,其敬业精神、创新努力和严格的科学态度实在难能可贵,令人折服。

我认为该书无论是在内容的框架的完备性、理论体系结构的逻辑性、知识的



严密性、实际案例的实践指导性方面,还是在提高研究生科研能力和新技术的应用能力等方面,都是一本难得的好教材、好参考书,值得在校研究生及各类医学科研工作者参考学习。

叶冬青

2012年6月16日



前 言

卫生统计学是医学研究生的公共必修课。在医学领域里,从事流行病学调查、临床基础研究、实验研究、药物疗效评价、社区卫生服务等各项工作,都离不开科研设计和数据的统计分析,如果辛苦获得的数据,由于采用的统计学方法错误而导致统计结论错误,那将是非常可惜的事情。医学研究生是各类医学院校科学研究的主要潜在力量,在医学研究中发挥着重要作用,如何提高该群体的科研设计和数据统计分析能力,是我们从事医学统计学教学的老师和科研工作者一直关注的课题。

目前图书市场上针对面向医学研究生开设的卫生统计学课程编写的教材和专著不少,但其中多数主要侧重于让学生掌握卫生统计学的基本原理和方法,而强调科研设计和统计结果分析与处理的却很少,尤其是与统计软件结合的内容更少,有的教材中虽然附有 SAS 软件运行程序,但这些内容对于非统计专业的研究生来说很难深入掌握,因而形同虚设。此外,关于 SPSS 软件应用的教程也不少,且内容非常好,但没有和“卫生统计学”课程教材的章节相对应,学生学习起来针对性不强,效果不好。很多院校的统计学老师都为在医学研究生卫生统计学教学方面选不到合适的教材和参考书而苦恼。鉴于此,我们组织安徽省几所医学院校从事卫生统计学教学的一线教师编写了《医用统计方法及其 SPSS 软件实现》一书,并邀请了国内部分重点院校的教师参与编写相关章节。主要目的是提高在校研究生利用统计软件分析和处理数据的能力,同时也为相关医学科研人员提供数据管理与分析处理的入门教材和参考资料。

本书共 16 章,包括:绪论,SPSS 软件数据管理及常用菜单功能,统计描述与变量分布,抽样误差,假设检验,定量资料的 t 检验,方差分析,定性资料的统计推断,统计图表,有序定性资料的分析方法,秩和检验,简单线性相关与回归,调查设计,几种特殊实验设计方案及分析方法,医学多元统计分析及同类研究结果的 Meta 分析等。本书以传统的卫生统计学教材为主线,在不打破传统理论教学框架的情况下,把 SPSS 软件实践教学内容纳入到研究生的教学中,在每个章节中针对每一个统计学方法都给出了相对应的 SPSS 软件菜单操作指南及主要的统计分析结果与解释,便于学生系统学习和掌握,特别是增加了有序资料的分析方法和



常见实验设计方案及相应的统计学分析方法等诸多内容。

在本书的编写过程中,作者参考了大量的文献、资料。与同类图书相比,本书既保持了传统卫生统计学教材的理论框架,又体现了体系创新,特别突出了教材的实用性,即把抽象的理论和实践结合在一起,我们相信这本书对于提高研究生科研设计水平及正确选择统计学方法的能力具有很好的指导作用。另外,本教材的出版得到安徽省重点学科和安徽医科大学卫生统计学精品课程项目的支持和资助(编号 jpkc201101)。

在本书即将出版之际,我衷心地感谢安徽医科大学叶冬青教授、第四军医大学夏结来教授、中山大学郝元涛教授,他们在百忙中挤出时间认真审阅了本书的全部内容,提出了许多宝贵的修改意见和建议,叶冬青教授还欣然为本书作序,对本书的出版给予了大力的支持和热情的鼓励;安徽医科大学、蚌埠医学院、皖南医学院、安徽中医学院、广东医学院等兄弟院校具有丰富研究生教学经验的卫生统计学专家能够抽出宝贵时间参加编写,他们利用寒暑假和其他休息时间,精心编写,字斟句酌,将自己的心血和平时所积累的宝贵的教学与实践经验都倾注到本书的编写中,在此一并表示诚挚的感谢。最后,我还要特别感谢我的家人和所有参编作者,没有他(她)们对我们的理解和支持,本书是很难保证如期出版的。

尽管我们做了很大的努力,争取使该书成为一部理实交融的成功之作,但限于知识创新和学科交叉与综合的能力,书中难免存在瑕疵和疏漏,恳请同行专家、学者及读者不吝赐教,以便本书在将来再版时更臻完美。

潘发明

2012年7月于合肥



目 录

序言	(i)
前言	(iii)
第 1 章 绪论	(1)
1.1 统计学的产生和发展	(1)
1.2 医学统计学在医学科研中的地位	(3)
1.3 医学统计学的基本概念	(3)
1.4 医学统计工作的基本步骤	(7)
第 2 章 SPSS 软件数据管理及常用菜单功能	(10)
2.1 SPSS 软件简介	(10)
2.2 SPSS 软件的数据库建立	(13)
2.3 常用菜单功能介绍	(20)
2.4 SPSS 结果窗口的使用与编辑	(30)
第 3 章 统计描述与变量分布	(36)
3.1 频数分布表与频数分布图	(36)
3.2 定量资料的统计指标	(38)
3.3 定性资料与等级资料的统计指标	(44)
3.4 正态分布及其应用	(47)
3.5 二项分布和 Poisson 分布及其应用	(50)
3.6 SPSS 软件实现	(54)
第 4 章 抽样误差	(70)
4.1 抽样误差的概念	(70)
4.2 均数的抽样误差及标准误	(71)
4.3 t 分布	(72)
4.4 频率的抽样误差及标准误	(74)
4.5 均数和率的参数估计	(75)
4.6 SPSS 软件实现	(77)



第 5 章 假设检验	(92)
5.1 假设检验的意义及基本原理	(92)
5.2 假设检验的基本步骤	(93)
5.3 假设检验的正确应用	(94)
第 6 章 定量资料的 t 检验	(96)
6.1 单样本均数 t 的检验	(96)
6.2 配对样本均数的 t 检验	(97)
6.3 两独立样本均数的 t 检验	(99)
6.4 假设检验的两类的两类错误	(103)
6.5 假设检验应注意的问题	(104)
6.6 t 检验的 SPSS 软件实现方法	(104)
第 7 章 方差分析	(123)
7.1 完全随机设计资料的方差分析	(123)
7.2 随机区组设计资料的方差分析	(125)
7.3 多组均数比较方差齐性检验	(126)
7.4 多组均数间的两两比较	(127)
7.5 析因设计和重复测量设计资料的方差分析	(128)
7.6 SPSS 软件实现.....	(132)
第 8 章 定性资料的统计推断	(148)
8.1 成组设计四格表资料的 χ^2 检验.....	(148)
8.2 配对设计四格表资料的 χ^2 检验.....	(153)
8.3 成组设计行乘列表资料的 χ^2 检验.....	(155)
8.4 似然比检验和确切概率法	(157)
8.5 χ^2 检验的多重比较	(161)
8.6 不同设计类型 χ^2 检验 SPSS 软件实现	(163)
第 9 章 统计图表	(195)
9.1 统计表	(195)
9.2 统计图	(198)
9.3 SPSS 软件实现.....	(212)
第 10 章 有序定性资料的分析方法	(231)
10.1 单向有序行 \times 列表数据的分析.....	(231)
10.2 双向有序属性相同行 \times 列表数据的分析.....	(235)
10.3 双向有序属性不同行 \times 列表数据的分析	(237)



10.4	SPSS 软件实现	(239)
第 11 章	秩和检验	(251)
11.1	配对设计资料的秩和检验	(251)
11.2	两组定量数据比较的秩和检验	(254)
11.3	多组定量数据比较的秩和检验	(257)
11.4	配伍组设计资料的秩和检验	(261)
11.5	不同设计类型秩和检验 SPSS 软件实现	(263)
第 12 章	简单线性相关与回归	(278)
12.1	线性相关	(278)
12.2	等级相关	(282)
12.3	线性回归	(283)
12.4	线性回归与相关的区别及联系	(289)
12.5	SPSS 软件实现	(290)
第 13 章	调查设计	(301)
13.1	调查研究的分类	(301)
13.2	调查设计	(302)
13.3	调查表的设计与考评	(306)
13.4	敏感问题调查方法	(310)
13.5	调查的质量控制	(312)
13.6	SPSS 软件实现	(314)
第 14 章	几种特殊实验设计方案及分析方法	(323)
14.1	实验设计的基本原则和基本要素	(323)
14.2	交叉设计及其方差分析	(327)
14.3	拉丁方设计及其方差分析	(331)
14.4	正交试验设计及其方差分析	(332)
14.5	临床试验设计及其统计分析	(337)
14.6	临床随访研究及其统计分析	(340)
14.7	SPSS 软件实现	(344)
第 15 章	医学多元统计分析	(366)
15.1	多元线性回归模型	(366)
15.2	二分类 Logistic 回归模型	(373)
15.3	Cox 比例风险回归模型	(378)
15.4	SPSS 软件实现	(384)



第 16 章 同类研究结果的 Meta 分析	(402)
16.1 Meta 分析的基本思想	(402)
16.2 定量资料的 Meta 分析	(402)
16.3 二分类资料的 Meta 分析	(407)
附录 卫生统计学附表	(415)
附表 1 标准正态分布曲线下的面积, $\Phi(-z)$ 值	(416)
附表 2 t 界值表	(417)
附表 3 百分率的可信区间(包括附表 3(1)~3(2))	(419)
附表 4 F 界值表(方差齐性检验界值表)(包括附表 4(1)~4(2))	(425)
附表 5 F 界值表(方差分析界值表)	(427)
附表 6 q 界值表(Student-Newman-Keuls 法)	(436)
附表 7 Dunnett- t 界值表(包括附表 7(1)~7(2))	(438)
附表 8 χ^2 界值表	(442)
附表 9 配对 T 界值表(配对比较的符号秩和检验用)	(444)
附表 10 成组设计 T 界值表(两组比较的秩和检验用)	(446)
附表 11 多组设计 H 界值表(三组比较的秩和检验 Kruskal-Wallis 法)	(448)
附表 12 M 临界值表(随机区组比较的秩和检验用)	(450)
附表 13 相关系数临界值表	(451)
附表 14 r_s 界值表	(454)



第1章 绪论

1.1 统计学的产生和发展

统计学的产生与统计实践活动是密不可分的,统计作为一种社会实践活动,已有四五千年的历史。早在原始社会,人们按部落居住在一起,打猎、捕鱼后就要算算有多少人、多少食物,以便分配食物;我国夏禹时代就有了人口数据的记载;为了赋税、徭役和兵役的需要,历代都有田亩和户口的记录。而统计学的理论和方法,则是在长期统计实践活动的基础上发展起来的,距今已有三百多年的历史。从统计学的发展过程来看,可以大致分为三个阶段。

1.1.1 古典统计学时期

17世纪中叶至18世纪中叶是古典统计学时期,在这一时期,统计学理论初步形成了一定的学术派别,主要有国势学派和政治算术学派。

1. 国势学派

国势学派又称记述学派,产生于17世纪的德国。由于该学派主要以文字记述国家的显著事项,故称记述学派。其主要代表人物是海爾曼·康令(Hermann Conring, 1606~1681)和高特弗里特·阿亨瓦尔(Gottfried Achenwall, 1719~1772)。康令于1660年把国势学从法学、史学和地理学等学科中独立出来,在大学中讲授“实际政治家所必需的知识”;阿亨瓦尔在哥廷根大学开设“国家学”课程,其主要著作是《近代欧洲各国国势学纲要》,书中讲述“一国或多数国家的显著事项”,主要用对比分析的方法研究了国家组织、领土、人口、资源财富和国情国力,比较了各国实力的强弱,为德国的君主政体服务。因在德文中“国势”与“统计”词义相通,后来正式命名为“统计学”。国势学派只是对国情的记述,偏重事物性质的解释,未能进一步揭示社会经济现象的规律,也不研究事物的计量分析方法,不注重数量对比和数量计算,只是用比较级和最高级的词汇对事物的状态进行描述。所以,人们也把它叫做记述学派(旧学派或德国学派),并认为国势学派有统计学之名而无统计学之实。

2. 政治算术学派

政治算术学派产生于17世纪中叶的英国,创始人为威廉·配第(William Petty, 1623~1687),其代表作是他于1676年完成的《政治算术》一书,这本书是经济学和统计学史上的重要著作,这里的“政治”是指政治经济学,“算术”是指统计方法。在这部书中,他利用实际资料,运用数字、重量和尺度等定量分析工具对英国、法国和荷兰三国的国情国力,作了系统的数量对比分析,其所采用的方法是前所未有的,为统计学的形成和发展奠定了方法论基础。因此马克



思说：“威廉·配第——政治经济学之父，在某种程度上也是统计学的创始人。”政治算术学派的另一个代表人物是约翰·格朗特(John Graunt, 1620~1674)，他以 1604 年伦敦教会每周一次发表的“死亡公报”为研究资料，在 1662 年发表了《关于死亡公报的自然和政治观察》的论著。书中通过大量观察发现了人口各年龄组的死亡率、性别比例等重要数量规律，并对人口总数进行了较为科学的估计，第一次编制了“生命表”，对死亡率与人口寿命作了分析，从而引起了人们的普遍关注，因此，他被认为是人口统计学的创始人。

1.1.2 近代统计学时期

18 世纪末至 19 世纪末是近代统计学时期，在这个时期，各种学派的学术观点已经形成，并且形成了两个主要学派，即数理统计学派和社会统计学派。

1. 数理统计学派

在 18 世纪，概率理论日益成熟，为统计学的发展奠定了基础。19 世纪中叶，概率论被引进统计学，从而形成数理学派，其奠基人是比利时的阿道夫·凯特勒(Lambert Adolphe Jacques Quetelet, 1796~1874)，他在其著作《社会物理学》中将古典概率论引入统计学，使统计学进入一个新的发展阶段。他认为概率论是适于政治及道德科学中以观察与计数为基础的方法，并以此方法对自然现象和社会现象的规律性进行观察，并认为要促进科学的发展，就必须更多地应用数学。总之，他把概率论引入统计学，为数理统计学的形成与发展奠定了基础。

2. 社会统计学派

社会统计学派产生于 19 世纪后半叶，创始人是德国经济学家、统计学家克尼斯(K. G. A. Knies, 1821~1898)，主要代表人物有厄恩斯特·恩格尔(Christian Lonrenz Ernst Engel, 1821~1896)、乔治·冯·梅尔(Georg von Mayr, 1841~1925)等人。他们融合了国势学派与政治算术学派的观点，沿着凯特勒的“基本统计理论”向前发展，但在学科性质上认为统计学是一门社会科学，是研究社会现象变动原因和规律性的实质性科学，以此同数理统计学派通用方法相对立。社会统计学派在研究对象上认为统计学是研究总体而不是个别现象，而且认为由于社会现象的复杂性和整体性，必须对总体进行大量观察和分析，研究其内在联系，才能揭示现象内在规律。这是社会统计学派的“实质性科学”的显著特点。

1.1.3 现代统计学时期

20 世纪至今为现代统计学时期，这一时期的主要特征是描述统计学已转向推断统计学，1907 年，英国人戈塞特(Gosset, 1876~1937)提出了小样本 t 统计量理论，丰富了抽样分布理论，为统计推断奠定了基础。英国的罗纳德·费雪(R. A. Fisher, 1890~1962)提出了极大似然估计量的概念，迅速地成为估计参数的重要方法，他还提出样本相关系数的分布、实验设计和方差分析等方法。英国科学家弗朗西斯·高尔顿(Francis Galton, 1822~1911)提出了相关与回归思想，并给出计算相关系数的明确公式。英国统计学者 K·皮尔逊(K. Pearson, 1875~1936)发展了拟合优度检验，还给出了卡方统计量及其极限分布，波兰学者奈曼(J. Neyman, 1894~1981)创立了区间估计理论，并和 K·皮尔逊发展了假设理论。美国学者瓦尔德(Wald, 1902~1950)提出决策理论和序贯抽样方法。美国化学家威尔科克松(Frank Wilcoxon-



on)发展了一系列非参数统计方法,开辟了统计学的新领域。由马哈拉诺比斯领导的印度统计研究所和20世纪30年代后期奈曼发表的两篇论文,使抽样的数学理论在20世纪30年代得到了迅速发展。

统计学大致经过以上三个阶段发展到今天,随着统计学理论知识的发展与健全,统计学的应用领域将会进一步扩大,将出现许多新型的交叉学科,比如统计应用到法律、文学等学科。同时,伴随着计算机技术的飞速发展,统计学还将在模糊现象、突变现象及混沌现象等方面开辟新的研究领域。

1.2 医学统计学在医学科研中的地位

统计学原理和方法几乎应用到自然科学和社会科学的各个领域,目前产生了许多应用性分支学科,诸如社会经济统计学、工业统计学、生物统计学、教育统计学、药物统计学等。医学领域的研究对象主要是人体以及与人体健康相关的各种因素,具有其特殊性,并受到社会、经济和心理等诸多因素的影响。这些影响具有不确定性,必须透过这些不确定性来探测其内部蕴涵的规律性,统计学便有了用武之地。医学统计学(medical statistics)就是运用概率论和数理统计原理、方法结合医药卫生工作的实际情况,阐述医学科研设计的基本原理,研究医学资料(信息)的搜集、整理和分析的方法学总称,它是认识医学现象数量特征的重要工具。

医学研究生学习统计方法,主要是因为:医学上许多现象(如血压、脉搏、SGPT等生理、生化指标测定)都是随机现象。随机现象广泛存在于生物医学的各个领域,对于这些“随机现象”,由于其不确定性,只有借助概率论原理,运用统计学方法,帮助我们透过偶然性来认清事物内部潜在的客观规律。现在可获得的书刊资料很多,如何识别错误信息,必须掌握一定的统计学知识。20世纪70年代以后发展起来的DME(Design, Measurement and Evaluation)就是应用统计学原理和分析方法,结合流行病学等相关学科,帮助临床医师阅读文献资料、评价医学文献、开展医学科研和总结工作经验,使统计方法的应用范围更加广泛。研究生通过本学科的学习,充分认识研究设计、收集原始资料及借助医学统计学的原理和方法进行统计推断的重要性,有助于培养医学研究生在科研中发现问题、分析问题、解决问题的能力,正确掌握医学统计学的原理和方法及其应用是医学科研人员必备的品质。

1.3 医学统计学的基本概念

1.3.1 变量和变量值

统计分析最基本的单位是变量,即观察对象个体的特征或测量的结果。由于个体的特征或指标存在个体差异,在测量前不能准确预测,故称为随机变量(random variable),简称为变量(variable),如患者的年龄、性别、职业等。变量的取值称为变量值或观察值(observation),如实际的年龄、性别等。



1.3.2 同质和变异

一个总体中有许多个体,他们之所以共同成为人们研究的对象,必定存在共性。性质相同的事物称为同质(homogeneity)的,否则称为异质(heterogeneity)。没有同质性就构不成一个总体供人们研究,如不同年龄组男童的身高不能计算平均数,因为所得结果没有意义。

不同研究中、或同一研究中不同观察指标对观察对象的同质性的要求不同,即同质是相对的。例如,男性身高与女性身高有着本质的差别,因此,在考虑身高这一指标时,不能把不同性别的人混在一起,此时,不同性别表示不同质;而在研究白细胞计数这一指标时,因性别对该指标没有影响,故可以把不同性别的人放在一起分析。又如,在某新药的临床试验中,计算有效率的观察病例必须患同一疾病,甚至具有相同的病型、病情、病程等,对同质性的要求是很严格的;而计算不良反应发生率,通常可将不同病种的病例合起来统计,此时对同质性的要求只有一条:按规定服用该新药。

宇宙中的事物千差万别,各不相同,即使是同质事物,就某一观察指标来看,各观察单位(亦称个体)之间也有差别,这种同质事物间的差别称为变异(variation)。例如,研究儿童的身体发育,同性别、同年龄儿童的身高,有高有矮,各不相同,称为身高的变异。由于观察单位通常是观察个体,故变异亦称个体变异(individual variation)。变异表现在两个方面:其一,个体与个体间的差别;其二,同一个体重复测量值间的差别。变异是宇宙事物的个性反映,在生物学和医学现象中尤为明显。

变异是由于一种或多种不可控因素(已知的和未知的)以不同程度、不同形式作用于生物体的综合表现。如果我们掌握了所有因素对生物体的作用机制,那么,生物体的某指标之观察值就是可预测的了。有些指标的变异原因已被人们认识,例如,染色体决定了新生儿的性别;有些指标的变异原因已被认识一部分,比如,人的身高受遗传和后天营养的影响,但尚有一部分影响因素是未知的;更多的情况下,影响变异的因素是未知的。就每个观察单位而言,其观察指标的变异是不可预测的,或者说是随机的。观察指标用变量(variable)表述。当观察值的个数达到足够多时,其分布将趋于稳定,并最终服从于总体分布(distribution of population)。

个体变异现象广泛存在于人体及其他生物体,是个性的反映。虽然每个个体的变异表现出一定的随机性和不可预测性,但变异并不等于杂乱无章,指标的变异是有规律的,当所观察的个体数足够多时,观察值的分布将呈现一定的规律性,这是总体的反映。统计学就是探讨变异规律,并运用其规律性进行深入分析的一门学科。可以这么说,没有变异就没有统计学。

1.3.3 总体和样本

总体(population)是根据研究目的所确定的同质观察单位的全体,确切地说,是同质的所有观察单位某种变量值的集合;个体(individual)是构成总体的最基本的观察单位;样本(sample)是从总体中随机抽取部分观察单位,其变量值的集合;样本中所包含的个体数称为样本含量(sample size)。

例如,调查某地某年正常成年男子的血红蛋白水平,则观察对象是该地的正常成年男子,全部正常成年男子构成了研究总体(study population),其同质基础是同一地区,同一年份,同