



全国医学成人高等教育专科规划教材

医学统计学

YIXUE TONGJIXUE

主编/唐 军 崔香淑

(第3版)



人民军医出版社

PEOPLE'S MILITARY MEDICAL PRESS

全国医学成人高等教育专科规划教材

医学统计学

YIXUE TONGJIXUE

(第3版)

主 编 唐 军 崔香淑
副主编 赵拥军 文育锋 桂立辉
编 者 (以姓氏笔画为序)
王守英 文育锋 刘凤阁
吴善玉 张合喜 张红杰
周纯先 赵拥军 袁尚华
桂立辉 唐 军 崔香淑



人民军医出版社

PEOPLE'S MILITARY MEDICAL PRESS

北 京

图书在版编目(CIP)数据

医学统计学/唐 军,崔香淑主编. —3 版. —北京:人民军医出版社,2009.12

全国医学成人高等教育专科规划教材

ISBN 978-7-5091-3069-8

I. 医… II. ①唐…②崔… III. 医学统计—成人教育:高等教育—教材 IV. R195.1

中国版本图书馆 CIP 数据核字(2009)第 201592 号

策划编辑:杨磊石 文字编辑:谢秀英 责任审读:张之生

出 版 人:齐学进

出版发行:人民军医出版社

经销:新华书店

通信地址:北京市 100036 信箱 188 分箱

邮编:100036

质量反馈电话:(010)51927290;(010)51927283

邮购电话:(010)51927252

策划编辑电话:(010)51927292

网址:[www. pmmp. com. cn](http://www.pmmp.com.cn)

印刷:北京天宇星印刷厂 装订:京兰装订有限公司

开本:787mm×1092mm 1/16

印张:10.75 字数:254 千字

版、印次:2009 年 12 月第 3 版第 1 次印刷

印数:52101~58100

定价:20.00 元

版权所有 侵权必究

购买本社图书,凡有缺、倒、脱页者,本社负责调换

全国医学成人高等教育专科规划教材

(第3版)

编审委员会名单

主任委员 文历阳 毛兰芝 王庸晋

常务副主任委员 金青松 姚磊 周海兵

副主任委员 (以姓氏笔画为序)

尹卫东	石增立	朱启华	朱漱玉	李贞保
李佃贵	李朝品	杨宝胜	宋国华	张纯洁
陈健尔	金秀东	武秋林	赵富玺	唐世英
常唐喜				

委员 (以姓氏笔画为序)

万新顺	王子寿	王长虹	王建立	王桂云
王庸晋	丰慧根	牛春雨	申保生	申素芳
玄云泽	玄英哲	戎华刚	刘凤芹	刘恒兴
刘新民	关利新	安丰生	李伟扬	李佃贵
李朝品	杨金香	宋景贵	张文彬	张忠元
张承刚	张洪福	范忆江	金政	金东洙
金秀东	金顺吉	金哲虎	赵卫星	赵志梅
赵富玺	栾希英	郭学鹏	席鸿钧	唐军
崔香淑	崔新宇	盖立起	梁玉	彭力辉
韩春姬	魏武			

编辑办公室 郝文娜 杨磊石 秦速励 徐卓立

全国医学成人高等教育专科规划教材

(第3版)

教材目录

1 医用化学	主编 杨金香等	18 外科学	主编 席鸿钧等
2 医学遗传学	主编 丰慧根等	19 妇产科学	主编 申素芳等
3 系统解剖学	主编 金东洙等	20 儿科学	主编 郭学鹏等
4 局部解剖学	主编 刘恒兴等	21 传染病学	主编 申保生等
5 组织胚胎学	主编 金政等	22 眼科学	主编 万新顺等
6 生物化学	主编 王桂云等	23 耳鼻咽喉科学	主编 金顺吉等
7 生理学	主编 金秀东等	24 口腔科学	主编 玄云泽等
8 病理学	主编 赵卫星等	25 皮肤性病学	主编 金哲虎等
9 病理生理学	主编 牛春雨等	26 神经病学	主编 宋景贵等
10 药理学	主编 关利新等	27 精神病学	主编 王长虹等
11 医学微生物学	主编 赵富玺等	28 急诊医学	主编 魏武等
12 医学免疫学	主编 栾希英等	29 影像诊断学	主编 赵志梅等
13 人体寄生虫学	主编 李朝品等	30 中医学	主编 李佃贵等
14 预防医学	主编 韩春姬等	31 医学心理学	主编 刘新民等
15 医学统计学	主编 唐军等	32 医学伦理学	主编 张忠元等
16 诊断学	主编 李伟扬等	33 卫生法学概论	主编 崔新宇等
17 内科学	主编 王庸晋等		

全国医学成人高等教育专科规划教材

(第3版)

修订说明

《全国医学成人高等教育专科规划教材》是全国第一套医学成人高等教育教材,第1版于1997年出版,第2版于2003年出版。本套教材出版以来在众多学校和师生的热情关心和支持下,已经逐步成为在全国具有影响力的品牌教材。人民军医出版社对所有在本套教材出版和推广过程中给予大力支持和帮助的相关院校,尤其是曾在第1版、第2版教材出版中做出贡献的编写专家们表示深切的感谢。

本套教材的第2版出版6年来,随着医学领域科技的迅速发展,成人教育开办的教学方针和招生规模都有了很大的变化,教师队伍也有部分新老更替,为了使我们的教材与时俱进,更加体现现代医学“以人为本”的教育理念,体现当前教学改革的新方法、新思路,及时补充修订一些新知识、新进展、新标准,我们决定组织修订出版第3版。

第3版的修订再版工作从2009年3月开始,遵照“延续品牌、调整作者、提升质量”的原则进行,共有20余所院校的上百位老师参加了编写工作。第3版编审委员会主任由我国著名的医学教育家文历阳校长、新乡医学院的毛兰芝院长和长治医学院的王庸晋院长共同担任。参编单位主要有新乡医学院、长治医学院、延边大学医学部、牡丹江医学院、皖南医学院、蚌埠医学院、安徽理工大学医学院、滨州医学院、成都中医药大学、承德医学院、河北北方学院、大同大学医学院、河北医科大学、河北大学医学部、河南职工医学院、潍坊医学院、漯河医学高等专科学校、南阳医学高等专科学校、盐城卫生职业技术学院、宁波天一职业技术学院、赣州卫生学校、河南省卫生学校、焦作中医药学校等。大家本着“共同参与,共同建设,共同受益”的方针,认真遴选出各书主编,精心组织作者队伍,讨论落实了编写大纲,有序展开了相关工作。

现在,在出版社和有关院校与老师们的共同努力下,《全国医学成人高等教育专科规划教材(第3版)》共33本正式出版了。希望本套教材能在医学成人高等教育中为我国卫生事业的发展输送更多合格人才,发挥出更多更好的作用,也希望有关院校和广大师生们在使用中多提宝贵意见,以利本套教材的进一步成熟提高。

人民军医出版社

2009年10月

第 3 版前言

2009 年 4 月在河南省新乡市召开了全国医学成人高等教育专科教材编审委员会会议。根据会议精神和要求,我们对第 2 版《医学统计学》教材进行了修订,编写了这本第 3 版《医学统计学》。

本版教材总结吸取了第 2 版的编写经验,继续坚持突出了“三基”(基本知识、基本理论和基本技能)训练,加强统计思维方法和能力的培养,贯彻“五性”(思想性、科学性、先进性、启发性和实用性)原则,同时,照顾到成人高等医学教育的特点与规律。在第 2 版基础上,根据医学统计学学科的发展和教学需要,适当增补了新内容,采用最新资料和研究成果,反映了医学统计学的前沿动态。

在内容安排上,本书按先易后难、循序渐进的原则进行编排,注重构建科学的医学统计学知识体系。全书共分 10 章,即绪论、计量资料的统计描述、计量资料的统计推断、计数资料的统计描述、计数资料的统计推断、秩和检验、直线相关与回归、医学科研设计、统计表与统计图、综合评价。计量资料统计推断中的方差分析和第 10 章综合评价可作为自学内容,目的在于丰富和扩展学生的视野。每章后面附有各种类型的练习题,有利于活跃思想,培养学生智能,使本书实用性更强。附录中除计算器使用方法和统计用表外,增加了常用统计软件介绍和统计图表的制作方法,以供学生参考。医学统计学为考试课程,第三学期开课,采取理论与实践相结合的教学方法,按 36 学时教学工作量设计,理论与实习课的比例为 1:0.6。

本书既是医学统计学的入门教材,也可作为广大医务工作者进行继续教育或业余自学的参考书。

本次修订,除吸收第 1、2 版的编写经验外,还引用了有关教材及专著的部分资料,在此一并致以衷心感谢。

本书在编写过程中,得到了人民军医出版社和滨州医学院领导的大力支持,特此致谢。

由于编者水平有限,时间仓促,书中缺点和错误在所难免,恳请使用本书的师生和广大读者批评指正。

编 者

2009 年 8 月

目 录

第1章 绪论.....	(1)	二、几何均数	(9)
第一节 医学统计学的概念及基本		三、中位数和百分位数	(11)
内容	(1)	第三节 离散趋势的指标	(12)
一、医学统计学的概念	(1)	一、全距	(12)
二、医学统计学的研究内容	(1)	二、四分位数间距	(12)
第二节 医学统计学中的几个基本		三、方差	(13)
概念	(2)	四、标准差	(14)
一、同质与变异	(2)	第四节 正态分布及其应用	(16)
二、总体与样本	(2)	一、正态分布的概念	(16)
三、参数与统计量	(2)	二、正态分布的特征	(16)
四、抽样误差	(2)	三、正态分布的应用	(17)
五、随机化	(3)	第3章 计量资料的统计推断	(21)
六、概率	(3)	第一节 均数的抽样误差与标准误	
第三节 医学统计资料的类型	(3)		(21)
一、计量资料	(3)	一、均数的抽样误差	(21)
二、计数资料	(3)	二、标准误	(21)
三、等级资料	(4)	第二节 t 分布	(22)
四、不同类型资料的转化	(4)	一、 t 分布的概念	(22)
第四节 医学统计工作的基本步骤		二、 t 界值表	(22)
.....	(4)	第三节 总体均数的估计	(23)
一、统计设计	(4)	一、可信区间的概念	(23)
二、搜集资料	(5)	二、总体均数可信区间的计算	(24)
三、整理资料	(5)	第四节 假设检验的基本步骤	(25)
四、分析资料	(5)	一、假设检验的基本思想	(25)
第2章 计量资料的统计描述	(6)	二、假设检验的基本步骤	(26)
第一节 计量资料的频数表	(6)	第五节 t 检验	(27)
一、频数表的概念	(6)	一、样本均数与总体均数比较的 t	
二、频数表的编制	(6)	检验	(27)
第二节 集中趋势的描述指标	(8)	二、配对计量资料的 t 检验	(28)
一、算术平均数	(8)	三、两样本均数比较的 t 检验	(30)



四、 t 检验时应注意的问题	(33)	(Wilcoxon 配对法)	(67)
第六节 方差分析	(34)	一、检验步骤	(67)
一、完全随机设计资料的方差分析	(34)	二、本法的基本思想	(69)
二、随机区组设计资料的方差分析	(37)	第二节 两样本比较的秩和检验 (Wilcoxon 两样本比较法)	(69)
三、多个样本均数间的两两比较	(40)	一、原始数据的两样本比较的秩和检验	(69)
第 4 章 计数资料的统计描述	(45)	二、频数表资料(或等级资料)的两样本比较秩和检验	(70)
第一节 相对数	(45)	第三节 多个样本比较的秩和检验	(71)
一、计数资料的频数分布	(45)	一、直接法	(71)
二、常用的相对数指标	(45)	二、频数表法	(73)
三、临床工作中常用的相对数指标	(47)	第四节 多个样本间两两比较的秩和检验	(73)
四、应用相对数时的注意事项	(50)	一、各样本例数相等的两两比较秩和检验	(74)
第二节 率的标准化法	(51)	二、各样本例数不等或不全相等的两两比较的秩和检验	(74)
一、率的标准化法的意义	(51)	第 7 章 直线相关与回归	(78)
二、标准化法中标准的选择	(51)	第一节 直线回归	(78)
三、标准化率的计算	(52)	一、直线回归的概念	(78)
第 5 章 计数资料的统计推断	(56)	二、直线回归方程的计算	(79)
第一节 率的抽样误差与标准误	(56)	三、回归系数的假设检验	(80)
一、率的抽样误差与标准误的概念	(56)	四、直线回归的区间估计	(82)
二、总体率的估计	(56)	五、回归方程的应用	(82)
第二节 u 检验	(57)	六、应用直线回归应注意的问题	(82)
一、样本率与总体率比较的 u 检验	(57)	第二节 直线相关	(83)
二、两个样本率比较的 u 检验	(58)	一、直线相关的概念	(83)
第三节 卡方检验	(59)	二、相关系数的意义及计算	(83)
一、四格表资料的 χ^2 检验	(59)	三、相关系数的假设检验	(84)
二、四格表的专用公式	(60)	第三节 直线相关与回归的区别与联系	(84)
三、四格表资料 χ^2 值的连续性校正	(60)	一、区别	(84)
四、配对计数资料的 χ^2 检验	(61)	二、联系	(85)
五、行 $\times$ 列表的 χ^2 检验	(62)	第四节 等级相关	(85)
第四节 多个样本率间多重比较——行 $\times$ 列表的 χ^2 分割	(63)	第 8 章 医学科研设计	(88)
第 6 章 秩和检验	(67)		
第一节 配对资料的符号秩和检验			



第一节 医学科研设计的基本原则	(88)	一、对评价指标(因子)的要求 ...	(112)
一、对照的原则	(88)	二、评价指标的分类	(112)
二、均衡的原则	(89)	第四节 评价指标的筛选	(114)
三、随机的原则	(89)	一、经验选择法	(114)
四、重复的原则	(90)	二、单因素选择法	(115)
第二节 实验设计的基本要素	(91)	三、多因素选择法	(116)
一、受试对象	(91)	第五节 综合评分法	(116)
二、处理因素	(91)	一、各评价指标分值的确定	(117)
三、实验效应	(92)	二、综合评价总分的计算	(117)
第三节 几种简单的实验设计方法	(92)	三、综合评分法实例	(118)
一、完全随机设计	(92)	第六节 综合指数法	(120)
二、配伍组设计	(92)	一、指数分类及计算步骤	(120)
三、盲法设计	(93)	二、应用实例	(121)
第四节 调查设计	(93)	附录 A 计算器使用方法	(123)
一、调查设计的意义	(93)	附录 B 统计用表	(127)
二、调查设计的内容	(94)	附表 B-1 标准正态分布曲线下的面积	(127)
三、调查的方法	(94)	附表 B-2 t 界值表	(129)
第 9 章 统计表和统计图	(97)	附表 B-3 F 界值表(方差分析用)	(130)
第一节 统计表	(97)	附表 B-4 q 界值表(Newman-Keuls 检验用)	(134)
一、制表原则和基本要求	(97)	附表 B-5 百分率的可信区间	(135)
二、统计表的种类	(98)	附表 B-6 χ^2 界值表	(138)
三、统计表常见错误分析	(98)	附表 B-7 T 界值表(配对比较的符号秩和检验用)	(139)
第二节 统计图	(99)	附表 B-8 T 界值表(两样本比较的秩和检验用)	(140)
一、统计图绘制的基本要求	(99)	附表 B-9 H 界值表(三样本比较的秩和检验用)	(142)
二、统计图的绘制方法	(99)	附表 B-10 D 界值表(各样本例数相等的 Nemenyi 法用)	(143)
第 10 章 综合评价	(107)	附表 B-11 r 界值表	(144)
第一节 综合评价的概念与内容	(107)	附表 B-12 r_s 界值表	(146)
一、综合评价的概念	(107)	附表 B-13 随机数字表	(147)
二、综合评价的内容	(107)	附表 B-14 随机排列表($n=20$)	(148)
三、综合评价的条件	(108)	附录 C 常用统计软件简介	(149)
四、综合评价信息的收集	(108)	附录 D 用 Excel 软件绘制统计图的方法	(157)
第二节 信息的预处理	(109)		
一、原始数据的审核、汇总和储存	(109)		
二、清除不合理的离群值	(109)		
第三节 评价指标及其选择	(112)		

第一节 医学统计学的概念及基本内容

一、医学统计学的概念

医学统计学(medical statistics)是运用数理统计、概率论的原理和方法于医学科研和实践,研究医学资料 and 信息的搜集、整理和分析的一门应用学科。用数量反映质量和利用样本推断总体是统计方法的两个特点。医学研究的对象主要是人体及与人体健康有关的因素。由于生物现象变异性较大,各种影响因素错综复杂,生物学因素、环境因素、心理因素、社会行为因素、医疗保健因素等均可影响人体及人体健康。例如用某药治疗某病,尽管年龄、性别、病情、病程均相同,其治疗效果也不一样;同年龄、同性别的人群,其身高体重和各种生理生化指标的数值均有所不同,即使同一个人的血压值在昼夜的不同时段也不一样,这些都是个体存在变异(差异)的结果。为了预防疾病,研究病因,促进健康,必须运用医学统计学方法透过偶然现象来探测其规律性,得出科学推断。因而医学统计学是进行医学科学研究的重要手段,是医学生认识世界的重要工具。电子计算机技术的发展、普及和应用,为医学统计学的应用和发展提供了广阔的前景。

二、医学统计学的研究内容

医学统计学的基本内容包括以下4个方面:

1. 医学统计学的基本原理和方法 包括科学、准确而及时地搜集、整理统计资料的方法,反映事物实际情况和特征的统计指标,反映事物规律性的数据处理方法,科学地进行描述、估计、比较和预测等,计量、计数资料的统计描述及推断,直线相关与回归等(详见第1~第7章),这既是本书的主要内容,也是要掌握的基本内容。

2. 医学统计设计 包括科研设计的原则、要素及实验设计方法、调查设计方法等。见本书医学科研设计章节。

3. 居民健康统计 包括研究发病情况的疾病统计,研究人口数量、人口构成、出生、死亡等的人口统计,以及研究人群发育水平的发育统计等。

4. 医用多因素分析 包括多元回归、逐步回归、Logistic 回归分析、判别分析、聚类分析、主成分分析等,属于比较高级的统计分析方法。



第二节 医学统计学中的几个基本概念

一、同质与变异

同质(homogeneity)是指事物的性质、影响条件或背景相同或相近。例如,研究某种新药对肝癌的疗效,所有研究对象都必须是确诊为肝癌的患者,不可包括疑似患者或根本不是肝癌的患者在内。要求研究对象是同质的,否则研究结果就不能反映新药对治疗肝癌疗效的客观规律。再如,研究儿童生长发育情况,应选取同性别、同年龄儿童的身高进行观察。

宇宙中的事物,千差万别,各不相同。即使是性质相同的事物,就同一观察指标来看,各观察单位(亦称个体)之间,也各有差异,称为变异(variation)。如前所述,研究儿童的身体发育,即使同质(同性别、同年龄)儿童的身高,有高有低,各不相同,称为身高的变异。再如同一种药物,同样的剂量治疗同样病种的患者,其疗效亦不一样。

同质观察单位之间的个体变异,是生物的重要特征,是由于生物体内外环境中多种因素的综合影响造成的。统计的任务就是在同质分组的基础上,通过对个体变异的研究,透过偶然现象,反映同质事物的本质特征和规律。

二、总体与样本

总体(population)是根据研究目的所确定的同质的研究对象的全体。更确切地说,是性质相同的所有观察单位某种变量值的集合。总体所包含的个体数,可以为无穷大,而且往往只是设想的或是抽象的,称为无限总体。如果总体只包括有限个体观察单位,称为有限总体。例如,我们对食管癌的防治研究,所有患食管癌的患者都具有患食管癌这个同质的特征,是一个总体,每一个食管癌患者叫做个体。但我们要研究食管癌的规律,事实上并不可能将所有食管癌患者都观察到,而只能对一部分个体进行观察。这种从总体中取出部分个体的过程称为抽样(sampling)。所抽得的部分称为样本(sample)。所谓样本,是根据随机的原则从总体中抽出有代表性的一部分观察单位。该样本所包含的个体数目,称为样本含量或样本大小(sample size)。对样本进行观察,用样本的特征推论总体的特征称为统计推断。

三、参数与统计量

描述总体特征的有关指标,称为参数(parameter),如总体平均数、总体标准差和总体率等,反映样本特征的有关指标,称为统计量(statistic),如样本均数、样本标准差和样本率等。如研究苗族 10 岁男孩身高的情况,把所有苗族 10 岁男孩作为观察对象,计算的身高均数为参数。若进行抽样研究,用随机的方法从总体中抽出一部分苗族 10 岁男孩,计算出的身高均数则称为统计量。

四、抽样误差

由于总体中各观察单位间存在个体变异,抽样研究中抽取的样本,只包含总体的一部分观察单位,因而,样本指标不可能恰好等于相应的总体指标。这种由于抽样而引起的样本指标与总体指标的差异,统计学上称为抽样误差(sampling error)。



一般说来,样本越大,则抽样误差越小,越和总体的情况相接近,用样本推断总体的精确度越高,反之亦然。但在实际工作中,我们往往不可能观察太多的对象,而只能对一个较小的样本资料进行分析研究,计算出一系列的统计指标。由于个体变异是客观存在的,因而抽样误差是不可避免的,但抽样误差是可以控制的。

五、随机化

随机化(randomization)是抽样研究和抽样分配时十分重要的原则。从研究对象的总体中抽取样本进行研究时,为了使样本对总体有较好的代表性,并使其抽样误差大小可以用统计学方法来估计,必须遵循随机化的原则,使总体中的每一个个体都有同等的机会被抽取。如在动物实验中,要按随机化原则使各个动物都有同等机会被分配到各处理组去,只有这样,才能避免人为因素所造成的偏差,以便于用统计方法处理各种数据。因此,统计分析和统计推断方法是根据随机化抽样而建立的。随机化的方法有抽签法和随机数字表法等。

六、概 率

概率(probability),亦称几率,是描述某事件发生的可能性大小的一个量度。如事件 A 发生的可能性大小,用实数来表示,即称为事件 A 的概率,常记为 $P(A)$,或简记为 P 。例如,在大量的临床研究发现,某种新药治疗某病的治愈率为 80%(0.80),该数值说明该药治愈某病的可能性,也就是说该药治愈某病的概率估计为 0.80。

在一定条件下,肯定发生的事件称为必然事件,肯定不发生的事件称为不可能事件,可能发生也可能不发生的事件,称为随机事件或偶然事件。必然事件的概率等于 1,不可能事件的概率等于 0,随机事件的概率介于 0 与 1 之间。概率越接近 1,表明事件发生的可能性越大。概率越接近 0,发生的可能性越小。我们在医学科研论著中常见到 $P \leq 0.05$,或 $P \leq 0.01$,前者表示事件发生的可能性等于或小于 0.05,后者表示发生的可能性等于或小于 0.01,习惯上把 $P \leq 0.05$ 或 $P \leq 0.01$ 称为小概率事件,作为事物差别有统计意义和高度统计意义的界限。

第三节 医学统计资料的类型

医学统计资料可分为计量资料、计数资料和等级资料 3 种类型。不同类型的资料应采用不同的分析方法。

一、计量资料

计量资料(measurement data)亦称定量资料,是对每个观察单位用定量方法测定某项指标量的大小,所得的资料称为计量资料。一般有度量衡单位,如调查某地某年 8 岁男孩的身体发育状况,以人为单位,每个人的身高(cm)、体重(kg)、血压(kPa)等,都属于计量资料。计量资料常用平均数、标准差、标准误等指标进行描述;用 t 检验、方差分析、相关回归等统计方法作资料的比较和分析。

二、计数资料

计数资料(enumeration data)亦称定性资料,是将观察单位按某种属性或类别分组,所得



各组的观察单位数,称为计数资料。如测定儿童身高时可以按年龄、性别分组,然后清点各年龄、性别组的人数。又如调查某民族人群的血型时,可以按 A、B、AB、O 4 个血型统计人数。计数资料常用相对数、率的标准误等指标进行描述;用 u 检验、 χ^2 检验等统计方法作资料的比较和分析。

三、等级资料

等级资料(ranked data)亦称半定量资料,是将观察单位按某种属性的不同程度分组,所得各组的观察单位数,称为等级资料。如临床上将患者按痊愈、显效、好转、无效 4 个疗效等级的人数分组;又如测定某项血清反应的一、 $\pm$ 、 $+$ 、 $++$ 的人数。这类资料与计量资料不同的是,每个观察单位未确切定量。与计数资料不同的是,属性的分组有程度的差别,各组按大小顺序排列。这种资料可以用非参数统计方法进行分析比较。

四、不同类型资料的转化

根据研究目的,计量资料、计数资料和等级资料可以互相转化。例如,每个人的血红蛋白含量,原属计量资料。若按血红蛋白正常与异常分为两组,得各组人数,是计数资料。若将血红蛋白按量(g/L)的多少分为 5 个等级:

血红蛋白含量(g/L)	等级
<60	重度贫血
60~	中度贫血
90~	轻度贫血
120~160	正常
>160	血红蛋白增高

计算各等级人数,就是等级资料。有时为了分析的需要,也可将计数资料或等级资料量化。例如将正常和异常分别取 0、1;将检验结果中的一、 $+$ 、 $++$ 、 $+++$ 、 $++++$ 分别取 0、1、2、3、4,这时计数资料或等级资料就转化为计量资料了。

第四节 医学统计工作的基本步骤

医学统计工作一般分为 5 个步骤,即先有一个精密的设计、然后根据设计的要求搜集资料、整理资料和分析资料。这 5 个步骤是相互联系、前后呼应、不能截然分开的整体,任何一个步骤存在不足都会影响统计结果的质量。

一、统计设计

统计工作和科研工作类似,设计是最关键的第一步,统计设计(statistical design)的意义在于能用比较经济的人力、物力和时间,得到较为科学、准确、可靠的结果。设计之前,应先对研究的问题有较多的了解。为此,不仅需要广泛查阅文献,了解实际情况,而且需要与有关专家共同协作,掌握尽可能多的信息。统计设计是整个研究过程的总体规划,设计的内容包括资料搜集、整理和分析全过程总的设想和安排。良好的设计是顺利进行实验、调查和处理结果的先决条件,也是使实验研究获得预期结果的重要保证。



二、搜集资料

根据统计设计的要求,及时取得准确、完整、可靠的原始数据的整个过程,称为搜集资料(collection of data)。只有原始数据可靠,才能得出科学的结论。因此,这一步骤具有极其重要的基础意义。

医学统计资料的来源主要有以下3个方面:

1. 统计报表 如医院工作月报表、年报表、疫情报表、居民的病伤死因报表等。这些报表根据国家规定的报告制度,由医疗卫生机构定期逐级上报。它较全面地提供居民健康状况和医疗卫生机构工作的主要数字,为拟定卫生工作计划与措施、检查与总结工作提供依据,也给科学研究提供基础资料。

2. 医疗卫生工作记录和报告卡片 如门诊、住院、临床检验、健康检查记录等。病历是医院工作的重要记录,分析时应注意其局限性(如不能反映一般人群特征)。这些资料要做到登记完整、准确、及时。

3. 专题调查或实验 当上述两方面不能满足需要时,经常采用专题调查或实验的方法获得资料(详见第8章)。

三、整理资料

整理资料(sorting data)的任务是净化原始数据,使资料系统化、条理化,便于进一步计算指标和分析。无论是报表、调查和实验的原始数据,都必须进行认真反复检查和核对,注意核查数据是否符合逻辑及记录是否有误,特别是在数据较多时,一定要在修正错误,去伪存真以后再进行处理。原始数据检查无误后,根据资料的属性进行分类,即区别原始资料是计量资料、计数资料或等级资料。然后根据资料的类型,列成整理表。如是计量资料可以汇总成频数表,以便进行指标的计算及进一步进行统计分析。

四、分析资料

分析资料(analysis of data)的任务是按研究设计的要求,计算有关指标,反映数据的综合特征,阐明事物的内在联系和规律。分析资料包括:①统计描述:指用统计指标、统计表、统计图等方法对资料的数量特征及其分布规律进行测定或描述;②统计推断:指如何抽样以及如何由样本信息推断总体特征问题。统计推断是研究人群健康过程中常用的方法。

(唐 军 赵拥军)

【思考题】

1. 名词解释

- | | | | |
|----------|----------|-----------|---------|
| (1)医学统计学 | (2)总体与样本 | (3)参数与统计量 | (4)随机化 |
| (5)概率 | (6)计量资料 | (7)计数资料 | (8)等级资料 |

2. 问答题

- (1)医学统计资料分为哪几类?
- (2)医学统计工作有哪几大步骤?

统计描述即用恰当的统计指标描述资料的数量特征。计量资料的统计描述分为两个方面：集中趋势(central tendency)和离散趋势(tendency of dispersion)。

第一节 计量资料的频数表

一、频数表的概念

频数是指观察值的个数。频数分布(frequency distribution)就是观察值在其所取值的范围内,于各组段中分布的情况。所谓频数表(frequency table),是指一种统计表:即同时列出观察值的可能取值及其出现的频数。具体做法是,先根据观察值数量大小进行分组,然后求出每组中观察值出现的次数。由于这种资料的表达方式较完整地体现了观察值的分布规律,所以称为频数分布表。简称频数表。

二、频数表的编制

例 2.1:随机抽查了某山区健康成年男性 100 名,身高的测量结果见表 2-1,试编制频数表。

表 2-1 某山区 100 名健康成年男性身高(cm)测量结果

168.5	170.0	172.5	165.0	175.5	155.5	157.0	180.0	179.5	161.5
175.0	175.0	167.0	171.5	171.0	172.5	168.5	168.0	159.5	164.0
167.0	162.0	176.0	177.5	166.0	167.5	165.0	175.5	169.0	173.5
167.5	172.0	168.5	165.5	161.0	175.5	165.0	169.5	166.5	158.0
160.0	168.5	178.5	172.0	160.0	170.0	167.0	146.5*	155.0	170.5
170.0	175.0	172.0	165.0	156.5	162.0	168.5	165.0	159.0	170.5
162.0	162.5	167.5	166.0	157.5	163.5	169.0	155.5	167.5	169.0
165.0	184.0**	168.0	160.0	179.5	150.0	165.0	173.0	168.0	152.0
170.0	165.0	173.0	156.0	167.5	161.0	176.5	171.5	162.0	158.0
164.0	154.5	168.5	158.0	163.5	170.0	162.5	166.0	152.5	160.0

* 最小值;** 最大值



编制频数表的具体步骤如下:

1. 计算全距:从表 2-1 的 100 个测量值中,找出最大值(maximum,简称 $X_{\max}$)184.0 和最小值(minimum,简称 $X_{\min}$)146.5,二者之差即为全距(range,R)。即:

$$R = X_{\max} - X_{\min} \quad (\text{公式 } 2 \cdot 1)$$

本例为

$$R = 184.0 - 146.5 = 37.5(\text{cm})$$

2. 确定拟分组数(k)和组距(class interval):首先决定组段数(k),一般取 8~15 个组段为宜。观察值个数(n)较多时,可适当增加组段数,但组段数过多,编制过程和计算较繁琐;组段数过少,计算误差较大。

根据全距(R)的大小和组段数(k)计算组距(interval,简称 i)。

$$i = \frac{R}{K} \quad (\text{公式 } 2 \cdot 2)$$

本例全距为 37.5cm,拟分 10 个组段,则组距为:

$$i = R/10 = 37.5/10 = 3.75(\text{cm})$$

为了分组和计算方便,一般取接近的整数作为组距,本例取 $i=4.0\text{cm}$ 。

3. 划分组段:划分组段的基本要求是第一个组段应包括最小值,最后一个组段应包含最大值;各组段只包含下限值但不包含上限值,故在列组段时只列出下限值,不列出上限值,但最后一个组段要依据具体情况进行封口,即要同时列出下限值和上限值。本例,最小值为 146.5cm,组距为 4.0 cm,故第一个组段为 146~,第二个组段为 150~,第三个组段为 154~,……第十个(最后一个)组段为 182~186。如某观察值为 150.0cm,则应归到第二个组段中,而不是归到第一个组段中。

4. 画记并列出频数表:根据已确定的组段,按照表 2-2 形式列出画记表。将原始资料进行归纳计数的方法,可采用画“正”字计数,最后合计出各组段的例数,即为频数表。也可将原始记录输入计算机,由计算机按事先编好的程序进行整理汇总及统计分析。

表 2-2 100 名健康成年男性身高(cm)的频数分布

身高 (cm)	画记	频数	频数%	累计频数%
146.0~	—	1	1.0	1.0
150.0~	下	3	3.0	4.0
154.0~	正下	8	8.0	12.0
158.0~	正正下	12	12.0	24.0
162.0~	正正正正	19	19.0	43.0
166.0~	正正正正正	25	25.0	68.0
170.0~	正正正下	18	18.0	86.0
174.0~	正正	9	9.0	95.0
178.0~	正	4	4.0	99.0
182.0~186.0	—	1	1.0	100.0
合计	—	100	100.0	—

也可以根据各组段的频数绘制直方图,即频数分布图(图 2-1)。从图中可以看出,该地