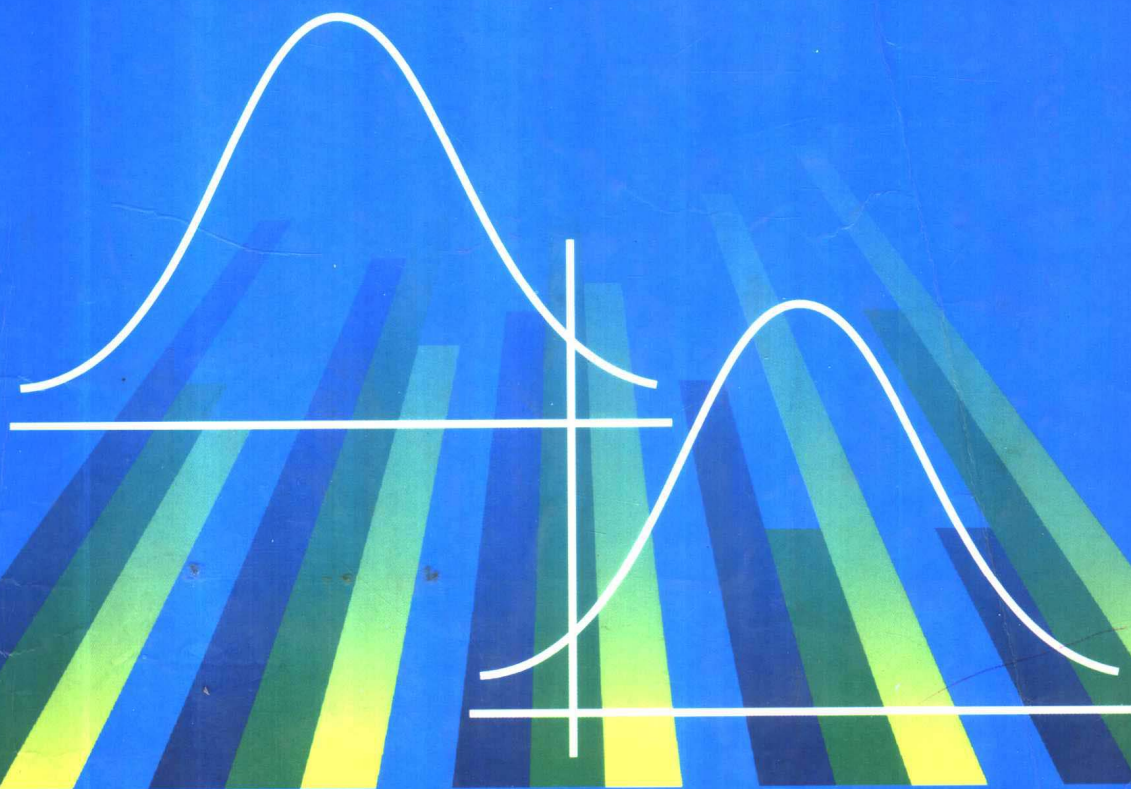


医学统计学

刘筱娴 主编



科学出版社

医学统计学

刘筱嫔 主编

周有尚 主审

科学出版社

2000

内 容 简 介

本书是按照高等医学院校卫生统计专业教学大纲的要求组织编写的,全书共分24章,重点在于阐明各种常用统计方法的意义、原理、应用原则和计算方法,编者结合多年来在教学、科研中所应用的统计方法,尽量多用直观性逻辑说明、实例,使理论与实际结合,内容力图适应广大医务人员目前工作中的需要。

本书既可作为高等医学院校研究生、预防医学专业教材,又可作为医学院校师生、医药卫生科研技术人员及广大医药卫生工作者的参考书。

医 学 统 计 学

刘筱娴 主编

周有尚 主审

责任编辑 王 军

科 学 出 版 社 出 版

北京东黄城根北街16号
邮政编码:100717

武汉大学出版社印刷总厂印刷

新华书店北京发行所发行 各地新华书店经售

2000年10月第 版 开本:787×1092 1/16
2000年10月第 次印刷 印张:26
印数:1~3000 字数:645 000

ISBN 7-03-001014-0/R·38

定价:40.00元

《医学统计学》编委会

主 编 刘筱娴

副主编 董时富 王增珍 尹 平

编 委 (以姓氏笔画为序)

王友洁 王增珍 尹 平 方鹏骞 卢 萍 孙 奕

刘筱娴 余松林 李朝阳 杜树发 陈 辉 张昌敏

赵庆国 梅文华 梅祖国 龚 洁 董时富

前 言

医学统计学是运用统计学的原理和方法,研究和应用于整个医学,包括基础医学、临床医学、预防医学和卫生事业管理的设计、数据的搜集、整理与分析的一门科学.它是认识社会和自然现象数量特征的重要工具,是研究事物运动的客观规律,有效地进行医学科学实验,推动医药卫生工作不断发展的重要科学手段.医务人员肩负着医疗、预防和社区保健服务的重任,需要经常运用统计学方法开展医学科学研究,掌握有关数据,提出各种医疗和预防保健措施,评价其效果。

本书是根据高等医学院校卫生统计专业教学大纲的要求,在修编我校历次编写教材的基础上,结合我们多年来在教学、科研中所应用的统计方法编写的.本书编写时尽量用直观性逻辑说明,多举一些实例,使理论与实际结合,以便于学习和运用。

本书共分 24 章.重点在于阐明各种常用统计方法的意义、原理、应用原则和计算方法.随着医学研究内容的拓宽和深化,要求通过更多、更新的统计方法揭示、分析影响人体的各种复杂因素及其相互关系.为此,本书用较多的章节系统地介绍了多因素分析法的基本概念、类型和计算方法.在计算机技术迅速发展和日益普及的今天,这些多变量分析方法可得到充分的应用.本书编撰内容力图适应广大医务人员目前工作的需要。

本书既可作为高等医学院校研究生、预防医学专业教材,又可作为医学院校师生、医药卫生科研技术人员及广大医药卫生工作者的参考书.鉴于我们编撰教材水平有限,缺点错误之处,恳望同道和读者批评指正。

本书承蒙科学出版社编辑同志悉心审校.龚洁老师完成了汉英、英汉医学统计学词汇的编排工作.贾桂珍同志完成本书全部统计图的绘制.陈若霞老师为编撰本教材做了大量后勤工作.他们为本书的出版作出了贡献,在此一并致以衷心的感谢。

刘筱嫻

目 录

第一章 绪论	(1)
第一节 医学统计学的地位和作用	(1)
第二节 统计工作的步骤	(1)
第三节 常用统计学名词与术语	(2)
第四节 医学统计学的学习方法	(5)
第二章 计数资料的描述性指标	(6)
第一节 相对数的意义	(6)
第二节 率	(6)
第三节 比	(7)
第四节 动态数列	(9)
第五节 应用相对数的注意事项	(10)
第六节 率(比)标准化法	(11)
第三章 计量资料的描述性统计与正态分布	(16)
第一节 频数分布	(16)
第二节 集中趋势指标	(18)
第三节 离散趋势的指标	(23)
第四节 正态分布	(28)
第四章 统计推断	(33)
第一节 抽样误差	(33)
第二节 总体参数估计	(34)
第三节 总体参数比较	(37)
第五章 t 检验	(41)
第一节 t 检验的基本思想与应用条件	(41)
第二节 t 值与 t 分布	(41)
第三节 样本与总体均数比较	(43)
第四节 配对资料比较	(44)
第五节 两样本均数比较	(45)
第六节 两样本几何均数的比较	(46)
第七节 方差齐性检验	(47)
第八节 校正 t' 检验	(48)
第九节 两大样本均数比较的 u 检验	(49)
第六章 方差分析	(51)
第一节 方差分析的基本思想	(51)
第二节 F 值与 F 分布	(52)
第三节 完全随机设计资料的方差分析	(54)

第四节	随机区组设计资料的方差分析	(59)
第五节	均数间的多重比较	(62)
第六节	拉丁方设计资料的方差分析	(67)
第七节	复因设计资料的方差分析	(69)
第八节	缺项估计	(72)
第九节	变量转换	(76)
第七章	χ^2 检验	(81)
第一节	χ^2 统计量及 χ^2 分布	(81)
第二节	四格表资料的 χ^2 检验	(84)
第三节	Fisher's 精确概率检验	(86)
第四节	配对四格表资料的 χ^2 检验	(88)
第五节	$R \times C$ 表资料的 χ^2 检验	(89)
第六节	$R \times C$ 表 χ^2 分割	(91)
第七节	频数分布拟合优度检验	(93)
第八节	列联表交互作用假设检验	(94)
第八章	二项分布与泊松分布	(96)
第一节	二项分布	(96)
第二节	泊松分布	(101)
第九章	非参数检验	(106)
第一节	单一样本和配对样本的分布位置检验	(106)
第二节	两独立样本分布位置相同的假设检验	(108)
第三节	多个独立样本的秩检验	(111)
第四节	中位数检验	(116)
第十章	Ridit 分析	(119)
第一节	Ridit 分析的基本原理	(119)
第二节	Ridit 值总体均数置信区间的估计	(123)
第三节	假设检验	(124)
第十一章	线性相关与回归分析	(127)
第一节	线性相关关系	(127)
第二节	线性回归	(133)
第三节	线性相关与回归的区别与联系	(142)
第十二章	协方差分析	(144)
第一节	协方差分析的意义及其作用	(144)
第二节	完全随机设计资料的协方差分析	(144)
第三节	调整均值的假设检验	(150)
第四节	随机区组设计资料的协方差分析	(151)
第十三章	半数效量	(156)
第一节	概率单位法	(157)
第二节	寇氏法	(160)
第三节	两个半数致死量的差异显著性检验	(163)

第四节	半数致死量的实验设计	(164)
第十四章	正常值范围的估计	(166)
第一节	正常值范围估计的步骤	(166)
第二节	正态性检验	(170)
第三节	正态分布法	(174)
第四节	对数正态分布法	(175)
第五节	百分位数法	(176)
第十五章	统计表和统计图	(178)
第一节	统计表	(178)
第二节	统计图	(182)
第十六章	医学研究设计	(190)
第一节	医学研究设计概述	(190)
第二节	实验设计	(206)
第三节	临床试验设计	(214)
第四节	专题调查设计	(217)
第十七章	多元线性回归分析	(224)
第一节	多元线性回归的意义	(224)
第二节	多元回归模型参数估计	(224)
第三节	多元回归方程的计算举例	(226)
第四节	多元线性相关分析	(233)
第十八章	逐步回归分析	(236)
第一节	“最优”回归方程的选择	(236)
第二节	指标的数量化	(237)
第三节	逐步回归的计算步骤	(238)
第四节	计算举例	(241)
第五节	预报区间估计	(246)
第六节	自变量的选择准则	(247)
第十九章	曲线回归	(250)
第一节	曲线回归的意义与一般步骤	(250)
第二节	常见曲线类型与图像	(250)
第三节	指数曲线拟合	(252)
第四节	对数曲线拟合	(255)
第五节	幂曲线拟合	(256)
第六节	Logistic 曲线拟合	(260)
第七节	曲线拟合优度的比较	(263)
第二十章	判别分析	(266)
第一节	判别分析概念及其医学应用	(266)
第二节	计量资料的二类判别	(266)
第三节	二值回归判别	(274)
第四节	多类线性判别	(276)

第五节	计数判别·····	(283)
第二十一章	聚类分析·····	(288)
第一节	概 述·····	(288)
第二节	聚类统计量·····	(288)
第三节	数据变换·····	(291)
第四节	系统聚类法·····	(292)
第五节	Q 型(样品)聚类实例·····	(300)
第六节	R 型(指标)聚类实例·····	(303)
第二十二章	主成分分析与因子分析·····	(305)
第一节	主成分分析·····	(305)
第二节	因子分析·····	(312)
第三节	主成分分析与因子分析的关系·····	(318)
第二十三章	列联表资料的对数线性模型拟合·····	(320)
第一节	对数线性模型的基本概念·····	(320)
第二节	对数线性模型的参数估计·····	(325)
第三节	对数线性模型的配合适度检验·····	(327)
第四节	对数线性模型的应用·····	(329)
第二十四章	综合评价法·····	(334)
第一节	TOPSIS 法·····	(334)
第二节	主成分分析法·····	(337)
第三节	量表结构效度的因子评价法·····	(339)
附录 1	统计用表·····	(343)
附录 2	汉英医学统计学词汇·····	(372)
附录 3	英汉医学统计学词汇·····	(387)
主要参考文献	·····	(402)

第一章 绪 论

第一节 医学统计学的地位和作用

统计学(statistics)是研究数据的搜集、整理、分析与推断的科学.它是认识社会和自然现象的数量特征,研究事物运动的客观规律的重要手段.我们无论做什么工作,都应当掌握基本情况,胸中有数,才能有的放矢地做好工作.在现代信息社会中,各行各业、各种部门都必须广泛地搜集有关数据,进行准确分析和可靠推断,以便于作出科学决策.因而,统计学原理和方法被科技、生产、生活及国民经济各个领域所广泛应用,产生了一系列应用性分支,如工业统计、农业统计、医学统计等等.

医学统计学(medical statistics)是把概率论与数理统计学的基本原理和方法应用于医学科学领域,涉及到医学科学研究的设计、资料搜集、归纳、分析和解释的一门应用性基础学科.

医学的研究对象主要是人体以及与人体健康有关的因素.由于生物现象变异较大,各种影响因素错综复杂,使得在一个环境里同一类个体对同一种刺激得出不同的数据和反应.如用药物治疗某病,尽管年龄、性别、病程、病情均相同,其治疗效果也不一.又如,同年龄、同性别的人群,其身高、体重和各种生化指标的数值均有所不同.这就是个体存在变异的结果.为了达到医学研究目的,须通过样本推断总体,了解某事物客观存在的规律.实践中还强调研究对象必须具有一定的观察数量.综上所述,医学研究必须运用统计方法透过偶然现象探测其客观规律性,使之得出科学推断.

电子计算机技术的发展和普及应用,给医学统计学的应用和发展提供了广阔的前景.电子计算机为大量数据资料的搜集、信息贮存、传输与检索、复杂的数据处理,特别是对多因素分析、抽样模拟等提供了极好的条件,大大促进了医学统计学的发展.

第二节 统计工作的步骤

统计工作一般分为四个步骤,即先有一个严密的设计,然后根据设计的要求搜集资料、整理资料和分析资料.这四个步骤是相互联系、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.在实际工作中,无论是进行调查研究、实验研究或临床试验研究,首先应根据研究的目的和要求以及内容的可行性作出周密、完善的设计,据此进行统计工作.

一、设计

统计工作和科研工作类似,设计(design)是最关键的一步.根据研究的目的和要求,通过参考有关专著、文献和杂志,了解这方面研究的动向、国内外水平或通过预备试验和试查的方法,获取有关数据和信息.譬如,采取什么方式、方法来搜集资料,如何保证资料的完整、准确,怎样对获取的原始资料整理、汇总,采取哪些统计指标,如何进行统计处理,预期会得到什么结

果?凡此种种,都要经过周密的考虑,结合实际情况,作出科学细致的安排.尤其应充分利用电子计算机来整理、分析资料.这个全过程的设计,是统计工作最重要的开端,以利于用较少的人力、财力取得最好的效果.

二、搜集资料

搜集资料(collection of data)是统计工作的第二步骤,亦是统计工作的基础.只有保证资料的完整性(原始资料无遗漏或重复)、正确性(原始资料准确反映实际情况,项目之间合乎逻辑、无矛盾),才能整理分析出可靠的指标,反映事物的特征和实际发展水平.统计资料来源于三个主要方面:①统计报表:它是由国家卫生主管部门制订、指定各级卫生机构填报、自下而上逐级上报的有关卫生方面的统计资料,为制订医疗卫生工作计划,总结工作提供了科学依据.②医疗卫生工作记录和报告卡:例如病历,健康检查记录,门诊和病房工作日志,传染病报告卡,出生、死亡、职业病报告等,以上提供的资料是否准确、完整,都会影响到整个统计工作质量.因此,必须加强检查、核对及全面管理,避免发生漏填、重复和项目填写不清等.③专题调查或实验:除上述两种来源外,根据工作需要,还可以进行专题统计调查研究和实验研究,例如孕产妇死亡流行病学统计调查,儿童意外伤害监测控制效果调查评价等.

三、整理资料

整理资料(sorting data)即主要把大量的原始资料进行科学的加工与综合,使之系统化.具体步骤是,通过分组,拟订整理表归纳汇总,编制统计表.例如研究医院工作质量指标之一——治愈率时,不能只看个别病例的疗效,而必须将所有患者的疗效进行分组整理,归纳汇总,才能得到总体疗效、分科的疗效、各种疾病的疗效等必要的资料.因此,任何统计资料不经过科学的整理,是不能进行科学分析、得出正确结论的.

四、分析资料

资料经过整理,得出的一系列不同类型的数据,此时根据资料的性质计算出有关统计指标(如相对数、平均数),并采用单因素或多因素分析方法进行统计处理.为使指标形象表达,可绘制统计图表,最后结合专业理论,对资料内容进行综合或专题分析,总结出符合客观事物规律、具有科学依据的结论.

第三节 常用统计学名词与术语

一、资 料

医学统计资料(data)一般分为三类,即计量资料、计数资料以及介于两者之间的等级资料.不同类型的资料应当采用不同的分析方法.

1. 计量资料(measurement data) 对每个观察单位用计量方法测定某项指标数值大小的资料,称为计量资料.这类资料一般是有度量单位的,且随计量工具精度的提高可测量出一

定位数的数据.例如调查某幼儿园儿童的生长发育状况,测量每个儿童的身高(cm)、体重(kg)、血压(kPa)等即是计量资料.

2. 计数资料(enumeration data) 先将观察单位按不同属性或类别进行分组,所得各组的观察单位数,称为计数资料.例如对某小学校的学生进行蛔虫卵粪检,每个学生是一个观察单位,将每个学生按粪检结果阳性与阴性进行分组,得出每组的人数,就是计数资料.计数资料的观察值是整数,没有小数位.

3. 等级资料(ranked data) 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.例如观察某种新药治疗小儿肺炎的疗效状况,其中治愈、好转、无效的例数,即为等级资料.这类资料因其属性的分组存在程度的差别,各组按数量大小顺序排列,故不同于计数资料;而且每个观察单位未确切定量,所以亦不同于计量资料,故称为半计量资料.

二、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.例如研究儿童生长发育状况,选取同性别、同年龄儿童的身高进行观察,在统计学上称“同质”观察单位.由于宇宙中的事物千差万别,各不相同,即使是性质相同的事物,也是相对的,每个观察单位(亦称个体)之间,各有差异,称为变异(variation).如在同性别同年龄的儿童中,身高有的高,有的低,各不相同,称为身高的变异.在同性别中,亦有低年龄儿童的身高高于高年龄儿童的情况.但总的趋势,儿童身高是随年龄的增加而增高,这是客观规律.

同质观察单位之间的个体差异,是生物的重要特征,是偶然性的表现,是由于生物体内外环境中,多种因素的综合影响造成的,其中有的因素亦是难于控制的.统计研究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与规律.

三、总体与样本

总体(population)是指被研究事物的全体,又称为母体.更确切地说,总体是指性质相同的所有观察单位(observed unit)某种变量值的集合.样本(sample)是从总体中抽取供研究的部分个体,又称子样.总体所包含的个体数,可以无穷大,而且往往只是设想的或是抽象的.例如对婴儿腹泻的防治研究,所有患腹泻的婴儿,都具有患腹泻这个同质的特征,是一个总体,每一个患腹泻的婴儿叫做个体.我们研究婴儿腹泻的规律,不可能把每个患儿全部观察到,只能对一部分个体进行观察.这种从总体中抽取部分个体观察的过程,称为抽样(sampling),所抽得的部分个体,称为样本.该样本所包含的个体数目,称为样本含量(sample size).

在医学研究中,很多是无限总体,要直接研究总体的情况是不可能的,即使是有限总体,如果包含的观察单位太多,也需花费很多的人力、物力和财力,有时亦是不必要和不可能的.因此,研究工作所掌握的资料经常只是总体中的一小部分,仅是一个样本.科学研究工作的目的,不能只满足于依据一个样本所得到的结果,而应期望明确总体的情况.所以,如何正确地从样本特征来推测总体水平,是科学研究中统计所要解决的问题.

四、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用统计学方法来估计,必须遵循随机化抽样的原则,使总体中的每个个体都有同等的机会被抽取.在动物实验中,对动物的分组亦要随机化,使各个动物都有同等机会被分配到各组中去.只有这样,才能避免人为的因素所造成的偏差,以便于运用统计方法处理各种数据.因此,统计分析和推断方法是根据随机化抽样而建立的.随机化的具体方法有抽签法和随机数字表法等,亦可利用计算机产生的随机数进行随机化.

五、抽 样 误 差

由于总体中各观察单位之间存在个体变异,随机抽取的样本,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)**.一般来说,样本愈大,抽样误差愈小,用样本推断总体的精确度愈高,愈能阐明事物的客观规律性.反之亦然.但在实际研究工作中,观察对象往往不可能太多,只能从一个较小的样本资料来进行分析研究,计算出一系列的统计指标.由于个体变异是客观存在的,因而抽样误差亦是不可避免的.样本愈小,抽样误差相应地愈大,若仅凭数字表面值进行判断,不考虑偶然性问题,将会导致错误的结论.抽样误差有其一定的规律性,应用概率理论所导出的一些统计处理方法,就能够辨别在实验研究中哪些有统计意义,哪些可能是由于偶然性所造成的,从而有助于得出正确的结论.这一类统计处理(或检验)方法,本书将分别详细介绍.

六、概 率

概率(probability),又称几率或或然率,是反映某一事件发生的可能性大小的量度.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,在大量临床研究中发现,某种新药治疗某病的治愈率为 75%(0.75),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 0.75.统计学上常用符号“ P ”来表示概率.

在一定的条件下,肯定发生的事件称为必然事件,其概率等于 1;肯定不发生的事件称为不可能事件,其概率等于 0;可能发生亦可能不发生的事件,称为随机事件或偶然事件,其概率介于 0 与 1 之间.概率越接近 1,表明发生的可能性越大;概率越接近 0,发生的可能性越小.统计上的很多结论都是带有概率性的.在医学科研论著中常常见到 $P \leq 0.05$ 或 $P \leq 0.01$,前者表示某事件发生的可能性等于或小于 0.05,后者表示某事件发生的可能性等于或小于 0.01,习惯上称为小概率事件,以此作为事物差别有统计意义或有高度统计意义的界限.

七、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)。反映样本特性的有关指标,称为统计量(statistic)(如样本均数、样本标准差和样本率等),亦即样本观测值的函数。例如,为了解7岁健康男童的身高而进行测量,将所测得的一系列数值算出一个平均数(即样本平均数),就是一个统计量。通过随机化抽样,研究样本特性的有关指标,以估计总体参数,是统计推断的基本任务。统计学上用希腊字母代表总体参数,用拉丁字母代表样本统计量。

第四节 医学统计学的学习方法

学习医学统计学必须结合学科特点,应用科学的方法,才能收到良好的效果。

(1) 要正确运用统计的思维方法,以唯物辩证思想为指导,正确认识客观事物,阐明事物固有的规律,把感性认识提高到理性认识,用于指导卫生工作实践。例如,由于我们研究的对象是有生命的个体,而生物个体变异是客观存在的,抽样过程产生的误差是不可避免的。但是,这种误差是有规律性的。因而,在进行样本指标的比较时,不能仅从数字表面大小看问题,而必须用辩证的观点去认识事物的偶然性和必然性、随机性和确定性对立统一的关系,通过随机性去认识确定性,通过偶然性去认识必然性,正确地搜集和分析取得的资料,才能获得科学正确的结论。

(2) 要重点理解医学统计学的基本原理与基本概念,掌握搜集、整理与分析资料的基本知识与技能,正确应用常用的统计指标与基本统计方法。要重视原始资料的完整性与可靠性,对数据处理必须持严肃、认真、实事求是的科学态度,反对伪造和篡改统计数据。对数理统计公式,只要求了解其意义、用途、应用条件及运用方法,一般不需深究其数学推导。

(3) 掌握群体健康的评价方法,学习运用医学统计指标,综合评价人群健康状况,为卫生工作的决策提供有关统计信息。

(4) 必须联系实际进行学习,要结合专业,对卫生工作实践、医学科学文献以及医学科研中的有关统计问题进行分析评价,以巩固学习理论,提高实际工作能力。

(刘筱娴)

第二章 计数资料的描述性指标

第一节 相对数的意义

描述计数资料的主要统计指标是相对数。相对数(relative number)是两个有关联的数据之比,用以说明事物的相对关系,便于对比分析。常用相对数指标甚多,按其作用不同可分为率、构成比和相对比三类。

研究计数资料首先获得的是某种性质类别的计数,我们把这些原始数据叫做绝对数(absolute number)。绝对数反映事物在某时某地发生的实际水平,是制订计划、总结工作不可缺少的基本数据,也是进一步统计处理的基础依据。但在进行深入统计分析时,仅有绝对数往往是不够的,因为绝对数不便于进一步的相互比较。

例如甲乙两地农村蛲虫感染情况调查,甲地蛲虫感染 156 人,乙地蛲虫感染 101 人,我们只能说甲地蛲虫感染人数比乙地多 55 人,却不能肯定甲地比乙地人群蛲虫感染程度更为严重,因为感染程度与该地人口数的大小有关,而两地调查的总人数不一定相等。已知甲地调查 244 人,乙地调查 158 人,则

$$\text{甲地蛲虫感染率} = 156/244 \times 100\% = 63.9\%$$

$$\text{乙地蛲虫感染率} = 101/158 \times 100\% = 63.9\%$$

这里甲乙两地都用 100 做基数,便于进一步分析比较。从计算结果来看,甲乙两地蛲虫感染的严重程度是一样的。因此,如果要深入进行统计分析,就需要计算相对数指标,消除分母不同的影响,产生一个无量纲的新指标。

相对数便于进一步分析比较,但却掩盖了绝对数所说明的情况。因此,分析相对数时,不能忽略绝对数,两者应互相补充。

第二节 率

率(rate)又称频率指标,表示在特定条件下,某现象实际发生例数与可能发生该现象的总例数之比,用来说明某现象发生的强度或频率。常用百分、千分、万分或十万分作为比例基数,称为百分率(%),千分率(‰),万分率(1/万)或十万分率(1/10 万)等。其计算公式如下:

$$\text{率} = \frac{\text{某现象实际发生的例数}}{\text{可能发生该现象的总例数}} \times K \quad (2.1)$$

式中, K 为比例基数,其选择主要依据习惯和使算得的率呈现一定的有效位数、方便阅读为宜,如常见病的发生率习惯用百分率(%)或千分率(‰),罕见病的发生率多用万分率(1/万)或十万分率(1/10 万),出生率、粗死亡率习惯用千分率(‰),肿瘤死亡率习惯用十万分率(1/10 万)等。

如上例中甲地蛲虫感染 156 人,调查人数为 244 人,计算蛲虫的感染率为:

$$\text{甲地蛲虫感染率} = 156/244 \times 100\% = 63.9\%$$

即该地每 100 人中有 63.9 人感染蛲虫。同理计算出乙地的蛲虫感染率也是 63.9%,可见两地

人群蛲虫感染的严重程度是一样的。

医学科学研究中常用的率有：

$$\text{某病发病率} = \frac{\text{某地某时期某病新发病例总数}}{\text{该地同期平均人口数}} \times K \quad (2.2)$$

$$\text{某病患病率} = \frac{\text{某地某时点某病现患病例总数}}{\text{该地该时点受检人数}} \times K \quad (2.3)$$

$$\text{死亡率} = \frac{\text{某地某时期死亡总数}}{\text{该地同期平均人口数}} \times K \quad (2.4)$$

$$\text{某病治愈率} = \frac{\text{某病治愈人数}}{\text{该病受治病人数}} \times K \quad (2.5)$$

$$\text{某病有效率} = \frac{\text{某病治疗有效人数}}{\text{该病受治病人数}} \times K \quad (2.6)$$

$$\text{某病病死率} = \frac{\text{某时期某病死亡人数}}{\text{同期该病病人数}} \times K \quad (2.7)$$

$$\text{年龄别死亡率} = \frac{\text{某年龄组死亡人数}}{\text{该年龄组人口数}} \times K \quad (2.8)$$

例 2-1 某县某年各年龄组恶性肿瘤死亡情况统计资料见表 2-1 的前 3 栏。各年龄组死亡率为：

$$0 \sim 14 \text{ 岁死亡率} = \frac{11}{356980} \times 100000/10 \text{ 万} = 3.08/10 \text{ 万}$$

$$15 \sim 29 \text{ 岁死亡率} = \frac{22}{232505} \times 100000/10 \text{ 万} = 9.46/10 \text{ 万}$$

余类推，结果见表 2-1 的末栏。

表 2-1 某县某年恶性肿瘤死亡统计

年 龄 组	人口数	死亡数	死亡率/1/10 万
0~	356 980	11	3.08
15~	232 505	22	9.46
30~	205 032	142	69.26
50~	121 882	443	363.47
≥85	20 047	107	533.75
合 计	936 446	725	77.42

从计算结果可以看出，随着年龄的增大，恶性肿瘤死亡率逐渐升高，即年龄愈大，恶性肿瘤死亡率愈高。年龄别死亡率反映各年龄组的死亡水平。某年龄组死亡率，由该年龄组死亡的例数和人口数所决定，它不受其他年龄组死亡率的影响。

合计全人口的死亡率为死亡数的合计除以人口数的合计，即

$$\frac{725}{936446} \times 100000/10 \text{ 万} = 77.42/10 \text{ 万}$$

第三节 比

比(ratio)是两事物数量的比值，以反映相对比例、比重或倍数。按照比的构成及意义可分为相对比指标与构成比指标。

一、构成比

构成比(constituent ratio)表示事物内部某一构成成分在全部构成中所占的比例或比重。常以 100 为比例基数,又称为百分比,计算公式为:

$$\text{构成比} = \frac{\text{事物内部某一构成成分的数值}}{\text{事物各构成成分数值的总和}} \times 100\% \quad (2.9)$$

例 2-2 某地区 1985 年和 1990 年 5 种急性传染病死亡人数见表 2-2,1985 年急性传染病死亡人数 301 人,其中死于伤寒、副伤寒者 33 人,则伤寒、副伤寒死亡人数占全部急性传染病死亡人数的百分比为:

$$\frac{33}{301} \times 100\% = 10.96\%$$

同理可分别计算出 1985 年和 1990 年流脑、痢疾、白喉……等死亡占全部急性传染病死亡人数的百分比,见表 2-2.

表 2-2 某地区 5 种急性传染病死亡情况

病 名	1985 年		1990 年	
	死亡人数	构成/%	死亡人数	构成/%
伤寒、副伤寒	33	10.96	14	10.61
流 脑	27	8.97	23	17.42
痢 疾	210	69.77	71	53.79
白 喉	26	8.64	13	9.85
百日咳	5	1.66	11	8.33
合 计	301	100.00	132	100.00

从表 2-2 中可以看出,1985 年及 1990 年因痢疾而死亡占的比重最大,占传染病全部死亡人数的一半以上;百日咳占的比重最小,从中看出我们防治传染病的重点是痢疾.

在计算构成比时,各个分子的总和等于分母,即构成比总和为 100%. 由于构成比的这种分子分母关系,所以某部分比重的变化受着两方面因素的影响,其一是这个部分自身数值的变化,其二是受其他部分数值变化的影响. 构成比只是说明某一组成部分在一定条件下(即一定分母条件下)所占的比重. 表 2-2 中,1985 年和 1990 年的 5 种传染病的死亡总数不相同,两年的伤寒、副伤寒虽例数相差很大,但构成比却很接近. 再看两年的流行性脑脊髓膜炎的死亡人数相差不很大,但构成比却相差近 1 倍,这不能说明 1990 年流行性脑脊髓膜炎的病死严重程度较 1985 年高. 由于各个构成比的大小受到其他组成部分数值的影响,在 1990 年其他 4 种传染病死亡人数有明显减少,而流行性脑脊髓膜炎死亡人数减少不多,致使其在总死亡数中所占的比重相应地增高. 因此,死因构成比只能说明某病死亡人数在总死亡人数中所占比重,如需要比较其病死的严重程度,则需要计算其病死率.

二、相对比

相对比(relative ratio)是两个有关指标之比. 两个指标可以性质相同,也可以性质不同,反映两者之间的差别变化情况. 通常以倍数或百分数(%)表示,计算公式为: