



复旦卓越·医学职业教育教材

YIXUE
ZHIYIE JIAOYU
JIAOCAI

卫生统计与流行病学 实践教程



主编 施 榕



复旦大学出版社

www.fudanpress.com.cn

责任编辑 肖 英
封面设计 马晓霞

内 容 提 要

《卫生统计与流行病学实践教程》是一本应用范围广泛的教材和参考书，几乎涉及医学众多学科各个领域，而且随着新技术的出现，已经发展成为一门重要的主干课程，并成为从事数理统计特别是涉及群体研究的学生们的必修课。

该书强调了两方面的特点：一是突出医学高等职业教与学的实际要求，在内容上有针对性地进行了梳理与调整；二是强调“务实”，关注理论的实际应用与操作，为将来学生的社会实践打下良好基础。

序 Preface

随着医药卫生事业研究的不断深入,对运用流行病学方法进行实证研究的要求不断提高,同时,越来越多的卫生统计学方法被积极运用于数据分析之中,流行病学和统计学结合的重要性日益凸显。与此相对应的“卫生统计和流行病学”也成为医学科学中一门重要的基础学科和方法学科。

《卫生统计和流行病学实践教程》应用范围广泛,几乎涉及医学众多学科各个领域,而且随着新技术的出现,已经发展成为一门重要的主干课程。《卫生统计和流行病学实践教程》已成为从事临床医学、预防医学各专业特别是涉及群体健康研究的学生们的必修课。

上海震旦职业学院隶属震旦教育,其秉承著名爱国教育家马相伯的教育理念和教学优势,坚持“培养学生综合素养、创造能力、高尚品格”的培养目标。2001年,学院开设营养与食品卫生专业;2007年该专业成为上海市民办高校教学高地的建设项目,得到上海市教育委员会的支持与肯定。

为培养拥有扎实理论知识与较强实践能力的大学生,在编写本教材的过程中,我们强调了两方面的特点:一是突出高职高专教与学的实际要求,在内容上有针对性地进行了梳理与调整;二是强调“务实”,关注卫生统计和流行病学方法在实际工作中的应用与操作,为学生将来的工作实践打下良好基础。

希望学生们通过学习,学会透过众多的偶然因素阐明事物客观存在的规律性,从而得出正确的结论并应用于工作实际。同时,希望通过本课程学习提高发现问题、解决问题的能力,并善于把卫生统计和流行病学的理论与方法应用于广泛的实践中,不断提高工作效率和工作水平。

张惠莉

2008年6月于上海

目 录 Contents

第一章 卫生统计学方法的基本概念和基本步骤..... 1

第一节 卫生统计学方法的基本概念 / 2

- 一、总体与样本 / 2
- 二、抽样误差 / 3
- 三、变量及其分类 / 3
- 四、概率 / 3
- 五、参数与统计量 / 4

第二节 卫生统计工作的基本步骤 / 4

- 一、设计 / 4
- 二、收集资料 / 4
- 三、整理资料 / 4
- 四、分析资料 / 5

第二章 数值变量的统计描述..... 6

第一节 数值变量资料的频数分布表与频数分布图 / 7

- 一、频数分布表的编制 / 7
- 二、频数分布图 / 8
- 三、频数分布的特征 / 9
- 四、频数分布的类型 / 9

第二节 集中趋势的描述 / 9

- 一、算术均数 / 9
- 二、几何均数 / 11
- 三、中位数 / 11

第三节 离散趋势的描述 / 13

- 一、标准差 / 13
- 二、变异系数 / 14



三、四分位数间距 / 15	
第四节 正态分布和医学参考值范围的估计 / 16	
一、正态分布 / 16	
二、医学参考值范围的估计 / 18	

第三章 数值变量的统计推断..... 21

第一节 均数的抽样误差与标准误 / 22	
一、均数的抽样误差 / 22	
二、标准误 / 22	
第二节 t 分布 / 23	
第三节 总体均数的估计 / 24	
第四节 假设检验 / 25	
一、假设检验的基本原理 / 25	
二、假设检验的一般步骤 / 26	
第五节 t 检验 / 27	
一、样本均数与总体均数比较 / 27	
二、配对设计的差值均数与总体均数的比较 / 28	
三、完全随机设计的两个样本均数的比较 / 29	
第六节 方差分析 / 30	
一、方差分析的基本思想 / 31	
二、完全随机设计的多个样本均数比较 / 31	
三、随机区组设计的多个样本均数比较 / 33	
四、多个样本均数间两两比较的 q 检验 / 34	
第七节 假设检验中的两类错误及注意事项 / 35	
一、I 类错误和 II 类错误 / 35	
二、假设检验的注意事项 / 36	

第四章 分类变量的统计描述..... 38

第一节 常用相对数 / 39	
一、率 / 39	
二、构成比 / 39	
三、相对比 / 40	
第二节 应用相对数的注意事项 / 40	
一、计算率和构成比时分母不宜过小 / 40	
二、分析时正确区分构成比和率 / 41	
三、注意相对数的可比性 / 41	
四、注意使用率的标准化 / 41	



五、两样本率比较时应进行假设检验 / 41	
第三节 率的标准化 / 42	
一、率标准化的概念与基本思想 / 42	
二、标准的选择 / 42	
三、标准化率的计算 / 43	
第五章 分类资料的统计推断	45
第一节 率的抽样误差和总体率的估计 / 46	
一、率的抽样误差和标准误 / 46	
二、总体率的可信区间估计 / 46	
第二节 χ^2 检验 / 47	
一、四格表的 χ^2 检验 / 47	
二、配对四格表 χ^2 检验 / 50	
三、行 $\times$ 列表的 χ^2 检验 / 51	
第六章 秩和检验	53
第一节 配对设计的符号秩和检验 / 54	
一、基本原理与方法 / 54	
二、检验步骤 / 55	
第二节 两独立样本的秩和检验 / 56	
一、两样本比较法 / 56	
二、正态近似法 / 57	
第三节 成组设计多个样本比较的秩和检验 / 57	
第四节 等级资料的秩和检验 / 58	
第七章 直线相关与回归分析	61
第一节 直线相关 / 62	
一、直线相关的概念 / 62	
二、直线相关系数 / 64	
三、相关系数的假设检验 / 65	
第二节 直线回归 / 65	
一、直线回归的概念 / 65	
二、直线回归方程的求法 / 66	
三、回归系数的假设检验 / 67	
四、直线相关与回归分析的区别和联系 / 68	
五、直线相关与回归分析的注意事项 / 68	



第八章	统计表和统计图	70
	第一节 统计表 / 71	
	一、统计表的结构与列表原则 / 71	
	二、常用统计表的种类 / 72	
	第二节 统计图 / 72	
	一、统计图的结构与制图原则 / 72	
	二、统计图种类的选择及绘制 / 73	
第九章	SPSS 统计分析软件简介	77
	第一节 SPSS 软件简介 / 78	
	一、概述 / 78	
	二、SPSS 的操作简介(以 SPSS 11.0 为例) / 78	
	第二节 SPSS 软件应用实例 / 79	
	一、某临床实验课题 / 79	
	二、建立数据库 / 80	
	三、数据处理分析 / 82	
第十章	流行病学概述	87
	第一节 流行病学的发展 / 88	
	第二节 流行病学的定义和研究方法 / 90	
	一、流行病学的定义 / 90	
	二、流行病学研究方法 / 91	
	第三节 流行病学的研究范围与用途 / 92	
	一、描述疾病或健康状况的分布 / 92	
	二、探索病因 / 92	
	三、用于疾病的诊断、疗效评价和预后分析 / 92	
	四、用于疾病防制和卫生决策的评价 / 93	
第十一章	疾病的分布	94
	第一节 描述疾病分布的常用指标 / 95	
	一、描述疾病发生频率的指标 / 95	
	二、描述疾病存在频率的指标 / 96	
	三、描述疾病死亡频率的指标 / 97	
	四、描述疾病流行水平的指标 / 98	
	第二节 疾病分布的形式 / 98	
	一、人群分布 / 98	



- 二、时间分布 / 101
- 三、地区分布 / 103
- 四、疾病三间分布的综合描述 / 104

第十二章 描述性研究 106

第一节 描述性研究概述 / 107

- 一、概念 / 107
- 二、描述性研究的种类 / 107
- 三、描述性研究的主要用途 / 107

第二节 现况研究概述 / 108

- 一、研究实例 / 108
- 二、概念 / 108
- 三、现况研究的目的 / 108
- 四、现况研究的特点与种类 / 109

第三节 现况研究的设计与实施 / 110

- 一、明确调查目的和类型 / 110
- 二、确定研究对象 / 110
- 三、样本含量和抽样方法 / 110
- 四、调查表的编制 / 113
- 五、现况研究资料的整理、分析和结果解释 / 115

第四节 生态学研究 / 117

- 一、概念 / 117
- 二、种类 / 117
- 三、优缺点 / 118

第五节 暴发调查 / 118

- 一、调查实例 / 118
- 二、概念 / 120
- 三、暴发调查的方法与步骤 / 120
- 四、资料分析与结果报告 / 122

第十三章 病例对照研究 124

第一节 病例对照研究概述 / 125

- 一、研究实例 / 125
- 二、概念 / 126
- 三、研究特点 / 126
- 四、种类 / 127

第二节 病例对照研究的设计与实施 / 128



一、明确研究目的 / 128

二、选择研究对象 / 129

三、估计样本含量 / 130

四、研究因素的确定 / 131

五、资料的收集 / 132

第三节 病例对照研究资料的分析 / 132

一、资料的整理 / 132

二、资料分析 / 132

第四节 病例对照研究的优缺点 / 136

一、病例对照研究的优点 / 136

二、病例对照研究的缺点 / 136

第十四章 实验流行病学 137

第一节 概述 / 138

一、实验流行病学简介 / 138

二、实验流行病学研究的主要类型 / 139

三、实验流行病学研究的基本原则 / 141

四、实验流行病学研究的三个基本要素 / 141

五、实验流行病学研究的优缺点 / 141

六、实验流行病学的设计和实施 / 142

七、社区试验案例 / 144

第二节 临床试验 / 144

一、临床试验的主要类型 / 145

二、临床试验设计的基本原则 / 145

三、设计和实施时应注意的问题 / 147

四、涉及的医学伦理问题 / 149

五、临床试验案例 / 149

第十五章 病因推断与研究中偏倚的控制 152

第一节 概述 / 153

一、病因的概念 / 153

二、病因模型 / 154

第二节 病因推断 / 155

一、病因推断的过程 / 155

二、病因推断的逻辑方法 / 157

三、从统计学关联到因果关联 / 158

四、病因推断的标准 / 159



五、病因推断案例 / 161

第三节 流行病学研究中的常见偏倚及其控制 / 162

一、误差的概念和分类 / 162

二、偏倚的分类 / 163

第十六章 疾病筛检 169

第一节 筛检概述 / 170

一、筛检的定义 / 170

二、筛检的类型 / 171

三、实施筛检的原则 / 171

四、疾病筛检的评价 / 171

第二节 筛检的评价方法 / 172

一、筛检的评价程序 / 172

二、筛检试验的评价 / 173

第十七章 营养流行病学 179

第一节 概述 / 180

一、营养流行病学的定义及研究目的 / 180

二、营养流行病学的应用 / 180

第二节 膳食测量 / 181

一、膳食测量的范围 / 182

二、膳食测量的时间 / 183

三、膳食测量的方法 / 183

第三节 营养流行病学的研究方法 / 187

一、生态学研究 / 187

二、现况研究 / 188

三、病例对照研究 / 190

四、干预试验 / 191

[附录] 195

附表1 t 界值表 / 195

附表2 F 界值表 / 195

附表3 Student - Newman - Keuls 检验用 q 界值表 / 197

附表4 χ^2 界值表 / 198

第一章

DYZ

卫生统计学方法的 基本概念和基本步骤

第一节 卫生统计学方法的基本概念

- 一、总体与样本
- 二、抽样误差
- 三、变量及其分类
- 四、概率
- 五、参数与统计量

第二节 卫生统计学的基本步骤

- 一、设计
- 二、收集资料
- 三、整理资料
- 四、分析资料



卫生统计学方法是应用概率论和数理统计的基本原理和方法,研究医学领域中数据的收集、整理和分析的一门应用性学科。医疗卫生领域中存在着大量的随机现象,即在一定的条件下结果不能确切预测的现象,如同样疾病的患者,即使用同样的药物,疗效也可能不同。例如,某医生比较两种疗法对活动期十二指肠球部溃疡的疗效,一组 100 例患者口服呋喃硝胺,另一组 100 例患者口服西咪替丁(甲氰咪胍)。结果是呋喃硝胺治疗的有效率为 90%,西咪替丁治疗的有效率为 85%。据此能否立即下结论,呋喃硝胺的疗效高于西咪替丁? 其实,由于治疗效果的不确定性,用呋喃硝胺再治疗 100 例活动期十二指肠球部溃疡患者,有效率有可能低于 85%。这时我们需要采用卫生统计学中的假设检验方法,才能对上述两组治疗效果作出合理的评价。由此可见,卫生统计学方法对医疗卫生科学研究和疾病防治工作有着重要意义,是医学相关专业学生必须掌握的基本知识。

第一节

卫生统计学方法的基本概念

一、总体与样本

(一) 总体

在医学科学研究中需要根据研究目的确定观察单位,观察单位可以是一个人、一个家庭、一个地区、一个样品等。总体就是指根据研究目的确定的性质相同观察单位的全体。例如,研究某地正常成人红细胞数的情况,观察对象是该地 2002 年全部正常成人,观察单位是每个人,观察值是每人测得的红细胞数,而该地该年全部正常成人的红细胞数就构成了研究总体,它的同质基础是同一地区、同一年份、同为正常成人。该总体只包括有限个观察单位,称为有限总体。有时总体是设想的或是抽象的,例如研究用某药治疗胃溃疡患者的疗效,总体为设想用该药治疗所有胃溃疡患者的治疗结果,这里没有确定的时间和空间范围的限制,称为无限总体。

(二) 样本

从总体中随机抽取部分观察单位的某项指标实测值组成的样本。从上述的某地某年正常成人中随机抽取 200 人,这 200 位正常成人的红细胞数就是样本。医学研究的总体许多是无限总体,要直接研究总体是不可能的。即使是有限总体,若包含观察单位太多,也要花费大量人力、财力,通常是不必要的和不可能,因此只能从总体中用随机的方法抽取部分样本进行研究。抽样研究的目的是用样本的信息来推论总体特征。从总体中抽取部分个体的过程称为抽样。抽样必须遵循随机化原则,即总体中每一个体有同等的机会被抽取,这样



的样本对总体有较好的代表性。

二、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,因而样本指标不一定恰好等于相应的总体指标。例如从某市某年3岁男童的总体中随机抽取150名儿童,调查得此样本的身高均数为103.8 cm,这个数值不一定恰好等于该市7岁男童的总体均数。又如从某地随机抽取1500人,经检测乙型肝炎病毒携带率为9.5%,但这个数字不一定恰好等于该地人群乙型肝炎病毒携带率。样本指标与总体指标的差异称为抽样误差。显然,抽样误差越小,用样本推断总体的准确性就越高,反之亦然。由于生物的个体变异是客观存在的,因而在抽样过程中抽样误差是不可避免的,但抽样误差是可以控制和进行估计的。

三、变量及其分类

我们研究某种医学现象时,一般需从掌握其具体的特征入手。这些特征称为观察指标,如人的年龄、体重、身高等。这些观察指标在统计学上统称为变量,对变量的测定值称为变量值。如测量同为5岁男孩的身高,身高值有高有矮。这种个体间的差异,称为变异。按变量值是定量还是定性,可将它们分成以下类型。

(一) 数值变量

又称计量资料,其变量值是用定量方法测量的,表现为数值的大小,一般有计量单位。如调查5岁男童生长发育状况时,每个人的身高(cm)、坐高(cm)、体重(kg)等都是数值变量。

(二) 分类变量

其变量值是用定性方法得到的,表现为互不相容的类别或属性。根据类别是否有程度上的差别,又可分为以下两种类型。

1. 无序分类 无序分类的各类别间无程度上的差别,又称计数资料,包括:①二项分类,如某药治疗菌痢,治疗结果按治愈和未愈分为两类,两类间相互对立。②多项分类,如测量某人群的血型,结果分为A型、B型、AB型、O型,表现为多个互不相容的类别。

2. 有序分类 有序分类的各类别之间有程度上的差别,又称等级资料,或半定量资料。例如,临床疗效按治愈、显效、好转和无效分为四级,每级治疗效果有程度上差别。

不同类型的变量其统计处理方法不同,在统计分析中分清变量类型是很重要的基础工作。在实际应用中,根据分析的需要,各类型变量间可以互相转化。如观察某人群成年女子的血红蛋白量(g/L),属数值变量;若按血红蛋白正常与异常分为两类,属二项分类变量;若按血红蛋白量的多少分为5个等级,即重度贫血、中度贫血、轻度贫血、正常、血红蛋白增高,又属有序分类变量。有时也可将分类变量数量化,如治疗效果中治愈和未愈可以分别用1、0表示;有序分类资料中的临床疗效(无效、好转、显效、治愈)可用0、1、2、3表示。

四、概率

在一定条件下某一现象可能发生也可能不发生,这种事件叫做随机事件。医学研究的现象绝大多数是随机现象,例如用相同方法治疗某病患者,我们只知道治疗转归可能是治



愈、好转、无效、死亡四种结果;但对一个正准备接受治疗的该病患者,治疗后究竟将会发生哪一种结果是不知道的。这里每一种可能的结果都是一个随机事件。概率(probability)是反映某一随机事件发生的可能性大小的量,用符号 P 表示,概率 P 的取值范围在 0 与 1 之间。概率越接近于 1,表明事件发生的可能性越大;概率越接近 0,表明事件发生的可能性越小。统计学上一般把 $P \leq 0.05$ 或 $P \leq 0.01$ 的事件称为小概率事件,表示某事件发生的可能性很小,在实际的一次抽样中可认为不会发生。

五、参数与统计量

根据总体个体值计算出来的描述总体特征的指标称参数(parameter)。参数一般用希腊字母表示,如总体均数 μ 、总体 π 等。根据样本个体值计算出来的描述样本特征的指标称统计量(statistic)。统计量用拉丁字母表示,如样本均数 $\bar{X}$ 、样本率 P 。总体参数一般是未知的或假设的,而样本统计量是研究者从样本中计算得到的。总体参数是固定不变的,样本统计量随样本不同而不同。

第二节

卫生统计工作的基本步骤

统计工作一般分为设计、收集资料、整理资料和分析资料 4 个步骤。这 4 个步骤是相互联系、不可分割的。任何一个步骤的缺陷都会影响后续步骤,使统计分析的结果不可靠。

一、设计

在从事医学科学研究工作之前,要事先作好研究计划,即研究设计。一个完整的医学科研设计应包括研究目的、意义、研究对象、研究方法 with 内容,以及研究进度和预期结果等基本内容。因此设计就是对整个研究工作进行全盘规划。所以设计是后三个步骤的依据,是统计工作最重要的一个环节。根据设计中是否对观察单位施加处理因素,医学科研设计分为调查设计和实验设计。这两种设计的具体内容将在有关章节介绍。

二、收集资料

根据研究目的收集准确、完整的原始资料,这是统计分析的基础。卫生统计资料主要来自于以下 4 个方面:①统计报表,如疫情报表、医院工作报表等,这些都是根据国家规定的报告制度,由医疗卫生机构定期逐级上报的。②登记和报告卡(单),如出生报告单和出生登记、死亡报告单和死亡登记、传染病和职业病报告卡、肿瘤发病和肿瘤死亡报告卡等。③日常医疗卫生工作记录,如门诊病历、住院病历、健康检查记录等。④专题调查或实验,这是开展医学科研的主要资料来源。上述前三种卫生统计资料的内容都有局限性,如需进行深入分析,常采用专题调查或实验。

三、整理资料

资料整理的目的是把原始资料系统化和条理化,便于下一步计算统计指标和统计分析。



资料整理的过程包括:①首先对原始数据进行核对和检查。②设计分组。分组有两种,一是质量分组:将观察单位按其属性或类别(如性别、职业、疾病分类等)归类分组;二是数量分组:将观察单位按数值大小(如年龄大小、血压高低等)分组。两种分组常结合应用,一般是在质量分组基础上进行数量分组,如先按性别分组、再按体重的数值大小分组。③按分组要求设计整理表,汇总资料(手工记分卡或计算机汇总)。

四、分析资料

统计分析目的是计算有关指标,反映资料的综合特征,揭示资料的内在联系和规律,是统计工作中最主要的一个环节。但是,这个环节的质量不仅取决于统计分析方法是否正确,也取决于前面环节的工作质量。资料分析包括统计描述和统计推断。统计描述是指用统计指标和适宜的统计表、图描述资料的分布规律及其数量特征;统计推断包含总体参数估计和假设检验两个方面内容。

[小结]

卫生统计学是基于概率论和数理统计学关于资料收集、分析的学科,在医学研究与实践和疾病预防中发挥着重要作用。本章介绍了卫生统计学中的基本概念,包括统计资料的类型、总体与样本、抽样误差、概率和小概率事件等。本章还介绍了统计工作的基本步骤:设计,资料收集、整理和分析。

(施 榕)

第二章

DEZ

数值变量的统计描述

第一节 数值变量资料的频数分布表与频数分布图

- 一、频数分布表的编制
- 二、频数分布图
- 三、频数分布的特征
- 四、频数分布的类型

第二节 集中趋势的描述

- 一、算术均数
- 二、几何均数
- 三、中位数

第三节 离散趋势的描述

- 一、标准差
- 二、变异系数
- 三、四分位数间距

第四节 正态分布和医学参考值范围的估计

- 一、正态分布
- 二、医学参考值范围的估计