

高职高专工学结合教改规划教材系列

社区卫生 诊断技术

Diagnostic Techniques
in Community Health

主 编 范春红 曲彤薇
副主编 高越明 赵海军



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

高职高专工学结合教改规划教材系列

社区卫生诊断技术

主 编 范春红 曲彤薇
副主编 高越明 赵海军



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

图书在版编目(CIP)数据

社区卫生诊断技术 / 范春红, 曲彤薇主编. —杭州:
浙江大学出版社, 2013.3

ISBN 978-7-308-11058-7

I. ①社… II. ①范…②曲… III. ①社区医学—诊断学 IV. ①R44

中国版本图书馆 CIP 数据核字 (2013) 第 014446 号

社区卫生诊断技术

范春红 曲彤薇 主编

丛书策划 阮海潮(ruanhc@zju.edu.cn)

责任编辑 阮海潮

封面设计 俞亚彤

出版发行 浙江大学出版社

(杭州市天目山路 148 号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排版 杭州中大图文设计有限公司

印刷 杭州日报报业集团盛元印务有限公司

开本 787mm×1092mm 1/16

印张 11.5

字数 294 千

版印次 2013 年 3 月第 1 版 2013 年 3 月第 1 次印刷

书号 ISBN 978-7-308-11058-7

定价 29.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部邮购电话 (0571)88925591

前 言

2009 年中共中央、国务院颁布《关于深化医药卫生体制改革的意见》,随后国务院出台了《医药卫生体制改革近期重点实施方案(2009—2011 年)》,明确将基层卫生服务定位于公共卫生服务和基本医疗服务。2010 年国家发展改革委等部门印发了《关于开展农村订单定向医学生免费培养工作的实施意见》,要求各地实施定向培养农村社区医生工作,加强基层卫生人才队伍建设,缓解农村卫生人才紧缺状况。为适应基层人才培养的需求,我们在教学实践和基层调研的基础上,编写了此教材——《社区卫生诊断技术》。

本教材根据我国医疗体制改革精神和《国家基本公共卫生服务规范》,结合临床执业医师助理医师资格考试《预防医学》的考试大纲,在参考了传统的《统计学基础》和《流行病学》课程教学内容和知识结构的基础上,建立高职高专《社区卫生诊断技术》的教学课程体系。本教材包括下面两个主要特点:

1. 强调基本理论和基本技能的实际应用。全书共 16 章,其中第 1~7 章介绍了卫生统计学的基本理论、基本知识和基本技能,第 8 章介绍了常用的卫生统计学指标,第 9~15 章介绍了流行病学常用的调查方法和研究指标,最后一章则介绍了 SPSS 软件中基本的分析模块。教材的内容选取着重于基本理论和方法的介绍,并在理论和方法的基础上结合医学研究和实践的案例进行分析和调查研究,以增强学生的分析问题和解决问题的能力。

2. 统计的实验教学融入 SPSS 软件的应用。编者认为,高职高专的教学目的重在技能的培养,在统计教学中遵循“弱化公式、任务为导向、结合软件、强调应用”的理念。因此,第 16 章重点介绍了统计理论部分的实践应用,并用具体案例来介绍 SPSS 的分析过程及结果的阅读和理解,使学生更容易掌握统计学的基本技能。该章最后一节是专题介绍统计分析基本技能的综合应用,是基于前面理论和技能的掌握基础上的进一步深化与扩充。

由于编者的经验和水平有限,在教材内容的选取和组织上难免有不足之处,恳请各位专家、读者批评指正,以便本书的不断完善。

编 者

2013 年 1 月

目 录



CONTENTS

第一章 社区卫生诊断概述	1
第二章 卫生统计学概述	4
第一节 卫生统计工作的步骤 /4	
第二节 统计学的基本概念 /5	
第三章 数值变量的统计描述	9
第一节 数值变量资料的频数分布 /9	
第二节 集中趋势的统计描述指标 /11	
第三节 离散程度的统计描述指标 /14	
第四节 正态分布及其应用 /17	
第四章 分类资料的统计描述	21
第一节 相对数 /21	
第二节 应用相对数的注意事项 /23	
第三节 标准化法及其应用 /24	
第五章 统计表和统计图	27
第一节 统计表 /27	
第二节 统计图 /29	
第六章 数值变量的统计推断	33
第一节 均数抽样误差与标准误 /33	
第二节 t 分布 /34	
第三节 总体均数的估计 /35	
第四节 假设检验的基本步骤 /37	



第五节	t 检验和 Z 检验 /38	
第六节	秩和检验 /42	
第七节	应用假设检验需要注意的问题 /45	
第七章	分类资料的统计推断	47
第一节	率的抽样误差和标准误 /47	
第二节	总体率的估计 /48	
第三节	χ^2 检验 /48	
第八章	常用的统计指标	54
第一节	医学人口统计常用指标 /54	
第二节	疾病统计常用指标 /60	
第三节	其他常用指标 /61	
第九章	流行病学概述	62
第一节	流行病学的概念 /62	
第二节	流行病学的研究方法 /63	
第三节	流行病学与其他学科的关系 /64	
第四节	流行病学的用途 /65	
第五节	流行病学研究的重要观点 /66	
第十章	疾病的分布	68
第一节	流行病学常用的测量指标 /68	
第二节	疾病流行的强度 /71	
第三节	描述疾病的分布形式 /72	
第十一章	描述性研究	82
第一节	现况调查 /82	
第二节	筛 检 /88	
第三节	个例调查 /91	
第四节	暴发调查 /91	
第十二章	病例对照研究	94
第一节	概 述 /94	
第二节	研究设计 /95	
第三节	资料整理与分析 /98	

第四节	偏倚及其控制 /100	
第五节	应用范围及其优缺点 /102	
第十三章	队列研究	104
第一节	概 述 /104	
第二节	研究步骤 /105	
第三节	队列研究的优缺点 /111	
第十四章	社区干预试验	112
第一节	实验研究 /112	
第二节	社区干预试验 /113	
第十五章	疾病监测	119
第一节	疾病监测的概念 /119	
第二节	疾病监测系统 /119	
第三节	疾病监测的方法与内容 /120	
第四节	我国疾病监测工作开展概况 /123	
第十六章	SPSS for Windows 实习	124
第一节	数据文件的建立 /124	
第二节	数值变量的统计描述 /128	
第三节	t 检验 /140	
第四节	秩和检验 /148	
第五节	χ^2 检验 /153	
第六节	实例综合分析 /160	
附录	统计用表	169
附表 1	标准正态分布曲线下的面积 /169	
附表 2	t 界值表 /170	
附表 3	F 界值表(方差齐性检验用) /171	
附表 4	F 界值表(配对比较的符号秩和检验用) /173	
附表 5	T 界值表(两样本比较的秩和检验用) /174	
附表 6	χ^2 界值表 /175	
参考文献		176

第一章 社区卫生诊断概述

一、社区卫生诊断概念与意义

社区卫生诊断是运用社会学、人类学和流行病学的研究方法对一定时期内,社区的主要健康问题及其影响因素、社区卫生服务的供给与利用以及社区综合资源环境进行客观、科学的确定和评价,发现和分析问题,提出优先干预项目,并针对性地制定社区卫生服务工作计划,从而充分利用现有卫生资源,提高社区卫生服务质量和效率,满足社区居民基本卫生服务需求,动员社区参与,实施社区干预,逐步解决社区主要卫生问题,不断提高居民健康水平 and 生活质量。

社区卫生服务是卫生工作的重要组成部分,是实现人人享有初级卫生保健目标的基础环节。发展社区卫生服务作为政府履行社会管理和公共服务职能的重要内容,是促进社会公平、维护居民健康、构建和谐社会的重大举措。

要提供优质高效的社区卫生服务,首位的基础工作就是需要有一个全面、正确的社区卫生诊断。开展社区卫生诊断对于政府及有关社会部门编制社区卫生规划、合理配置卫生资源以及发挥社区各类相关资源的综合利用效益,对于提高社区卫生服务质量与效率、切实落实社区卫生服务机构的公共卫生和基本医疗双重网底功能、满足社区居民基本卫生服务需求,从而保证和促进社区卫生服务健康、可持续发展,构建社会主义和谐社会,达到提高社区居民整体健康水平和生活质量的最终目的具有极其重要的意义。

二、社区卫生诊断的目的

社区卫生诊断的目的主要包括:

- (1)发现并确定社区主要健康问题及其危险因素;
- (2)总结并评价社区卫生资源特别是社区卫生服务中心的供给与利用效率;
- (3)了解并分析社区环境及其相关资源现状;
- (4)调查并分析居民卫生服务需求与利用、满意度及其卫生知识水平;
- (5)分析并提出本社区优先干预的卫生问题;
- (6)制定本社区卫生服务工作规划,并为评估社区卫生服务的综合效果提供基线数据。

三、社区卫生诊断流程

社区卫生诊断流程分为四步,即设计准备、资料收集、资料统计和分析报告。

1.设计准备 社区卫生诊断工作需要周密的计划,制订实施方案,确定资料的收集、整理、分析的方法以及时间进度,并进行必要的组织准备和物资准备。

2.资料收集 资料收集是社区卫生诊断的重要内容,是做好社区卫生诊断的关键环节。

开展社区卫生诊断要尽量收集可能收集到的资料,力求资料翔实可靠,为社区卫生诊断提供较高利用价值的客观数据。资料收集方法包括收集现有资料和进行社区卫生专项调查。

社区卫生诊断中的专项调查主要包括居民卫生调查、服务对象满意度调查和社区卫生服务机构调查。关于社区访谈调查视社区具体情况决定是否进行。

3.资料统计 通过对资料审核、计算机录入后,利用统计软件进行统计,客观描述社区环境特征、社区人群特征以及社区卫生服务资源特征。

4.分析报告 通过对收集到的资料的统计,全面总结、分析本社区人群的主要健康问题及其危险因素,评价卫生资源的供给与利用效率以及社区环境的支持保障能力,从而确定本社区优先干预项目,最后要撰写社区卫生诊断报告以及制定今后五年的社区卫生服务工作计划。

社区卫生诊断流程中涉及的方法和内容等详见图 1-1 所示。

四、社区卫生诊断所涉及的方法

从社区卫生诊断的流程可以看出,社区卫生诊断所涉及的方法很广泛,其综合采用社会学、市场学、人类学、测量评估学、营养学、临床医学、流行病学、卫生统计学、卫生服务管理、卫生经济学等手段针对社区的基本现况以及存在的公共卫生问题进行综合性的调查和评估。本书从社区卫生诊断的要求出发,主要介绍社区卫生诊断方法中涉及的流行病学调查和卫生统计学分析的方法,主要包括:一是卫生统计的基本理论和技能;二是常用的健康统计指标;三是流行病学的常用调查方法和研究指标;四是介绍 SPSS 软件中基本的分析模块,以适应社区卫生诊断的需求。

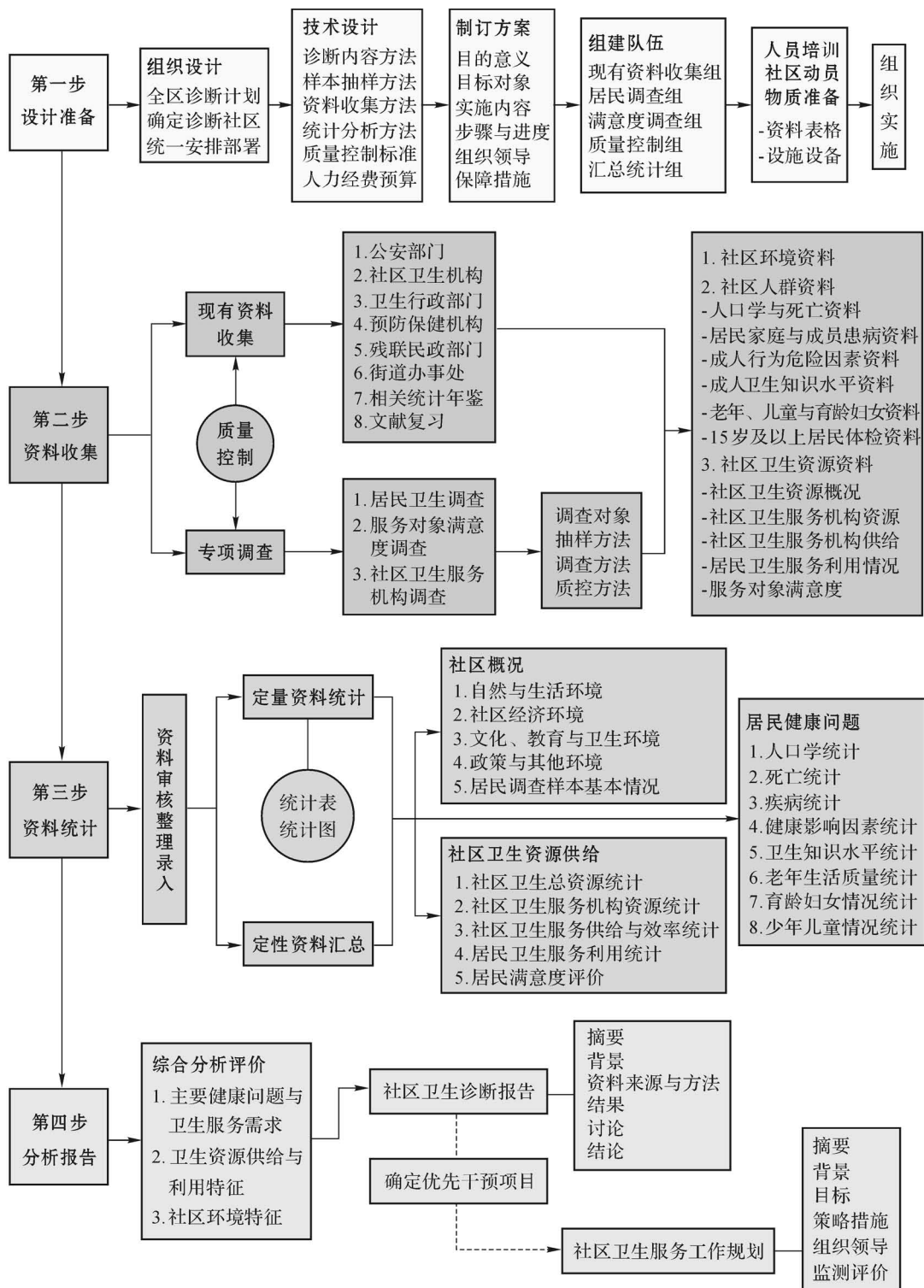


图 1-1 社区卫生诊断流程

(资料来源:www.jlws.gov.cn/JlsWst/eWebEditor/Up)

第二章 卫生统计学概述

统计学(statistics)是研究数据的收集、整理、分析和解释,以显示其总体特征和内在规律性的科学。统计学的研究主要有两个领域:数理统计学和应用统计学。数理统计学侧重于建立统计方法和阐述统计方法的原理;而应用统计学则是把数理统计学的原理和方法与特定的领域结合。卫生统计学(health statistics)属于应用统计学的范畴,是应用数理统计学的原理和方法研究居民健康状况以及卫生服务领域中数据的收集、整理和分析。

第一节 卫生统计工作的步骤

统计工作包括统计设计、收集资料、整理资料和分析资料四个基本步骤,这四个步骤相互联系、不可分割,任何步骤的缺陷都会影响统计分析的结果。但在实际工作中,许多人往往是忽略了设计、收集和归类(整理),到了分析数据时才想到统计学,使实验的结果往往偏离预期的设想。

一、统计设计

统计设计(design)是统计工作中最关键的一步,是在广泛查阅文献或通过预备试验,在全面了解现状、充分征询意见的基础上,对将要进行的研究工作所做的全面设想。其内容包括:明确研究目的和研究假说,确定观察对象、观察单位、样本含量和抽样方法,拟定研究方案、预期分析指标、误差控制措施、进度与费用等。一个好的统计设计就是在运用较少的人力、物力和时间的基础上,获得更可靠、更科学的结果资料。

二、收集资料

收集资料(collection of data)的任务是按研究设计的要求,及时取得准确、完整的原始数据。只有原始数据可靠,才能得出可靠的结论,因此,这一步骤具有极重要的基础意义。卫生统计资料主要来源于以下两大类:

- 1.经常性资料 一般指医疗工作中的记录,包括:①统计报表,根据国家规定的报告制度,由医疗卫生机构定期逐级上报,如医院工作年报表、月报表、卫生基本情况年报表、疫情报表、居民的病伤死因报表等。②工作记录与报告卡,工作记录包括卫生监测记录、健康检查记录、门诊记录、住院记录等,报告卡包括传染病、职业病报告卡、出生、死亡报告卡等。

- 2.专题调查或实验资料

三、整理资料

整理资料(sorting data)的任务是按研究设计的要求,使原始数据系统化、条理化,便于进一步的计算指标和分析。其过程是:首先对原始资料进行准确性审查(逻辑审查与技术审查)和完整性审查;再拟定整理表,按照“同质者合并,非同质者分开”的原则对资料进行分组,并在同质基础上根据数值大小进行数量分组;最后汇总归纳。

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,阐明事物的内在联系和规律。统计分析包括统计描述(descriptive statistics)和统计推断(inferential statistics)。前者是用统计指标与统计图(表)等方法对样本资料的数量特征及其分布规律进行描述;后者是指如何抽样,以及如何用样本信息推断总体特征。进行资料分析时,需根据研究目的、设计类型和资料类型选择恰当的描述性指标和统计推断方法。

第二节 统计学的基本概念

一、变量的类型

观察单位(observed unit),亦称个体(individual),是统计研究中的最基本单位。观察单位可以是一个人,一只动物,一份样品或一个地区等。研究者对每个观察单位的某项特征进行测量或观察,观察单位的这项特征就称为变量,测量或观察得到值称为变量值。

统计资料按变量的类型,一般分为数值变量资料与分类变量资料两大类(图 2-1)。不同类型的变量应采用不同的分析方法。

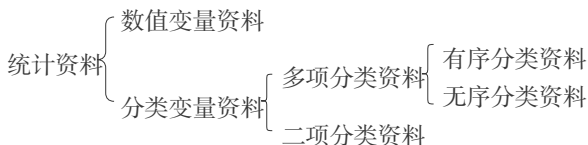


图 2-1 统计资料的分类

(一)数值变量

数值变量(measurement data)也称为计量资料、定量资料,是对每个观察单位的各项指标用定量的方法,通过测量得到的数值,一般有度量衡等单位。如每个人的身高(cm)、体重(kg)和血压(mmHg)等。大多数的数值变量为连续型变量。描述数值变量常用的统计指标有平均数、标准差等,假设检验的方法有 t 检验、 Z 检验等。

(二)分类变量

分类变量(enumeration data)也称为计数资料,先将观察单位按某种属性或类别分组,然后清点所得各组的观察单位数。取值表现为互不相容的类别,一般用符号表示,无度量衡单位。分类变量资料用相对数进行统计描述,用卡方检验进行统计推断。这类资料可分为以下两类:①只有两个类别,如性别分为男和女,药物疗效分为有效和无效等。②可分为两

个以上类别,各观察单位间或者相同,或者存在质的差别,有质的差别者之间没有连续性。多项分类资料可分为两大类:

1.无序分类资料 指各类别间无程度上的差别和等级顺序关系,如血型可分为 A 型、B 型、AB 型和 O 型,职业包括工人、农民、学生、军人等。

2.有序分类资料 也称等级资料,指各类别之间存在程度上的差别和等级顺序关系,如疾病预后分为无效、好转、显效、痊愈,尿蛋白检测结果分为一、+、++、+++、++++等。各等级之间只有程度上的区别,而无数值大小,不可度量。

根据研究目的和记录方法不同,各类资料之间有时可以相互转换。如年龄是数值变量资料,按照年龄大小分为未成年人和成年人,转化为分类变量资料中的二项分类变量资料;若再分为婴儿、幼儿、学龄前儿童、学龄儿童、青少年、壮年、老年人,则转换为多项分类资料中的等级资料。研究中为方便起见,也可以对分类变量资料赋值,将其转化为数值变量资料。如在多变量研究中,将性别为男时取 1、为女时取 2,此时 1 和 2 没有数值大小的区别,只起到标签的作用;又如将上述 7 个年龄段的人分别取值 1、2、3、4、5、6、7,将等级变量资料转化为数值变量资料,此时数值的大小只表示程度上的差别,不能进行各类运算。

二、同质与变异

严格意义上的同质(homogeneity),是指被研究指标的影响因素完全相同。但在医学领域,某些影响因素常难以控制(如遗传)甚至是未知的,因此在医学统计中,常把同质理解为对研究指标影响较大、可以控制的主要因素尽可能相同。如研究青少年身高时,要求性别、年龄、民族、地区等对身高影响较大而又容易控制的要素要相同。

在同质条件下,就同一观察指标来说,各观察单位表现出来的数量间存在着差异,这种客观存在的差异性称为变异(variation)。例如,同样是研究青少年的身高,同性别、同年龄、同民族、同地区青少年(即“同质”观察单位)的身高,有高有低,各不相同,称为身高的变异。同质观察单位之间的个体变异,是生物的重要特征。同质是相对的,没有同质性就不能构成一个总体;而变异是绝对的,没有变异就无需统计学。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。

三、总体与样本

总体(population)是根据研究目的确定的同质的研究对象的某次变量值的集合。例如,研究某地 2012 年正常成人的血压值,则研究对象是该地 2012 年的正常成人,观察单位是每个人,变量是血压,变量值是测得的血压值,该地 2012 年全部正常成人的血压值就构成一个总体。它的同质基础是同一地区,同一年份,同为正常成人。这里的总体只包括有限个观察单位,称为有限总体(finite population)。

有时总体是设想的,如研究高血压患者用某药治疗后的血压,这里的总体同质基础应是同为高血压的所有患者,其观察单位数显然是不确定的,无限的,称为无限总体。

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。即使对有限总体来说,若包括的观察单位数过多,直接研究总体也是耗费人力、财力很大的,有时也是不可能和不必要的,因此我们通常都用样本。所谓样本(sample),是指从总体中随机抽取的有代表性的一部分研究对象。对抽样所得到的样本进行深入观察与测量,获取数据;再通过统计学知

识,用样本数据推断总体特征。

四、参数和统计量

统计学中把总体的特征指标统称为参数(parameter)。而由样本算得的相应的特征指标称为统计量(statistics)。如研究某地成年男子的平均脉搏数(次/分),并从该地抽取 1000 名成年男子进行测量,所得的样本平均数即称为统计量。习惯上用希腊字母表示总体参数,例如 μ 表示总体均数, π 表示总体率, σ 表示总体标准差等。以拉丁字母表示统计量, $\bar{X}$ 表示样本均数, P 表示样本率, S 表示样本标准差等。

同质与变异、总体与样本、参数与统计量之间的关系见图 2-2 所示。

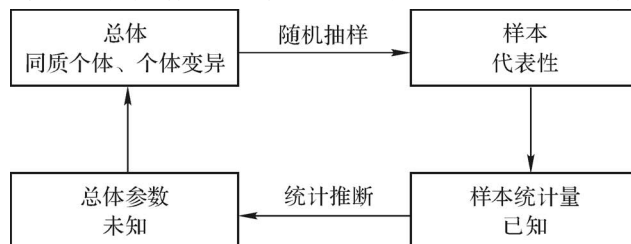


图 2-2 各概念之间的关系

五、常用的实验设计方法

1. 完全随机设计(completely random design) 不考虑个体差异的影响,仅涉及一个处理因素,但可以有多个或两个水平,所以亦称单因素实验设计。该设计常用于将受试对象按随机化原则分配到处理组和对照组中,各组样本例数可以相等,也可以不等,但相等时效率高。完全随机设计的优点是设计和统计分析方法简单易行;缺点是只分析一个因素,没有考虑个体间的差异,因而要求各观察单位要有较好的同质性,否则,需扩大样本含量。

2. 配对设计(paired design) 配对设计是将受试对象按配对条件配成对子,每对中的个体接受不同的处理。配对设计一般以主要的非实验因素作为配比条件,而不以实验因素作为配比条件。动物实验中,常将同性别、同窝别、体重相近的两个动物配成一对;人群试验中,常将性别和年龄、生活条件、工作条件相同或相近的两个人配成对子,再按随机化原则把每对中的受试对象分别分配到实验组和对照组,或不同处理组。此外,某些医学实验研究中的自身对照也可看作配对设计,如某指标治疗前后的比较,同一受试对象不同部位、不同器官的比较,同一标本不同检测方法的比较。

六、统计描述与统计推断

统计描述(statistical description),主要描述样本特征,不考虑抽样误差问题。描述的形式有:①列表描述,如一览表、频数表;②图示描述,如直条图、直方图、构成图;③指标描述,如平均数、标准差。统计描述是对原始资料的一种概括,即把分散而不好理解的原始数据,通过计算有关指标,阐明事物的内在联系和规律。

统计推断(statistical inference)是通过样本所提供的信息来推断总体特征,考虑了抽样误差问题,推断的内容有:①总体参数的估计;②假设检验。



七、概率

概率(probability)是描述某事件发生的可能性大小的一个度量。如事件 A 发生的可能性大小,用实数来表示,就称为事件 A 的概率,常记为 $P(A)$,大量观察某病病例的病亡情况时,就可以得出病亡例数 f 占总病例数 n 的比值,称为频率 f/n 。当 n 充分大时,就以频率作为概率,概率常用 P 来表示。随机事件概率的大小在 0 与 1 之间,即 $0 < P < 1$,常用小数或百分数表示。 P 越接近 1,表示某事件发生的可能性越大; P 越接近 0,表示某事件发生的可能性越小。 $P=1$ 表示事件必然发生, $P=0$ 表示事件不可能发生,它们是确定性的,不是随机事件,但可以把它看成随机事件的特例。若随机事件 A 的概率 $P(A) \leq \alpha$,习惯上,当 $\alpha=0.05$ 时,就称 A 为小概率事件,其统计学意义是小概率事件在一次随机试验中不可能发生。此为小概率原理,是统计推断的理论根据之一。

第三章 数值变量的统计描述

统计描述是用统计指标、统计图或统计表描述资料的分布规律及其数量特征。通过实验或调查收集到的数值变量资料,例数较多时,可以先将数据编制成频数表,了解变量值的分布情况,然后计算平均数描述其集中位置,计算变异程度指标描述其离散程度;例数较少时,亦可直接计算平均数和变异指标。

第一节 数值变量资料的频数分布

一、频数表的编制

频数就是观察值的个数,将各个类别及其相应的频数(或频率、百分比)用表格的形式全部列出来,就是频数分布表,简称频数表。

以某地某年 100 名成年男子血清总胆固醇测定结果为例,说明频数分布表的编制步骤。

【例 3.1】 某地某年 100 名健康成年男子血清总胆固醇浓度见表 3-1,试编制频数分布表。

表 3-1 某地某年 100 名成年男子血清总胆固醇浓度(mol/L)

3.37	4.79	5.10	4.77	5.32	4.50	5.10	4.70	4.44	5.16
4.37	6.25	5.55	4.56	3.35	4.08	4.63	3.61	4.97	4.17
5.77	5.09	4.38	5.18	4.79	5.15	4.79	5.30	4.77	4.40
4.89	5.86	3.40	3.38	4.55	5.15	4.24	4.32	5.85	3.24
5.85	3.04	3.89	6.16	4.58	5.72	4.87	5.17	4.61	4.12
4.43	4.31	6.14	4.88	2.70	4.60	6.55	4.76	4.48	6.51
5.18	3.91	5.39	4.52	4.47	3.64	4.09	5.96	6.14	4.69
6.36	4.60	5.09	4.47	3.56	4.23	4.34	5.18	5.69	4.25
6.30	3.95	4.03	5.38	5.21	7.22	4.31	4.71	5.21	3.97
5.12	4.55	4.90	3.05	5.20	4.74	5.54	3.93	3.50	6.38

1. 计算全距(range) 全距也称极差,用 R 表示。

$$R = X_{\max} - X_{\min} \tag{3-1}$$

式中: R 是全距, $X_{\max}$ 为所有数据中最大值, $X_{\min}$ 为所有数据中最小值。

本例中, $X_{\max} = 7.22(\text{mol/L})$, $X_{\min} = 2.70(\text{mol/L})$, $R = 7.22 - 2.70 = 4.52(\text{mol/L})$ 。

2. 确定组数(k) 组数的确定应以能够显示数据的分布特征和规律为目的,根据样本含量的大小确定组数,一般设 8~15 个组段。观察单位较少时组数可相对少些,观察单位较多

时组数可相对多些。

本例中, $R=4.52(\text{mol/L})$, $N=100$,可分 10 组,即 $k=10$ 。

3.确定组距(i) 根据全距和组数确定组距(i),组距可用公式(3-2)进行估算。

$$i=R/k \tag{3-2}$$

本例中, $i=4.52/10=0.452(\text{mol/L})$,为方便计算,取 0.5mol/L 作为组距。

4.划分组段 各组段的界限应清晰分明,第一组段应包括最小值,最后一组段应包括最大值。每一组段的起始值称下限,终止值称上限。为避免交叉,各组段从下限开始(包括下限),到本组段上限为止(不包括上限),即遵循“不重不漏,含组下限,不含组上限(最后一组除外)”原则。各组用下限+“~”表示,最后一组同时写出上下限。

5.计算频数,形成频数分布表 按照确定的组段设计计划记表(表 3-2),将原始数据进行归纳计数,可用划“正”字的方式(见表 3-2 第 2 列),并给出各组段的频数 f (见表 3-2 第 3 列),完成频数表的编制。

表 3-2 某地某年 100 名健康成年男子血清总胆固醇浓度(mol/L)频数表

组段(1)	划记(2)	频数(3)	组中值(4)	fX (5)	fX^2 (6)
2.50~	—	1	2.75	2.75	7.56
3.00~	正丁	7	3.25	22.75	73.94
3.50~	正正	9	3.75	33.75	126.56
4.00~	正正正正	20	4.25	85.00	361.25
4.50~	正正正正正	25	4.75	118.75	564.06
5.00~	正正正正	19	5.25	99.75	523.69
5.50~	正正	9	5.75	51.75	297.56
6.00~	正丁	7	6.25	43.75	273.44
6.50~	丁	2	6.75	13.50	91.13
7.00~7.50	—	1	7.25	7.25	52.56
合计		100		479.00	2371.75

二、频数分布图的绘制

根据数值变量资料的频数分布表,以观察值为横轴,以各组频数为纵轴,每一组画一等宽矩形,矩形面积与该组频数成正比,如图 3-1 所示,称为频数分布图,简称频数图。

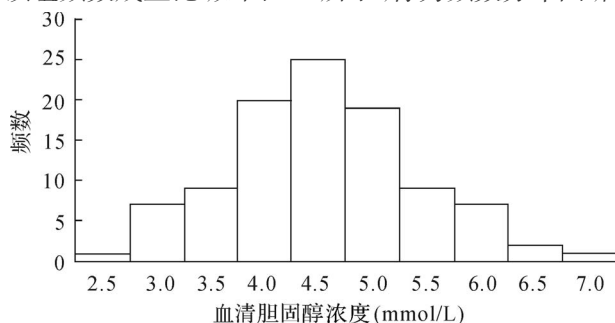


图 3-1 某地某年 100 名健康成年男子血清胆固醇浓度的频数分布图