

卫生部规划教材

WeiShengBuGuiHuaJiaoCai

全国高等医药院校教材  
供预防医学类专业用

# 卫生统计学

第四版

倪宗瓚 主编



人民卫生出版社

全国高等医药院校教材  
供预防医学类专业用

# 卫生统计学

第四版

人民卫生出版社

图书在版编目 (C I P) 数据

卫生统计学 / 倪宗瓚主编, —4 版. —北京:  
人民卫生出版社, 2000.12

ISBN 7-117-03918-3

I. 卫… II. 倪… III. 卫生学: 统计学  
IV. R 195.1

中国版本图书馆 CIP 数据核字 (2000) 第 55065 号

---

主 编: 倪 宗 瓚

出版发行: 人民卫生出版社 (中继线 67616688)

地 址: (100078) 北京市丰台区方庄芳群园 3 区 3 号楼

网 址: [http://www. pmph. com](http://www.pmph.com)

E - mail: [pmph @ pmph. com](mailto:pmph@pmph.com)

印 刷: 北京市安泰印刷厂

经 销: 新华书店

开 本: 850×1168 1/16 印张: 20.75

字 数: 451 千字

版 次: 1978 年 12 月第 1 版 2001 年 11 月第 4 版第 32 次印刷

印 数: 543 971—573 985

标准书号: ISBN 7-117-03918-3/R·3919

定 价: 26.50 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究

(凡属质量问题请与本社发行部联系退换)

# 全国高等医药院校预防医学专业第四轮 规划教材修订说明

为适应我国高等医药院校预防医学教育的改革和发展需要，在总结前三轮预防医学专业教材编写经验的基础上，卫生部教材办公室于1997年5月决定进行第四轮教材修订，根据预防医学专业的培养目标，确定了第四轮教材的品种和修订的指导思想。本轮教材修订紧密围绕培养目标，突出预防医学专业各学科的基本理论、基本知识、基本技能，同时又反映学科的新进展。

全套教材共9种，均经卫生部聘任的全国预防医学专业教材评审委员会审定。  
教材目录如下：

- |                   |       |
|-------------------|-------|
| 1. 流行病学（第四版）      | 李立明主编 |
| 2. 劳动卫生与职业病学（第四版） | 梁友信主编 |
| 3. 营养与食品卫生学（第四版）  | 陈炳卿主编 |
| 4. 环境卫生学（第四版）     | 陈学敏主编 |
| 5. 儿童少年卫生学（第四版）   | 叶广俊主编 |
| 6. 卫生化学（第四版）      | 胡曼玲主编 |
| 7. 卫生统计学（第四版）     | 倪宗瓚主编 |
| 8. 卫生毒理学基础（第三版）   | 张桥主编  |
| 9. 社会医学           | 龚幼龙主编 |

## 全国预防医学专业教材第二届评审委员会

名誉主任委员 刘世杰

主任委员 陈学敏

副主任委员 王天根

委员（以姓氏笔画为序）

孙秀发 刘宝林 郑履康 姚志麒 倪宗瓚

## 再版说明

根据全国高等医药院校预防医学专业教材评审委员会的要求，在总结前三版教材经验的基础上，为适应当前我国改革开放和社会发展的要求，我们本着继承和发展的精神对第三版教材进行了改编，主要有以下特点：

1. 本着既要避免学科之间不必要的重复，又要将一些基本概念、基本知识讲透，以利于学生自学的原则，近 10 年学科发展迅速，许多新的内容有待补充。因此，本版内容有所扩大，由原来的十六章增加为十八章。将随访资料的生存分析单列，并增加了常用的综合评价方法一章，从方法学的角度介绍各种方法的意义和步骤；增加了当前医学领域中常用的多元回归、logistic 回归、Cox 回归；将原有的医学人口统计和疾病统计合并。保留常用的基本内容，其他各章的内容根据当前学科的发展，也作了相应的增减，以利于学生在学习相关专业知识和实际工作时参考。

2. 培养学生综合分析和独立思维能力。本书继承了卫生统计学一贯的叙述方法，从实例入手，给出概念和推理，重点介绍运用，对数理统计的公式推导和理论不作深究，书中打 \* 号的章节可作为学生自学参考的内容，和本书同时出版的还有卫生统计学习题集等配套教材，有关计算器和计算机的使用方法，以及例题分析，将在习题集中着重介绍。

本书在编写过程中，得到华西医科大学和湖南医科大学校领导的高度重视和大力支持，华西医科大学教务处和公共卫生学院领导在条件和物力上给予了高度重视和支持。前三版主编杨树勤教授在整个编写过程中给予指导和帮助；湖南医科大学黄振南教授、浙江医科大学谢隆化教授、新疆医学院徐秦副教授在百忙中抽出时间参加编写会议，对本教材编写提出宝贵意见；华西医科大学的张菊英副教授，在读博士生安徽医科大学吕桦副教授、朱彩蓉硕士、田晓燕硕士、王柏松硕士对教材书稿的计算机录入、修订以及书中例题的复核做了大量细致的工作。本书学术秘书李晓松副教授在联络各编委、筹备会议以及教材定稿编排等方面做了大量工作，谨此致以衷心的感谢。

本次改稿全体编委力图有所创新和突破，做了大量认真细致的工作，但限于编者的水平，难免存在缺点和错误，欢迎广大读者批评指正。

倪宗瓚

2000 年 5 月

# 目 录

<b>第一章 绪论</b> .....	1
第一节 卫生统计学的定义和内容.....	1
第二节 统计工作的步骤.....	2
第三节 统计学中的几个基本概念.....	2
第四节 学习卫生统计学应注意的问题.....	4
<b>第二章 定量资料的统计描述</b> .....	6
第一节 频数分布表与频数分布图.....	6
第二节 集中趋势的描述.....	10
第三节 离散程度的描述.....	14
<b>第三章 正态分布</b> .....	19
第一节 正态分布的概念和特征.....	19
第二节 标准正态分布.....	21
第三节 正态分布的应用.....	23
<b>第四章 总体均数的估计和假设检验</b> .....	26
第一节 均数的抽样误差与标准误.....	26
第二节 $t$ 分布.....	27
第三节 总体均数的估计.....	28
第四节 假设检验的基本步骤.....	31
第五节 $t$ 检验和 $u$ 检验.....	34
第六节 I 型错误和 II 型错误.....	43
第七节 假设检验应注意的问题.....	45
第八节 可信区间与假设检验的区别和联系.....	45
<b>第五章 方差分析</b> .....	48
第一节 方差分析的基本思想.....	48
第二节 完全随机设计的单因素方差分析.....	50
第三节 随机区组设计的两因素方差分析.....	53
第四节 多个样本均数间的多重比较.....	57

第五节 变量变换 .....	61
<b>第六章 分类资料的统计描述 .....</b>	<b>63</b>
第一节 常用相对数 .....	63
第二节 应用相对数时应注意的问题 .....	64
第三节 标准化法 .....	66
第四节 动态数列及其分析指标 .....	70
<b>第七章 二项分布与 Poisson 分布及其应用 .....</b>	<b>72</b>
第一节 二项分布的概念 .....	72
第二节 二项分布的应用 .....	75
第三节 Poisson 分布的概念 .....	78
第四节 Poisson 分布的应用 .....	81
<b>第八章 <math>\chi^2</math> 检验 .....</b>	<b>84</b>
第一节 四格表资料的 $\chi^2$ 检验 .....	84
第二节 配对四格表资料的 $\chi^2$ 检验 .....	88
第三节 行 $\times$ 列表的 $\chi^2$ 检验 .....	89
* 第四节 行 $\times$ 列表的 $\chi^2$ 分割法 .....	92
第五节 频数分布拟合优度的 $\chi^2$ 检验 .....	92
附：四格表的确切概率法 .....	93
<b>第九章 秩和检验 .....</b>	<b>96</b>
第一节 配对设计差值的符号秩和检验 (Wilcoxon 配对法) .....	96
第二节 成组设计两样本比较的秩和检验 (Wilcoxon 两样本比较法) .....	98
第三节 成组设计多个样本比较的秩和检验 (Kruskal-Wallis 法) .....	100
第四节 多个样本两两比较的秩和检验 (Nemenyi 法) .....	103
第五节 随机区组设计资料的秩和检验 .....	104
第六节 随机区组设计资料的两两比较 .....	107
附：Ridit 分析 .....	108
<b>第十章 回归与相关 .....</b>	<b>114</b>
第一节 直线回归 .....	114
第二节 直线相关 .....	122
第三节 直线回归与相关的区别和联系 .....	124
* 第四节 曲线直线化 .....	125
第五节 秩相关 .....	128

第六节	秩回归	129
<b>第十一章</b>	<b>多元线性回归与 logistic 回归</b>	<b>132</b>
第一节	多元线性回归	132
第二节	logistic 回归	138
<b>第十二章</b>	<b>常用统计图表</b>	<b>143</b>
第一节	常用统计表	143
第二节	常用统计图	144
<b>第十三章</b>	<b>实验设计</b>	<b>148</b>
第一节	实验设计的特点和分类	148
第二节	实验设计的基本原则	149
第三节	实验设计的基本内容和步骤	150
第四节	常用的实验设计方法	156
第五节	确定样本含量	159
附：1.	临床科研设计书的主要内容	163
2.	临床试验框图模式	164
3.	一致性检验	165
<b>第十四章</b>	<b>调查设计</b>	<b>168</b>
第一节	调查研究的特点	168
第二节	调查设计的基本原则与内容	168
第三节	常用的抽样方法	174
第四节	调查的质量控制	179
附：	调查设计的基本步骤框图	181
<b>第十五章</b>	<b>医学人口统计与疾病统计常用指标</b>	<b>182</b>
第一节	医学人口统计常用指标	182
第二节	疾病统计常用指标	192
<b>第十六章</b>	<b>寿命表</b>	<b>197</b>
第一节	寿命表的概念	197
第二节	寿命表的编制原理与方法	198
第三节	简略寿命表	199
第四节	去死因寿命表	201
第五节	寿命表的分析与应用	205



第十七章	随访资料的生存分析	209
第一节	生存分析的基本概念	209
第二节	生存率及其标准误	211
第三节	对数秩检验	214
*第四节	Cox 回归分析	216
第十八章	常用综合评价方法	223
第一节	综合评价概述	223
第二节	几种综合评价方法简介	231
附 1	统计用表	245
附表 1	标准正态分布曲线下的面积, $\Phi(-u)$ 值	245
附表 2	$t$ 界值表	246
附表 3	$F$ 界值表	247
附表 4	$q$ 界值表 (Newman-Keuls 法用)	251
附表 5	Dunnet $t$ - $t$ 检验临界值表	252
附表 6	百分率的可信区间	254
附表 7	Poisson 分布 $\mu$ 的可信区间	257
附表 8	$\chi^2$ 界值表	258
附表 9	$T$ 界值表 (配对比较的符号秩和检验用)	259
附表 10	$T$ 界值表 (两样本比较的秩和检验用)	260
附表 11	$H$ 界值表 (三样本比较的秩和检验用)	261
附表 12	$M$ 界值表 (随机区组比较的秩和检验用)	262
附表 13	$D$ 界值表 (各样本例数相等的 Nemenyi 法用)	263
附表 14	$r$ 界值表	264
附表 15	$r_s$ 界值表	265
附表 16	随机排列表 ( $n=20$ )	266
附表 17	随机数字表	267
附 2	实习题	268
第一单元	数值变量的统计描述	268
第二单元	数值变量的统计推断	270
第三单元	分类变量的统计描述与统计推断	276
第四单元	秩和检验	283
第五单元	直线回归与相关	286
第六单元	统计图表	291
第七单元	实验设计和调查设计	293

第八单元 健康统计.....	296
第九单元 多因素分析.....	302
第十单元 综合分析及医学综合评价.....	307
附 3 英汉卫生统计学词汇 .....	312

# 第一章 绪 论

## 第一节 卫生统计学的定义和内容

当今的人类社会进入了科技高速发展的时代,信息的传播和交流日益加快,为适应时代发展的要求,必须加快认识世界、改造世界的步伐。统计学(statistics)正是研究数据的收集、整理、分析的一门科学,帮助人们分析占有信息,达到去伪存真、去粗取精、正确认识世界的一种重要手段。

医学、卫生领域的研究对象是人,故有其特殊性,具有较大的生物变异性,并受诸多社会、心理因素的影响,更需要借助统计分析,透过偶然现象认识其内在的规律性,以推动各项医疗卫生事业的发展。

卫生统计学(health statistics)是研究居民健康状况以及卫生服务领域中数据的收集、整理和分析的一门科学,它的内容包括:①健康统计:医学人口统计、疾病统计和生长发育统计;②卫生服务统计:包括卫生资源利用、医疗卫生服务的需求、医疗保健体制改革等方面的统计学问题。但是,上述卫生统计学的研究内容都必须建立在统计学的基本原理和方法的基础上。因此,学生必须首先掌握它的基本原理和方法,才能透彻理解和应用卫生统计学的基本内容,解决各自专业领域中的实际问题。而统计研究设计(包括调查研究设计和实验研究设计)是公共卫生领域的学生必须学习的内容。随着社会的进步和医学科学的发展,对公共卫生(包括预防医学、妇幼卫生、卫生管理)医师提出了更高的要求,不仅要懂得各种统计方法的应用,还应具备统计研究设计的能力,以适应今后社会对人才的多方面的需求。

这本教材是根据预防医学专业的教学计划而编写的,其内容包括:①卫生统计学的基本原理和方法,含研究设计、统计分析和评价中的理论和方法(第二至十四章);②健康统计(第十五章),医学人口与疾病统计中常用的指标(第十五章),寿命表(第十六章),生存率分析(第十七章);③常用的综合评价方法(第十八章)。根据教学计划的安排,将卫生统计学中的生长发育统计加入儿少卫生学,医疗保健制度以及卫生管理中的部分统计内容列入社会医学。

随着电子计算机的发展和普及,使得大量信息的存储成为现实,许多必须通过大量繁杂的计算才能实现的统计分析方法,能够应用于医学领域,进一步促进了卫生统计学的发展。

要做好预防保健工作,公共卫生医师必须经常深入现场进行调查研究,了解人群的健康状况,发现存在的问题,提出有针对性的预防措施,评价、总结各种预防措施的效果,并且能够根据工作的需要开展相应的研究工作,为此预防医学专业学生必须学好卫生统计学这门课程。

## 第二节 统计工作的步骤

统计学是统计工作实践经验的总结,又对统计工作的全过程起指导作用,任何统计工作和统计研究的全过程都可分为以下四个步骤:

1. 设计(design) 在进行统计工作和研究工作之前必须有一个周密的设计。设计之前,当然应先对研究的问题有较多的了解。为此,需要广泛查阅文献,了解实际情况,而且常要与有关专家共同协作。设计的内容包括资料收集、整理和分析全过程总的设想和安排。例如,明确研究目的和研究假说、观察对象和观察单位,收集哪方面的原始资料,收集资料的方法等,列出拟分析的统计指标以及控制误差和偏倚的措施,预期拟得到的结果,经费预算等等,都要结合实际,周密考虑,妥善安排。设计是整个研究中最关键的一环,也是今后工作应遵循的依据。详见第 13 章和第 14 章。

2. 收集资料(collection of data) 应采取措施使能取得准确可靠的原始数据。卫生工作中的统计资料主要来自三个方面:①统计报表。如法定传染病报表,职业病报表,医院工作报表等。由国家统一设计,要求有关医疗卫生机构定期逐级上报,提供居民健康状况和医疗卫生机构工作的主要数据,作为制定卫生工作计划与措施、检查与总结工作的依据,报表要做到完整、准确、及时。首先要保证基础资料的质量,如法定传染病报告卡是法定传染病报表的基础资料,要提高基层卫生人员的认识和责任感,要重视对漏报、重报和错报的检查。②经常性工作记录。如经常性的卫生监测记录、健康检查记录等。要做到登记的完整、准确。病历是医疗工作的重要记录,分析时应注意其局限性(如不能反映一般人群特征)。③专题调查或实验(详见第 13 章和第 14 章)。

3. 整理资料(sorting data) 任务是净化原始数据,使其系统化、条理化,便于进一步计算和分析。首先是资料清理(data cleaning),因为无论是调查或实验的原始记录和计算机录入过程,常会有错误,必须经过反复地检查和核对,这是需要耐心从事的基础工作,特别是数据较多时。一定对数据进行逻辑检查,以便纠正错误,再按分析要求,用计算机汇总资料。

4. 分析资料(analysis of data) 目的是计算有关指标,反映数据的综合特征(亦称综合指标),阐明事物的内在联系和规律。统计分析包括:①统计描述(descriptive statistics),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和描述;②统计推断(inferential statistics),指如何抽样,以及如何由样本信息推断总体特征问题。详见第 2 章至第 11 章以及第 17 章。

虽然人为地将统计工作分为以上四个步骤,但它们是紧密联系、不可分割的整体,缺少任何一步,都会影响整个研究的结果。

## 第三节 统计学中的几个基本概念

1. 总体(population)与样本(sample)任何统计研究都必须首先确定观察单位(ob-

served unit),亦称个体(individual),它是统计研究中最基本的单位,它可以是一个人、一个家庭、一个地区、一个样品、一个采样点等,这些根据研究目的确定的同质观察单位的全体称为总体。总体是根据研究目的确定的同质观察单位的全体,更确切地说,是同质的所有观察单位某种观察值(变量值)的集合。例如调查某地 1999 年正常成年男子的红细胞数,则观察对象是该地 1999 年的正常成年男子,观察单位是每个人,观察值(变量值)是每人测得的红细胞数,该地 1999 年全部正常成年男子的红细胞数就构成一个总体。它的同质基础是同一地区、同一年份、同为正常成人,同为男性。这里的总体明确了确定的时间,空间范围内有限个观察单位,称为有限总体(finite population)。有时总体是抽象的,如研究用某药治疗缺铁性贫血的疗效,这里总体的同质基础是贫血患者,同时用某药治疗,该总体应包括用该药治疗的所有贫血患者的治疗结果,是没有时间和空间范围限制的,因而观察单位数无限,称为无限总体(infinite population)。

在许多情况下医学研究的总体是无限总体,要直接观察总体的情况是不可能的。即使对有限总体来说,若包含的观察单位过多,也要花费很大的人力、财力,有时也是不必要的和不可能的。又如乳制品(奶粉、奶酪等)的卫生检查,不可能将所有生产的乳制品打开一一加以检验。所以在实际工作中经常是从总体中抽取样本,是用样本信息来推断总体特征。样本是总体中随机抽取部分观察单位,其实测值的集合。如上例,可从某地 1999 年的正常成年男子中,随机抽取 200 人,分别测得其红细胞数,组成样本。所谓随机抽样,就是按随机化原则(即总体中每一个观察单位都有同等的机会被选入到样本中来)获取样本,以避免误差和偏倚对研究结果有所影响。抽样的方法有多种,详见第 14 章。样本包含的观察单位数称为样本含量或样本大小(sample size),医学上也常称为样本例数。

2. 资料 and 变量(data and variable)在确定总体之后,研究者则应对每个观察单位的某项特征进行测量和观察,这种特征称为变量。对变量的测得值称为变量值(value of variable)或观察值(observed value),亦称为资料。例如,以人为观察单位调查三岁儿童的生长发育情况,儿童的性别变量分为男性和女性,身高变量有高有矮。又如同窝别的小白鼠,用同样的饲料喂养,经过一段时间观察,各只体重增加的重量是不等的,这种个体间的差异,称为变异(variation)。这些变异来源于一些已知的或未知的、甚至是不可控制的因素所导致的随机误差。变量的观察结果可以是定量的,也可是定性的,通称为变量值(value of variable)或观察值(observed value, observation),亦称为资料。按资料(变量值)是定量的还是定性的,可将它们分为以下类型,以便研究者根据不同类型的资料采用相应的统计分析方法。

(1) 定量资料(quantitative data):亦称数值变量(numerical variable),其变量值是定量的,表现为数值大小,一般有度量衡单位,如调查某地某年七岁女童的身体发育状况,以人为观察单位,每个人的身高(cm)、体重(kg)、血压(kPa)、坐高指数(% ,坐高/身高)等均属定量资料。

(2) 分类资料(categorical data)或分类变量(categorical variable):亦称定性资料(qualitative data),其观察值是定性的,表现为互不相容的类别或属性,有两种情况:

1) 无序分类(unordered categories): 包括: ①二项分类。如检查某小学学生大便中的蛔虫卵, 以每个学生为观察单位, 结果可以是蛔虫卵阳性或阴性; 又如观察用某药治疗某病患者的治疗结果, 以每个患者为观察单位, 结果分为治愈与未愈两类。两类间互相对立。②多项分类。如观察某人群的血型, 以人为观察单位, 结果分为 A 型、B 型、AB 型与 O 型, 为互不相容的多个类别。无序分类变量的分析, 应先分类汇总, 计观察单位数, 编制分类资料的频数表, 亦称计数资料。

2) 有序分类(ordinal categories): 各类之间有程度的差别, 给人以“半定量”的概念, 亦称等级资料(ordinal data)。如测定某人群血清反应, 以人为观察单位, 结果可分 -、±、+、++ 四级; 又如观察用某药治疗某病患者的治疗结果, 以每个患者为观察单位, 结果分为治愈、显效、好转、无效四级。有序分类资料的分析, 应先按等级顺序, 分类汇总, 计观察单位数, 编制等级资料的频数表。

实际上, 资料类型的划分, 是根据研究目的而确定的。根据需要, 各类变量可以互相转化。如以人为观察单位观察某人群成年男子的血红蛋白量(g/L), 属数值变量; 若按血红蛋白正常与偏低分为两类, 可按二项分类资料处理; 若按贫血的诊断标准将血红蛋白含量分为五个等级: 重度贫血、中度贫血、轻度贫血、正常、血红蛋白增高, 可按等级资料处理。有时亦可将分类资料数量化, 如将多项分类资料的治疗结果赋以分值, 分别用 0、1、2... 等表示, 则可按定量资料处理。

3. 概率(probability) 医学研究的现象, 大多数是随机现象。例如用相同治疗方法治疗某病的一群患者, 只知道治疗转归可能为治愈、好转、无效、死亡四种结果, 但对一个刚入院的该病患者, 治疗后究竟发生哪一种结果是不确定的。这里的每一种可能发生的结果都是一个随机事件, 亦称偶然事件, 简称事件。概率是描述随机事件发生的可能性大小的数值, 常用  $P$  表示。比如某例患者的结果为“治愈”, 这个事件记为  $A$ , 则该患者治愈的概率可记为  $P(A)$ , 或简记为  $P$ , 这是一个很有意义的, 医生颇为关心的数值。假如我们用 200 例的样本, 求得治愈率为 75%, 这只是一个频率。在实际工作中, 当概率不易求得时, 只要观察单位数充分多, 可以将频率作为概率的估计值。但在观察单位数较少时, 频率的波动性是很大的, 用于估计概率是不可靠的。

随机事件概率的大小在 0 与 1 之间, 即  $0 \leq P \leq 1$ , 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。严格说,  $P=1$ , 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, 但可把它们看成随机事件的特例。统计分析中的很多结论都是带有概率性的。习惯上将  $P \leq 0.05$ , 称为小概率事件, 表示在一次实验或观察中该事件发生的可能性很小, 可以视为很可能不发生。关于概率的基本概念和有关的知识在本书的其他章节还会进一步深入讨论。

#### 第四节 学习卫生统计学应注意的问题

本课程的教学目的是为学生在校学习专业课程, 毕业后从事公共卫生领域的研究和

实际工作,打下必要的卫生统计学基础。为此,学习本课程时,应注意:

1. 掌握卫生统计学的基本知识、基本技能、基本概念和基本方法。本课程许多内容都能帮助医学生建立逻辑思维方法和提高分析问题的能力。例如,由于事物存在个体差异,用样本推断总体就会出现误差,但这种误差是有规律性的,据此可引出统计推断的理论;懂得了假设检验的逻辑推理,就能理解统计结论的概率性。

2. 掌握调查设计和实验设计的原则,培养收集、整理、分析统计资料的系统工作能力。首先要重视原始资料的完整性和准确性,对数据处理持严肃、认真、实事求是的科学态度,反对伪造和篡改统计数字;要能够应用恰当的统计方法处理各种数据,对统计公式只要求了解其意义、用途和应用条件,不必深究其数学推导。

3. 掌握群体健康的评价方法,学会用医学人口统计和疾病统计等方面的统计指标,综合评价人群健康状况,为卫生决策提供统计信息。

总之,应多联系实际,结合专业,分析评价实际工作、医学文献和医学科研中的统计问题,才能学好卫生统计学。

(倪宗瓚)

## 第二章 定量资料的统计描述

定量资料又称为测量资料、计量资料,它是测量每个观察单位某项指标值大小所得的资料,一般均有计量单位。例如,身高(cm),体温(℃),年龄(岁),牙齿数(个),已婚育龄妇女现有子女数(个)等。

实际工作中,往往掌握较多的观察数据,为了解数据的分布规律,需借助统计方法加以描述。常用的描述定量资料分布规律的统计方法有两大类,一类是用统计图表,主要是频数分布表(图),这是本章第一节讲述的内容;另一类是选用适当的统计指标,本章第二、三节将分别介绍常用的、描述定量资料集中趋势和离散程度的统计指标。

### 第一节 频数分布表与频数分布图

#### 一、频数分布表

##### (一)频数表的编制

例 2.1 某市 1995 年 110 名 7 岁男童的身高(cm)资料如下,请作统计描述。

114.4 119.2 124.7 125.0 115.0 112.8 120.2 110.2 120.9 120.1  
125.5 120.3 122.3 118.2 116.7 121.7 116.8 121.6 115.2 122.0  
121.7 118.8 121.8 124.5 121.7 122.7 116.3 124.0 119.0 124.5  
121.8 124.9 130.0 123.5 128.1 119.7 126.1 131.3 123.8 114.7  
122.2 122.8 128.6 122.0 132.5 122.0 123.5 116.3 126.1 119.2  
126.4 118.4 121.0 119.1 116.9 131.1 120.4 115.2 118.0 122.4  
114.3 116.9 126.4 114.2 127.2 118.3 127.8 123.0 117.4 123.2  
119.9 122.1 120.4 124.8 122.1 114.4 120.5 115.0 122.8 116.8  
125.8 120.1 124.8 122.7 119.4 128.2 124.1 127.2 120.0 122.7  
118.3 127.1 122.5 116.3 125.1 124.4 112.3 121.3 127.0 113.5  
118.8 127.6 125.2 121.5 122.5 129.1 122.6 134.5 118.3 132.8

直接从上面 110 个数据难以得出关于这 110 名 7 岁男童身高的概括印象。现制作频数分布表,步骤如下:

1. 求数据的极差(range) 极差是全部数据中的最大值与最小值之差,它描述了数据变异的幅度。本例最大值为 134.5cm,最小值为 110.2cm,故极差

$$R = 134.5\text{cm} - 110.2\text{cm} = 24.3\text{cm}$$

2. 划分组段



(1)确定组数:制作频数表是为了简化资料,显示出数据的分布规律,故组数不宜过多,但组数也不能太少,组数太少会掩盖数据分布的规律。适宜的分组数与观察值个数  $n$  的多少有关,大体上当观察值个数  $n$  在 30 左右时,可分 5 到 6 组,随着  $n$  的增加,分组数适当增加。较大样本时,一般取 10 组左右。

(2)确定组距:组距可以相等,也可以不相等。实际应用时一般采用等距分组,其组距近似等于(极差/组数),本例极差/组数 =  $(24.3/10) \approx 2$ ,故取组距 = 2cm。

(3)确定各组段的上下限:每个组段的起点被称为该组的下限(low limit),终点被称为上限(upper limit),上限 = 下限 + 组距。显然第一组段必须包含最小值,其下限一般取包含最小值的、较为整齐数值。本例最小值为 110.2,故取 110 为第一组段的下限,其上限 =  $110 + 2 = 112$ 。值得注意的是各组段不能重叠,故每一组段均为半开半闭区间。表 2.1 第(1)栏中的“110~”表示  $[110, 112)$ ,”112~”表示  $[112, 114)$ ,以此类推。

3. 统计各组段内的数据频数 统计出各组段内的数据个数(频数)即得频数表。本例的频数表见表 2.1 第(1)、(2)栏。

表 2.1 110 名 7 岁男童身高频数分布

身 高 组 段 (1)	频 数 (2)	频 率 (%) (3)	累 计 频 数 (4)	累 计 频 率 (%) (5)
110~	1	0.91	1	0.91
112~	3	2.73	4	3.64
114~	9	8.18	13	11.82
116~	9	8.18	22	20.00
118~	15	13.64	37	33.64
120~	18	16.36	55	50.00
122~	21	19.09	76	69.09
124~	14	12.73	90	81.82
126~	10	9.09	100	90.91
128~	4	3.64	104	94.55
130~	3	2.73	107	97.27
132~	2	1.82	109	99.09
134~136	1	0.91	110	100.00
合计	110	100.00	-	-

例 2.1 中的指标身高是一个连续变化的量。这种连续变化的变量被称为连续型变量。已婚育龄妇女的现有子女数、幼儿的牙齿数等则不然,其取值是 0、1、2 等不连续的量,这种变量被称为离散型变量。上面所讲的频数表的制作方法是针对连续型变量的数据资料而言的。对于离散型变量,其频数表的编制较为简单,每一组段往往是一个取值。如表 2.2 的第(1)、(2)栏。