

卫生部
规划教材

全国成人

高等医学学历(专科) 教育教材

供预防医学专业用

卫生统计学

卫生部教材办公室组织编写
马燕 主编



人民卫生出版社

全国成人高等医学学历(专科)教育教材

供预防医学专业用

卫生统计学

卫生部教材办公室组织编写

马燕 主编

编委 (以姓氏笔画为序)

丁瑾瑜 (上海职工医学院)

刘美娜 (哈尔滨医科大学)

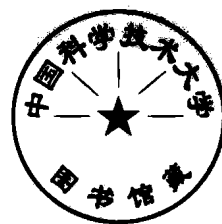
张建国 (湖南医科大学)

陈冠民 (湖北医科大学)

骆福添 (中山医科大学)

谢隆化 (浙江大学)

秘书:姜树林 (哈尔滨医科大学)



人民卫生出版社

图书在版编目 (CIP) 数据

卫生统计学/马燕主编. —北京: 人民卫生出版社,
2000

ISBN 7 - 117 - 03956 - 6

I. 卫… II. 马… III. 卫生统计 IV. R195.1

中国版本图书馆 CIP 数据核字 (2000) 第 26448 号

卫生统计学

主 编: 马 燕

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>

E - mail: pmph@pmph.com

印 刷: 三河市宏达印刷有限公司

经 销: 新华书店

开 本: 850×1168 1/16 印张: 18.5

字 数: 380 千字

版 次: 2000 年 7 月第 1 版 2003 年 2 月第 1 版第 5 次印刷

标准书号: ISBN 7-117-03956-6/R·3957

定 价: 27.00 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)

出版说明

成人医学教育是我国医学教育的重要组成部分,为加强成人医学教育教学管理,完善教学基础建设,保证教育质量,卫生部、教育部联合颁发了《全国成人高等医学学历教育主要课程目录及课程基本要求(试行)》,这是国家为实现成人医学教育培养目标和要求,根据各门课程在某一专业中地位和作用而确定的,是该专业学生在学习课程时必须达到的基本合格标准,是编审规划教材、组织对教学水平进行监督检查和评价的重要依据,是规范我国成人高等医学学历教育的重要指导性文件。为了配合这一要求的顺利实施,卫生部教材办公室成立了全国成人高等医学教育教材评审委员会,组织编写全国成人高等医学学历(专科)教育规划教材。本套教材的主编、编者从全国推荐的 600 名候选人中选出,均为一线教学人员,具有丰富的成人医学学历教育教学经验;教材内容根据《全国成人高等医学学历教育主要课程目录及课程基本要求(试行)》确定,由全国成人高等医学教育教材评审委员会审定,突出成教学员在一定工作经验基础上学习的特点,篇幅适中,针对性强。

本套教材包括 4 个专业(临床医学、预防医学、护理学、药学),共 38 种,均由人民卫生出版社出版。

临床医学、预防医学、护理学、药学专业共用

人体解剖学	孙荣鑫主编	生物化学	查锡良主编
生理学	倪江主编	卫生法学概论	樊立华主编
药理学	李元建主编		

临床医学、预防医学、护理学专业共用

病理学	李玉林主编
-----	-------

临床医学、预防医学、药学专业共用

医学微生物学与免疫学	刘晶星主编
------------	-------

临床医学、预防医学专业共用

内科学	吕卓人主编	儿科学	徐立新主编
外科学	孙靖中主编	诊断学	汤美安主编
妇产科学	李荷莲主编		

临床医学专业用

预防医学	仲来福主编	全科医学概论	顾浚主编
------	-------	--------	------

预防医学专业用

卫生化学	计时华主编	环境卫生学	王振刚主编
卫生统计学	马燕主编	营养与食品卫生学	凌文华主编
卫生毒理学	石年主编	劳动卫生与职业病学	陈自强主编
儿童少年卫生学	孙江平主编	社会医学	肖水源主编
流行病学	王建华主编		

护理学专业用

护理学基础	张景龙主编	儿科护理学	童秀珍主编
内科护理学	李改焕主编	护理管理学	成翼娟主编
外科护理学	鲁连桂主编	护理心理学	张树森主编
妇产科护理学	何仲主编		

药学专业用

高等数学	马湘玲主编	天然药物化学	吴立军主编
有机化学	田昌荣主编	药物化学	徐文方主编
物理化学	曹宗顺主编	药剂学	梁文权主编
分析化学	李发美主编	药物分析	晁若冰主编

全国成人高等医学教育教材评审委员会

主任委员：唐建武

委员：（以姓氏笔画为序）

王怀良 冯美丽 白继荣 朱立华 汤恢焕 吴仁友 吴坤
张爱珍 张鹏 李守国 李继坪 沈彬 陈金华 梁万年
董崇田 樊小力

秘书：郭明

前 言

本书是由卫生部规划、卫生部教材办公室组织全国各类高等医学院校编写的成人高等医学学历教育系列教材之一。本教材的读者对象是经过中等医学学历教育的在职卫生人员,其目的是通过本教材的学习达到预防医学专业卫生统计学专科水平。

本教材以《全国成人高等医学学历教育主要课程目录及课程基本要求(试行)》中《卫生统计学》课程基本要求和卫生统计学考试大纲为依据;注意与中专和本科教材相衔接;注重加强基本理论知识、基本思维方法、基本实践技能。力争做到从实际出发,适合成人自学的需要,循序渐进,由浅入深。

本书共分二十一章,其中基本统计方法部分共十二章,多元统计方法二章,调查与实验设计二章,人口、疾病与死亡统计三章,计算机统计软件简介和计算器的使用二章。有“*”号的章节可供选学。为适应自学的需要本书选编了部分练习题。

本书全体编委力求在科学性、实用性、可操作性等方面写出自己的特色,但限于水平,一定存在不少缺点和错误,欢迎广大读者批评指正。

本书在编写过程中得到哈尔滨医科大学成人教育学院和上海职工医学院的帮助,在此表示衷心的感谢。

编 者

2000年2月

目 录

第一章 绪论	(1)
第一节 卫生统计学的定义和内容	(1)
第二节 卫生统计工作的步骤	(1)
第三节 卫生统计中的几个基本概念	(2)
第四节 学习卫生统计学应注意的问题	(5)
第二章 数值变量资料的统计描述	(6)
第一节 数值变量资料的频数表	(6)
第二节 平均水平的描述	(8)
第三节 变异程度的描述	(12)
第三章 正态分布及其应用	(16)
第一节 正态分布的概念	(16)
第二节 正态曲线下面积分布规律	(17)
第三节 正态分布的应用	(19)
第四章 总体均数估计和假设检验	(21)
第一节 均数的抽样误差与标准误	(21)
第二节 t 分布	(22)
第三节 总体均数估计	(23)
第四节 假设检验的基本思想和步骤	(25)
第五节 t 检验和 u 检验	(27)
第六节 正态性检验	(34)
第七节 假设检验的两类错误	(36)
第八节 假设检验应注意的问题	(37)
第五章 方差分析	(38)
第一节 方差分析的基本思想	(38)
第二节 完全随机设计的多个样本均数的比较	(39)
第三节 配伍组设计的多个样本均数的比较	(41)
第四节 多个样本均数间的两两比较	(43)
第五节 基于统计量的方差分析*	(46)

第六章 分类变量资料的统计描述	(49)
第一节 常用相对数	(49)
第二节 应用相对数应注意的问题	(50)
第三节 标准化法	(51)
第四节 动态数列	(56)
第七章 二项分布及其应用	(58)
第一节 二项分布的概念	(58)
第二节 总体率的估计与假设检验	(61)
第八章 泊松分布及其应用	(65)
第一节 泊松分布的概念	(65)
第二节 泊松分布总体均数的估计与假设检验	(68)
第九章 χ^2 检验	(71)
第一节 四格表 χ^2 检验	(71)
第二节 行 \times 列表资料的 χ^2 检验	(74)
第三节 列联表资料的 χ^2 检验	(76)
第四节 频数分布的拟合优度检验*	(80)
第五节 四格表的确切概率法*	(81)
第十章 秩和检验	(84)
第一节 配对设计资料的符号秩和检验(Wilcoxon 配对法)	(84)
第二节 完全随机设计两样本比较的秩和检验(Wilcoxon 两样本比较法)	(86)
第三节 完全随机设计多样本比较的秩和检验(Kruskal-Wallis 法)	(89)
第四节 完全随机设计多样本两两比较的秩和检验(Nemenyi 法)	(93)
第五节 随机区组设计多样本比较的秩和检验*	(94)
第十一章 直线相关与回归	(96)
第一节 直线相关	(96)
第二节 直线回归	(99)
第三节 应用直线相关和回归应注意的问题	(104)
第四节 等级相关	(105)
第十二章 多元线性回归简介*	(107)

第一节	多元线性回归的基本概念	(107)
第二节	多元线性回归分析实例	(108)
第三节	多元线性回归分析的常用评价指标	(110)
第四节	回归分析中自变量的筛选	(111)
第五节	多元线性回归的应用	(112)
第十三章	logistic 回归简介 *	(113)
第一节	logistic 回归基本概念	(113)
第二节	非条件 logistic 回归分析实例	(114)
第三节	非条件 logistic 回归的应用	(116)
第十四章	常用统计表和统计图	(118)
第一节	统计表	(118)
第二节	统计图	(121)
第十五章	调查设计	(128)
第一节	调查设计的意义	(128)
第二节	调查的统计设计	(129)
第三节	抽样方法	(131)
第四节	样本含量的估计	(135)
第五节	调查研究的误差及控制	(137)
第十六章	实验设计	(139)
第一节	实验设计的意义	(139)
第二节	实验设计的要素及原则	(139)
第三节	样本含量的估计	(144)
第四节	常用医学实验设计方法	(147)
第五节	试验误差及其控制	(150)
第十七章	人口统计	(153)
第一节	人口统计的意义及其资料的收集	(153)
第二节	静态人口统计	(154)
第三节	出生与计划生育统计	(156)
第十八章	疾病统计	(161)
第一节	疾病统计的意义及其资料的搜集	(161)
第二节	疾病和死因分类(ICD-10)	(162)

第三节	疾病统计常用指标	(169)
第四节	随访资料的生存率分析	(170)
第十九章	死亡统计	(175)
第一节	死亡统计的意义及其资料的搜集	(175)
第二节	死亡统计的常用指标	(176)
第三节	简略寿命表	(180)
第二十章	统计软件简介	(183)
第一节	常用统计软件包简介	(183)
第二节	SPSS 软件的基本使用方法	(188)
第三节	SAS 软件的基本使用方法	(195)
第二十一章	计算器的使用	(204)
第一节	键盘功能及名称	(204)
第二节	统计计算应用举例	(208)
第三节	其他类型计算器使用方法简介	(210)
第四节	使用计算器的注意事项	(211)
附 I	统计用表	(212)
附表 1	标准正态分布曲线下的面积($\Phi(-u)$ 值)	(212)
附表 2	t 界值表	(213)
附表 3	正态性检验计算 W 系数表	(215)
附表 4	正态性检验计算 W 界值表	(218)
附表 5	F 界值表(方差齐性检验用)	(219)
附表 6	F 界值表(方差分析用)	(221)
附表 7	q 界值表	(229)
附表 8	Dunnett 检验用 t_D 界值表	(230)
附表 9	百分率的可信区间	(232)
附表 10	Poisson 分布的可信区间	(235)
附表 11	χ^2 界值表	(236)
附表 12	T 界值表(配对比较的符号秩和检验用)	(237)
附表 13	四格表的显著性检验用表(C 值表)	(238)
附表 14	T 界值表(两样本比较的秩和检验用)	(254)
附表 15	H 界值表(三样本比较的秩和检验用)	(255)
附表 16	D 界值表(各样本例数相等的 Nemenyi 法用)	(256)
附表 17	M 界值表	(258)

附表 18	r 界值表	(259)
附表 19	r_s 界值表	(261)
附表 20	配对比较(t 检验)时所需样本例数	(263)
附表 21	两样本均数比较(t 检验)时所需样本例数	(264)
附表 22	两样本率比较时所需样本例数	(265)
附表 23	随机数字表	(267)
附表 24	随机排列表($n = 20$)	(269)
附 II	练习题	(270)
第一单元	数值变量资料(计量资料)的统计描述与推断(第二至五章)	(270)
第二单元	分类资料的统计描述与推断(第六至九章)	(274)
第三单元	秩和检验(第十章)	(277)
第四单元	直线回归与相关(第十一章)	(279)
第五单元	统计表与统计图(第十四章)	(281)
第六单元	调查设计与实验设计(第十五、十六章)	(282)

第一章 绪 论

第一节 卫生统计学的定义和内容

统计学(statistics)是研究随机现象的统计规律性的科学。经常有这样的情况,在同样条件下多次进行同一试验或观察同一现象,结果总是不完全一样,比如观察同性别、同年龄的健康人,他们的身高却不一样,类似这样的现象叫随机现象,这种个体间的差异称为变异。条件一致,结果不确定是随机现象的一个特点。随机现象产生的原因是由于受许许多多无法控制的偶然因素的影响所导致的个体差异和偶然测量误差,将其统称为随机误差。随机误差虽不可避免但却有很强的统计规律性,统计规律性是随机现象的另一个特点。这种规律性只有对同一现象在相同条件下做大量观察或实验时,才能表现出来。正象我们在相同条件下对某一年龄的男孩身高进行观察一样,从几个零星的观测数据看不出有什么规律,但当观察的人数逐渐增多时,就会看到这些男孩身高的分布呈现某种规律性,观察的人数越多,这种规律性就越明显。统计学就是在搜集、整理、分析和解释大量数据的过程中,探索这种偶然变化的规律性,并作出科学推断的一门学科。可见统计学是认识世界的一个重要手段。

卫生统计学(health statistics)是运用统计学的基本原理和方法,探索生命科学领域的内部规律,侧重于研究居民健康状况以及与健康有关的各种因素及其相互关系和卫生事业管理中数据的搜集、整理与分析的一门应用科学。卫生统计学的主要内容包括:卫生统计学的基本原理和方法、健康统计、卫生服务统计以及调查设计和实验设计等。随着医学科技的进步,学科间的相互渗透以及边缘学科的兴起,数学与电子计算机在医学研究中的应用日益广泛,使数理统计方法成为必不可少的手段。学习卫生统计学的目的在于建立统计思维方法;提高科研和工作能力,是继续教育的需要。

近二十年来我国各级卫生部门逐步建立、健全了卫生统计信息系统,全面发挥统计工作的统计信息功能、统计咨询功能和统计监督功能。使用先进的统计分析方法和计算手段,搜集、整理、传递和提供信息;为科学管理与决策提供依据、建议和对策;对存在的问题及时进行监督评价。这些对卫生统计工作提出了更高的要求。为适应医学和卫生管理的科技进步,为以后能够更好地开展工作,必须要学好卫生统计学。

第二节 卫生统计工作的步骤

统计是一门科学,它是对观察到的原始数据资料进行加工、解释、作出科学判断

的全过程。统计工作的基本步骤是：

1. 设计(design) 首先明确研究目的。根据研究目的,从统计角度对资料的搜集、整理和分析全过程提出全面具体的计划和要求,作为统计工作实施的依据,以使用尽可能少的人力、物力和时间获得准确可靠的结论。统计设计是保证统计工作顺利地完成最关键的一个环节,实施中的细节见第十五章和十六章。

2. 搜集资料(collection of data) 及时、准确、完整地搜集原始资料是统计工作最重要的一步,它直接关系到统计结论的质量。卫生统计资料的来源主要有:①报表资料:如医院工作年报表,传染病、疫情、居民出生、病伤死亡报表等。这些都是依据国家规定的报表制度由医疗卫生机构逐级上报。统计报表是了解居民健康状况的基础资料,为各级卫生决策部门提供了科学依据。②医疗、预防机构的日常工作记录。③专题研究实验数据:指根据研究工作的需要,通过实验的方法获取的数据。一般实验均有较好的设计,数据量不大,问题较少。④现场调查资料:通常是依据一定的研究目的制定调查表格,通过现场调查获取的资料。由于数据量大,参加调查的人员多,容易出现的问题也较多。在统一审核和规范调查表的内容、统一调查方法等方面均需作精细的安排。

3. 整理资料(sorting data) 整理资料是把搜集到的资料进行适当的分组,把性质相同的资料归纳到一起,用表格或图形的方式展示出来,以反映研究对象的规律性。整理的过程中要核对原始资料的准确性、完整性和可靠性,注意看其是否合乎逻辑,合计是否有误。出现差错要及时纠正。

4. 分析资料(analysis of data) 资料的分析过程是通过计算有关的统计指标,对资料进行概括的、全面的描述,以及从样本信息推断总体特征,如估计总体参数、进行假设检验、进行预测等等。分析资料就是从获取的资料中提取有关信息的过程。

第三节 卫生统计中的几个基本概念

一、变 量

进行统计研究,首先要确定观察单位(observation unit),即个体(individual),然后对每个观察单位的若干属性进行试验和观测,我们称观察单位的这些属性或特征为变量(variable)。例如身高、体重、肺活量、白细胞计数、血红蛋白、性别、血型等等都是变量。变量一般可分为两类:

1. 数值变量(numerical variable) 数值变量为连续变量,如身高、体重、血压等。数值变量均可通过对观察单位测量取得数值,其值一般有度量衡单位。虽然如脉搏、白细胞计数等测量值只能是正整数,但为研究方便习惯上也视为连续变量。数值变量资料也称作计量资料。

2. 分类变量(categorical variable) 分类变量的可能取值是离散的。表现为互不

相容的类别。比如性别、血型、民族、职称等。对分类变量资料的分析一般是先按类别汇总,然后清点各类的观察单位数,列出分类变量资料的频数表。分类变量资料又称为计数资料。分类变量有两种:

(1)无序分类(unordered categories)变量:表现为变量的不同取值间没有大小、强弱、优劣等程度之别。如性别、血型等。无序分类包括:①两项分类。如性别分为男、女,虫卵化验结果分为阴性、阳性等。②多项分类。如血型分为A型、B型、AB型、O型等。

(2)有序分类(ordinal categories)变量:表现为变量的不同取值间有大小、强弱、优劣等程度之别。如疗效可分为治愈、显效、好转、无效,尿蛋白化验结果分为一、+、++、+++等。有序分类变量资料又称为等级资料。这类变量有时可通过给各类别赋值(如治愈为3、显效为2、好转为1、无效为0),作为数值变量处理。而数值变量也可转化为分类变量,如按血压测量值大小将血压分为低血压、正常、高血压;按血红蛋白量多少将血红蛋白量分为正常与异常,从而按分类变量处理。

变量类型不同与其相适应的统计分析方法也不同。进行处理时要分清是什么类型的资料,选择合适的统计分析方法进行分析。如数值变量资料一般选用 t 检验或方差分析,分类变量选用 χ^2 检验等。

二、总体与样本

总体(population)是根据研究目的确定的同质的观察单位的某个变量值的全体。例如,调查某年某地某批号菌苗接种后发热反应率,观察对象是该年该地该批号菌苗的全部被接种人员,观察单位是每个被接种者,同质基础是同一年份、同一地区、被接种同一批号菌苗,变量值是发热反应率。这类在时间和空间上都确定(只包括有限个观察单位)的总体称为有限总体(finite population)。有时总体是假想的,例如研究老年慢性气管炎用某药治疗后的疗效,这里总体的同质基础是老年、慢性气管炎患者、用该药治疗,变量值是疗效,因为没有时间和空间限制,因而包含的观察单位数是无限的,这样的总体称为无限总体(infinite population)。

对无限总体要观察其中的全部个体是不可能的,即使是有限总体,观察全部个体不仅要花费很大的人力、物力、财力,而且有时也是不必要和不可能的。所以在研究工作中,通常是从总体中随机抽取有代表性的一部分观察单位,我们称其为样本(sample),用样本信息去推断总体特征。怎样正确地抽取样本,用样本信息去推断总体特征是统计所要解决的问题。

抽样研究的目的是用样本信息推断总体特征,因此样本对它来自的总体必须有代表性,并且使抽样误差的大小可以用统计方法进行估计,为此,抽样必须遵循随机化的原则,即在抽取样本前,要使总体中的每个个体有同等的被抽取的机会。同样的,在实验中,要把观察对象分为几组,用随机化方法使每个观察对象有同等的机会被分配到各组去,以使各组间大体保持均衡。随机化的方法有抽签法或用随机数字

表,也可以用计算机产生的随机数字实现随机化。样本包含的观察单位数称为样本含量(sample size)。

三、抽样误差与统计推断

用于描述样本特征的指标称为统计量,而用于描述总体特征的指标称为参数。样本指标与总体指标间的差异在抽样研究中是不可避免的。我们把由随机抽样引起的样本指标与总体指标的差异称为抽样误差(sampling error)。由于个体变异的普遍存在,抽样误差是不可避免的。只要遵循随机化的原则,抽样误差的大小就可以用统计方法进行估计。一般情况下样本越大,抽样误差越小,反映事物客观规律的准确性越高,反之,样本越小,抽样误差越大。

进行抽样研究的时候,由于有抽样误差,如果仅仅根据样本指标表面数字的大小就下结论,往往会导致结论错误,只有正确应用统计方法,才能根据样本信息对研究对象的总体规律进行科学的推断,有效地由样本指标估计总体指标;比较可靠地辨别样本指标间的差异仅仅是由抽样误差造成的,还是因为除此之外总体间存在本质差别所致,从而作出较科学的推论。用样本信息推断总体特征叫统计推断(statistical inference)。统计推断包括总体参数估计和假设检验两部分。总体参数估计是按一定的概率估计总体指标在哪个范围,假设检验是首先建立一个关于总体特征的假设,然后根据统计量的抽样分布理论推断检验假设是否成立,并作出结论。

四、概 率

生活或工作中经常遇到这样的情况,对某一事件进行一次观察将出现什么结果事前不能确切知道,这种不确定的结果叫随机事件。概率(probability)是描述随机事件发生的可能性大小的数值。例如用某种中药治疗胃肠道感染,只知道治疗结果可能是治愈、显效、好转、无效,但对具体的某一个患者治疗后将发生什么结果是不能确切知道的。每一种可能的结果都是一个随机事件。比如治愈这个事件发生的可能性大小是患者和医生都非常关心的未知数值,怎样确定这个数值,通常是通过治疗一定数量的患者,随着样本例数的增大,治愈者出现的频率逐渐稳定在一个数值附近,这个稳定性揭示了治愈这个随机现象的规律性,这个数值就可以看做是治愈这个事件发生的概率。当概率不易求得时,只要观察单位充分多,就可以将频率作为概率的估计值。

事件 A 的概率记为 $P(A)$,随机事件的概率 P 取值在 $0\sim 1$ 之间, P 越接近 1,说明某事件发生的可能性越大; P 越接近 0,说明某事件发生的可能性越小。如果某事件的概率 $P=0$,表示该事件不可能发生,称其为不可能事件;如果 $P=1$,表示该事件必然发生,称其为必然事件。如果某随机事件发生的概率 $P\leq 0.05$,或 $P\leq 0.01$ 表示该事件发生的可能性很小,我们称其为小概率事件。

第四节 学习卫生统计学应注意的问题

学习卫生统计学最重要的是掌握统计学的基本理论知识、基本思维方法和基本实践技能。在由样本信息推断总体特征的时候应该考虑抽样误差的问题,进行样本指标的比较,一般不能仅凭表面数字上的差异就下结论。必须从概率角度出发经过统计推断,判断该差异是仅仅产生于抽样误差,还是因为总体间存在本质差异。

正确运用统计方法可以帮助我们科学地认识客观事物,阐明事物固有的规律,但统计决不能创造规律,统计工作最根本的一条是实事求是,必须重视原始资料的完整性、准确性,以严肃认真的科学态度对待数据处理,绝对不许伪造和篡改统计数字。必须注重应用统计方法解决实际问题,深刻理解各种统计方法的意义及条件,注重结果的解释,对统计公式可不必追究其数学推导。

学习卫生统计学,要掌握群体健康的评价方法,学会运用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,为科学决策服务。

只有理论联系实际,结合专业,分析评价预防工作、医学科研、卫生事业管理以及医学文献中的实际问题才能学好卫生统计学。

(马 燕)

第二章 数值变量资料的统计描述

第一节 数值变量资料的频数表

频数就是观察数据的个数。频数分布就是观察数据在其取值范围内分布的情况。数值变量资料的频数分布情况可以用频数表(frequency table)或直方图(histogram)表示。

一、频数表的编制

例 2.1 试根据某市某年 120 名 8 岁男孩的身高(cm)资料编制频数表

124.6	126.2	128.1	130.6	132.6	125.4	126.6	128.4	124.5	129.5
124.8	127.8	128.3	131.8	125.8	126.8	129.6	125.6	127.6	129.8
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	121.1	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

(1)找出观察值中的最大值、最小值和极差。本例最大值为 134.5cm,最小值为 112.3cm。最大值与最小值的差,称为极差,本例极差为 $134.5\text{cm} - 112.3\text{cm} = 22.2\text{cm}$ 。

(2)确定组数、组距和组段。组数通常根据观察例数的多少而定,观察例数多,组数可相对多些。一般以 8~15 组为宜。不足 100 例,组数可取 8~10 组,超过 120 例可取 12~15 组。组距(class interval)的选取以方便为准,可根据极差除以组数之商略加调整,本例 120 例,可分 12 组, $22.2/12=1.85$,为计算方便可取组距为 2cm。每个组段的起点称“下限”(low limit),终点称“上限”(upper limit),各组段必须首尾相接,从该组段的下限开始,不包括该组段的“上限”。例如身高在 112cm 至不满 114cm 的观察值应属于“112~”组段,满 114cm 则属于“114~”组段。最末组段应同