

全国中等卫生学校教材

卫生统计学

(供卫生医士专业用)

周士楷 主编

杨树勤 主审

人民卫生出版社

87
R195.1
15
3

全国中等卫生学校教材

卫生统计学

(供卫生医士专业用)

15731/28

周士楷 主编

严日树

杨天忠 编写

周士楷

杨树勤 主审

人民卫生出版社

北京...
地址...
电话...

人民卫生出版社

3 415072



78
07

卫生统计学

(供医学专业用)

主编 周士楷

副主编 李月波

参编 李月波

李月波

李月波

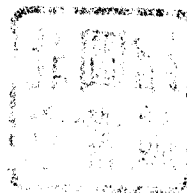
卫生统计学

周士楷 主编

人民卫生出版社出版
(北京市崇文区天坛西里 10 号)

人民卫生出版社印刷厂印刷
新华书店北京发行所发行

787×1092毫米16开本 18十印张 427千字
1987年5月第1版 1987年5月第1版第1次印刷
印数：00 001—20 100
ISBN 7-117-00217-4/R·218 定价：2.50元
统一书号：14048·5412



编写说明

本书是根据卫生部一九八二年颁发的中等卫生学校教学计划、一九八三年全国中等卫生学校教材编审工作会议提出的要求和新编《卫生统计学》教学大纲编写而成的，供中等卫校卫生医士专业用。

全书共十六章。第一章绪论；第二章至第十五章讲述卫生统计基本理论与方法，其中先讲统计描述和统计推断，并专章讨论半数致死量和统计表与统计图，在这个基础上，后讲调查和实验设计；第十六章讲述应用上述统计方法研究居民生老病死的健康统计。各章正文后附有复习思考题，供学生课后复习参考。书中有“*”号的内容供选学用。书末实习作业指导供各校安排课堂实习和课后练习选题参考。书中例题、练习题的数据资料皆引自有关统计书籍和杂志，仅供统计演算示范用，不作有关专业内容的解释依据。书中统计符号、缩写一律按照《中国医学百科全书·医学统计学分卷》的规定。

在本书编写过程中，黑龙江、河北、河南、陕西、山东、安徽、江苏、浙江、江西、湖北、湖南、广东、广西、四川、贵州、上海、武汉等省市有关卫生学校、卫生职工学院和医学院的教师们，对本书的编写工作给予了热情的支持和帮助，对教材初稿提出了许多宝贵意见。在此，谨向他们表示感谢。

在本书编写过程中，我们还参考了《中国医学百科全书·医学统计学分卷》，现行有关高校、中等卫校教材，在此谨向有关作者表示衷心的感谢。

编者

1986年4月

目 录

第一章 绪论	1
第一节 概述.....	1
第二节 卫生统计工作步骤.....	2
第三节 统计资料的类型.....	3
第四节 几个基本概念.....	4
第二章 计量资料频数分布的特征与规律	7
第一节 计量资料的频数分布.....	7
第二节 集中趋势指标.....	10
第三节 离散程度指标.....	18
第四节 正态分布.....	22
第三章 总体均数的估计和假设检验	33
第一节 均数的抽样误差.....	33
第二节 t 分布.....	34
第三节 总体均数可信区间的估计.....	35
第四节 假设检验.....	37
第五节 进行假设检验时应注意的问题.....	49
第四章 方差分析	52
第一节 完全随机设计资料的方差分析.....	52
第二节 随机区组(配伍组)设计资料的方差分析.....	58
第三节 多个均数间的两两比较.....	61
*第四节 方差分析的基本条件与数据变换.....	65
第五章 相对数	70
第一节 相对数的概念.....	70
第二节 常用相对数.....	70
第三节 应用相对数时应注意的问题.....	76
第四节 标准化法.....	80
第六章 二项分布及其应用	83
第一节 二项分布的概念及其应用条件.....	83
第二节 二项分布的图形.....	84
第三节 率的抽样误差与标准误.....	85
第四节 二项分布的应用.....	87
*第七章 泊松分布及其应用	91
第一节 泊松分布的概念.....	91
第二节 泊松分布的性质.....	92
第三节 泊松分布的应用.....	94

第八章 χ^2(卡方)检验	103
第一节 概述.....	103
第二节 样本率或构成比的比较的假设检验.....	104
第三节 计数资料相关分析.....	111
第四节 拟合优度检验.....	113
*第五节 四格表资料的确切概率法.....	115
第九章 非参数统计	119
第一节 非参数统计的概念.....	119
第二节 秩和检验.....	119
第三节 中位数的比较.....	130
第十章 直线相关与回归	133
第一节 直线相关.....	133
第二节 等级相关.....	139
第三节 作相关分析时应注意的问题.....	141
第四节 直线回归.....	142
第五节 作回归分析时应注意的问题.....	151
*第十一章 多元线性回归简介	153
第一节 多元线性回归的概念.....	153
第二节 二元线性回归.....	153
第三节 三元线性回归.....	157
第十二章 半数致死量	168
第一节 半数致死量的概念.....	168
第二节 目测概率单位法.....	169
第三节 寇氏法.....	172
第四节 点斜法.....	174
第五节 半数致死量实验设计要求.....	175
第十三章 统计表与统计图	178
第一节 统计表.....	178
第二节 统计图.....	182
第十四章 调查设计	195
第一节 搜集资料的计划.....	195
第二节 整理资料的计划.....	200
第三节 分析资料的计划.....	204
第十五章 实验设计	206
第一节 实验设计的原则.....	206
第二节 实验设计的方法.....	208
第三节 样本含量的估计.....	211
第十六章 居民健康统计	213
第一节 计划生育统计.....	213

第二节 人口死亡统计	216
第三节 疾病统计	219
卫生统计学实习作业指导	223
附表	249
附表 1 标准正态曲线下的面积表	249
附表 2 t 界值表	251
附表 3 F 界值表 (方差齐性检验用)	252
附表 4 F 界值表 (方差分析用)	253
附表 5 q 界值表	259
附表 6 百分率 p 与概率单位对照表	260
附表 7.1 百分率的可信限 ($1 \leq n \leq 50$)	262
附表 7.2 百分率的可信限 ($50 \leq n \leq 100$)	265
附表 8 泊松分布均数的可信限	266
附表 9 χ^2 界值表	267
附表 10 阶乘的对数表	268
附表 11 相关系数 r 界值表	269
附表 12 相关系数 r 与 z 值对照表	271
附表 13 等级相关系数 r_s 界值表	271
附表 14 配对比较的秩和检验 T 界值表	272
附表 15 两组比较的秩和检验 T 界值表	273
附表 16 三组比较的秩和检验 H 界值表	275
附表 17 M 界值表	275
附表 18.1 两样本率比较时所需样本含量 (单侧)	276
附表 18.2 两样本率比较时所需样本含量 (双侧)	277
附表 19 配对比较 (t 检验) 时所需样本含量	278
附表 20 两样本均数比较 (t 检验) 时所需样本含量	279
附表 21 随机数字表	280
附表 22.1 10×10 随机化拉丁方	281
附表 22.2 10×10 随机化拉丁方	281
附表 23 30 个自然数的随机排列表	282
《卫生统计学》教学大纲	283

第一章 绪 论

第一节 概 述

统计学是研究数据的搜集、整理、分析与推断的科学。卫生统计学是把统计理论与方法，应用于居民健康状况、医疗卫生实践和医学科学研究的一门学科。现代医学的研究，正由定性描述向定量研究方向发展。由于错综复杂的条件和不可控制的偶然因素的影响，医学研究所获得的数据，一般变异较大，不能从中直观地了解事物本来的面貌和其中的必然规律。卫生统计学的任务是借助于统计方法，从有限的观察中，从表现为偶然的数据中，把所研究的事物或现象的本质特征、整体情况和相互关系一一揭示出来。

卫生统计学的内容较广，主要包括三个部分：

1. 卫生统计学的基本理论和方法 讨论统计设计、统计描述、统计推断和统计联系研究的原理和方法。研究统计方法在医学科学中的应用。

2. 健康统计 包括出生、死亡、疾病、身体发育统计等。其任务是搜集有关资料，建立和应用有关指标，反映和评价居民健康状况，掌握疾病的流行规律，为疾病防治提供依据。

3. 卫生事业统计 搜集、整理和分析卫生事业发展的情报(如卫生机构、人员、设备情况)，为研究或规划卫生事业发展提供依据。

本教材将以较大部分篇幅介绍卫生统计的基本理论和方法，同时重点地介绍健康统计的部分内容。

卫生统计学的应用十分广泛。在卫生防疫工作中，为了了解居民健康状况和评价医疗卫生措施的效果；为了判断社会、环境和生物因素对居民健康的影响；为了改善卫生条件，提高居民健康水平，经常需要作流行病学、卫生学调查和实验研究。卫生统计学是完成这些任务的重要手段，它在调查与实验设计、数据处理、统计推断以及统计分析结果的表达方面都有重要的作用。因此做一名合格的卫生医士，必须学好卫生统计学。

随着数学和医学的发展，卫生统计学的应用范围也不断扩大。大量的卫生统计工作实践，积累了丰富的经验，也进一步充实了卫生统计学的内容，促进了卫生统计学本身的发展。由于电子计算机的普遍应用，资料搜集、贮存、整理和分析工作已经和正在实现自动化，人们正在摆脱繁琐的手工计算劳动，这必将把卫生统计工作提高到一个新的水平。卫生统计学的进一步的应用和发展，面临着一个广阔的前景。

卫生医士专业学生学习卫生统计学的目的是：掌握卫生统计学的基本理论、基本知识和基本技能。为学习各门专业课，阅读专业书刊，从事卫生防疫工作打下必要的统计学基础。为此要求：

1. 明确学习目的 充分认识卫生统计学在医疗卫生工作中的重要作用，发挥学习主动性和积极性。

2. 注意学习方法 要弄清统计方法的基本思想及有关公式的应用条件和用法，但不深究其数学原理；要理论联系实际，统计设计和分析都要结合专业知识，避免作专业上

无意义甚至错误的统计结论；要重视基本理论与知识的学习，把各章节的内容作为一个整体，前后呼应，纵横联系，融会贯通地学习。

3. 培养科学作风 通过学习，培养对数据处理的严肃认真、实事求是的科学作风，重视原始资料的搜集与整理的每一个细节，保证资料真实可靠。

4. 培养科学的统计思维方法 养成定量分析问题的习惯，不凭数据表面的大小作判断。重视统计设计，重视原始资料的完整性与可靠性；严格按照资料特点和统计方法的应用条件，选择适当的统计方法，通过严密推理，作出合理判断。

第二节 卫生统计工作步骤

统计全过程设计、搜集资料、整理资料和分析资料是卫生统计工作的四个基本步骤。

一、统计全过程设计

根据研究目的，从统计学角度，对搜集资料、整理资料和分析资料提出周密的计划和要求，作为统计全过程实施的依据，以便能用尽可能少的人力、物力和时间，获得准确可靠的结论。

对于统计全过程设计的要求是科学、周密、简明。关于这方面的内容将在第十四章和第十五章详细介绍。

二、搜集资料

根据统计全过程设计所提出的要求，实施搜集资料工作。

资料的来源，可分为经常性资料与一时性资料两种。经常性资料包括从日常医疗卫生工作原始记录（如病历），专门报告卡（如出生、死亡报告卡），统计报表（如疫情月报表、年报表）中搜集到的资料。一时性的资料常指由现场调查和实验室研究收集的资料。现场调查一般涉及范围较广，观察数量较多，影响因素较复杂；实验研究一般涉及范围较小，观察数不多，因而便于对实验条件作严格控制，结论也比较确切。现场调查与实验研究既有区别又有联系，它们常常结合应用。现场调查可提供进一步实验研究的线索，而实验研究的成果也需要再回到现场中去实践验证。

对于搜集资料的要求是完整、准确、及时。这是统计分析准确可靠的基础。完整是指搜集资料项目不遗漏；准确是指观察、测量准确，记录、计算无误；及时是指经常性资料的搜集，应按规定的时间内完成，一时性资料搜集者对数据的记录应在观察、测量的同时完成，不得以“回忆”方式记录数据。

三、整理资料

整理资料就是把搜集到的资料，有目的、有计划地进行科学加工，使分散的、零乱的资料变成系统化、条理化的资料，以便进一步统计分析。

四、分析资料

分析资料就是运用各种统计分析方法，结合专业知识，从经过整理的资料中，计算出各种指标，进行统计描述与统计推断，阐明事物的规律。

做好分析资料的条件有二：其一，对于各种统计分析方法能够融会贯通地理解，能够正确地选择、综合地运用各种统计分析方法；其二，对于所研究的事物本身及其与周围事物的联系具有丰富的专业知识，因而能作出合理的判断。

第三节 统计资料的类型

统计资料一般分为计数资料、计量资料和等级资料三种类型。

一、计数资料

先将观察单位按性质或类别进行分组，然后清点各组观察单位的个数，这样所得的资料，称为计数资料。计数资料没有计量单位。例如，对某小学全体学生进行粪便蛔虫卵检查，观察单位为每一名学生，因观察单位的性质不同，可分为阳性和阴性两组，然后分别计数各组观察单位个数即阳性人数和阴性人数；又如调查某人群血型分布的资料也是计数资料。观察单位为人群中每一个人，按观察单位性质不同（血型类别不同），可分为A、B、O、AB型各组，然后分别计数各组的观察单位个数；再如检查某人白细胞分类的资料，也是计数资料。观察单位是每一个白细胞，按白细胞类型不同分为中性、嗜酸性、嗜硷性、淋巴细胞、大单核细胞等各组，然后分别计数各组观察单位个数。分析计数资料常用第五、六、七、八等章所介绍的统计分析方法。

二、计量资料

用定量方法测定每一观察单位某项指标数值的大小，所得的资料称为计量资料。计量资料一般具有计量单位。同一总体内各观察单位的测量值常有差异。例如，测量身高（cm）、体重（kg）、浓度（mg/L）、血压（mmHg）等等，所获得的资料就是计量资料。分析计量资料常用第二、三、四、十等章所介绍的统计分析方法。

三、等级资料

有些计数资料带有计量资料的性质，统计上常称它们为等级资料或半定量资料。它是先将观察单位按某项指标的等级顺序分组，再清点各组观察单位的个数所得的资料。如检查尿蛋白，观察单位是每一份尿标本，按结果等级不同可分为-、+、++、+++、++++等各组（这里尿蛋白量一组比一组多），然后清点各组观察单位个数。又如观察某病治疗效果，观察单位是被观察的每一名患者。按疗效等级可分为痊愈、显效、无效、恶化、死亡等组（这里疗效一组比一组差），然后清点各组的观察单位个数。这样的资料带有“计量”的性质，是因为各观察单位具有含量或程度方面的等级差别。分析等级资料常用第九章所介绍的统计分析方法。

以上所讨论的资料类型的区分，并不是绝对的。有时根据研究目的，可以把计量资料变换为计数资料或等级资料。如白细胞总数属于计量资料，但可以按白细胞总数正常（4000~10000个/mm³）与白细胞总数不正常（多于10000个/mm³或少于4000个/mm³）分为两组，再清点各组人数；这就成为计数资料了。又如脉搏数也是计量资料，可以把脉搏数小于60次/分称为缓脉，60~100次/分称为正常脉，大于100次/分称为速脉，然后清点各组人数，即成为等级资料。

第四节 几个基本概念

一、变异与变量

世界中的事物、现象，普遍地存在变异性。在医学研究中，存在着两种变异，一种称为个体变异，即观察单位本身的变异，另一种称为随机测量变异。

个体变异表现为即使条件相同的个体（观察单位）间，各项特征仍存在差异。例如，同一年龄的男孩，身高、体重各不相同；同一年龄的成年男子，血清中胆固醇含量各不相同；同一种病的患者，即使病情一致，治疗方法和条件相同，疗效也未必相同等等。个体变异的原因在于个体与外界环境关系极为复杂，影响个体特征的内外因素很多，不同因素对于个体的影响偶然地结合在一起，使个体间表现出差异。

随机测量变异表现为同一个体（观察单位）多次观测结果不完全相同。例如，同一样品，用分析天平多次称量，各次所得量值不完全相同。随机测量变异是由于观测手段或条件在种种偶然因素影响下，发生轻微波动所引起的。

对某一变异的事物进行观测，所取得的某项指标的数值各不相同。我们把这种各不相同的数值称为变量值。

二、总体与样本

总体是根据研究目的确定的性质相同的所有观察单位某种变量值的集合。例如，欲研究某地40岁以上正常男子血清胆固醇含量，则该地所有40岁以上的正常男子的血清胆固醇含量的集合就是总体。在这里，性质相同的观察单位指的是该地的（不包括外地的）、所有的（不是部分的）、40岁以上的（不包括40岁以下的）、正常的（不包括有病的，这里的“病”是指某些影响胆固醇含量的疾病）男子（不包括女子）。总体的概念不仅限于个体的某种变量值的集合。如在同样条件下，对同一份样品多次重复测量，如果这一测量可以无限地进行下去，则所有可能的测量结果的集合，也是一个总体。

了解总体情况是统计工作的目的之一。但是，多数情况下，总体很大，我们不可能对所有观察单位逐个观察，有时由于观测带有破坏性，如检查一批鸡蛋中沙门氏菌污染情况，即使可能逐一检查，也不允许这样做。因此，我们总是通过样本情况去推断总体情况。

样本是从总体内随机抽取的一部分。例如，我们从某地抽取100名40岁以上的正常男子，测定其血清中胆固醇含量，这100个胆固醇值就组成了一个样本。为了使样本对总体具有充分的代表性，必须遵循随机抽样原则，即要使总体内每一个体都具有同等被抽取到的机会。关于随机抽样方法见第十四章。

三、概率与频率

概率是事件发生的可能性大小的度量，以符号 P 表示。可以把我们经常遇到的事件分为三种类型：①必然事件：指的是必然会发生的事件。如在一大气压下，纯水加热到 100°C ，必然会发生沸腾现象。必然事件的概率 $P = 1$ 。②不可能事件：指的是不可能发生的事件。如人的生长不老（不死）事件，必定不会发生。不可能事件的概率 $P = 0$ 。

③随机事件：指的是在一定条件下，可能发生，也可能不发生的事件。如某人在肝炎流行时，是否会得肝炎，回答是不能肯定的，可能得，也可能不得。随机事件的概率 P 在0与1之间。某事件发生的可能性愈大，则其概率 P 值愈接近1；某事件发生的可能性愈小，则其概率 P 值愈接近于0。

频率也是某事件出现的可能性大小的度量。只不过概率是对总体而言，频率是对样本而言。在相同的条件下，进行 n 次重复试验，事件 A 发生数为 a ($a \leq n$)。则 a 与 n 的比 a/n 为事件 A 的频率。如 n 逐渐增大则事件 A 的频率就越来越接近概率 P 。统计上，常以 n 充分大时事件 A 的频率作为该事件概率的近似值。

四、误差

统计上所说的误差，包括测得值（观察值）与真值之差和样本指标与总体指标之差。从误差的性质来看，可以把误差分为两大类，即偶然误差和系统误差。

（一）偶然误差

偶然误差包括抽样误差和随机测量误差。

1. 抽样误差 指样本指标（如样本均数、样本率）与总体指标（如总体均数、总体率）之差。如某地100名40岁以上正常男子血清胆固醇量的均数，一般不会恰好等于某地全体40岁以上正常男子血清胆固醇量的均数，两者之差就是抽样误差。产生抽样误差的原因，是总体内各观察单位存在着个体变异。样本内各个体变异情况和总体内各个体变异情况不会恰好相同。因此，由样本求得的指标值就不会恰好等于总体指标值。由于总体内个体变异是必定存在的，因此，抽样误差是不可避免的。但是，增加样本含量可以缩小抽样误差。

2. 随机测量误差 指的是同一个体（观察单位）多次观测结果之差。产生随机测量误差的原因是观测中存在着的随机测量变异。由于观测中随机测量变异必定存在，因此，随机测量误差也是不可避免的。但是，改善测量手段和测量条件，可以将随机测量误差控制在很小的范围内。

偶然误差的性质：抽样误差与随机测量误差具有共同的性质，即其误差值一般较小；方向是双向的，可正可负；重复观测时其误差的大小和方向不一定重现。偶然误差是由很多影响较小而又难以完全消除的因素综合影响的结果，它的出现是不可避免的，它表面上捉摸不定，实际上却有严格的规律性（如服从正态分布）。

应该指出，偶然误差是有上述确定涵义的统计学概念，不要把它与偶然的误差（偶然发生的误差）相混淆。后者仅仅是指误差发生的方式是偶然的而已。实际上，有些偶然发生的误差却具有系统误差的性质。例如，由于操作的偶然失误，电压的突然变化，仪器的突然故障等所造成的误差，虽然是偶然发生的，但却可能具有系统误差的性质。

（二）系统误差

系统误差是由确定的原因引起的测得值与真值的较大偏差。产生系统误差的原因很多，其中最常见的原因是试验条件不同。由试验条件不同而引起的系统误差，又称为条件误差。指的是因试验仪器、试剂、操作方法、诊断方法、治疗方法等等条件不同，而造成的结果偏大或偏小。例如，天平砝码未校正，其真实重量（严格地说应为质量）比其所示值偏小时，测得的物质重量将比物质的真实重量为大。

系统误差的性质：虽然具体的系统误差各种各样，但它们都具有共同的性质。即其误差值与偶然误差相比，一般较大；方向是单一的，或者偏大，或者偏小；在条件不变的情况下观测，同样大小和方向的误差将一再重复出现；具有某个（或几个）明确的原因（虽然这样的原因有时可能尚未被发现）；当引起该系统误差的原因消除以后，该系统误差就不再出现；在试验中，系统误差会不会出现，出现的大小和方向等并无统计规律性。例如，从上面说的天平称量的例子中可以看到，因砝码不准造成的误差是较大的，方向是单一的，即夸大了物质的重量。只要砝码未校正，这样的误差会一再出现。砝码一经校正，这样的误差即消失，其原因是明确的，但当不了解砝码真实的情况时，我们将无法预言它是否会出现以及如果出现的话，它的大小和方向如何。

偶然误差与系统误差性质的比较见表 1·1。

表 1·1 系统误差与偶然误差的性质比较

误差类型	大小	方向	大小和方向的重现性	产生的原因	可否消除	统计规律性
偶然误差	一般较小	双向	不一定重现	是各种偶然因素综合影响的结果	不可避免但可控制	有
系统误差	一般较大	单向	可重现	有少数确定的原因	消除原因即可避免	无

复习思考题

1. 卫生统计学的任务是什么？卫生医士专业学生为什么要学习卫生统计学？
2. 卫生统计工作的步骤，各步骤的意义与要求如何？
3. 统计资料的类型分哪几种？各举实例说明。
4. 观察值变异是由什么原因引起的？试举例说明。
5. 举例说明抽样研究的目的。
6. 系统误差与偶然误差的性质如何？各举例说明。
7. 为什么说偶然发生的误差不一定是偶然误差？试举例说明。
8. 研究误差性质与统计学有何关系？

(周士楷)

第二章 计量资料频数分布的特征与规律

第一节 计量资料的频数分布

所谓频数就是观察值的个数。频数分布就是观察值在其所取值的范围内，于各组段中分布的情况。频数分布情况可用频数分布表、频数分布图来表示。编制频数分布表，绘制频数分布图，是整理资料的基本步骤之一。了解频数分布情况是研究资料的第一步。

一、频数分布表的编制

例 2·1 某市 150 名 3 岁女孩身高 (cm) 资料如下，试编制频数分布表。

80.1	100.1	97.0	96.7	97.9	100.7	86.2	91.7	94.7	90.8
82.5	102.6	99.1	96.6	99.3	85.2	89.2	90.6	95.1	93.6
84.4	104.8	101.3	98.7	101.5	87.1	89.0	92.7	96.8	92.7
87.2	83.5	103.2	101.6	84.4	88.4	91.8	93.6	99.2	94.4
89.3	84.2	82.3	84.5	87.9	89.4	91.9	94.5	86.9	95.6
89.1	86.5	85.0	87.6	89.3	90.4	92.1	95.0	89.3	96.3
91.3	89.7	87.4	89.8	88.7	90.2	92.9	97.2	91.4	90.3
90.5	88.9	88.1	88.2	91.1	93.0	95.6	98.7	90.0	93.5
92.4	90.0	88.0	90.7	91.7	93.8	94.4	87.3	93.9	92.8
92.6	90.0	90.8	90.1	93.2	94.4	97.3	89.0	92.9	91.3
94.7	92.8	90.3	92.8	93.6	94.8	98.3	88.5	94.0	96.0
94.8	92.3	93.3	93.1	95.1	97.0	84.5	91.1	94.3	93.4
97.1	95.8	93.7	95.1	94.9	99.4	86.4	91.7	96.5	92.5
96.2	94.3	94.2	94.6	96.4	100.9	89.1	93.2	98.4	95.5
99.5	97.5	95.1	96.2	99.5	85.7	88.4	92.5	91.1	97.3

编制频数分布表步骤如下：

(一) 计算全距

找出观察值中的最大值与最小值，求它们的差，即全距。本例最大值为 104.8，最小值为 80.1，全距 = $104.8 - 80.1 = 24.7(\text{cm})$ 。

(二) 确定组段数、组距和组段

根据全距的大小、观察值个数，决定组段数。全距大，观察值个数多，则组段数可适当增加。一般以取 8~15 个组段为宜。组段数过多，编制过程和计算较繁，组段数过少，计算误差较大。

根据组段数和全距，决定组距。组距 = $\frac{\text{全距}}{\text{组段数}}$ 。本例全距为 24.7cm，组段数暂定 12 个，组距为 $\frac{24.7}{12} = 2.06$ ，为计算方便，取组距为 2cm。

合理划分组段。第一组段要包括最小的观察值，组段的下限要略低于或等于最小观察值。最后一组段要包括最大观察值，其上限应略大于最大观察值。

(三) 列表归组

列出频数分布表（表 2·1）。表中第(1)栏为组段，用划记法或分卡法将各观察值分别归入各组段，得第(2)栏。清点各组段内的观察值个数即得各组段频数，将各组段频数填入第(3)栏。合计各组段频数为总频数。

表 2·1 某市 150 名 3 岁女孩身高的频数分布

组 段 (cm) (1)	划 记 (2)	频数 f (3)
80~	—	1
82~	下	3
84~	正下	8
86~	正正	10
88~	正正正正	19
90~	正正正正下	23
92~	正正正正正—	26
94~	正正正正正	24
96~	正正正正	17
98~	正正	10
100~	正—	6
102~	丁	2
104~106	—	1
合 计		150

二、频数分布图

为了更加直观地了解频数分布情况，通常在编制频数表的基础上，绘制频数分布图。常见的频数分布图有两种，即直方图和多边图。

(一) 直方图

以横轴表示观察值，纵轴表示频数。将横轴等分为若干以组距为单位的小区间，注明各组段的下限值，在各小区间上作高度等于频数的长方形，即得直方图。直方图的纵轴也可以表示频率。频率(%) = $\frac{\text{各组段频数}}{\text{总频数}} \times 100\%$ 。

图 2·1 是根据表 2·1 资料绘制的直方图。

(二) 多边图

纵、横轴同直方图，标出各组段中点与相应的频数（或频率）之交点，然后将各交点连接起来，就是频数分布多边图。

图 2·2 是根据表 2·1 资料绘制的多边图。

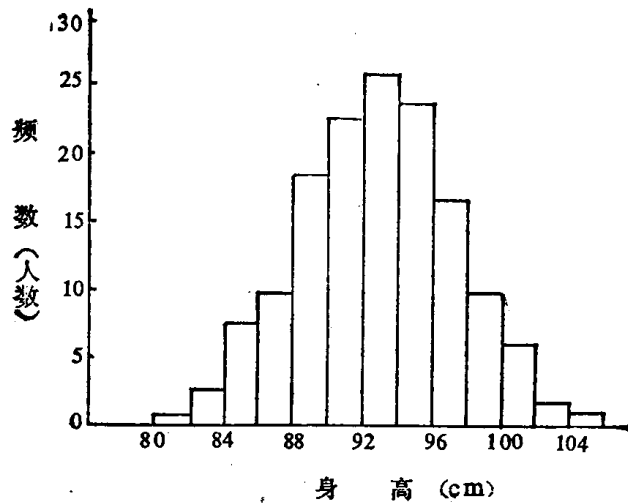


图 2·1 某市150名 3 岁女孩身高频数分布直方图

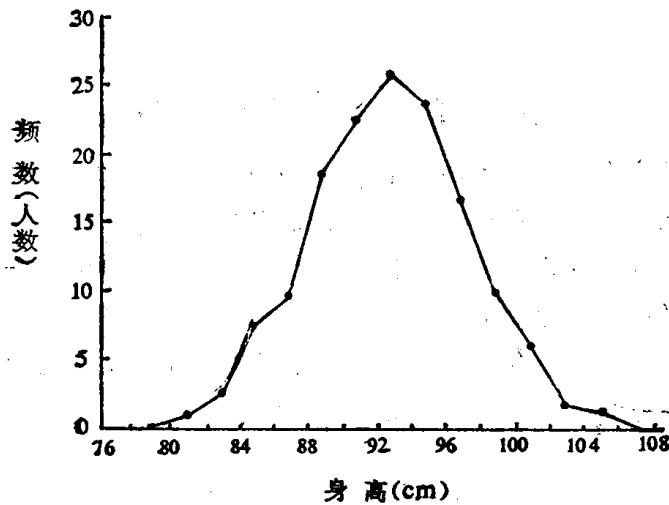


图 2·2 某市150名 3 岁女孩身高频数分布多边形

三、频数分布类型

从频数分布的图形来看，常见频数分布有三种类型：

(一) 对称分布型

如图 2·3(1)，整个图形以高峰所在处的垂线为中心，左右两侧逐渐下降并对称。在这类分布中，以正态分布为最典型（后详）。上面例 2·1 资料即属此类型分布。

(二) 正偏态分布型

如图 2·3(2)，整个图形不对称，高峰偏左，即观察值较小的这一端，集中了较多的频数。属于这类分布的资料并不少见。如正常人体中某些非必需元素含量的频数分布；一些传染病潜伏期的频数分布；实验动物对一些作用物质产生某种特征反应的剂量频数分布；粉尘粒子大小的频数分布等。

(三) 负偏态分布型

如图 2·3(3)，整个图形不对称，高峰偏右，即观察值较大的这一端，集中了较多的频数。属于这类分布的资料较为少见。如某班学生成绩的频数分布，多数学生得分较高，

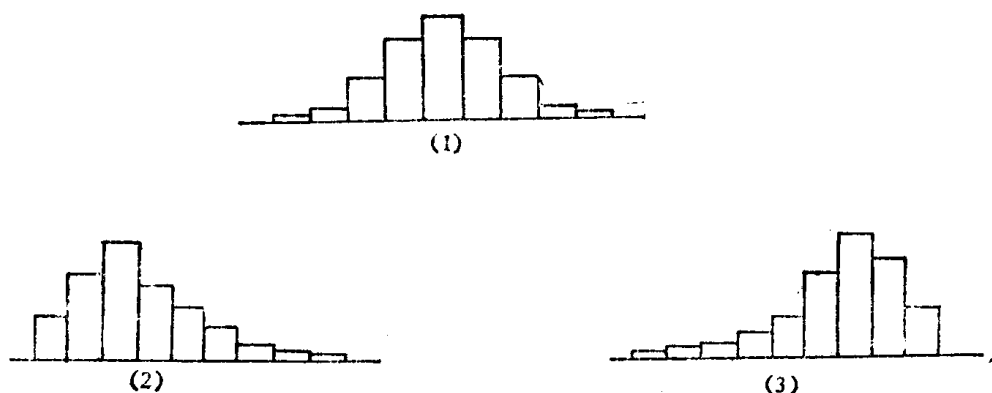


图 2·3 几种不同类型的频数分布示意图
(1)正态分布型 (2)正偏态分布型 (3)负偏态分布型

少数学生得分较低。又如儿童少年远视力、近视力频数分布，即多数学生视力较好，少数学生视力较差。

第二节 集中趋势指标

从频数表和频数分布图中，可以大致看出频数分布的规律。也可以大致了解频数分布的特征——集中趋势和离散程度，但这种认识是粗略的。如欲准确掌握频数分布的特征，就应作频数分布特征的定量描述，即计算集中趋势指标和离散程度指标。

集中趋势指标又称平均数。它反映了观察值的集中位置或平均水平。也可以说，是观察值的典型水平或代表值。集中趋势指标的使用场合不同，计算方法也不一样。常用的集中趋势指标有算术均数（均数）、几何均数和中位数等。

一、算术均数(均数)

算术均数又称均数，是各观察值相加后，除以观察值个数所得之商，以符号 \bar{X} 表示。

(一) 应用条件

均数最适用于对称分布，尤其是正态分布资料。因为这时均数位于中央，能反映观察值的集中趋势。当观察值个数较少，而其频数分布基本对称或从专业上可推断其总体为正态、近似正态者，也可用均数作为其集中趋势指标。

(二) 计算方法

1. 直接法 按式 (2·1) 计算。

$$\bar{X} = \frac{\Sigma X}{n} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} \quad (2\cdot1)$$

式中 \bar{X} ——均数，读作 [eksba:]，

$X_1, X_2, X_3, \dots, X_n$ ——各观察值，

Σ ——求和之符号，读作 [sigmø]，

n ——观察值个数。

例 2·2 有 10 份面粉样品，测得蛋白质含量分别为：10.0, 9.5, 9.8, 8.9, 9.2, 10.1, 9.5, 9.7, 9.2, 9.3g/100g。求 10 份面粉样品的平均蛋白质含量。