教育部人才培养模式改革和开放教育试点教材

高等医学院校卫生事业管理专业教材

Shiyong Weisheng Tongji Xue

实用卫生统计学

主编：康晓平

北京大学医学出版社

•教育部人才培养模式改革和开放教育试点教材
•高等医学院校卫生事业管理专业教材

实用卫生统计学

主　编　康晓平
主　审　陈育德

编写人员：（按姓氏笔画排序）

李　凯　何平平　易伟宁
赵新胜　康晓平

北京大学医学出版社

SHIYONG WEISHENG TONGJIXUE

图书在版编目（CIP）数据

实用卫生统计学/康晓平主编 .—北京：北京医科大学出版社，2002（2007.11 重印）
ISBN 978-7-81071-268-2

Ⅰ.实… Ⅱ.康… Ⅲ.卫生学：统计学－医学院校－教材 Ⅳ.R195.1

中国版本图书馆 CIP 数据核字（2002）第 015442 号

实用卫生统计学

主　　编：康晓平
出版发行：北京大学医学出版社（电话：010-82802230）
地　　址：（100083）北京市海淀区学院路 38 号　北京大学医学部院内
网　　址：http：//www.pumpress.com.cn
E - mail：**booksale@bjmu.edu.cn**
印　　刷：莱芜市圣龙印务书刊有限责任公司
经　　销：新华书店
责任编辑：暴海燕　　**责任校对**：齐　昕　　**责任印制**：郭桂兰
开　　本：787mm × 1092mm　1/16　**印张**：14　**字数**：356 千字
版　　次：2002 年 4 月第 1 版　2007 年 11 月第 4 次印刷　**印数**：18001 - 21000 册
标准书号：ISBN 978-7-81071-268-2
定　　价：18.80 元

编 写 说 明

《实用卫生统计学》是为中央电大卫生事业管理专业学生而编写的教材。根据电大学生的特点，本教材的编写注意到以下几个方面：

1. 强调理论与实用结合。全书共分两篇十四章。第一篇为卫生统计学基本概念和基本方法，设十章，重点介绍卫生统计学的基本理论、基本知识和基本技能。第二篇为卫生服务与居民健康统计，设四章，主要介绍卫生服务与居民健康统计资料的来源、常用统计指标及其应用。两篇内容互相联系，互为补充。例如，第一章讲统计工作的四个步骤，第十二章则详细介绍卫生服务的调查设计并用实例具体阐述数据收集和处理的过程。在前十章，尽量多采用实例讲解概念和计算，避免不易理解的数学语言。除此以外，在每章最后一节采用综合例子概括本章节的重点内容，并且还用章末小结和思考题的形式，帮助学生进行归纳、理解各章节的重点和难点。书中打＊号的章节是学生在掌握基本概念和基本方法的前提下，作为自学参考的内容，不作重点要求。

2. 本书在部分内容的编写结构上与以往的《卫生统计学》教材略有不同，即将具有共性的内容归纳到一起，组成第五章“单个样本数据的统计推断”。例如，单个样本的抽样误差，用均数的标准误和率的标准误来估计，其概念完全一样，不同点只是资料性质不同，计算公式不同而已。现把它们集中到一章，既加深了对抽样误差概念的理解，又在对比学习中分别记住了计量资料和计数资料的标准误计算。这种编排便于读者在自学过程中将复杂的内容，对比整理出头绪。

3. 学习卫生统计学的最好方法是对照各章节的例题亲自计算一遍。很多学过卫生统计的人都有体会，如果只听课不作练习或者只阅读不计算，等于没学统计。因此书中介绍的每一个统计方法都至少附一个例子供读者练习，并且尽量将每一个计算步骤写的详细一些，便于读者的学习和理解。此外，我们还编写了一本学习指导作为本书的配套教材，为读者提供更多的练习机会。

本书作者都是从事卫生统计教学工作的中青年教师，我们将收集到的一些经典例子和教学中的个人体会写入此教材，不仅是让大家与我们一起分享成果，更主要的是以此表达我们对卫生统计学界老前辈的谢意，因为我们是在读了他们的书后成长起来的。同时我们感谢所有被引用为参考文献的原作者。感谢曹卫华、安琳副教授为初稿提出的修改建议。特别要感谢的还有陈育德教授，他在百忙中多次认真审阅全部书稿，并提出了许多宝贵意见。由于作者的水平和实践经验有限，书中不足和错误之处在所难免，恳请读者批评指正。

康晓平

2002 年 1 月 29 日

目　录

第一篇　卫生统计学基本概念和方法

第二篇 卫生服务与居民健康统计

第一篇

卫生统计学基本概念和方法

第一章 绪 论

第一节 卫生统计学的任务及其内容

在医学实践与研究以及卫生管理工作中，卫生统计学（health statistics）作为一种认识事物数量特征的重要工具，已越来越被人们所接受。例如，将实际工作中的原始数据转变成有价值的信息，需要统计；作流行病学调查，研究各种危险因素与疾病的关系，也需要统计；阅读医学杂志评价别人的研究结果，需要懂统计；进行新药临床试验和各种疗法疗效的比较研究，需要用统计；总之，无论在临床医学、预防医学和卫生管理各个方面的科学研究以及防治工作计划的拟定和成果评价，只要作数量分析都要用到统计。卫生统计学是运用数理统计的基本原理和方法研究预防医学和卫生事业管理中资料的收集，整理和分析的一门应用科学。具体地讲，是将按照设计方案收集上来的数据进行整理分析，透过众多的偶然的次要的因素阐明事物客观存在的规律性，辨别事物间在数量上的差别是否仅是偶然现象，从而作出比较正确的结论。

卫生统计学的基本内容包括三个方面：①卫生统计学的基本原理和方法，包括研究设计和数据处理中的统计理论和方法。②健康统计，包括医学人口统计、疾病统计和生长发育统计等。③卫生服务统计，包括卫生资源、医疗卫生服务的需求和利用、医疗保健制度和管理等的统计问题。根据教学目的的需要，本教材包括了卫生统计学的主要内容，即①数据处理中的基本统计概念和方法（第一章至第九章），以及常用的综合评价方法（第十章）。②卫生服务统计中的卫生服务调查统计（第十二章）。③健康统计中的疾病统计、医学人口统计和寿命表方法（第十一章、第十三章、第十四章）。研究设计不仅是卫生统计学的基本内容，也是卫生管理过程中的计划、实施、监督、评价等基本环节的基础。为了减少重复，未将研究设计作为独立章节列入本书，而是根据卫生管理专业的特点，将调查设计的一般原则在本书第十二章卫生服务调查统计中作了介绍。

第二节 卫生统计在卫生事业管理中的作用

同医学研究对象一样，卫生事业管理的研究对象也存在许多不确定性，例如，同样规模的医院，他们的病人住院天数不同、各种疾病的病死率也可能不同。面对这些不确定现象，如果管理者凭经验作出某项决策必然带有盲目性。如果按科学管理的原则作决策，就应该先做调查，收集有关数据，并在对这些数据进行处理和分析的基础上，找出有用的信息，从而作出具有定量依据的决策。如何去收集和分析资料，以及如何对分析结果作合理解释，卫生统计正是处理这一问题的有效工具和手段。随着政府部门职能转变，宏观管理与科学决策日益提到各级卫生行政部门的议事日程。科学的决策取决于是否拥有必要的统计信息，卫生事业改革发展以及卫生服务过程的管理均需要统计信息的支持。例如，卫生资源发展规模、速度、结构、效益，人群健康状况、卫生服务的需求状况、卫生服务利用程度、效率与效果等

方面都要有相应的客观指标来描述和解释。

卫生统计信息系统是卫生事业管理的重要组成部分，它在卫生事业宏观管理与科学决策中将发挥越来越重要的作用。随着电子计算机和大型统计软件包（如 SPSS，SAS 统计软件包）的迅速发展与广泛应用，不但使基础的、单变量的统计分析工作大为简化，也为解决复杂问题如多指标的综合分析、多变量分析等提供了有力的工具；而且最终实现了卫生统计信息系统的自动化，包括数据录入、存储、传输、数据处理和分析整个管理过程。

第三节　统计资料的类型

卫生统计资料一般分为三大类，即计量资料，计数资料和等级资料。不同类型的资料选用不同的统计指标和统计分析方法。根据分析需要，各类资料可进行互相转化。

一、计量资料（quantitative data）

用度量衡的方法测定每个观察单位的某项研究指标量的大小，所得到的数据（即测量值）称为计量资料。计量资料通常是有度量衡单位的，属于连续性资料。例如，调查某地 12 岁男孩的身体发育状况。这时，每个男孩就是一个观察单位，身高（cm），体重（kg），血压（mmHg 或 kPa）均可做为观测指标。通过测定每个男孩的这三项指标量的大小，所得到的身高值，体重值和血压值为计量资料。描述计量资料常用的统计指标有平均数，标准差等（见第二章）。统计分析方法有 t 检验，u 检验，方差分析，直线相关与回归（见第六章，第九章）。

二、计数资料（categorical data）

将全体观察单位按照某种性质或类别进行分组，然后分别清点各组中的例数，这样得到的数据称为计数资料，也称分类资料。计数资料一般没有度量衡单位，是一种间断性的资料。例如，对某卫生机构做人力资源调查。这里每个工作人员作为一个观察单位，将全体工作人员按技术人员和非技术人员分为两组，清点每组中的人数，所得资料为计数资料，也称二分类资料；又如，将这个例子中的工作人员再重新分为医护人员、非医护技术人员和管理人员三组，清点每组中的人数，所得计数资料称多分类资料，或称无序分类资料。计数资料常用的统计指标有率、构成比等（见第三章）。统计分析方法有 u 检验，χ^2 检验（见第七章）。

三、等级资料（ordinal data）

将全体观察单位按照某种性质的不同程度分为若干组，分别清点各组中观察单位的个数，这种数据资料称为等级资料。等级资料是介于计量资料和计数资料之间的一种有序分类资料，一般没有度量衡单位，也是一种间断性的资料。例如，为了观察黄连素对细菌性痢疾的疗效，以菌痢患者作为观察单位，按疗效的不同程度，将接受治疗的菌痢患者分为治愈、显效、有效和无效四组，分别计算各组中的菌痢患者人数。所得资料在疗效分组上有定量的性质（按程度排列），但不确切；在清点各组人数上又有定性的特征，因此属于等级资料，也称半定量资料或有序分类资料。等级资料的统计指标也可用构成比表示（见第三章）。统计分析方法可以用秩和检验（见第八章）。

四、数据转换（data transformation）

根据分析的需要，计量资料，计数资料和等级资料之间经常要做转换。

1. 定量数据的性质化转换　例如，观察得到100名婴儿出生体重（克），这是计量资料，可以计算他们的平均出生体重（克）。如果想分析有多少婴儿属于低出生体重，多少婴儿是正常出生体重，可以将这100名婴儿的出生体重分为<2500克（低出生体重）和≥2500克（非低出生体重）两组，这时就成了两分类的计数资料。如果分组再细一些，将出生体重分为<2500克（低出生体重），2500～3999克（正常出生体重），≥4000克（高出生体重），这时计量资料就成了等级资料。

2. 定性数据的数量化转换　很多情况下，数据需要计算机处理。为了便于计算机的识别和运算，对定性数据可以赋值进行数量化转换。例如，性别是属于计数资料的两分类变量，可将男女分别取值为1和2。取值1和2之间没有量的差别，只是一种“数据代码”。如果文化程度是按文盲，小学，初中，高中，大学及以上分组，此变量属于等级资料，可分别取值为0，1，2，3，4。取值0，1，2，3，4之间不仅是一种“数据代码”，而且也有量的差别。

第四节　统计学基本概念

一、总体与总体研究（population and population study）

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变量值的集合。这里的观察单位，亦称个体，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、一个样品等，无论何种研究都要先确定观察单位。只有观察单位明确，才能确定总体范围。例如，调查某地1998年20岁健康男大学生的身高。该地区具体的每个20岁健康男大学生就是一个观察单位，该地1998年所有20岁健康男大学生的身高值就构成一个总体。又如，了解某市某年三级甲医院的病床数。该市每个三级甲医院就是一个观察单位，该市某年所有三级甲医院的病床数就构成一个总体。这里的总体明确了一定时间、一定空间的有限个观察单位，称为有限总体。对有限总体中的每个个体都作观察就称总体研究。有时总体是抽象的，如观察用某药治疗过敏性哮喘的效果，这里总体的同质基础是用某药的过敏性哮喘病人，但没有治疗时间和地点的限制，观察单位数是无限的，该总体称无限总体。而无限总体是无法作总体研究的。

二、变量与变量值（variable and value of variable）

观察单位（或个体）的某种属性或标志称为变量；对变量进行测量或观察的值称为变量值（或测量值、观察值）。如调查某市某年三级甲医院的病床数。病床就是变量，而每一个三级甲医院的病床数就是变量值。又如，调查某地成年人的高血压患病情况。调查问卷中的年龄、性别、体重、血压等项就是变量，这些变量中的年龄、体重、血压属于计量资料或者称数值变量，性别则属于计数资料或者称分类变量。而测得每一个成年人的具体年龄、性别、体重、血压值就是变量值。

三、同质与变异（variation）

研究对象具有相同的背景、条件、属性称同质；同一性质的事物，其个体观察值（变量值）之间的差异，在统计学上称为变异。统计学所研究的对象是以同质为基础，并具有变异的事物或现象。例如，调查某地1998年所有20岁健康男大学生的身高。它的同质基础是同一地区、同一年份、同为20岁健康男大学生；这些20岁健康男大学生的身高值有的相同，有的不尽相同，存在差异，这种身高值之间的差异就是变异。又如，研究某种新药治疗胃溃疡的效果，所有研究对象都必须是确诊为胃溃疡的病人，而且病情相同，不可包括疑似病人或根本不是胃溃疡的病人。在这种同质的基础上观察治疗效果，有的人治愈，有的人未愈，这种差异就是变异。

四、样本与随机抽样（sample and random sampling）

从总体中随机抽取有代表性的一部分个体，其测量值（或观察值）的集合称为样本。所谓随机抽样，就是总体中每个个体都有均等机会被抽取，抽到谁具有一定的偶然性。随机抽样的方法很多：有单纯随机抽样、整群抽样、系统抽样、分层抽样等（参见流行病学教材）。例如，要了解某地1998年所有20岁健康男大学生的身高。从该地区用系统抽样或其他方法随机抽取120名20岁健康男大学生，分别测其身高值。这120名20岁健康男大学生的身高值就是样本。

五、抽样研究（sampling study）

对从所研究的总体中随机抽取有代表性的一部分个体构成的样本进行研究称为抽样研究。抽样研究的目的是通过用样本资料计算的指标去推论总体。由于总体较大，要收集所有观察单位的数据既费时、费力还容易产生误差；对于无限总体，又不可能观察到每一个个体，所以医学研究的资料多数是通过抽样研究去获得。如要了解某地1998年20岁健康男大学生的平均身高。该地1998年所有20岁健康男大学生的身高值是一个总体，但是我们不可能，而且也没有必要把每个20岁健康男大学生都找到测其身高值。因此可以从总体中随机抽取一定数量的20岁健康男大学生的身高值做为样本（例如样本量为120），并计算样本的平均身高（$\bar{x}$）。如果这个样本均数是有代表性的，而且是可靠的，即可用该样本的平均身高（$\bar{x}$）推论该地1998年20岁健康男大学生的平均身高（μ）。

六、参数和统计量（parameter and statistic）

参数是指总体指标。如总体均数（μ），总体率（π），总体标准差（σ）等。统计量是指样本指标。如样本均数（$\bar{x}$），样本率（p），样本标准差（s）等。如某地1995年全部正常成年男子的平均红细胞数（μ）即为总体参数，而从该总体中随机抽取的144名正常成年男子的平均红细胞数（$\bar{x}$）为样本统计量。一般情况下，参数是未知的，需要用统计量去估计。用统计量推论参数的方法，统计学上称为参数估计（例如，总体均数的区间估计见第五章）和参数检验（例如，t检验见第六章）。

七、统计描述与统计推断（statistical description and statistical inference）

用统计图表或计算统计指标的方法表达一个特定群体（这个群体可以是总体也可以是样

本）的某种现象或特征，称统计描述；根据样本资料的特性对总体的特性作估计或推论的方法称统计推断，常用方法是参数估计和假设检验。需要注意：随机抽样得到的样本资料既可做统计描述，也要做统计推断；而总体资料只作统计描述，无须作统计推断。例如，用某地1998年120名20岁健康男大学生的身高值绘制直方图表示频数分布的类型，或计算身高的平均数表示平均水平的方法即为统计描述；用120名20岁健康男大学生的身高的平均值去估计该地1998年所有20岁健康男大学生的身高的平均值的方法为统计推断。又如，比较两个县某年的婴儿死亡情况，资料分别来自该年全县的婴儿死亡和出生登记（忽略漏报因素）。此时可计算两个县的婴儿死亡率，直接比较他们的死亡水平，而不必作假设检验。因为资料是来自某年全县的常规报表，不是抽样调查得到的样本。

八、误差（error）

任何周密设计的科学研究，都不可能没有误差。医学科学研究中的误差通常指测量值与真值之差，其中包括系统误差和随机测量误差；以及样本指标与总体指标之差，即抽样误差。随机测量误差及抽样误差都属于随机误差，其中抽样误差是统计学研究和处理的重要内容。

（一）系统误差（systematic error）

这种误差不是偶然机遇造成的，而是某种必然因素所致，具有一定的倾向性。其特点是观察结果一惯性的往一边偏，要高都高，偏低都低。系统误差一旦发生，统计学是无能为力的，因此要尽可能避免。而大多数系统误差可以通过周密的研究设计和调查（或测量）过程中的严格质量控制措施得以解决。常见情况：①操作方法不正确或对调查问卷理解有误；②医生掌握疗效标准偏高或偏低；③周围环境的改变。如实验室内室温过高或过低，作用时间掌握不够一致；以及现场调查时出现不必要的行政干预；④仪器不准或试剂不合格。例如，测量血压，要求血压计的水银面与0平行。如果使用的血压计没校正，高出4 mmHg，那么测定出的血压值都高4 mmHg。

（二）随机测量误差（random measurement error）

这种误差是偶然机遇所致，故无方向性，对同一样品多次测定，结果有高有低，不完全一致。随机测量误差是不可避免的，再精确的测量仪器也会存在误差，但只要将误差控制在一定的允许范围内，读出的数据都可以使用。

（三）抽样误差（sampling error）

在抽样研究中，即使消除了系统误差，控制了随机测量误差，样本统计指标和总体参数间仍会存在差别。这是由于个体变异造成的，是抽样机遇所致，是客观存在，不可避免的。这种误差可以通过统计方法估计，也可通过增大样本使其减小。我们可以通过一个实验来理解什么是抽样误差。假定已知某年某地所有13岁女学生身高的总体均数（μ）是155.4cm，总体标准差（σ）是5.3cm。该地每一个13岁女学生都有一个身高测量值，我们将她们每个人的身高测量值（cm）都录入计算机，存在数据库里做为一个有限总体。然后在这样一个有限的总体中作多次重复抽样，每次均抽取100例（n_i = 100）组成一个样本，可以算出每一个样本的平均身高（$\bar{x}_i$）。因为是完全随机抽样，数据库中的每一个女学生的身高值都有可能被抽到。最终得到的样本均数（$\bar{x}_i$）可能是153.6，153.1，154.9，···158.7等。这是在一个人为的控制得非常好的条件下进行的。我们看到每个样本均数（$\bar{x}_i$）与总体均数（μ）间仍有一个差，而且样本均数与样本均数间也有差别。这种误差既不是系统误差，也不是测

量误差，完全是由抽样造成，是偶然的机遇。因此我们讲，只要是抽样研究，必然存在抽样误差。这种误差虽然是不可避免的，但可以认识它，估计它，并可缩小它。

九、概率与频率（probability and frequency）

概率是对总体而言，频率是对样本而言。概率是指某随机事件发生的可能性大小的数值，常用符号 P 来表示。随机事件的概率在0与1之间，即 $0 \leqslant P \leqslant 1$，常用小数或百分数表示。$P$ 越接近1，表明某事件发生的可能性越大，P 越接近0，表明某事件发生的可能性越小。如用某药治疗200个病人，其治愈率为80%，这是一个频率指标。频率是指一次试验结果计算得到的样本率。若经过多次试验和许多人的治疗，其治愈率稳定在80%，这时可以说，某药治愈某病的可能性，即概率为80%。统计中的许多结论都是带有概率性的。一般常将 $P \leqslant 0.05$ 或 $P \leqslant 0.01$ 称为小概率事件，表示某事件发生的可能性很小。具体的应用，在以后的章节中将会介绍。

第五节　统计工作的基本步骤

计划与设计、收集资料、整理资料和分析资料是统计工作的四个基本步骤。这四个步骤是紧密联系不可分割的，某一环节发生错误，都可影响统计分析结果。

一、计划与设计（plan and design）

计划是开展研究工作的前提和依据。一个全面完整的计划应包括研究设计和组织管理两方面的内容，即资料收集、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设、研究对象、抽样方法与样本含量、研究内容和问卷设计、研究方法和技术路线、统计指标和分析方法、质量控制和预期结果、经费预算、人员安排和进度等等。按研究者是否对观察对象施加干预（即处理因素），研究设计可分为调查设计和实验设计两大类。调查设计（不加干预）主要是了解客观实际情况的现场工作。实验设计（加干预）根据研究对象不同分为动物实验和临床试验（或现场试验）。无论是调查设计，还是实验设计均包括专业设计和统计学设计两个方面。专业设计是运用专业理论技术知识进行设计，统计学设计是运用统计学知识和方法进行设计。两者应相互结合，缺一不可。

二、收集资料（collection of data）

其任务是取得准确可靠的原始数据。

（一）统计资料的来源

统计资料的来源是多方面的，可概括为经常性资料和一时性资料两大类：

1. 经常性资料：一般指医疗卫生工作中的原始记录。①统计报表。如医院工作报表，居民病伤死亡原因报表，疫情报表等。②医疗卫生工作记录和报告单（卡）医院各科门诊病历，住院病历，健康检查记录，各种医疗和检验记录及传染病报告卡等。

2. 一时性资料：根据专题调查或实验研究的需要而临时设计的调查表或调查问卷，如卫生服务调查、卫生人力资源调查等。

（二）统计资料的要求

原始资料是统计工作的基本依据，把好收集资料这一关，要求做到：

1. 资料必须完整、正确和及时。完整是指调查项目填写完整无空项。若数字不详可用代码填写，如年龄不详，填“99”。正确是指填写的内容准确无误，保证资料的真实可靠。及时是指资料的时间性，要按规定时间完成资料的收集，尽快反馈信息。

2. 要有足够的数量。原始数据要有一定的数量才能反映事物的规律性。但不是越多越好，足够即可。多少数量算足够？①根据研究目的确定：制定参考值范围要求样本量至少上百例，观察药物疗效要求样本量至少数十例。②根据资料类型确定：计量资料样本数量可少些，计数资料样本数量应多些。③根据允许误差计算样本数量（参考有关卫生统计学书籍）。

3. 注意资料的代表性和可比性。代表性是指做专题研究时遵循随机化原则收集资料，即总体中每一个体都有同等的机会被抽取。可比性是指在统计比较时，对比的各组之间，除观察问题或实验因素不同外，其它条件都要求尽量一致。因此对于做两个样本或多个样本的比较研究，收集资料时一定要注意资料间的可比性问题。

三、整理资料（sorting data）

任务是清理原始数据，使其系统化、条理化，以便进一步计算指标和分析。

（一）原始数据的检查与核对

检查核对原始数据有无错漏，以及数据间的相互关系是否合乎逻辑，并予以必要的补充，修正与合理的剔除。对原始记录的检查核对，应在调查现场完成，而整理资料过程则是从不同角度、用不同方法进一步净化数据。

1. 统计数据的常规检查。①检查原始记录的数据有无错误和遗漏。②调查项目是否按要求或填表说明填写。③统计报表的行栏合计应与总计相符。

2. 数据的取值范围检错。可利用频数分布表检查是否有异常值的出现，如在“结婚年龄”的频数表中有 15 岁、16 岁结婚的妇女，这时就要返回原始数据核查，确认这些异常值是真实情况还是因某一环节出错所致。

3. 数据间的逻辑关系检错。逻辑检查是为了查明资料项目之间是否有矛盾。例如吸烟的调查，某一被调查者的年龄项目填写“23 岁”，吸烟史项目填写“20 年”，这意味此人 3 岁就开始吸烟，显然是填写错误。这时再结合其它项目进一步核实，确认是年龄项目填错了，或是吸烟史项目填错了。

（二）数据的分组设计和归纳汇总

按资料的性质和数量特征分组，以反映事物的特点。例如，整理某地居民高血压病发病资料时，除了求出总的发病人数外，还要按年龄、性别、地区、劳动环境和生活环境等多种特征进行分组，得出各组的发病人数和发病率，才能对发病的重点人群、多发地区、与疾病有关的环境等进行研究。常用的分组方法有以下两类：

1. 质量分组　按事物的性质或类型分组，这种方法多适用于计数资料或等级资料。如病人按性别、职业等分组；疗效按治愈、好转和无效等分组。根据研究需要，有时也可将计量资料转换成计数资料或等级资料，进行质量分组。例如，舒张压 <90mmHg，为正常血压，≥ 90mmHg，为高血压，这样将测定到的血压值（计量资料）分为正常和非正常两组（计数资料）。

2. 数量分组　按观察值的大小进行分组，这种方法多适用于计量资料。分几组合适？要根据研究内容的特点和分析目的来定。例如，冠心病多发于中、老年人。年龄分组时，应把中、老年组分得细些，如 5 岁一组；青、少年组分得粗些，如 10 岁一组。另外也要根据

观察数据的多少来定。例如，当观察例数在100例以上，分8~15组较合适。以1998年某地120名20岁健康大学生身高为例说明数量分组的步骤，即编制频数表（见本书第二章表2.1）。

四、分析资料（analysis of data）

其任务是按研究设计的要求，结合资料的类型计算有关指标，阐明事物的内在联系和规律。统计分析主要包括：①用一些统计指标，统计图表等方式表达和描述资料的数量特征和分布规律，不涉及由样本推论总体的问题。②对样本统计指标作参数估计和假设检验，并结合专业知识解释分析结果，目的是用样本信息推断总体特征。

由于计算机的普及，除了计划设计和资料收集需要大量的人工操作外，整理资料和分析资料都可在计算机上完成。一些统计软件，例如，SPSS、SAS等都可以作数据的录入、检错、整理和分析。用计算机替代手工计算，速度和效率确实提高了许多，但统计结果的正确性并不像速度一样成倍数增长。例如，计算机可以将散乱的数据整理成分组资料。但分几组合适？按数量分组还是按质量分组，这些仍需要人们事先根据统计知识和专业知识将组数确定，最后由计算机完成。另外，任何数据进了计算机都能很快做显著性检验、甚至多因素分析。但如果数据质量不好，或者统计方法选择不正确，都将影响到最终的统计结果。由此可见，无论计算工具多么先进，统计工作的四个步骤都是不可被忽略、被替代的。

小　结

1. 卫生统计学是运用数理统计学的原理和方法研究预防医学和卫生事业管理中资料的收集，整理和分析的一门应用科学。它的主要内容包括：卫生统计学的基本原理和方法、健康统计和卫生服务统计。

2. 卫生事业管理的研究对象也存在许多不确定性，因此，要利用卫生统计这个有效工具，充分发挥卫生统计的信息、咨询、监督的整体功能，为满足决策管理和卫生服务研究的需要。

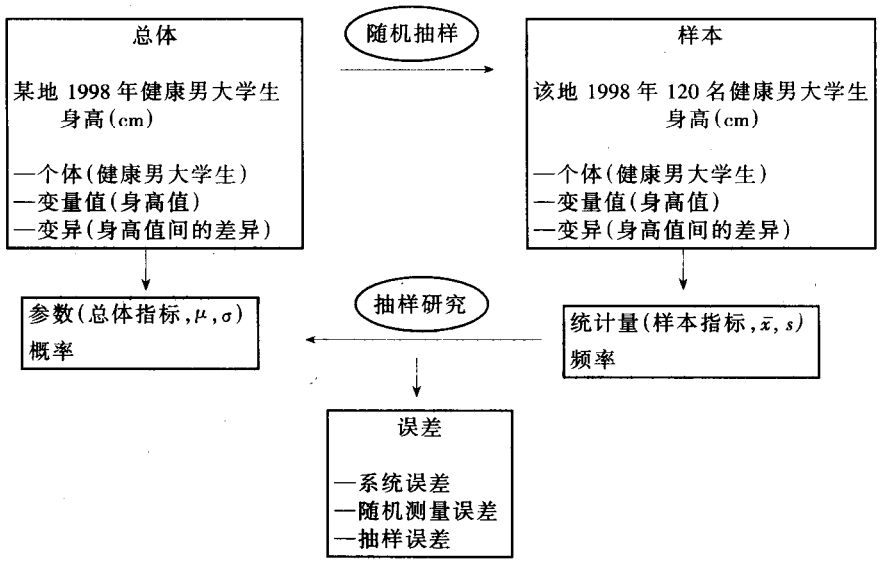
3. 卫生统计资料一般分为三大类，即计量资料，计数资料和等级资料。不同类型的资料选用不同的统计指标和统计分析方法。根据分析需要，各类资料可进行互相转换。

4. 几个基本概念间的关系（见示意图）

5. 收集资料过程中，系统误差尽可能避免或通过周密的设计解决。随机测量误差及抽样误差都属于随机误差，是不可避免的。随机测量误差应控制在一定范围内；抽样误差可通过统计方法估计并减小。

6. 统计工作一般分为计划与设计、收集资料、整理资料和分析资料四个基本步骤。任何一步发生错误，都会影响统计结果及结论的正确性。

7. 学习卫生统计学重点是掌握统计的基本概念、基本知识、基本方法和基本计算。对统计公式的推导不作深究，只要求正确计算、了解其含义、用途和适用条件。培养统计思维和分析问题的能力，学会用卫生服务和居民健康统计的指标综合评价居民健康状况，为科学管理提供统计信息。



几个基本概念之间关系的示意图

思 考 题

1. 卫生事业管理专业与卫生统计学的关系是什么。
2. 三种类型的统计资料有何不同。
3. 总体和样本的关系是什么。
4. 统计描述和统计推断的内容是什么。
5. 抽样研究的目的是什么？为什么有抽样误差。
6. 简述统计工作四个基本步骤。
7. 收集资料时，对统计资料的要求是什么？

（康晓平）

第二章　计量资料的统计描述

计量资料统计描述的目的是了解资料的分布类型，并根据分布类型选用适当的指标描述其集中趋势和离散趋势。若资料服从正态分布，可用正态法估计其参考值范围。

第一节　计量资料的频数表

当计量资料的观察值较多时，为了解其分布规律和类型，需对资料进行整理，编制频数分布表，简称频数表（frequency table）。它包括一些有序的区间（或组段）及落在各区间（或组段）内的观察值的个数即频数。

一、频数表的编制

例 2.1　某地 1998 年随机抽查 120 名 20 岁健康男大学生身高（cm），资料如下，试编制频数表。

175.7	171.6	172.4	170.5	172.3	163.8	172.4	167.5	173.6	175.0
178.4	170.4	169.9	173.6	172.0	172.1	179.4	173.1	172.4	180.5
170.4	178.2	172.9	172.7	179.6	174.5	174.8	172.0	175.8	172.7
170.0	168.5	173.8	168.9	179.9	172.4	166.5	171.6	177.0	171.4
170.3	167.4	174.3	172.3	175.3	170.4	171.6	174.1	171.6	173.8
162.8	172.7	174.0	179.6	166.7	166.6	164.3	177.8	182.7	171.4
168.9	175.2	176.7	169.5	176.3	177.7	172.1	166.6	177.1	176.1
171.5	170.1	176.5	174.4	175.3	181.5	174.1	168.4	174.9	167.9
175.3	172.3	174.2	174.4	173.5	171.9	167.4	181.7	179.5	177.3
166.9	168.4	175.2	173.3	172.9	173.6	165.3	171.9	169.1	168.9
178.2	169.5	172.1	178.4	166.6	165.8	171.1	174.9	176.7	174.8
168.2	178.1	170.5	172.3	172.3	169.8	168.1	172.1	180.0	171.2

（一）求极差（range）

极差也称全距，用符号 R 表示，即最大值与最小值之差。本例最大值为 182.7cm，最小值为 162.8cm，极差 $R = 182.7 - 162.8 = 19.9$（cm）。

（二）确定组数、组距和组段

1. 组数：不宜过多也不宜过少。百余例的资料一般设 8～15 个组。可根据观察单位的多少酌情增减组数。

2. 组距：用符号 i 表示，相邻两个组的下限之差为组距。各组之间可以等组距，也可以不等组距，一般多采用等组距。组距 i =（极差/组数）取整，取整只是为了便于资料整理及运算。本例先将组数初步定为 10，则以极差的 1/10 取整作为组距，组距 $i = 19.9/10 \approx 2$cm，该资料最后共分 11 个组。

3. 组段的上、下限：每个组段的起点被称为该组的下限，终点为上限。因为是连续性资料，为避免混淆，常用表 2.1 第（1）栏的表示法，即各个组段从本组段的“下限”开始，不包括本组段的“上限”，最后一个组段应同时写出“下限”和“上限”，如本例最后一个组