附：实用卫生统计学自学考试大纲

实用卫生统计学

[2006年版]

组编／全国高等教育自学考试指导委员会
主编／康晓平

全国高等教育自学考试指定教材 营养、食品与健康专业（专科）

北京大学医学出版社

全国高等教育自学考试指定教材
营养、食品与健康专业（独立本科段）

实用卫生统计学

（2006年版）

（附：实用卫生统计学自学考试大纲）

全国高等教育自学考试指导委员会　组编

主　编　康晓平
编　者　（按姓氏笔画为序）
何平平　郑迎东
易伟宁　康晓平
主　审　陈育德
参　审　安　琳　韩少梅

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（2006年版）/康晓平主编. —北京：北京大学医学出版社，2006.8

全国高等教育自学考试指定教材

ISBN 7-81116-089-7

Ⅰ.实… Ⅱ.康… Ⅲ.卫生统计—高等教育—自学考试—教材 Ⅳ.R195.1

中国版本图书馆CIP数据核字（2006）第075127号

实用卫生统计学（2006年版）

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：莱芜市圣龙印务有限责任公司

责任编辑：药　蓉　　**责任校对**：杜　悦

开　　本：787mm×1092mm　1/16　　**印张**：13.25　　**字数**：328千字

版　　次：2006年8月第1版　2006年12月第1次印刷　**印数**：1—3000册

书　　号：ISBN 7-81116-089-7/R·089

定　　价：21.00元

组编前言

21世纪是一个变幻莫测的世纪，是一个催人奋进的时代。科学技术飞速发展，知识更替日新月异。希望、困惑、机遇、挑战，随时随地都有可能出现在每一个社会成员的生活之中。抓住机遇，寻求发展，迎接挑战，适应变化的制胜法宝就是学习——依靠自己学习，终身学习。

作为我国高等教育组成部分的自学考试，其职责就是在高等教育这个水平上倡导自学、鼓励自学，为每一个自学者铺就成才之路。组织编写供读者学习的教材就是履行这个职责的重要环节。毫无疑问，这种教材应当适合自学者增强创新意识、培养实践能力、形成自学能力，也有利于学习者学以致用，解决实际工作中所遇到的问题。具有如此特点的书，我们虽然沿用了“教材”这个概念，但它与那种仅供教师讲、学生听，教师不讲、学生不懂，以“教”为中心的教科书相比，已经在内容安排、形式体例、行文风格等方面都大不相同了。希望读者对此有所了解，以便从一开始就树立起依靠自己学习的坚定信念，不断探索适合自己的学习方法，充分利用已有的知识基础和实际工作经验，最大限度地发挥自己的潜能，达到学习的目标。

祝每一位读者自学成功。

本教材由全国考委医药学类专业委员会遴选作者、安排编写、组织审稿，保证了医药学类自考教材的质量。

欢迎读者提出意见和建议。

全国高等教育自学考试指导委员会

2006年4月

编者前言

《实用卫生统计学》是为全国高自考营养、食品与健康专业（独立本科段）学生编写的教材。根据高自考学生的特点，本教材的编写注意到以下两个方面：

1. 强调理论与实用结合。全书共分十章，第一章至第九章介绍卫生统计学的基本理论、基本知识和基本技能，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。第十章主要对科研论文写作中有关统计方法的应用进行了简介，为学生做毕业论文提供参考。前九章的编写重点突出了三个方面的内容：（1）尽量举一些与营养、食品与健康相关的实际例子帮助理解各章的统计基本概念和方法，目的是让学生通过本章内容的学习能够举一反三解决类似的实际问题；（2）书中统计公式较多，为了突出重点将常用的公式和内容进行了加框标记，方便学生在自学过程中认读、理解和代入数据计算；（3）用章末小结的形式，帮助学生进行归纳、理解各章节的重点和难点。

2. 各章都备有大量的习题供学生练习。为了方便学生的自学，我们对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，并希望记住一些有代表性的例题会做到举一反三。另外，将各章节重点和难点的内容以各种题型编写成练习，学生可以参照考试大纲的要求选做练习，并从不同角度去理解和熟悉常用的基本概念和方法。学习卫生统计学的最好方法就是亲自做题，首先对照各章节的例题按计算步骤做一遍，然后做书中的练习题。教材中加框标记的公式不要求死记硬背，但要会认读，会根据所给的资料选择适合的公式代入计算，并作出解释。

本教材和全日制在校生的教材要求大不相同。为编写一本适合高自考学生自学的书，参加编写教材的老师反复修改稿件，使抽象的概念具体化，复杂的计算简单化，尽量做到通俗易懂。但由于卫生统计学是一门比较抽象的课程，我们建议学生在自学的过程中要静下心、花时间、多做练习，除此之外没有捷径可走。

在此，特别感谢本教材的主审：陈育德教授。他在百忙中认真审阅了书稿和大纲，并提出了许多宝贵意见。感谢安琳教授和韩少梅副教授对书稿和大纲提出的修改意见，并感谢缪之文同学在第十章文献检索和编写方面的贡献。由于作者的水平和实践经验有限，书中不足和错误在所难免，恳请读者批评指正。

康晓平

2006 年 5 月 18 日

目　录

实用卫生统计学

附　实用卫生统计学自学考试大纲

第一章 绪 论

第一节 卫生统计学的内容及其应用

在公共卫生实践以及营养健康科学研究中，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；进行两组样本数据的比较或两组动物实验结果的比较时，需要用统计。总之，在临床医学、预防医学和公共卫生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中，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对预防医学和公共卫生领域中的科学研究进行设计，以及研究资料的收集、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，从而得出比较正确的结论。

卫生统计学的基本内容包括三个方面：①卫生统计学的基本理论和方法，包括研究设计和数据分析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，本教材主要包括卫生统计学第一个方面的内容，重点介绍了统计描述的常用指标、统计推断的参数估计和假设检验，以及实验设计和调查设计的一般原则。

卫生统计学在实际应用中也被狭义地分为设计和数据分析两个部分。第一部分是设计，通常分为实验设计和调查设计。例如，某研究者为了解螺旋藻的保健功能对患有糖尿病的小鼠作降血糖实验，按初始血糖浓度将 20 只小鼠随机分为两组，一组为空白对照，另一组给螺旋藻，然后观察血糖是否有变化。为了得到这个实验结果，首先要作研究设计。这种研究设计，即研究对象接受了某种干预（或处理）后获得的数据属于实验设计。调查设计一般是指为了对某个特定人群的现况作调查而进行的研究设计。例如，“2005 年某地区小学生营养膳食调查”属于现况调查的研究，研究者希望通过调查研究知道这个地区小学生的饮食结构是否合理，饮食习惯是否良好？如何获得这些数据并整理成最终结果，需要在调查设计中考虑。无论实验设计还是调查设计都包括统计学设计，两者在统计学设计上不尽相同，有关内容将在本教材的第九章介绍。第二部分是数据分析，即对那些按照不同研究设计收集上来的数据进行分析，主要内容是进行统计描述和统计推断。例如，计算每组 10 只小鼠的平均血糖值，或者计算 200 名小学生不良饮食习惯的发生率，这些都属于统计描述的内容。如果有两个样本数据要比较它们的样本均数或样本率，就要作假设检验，例如比较两组小鼠使用螺旋藻后的平均血糖变化值用 t 检验，或者比较男女两组小学生不良饮食习惯的发生率用 χ^2 检验，这些都属于统计推断的内容。在本教材的第二章到第四章主要介绍统计描述中统计指标的计算和统计指

标的表达形式；第五章到第八章介绍了统计推断的参数估计和假设检验。

第二节　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，属于连续型资料。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）均可作为观测指标。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）。根据资料的总体分布类型不同，应选用不同的统计分析方法（见第六章）。

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，对某卫生机构作人力资源调查。这里每个工作人员被看作一个观察单位，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）。统计分析方法主要有 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，分别计算各疗效组的菌痢患者人数。这类资料在疗效分组上有定量的性质（按程度排列），但不精确；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有序分类资料的秩和检验（见第七章）。

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作转换。

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。

第三节　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2004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2004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等医院的病床数。该市每个三级甲等医院就是一个观察单位，该市某年所有三级甲等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等医院的病床数。病床就是变量，而每一个三级甲等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、职业、文化程度、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；性别和职业均属于计数资料或者称分类变量，再细分，性别可以看成二分类变量，职业为多分类变量或者称无序分类变量；文化程度属于等级资料或者称有序分类变量。而测得每一个成年人的具体年龄、职业、文化程度、体重、血压值就是变量值。

三、同质与变异（homogeneity and 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2004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（见第九章）。例如，要了解某地2004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。如欲了解某地2004年20岁健康男大学生的平均身高。该地2004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作为样本（例如样本量为120），并计算样本的平均身高（$\overline{X}$）。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（$\overline{X}$）推论该地2004年20岁健康男大学生的平均身高（μ）。

六、参数与统计量（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（$\overline{X}$）、样本率（p）、样本标准差（S）等。如某地2002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（$\overline{X}$）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要作统计描述，也要作统计推断；而总体资料只作统计描述，无需作统计推断。例如，用某地2004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 2004 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。其特点是观察结果往一边偏，要高都高，要低都低。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 4mmHg，那么测定出的血压值都高 4mmHg。

（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，存在数据库里作为一个有限总体。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，完全是由抽样造成的，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。

第四节　统计工作的基本步骤

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、设计（design）

设计是开展研究工作的前提和依据。一个全面完整的设计应包括专业设计和统计学设计两个方面的内容。专业设计是运用专业理论技术知识进行设计，还要考虑研究工作实施的组织管理。例如，研究意义、研究目的和研究假设、研究对象纳入或剔除标准、研究内容和问卷设计、预期结果和经费预算、人员安排和进度等。统计学设计是运用统计学知识和方法进行设计，即资料收集、整理和分析全过程总的设想和安排。例如，研究方法、抽样方法与样本含量、数据收集和整理、统计指标和分析方法、质量控制等。两者的知识有很多交叉，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，要在规定时间内完成资料的收集，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，足够多即可。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（见第九章）。

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，其他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。这时再结合其他项目进一步核实，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$<$90mmHg，为正常血压，$\geqslant$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。

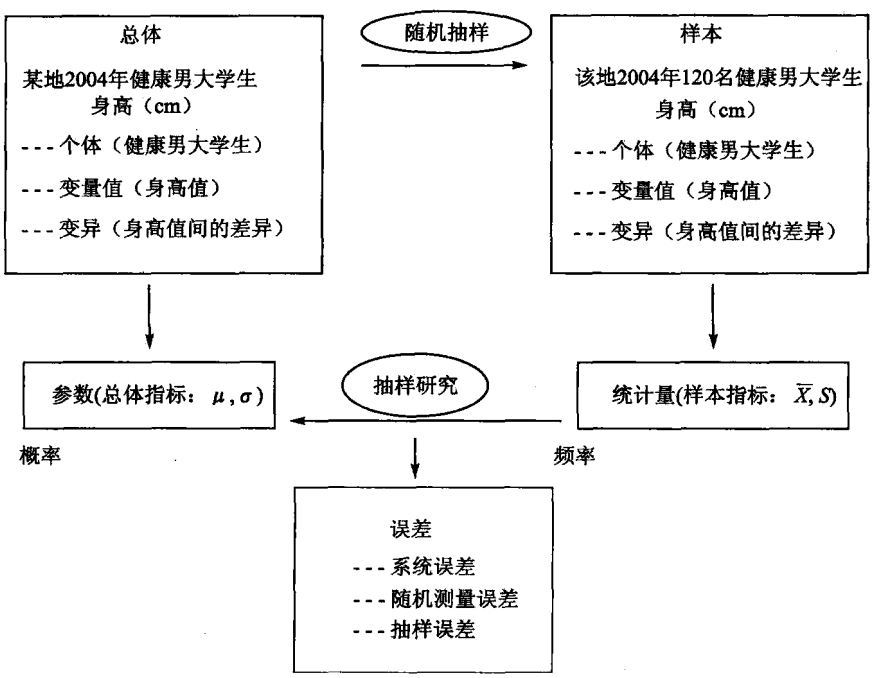
由于计算机的普及，除了计划与设计、资料收集需要大量的人工操作外，整理资料和分析资料都可在计算机上完成。一些统计软件，例如，SPSS、SAS 等都可以作数据的录入、检错、整理和分析。用计算机替代手工计算，速度和效率确实提高了许多，但统计结果的正确性并不像速度一样成倍数增长。例如，计算机可以将散乱的数据整理成分组资料。但分几组合适？按数量分组还是质量分组？这些仍需要人们事先根据统计知识和专业知识将组数确定，最后由计算机完成。另外，任何数据进了计算机都能很快作显著性检验，甚至多因素分析。但如果数据质量不好，或者统计方法选择不正确，都将影响到最终的统计结果。由此可见，无论计算工具多么先进，统计工作的四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

小　结

1. 卫生统计学是运用数理统计的基本原理和方法对预防医学和公共卫生领域中的科学研究进行设计，以及研究资料的收集、整理和分析的一门应用科学。本教材主要介绍卫生统计学的基本理论和方法，包括统计描述的常用指标、统计推断的参数估计和假设检验，以及实验设计和调查设计的一般原则。

2. 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。

3. 几个基本概念间的关系



4. 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。

5. 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。

6. 学习卫生统计学的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。

练习题

一、名词解释

1. 总体
2. 计量资料
3. 变异
4. 抽样研究
5. 统计描述
6. 统计推断

二、单项选择题（下列四个备选答案中只有一个是正确的，请选出并将其代码写在题干后面的括号内）

1. 抽样研究中的样本是［　　］
   A. 研究对象的全体
   B. 总体中特定的一部分
   C. 总体中随机抽取的一部分
   D. 随意收集的一些观察对象
2. 对某地 200 名 16 岁中学生口腔检查，发现患龋齿的人数为 54 人，该资料属于［　　］
   A. 计量资料
   B. 计数资料
   C. 等级资料
   D. 经变量转换也可作为计量资料
3. 对某样品进行测量时，由于测量仪器事先未校正，造成测量结果普遍偏高，这种误差属于［　　］
   A. 系统误差
   B. 随机测量误差
   C. 抽样误差
   D. 随机误差
4. 欲了解某市某年所有三级甲等医院的病床数，该市每个三级甲等医院就是一个［　　］
   A. 有限总体
   B. 观察单位
   C. 无限总体
   D. 观察值
5. 下面的变量中哪个是数值变量［　　］
   A. 每个病人的就诊科室
   B. 每个病人的就诊次数
   C. 每个病人就诊的疾病
   D. 每个病人就诊的医院
6. 下面哪个指标是样本指标［　　］
   A. μ
   B. σ
   C. π
   D. $\bar{X}$
7. 对男女两个样本的小学生饮食习惯的不良发生率作假设检验，这种方法属于［　　］
   A. 实验设计
   B. 总体研究
   C. 统计描述
   D. 统计推断
8. 下面哪一种统计资料的来源不属于经常性的资料［　　］