中国科学院教材建设专家委员会规划教材

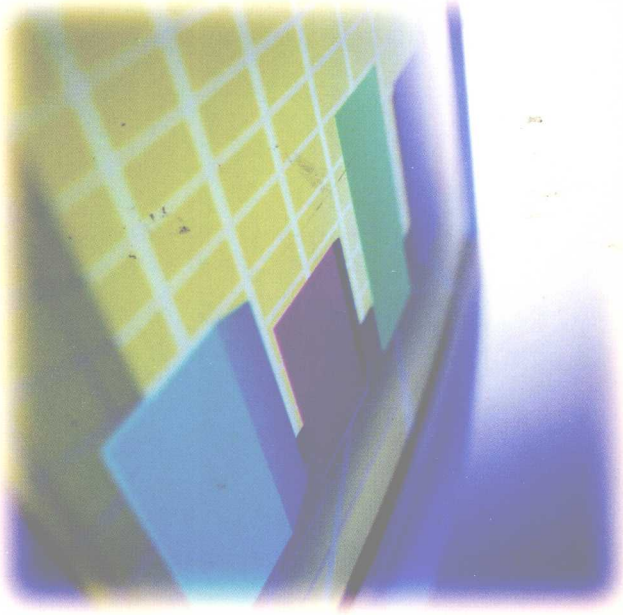
全国高等医药院校规划教材

供预防医学类、卫生管理类本科及非预防医学、非卫生管理专业研究生使用

TM 案例版

卫生统计学

丁元林 高 歌 主编

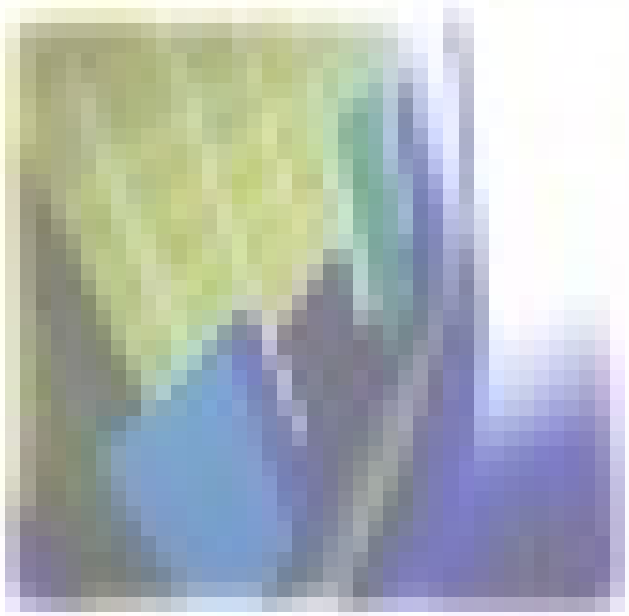


科学出版社
www.sciencep.com

[redacted]

卫生统计学

[redacted]



[redacted]

中国科学院教材建设专家委员会规划教材
全国高等医药院校规划教材

案例版™

供预防医学类、卫生管理类本科及非预防医学、
非卫生管理专业研究生使用

卫生统计学

主　编　丁元林　高　歌
副主编　刘启贵　郭秀花　黄水平　罗家洪
编　委　（按姓氏笔画排序）

丁元林（广东医学院）　易　静（重庆医科大学）
尹素凤（华北煤炭医学院）　罗家洪（昆明医学院）
孔丹莉（广东医学院）　赵宏林（内蒙古民族大学医学院）
田　俊（福建医科大学）　赵若望（包头医学院）
刘启贵（大连医科大学）　胡利人（广东医学院）
孙　忠（天津医科大学）　高　歌（苏州大学）
杨江林（郧阳医学院）　郭秀花（首都医科大学）
李向云（潍坊医学院）　黄水平（徐州医学院）
余金明（同济大学医学院）　潘发明（安徽医科大学）
沈月平（苏州大学）　潘秀丹（沈阳医学院）

秘　书　修良昌　潘海燕

科学出版社

北京

郑 重 声 明

为顺应教育部教学改革潮流和改进现有的教学模式，适应目前高等医学院校的教育现状，提高医学教学质量，培养具有创新精神和创新能力的医学人才，科学出版社在充分调研的基础上，引进国外先进的教学模式，独创案例与教学内容相结合的编写形式，组织编写了国内首套引领医学教育发展趋势的案例版教材。案例教学在医学教育中，是培养高素质、创新型和实用型医学人才的有效途径。

图书在版编目(CIP)数据

卫生统计学：案例版 / 丁元林，高歌主编．—北京：科学出版社，2008
中国科学院教材建设专家委员会规划教材·全国高等医药院校规划教材
ISBN 978-7-03-022143-8

Ⅰ．卫… Ⅱ．①丁… ②高… Ⅲ．卫生统计学-医学院校-教材
Ⅳ．R195.1

中国版本图书馆 CIP 数据核字(2008)第 076443 号

策划编辑：李国红 周万灏 / 责任编辑：周万灏 李国红 / 责任校对：陈玉凤
责任印制：刘士平 / 封面设计：黄 超

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
新蕾印刷厂印刷
科学出版社发行 各地新华书店经销
*
2008 年 7 月第 一 版 开本：850×1168 1/16
2008 年 7 月第一次印刷 印张：23 1/4
印数：1—4 000 字数：715 000
定价：42.00 元
(如有印装质量问题，我社负责调换〈新欣〉)

全国高等医药院校预防医学专业教材建设指导委员会

前　言

案例启发式教学是培养高素质、创新型和实用型医学专门人才的有效途径，已为医学教育界所共识。但目前国内与案例启发式教学相配套的案例版教材的编写才刚起步，卫生统计学教材的编写更是如此。本教材是根据全国高等医药院校预防医学专业案例版规划教材编委会的要求而编写的第一本《卫生统计学》案例版教材。

卫生统计学是一门实践性很强的学科，而目前国内现有的教材，在体系结构、编写形式和内容上，普遍存在与医学科研联系不够紧密的现象，主要问题表现为：重统计分析轻统计设计；重统计方法的计算过程轻统计方法的选择；重统计方法的软件实现轻统计方法的用途、适用条件、应用注意事项及结果的正确解释。尤为突出的是，重典型实例和“正例”的列举轻不典型实例和“反例”的剖析。其结果是：一方面，学生重视统计方法、轻视统计设计，忽略了统计方法的用途、应用条件、应用注意事项及结果的正确解释，因此，在统计方法的复杂计算过程中迷失方向，感到学习枯燥乏味进而产生畏难情绪。当碰到实际问题时，“照本宣科、依葫芦画瓢”，不能举一反三，创造性地运用统计学原理和方法解决实际问题的能力低下，在医学科研实践中误用或不恰当运用统计设计方案和统计分析方法或是不恰当地解释和报告统计分析结果。国内医学期刊中统计学误用率长期居高不下，便是有力的佐证。另一方面，教师要改变这种传统的结构式课堂教学方法需花费大量的时间和精力重新对教材进行“补充”和“调整”。

本书顺应教学改革的潮流，对内容进行精心筛选，突出“以问题为中心”的特色，在不打破原有学科体系的前提下，对课程名称及教学核心内容均不改变。本教材主要特点如下：

1. 基础性　突出“三基”和重点、难点内容，知识点明确，学生易学，教师易教，使学生在尽可能短的时间内掌握所学内容。本教材以预防医学专业和卫生管理专业为重点对象，同时可用于临床医学、基础医学、口腔、麻醉、药学、医学检验、护理学等本科专业。

2. 实用性　在编写内容和结构上，考虑教师授课的实际需要。使教师使用本教材组织教学时，既可以按传统模式讲授、案例作为补充，也可以以案例为先导进行教学，使课堂讲解内容更加形象、生动。

3. 创新性　突出“以问题为中心”的编写要求。融典型实例、“正例”和不典型实例、“反例”于教材中，由案例引出某章或节的基本知识点和重点、难点内容。对每一章或节，先列举医学或公共卫生研究中的实际案例，并根据案例提出相关问题，引导学生思考，然后再结合研究目的、设计方案和统计方法的用途、适用条件来解决实际问题。

本教材是常年从事卫生统计学教学工作的多位专家经验与智慧的结晶。在教材编写过程中，得到了科学出版社、广东药学院公共卫生学院以及各参编兄弟医学院校的大力支持。在此，我谨代表全体编委一并致谢。

限于编写人员水平所限，书中难免存在缺点或错误，欢迎读者批评指正。

丁元林

2008 年 4 月于广东东莞

目　录

第1章 绪 论

第一节 卫生统计学的作用和地位

在信息化时代的今天，或许你读了几个月的报纸，看了几个月的电视，浏览了几个月的网页，都几乎见不到一个统计公式，但几乎每天你都会接触到一些统计信息。如，第三季度国民生产总值与去年同期相比增长7.6%，12月份的CPI（消费者物价指数）涨幅为6.5%，某电视节目的收视率为22%，某感冒药的有效率达95%以上，截至2007年底中国艾滋病病毒感染者和艾滋病患者的人数将达到70万等等。对于这些统计信息，多数人都会欣然接受，而很少会有人问：这些数据是怎么来的？这些数据可靠吗？要回答这些问题，我们就要掌握一些统计学（statistics）知识，运用统计学的基本原理和方法来辨别真伪、去粗取精，正确认识客观事物的规律性。

统计学的基本原理和方法应用于不同的学科领域，产生了不同的统计学分支。如，在生物、卫生、医学等领域的应用，就产生了生物统计学（biostatistics）、卫生统计学（health statistics）、医学统计学（medical statistics）。一般认为，生物统计学应用于生物学研究领域，医学统计学和卫生统计学均应用于医学领域，前者侧重应用于临床医学，而后者侧重应用于预防医学和公共卫生。但三者存在诸多的联系和交叉，难以截然分开。卫生统计学是应用概率论和数理统计学的基本原理和方法，研究居民卫生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，是卫生及其相关领域研究中不可缺少的分析问题和解决问题的重要工具。

【案例1-1】 某研究者探讨银屑病的发病与血型的关系，对64例银屑病患者的血型进行观察，结果发现O型30例，占46.88%，居首位；A型和B型均为17例，各占26.56%；AB型0例，居末。由此，研究者认为银屑病的发病与血型有明显关系，O型血的人最容易患银屑病。

【问题1-1】

（1）该研究存在什么缺陷？

（2）研究结果是否可靠？

【分析】 正常人群中血型构成本身就存在较大差异，O型所占比例较高，AB型所占比例最少，这是一般的医学常识。银屑病患者的血型分布也存在这种差异，按一般的逻辑推理，研究者认为“银屑病的发病与血型有明显关系，O型血的人最容易患银屑病”的结论是站不住脚的。从统计学角度看，其缺陷有二：其一，没有设置对照组。有比较才有鉴别，应设置正常人群作为对照组。其二，没有统计分析，仅根据数据的大小直接下结论。本研究属于抽样研究，存在抽样误差，64例银屑病患者仅是一个样本，并不一定能代表所有的银屑病患者，需要采用适当的方法进行统计分析，然后根据分析结果下结论。所以，该研究结果是不可靠的。

【案例1-2】 某研究者欲研究其所在地区居民对实施家庭病床的认同态度，拟从所有的居民小区中随机抽取三个小区的住户为样本，以户为单位进行入户调查。调查员在小区门口对出入的居民进行了调查，然后对所得数据进行统计分析，结果发现“该地区居民认为不需要设置家庭病床”。

【问题1-2】

（1）该研究存在什么缺陷？

（2）研究结果是否可靠？

【分析】 家庭病床主要是为长期患病的或不能自由活动的老人设置的，老人及与老人关系密切的家庭成员（当老人不能回答调查提问时，可由家庭成员代为回答）应为调查的主要目标人群，而经常出入小区的居民大部分都是年龄相对较轻的人群或是能够自由活动的老人。因此，研究者仅对这部分居民进行调查，所得到的数据是有偏差的。此外，可能还会出现对同一家庭中的成员进行多次调查的情况，这样也会造成偏差，因为同一家庭中的成员对某些问题的态度会有某种相似性。

笔记栏

该研究者对调查的主要目标人群不明确，且收集资料所采取的方法欠妥当(应入户调查，而研究者只是在小区门口对出入的居民进行调查)，导致调查得到的数据不准确，不能较好地反映真实情况，所以结果是不可靠的。

【案例 1-3】 某疾病控制中心开展了一项研究，以了解当地肺癌的患病情况，从 10 万人口中随机抽取 2000 人进行调查，调查内容包括流行病学资料和临床实验室检查资料。其中男性 1100 人，患肺癌者 6 人；女性 900 人，患肺癌者 3 人。由此，研究者计算得出，男性肺癌发病率为 0.55%，女性肺癌发病率为 0.33%，并认为男性肺癌的发病率高于女性。

【问题 1-3】

(1) 该研究者所选择的统计指标正确吗？为什么？应该选择何种指标？

(2) 该研究者认为“男性肺癌的发病率高于女性”的结论是否可靠？

【分析】

(1) 该研究者所选择的统计指标是不正确的。因为现况调查只能得到现在有多少人患病，而不知道哪些人是新近发病和哪些人是以前发病的，一些慢性非传染性疾病的发病时间更难确定，所以无法计算发病率，而只能计算患病率。因此，研究者应选用患病率这一指标。

(2) 该研究者认为“男性肺癌的发病率高于女性”的结论不可靠。本研究属于抽样研究，抽取的 2000 人仅为一个随机样本，存在抽样误差，需要采用适当的方法进行统计分析，推断男性和女性肺癌发病率的差别是因为抽样误差引起的，还是因为男性肺癌的发病率本身就高于女性，然后再下结论，而不能仅根据数据的结果直接下结论。

类似的实例还有很多。从这些实例中，我们可以看出，研究设计、数据收集和统计分析的任一环节存在缺陷都有可能导致整个研究的失败。但这一点并非从一开始就为人们所共识，而是在断送了许多前沿的医学研究成果甚至付出了生命的代价后才逐步被人们认识到的。英国著名统计学家 F. Yates 和 M. J. R. Healy 就曾说过：“非常痛心地看到，因为数据分析的缺陷和错误，那么多好的生物研究工作面临着被葬送的危险”。可见，卫生统计学在卫生及其相关领域研究中的地位是举足轻重的，是卫生工作者从事科学研究必须掌握的一门基本技能。

【知识点 1-1】

卫生统计学是应用概率论和数理统计学的基本原理和方法，研究居民卫生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，是卫生及其相关领域研究中不可缺少的分析问题和解决问题的重要工具。

第二节 卫生统计学的主要内容和基本步骤

一、卫生统计学的主要内容

卫生统计学的主要内容包括以下几个方面：

(一) 统计设计

统计设计是卫生统计学的重要内容，包括资料收集、整理和分析全过程总的设想和安排。

(二) 统计分析

1. 统计描述 定量资料和定性资料的统计描述，统计表和统计图。

2. 统计推断 主要包括参数估计和假设检验。常用假设检验方法有 t 检验、z 检验、方差分析、χ^2 检验、秩和检验、双变量关联性分析和直线回归分析、对数秩检验、多重线性回归、logistic 回归、Cox 比例风险回归及 Meta 分析等。

(三) 其他内容

常用医学人口统计、生育统计、疾病统计与死亡统计指标，寿命表的编制及其应用

（四）常用统计分析软件简介

本书主要介绍 SAS 和 SPSS 两种权威统计分析软件。

二、统计工作的基本步骤

统计工作可分为以下四个基本步骤：

（一）设计（design）

设计是统计工作的第一步，也是最关键的一步，关系到整个研究的成败。一般包括专业设计和统计设计，本书重点介绍统计设计。如果统计设计存在缺陷，任何高深的统计方法都于事无补，所进行的统计分析只是数字游戏而已，所得出的结论也是不可靠的。但许多研究人员并不重视设计，等到数据收集完成后再去咨询或求助于统计专业人员，不过此时为时已晚。英国著名统计学家与遗传学家、现代统计学的奠基人之一费希尔（R. A. Fisher，1890～1962 年）曾精辟地指出："做完实验后才找统计学家无异于请他做尸体解剖，他能做的全部事情就是告诉你这实验死于什么原因。"因此，在研究之前一定要查阅大量文献、必要的时候咨询统计学专家，做好周密的设计。本书中统计设计包括调查设计和实验研究设计，详见第 2 章和第 3 章。

（二）收集资料（collection of data）

设计完成后，研究进入实施阶段，需要收集准确可靠的原始数据。资料的来源是多方面的，大致可分为以下几类：

1. 统计报表 如国家法定的有关卫生工作报表、传染病报表、职业病报表、医院工作报表等。这些报表是由国家统一设计，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，报表要求做到完整、准确、及时。

2. 日常工作记录 如医院的病历、经常性的卫生监测记录、健康检查记录等。

3. 专题调查或实验 是指针对某个专题做的调查或实验研究所收集的资料。

（三）整理资料（sorting data）

收集的资料通常是杂乱无章的，需要进行清理，使其系统化和条理化，便于进一步的统计分析，这便是整理资料的过程，包括数据的录入、核查和汇总，一般应用计算机软件来完成。在输入计算机前，需要对数据进行编码，如用"1"代表男性，"2"代表女性或用"M"代表男性，"F"代表女性等。

（四）分析资料（analysis of data）

分析资料是根据研究目的计算有关指标描述数据的基本特征，选择适当统计方法对资料进行分析，阐明事物的内在联系和规律的过程。统计分析包括：

1. 统计描述（descriptive statistics） 是指选用统计指标、统计表或统计图等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） 是指选择恰当的统计方法由已知的样本信息推断总体的特征，包括参数估计和假设检验。

值得注意的是，虽然我们把统计工作人为地分为设计、收集资料、整理资料和分析资料四个步骤，但是它们之间并不是孤立的，而是紧密联系、不可分割的一个整体。任何一项研究，如果缺少其中的任何一步，都可能会影响到整个研究结果。

【知识点 1-2】

卫生统计学的主要内容包括：①统计设计；②统计分析；③常用医学人口统计、生育统计、疾病统计与死亡统计指标，寿命表的编制及其应用；④常用统计分析软件简介等几个部分。

统计工作可分为设计、收集资料、整理资料和分析资料四个基本步骤。

笔 记 栏

第三节 卫生统计学的几个基本概念

1. 同质与变异 俗语说“物以类聚，人以群分”，那么“类”或“群”是用什么标准来划分呢？当然是一些本质特征或属性。在统计学中，若某些观察对象具有相同的特征或属性，我们就称之为同质(homogeneity)，或具有同质性。如研究某地区5岁男童的生长发育情况，那么“该地区、男性、5岁”就是这些观察对象共同具有的特征或属性，每个男童我们称之为同质的个体。同质的个体之间是不是所有特征或属性的观察值都相同呢？显然不是，如该地区5岁男童的身高、体重、血压、肺活量等特征或属性的观察值不尽相同。我们将同质个体的某项特征或属性的观察值或测量值之间的差异称为变异(variation)。统计学的任务就是在同质的基础上对变异进行研究，从而揭示事物内在的规律性。

2. 总体与样本 总体(population)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特征或属性的观察值或测量值的集合，如研究某地2007年正常成年男子的脉搏数，则该地2007年所有的正常成年男子的脉搏数就构成了一个总体。该总体明确了特定的时间和空间范围且包含有限个观察单位，称为有限总体(finite population)。若总体没有特定的时间和空间范围的限制，且所包含的观察单位个数是无限的或几乎是不可能准确计数的，称该总体为无限总体(infinite population)，如研究某新药治疗高血压病的疗效，总体包含了接受该新药治疗的所有高血压病患者，没有时间和空间范围的限制，且观察单位个数几乎是不可确定的，因而是无限总体。

医学研究中的总体大多是无限总体，要直接观察总体的情况几乎是不可能的。即使对于有限总体来说，若包含的观察单位过多，对每个个体进行观察，一方面需要花费大量的人力、物力和财力；另一方面，这种观察有时也是不可能实现的，如检验一批鸡蛋的坏蛋率，不可能将所有的鸡蛋都一一打破。因此，经常是从总体中抽取样本(sample)，用样本信息来推断总体特征。样本是从总体中随机抽取的具有代表性的部分观察单位的集合。如上例，可从该地2007年正常成年男子中，随机抽取300人组成样本。样本中包含的观察单位个数称为样本含量(sample size)。

3. 参数与统计量 反映总体特征的指标称为参数(parameter)。参数一般是未知的，常用希腊字母表示，如总体均数μ、总体率π等。根据样本观察值计算出来的指标称为统计量(statistic)，统计量常用拉丁字母表示，如样本均数$\bar{x}$、样本率p等。如果样本对总体具有较好的代表性，那么样本的某项观察指标的统计量就与总体相应指标的参数较为接近。因此，可以把样本的统计量作为总体参数的估计值，如我国人群糖尿病的患病率为4%，由随机抽取的5万人样本计算出来的糖尿病患病率为3.9%，可认为该样本对总体具有较好的代表性。

4. 变量与资料 确定总体之后，研究者需要对每个观察单位的某项特征或属性进行观察或测量，这种特征或属性称为变量(variable)。变量的观察值或测量值称为变量值(value of variable)或观察值(observed value)。变量值的集合称为资料(data)，例如，抽样调查某年某地区50岁及以上人群的年龄、性别、身高、体重、血型等一般情况，年龄、性别、身高、体重、血型就是变量，其测量值的大小就是变量值，这些测量值的集合就组成了资料。根据变量值是定量的还是定性的，资料可分为以下两大类。

(1) 定量资料(quantitative data)：亦称计量资料，其变量值是定量的，表现为数值大小，一般有度、量、衡单位，如上例中每个对象的身高(cm)、体重(kg)资料等均为定量资料。

(2) 定性资料(qualitative data)：亦称分类资料(categorical data)，其观察值是定性的，表现为互不相容的类别或属性，一般无度、量、衡单位。可进一步细分为以下两种资料：

1) 计数资料(count data)：是指将观察单位按某种类别或属性进行分组，清点各组观察单位数所得的资料。它包括两种类型：①二项分类资料，是指观察单位的某种特征或属性表现为“互不相容的两个类别”的资料，如上例中的性别资料(每个观察单位的取值为“男”或“女”)，临床化验结果(“阳性”或“阴性”)，疾病统计资料(“发病”或“未发病”，“患病”或“未患病”)等。②无序多项分类资料，是指观察单位的某种特征或属性表现为“互不相容的多个类别”的资料，如上例中的血型资料，每个观察单位的结果为“O型”、“A型”、“B型”或“AB型”中的一种。

2) 等级资料(ordinal data)：亦称有序多分类资料，是将观察单位按某种特征或属性的程度或等级顺序分组，清点各组观察单位数所得的资料。各属性之间互不相容且有程度的差别，给人以“半定量”的感觉，如研究某药治疗糖尿病的疗效，以每个患者为观察单位，结果可分为“治愈、显效、好转、

笔记栏

无效”四个有顺序的等级，这类资料就属于等级资料。

当然，资料类型的划分是相对的，它们之间可以相互转化。定量资料可以转化为定性资料（计数资料或等级资料），定性资料也可以数量化，如，健康调查简表 SF-36 中把健康状况分为“非常好、较好、一般、差、非常差”五个等级，应划归为等级资料。但若将这五个等级数量化，分别将它们赋值为5、4、3、2、1，就可按定量资料处理。资料类型不同所采用的统计方法亦不同，这在后面的有关章节中还会强调。

5. 抽样研究与抽样误差 从总体中随机抽取样本，通过样本信息推断总体特征的研究方法称为抽样研究（sampling research）。由随机抽样造成的样本统计量与总体参数之间、样本统计量之间的差异称为抽样误差（sampling error）。产生抽样误差的根源在于个体变异，由于个体变异是普遍存在的，因此在抽样研究中抽样误差是不可避免的，但它具有一定的规律性，可以用统计方法估计其大小。

6. 概率 概率（probability）是随机事件发生可能性大小的数值度量。例如，在每场足球比赛开始之前，我们都可以看到裁判掷硬币的一幕，由猜中的队决定己方上半场比赛的进攻方向，其目的是确保公平，因为我们相信硬币出现正面或反面的概率都是 50%（或 0.5）。但事实上能确保公平吗？法国自然主义者布方伯爵（Count Buffon）曾掷铜板 4040 次，结果 2048 次出现正面，即出现正面的概率为 0.5069；南非数学家柯瑞屈（John Kerrich）在二战期间被关在德国集中营的时候掷铜板 10000 次，结果 5067 次出现正面，出现正面的概率为 0.5067；英国统计学家皮尔逊（Karl Pearson）掷铜板 24000 次，结果 12012 次出现正面，即出现正面的概率为 0.5005。从这几位科学家的实验我们可以看到，随着抛掷次数的增加铜板出现正面的比例越来越接近 0.5。假设无限次地抛掷下去，出现正面的概率最终会是 0.5。因此，概率大小的估计是要以足够大的样本含量为前提，换言之，当某实验在相同条件下独立地重复无数次时，某事件发生次数的比例才是该事件发生的概率。

概率通常用 P 表示，其大小介于 0 与 1 之间，即 $0 \leqslant P \leqslant 1$。$P$ 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；$P=1$，表示某事件为必然事件，即一定要发生的事件；$P=0$，表示某事件为不可能事件，即一定不可能发生的事件。当某事件发生的概率 $P \leqslant 0.05$ 时，统计学中习惯上称该事件为小概率事件（small probability event），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。

【知识点 1-3】

1. 总体是根据研究目的确定的同质观察单位的全体。样本是从总体中随机抽取的具有代表性的部分观察单位的集合。

2. 医学研究中的总体大多是无限总体，要直接观察总体的情况几乎是不可能的。即使对于有限总体来说，若包含的观察单位过多，对每个个体进行一一观察，一方面需要花费大量的人力、物力和财力；另一方面，这种观察有时也是不可能实现的。因此，经常是从总体中抽取样本，用样本信息来推断总体特征，即进行抽样研究。抽样研究中抽样误差是不可避免的。

3. 根据变量值是定量还是定性，可将资料分为定量资料和定性资料两大类。资料类型的划分是相对的，它们之间可以相互转化。资料类型不同所采用的统计方法亦不同。

4. 概率是随机事件发生可能性大小的数值度量，是统计学中一个很重要的基本概念。

第四节 学习卫生统计学应注意的问题

本门课程的教学目的是，培养学生的统计学思维，为学生学习其他专业课程打下必要的统计学基础，提高学生应用统计学分析和解决实际问题的能力。为此，学习本门课程时，应注意以下几个问题：

1. 重点应放在卫生统计学基本概念和基本原理的理解和掌握上 对于任何一门学科来说，其基本概念和基本原理都是整个学科体系的基石，统计学当然也不例外。只有深刻理解和掌握这些基本概念和基本原理，才能举一反三，运用这些原理和方法解决卫生实践中的实际问题。

2. 重点应放在基本统计方法的适用条件、用途及注意事项的理解和掌握上 对于一般的卫生工作者而言，并不需要掌握太多高深的统计学方法，更不必深究统计公式的推导过程和死记硬背公式，重点要放在对一些基本统计方法的适用条件、用途及注意事项的理解和掌握上。换言之，应掌握

一些基本统计方法在资料具备什么条件下可用、用来解决什么问题、使用时应注意什么问题等。

3. 重点应放在运用卫生统计学知识解决实际问题能力的培养上 我们不是为了学习而学习，而是为了通过学习卫生统计学的知识来提高解决卫生及其相关领域中实际问题的能力，这是我们学习这门课程的出发点和终极目标。如果学了一学期的卫生统计学，相关的概念和理论背得滚瓜烂熟，但遇到实际问题就束手无策，那么教与学都应该是失败的。所以，学习时要结合书中的案例，认真思考，做到“知其然更知其所以然”，学以致用。

在学习本课程之前，同学们可能会听到诸如“这门课程是如何如何的难，统计公式是‘头上长角身上长刺’，统计理论是‘云山雾罩不知所云’”等。卫生统计学对于已习惯了形象思维和死记硬背的医学生来说，的确是有一定的难度，但并不是想象的那么“高处不胜寒”。只要同学们克服畏难情绪，注意以上几个方面的问题，采取正确的学习方法，相信一定能够学好这门课程。最后用“人之为学有难易乎？学之，则难者亦易矣；不学，则易者亦难矣”这句名言与大家共勉，希望同学们学习这门课程时有一个轻松愉快的体验。

思考与练习题

一、选择题（从 a～e 中选出一个最佳答案）

1. 观察单位为研究中的 （ ）
   a. 样本　b. 研究对象　c. 全体　d. 影响因素　e. 个体
2. 总体是由 （ ）
   a. 个体组成　b. 研究对象组成　c. 同质个体组成
   d. 研究指标组成　e. 样本组成
3. 统计学中的样本是指 （ ）
   a. 随意抽取的总体中任意部分
   b. 有意识的选择总体中的典型部分
   c. 依照研究者要求选取总体中有意义的一部分
   d. 依照随机原则抽取总体中有代表性的部分
   e. 以上说法均不正确
4. 抽样研究的目的是 （ ）
   a. 研究样本统计量　b. 由样本统计量推断总体参数
   c. 研究典型案例　d. 研究误差　e. 研究总体统计量
5. 参数是指 （ ）
   a. 参与个体数　b. 总体的统计指标　c. 样本的统计指标
   d. 样本的总和　e. 总体中的观察单位数

二、思考题

1. 统计资料可分哪几种类型？举例说明不同类型资料之间是如何转换的？
2. 统计工作分为哪几个步骤？
3. 举例说明小概率事件的含义。

（丁元林）

第2章 调查研究设计

第一节 调查研究的特点和类型

【案例 2-1】 苏州大学公共卫生学院卫生统计学教研室于 2006～2008 年主持了国家自然科学基金项目“敏感问题的调查设计研究”(项目编号:3057162)。该项目从 2007 年的每个月中抽取 1 周,于抽中周的每日 9 点至次日 3 点,调查上海市区随机抽取的 10 个男同性恋活动场所的全部对象(各对象该周内只调查一次),调查内容包括同性恋者的年龄、学历、职业、经济状况、户籍地、性行为的方式、频繁程度、安全套使用率、每月更换性对象的人数、性病知识的知晓程度、是否从事同性性服务及收费情况等,并对调查对象免费检测艾滋病病毒。

【问题 2-1】

(1) 案例 2-1 中采用的是什么研究方法?

(2) 案例 2-1 中采用的研究方法有何特点?

(3) 案例 2-1 中采用了何种抽样方法?

【分析】

(1) 案例 2-1 采用的是调查研究。

(2) 调查研究的特点是:不能对调查对象人为施加干预措施,不能将调查对象随机分组,很难控制干扰因素,一般不可下因果结论。

(3) 案例 2-1 项目中,首先在上海市区全部男同性恋活动场所随机抽取 10 个场所;再在随机抽取的调查时间里调查 10 个场所的全部对象。调查对象分布在场所-时间两维空间里,属于从场所-时间两维空间里整群随机抽样。

一、调查研究的特点

医学研究方法主要有调查研究(survey research)、实验研究(详见第 3 章)和文献研究(如 Meta 分析,见第 16 章)。调查研究又称为观察性研究,具有以下特点:

1. 不能人为施加干预措施 调查研究不能对调查对象人为施加干预措施(处理因素)。如案例 2-1 中,研究者不可能人为给男同性恋调查对象施加某种性行为方式。

2. 不能随机分组 调查研究不能将调查对象随机分组。如案例 2-1 中,不可能将 10 个男同性恋活动场所的同性恋者随机分配到各年龄、学历、职业等中去,只能在调查过程中客观地记录下每个人的年龄、学历、职业等。

3. 很难控制干扰因素 调查研究很难控制干扰因素。如案例 2-1 中,不同年龄、学历、职业等男同性恋者每月更换性对象的平均人数,受各同性恋者的生理及心理健康状况的影响,而这些干扰因素是很难控制一致的。

4. 一般不能下因果结论 调查研究一般不能下因果结论。如案例 2-1 的调查结果中,在调整了月份、年龄、学历、经济状况等因素的影响后,男同性恋活动场所在校大学生组每月性对象更换的平均人数比其他职业组高,虽然有统计学意义,一般也不能说明在读大学就是性对象更换频繁程度高的原因,因为在调查中无法控制调查对象的生理、心理、家庭管束等其他因素的影响。

二、调查研究的类型

调查研究的类型有不同的划分方法。根据调查的抽样比例可划分为全面调查及抽样调查;根据

笔记栏

调查时间可划分为横断面(现况)调查、病例对照研究、队列研究及回顾性队列研究;根据调查的抽样概率可划分为概率抽样调查及非概率抽样调查。

(一) 根据调查的抽样比例划分

1. 全面调查(complete survey) 即对研究总体全部(抽样比例为100%)进行调查,如某病患病率普查、全国人口普查等。全面调查的优点是能得到总体的参数,不存在抽样误差。但由于总体数量庞大,操作时可能会引入一些非抽样误差,且消耗较多的人力、物力和财力。

2. 抽样调查(sampling survey) 即从总体中抽取一定数量的观察单位组成样本,然后用样本资料的信息对总体进行研究。抽样调查与全面调查相比,因观察例数较少,可节省人力、物力和时间,并可获得较为深入细致和准确的资料,大大减少了系统误差和过失误差产生的机会,往往可达到事半功倍的效果,值得大力推广应用。

抽样调查又分为概率抽样调查和非概率抽样调查。所谓概率抽样调查,即在抽样过程中必须保证总体中的每个观察单位都有同等的概率被抽到样本中来,然后根据样本信息来推断总体特征。基本的概率抽样方法有单纯随机抽样、系统抽样、分层抽样、整群抽样。所谓非概率抽样调查,即当总体不明、为特殊调查目的或无法进行概率抽样时,常常采用典型调查(选择个别典型的人和事物在深度方面进行详细的调查)、方便抽样(最为便利的方式抽样)、雪球抽样(通过样本对象介绍样本对象)、配额抽样(按相同抽样比例在总体各类中方便抽样或雪球抽样)等方法获取样本,此时每个个体被抽中的概率是未知的或无法计算的。非概率抽样调查不能进行统计推断,社会医学定性调查常采用非概率抽样方法。

(二) 根据时间划分

1. 横断面调查(cross-sectional survey) 又称现况调查,调查总体某时间断面上的情况,通常是针对一个人群的描述性调查,目的是了解该人群中疾病或卫生事件的现状以及相关各种因素的分布情况。

2. 病例对照研究(case-control study) 是以患所研究疾病的为病例,以未患该病的合适对象为对照,分别调查其既往暴露于某个(或某些)危险因子的情况及程度,以判断暴露危险因子与某病有无关联及其关联程度大小的一种观察性研究方法。因为这种调查始于疾病发生之后,是由结果到原因的顺序,所以也称为"回顾性调查"。例如,调查某精神疾病患者与正常对照组过去一年内所遭遇的生活事件即为一病例对照研究。

3. 队列研究(cohort study) 是选定暴露及未暴露于某个(或某些)因素的两组人群,追踪其各自的结局,并比较两组人群发病的差异,从而判定暴露因子与发病有无因果关联及关联大小的一种观察性研究方法。该类研究设计的特点是由原因到结果的顺序,也称为"前瞻性研究",例如调查某不良生活方式对人群健康的影响可采用队列研究。

4. 回顾性队列研究 是回顾性地收集人群既往暴露于危险因子的情况及程度,再随访不同暴露人群的疾病发生情况,是将病例对照研究与队列研究相结合的一种研究方法。

【知识点 2-1】

1. 调查研究通常是在对研究事物或现象不太了解或在研究的初始阶段进行的,只是客观地观察和记录调查对象的真实情况,不能人为施加干预措施,也不能将调查对象随机分组,所以,很难控制干扰因素,一般不可下因果结论。

2. 调查研究的类型 根据抽样比例可划分为全面调查及抽样调查;根据调查时间可划分为横断面(现况)调查、病例对照研究、队列研究及回顾性队列研究;根据抽样概率可划分为概率抽样调查及非概率抽样调查。

第二节 常用抽样方法

【案例 2-2】 苏州大学公共卫生学院卫生统计学教研室于2005～2007年主持了国家社会科学基金项目"连续调查的抽样设计研究"(项目编号:04BTJ001)。该项目分别采用几种基本的常用抽样方法,随机抽取苏州大学新校区的部分本科生与硕士研究生,于2006年9月～2007年7月连续3次

笔记栏

（前后两次调查时间间隔100天），调查他们的身体健康状况、学习成绩、考试作弊、生活消费、婚前性行为等指标。

【问题2-2】

（1）案例2-2采用了哪几种抽样方法？

（2）基本的常用抽样方法的概念、特点是什么？

（3）对各种基本的常用抽样方法，如何具体操作？

【分析】

（1）案例2-2中的研究项目分别采用了单纯随机抽样、系统抽样、分层抽样、整群抽样四种基本的常用抽样方法及分层多阶段抽样。

（2）四种基本的常用抽样方法的概念、特点分别见本节“一、二、三、四”的第“（一）、（三）”项内容。

（3）四种基本的常用抽样方法的具体操作分别见本节“一、二、三、四”的第“（二）”项内容。

一、单纯随机抽样

（一）概念

单纯随机抽样（simple random sampling）是先将调查总体的全部观察单位统一编号，然后采用随机数字表、统计软件或抽签等方法之一随机抽取n（样本大小）个编号，由这n个编号所对应的n个观察单位构成研究样本。

（二）操作

在案例2-2的项目中，苏州大学新校区共有硕士研究生5200名，采用单纯随机抽样从中抽取800名调查其婚前性行为。采用常用的随机数字表法，具体操作如下：

1. 统一编号　将5200名研究生统一编号：1、2、…、5200。

2. 确定随机数　从附表1随机数字表中任一行、任一列开始向任何方向抄录略多余800个4位数（因最大编号为4位数），例如从第36行第1列开始向右抄录：0526、9370、6022、3585、1513、9203、…。

3. 确定样本　将重复数字、首数大于6的数值弃用，得到符合要求的800个数：0526、3585、1513、…，然后将与这800个编号对应的学生抽出即可。在此需指出：不能按一般卫生统计学书中介绍的“首数≥6000减6000，≥8000减8000形成新数字”的做法，因这样做的结果首数为6、7、8、9的四位数分别变为首数为0、1、0、1的四位数，致使首数为0、1编号的学生被抽中的概率将是其他学生的3倍。

（三）特点

单纯随机抽样是最基本的抽样方法，也是其他抽样方法的基础。该方法的优点是操作简单，统计量的计算较简便；缺点是当总体观察单位数量庞大时，给观察单位逐一编号甚为繁复，有时难以做到。

二、系统抽样

（一）概念

系统抽样（systematic sampling），又称机械抽样或等距抽样。事先将总体内全部观察单位按某一顺序号等距分隔成n（样本大小）个部分，每一部分内含m个观察单位；然后从第一部分开始，从中随机抽出第i号观察单位，依此用相等间隔m机械地在第2部分、第3部分直至第n部分内各抽出一个观察单位组成样本。

笔记栏