



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辅导教材

Shiyong Weisheng

Tongji Xue

实用卫生统计学

XUEXI ZHIDAO
学习指导

■ 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辅导教材

实用卫生统计学学习指导

主编：康晓平

编写人员（按姓氏笔画排序）

李 凯 赵新胜
何平平 康晓平
易伟宁

北京大学医学出版社

SHIYONG WEISHENG TONGJIXUE XUEXIZHIDAO

图书在版编目(CIP)数据

实用卫生统计学学习指导/康晓平主编. —北京:北京大学医学出版社, 2002. 11

ISBN 7-81071-254-3

I. 实… II. 康… III. 卫生学:统计学—医学院校—教学参考资料 IV. R195.1

中国版本图书馆 CIP 数据核字(2002)第 085106 号

北京大学医学出版社出版发行

(100083 北京海淀区学院路 38 号 北京大学医学部院内)

责任编辑:暴海燕

责任校对:李月英

责任印制:郭桂兰

怀柔师范学校印刷厂印刷 新华书店经销

开本 787mm×1092mm1/16 印张:9.5 字数:240 千字

2002 年 11 月第 1 版 2002 年 11 月第 1 次印刷 印数:1-8000 册

定价:14.50 元

版权所有 不得翻印

编写说明

为配合《实用卫生统计学》主教材的学习,我们编写了这本《学习指导》。书中主要包括三个内容:

1. 重点难点。根据教学要求,我们将主教材中各章节的重点和难于理解的内容进行了提炼和归纳,目的是帮助读者将抽象的概念具体化,复杂的计算简单化,作出比较条理和规律的总结。

2. 练习题和参考答案。围绕主教材的内容,我们从不同角度对统计的基本概念、基本计算和分析方法编写了名词解释、简答、填空、选择、是非、计算分析等多种题型供读者练习。涉及到统计基本概念和方法较多的章节(如第一至第九章),其选择题和是非题的部分参考答案中,不仅给出答案,还给出答案的解释,目的是使读者加深对统计基本概念的理解,较好地应用基本统计指标和统计分析方法。

3. SPSS10.0 统计软件的应用。为满足部分读者使用计算机处理数据的需要,本书选择《实用卫生统计学》部分章节的例题,用 SPSS10.0 统计软件进行了均数和比例数的假设检验,并对 SPSS 输出结果进行解释。本书不是专门介绍 SPSS 统计分析软件的,因此内容并不系统和全面。但对于那些已有 SPSS 基本操作基础,希望进一步了解 SPSS 在统计学中应用的人,可作为参考内容。

康晓平

2002 年 7 月 26 日

目 录

第一篇 卫生统计学基本概念和方法

第一章 绪 论	(2)	教学要求	(49)
教学要求	(2)	重点难点	(49)
重点难点	(2)	练习题	(53)
练习题	(5)	参考答案	(59)
参考答案	(7)	第七章 样本率(或构成比)比较的假设	
第二章 计量资料的统计描述	(10)	检验	(69)
教学要求	(10)	教学要求	(69)
重点难点	(10)	重点难点	(69)
练习题	(13)	练习题	(72)
参考答案	(16)	参考答案	(74)
第三章 计数资料的统计描述	(20)	第八章 秩和检验	(79)
教学要求	(20)	教学要求	(79)
重点难点	(20)	重点难点	(79)
练习题	(22)	练习题	(80)
参考答案	(25)	参考答案	(83)
第四章 统计表与统计图	(29)	第九章 直线相关与回归	(86)
教学要求	(29)	教学要求	(86)
重点难点	(29)	重点难点	(86)
练习题	(30)	练习题	(89)
参考答案	(34)	参考答案	(91)
第五章 单个样本数据的参数估计	(39)	第十章 常用综合评价方法	(97)
教学要求	(39)	教学要求	(97)
重点难点	(39)	重点难点	(97)
练习题	(42)	练习题	(100)
参考答案	(44)	参考答案	(101)
第六章 样本均数比较的假设检验	(49)		

第二篇 卫生业务与居民健康统计

第十一章 疾病统计	(104)	教学要求	(112)
教学要求	(104)	重点难点	(112)
重点难点	(104)	练习题	(115)
练习题	(107)	参考答案	(116)
参考答案	(108)	第十三章 医学人口统计	(118)
第十二章 卫生服务调查统计	(112)	教学要求	(118)

重点难点	(118)	练习题	(126)
练习题	(121)	参考答案	(128)
参考答案	(122)	附录:SPSS for Windows 与统计应用	
第十四章 寿命表	(125)		(130)
教学要求	(125)	参考书目	(145)
重点难点	(125)		

第一篇

卫生统计学基本概念和方法

第一章 绪 论

【教学要求】

了解:知道卫生统计工作的基本步骤。

熟悉:总体与总体研究、变量与变量值、同质与变异、样本与随机抽样、参数与统计量、统计描述与统计推断、抽样研究与小概率事件。

掌握:统计资料的类型和误差的种类。

【重点难点】

第一节 卫生统计学的任务及其内容

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,是按照设计方案去收集、整理、分析数据,并对数据结果进行解释,从而做出比较正确的结论。

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;②健康统计;③卫生服务统计。《实用卫生统计学》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)。

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,即个体间存在变异。例如,同样规模的医院,他们的病人住院天数不同、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。如果按科学管理作决策,就应该先做调查研究收集有关数据,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。

第三节 统计资料的类型

卫生统计资料一般划分为三大类,不同类型的资料对应不同的变量类型,应选用不同的统计指标和统计分析方法。

表 1.1 统计资料的类型

资料类型	变量类型	例 子	统计指标	统计分析方法
计量资料	数值变量	12 岁男孩身高(cm): 142.3 134.2 145.2 151.2	平均数,标准差等	t 检验, u 检验,方差分析, 直线相关与回归
计数资料	分类变量	某药治疗 70 个病人的效果: 有效:43 例 无效:27 例	率,构成比等	χ^2 检验, u 检验
等级资料	有序分类变量	某药治疗 70 个病人的效果: 显效:13 例 好转:30 例 无效:27 例	可计算构成比	秩和检验

第四节 统计学基本概念

一、总体和总体研究

概念:总体是根据研究目的确定同质的所有观察单位某种变量值的集合。注意,此处强调的是变量值的集合,而不是观察单位的集合。总体研究是对有限总体中的每个个体都作观察。

举例:调查某地 2000 年正常成年男子的红细胞数。该地 2000 年全部正常成年男子的红细胞数就构成一个总体。它的同质基础是同一地区、同一年份、同为正常成人、同为男性。它的所有观察单位某种变量值的集合是指全部正常男子的红细胞数,而不是所有正常男子。若对该地 2000 年每一名正常成年男子的红细胞数都作测量就是总体研究。

二、变量与变量值

概念: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)。

举例:调查某地 2000 年正常成年男子的红细胞数,红细胞就是变量,每一个正常成年男子的红细胞数就是一个测量值或变量值。

三、同质与变异

概念: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,其观察值(变量值)之间的差异,统计上称为变异。

举例:某年从某地随机抽取 144 名正常成年男子分别测其红细胞数。每个正常成年男子的红细胞数不尽相同,存在差异。这种红细胞数间的差异就是变异。而这 144 人的同质基础是来自同一年,同一个地区的所有正常成年男子。

四、样本与随机抽样

概念:总体中随机抽取有代表性的一部分个体,其实测值的集合称为样本。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。

举例:调查某地 2000 年正常成年男子的红细胞数。从该地区用系统抽样的方法随机抽取 144 名正常成年男子,分别测其红细胞数。这 144 名正常成年男子的红细胞数就是样本。系

统抽样就是随机抽样方法中的一种。

五、抽样研究

概念:从所研究的总体中随机抽取一部分有代表性的样本进行研究,用样本指标推论总体,最终达到了解总体的目的。这种用样本指标推论总体参数的方法称为抽样研究。

举例:欲了解某地 2000 年正常成年男子的平均红细胞数。该地 2000 年全部正常男子的红细胞数是一个总体,但是我们不可能,而且也没有必要去测定每个正常成年男子的红细胞数。因此只能从总体中随机抽取 144 名正常成年男子的红细胞数作为样本,计算平均红细胞数($\bar{x}$)。如果这个样本均数具代表性,且可靠,即可用该样本的平均红细胞数($\bar{x}$)推论该地 2000 年正常成年男子的平均红细胞数(μ)。

六、参数和统计量

概念:参数是指总体指标,如总体均数(μ),总体标准差(σ),总体率(π)等。统计量是指样本指标,如样本均数($\bar{x}$),样本标准差(s),样本率(p)等。

举例:某地 2000 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。

七、统计描述与统计推断

概念:用统计图表或计算统计指标的方法表达一个特定群体的某种现象或特征,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: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。

举例:用某地 2000 年 144 名正常成年男子的红细胞数绘制直方图表示频数分布的类型,或计算红细胞数的平均数表示平均水平的方法即为统计描述;用 144 名正常成年男子的红细胞数的平均值去估计该地 2000 年所有正常成年男子的红细胞数平均值的方法为统计推断。

八、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。研究抽样误差的规律,是医学统计的重要内容之一。

表 1.2 医学科学研究中常见的三种误差

误 差	解 释	解决办法
系统误差	不是偶然机遇所致,而是某种必然因素所致,具有一定的倾向性	尽可能避免,或通过周密的研究设计解决
随机测量误差	偶然机遇所致,是不可避免的	控制在一定的允许范围内,数据仍可以使用
抽样误差	抽样研究中,即使消除了系统误差,控制了随机测量误差,样本指标和总体参数间仍会存在差别。这是由个体变异造成,是抽样波动所致,不可避免的	可以通过统计方法估计并减小

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。

表 1.3 统计工作的基本步骤

基本步骤	内 容
计划与设计	是开展医学研究的前提和依据。调查设计和实验设计均包括专业设计和统计学设计两个方面,主要内容有资料的收集、整理和分析
收集资料	概括为经常性资料 and 一时性资料两大类。收集的资料要求①完整、正确和及时;②要有足够的数量;③资料的代表性和可比性
整理资料	原始资料的检查与核对①数据的取值范围检错;②数据间的逻辑关系检错。以及资料的分组设计与归纳汇总①质量分组或数量分组;②编制频数分布表
分析资料	①统计描述:用一些统计指标,统计图表等方法对资料的数量特征和分布规律进行测定和描述,不涉及由样本推论总体问题。②统计推断:对样本统计指标作参数估计和假设检验,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。

【练习题】

一、名词解释

1. 变异
2. 抽样研究
3. 统计描述
4. 统计推断

二、简答题

1. 卫生事业管理专业与卫生统计学的关系是什么。
2. 总体和样本的关系是什么。
3. 收集资料时,对统计资料的要求是什么?

三、选择题

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医院的病床数,该市每个三级甲医院就是一个 ____。
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、每个病人就诊的科室

☒ B、每个病人就诊的次数

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. 用某年全市的出生数和婴儿死亡数计算婴儿死亡率,这种方法是属于_____。

A、抽样研究

☒ B、参数估计

☒ C、统计描述

D、统计推断

8. 调查某单位科研人员论文发表的情况,统计每人每年的论文发表数应属于_____。

A、计数资料

☒ B、计量资料

C、总体

D、个体

☒ 9. 医学人口统计应属于卫生统计学中的哪部分内容_____。

A、卫生统计学基本原理

☒ B、卫生统计学基本方法

C、健康统计

☒ D、卫生服务统计

☒ 10. 下面哪一种统计资料的来源不属于经常性的资料_____。

☒ A、统计报表

B、门诊病例

C、出生报告卡

☒ D、调查问卷

四、填空题

1. 计量资料是指_____,常用的统计指标有_____,_____,常用的统计方法有_____。

2. 计数资料是指_____,常用的统计指标有_____,_____,常用的统计方法有_____。

3. 在研究工作中随机抽样误差是不可避免的,系统误差是可以而且应该避免的,统计方法能处理的是抽样误差。

4. 一般情况下,参数是未知的并需要用统计量去估计。用统计量推论参数的方法,统计上称为参数估计。

5. 频率是对样本而言,概率是对总体而言。

6. 统计工作的四个步骤设计,收集资料,整理,分析。

7. 统计资料的两个主要来源_____。

8. 收集统计资料的三个基本要求_____。

9. 主要从哪两个方面对原始资料进行检查与核对_____。

10. 统计分析主要包括:描述,推断。

五、应用分析题

1. 某种疫苗通过皮下注射,对20名观察对象进行免疫,21天后观察结果,分别采用三种原始记录形式,结果如表1.4

表 1.4 用某疫苗对 20 名对象作皮下注射免疫结果

观察对象	抗体滴度	目测判断 抗体水平	免疫效果 分 类	观察对象	抗体滴度	目测判断 抗体水平	免疫效果 分 类
1	1:40	++	有效	11	1:80	+++	有效
2	1:20	+	无效	12	1:160	++++	有效
3	1:160	++++	有效	13	1:160	++++	有效
4	1:40	++	有效	14	1:80	+++	有效
5	1:320	+++++	有效	15	1:40	++	有效
6	1:80	+++	有效	16	1:40	++	有效
7	<1:20	±	无效	17	1:20	+	无效
8	<1:20	±	无效	18	1:80	+++	有效
9	1:40	++	有效	19	1:40	++	有效
10	1:40	++	有效	20	1:160	++++	有效

(1) 以上三种记录各属何种类型的统计资料?

(2) 怎样对这些资料进行合理的分组整理,进一步的统计分析方法是什么?

2. 对某医院的住院病历进行整理,请指出表 1.5 中的资料各属于何种变量。

表 1.5 某医院的住院病历资料

变量代码	变量名	变量的取值
x_1	入院情况	公费 1 劳保 2 自费 3
x_2	住院天数	天
x_3	手术	有 1 无 2
x_4	治疗效果	治愈 1 好转 2 未愈 3 未治 4 死亡 5
x_5	人出院诊断符合	符合 1 不符 2 待查 3

【参考答案】

一、名词解释

1. 同一性质的事物,其观察值(变量值)之间的差异,统计上称为变异。
2. 从所研究的总体中随机抽取一部分有代表性的样本进行研究,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。
3. 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。
4. 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。

二、简答题

1. 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。

2. 样本是从总体中随机抽取的一部分有代表性的个体组成,除了数量比总体少,其他构成均与总体一样,是总体具体而微的缩影。

3. 要做到①资料必须完整、正确和及时;②要有足够的数量;③注意资料的代表性和可比性

三、选择题

1. C 2. B 3. A 4. B 5. B 6. D 7. C 8. B
9. C 10. D

四、填空题

1. 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获得的连续型资料;平均数,标准差; t 检验, u 检验,直线相关与回归

2. 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获得的间断性资料;率,构成比; χ^2 检验, u 检验

3. 随机测量 系统 抽样

4. 参数 统计量 统计量 参数

5. 样本 总体

6. 计划与设计 收集资料 整理资料 分析资料

7. 经常性资料 一时性资料

8. 完整、正确和及时 要有足够的数量 资料的代表性和可比性

9. 数据的取值范围检错 数据间的逻辑关系检错

10. 统计描述 统计推断

五、应用分析题

1. ①表中给出的抗体滴度属于计量资料,可以进行数量分组绘制频数分布表,并可计算几何平均数和几何标准差。

20 名观察对象抗体滴度的频数分布

组段	划记	频数
<1 : 20		2
1 : 20		2
1 : 40		7
1 : 80		4
1 : 160		4
1 : 320		1
合计		20

②表中给出的目测判断抗体水平属于等级资料,可以进行质量分组清点各组人数,并可计算百分构成比指标。

20 名观察对象抗体滴度的目测判断抗体水平

目测判断抗体水平	例数	构成比(%)
±	2	
+	2	
++	7	
+++	4	
++++	5	
合计	20	

③表中给出的免疫效果分类属于计数资料,可以进行质量分组清点各组人数,并可计算百分构成比指标。

20 名观察对象抗体滴度的免疫效果分类

免疫效果分类	例数	构成比(%)
有效	16	
无效	4	
合计	20	

2. 表 1.5 中的变量 x_1 、 x_5 为无序分类变量, x_2 为数值变量, x_3 为二分类变量, x_4 为有序分类变量。

(康晓平)

第二章 计量资料的统计描述

【教学要求】

了解:频数表的编制及资料的频数分布特征,了解正态分布及标准正态分布的概念。

熟悉:标准正态变换。

掌握:各种平均数指标、离散指标的适用条件及其直接法的计算,尤其是均数和标准差,掌握标准正态分布曲线下面积的分布规律,掌握正态分布资料 95%和 99%的个体观察值所在范围估计。

【重点难点】

第一节 计量资料的频数表

一、频数表编制的步骤

当计量资料的观察值较多时,需编制频数分布表,其目的是了解数据分布规律和类型。频数表编制的步骤及注意事项见表 2.1。

表 2.1 频数表编制的步骤

步骤	具体操作	注意事项
1	极差 $R = \text{最大值} - \text{最小值}$	
2	①组数:一般 8~15 组 ②组距 $i = \text{极差} / \text{组数}$ ③组段:指每组的下限~上限	根据观察单位的多少酌情增减组数 组距一般取整数,可等组距,也可不等组距,一般多采用等组距。 一般只写下限,不用写上限。第一组段要包括最小观察值,最后一个组段包括最大观察值。
3	列表划记:采用划记法或计算机汇总 将原始数据归入各组	

二、频数分布的类型

1. 对称分布:是指集中位置在正中,左右两侧频数分布大体对称的分布。
2. 偏态分布:偏态分布是指集中位置偏向一侧,两侧频数分布不对称。
3. 对数正态分布:有些偏态分布的资料,其原始数据经过对数转换后(如用原始数据的对数值 $\lg X$ 代替 X)服从正态分布,称为对数正态分布。

三、频数表的用途

1. 便于观察资料的分布类型和分布特征。根据分布类型可以选择合适的统计指标进行计算和分析。

2. 常作为大样本资料的陈述形式。

3. 当数据不是用计算机处理分析时,大样本资料常整理成频数表的形式,计算相应的描述指标,并进行统计分析等。

4. 便于发现某些特大或特小的异常值,必要时经检查、核实后决定取舍。

第二节 描述集中趋势的指标

描述一组计量资料的集中趋势指标叫平均数。常用的平均数包括算术均数、几何均数和中位数。由于原始数据分布类型的不同,应选用不同的平均数表示其平均水平和集中趋势,见表 2.2。

表 2.2 常用描述集中趋势的指标

指 标	计算公式	适 用 资 料
算术均数 $\bar{x}$	①直接法(未分组资料)	适用于对称分布,尤其正态分布
	$\bar{x} = \frac{\sum x}{n},$	
	②加权法(分组资料)	
	$\bar{x} = \frac{\sum fx}{\sum f}$	
几何均数 G	①直接法(未分组资料)	①等比资料,②对数正态分布
	$G = \lg^{-1} \left(\frac{\sum \lg x}{n} \right)$	
	②加权法(分组资料)	
	$G = \lg^{-1} \left(\frac{\sum f \lg x}{\sum f} \right)$	
中位数 M	①直接法(未分组资料)	①偏态分布,②分布类型不详, ③末端无确定值
	位次居中的数值。 计算公式为: n 为奇数时 $M = x_{(\frac{n+1}{2})}$ n 为偶数时 $M = \left[x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right] / 2$	
	②频数表资料(分组资料)	
	$M = L + \frac{i}{f_M} \left(\frac{n}{2} - \sum f_L \right)$	

[附]百分位数

百分位数(percentile)是一个位置指标。将由小到大顺序排列的观察值分成 100 等份,对应于第 $x\%$ 位的观察值即为第 x 百分位数,用 P_x 表示。百分位数常用于描述偏态分布资料在某百分位置上的水平及确定偏态分布资料医学参考值范围,其计算公式

$$P_x = L + \frac{i}{f_x} (n \cdot x\% - \sum f_L)$$
 (2.1)

公式中 L 、 i 、 f_x 分别为 P_x 所在组段的下限、组距和频数, $\sum f_L$ 为小于 L 的各组段的累计频数。