

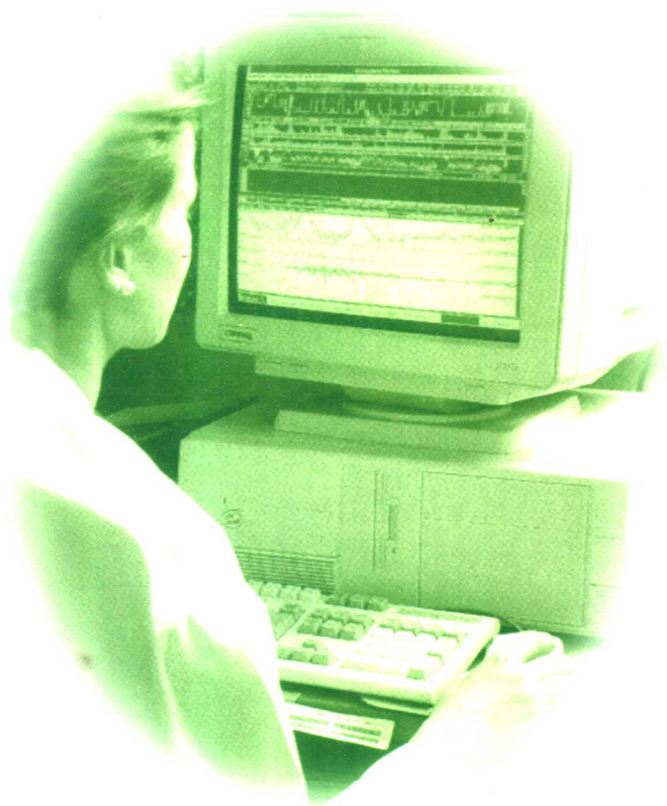
预防专业必修课考试辅导教材

供 预 防 类 专 业 用

梳理教材知识体系 精讲重点难点考点 揭示名校命题规律

卫生统计学

主编 陈长生



 科学技术文献出版社

预防专业必修课考试辅导教材
供预防类专业用

卫生统计学

主 编 陈长生
副主编 门 可 王 彤

科学技术文献出版社

Scientific and Technical Documents Publishing House

北 京

图书在版编目(CIP)数据

卫生统计学/陈长生主编.-北京:科学技术文献出版社,2005.1

(预防专业必修课考试辅导教材)

ISBN 7-5023-4875-1

I. 卫… II. 陈… 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(2004)第 115658 号

出 版 者 科学技术文献出版社

地 址 北京市复兴路 15 号(中央电视台西侧)/100038

图书编务部电话 (010)68514027,(010)68537104(传真)

图书发行部电话 (010)68514035(传真),(010)68514009

邮 购 部 电 话 (010)68515381,(010)58882952

网 址 <http://www.stdph.com>

E-mail: stdph@istic.ac.cn

策 划 编 辑 薛士滨

责 任 编 辑 马永红

责 任 校 对 唐 炜

责 任 出 版 王芳妮

发 行 者 科学技术文献出版社发行 全国各地新华书店经销

印 刷 者 北京金鼎彩色印刷有限公司

版 (印) 次 2005 年 1 月第 1 版第 1 次印刷

开 本 787×1092 16 开

字 数 618 千

印 张 21.5

印 数 1~6000 册

定 价 32.00 元

© 版权所有 违法必究

购买本社图书,凡字迹不清、缺页、倒页、脱页者,本社发行部负责调换。

前言

第四军医大学徐勇勇教授主编的《医学统计学与卫生统计学》于 2002 年由科学技术文献出版社作为医学专业必修课考试辅导丛书之一出版,实用易懂,已先后三次印刷,受到各医学院校广大师生的普遍欢迎。考虑到医学统计学与卫生统计学在内容上的差异,以及不同医学专业统计学教学的需要,《医学统计学与卫生统计学》将分为《医学统计学》和《卫生统计学》两本书由科学技术文献出版社再次出版。此次再版的《卫生统计学》是原《医学统计学与卫生统计学》的继承与发展。

统计学是所有医学院校普遍开设的一门重要的必修基础课程,目的是让学生掌握医学与卫生研究及工作实践中统计数据的收集、整理和分析的基本方法,统计学方法被医药卫生界比喻为整个医学大厦中的一个支柱。对于预防医学、卫生管理类专业来说,《卫生统计学》是一门专业基础课程,它是医学与卫生工作者在从事公共卫生实践和医药卫生科学研究过程中必须掌握和了解的基本知识,其在内容上涵盖了《医学统计学》,是研究生入学考试科目《预防医学综合》或《卫生综合》的重要内容之一。由于《卫生统计学》课程在理论知识体系和学习方法上有别于其他门类的医学卫生课程,因此被许多医学学生认为是较难学习和掌握的一门课程。为此,根据我们多年的统计学教学实践与体会,我们组织编写了这本《卫生统计学》考试辅导教程,旨在帮助医学学生更好地理解、消化、掌握卫生统计学的知识要点,提高学习成绩与培养实际应用能力。

本书是科学技术文献出版社出版的医学专业必修课考试辅导丛书之一,配套教材为人民卫生出版社出版的卫生部规划教材——《卫生统计学》(第 5 版,方积乾主编)。本书的编写得到科学技术文献出版社、第四军医大学、山西医科大学等兄弟医学院校和有关部门的大力支持。本书的编委会由国内 17 所知名高等医学院校的资深统计教师组成,他们多年从事卫生统计学的教学与科研工作,具有丰富的教学经验和广博深厚的统计理论知识,他们的积极参与和指导是本书具有科学性、理论性和实践性的保证。第四军医大学卫生统计学教研室的吴昊、王锐、杜晓哈三位同志核算了本书的所有典型试

题、习题以及模拟试题的解答。

本书的特点在于对各章的知识体系进行了精心梳理,突出重点与难点,结合典型试题进行剖析,通过大量的习题对各知识点予以加强和巩固,并结合模拟试题的自测、计算器的操作及 SPSS 统计软件的应用培养统计学方法的综合分析、运用及动手能力,使读者对统计学知识的理解更加深刻,处理统计学问题更加灵活,从而提高学习成绩和增强实际应用能力。因此,本书可作为高等医学院校各专业本科生、硕士研究生的教学配套教材,或《卫生统计学》考试的辅导教材;亦可作为考研、公共卫生执业医师资格考试和各类自学考试或函授考试的复习指南及模拟试题使用。

由于本人能力所限,书中肯定存在许多不足,诚恳希望广大师生提出宝贵修订意见。

陈长生

2004 年 9 月于西安

目 录



第一章 绪论	(1)
第二章 定量资料的统计描述	(8)
第三章 定性资料的统计描述	(28)
第四章 常用概率分布	(38)
第五章 参数估计基础	(52)
第六章 假设检验基础	(62)
第七章 χ^2 检验	(75)
第八章 实验设计	(95)
第九章 方差分析	(106)
第十章 基于秩次的非参数检验	(126)
第十一章 两变量关联性分析	(141)
第十二章 简单回归分析	(153)
第十三章 多重线性回归与相关	(164)
第十四章 调查设计	(178)
第十五章 总体特征抽样调查的设计与分析	(191)
第十六章 横断面研究资料的统计分析	(203)
第十七章 队列研究的设计与分析	(223)
第十八章 病例—对照研究的设计与分析	(236)
第十九章 Logistic 回归分析	(246)
第二十章 生存分析	(257)

第二十一章 综合评价.....	(268)
第二十二章 Meta 分析	(278)
附录一 计算器统计功能的使用方法.....	(289)
附录二 SPSS 计算程序	(294)
附录三 模拟试题.....	(303)
模拟试题一.....	(303)
模拟试题二.....	(306)
模拟试题三.....	(309)
模拟试题四.....	(313)
模拟试题五.....	(316)
模拟试题六.....	(319)
模拟试题一参考答案及评分标准	(322)
模拟试题二参考答案及评分标准	(324)
模拟试题三参考答案及评分标准	(326)
模拟试题四参考答案及评分标准	(328)
模拟试题五参考答案及评分标准	(330)
模拟试题六参考答案及评分标准	(332)

第一章

绪 论

教学大纲要求

(一)掌握内容

1.统计学的若干基本概念

总体与样本、同质与变异、随机变量、资料类型、参数与统计量、抽样误差、频率与概率。

2.统计工作的步骤

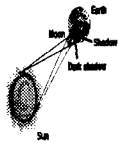
统计设计、收集资料、整理资料、分析资料。

(二)熟悉内容

1. 统计学、卫生统计学和医学统计学的含义。
2. 医学研究中的统计思想。
发展历史、设计与分析、因果与联系等。
3. 医学研究中统计方法的运用。

(三)了解内容

《卫生统计学》的学习方法与技能。



教学内容精要

(一) 统计学、医学统计学、卫生统计学

统计学是一门处理数据中变异性的科学与艺术,内容包括数据的收集、分析、解释和表达。医学统计学是用统计学的原理和方法研究生物医学现象的一门学科。

卫生统计学则是把统计学的理论和方法应用于居民健康状况研究、医疗卫生实践、卫生事业管理和医学科研的一门应用学科。

(二) 统计学中的几个基本概念

1. 总体与样本

总体(population)指根据研究目的所确定的同质的观察单位的全体。更确切地说,它是同质的所有观察单位某种观察值的集合,可分为有限总体和无限总体。总体中的所有观察单位都能够标识者为有限总体,反之为无限总体。

从总体中抽取部分观察单位,其观测结果的集合称为样本(sample)。样本应具有代表性。所谓有代表性的样本,是指用随机抽样方法从总体中获得的样本,也称随机样本。

2. 同质与变异

(1) 同质:一个总体中有许多个体,他们之所以汇集起来成为人们的研究对象,必定存在共性,即他们具有同质性(homogeneity)。医学研究中常见的同质总体有“正常人”总体、同性别总体、同年龄组总体、同职业总体等。没有同质性就构不成一个总体供人们研究。

(2) 变异:个体差异是生物医学研究领域普遍存在的现象。即使是同质总体的个体观察值之间也存在差异,这种现象称为变异(variation)。总体内没有变异就无须统计。

统计学的任务就是描述同一总体的变异规律,揭示不同总体之间存在的差别。

3. 随机变量

随机变量(random variable)指取值不能事先确定的观察结果,通常称为变量(variable)。例如,某地7岁男童的身高、体重;一次动物实验后动物的存活情况(存活或死亡)等。变量可分为定性变量(qualitative variable)(如人类的血型)和定量变量(quantitative variable)(如正常人血红蛋白含量)两种类型。定性变量又可分为分类变量(categorical variable)和有序变量(ordinal variable)。分类变量又可分为多分类变量和二分类变量。

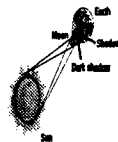
不同类型的变量可以进行转化,但变量通常由高级向低级转化,即连续变量→有序、多分类变量→二分类变量。

4. 资料类型

变量的实际观测结果构成资料(data)。资料按其性质可分为三种类型:

(1) 计量资料:对每个观察单位用定量的方法测定某项指标量的大小,所得的资料称为计量资料(measurement data)。计量资料亦称定量资料、测量资料。其变量值是定量的,表现为数值大小,一般有度量衡单位。如患者的身高(cm)、体重(kg)、红细胞计数($10^{12}/L$)、脉搏(次/分)、血压(kPa)等。

(2) 计数资料:将观察单位的观察结果按某种属性或类别分组,分别统计各组的观察单位数所得的资料称为计数资料(count data)。计数资料亦称定性资料或分类资料。其观察值是定性的,表现为两个或多个互不相容的类别或属性。例如,某地某时新生儿的男、女人数;一批少数民族居民的A、B、



AB、O 四种血型的人数等。

(3) 等级资料: 将观察单位的观察结果按某种属性的程度或等级分组, 分别统计各组的观察单位数所得的资料称为等级资料 (ordinal data)。等级资料又称有序资料。如患者的治疗结果可分为治愈、好转、有效、无效或死亡, 分别统计各种治疗结果的人数所得的资料即为等级资料。这里的各种结果既是属性分类, 又有顺序和等级差别, 但这种差别却不能准确测量。

等级资料与计数资料不同: 属性分组有程度或等级差别, 各组按大小顺序排列。

等级资料与计量资料不同: 每个观察单位的观察结果未确切定量, 故亦称为半计量资料。

5. 参数与统计量

(1) 参数: 参数 (parameter) 是反映总体特征的统计指标, 如总体均数、总体率等。总体参数是固定的常数。多数情况下, 总体参数是不易知道的, 但可通过随机抽样抽取有代表性的样本, 用算得的样本统计量估计未知的总体参数。

(2) 统计量: 统计量 (statistic) 是指与参数对应、通过样本计算的统计指标, 如样本均数、样本率等。样本统计量可用来估计总体参数。总体参数是固定的常数, 统计量是在总体参数附近波动的随机变量。

6. 抽样误差

抽样误差 (sampling error) 是指由抽样引起的样本统计量之间以及样本统计量与总体参数之间的差别。在总体确定的情况下, 总体参数是固定的常数, 统计量是在总体参数附近波动的随机变量。在抽样研究中, 抽样误差是不可避免的, 但可通过适当的统计方法进行估计。产生抽样误差的根本原因是生物个体间存在的变异性。

7. 概率与频率

概率 (probability) 又称几率, 是度量某一随机事件 A 发生可能性大小的一个数值, 记为 $P(A)$ 。 $0 < P(A) < 1$ 。在相同的条件下, 独立重复做 n 次实验, 事件 A 出现了 m 次, 则比值 m/n 称为随机事件 A 在 n 次实验中出现的频率 (frequency)。当实验重复很多次时, n 越大, m/n 越接近概率 $P(A)$ 。频率是就样本而言的, 而概率总是从总体的意义上说。当 A 发生的概率小于或等于 0.05 时, 统计学上称 A 为小概率事件。在一次实验中小概率事件发生时, 应特别引起注意。

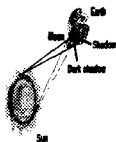
(三) 统计工作的步骤

1. 统计设计: 统计设计的内容包括对资料收集、整理和分析全过程的总的设想和安排。设计是整个研究中最关键的一环, 是今后工作应遵循的依据。
2. 收集资料: 按照设计的要求收集数据, 使获得的原始数据准确可靠。
3. 整理资料: 简化数据, 使其系统化、条理化, 便于进一步分析计算。
4. 分析资料: 计算有关指标, 反映事物的综合特征, 阐明事物的内在联系和规律。分析资料包括统计描述和统计推断。

(四) 医学研究中统计方法的运用

医学统计方法在医学研究中的运用主要有三个方面:

1. 以正确的方式收集数据。
2. 描述数据的统计特征。
3. 统计分析得出正确结论。



典型试题分析

(一)名词解释

总体与样本

答案:总体(population)是根据研究目的确定的同质的观察单位的全体,更确切地说,是同质的所有观察单位某种观察值(变量值)的集合。总体可分为有限总体和无限总体。总体中的所有单位都能够标识者为有限总体,反之为无限总体。

从总体中抽取部分观察单位,其测量结果的集合称为样本(sample)。样本应具有代表性。所谓有代表性的样本,是指用随机抽样方法从总体中获得的样本,即随机样本。

【评析】 本题考点:总体与样本的概念。

总体与样本是统计学中的重要概念。透过样本数据对总体的规律进行推断是统计学的核心内容。能否成功地由样本推断总体,关键是抽样方法、样本的代表性和推断的技术。

(二)单项选择题

1. 统计学中所说的有代表性的样本是指()。

- A. 随意抽取的总体中任意个体
- B. 有意识地选择总体中的典型个体
- C. 依照研究者要求选取总体中有意义的个体
- D. 最方便获取的部分个体
- E. 依照随机原则抽取总体中的个体

答案:E

【评析】 本题考点:统计学中样本概念的理解。

统计学中的样本是指从总体中随机抽取的部分观察单位的测量值的集合。这里的“随机抽取”并非通常所说的“随意抽取”,而是保证总体中每个观察单位等概率被抽取的科学方法。随机抽样是样本具有代表性的保证。

2. 下列变量的观测结果属等级资料的是()。

- A. 白细胞计数
- B. 住院天数
- C. 门急诊就诊人数
- D. 患者的病情分级
- E. 各临床科室的患者数

答案:D

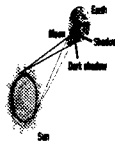
【评析】 本题考点:统计资料的分类。

统计资料按其性质可分为三种类型:计量资料、计数资料和等级资料。计量资料变量值是定量的,表现为数值大小,一般有度量衡单位,如本例中白细胞计数($10^9/L$),住院天数(天)。计数资料其观察值是定性的,表现为两个或多个互不相容的类别或属性的观察单位数,如门急诊就诊人数可按门诊、急诊分类清点各组人数。等级资料的属性分组有程度或等级的差别,各组按大小顺序排列,如患者的病情可按轻、中、重分级,然后分别统计各级的患者数所得的资料就是等级资料。

(三)简答题

一位研究人员欲做一项实验研究,研究设计应包括哪几方面的内容?

答案:一般来讲,研究设计应包括两方面的设计:专业设计和统计设计。专业设计是针对专业问题进行的研究设计,如选题、形成假说、干预措施、实验对象、实验方法等;统计设计是针对统计数据收集进行的设计,如样本来源、样本量、干预措施的分配、统计设计类型、测量指标的选择等。统计设计



是统计分析的基础,任何设计上的缺陷,都不可能在统计分析阶段弥补和纠正。

【评析】 本题考点:研究设计包含的内容。

研究设计是整个研究中最关键的一环,是整个研究过程中始终遵循的依据。正确、严谨、周密的设计是研究工作顺利进行、研究结果真实可靠的保证。因此,应深刻理解并掌握研究设计的内容及其意义。

(四)是非题

描述不确定现象,通过重复观察,发现生物医学领域的不确定现象背后隐藏的统计规律是医学统计的显著特征。()

答案:正确。

【评析】 本题考点:统计方法的特征。

在生物医学研究领域,由于存在较大的生物变异性,并受诸多因素的影响,使实验或观察结果往往成为不确定现象。在大量的重复实验中,这种不确定现象却呈现出明显的统计规律性。统计方法能够帮助人们分析数据,达到去伪存真、去粗存精,透过偶然现象认识其内在的规律性。这正是统计方法的显著特征。

习题

(一)名词解释

1. 随机抽样
2. 变异
3. 等级资料
4. 概率与频率
5. 变量
6. 参数
7. 统计量
8. 抽样误差

(二)单项选择题

1. 观察单位为研究中的()。

- A. 样本
- B. 全部对象
- C. 影响因素
- D. 个体
- E. 观察指标

2. 总体是由()组成。

- A. 部分个体
- B. 全部研究对象
- C. 全部个体
- D. 相同的观察指标
- E. 同质个体的所有观察值

3. 抽样的目的是()。

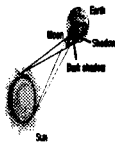
- A. 研究样本统计量
- B. 由样本统计量推断总体参数
- C. 研究典型案例
- D. 研究总体统计量
- E. 研究特殊个体的特征

4. 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. 总体中研究对象的总和

5. 关于随机抽样,下列哪一项说法是正确的()。

- A. 抽样时应使得总体中的每一个个体都有同等的机会被抽取
- B. 研究者在抽样时应精心挑选个体,以使样本更能代表总体



- C. 随机抽样即随意抽取个体
- D. 为确保样本具有更好的代表性,样本量应越大越好
- E. 选择符合研究者意愿的样本

6. 在探究疾病与危险因素的关系时,如果样本没有代表性,而统计分析发现某因素与疾病的发生在统计学上存在关联,则该因素与疾病之间()。

- A. 存在因果关系
- B. 存在数量上的联系
- C. 一定不存在因果关系
- D. 存在数量上的联系,说明该因素是导致疾病发生的原因
- E. 不一定存在数量上的联系

(三)是非题

- 1. 研究人员测量了 100 例患者外周血的红细胞数,所得资料为计数资料。()
- 2. 统计分析包括统计描述和统计推断。()
- 3. 定性变量、有序变量和定量变量可根据分析需要相互转化。()

(四)简答题

- 1. 某年级甲班、乙班各有男生 50 人。从两个班各抽取 10 人测量身高,并求其平均身高。如果甲班的平均身高高于乙班,能否推论甲班所有同学的平均身高高于乙班?为什么?
- 2. 一位研究人员欲做一项实验研究,研究设计应包括哪几方面的内容?

习题答案要点

(一)名词解释

1. 随机抽样:随机抽样(random sampling)是指按照随机化的原则(总体中每一个观察单位都有同等的机会被抽到到样本中),从总体中抽取部分观察单位的过程。随机抽样是样本具有代表性的保证。

2. 变异:在自然状态下,个体间测量结果的差异称为变异(variation)。变异是生物医学研究领域普遍存在的现象。严格地说,在自然状态下,任何两个患者或研究对象间都存在差异,其表现为许多生理测量值不是恒定的常数。

3. 等级资料:将观察单位按测量结果的某种属性的程度或等级分组,分别统计各组的观察单位数所得的资料称为等级资料(ordinal data)。等级资料又称有序资料。如患者的治疗结果可分为治愈、好转、有效、无效或死亡,分别统计各种治疗结果的人数所得的资料即为等级资料。这里的各种治疗结果既是属性分类,又有顺序和等级差别,但这种差别却不能准确测量。

4. 概率:概率(probability)又称几率,是度量某一随机事件 A 发生可能性大小的一个数值,记为 $P(A)$, $P(A)$ 越大,说明 A 事件发生的可能性越大。 $0 < P(A) < 1$ 。

频率:在相同的条件下,独立重复做 n 次实验,事件 A 出现了 m 次,则比值 m/n 称为随机事件 A 在 n 次实验中出现的频率(frequency)。当实验重复很多次时, n 越大, m/n 越接近概率 $P(A)$ 。当 $P(A) \leq 0.05$ 时,事件 A 称为小概率事件。

5. 变量:变量(variable)亦称随机变量(random variable),是指取值不能事先确定的观察结果。随机变量的具体内容虽然是各式各样的,但共同的特点是不能用一个常数来表示,而且,理论上讲,每个变量的取值服从特定的概率分布。

6. 参数:参数(parameter)是反映总体特征的统计指标,如总体均数、总体率等。总体参数是固定的常数。多数情况下,总体参数是不易知道的,但可通过随机抽样抽取有代表性的样本,用算得的样本统计量估计未知的总体参数。

7. 统计量:统计量(statistic)是指与总体参数对应、通过样本计算的统计指标,如样本均数、样本率等。样本统计量可用来估计总体参数。总体参数是固定的常数,统计量是在总体参数附近波动的随机变量。

8. 抽样误差:抽样误差(sampling error)是指样本统计量与总体参数的差别。在总体确定的情况下,总体参数是固定的常数,统计量是在总体参数附近波动的随机变量。

(二)单项选择题

1.D 2.E 3.B 4.B 5.A 6.E

(三)是非题

1. 错误。外周血的红细胞数是对血液中红细胞含量的测量值,其测量单位为($10^9/L$),属计量资料。

2. 正确。

3. 错误。虽然不同类型的变量可以进行转化,但变量通常由高级向低级转化,即连续变量→有序、多分类变量→两分类变量。

(四)简答题

1. 答案:不能。因为,从甲、乙两班分别抽取的 10 人,测量其身高,得到的分别是甲、乙两班的一个样本。样本的平均身高只是一个统计量,即使是按随机化原则进行抽样,由于存在抽样误差,样本均数与总体均数一般很难恰好相等。因此,不能仅凭两个样本均数高低就作出两总体均数孰高孰低的判断。

2. 答案:一般来讲,研究设计应包括两方面的设计:专业设计和统计设计。专业设计是针对专业问题进行的研究设计,如选题、形成假说、干预措施、实验对象、实验方法等;统计设计是针对统计数据收集和分析进行的设计,如样本来源、样本量、干预措施的分配、统计设计类型、测量指标的选择以及日后分析数据要用到的统计分析方法等。统计设计是统计分析的基础,任何设计上的缺陷,都不可能在统计分析阶段弥补和纠正。

(王 霞)

第二章

定量资料的统计描述

教学大纲要求

(一)掌握内容

1.频数与频数分布

- (1)频数表的编制。
- (2)频数分布的类型。

2.描述集中趋势的统计指标

掌握算术均数、几何均数、中位数、众数、调和均数的意义、计算方法和用途。

3.描述离散趋势的特征数

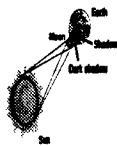
掌握极差、四分位数间距、方差、标准差、变异系数的意义、计算方法和用途。

4.统计表

- (1)统计表的结构。
- (2)统计表的种类及应用。
- (3)统计表的编制要求。

5.统计图

- (1)统计图的结构。



(2)统计图的种类及应用。

(3)统计图的编制要求。

(二)熟悉内容

1. 连续型定量变量的频数分布图编制。
2. 离散型定量变量的频数表和频数分布图的编制。
3. 偏度系数、峰度系数的计算方法和意义。
4. 常用的统计图表的绘制方法和注意事项。

教学内容精要

计量资料(measurement data)又称为测量资料,它是通过测量每个观察单位某项指标值的大小所得的资料,一般均有计量单位。它又可以分为两种类型:离散型变量(discrete variable)和连续型变量(continuous variable)。离散型变量只能取整数,连续型变量可以取任意实数。描述计量资料的分布规律通常有两种方法:统计图表和适当的统计指标。

(一)频数分布表的编制

频数是指当汇总大量的原始数据时,将数据按类型或组段分组,其中每个组出现的数据个数,称为该组的频数。频数分布表(frequency distribution table)又称频数表,可用来表示连续型变量各观察值在不同取值区间所出现的频繁程度,是对观察例数较多的大样本资料进行统计描述的常用方法。通过频数表可以显示数据分布的范围与形态。对于离散型定量变量,每个观察值就对应一个频数,把相应的观察值和频数列列表即可得到频数分布表。如某山区孕妇产前检查次数分别是0,1,2,⋯,>5的人数。对于散布区间很大的离散资料和连续型资料,数据散布区间由若干组段组成,每个组段对应一个频数。绘制连续型数据的一般步骤如下:

1. 找出数据的最小值($X_{\min}$)与最大值($X_{\max}$)

2. 计算全距(range, R)

$$R = X_{\max} - X_{\min} \quad (2-1)$$

3. 根据极差确定组段和组距

组段通常取10~15个,组距取 $1/R$ 。每个组段都有上限 U 和下限 L ,数据 X 归组统一规定为 $L \leq X < U$ 。第一组段和第末组段分别包括最小值和最大值。

4. 列出频数表

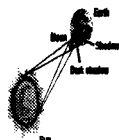
将各组段及所对应的频数用表格的形式列出来。频数表可揭示资料的分布特征和分布类型,便于进一步的统计分析处理,而且还便于发现一些特大或特小的可疑值。

(二)集中趋势的描述

平均数(average)常用于描述一组变量的集中位置和平均水平,根据资料的不同类型可选择不同的平均数指标。常用的有算术均数、几何均数和中位数。

1. 算术均数

算术均数(arithmetic mean)简称均数(mean),用于描述一组计量资料的平均水平。适合于描述对



称分布资料的集中位置,尤其是正态分布的资料。总体均数通常用 μ 表示,样本均数通常用 $\bar{X}$ 表示,其计算方法为:

(1)直接法:

$$\bar{X} = \frac{\sum X}{n} \quad (2-2)$$

(2)加权法(频数表法):适合于大样本资料,通常在频数表的基础上计算。

$$\bar{X} = \frac{\sum fX}{\sum f} \quad (2-3)$$

式中 X 为组中值, f 为频数。

2. 几何均数

几何均数(geometric mean, G)适用于观察值呈倍数变化或等比资料(其频数分布通常为正偏态)。其定义公式为

$$G = \sqrt[n]{X_1 X_2 \cdots X_n} \quad (2-4)$$

几何均数通常用来描述对数正态分布资料的集中位置。计算方法有以下两种:

(1)直接法:

$$G = \lg^{-1} \left(\frac{\sum \lg X}{n} \right) \quad (2-5)$$

(2)加权法:

$$G = \lg^{-1} \left(\frac{\sum f \lg X}{\sum f} \right) \quad (2-6)$$

3. 中位数

中位数(median, M)是指将 n 个观察值从小到大排列,位置居中的那个数。它特别适用于偏态分布,分布不明或分布末端无确定值的资料。计算方法如下:

当 n 为奇数时:

$$M = X_{\left(\frac{n+1}{2}\right)} \quad (2-7)$$

当 n 为偶数时:

$$M = \frac{1}{2} \left[X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)} \right] \quad (2-8)$$

4. 百分位数

百分位数(percentile)是一种位置指标,用 P_X 来表示。一个百分位数 P_X 将全部的变量分成两部分,在不包含 P_X 的全部变量中有 $X\%$ 的变量比它小,有 $(100 - X)\%$ 变量比它大。它的重要用途就是用来确定医学参考值范围。对于频数表资料,百分位数的计算公式为

$$P_X = L + \frac{i}{f_X} (nX\% - \sum f_L) \quad (2-9)$$

5. 众数

众数(mode)原指离散分布总体中出现机会最高的观察值。连续型变量众数的估计值通常是频数表中频数最多组段的组中值。