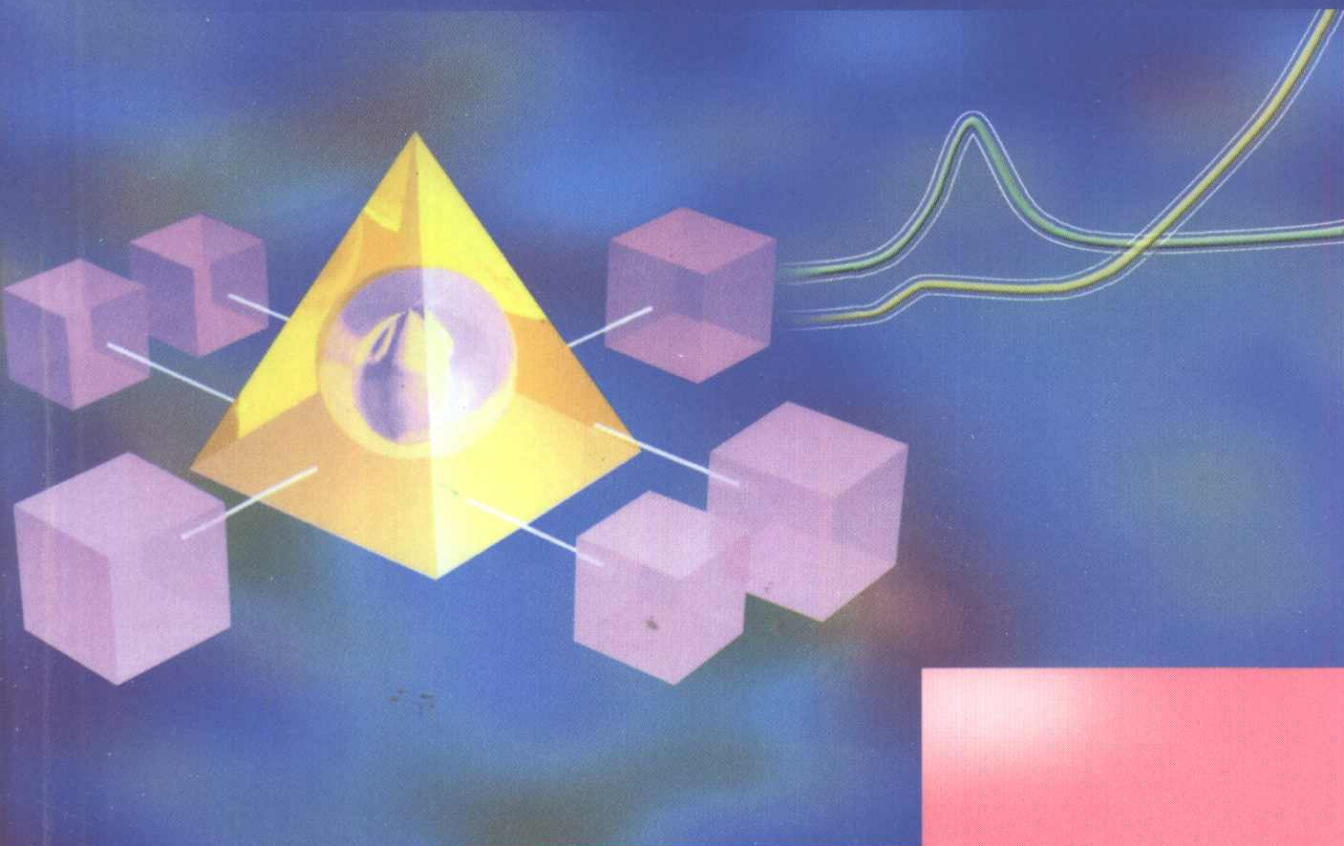


医学专业必修课考试辅导丛书



● 主编 徐勇勇 倪宗瓚

医学统计学 与卫生统计学

紧扣教学大纲 梳理知识体系 解读重点难点
网罗名校真题 精讲单项考点 引导复习路径

 科学技术文献出版社

医学专业必修课考试辅导丛书

医学统计学与 卫生统计学

主 编 徐勇勇 倪宗瓚

科学技术文献出版社

Scientific and Technical Documents Publishing House

北 京

编委会成员

(按姓氏笔划排序)

- 编委** 尹平(华中科技大学同济医学院)
王仁安(北京大学医学院)
王彤(山西医科大学)
孙振球(中南大学湘雅医学院)
周燕荣(重庆医科大学)
姚晨(解放军军医进修学院)
赵清波(第四军医大学)
骆福添(中山医科大学)
倪宗瓚(四川大学华西医学院)
夏结来(第四军医大学)
徐勇勇(第四军医大学)
曹素华(复旦大学医学院)
蒋知俭(哈尔滨医科大学)
詹绍康(复旦大学医学院)
潘晓平(四川大学华西医学院)
颜虹(西安交通大学医学院)
颜艳(中南大学湘雅医学院)
- 学术秘书** 王霞(第四军医大学)
(高等学校骨干教师资助计划资助)

编写组成员

(按姓氏笔划排序)

万 毅(第四军医大学)

马跃渊(第四军医大学)

王 霞(第四军医大学)

白玉祥(第四军医大学)

张玉海(第四军医大学)

杜晓晗(第四军医大学)

杨 鹏(第四军医大学)

陈 平(第四军医大学)

姜建辉(第四军医大学)

蒋红卫(第四军医大学)

潘 峰(第四军医大学)

薛富波(第四军医大学)

前 言

统计学是所有医学院校普遍开设的一门课程,目的是让学生掌握医学研究和临床实践中统计数据的收集、整理和分析的基本方法。对于非预防医学专业(如临床医学专业),统计学是一门基础课程,课程名称为“医学统计学”;对于预防医学专业来说,统计学是一门专业基础课程,课程名称为“卫生统计学”;对于医学研究生来说,则是所有专业的公共必修课,课程名称为“医学统计学”。无论是叫“医学统计学”还是叫“卫生统计学”,要求学生掌握的基本原理和方法是相同的。由于这门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课程,因此被许多医学生认为是较难学习和掌握的一门课程。鉴于此,我们组织编写了这本《医学统计学与卫生统计学》考试辅导教程,旨在帮助医学生更好地理解、消化、掌握医学统计学或卫生统计学的知识要点,提高学习成绩。

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》(第四

版,倪宗瓚主编)和高等教育出版社出版的“教育部面向 21 世纪课程教材”《医学统计学》(徐勇勇主编)。编委会由医学院校资深统计学专家教授组成,他们均为上述两本配套教材的主要编写人员,多年从事医学统计学或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。本书每一章作者的署名,第一为编委会的专家教授,第二为执笔者。诸位专家教授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。

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,并通过大量的习题对各知识点予以加强与巩固,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。因此,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,非医学专业毕业的硕士研究生《医学统计学》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及各类自学考试或函授考试的模拟试题使用。

由于编者水平与时间条件所限,书中难免存在缺点和不足,欢迎广大读者批评指正,以便再版时改正。

徐勇勇

2002 年 3 月于西安

目 录

第一章	绪论	(1)
第二章	计量资料的统计描述	(11)
第三章	正态分布	(26)
第四章	总体均数的估计和假设检验	(40)
第五章	方差分析	(55)
第六章	分类资料的统计描述	(75)
第七章	二项分布与 Poisson 分布及其应用	(89)
第八章	χ^2 检验	(101)
第九章	秩和检验	(116)
第十章	直线相关与回归	(132)
第十一章	多元线性回归与 logistic 回归	(152)
第十二章	统计表与统计图	(165)
第十三章	实验设计	(183)
第十四章	调查设计与资料分析	(202)
第十五章	医学人口统计与疾病统计常用指标	(224)
第十六章	寿命表	(240)
第十七章	随访资料的生存分析	(250)
第十八章	常用综合评价方法	(261)

第一章

绪论

一、教学大纲要求

+++++

(一) 掌握内容

1. 几个基本概念

样本与总体、频率与概率、资料类型、随机变量、误差。

2. 统计工作的步骤

设计、收集资料、整理资料、分析资料。

(二) 熟悉内容

医学统计学的含义、内容及其医学应用。

(三) 了解内容

医学统计的发展历史。

二、教学内容精要

(一) 统计学、医学统计学、卫生统计学

统计学是研究数据的收集、整理、分析与推断的科学。

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理论、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. 随机变量

随机变量(random variable)指取值不能事先确定的观察结果,通常简称为变量。随机变量有一个共同的特点是不能用一个常数来表示,而且从理论上讲,每个变量的取值服从特定的概率分布。

随机变量可分为两种类型:离散型变量和连续型变量。

2. 误差

误差(error)指实际观察值与观察真值之差、样本指标与总体指标之差。误差可分为系统误差和随机误差,两种误差的区别见表 1-1。

3. 资料类型

观察单位的某项特征的测量结果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。如某一患者的身高(cm)、体重(kg)、红细胞计数($10^{12}/L$)、脉搏(次/分)、血压(kPa)等。

表 1-1 系统误差与随机误差的区别

误差分类	产生原因	对观察值的影响	处理方法
系统误差	仪器未校正、测量者感官的某种偏差、医生掌握疗效标准偏高或偏低等因素所致。	使观察值不是分散在真值的两侧,而是有方向性、系统性或周期性地偏离真值。	通过实验设计的完善和技术措施的改进来消除或减少。
随机误差	排除系统误差后,其他多种不确定因素所致。	使观察值不按方向性、系统性而随机的变化,误差变量一般服从正态分布。	通过计算统计指标估计。

(2)计数资料:将观察单位按某种属性或类别分组,所得的观察单位数称为计数资料(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,表现为互不相容的类别或属性。如调查某地某时的男、女性人口数;治疗一批患者,其治疗效果为有效、无效的人数;调查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:将观察单位按测量结果的某种属性的不同程度分组,所得各组的观察单位数,称为等级资料(ordinal data)。等级资料又称有序变量。如患者的治疗结果可分为治愈、好转、有效、无效或死亡,各种结果既是分类结果,又有顺序和等级差别,但这种差别却不能准确测量;如:一批肾病患者尿蛋白含量的测定结果分为 +、++、+++ 三个等级,清点各等级的患者数所得资料即为等级资料。

等级资料与计数资料不同:属性分组有程度差别,各组按大小顺序排列。

等级资料与计量资料不同:每个观察单位未确切定量,但有大小程度不同,故亦称为半计量资料。

4. 总体与样本

总体(population)指特定研究对象中所有观察单位的测量值。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。

从总体中随机抽取部分观察单位,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获得的样本。

5. 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 次试验,事件 A 出现了 m 次,则比值 m/n 称为随机事件 A 在 n 次试验中出现的频率(frequency)。当试验重复很多次时 $P(A) = m/n$ 。

(三) 统计工作的步骤

1)设计:设计内容包括资料收集、整理和分析全过程总的设想和安排。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:应采取措施使能取得准确可靠的原始数据。

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应用

医学统计方法在医学研究中的应用主要有三个方面:

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。

(五) 医学统计发展历史

最初的统计主要是数据汇总。统计发展到今天,已经成为一种对客观现象数量方面进行的调查研究活动,是收集、整理、分析、判断等认识活动的总称,数据汇总仅仅是统计工作的一小部分。医学统计的发展起源于生物统计、遗传统计;其经过了描述统计、大样本统计、小样本统计推断、临床统

计和多变量统计几个发展阶段。

三、典型试题分析

+++++

(一) 名词解释

抽样误差。

答案: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。

【评析】 本题考点:抽样误差的概念。

抽样误差是统计学中的重要概念。在抽样研究中是不可避免的。产生抽样误差的根本原因是生物个体间存在的变异性。

(二) 单项选择题

1. 统计学中所说的样本是指()。

- 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一部分

答案:D

【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. 下列资料属等级资料的是()。

- A. 白细胞计数
- B. 住院天数
- C. 门、急诊就诊人数
- D. 病人的病情分级(轻、中、重)

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差别,各组按大小顺序排列,如病人的病情分级为轻、中、重。

(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。在大量的重复试验中,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、去粗取精,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。

四、习题

(一) 名词解释

1. 总体与样本
2. 随机抽样
3. 变异
4. 等级资料
5. 概率与频率
6. 随机误差
7. 系统误差
8. 随机变量
9. 参数
10. 统计量

(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
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
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. 研究典型案例研究误差
D. 研究总体统计量
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
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

(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. 计量资料、计数资料和等级资料可根据分析需要相互转化。

(四) 简答题

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大于乙班, 能否推论甲班所有同学的平均身高大于乙班? 为什么?

五、习题答案要点

+++++

(一) 名词解释

1. 总体: 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。

样本: 从总体中随机抽取部分观察单位, 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获得的样本。

2. 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选入到样本中), 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。

3. 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群体间都存在差异, 其表现为各种生理测量值的参差不齐。

4. 等级资料: 将观察单位按测量结果的某种属性的不同程度分组, 所得各组的观察单位数, 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、死亡, 各种结果既是分类结果, 又有顺序和等级差别, 但这种差别却不能准确测量。

5. 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 次试验,事件 A 出现了 m 次,则比值 m/n 称为随机事件 A 在 n 次试验中出现的频率(frequency)。当试验重复很多次时 $P(A) = m/n$ 。

6. 随机误差:随机误差(random error)又称偶然误差,是指排除了系统误差后尚存的误差。它受多种因素的影响,使观察值不按方向性和系统性而随机的变化。误差变量一般服从正态分布。随机误差可以通过统计处理来估计。

7. 系统误差:系统误差(systematic error)是指由于仪器未校正、测量者感官的某种偏差、医生掌握疗效标准偏高或偏低等原因,使观察值不是分散在真值的两侧,而是有方向性、系统性或周期性地偏离真值。系统误差可以通过实验设计和完善技术措施来消除或使之减少。

8. 随机变量: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,而且从理论上讲,每个变量的取值服从特定的概率分布。

9. 参数:参数(paramater)是指总体的统计指标,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。

10. 统计量:统计量(statistic)是指样本的统计指标,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。

(二)单项选择题

1. D 2. C 3. B 4. B 5. A

(三)是非题

1. 错。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. 正确。

(四) 简答题

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甲、乙两班所有同学平均身高的一个点估计值。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个总体均数孰高孰低的判断,而应通过统计分析,进行统计推断,才能作出判断。

(倪宗瓚 王 霞)