

姜 明 王国荃 主编

新编医学统计学

新疆科技卫生出版社(K)

新 编 医 学 统 计 学

姜 明 王国荃 主编

新疆科技卫生出版社(k)出版发行

(乌鲁木齐市延安路4号 邮政编码 830001)

小硕士电脑公司排版 乌鲁木齐铁路局印刷厂印刷

787×1092毫米 32开本 12.75印张 283千字

1991年12月第1版 1992年4月第1次印刷

印数:1—3000册

ISBN7—5372—0272—9/R·30 定价:5.50元

编著 (以姓氏笔画为序)

王学元 王国荃 汪艾翔 宋克征 何宏铨

姜 明 徐 泰 黄承平 隋枝叶

前 言

医学统计是整个医学大厦的柱石,也是每个医药卫生工作者必修的一门基本课程。为了帮助大家尽快掌握医学统计学原理和方法,提高医务工作效益,我们特编写了这本《新编医学统计学》。

所谓“新编”,是因为这本书既不同于全国医学院校统编教材,也不同于近几年来出版的一般医学统计辅导读物。它内容新颖,体系独特。在内容上着重实际应用,兼顾普及与提高,特别是阐明了某些新的观点,提出了某些新的见解;对容易混淆的概念多采用图示和列表的方法加以对比,便于理解和记忆;针对科研需要,对“统计方法在临床实验室质量控制中的应用”作了专章论述;还编有从63种医学杂志中精选出来的“综合实习题”,通过对有代表性的各类医学统计实例的分析,提高读者解决实际问题的能力。在体例编排上适应快速查阅的需要,把同一问题的各种解决方法集中在一起,读者只要查阅本书目录,根据自己所掌握的资料类型和研究目的,学会

“对号入座”，很快就可以找到相应的统计分析方法，犹如查字典一样。我们曾将本书初稿在几个统计培训班试用，收到了事半功倍之效。

本书可作为培训广大基层医药卫生人员及不同类型的成人医学教育的教材；对高、中等医学院校师生也有较大参考价值，可作为辅导教材；也可供自学之用。

本书在编写过程中，参考了许多学者的研究成果，并得到卫生统计界前辈、上海医科大学顾杏元教授的指导，杨振明主任医师亦对本书提出了宝贵意见，夏新业同志为本书绘图，在此一并致谢。

由于我们水平有限，不妥之处在所难免，敬请读者不吝指正。

姜 明 王国荃

1991年11月于新疆医学院

目 录

第一章 绪论	(1)
1·1 医学统计学的定义、研究的主要内容及对象	(1)
1·2 统计工作的步骤.....	(2)
1·3 统计学中的几个基本概念.....	(5)
1·4 统计资料的类型.....	(9)
第二章 单变量计量资料的统计分析 ——统计描述	(11)
2·1 频数分布	(11)
2·2 集中趋势	(14)
2·3 离散趋势	(20)
2·4 正态分布	(26)
2·5 正常值范围的估计	(35)
第三章 单变量计量资料的统计分析 ——统计推断(上)	(40)
3·1 均数的抽样误差与标准误	(40)
3·2 t 值和 t 分布	(45)

3·3	总体均数的估计	(50)
3·4	假设检验	(54)
3·5	配对资料(或同一批对象实验前后)的比较	(60)
3·6	两个样本平均数的比较	(66)
第四章 单变量计量资料的统计分析		
	——统计推断(下)	(80)
4·1	多个样本平均数比较的思考	(80)
4·2	多个样本比较的方差分析	(81)
4·3	多个样本比较的秩和检验	(99)
4·4	参数检验与非参数检验的比较	(109)
第五章 正态性检验 (111)		
5·1	正态性检验的意义	(111)
5·2	矩法(动差法)	(113)
5·3	D 检验法	(124)
5·4	正态概率纸目测法	(130)
第六章 双变量计量资料的统计分析 (133)		
6·1	直线相关	(133)
6·2	直线回归	(140)
6·3	直线相关与回归的区别和联系	(149)
第七章 计数资料的统计分析		
	——统计描述	(153)
7·1	相对数的种类及计算方法	(153)
7·2	应用相对数时的注意事项	(164)
7·3	标准化法	(166)
第八章 计数资料的统计分析		

——统计推断(上)·····	(170)
8·1 率的抽样误差与标准误·····	(170)
8·2 总体率的估计·····	(171)
8·3 样本率与总体率的比较·····	(172)
8·4 两个样本率的比较·····	(177)
8·5 四格表的确切概率法·····	(186)
第九章 计数资料的统计分析	
——统计推断(下)·····	(191)
9·1 多个样本率的比较·····	(191)
9·2 两个或多个样本构成比的比较·····	(195)
9·3 同一样本中两个或多个构成比的比较·····	(197)
9·4 对同一批对象进行两种实验的 χ^2 检验·····	(199)
9·5 计数资料的相关分析·····	(201)
9·6 分层分析的 Mantel—Haenszel 法·····	(206)
第十章 等级资料的统计分析 ·····	(213)
10·1 概述·····	(213)
10·2 两组等级资料的比较·····	(215)
10·3 多组等级资料的比较·····	(221)
10·4 等级相关分析·····	(226)
第十一章 生存率分析 ·····	(231)
11·1 概述·····	(231)
11·2 病例随访资料分析的寿命表法·····	(232)
11·3 应用生存率应注意的问题·····	(236)
第十二章 统计表与统计图 ·····	(238)
12·1 统计表·····	(238)
12·2 统计图·····	(242)

第十三章 调查研究设计	(251)
13·1 调查研究的定义和分类	(251)
13·2 制定调查计划	(253)
13·3 几种抽样方法介绍	(255)
13·4 样本量大小的确定	(269)
13·5 流行病学调查方法	(277)
第十四章 临床试验设计	(286)
14·1 概述	(286)
14·2 临床试验的四个基本要素	(287)
14·3 临床试验设计的三大原则	(295)
14·4 临床试验设计方案的选择及几种常用试验 设计方案的比较	(302)
14·5 一个正交试验设计的实例分析	(304)
第十五章 医学科研和论文中存在的主要统计问题	(311)
15·1 设计和搜集资料中存在的问题	(311)
15·2 资料整理和分析中存在的问题	(315)
15·3 资料表达和论文写作中存在的问题	(320)
第十六章 统计方法在临床实验室质量控制中的应用	(330)
16·1 概述	(330)
16·2 实验室内质量控制	(333)
16·3 室间质量评价	(350)
16·4 质量控制中应用统计方法的注意事项 ...	(353)
附： 综合实习题	(354)

第一章 绪论

1·1 医学统计学的定义、研究的主要内容及对象

医学统计学是运用概率论和数理统计的原理、方法,结合医学专业知识,按统计设计要求搜集、整理和分析资料的一门学科。这种分析可引出推断结论并导致决策,它归属于应用统计范畴。

本书研究的主要内容,侧重于医学统计的基本理论、基本概念和基本统计方法;侧重于现场调查、临床试验设计和医学论文中存在的主要统计问题;侧重于统计方法在科研工作中的应用,如控制混杂和偏倚,实验室的质量控制等。最终目的是通过学习解决医疗和预防工作中所遇到的实际问题。正如奥斯丁·比·希尔所说:人们对疾病有变异性,对疾病和治疗的反应也有变异性。我们必须承认,因为变异性的存在,

才有必要应用统计方法。由此可见,医学统计的产生与发展是由于生物具有变异性。因此,凡是在同质的基础上,有变异的大量随机现象都是统计学的研究对象。

1·2 统计工作的步骤

统计工作一般可分为统计设计、搜集资料、整理资料、分析资料和运用资料等五个步骤。

1. 设计(design)

设计是科研工作开始之前的先导,是科研进行过程中的依据,是所得结果准确可靠的保证。由此可见,它是最关键的一步。

设计有专业设计与统计设计之分,二者相辅相成,缺一不可。专业是基础,统计是使研究结果更可靠、更科学,而且经得起重复考验的手段。本书重点讨论统计设计。

2. 搜集资料(collection of data)

搜集资料是统计工作的前提与基础,因此资料要求完整、准确、及时。防止误差及偏倚影响观察结果的精确性。诊断标准和疗效判断标准要有科学依据并要求统一,避免研究者的主观偏见。

资料的来源包括靠登记获得的“经常性资料”(如统计报

表、医疗卫生工作记录及报告卡片等)和靠调查、实验获得的“一时性资料”(如专题调查及实验研究)两大类。

3. 整理资料(sorting data)

(1)对原始资料进行逻辑和计算审核。

(2)设计分组,包括质量分组(如病人按性别、职业、病情轻重分组)和数量分组(如按年龄、体重、身高分组)两种。

(3)拟订整理表。整理表是统计表的前身,是过渡性表格,作归纳汇总用。根据研究目的,应将有关联的几个项目制订在一个表内,以便说明事物之间相互联系的规律性。例如:

表 1·1〔原始表〕 ×××病感染按性别、年龄整理表

	性 别		年 龄 组				合 计
	男	女	0—	5—	10—	15—, 20—, ...	
调查人数							
感染人数							

该整理表由于将性别与年龄这两个有联系的项目分割开来,于是计算不出不同性别、年龄的感染率,丧失了有价值的信息。为了克服上述缺点,将上表作如下修改。

表 1·2〔修改表〕 ×××病感染按性别、年龄整理表

年龄组 (岁)	男		女	
	调查人数	感染人数	调查人数	感染人数
0—				
5—				
10—				
...				
...				
合计				

(4)编制统计表(详见第十二章)

4. 分析资料(analysis of data)

根据统计整理的结果,计算相应的指标,进行统计学处理(统计描述和统计推断)。利用对比分析、归纳、逻辑推理、假说的验证等科学方法研究分析资料,最后结合专业作出恰如其分的结论。

5. 运用资料(manipulation of data)

资料的搜集、整理和分析是属于对客观事物的认识过程,而资料的运用则是属于改造客观事物的过程。因此,资料的运用是理论指导实践的过程,是决策的兑现。

1·3 统计学中的几个基本概念

1. 变异

变异的含义是指在同质条件下的观察单位,其同一标志的数据间之差别。例如,同民族、同性别、同年龄的一批男孩,在类同的条件下生长,而他们的身高、体重不完全相同,统计学把这种个体间的差异称为变异。变异不但存在于人体之间,而且普遍存在于一切生物体之间。

2. 总体、样本与抽样研究方法

根据研究目的确定同质的研究对象的全体称为总体。总体又有无限总体与有限总体之分。如果被抽样的总体含量为无穷大,则称它为无限总体;反之,称之为有限总体。在实际工作中,如果样本在总体里占的比例很小,如 $n < \frac{1}{10}N$ 时,就可以把这个总体近似地当作无限总体。

从总体中随机(random)抽取的一部分个体称为样本。样本是总体中有代表性的一部分。样本中所含个体的多少,称为该样本的含量。例如,要调查某年某地汉族 12 岁健康男孩的平均身高水平,从中随机抽取 120 名进行身高测量。某地所有汉族 12 岁健康男孩是总体,而随机抽取的 120 名就是样本。

样本含量(常用 n 表示)为 120。

通过 120 名 12 岁汉族男孩的平均身高,便可用统计学方法推断某地所有 12 岁汉族男孩身高水平。这种从总体中随机抽样,用样本指标推断总体指标的方法叫抽样研究方法(图 1 · 1)

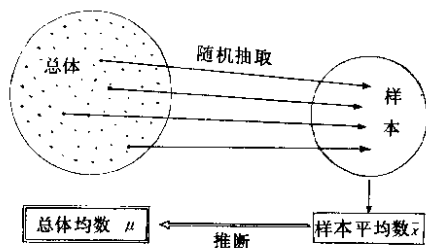


图 1 · 1 抽样研究方法示意图

从总体中随机抽取一部分个体,这是手段;用样本指标推断总体指标才是目的。

3. 随机

在抽样过程中为了避免研究者的主观意愿或客观无意识的偏性影响,必须严格遵循“随机化”抽样和分配的原则。

所谓随机,是指每个观察对象都有同等的机会(等概率)被抽中或分配,而不受研究者的主观意愿决定。

常用的随机化方法有两种,一是抽签法,二是随机数字表

法。

4. 抽样误差与偏倚(详述于后)

5. 概率(机率)

概率的理论是统计推断的依据。可从以下两个方面来理解概率：

(1) 概率俗称可能性，它是用于描述一个随机事件发生的可能性大小的一个度量，常用符号 P 来表示。

(2) 通常我们把概率理解为频率的稳定值。从下面的试验不难看出，当大量重复观察时，频率稳定在 0.5。这时，我们说“正面朝上”这个事件的概率为 0.5。

表 1·3 投掷硬币试验结果

受试者	投掷次数 (n)	出现“正面朝上” 的次数(z)	频率(z/n)
Demorgan	2048	1061	0.5180
Buffon	4040	2048	0.5069
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

频率和概率都是描述某一随机事件发生的可能性大小的一个度量。频率是对样本而言，概率则是总体的属性。

随机事件的概率介于 0 与 1 之间。习惯上将 $P \leq 0.05$ 或 $P \leq 0.01$ 称为小概率事件。小概率事件在一次抽样中是几乎不会发生的。假设检验就是依据这个小概率原理作为判断拒

绝还是不拒绝 H_0 的理论根据。由此可见,统计分析所作的任何推断结论,实际上都是概率估计。

概率有两个特性:

(1) 对任意事件 A , 有 $0 \leq P(A) \leq 1$ 。在特殊情况下, 如果

① $P(A) = 0$, 事件 A 不会出现(不可能事件);

② $P(A) = 1$, 事件 A 必然出现(必然事件)。

(2) 属于一个试验的所有互斥结果的概率之和为 1。这样, 如果一个试验有 k 个互斥结果, $A_1, A_2, \dots, A_k$, 各自的概率为 $P(A_1), P(A_2), \dots, P(A_k)$, 则 $\sum_{i=1}^k P(A_i) = 1$ 。

6. 参数与统计量

根据总体分布的特征而计算的总体指标为参数。一般用小写的希腊字母表示。如 ρ 表示总体相关系数, β 表示总体回归系数, ……。

从样本数据中算得的统计指标称之为统计量。一般用小写的拉丁字母表示, 如 r 表示样本相关系数, b 表示样本回归系数, ……。

7. 统计描述与统计推断

统计描述: 主要描述样本特征, 不考虑抽样误差问题。描述的形式有: (1) 列表描述, 如一览表、频数表; (2) 图示描述, 如直条图、直方图、构成图; (3) 数字描述, 如平均数、标准差。

统计描述是对原始资料的一种概括, 即把分散而不好理