

医学统计学概念、 方法及应用

蒋知俭 编著

黑龙江科学技术出版社

医学统计学概念、方法及应用

蒋知俭 编著

黑龙江科学技术出版社

1987年 • 哈尔滨

责任编辑：李 南

封面设计：岳大地

医学统计学概念、方法及应用

蒋知俭 编著

黑龙江科学技术出版社出版

(哈尔滨市南岗区建设街 35 号)

依安印刷厂印刷·黑龙江省新华书店发行

787×1092毫米32开本9.5印张·1 插页 194千字

1987年10月第1版·1987 年10月第1次印刷

印数：1—5,000册

书号：13217·187 定价：2.10元

ISBN7—5388—0051—4/N·7

前 言

广大医务工作者为向科学技术现代化进军，有效地进行医学科学实验研究，急需掌握一定的医学统计学方法。本书列举了大量临床、卫生、医药教学、科研实践中常用的统计方法，按概念、方法、应用三方面进行叙述，内容简明扼要，通俗易懂，实用性强，可供医药卫生科研人员在教学、科研及撰写论文时参考使用。

限于本人水平有限，书中一定有不少缺点和错误，恳请读者批评指正。

蒋知俭

目 录

第一章 绪论	(1)
第一节 几个基本概念	(1)
第二节 统计资料的收集和整理	(5)
第三节 统计表与统计图	(10)
第二章 平均数与标准差	(20)
第一节 平均数	(20)
第二节 标准差	(32)
第三节 平均数与标准差的应用	(35)
附：正常值范围的确定方法	(39)
第三章 相对数	(48)
第一节 常用的相对数	(48)
第二节 相对数的正确应用	(50)
第三节 标准化法	(54)
第四章 抽样误差与t检验	(61)
第一节 抽样误差	(61)
第二节 $t(u)$ 检验	(67)
第三节 应用 $t(u)$ 检验时的注意事项	(81)
附：其他几种常用的t检验	(83)
第五章 方差分析 (F 检验)	(88)
第一节 单因素多个样本均数的方差分析	(89)

第二节	按两个因素分组的多个样本均数的 方差分析·····	(95)
第三节	拉丁方资料方差分析·····	(98)
第四节	方差分析使用注意事项·····	(100)
第六章	χ^2 检验·····	(102)
第一节	四格表 (2×2 表) 资料的 χ^2 检验·····	(104)
第二节	$2 \times k$ 表资料的 χ^2 检验·····	(110)
第三节	行 $\times$ 列 ($R \times C$) 表资料的 χ^2 检验 ·····	(113)
第四节	配对资料的 χ^2 检验 ·····	(114)
第五节	χ^2 检验使用的注意事项·····	(117)
附:	相对危险性 ($R \cdot R$) 检验 ·····	(119)
第七章	直线相关与直线回归·····	(124)
第一节	直线相关·····	(124)
第二节	直线回归·····	(133)
第三节	直线相关与直线回归应用的 注意事项·····	(138)
附:	曲线回归方程·····	(143)
第八章	半数致死量 (LD_{50}) ·····	(145)
第一节	半数致死量的意义与 实验设计要点·····	(145)
第二节	目测概率单位法·····	(148)
第三节	寇氏法·····	(154)
第九章	非参数统计·····	(159)
第一节	配对资料的比较·····	(160)

第二节	两组资料的比较	(165)
第三节	多组资料的比较	(166)
第四节	按等级分组资料的秩和检验	(168)
第五节	等级相关	(170)
第六节	Ridit 分析 (简称R检验)	(171)
第七节	随机性检验	(178)
第八节	非参数统计的适用范围	(181)
第十章	医学调查	(183)
第一节	医学调查分析提纲	(183)
第二节	现场调查	(186)
第三节	关于样本规模的估计	(190)
第十一章	医学实验设计	(201)
第一节	实验设计的作用、内容、要素、误差及 实验设计的原则	(201)
第二节	几种常用的实验设计	(206)
第三节	实验设计中样本规模估计	(228)
第四节	实验设计参考提纲	(231)
第五节	临床实验设计	(233)
第十二章	多因素分析简介	(244)
第一节	多元线性回归	(245)
第二节	多元逐步回归	(247)
第三节	判别分析	(250)
第四节	聚类分析	(255)
附表1	随机数目表	(258)
附表2	标准正态曲线下的面积表	(260)

附表 3	t 值表	(262)
附表 4	F 值表 $4_{2(1-4)}$	(264)
附表 5	q 值表	(274)
附表 6	χ^2 值表	(276)
附表 7	相关系数界 值 表.....	(278)
附表 8	配对比较的秩和检验界 值 表.....	(282)
附表 9	等级相关系数界 值 表.....	(284)
附表10	两组比较的秩和检验界 值 表.....	(286)
附表11	百分率与概率单位换算 表.....	(290)
附表12	正态性 D 检验界 值 表.....	(295)
附表13	总复习参考表	

第一章 绪 论

医学统计方法是数理统计在医学科学研究和医疗预防实践中的应用。它运用概率论与数理统计的基本规律,研究社会条件、环境因素、生物因素等对人类健康的影响,可应用于评价医疗预防措施的质量和效果及制定医疗预防措施,亦可应用于医学科学研究的调查、实验设计和数据处理,以提高科学研究水平与效率。

医学统计方法是在质和量密切联系的基础上,认识人群的健康和疾病现象的数量特征,进行医药卫生科研的重要手段。主要包括收集、整理和分析资料的基本统计理论与方法,以及反映人民健康的疾病统计和死亡统计等。

近年来,医学统计方法的研究有了很大发展,特别是电子计算机的普及,对实验设计和分析方法都提出了新的要求。本书着重于基本统计方法、概念及应用方面介绍,并对实验设计、多元分析等方面作以扼要介绍。

第一节 几个基本概念

一、总体与样本

总体是同质的个体所组成的集体,即研究对象的全体。

样本是总体中有代表性的一部分。

实际工作中，经常是用掌握的总体中的一部分资料（即一个样本），来推测或说明总体，如用一滴血代表一个人的全血，用部分水样推测整个江、河的水质，用部分病人的疗效估计全体病人疗效等。

二、变异

研究对象中个体间的差异称为变异。例如，同性别、同年龄的一批儿童，虽在相同条件下生长，他们的身高、体重不尽相同、生长速度也各异；经严格精选的受试对象疗效、反应也未必相同等。

三、随机化

研究对象中，每个个体被抽取的机会完全相等。例如，动物实验中，动物分组必须遵循随机方法，将动物按同等的机会分配到各组中。随机化是抽样研究和抽样分配十分重要的原则，具体方法有抽签法和随机数目表等。

四、抽样误差

从总体中抽取部分个体的过程称为“抽样”。由于总体中各个体间存在变异，在同一总体中随机抽取若干样本，各样本间会有所不同；同时也反映样本与总体间的差异，这种由于抽样而引起的样本指标与总体指标间的差异称抽样误差。

一般而言，样本越大，抽样误差越小，越和总体指标近似，正确性越高。反之，样本越小，抽样误差越大。

五、概率（机率）

概率是指某事件发生的可能性的。统计上用符号“P”表示。P 介于 0 与 1 之间。某事件发生的概率越接近 0，表示该事件发生的可能性越小，概率越接近 1，表示事件发生的可能性愈大。例如，大量统计表明，生男或生女的机会各约为 0.5，即概率分别为 0.5。医学文献中，常见 $P \leq 0.05$ 或 $P \leq 0.01$ ，分别表示事件发生的可能性 ≤ 0.05 或 ≤ 0.01 ，作为判定事物间差别有、或有非常显著意义的界限。

六、计量资料与计数资料

计量资料是对每个观察单位测量所得的资料，一般用度量衡单位表示。例如，身高、体重、脉搏、血压、白细胞数等。

计数资料是将观察单位按其性质或类别进行分组后，清点个数所得的资料，例如，阳性数、治愈数、婴儿数、死亡数、龋齿数等。

在医学实践中，有些资料具有计数资料的特性，同时又兼有半定量的性质，我们称它为按等级分组的资料。例如，检验结果按一、±、+、++、+++ 等级，分别清点个数，观察疗效按痊愈、显效、无效、恶化等级分组清点。就其本质而言，该资料属计数资料。

有时计量资料和计数资料可以互相转化。例如，作为计量资料的舒张期血压规定 $\geq 90 \text{ mmHg}$ 为异常，可清点正常或异常个数等（表 1.1）。

表1.1 计数、计量资料举例

计数资料			计量资料		
实验结果	治疗效果	抗体反应	抗体滴度	住院天数	正常成人血压mmHg
阳性	痊愈	-	1 : 2	2	108
	好转	±	1 : 4	3	114
	无效	+	1 : 8	4	120
阴性	恶化	++	1 : 16	5	132
	死亡	+++	1 : 32	6	116
					

七、统计量与参数

样本的统计指标称为统计量或统计数。例如，样本平均数，样本标准差，样本率等。

总体的统计指标称为参数。例如，总体平均数，总体标准差，总体率等。

八、几种常见的分布

数理学家据积累大量资料，加以概括和抽象，结合概率理论，得出几种理论频数分布，并应用于实践。

1. 正态分布

为连续型、钟型光滑、中间高两侧低，左右对称的分布。在医学领域中，有不少资料的分布属于或近似正态分布。例如，儿童发育指标，健康成人红、白细胞等生理数值。

2. 二项分布

各事件彼此独立互斥，其中任何一事件发生的概率都是

各事件发生概率之和。适用于医学实验中某一现象只能分成对立的双方者。例如，动物的生与死，性别的雌与雄，检验结果的阴性与阳性等计算概率大小或观察疗效等。

3. 普哇松分布

具有出现甲事件的概率很小，而出现乙事件的概率很大的随机均匀分布，是二项分布中概率很小时的极限形式。个体间互不干扰，一次所用样本个数很大。医学科研或流行病学分析中，习惯上凡均数 < 5 者，如镜检或液体中细菌数或白细胞数，罕见的某种非传染病发生数的分布，放射性同位素计数的数据处理等。

4. t 分布：从正态总体中抽取若干个大小相同的样本，每个样本各计算一个 t 值。如抽取的样本很多时， t 值的分布是以0为中心，两侧对称接近正态分布的曲线，称 t 分布。 t 分布为一簇曲线，当 $n' \rightarrow +\infty$ ， t 分布趋向于标准正态分布。

第二节 统计资料的收集和整理

统计工作一般分三个阶段：收集资料、整理资料、分析资料。收集资料又叫统计调查，要求及时、准确、完整。整理资料是将收集的资料进行科学的加工整理，即分组、归纳、汇总和列表、图示等。分析资料是将整理的资料计算相应的指标，在初分析的基础上进行进一步综合分析。三者密切联系，不可偏废。

一、统计资料的来源

医学统计资料的来源主要有统计报表、医疗卫生工作的原始记录、专题调查或实验三个方面。

统计报表是用表格形式，按一定时间和程序，系统地收集统计资料的组织形式，如医院工作年报、传染病、疫情、居民病伤死亡报表等。要求填写报表认真、准确、及时上报，防止漏报、错报。

医疗卫生工作的原始记录，是业务管理和科学研究的重要资料，如门诊病历、住院病历、健康检验记录等，要求认真填写，注意积累和保存。

专题调查或实验，是根据业务或科研之需，临时选用某些调查或实验方法，有明确针对性的收集资料，如恶性肿瘤的流行病学调查、某种药物疗效观察等，是医学研究中常用的方法。

统计资料的收集靠调查获得，统计调查可分为常规调查和研究性调查两种类型（表1.2）。

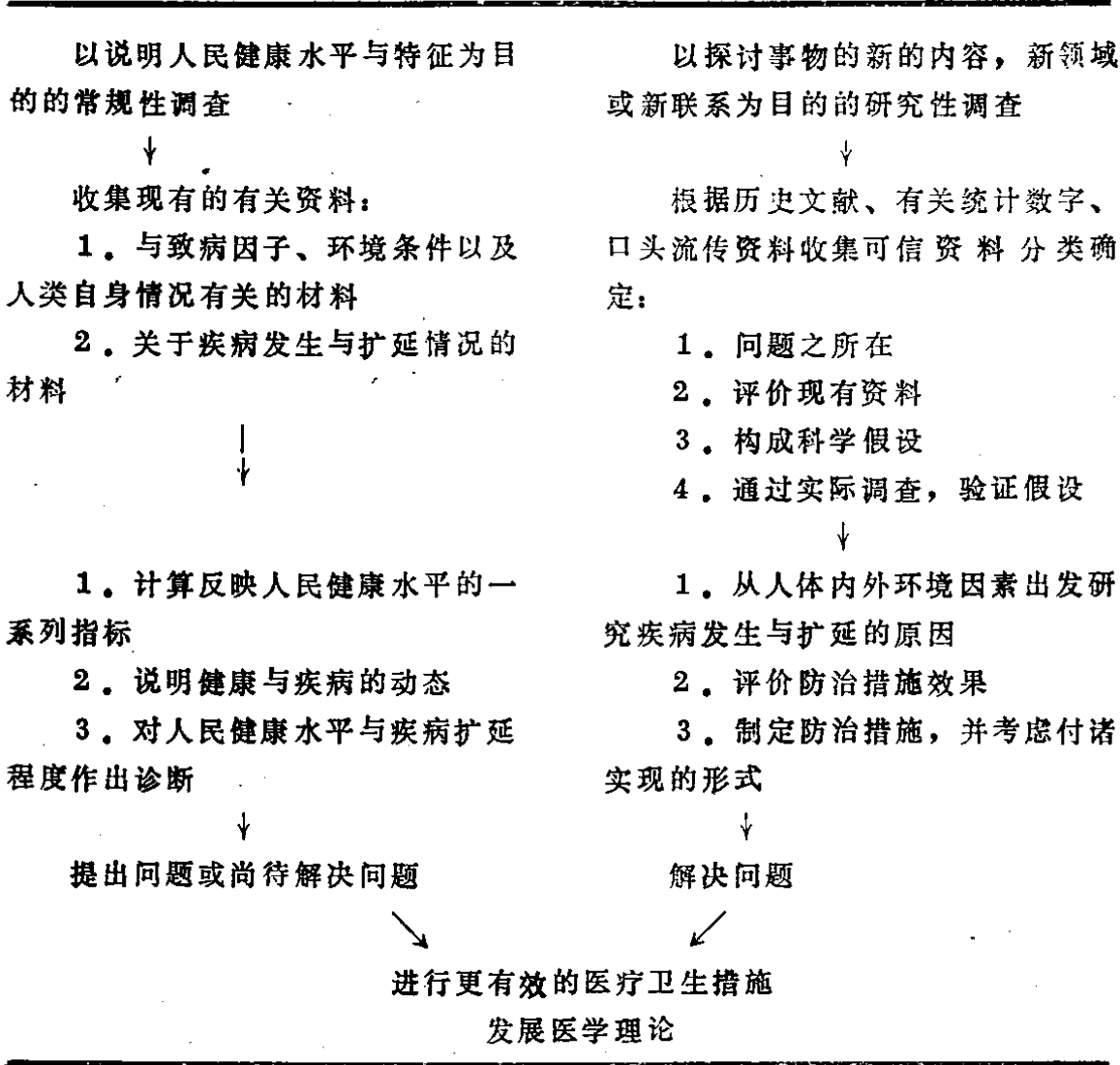
二、统计资料的整理

统计资料的整理是将收集来的原始资料进行科学的分组归纳，使资料系统化，以反映被研究事物的规律性。

整理工作包括审核资料、设计分组、拟整理表和归纳汇总四个步骤。

1. 审核资料

表 1.2 医学统计调查基本过程表解



主要是核对原始资料的准确性和可靠性。方法有两种：一是逻辑检查，即从资料的相互关系中检查是否合乎逻辑或自相矛盾；一是计算检查，即计算各横、纵行的合计有无错误。不合理的或填错的项目必须复查、补正或舍弃。

2. 设计分组

将性质相同（同质）的资料归纳到一起，使资料系统化，以反映事物的本质叫作设计分组。分组是统计工作的基础，只有坚持“同质”的原则，才能得出正确的结论。分组

可分为质量分组和数量分组两种形式。

(1) 质量分组：即按事物的性质、特征或类型分组，如疾病分类、死因分类，病人按性别、职业、工种分类，疾病或病情按轻、中、重分组，疗效按治愈、好转、死亡分组等。特点是同质、组间界限划分明确。

(2) 数量分组：在质量分组的基础上，按观察单位的数量大小来分组，如年龄、血压、脉搏、时间、身高、体重等。特点是从量的变化分析事物的差别和规律。数量分组的粗细及组数的多少，以能说明资料的规律性为准，为便于资料间相互比较，还必须注意到习惯分组法，如年龄分组由0岁开始，习惯按每5岁或10岁一组，身高多间隔2cm一组等。数量分组一般采用下限分组法，即用各组段的下限表示，如用0—，5—，10—，……等表示，由0开始到不满5者归入0—组，满5到不足10者归入第二组，余类推。

3. 拟整理表

本表是一种过渡性表格，可表达资料的分配情况和内部结构，初步显示各项目之间的联系。整理项目中要有“其他”或“不详”，“合计”或“总计”等项，以便互相核对（表1.3）。

4. 归纳汇总

一般常用手工归纳。手工归纳又分划记法和分卡法。

划记法是将调查表中同类资料逐个记入整理表中，记数的方式常用“正”字或“册”划记。其优点是经济方便、适用于少量资料，缺点是容易划错。

分卡法是将原始调查卡片按分组项目分别归组，或过录

表1.3 流脑各种病型与病情的关系（整理表）

病 型	病 情			合 计
	轻	中	重	
菌 血 型 脑膜炎型 混 型 合 不 详				
合 计				

在一张卡片上再行归组，后清点每组卡片数，填入整理表（表1.4）。

表1.4 过录卡片式样

编号				
1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20

过录卡片一般分二十个小格，每一小格代表一个标志。如第一小格示性别，第二小格示年龄，第三小格示阴性或阳性等，然后据不同小格分卡。

资料整理好后，可进一步列出统计表、制统计图、计算统计指标和进行统计分析，做出调查结论。