

医学统计方法

XUETONGJIFANGFA

中国医

学学报编辑室

数学统计方法

（第二版）

中国统计出版社

（北京）

医学统计方法

丁道芳 刘延令

大学学报编辑室

前 言

《医学统计方法》原是供我校教师开展科研实验而撰写的讲稿，曾于1974年连续两次刊载在我校的《医学研究》刊物上，受到读者们的欢迎。为了进一步满足我们师生和有关读者的需要，我们特请作者丁道芳、刘延令两同志又作了修改和增补，汇编成册，供同志们参考。由于时间仓促和水平所限，一定存在许多缺点和错误，望同志们提出批评指正。

中国医科大学学报编辑室

一九七八年六月八日

目 录

概 述	1
第一章 资料整理	4
§1 资料检查	4
§2 资料分组	5
§3 归纳汇总	7
§4 编制统计表	9
第二章 相对数	13
§1 比与率	13
§2 医学中常用的几种率	14
§3 标准化法	21
第三章 平均数与标准差	24
§1 平均数	24
§2 中位数	27
§3 众数	29
§4 标准差	30
§5 正态分布	35
第四章 抽样误差	41
§1 总体与样本	41
§2 抽样误差的概念	42
§3 抽样误差的表示方法	44
§4 置信限	45

第五章	平均数的显著性检验	47
§1	显著性检验的意义	47
§2	小样本平均数的显著性检验	49
§3	大样本平均数的显著性检验	54
§4	显著性检验应注意的几个问题	55
第六章	相对数的显著性检验	57
§1	χ^2 检验的原理	57
§2	四格表法	60
§3	二列多格表法	61
§4	交互表法	63
§5	小样本 χ^2 的校正	64
第七章	非参数统计方法	66
§1	符号检验	66
§2	秩和检验	68
§3	相关顺序检验	69
§4	Ridit 分析法	70
第八章	变异分析	75
§1	一个标识分组的变异分析	75
§2	两个标识分组的变异分析	80
§3	交互作用的变异分析	82
§4	三个标识分组的变异分析	85
第九章	相关与回归	90
§1	相关概念与回归系数	90
§2	相关系数的计算	92
§3	相关系数的显著性检验	96
§4	直线回归方程	97
第十章	正常值	101

§1 正态分布法	101
§2 直线回归法	104
§3 百分位数法	107
第十一章 半数致死量的测定	110
§1 剂量与反应	110
§2 目测法	112
§3 直线回归法	115
§4 寇氏法	117
§5 累计法	118
第十二章 疾病寿命表	121
§1 编制原理	121
§2 计算方法	122
§3 累积发病率的计算	125
§4 应用疾病寿命表中的几个问题	126
第十三章 试验设计	128
§1 试验中的三个基本组成部分	128
§2 均衡化与对照组	131
§3 随机化与随机数表的用法	132
§4 重复化与样本大小的估计	137
§5 几种常用的试验设计法	140
§6 现场调查设计	144
附表:	
1. t_i 值表	150
2. t 分布表	151
3. 二项分布置信限区间表	152
4. X^2 值表	156
5. 符号检验表	158

6. 秩和检验表	159
7. 相关顺序检验表	160
8. F 检验表	162
9. 相关系数检验表	168
10. 随机数表	169
11. 两组百分比相差显著时所需例数估计表	186
12. 平均数百分比差所需的例数估计表	185
13. 平方表	188
14. 开方表	192
15. 常用对数表	202
16. 正态分布表	206

概 述

在生产斗争和科学实验中，经常要接触许多数据，经过统计分析，可以帮助人们发现存在的问题，认识事物内在规律，是人们为进一步增加生产，提高质量而采取措施的依据。

在医学领域里，无论是卫生预防工作、临床医疗工作，还是医学科学实验，统计方法也都是不可缺少的。最为明显的事例就是医生在给病人作诊断时，都要掌握一些常用的临床检查正常值，如体温、脉搏、白细胞数、红细胞数等等。这些正常值的确定就是利用了统计方法。又如我们在鉴定或筛选某种药物的疗效时，也必须利用统计方法确定所鉴定的药物比其他药物的优越程度及其可信程度。在进行多发病、常见病和职业病的防治工作的调查研究中，统计方法是它的一个有效的、主要的方法。如我们研究老年慢性气管炎的发病率以及各种因素对它的影响，统计方法就是它的一个主要方法。至于在医学实验中的设计和数据处理，已经成为一个重要手段。实际上，我们经常运用一些统计方法来帮助我们分析问题和解决问题。因此，统计方法对我们并不生疏，但是能正确和有效地运用它来为医疗卫生工作服务却是极为重要的事，这首先决定于对它是否有一个正确的认识。

正确的数字统计，具有丰富的现实内容，它反映着客观事物的质与量，是揭示事物内部发展规律的重要依据。毛主席教导我们：“不可无根据地、主观地决定问题。”要做到“胸中有数”。“不悖得注意基本的统计、主要的百分比，不悖

得注意决定事物质量的数界界限，一切都是胸中无‘数’，结果就不能不犯错误。”做党的工作，做经济工作是如此，做医疗卫生工作也同样如此。如果我们在医疗卫生或科研工作中，统计数字和统计分析都不注意，都不成为依据，那么分析和结论要从那里来呢？我们必须是一切领域内坚持唯物论的反映论，彻底肃清唯心论的先验论，使统计为生产服务，为工农兵的健康服务。

自然界的一切事物都有其内在的规律性。可是往往由于受着许多偶然性因素所影响，使我们在观察自然界的事物时，即或在同样条件下，也出现互不一致的情况，这种情况就是我们所说的变异性。引起事物变异的原因不外是由于它们自身的变化或者由于在观察过程中环境因素的影响而引起的，不管这种变异的性质如何，它所具有的偶然性都和事物内在的必然性不可分地互相联系着，正如恩格斯所指出那样“……在表面上看去是发生着偶然性的地方，其实这种偶然性本身总是服从于内中隐藏着的规律的。”“必然性通过无数的偶然性为自己开辟道路。”这说明，科学研究工作的主要任务之一就是要从表现为偶然性的数据中整理分析出其中必然的规律。数理统计方法就是帮助人们解决这个问题的重要工具。

我们做任何一项研究，总是希望所得的结果是可靠的，重现性强的，也就是说希望在同样条件下重演时，本次所得的数据和前一次是一致的，但是实际是很难做到这一点的。例如某人进行某种药物疗效的观察，一共观察了20例病人，10例痊愈，10例未愈。在同样条件下有人有效，有人就无效。如果再行观察20例病人，还是用这种药物，其他条件也尽是相同的，那么还会出现10例痊愈10例无效吗？如果不出现这样结果又应做何解释，如何进行判断？象这样问题常使试验者

陷于迷惑，找不到一个解决问题的要领。其实这正是事物偶然性作为必然性的表现形式，它本身存在着一定的内在规律。数理统计就是透过带有偶然性的数据，说明事物内在的必然规律的一个科学分析方法。概率就是研究事物偶然性发生的可能性的一个方法。因此，数理统计是以概率论为其基础。现在我们使用的医学统计方法即属于数理统计的一个方面。

概率（机率、或然率）是统计中的一个重要概念，让我们做一个初步了解。例如我们抛掷一枚均匀的硬币，正面朝上的机会（概率）是 $1/2$ ，若抛掷次数少时，正面出现的机会不一定是 $1/2$ ，若抛掷次数多时，正面出现的机会就会接近 $1/2$ 。因此概率是某一现象客观上发生的可能性。概率的符号是 P ，概率的大小用分数、小数或百分比来表示，如以小数来表示时取值范围是 0 到 1 。某一事物必然不发生的概率为 0 ，某一事物必然发生的概率为 1 ，概率越接近于 0 ，发生的可能性越小，概率越接近于 1 ，发生的可能性越大。在医学领域里，许多生物现象是可以用统计方法来处理和研究的，这些现象的差异出现的可能性就可以用概率来表示。

统计资料大体可分为两类：一类称为计数资料，它是属于定性指标的。凡是按每个个体的某一属性来进行分类计数的都叫做计数资料。如动物实验中的生存与死亡，化验结果的阴性与阳性，临床治疗的有效与无效等。另一类称为计量资料，它是属于定量指标的。凡是以测量所得的记录，按数值大小来表示的都叫做计量资料。如身长、体重、脉搏、血压、体温、血细胞数、时间等。

计数资料和计量资料这两大类的统计分析方法都将作扼要的介绍。重点是这些分析方法的基本概念、计算方法和应用。公式的推演过程从略。

第一章 资 料 整 理

为便于分析研究，对于实验观测的数据或调查的原始资料：都应作科学的归纳与整理。资料整理包括检查、分组、归纳和列表等内容。

§1 资 料 检 查

在整理资料之前，应先对原始资料作详细、认真的检查，检查记录或登记卡的正确性和完整性。

记录经常会发生错误，错误的原因有两种：一种是偶然的错误，这是由于记录者一时的填写错误引起的，这种错误对结论的影响较小；另一种是一贯的错误，这是由于记录者对填写要求的错误理解所引起的，这种错误对结论的影响较大，有时甚至是很严重的。检查是减少错误的一个程序，而不是避免全部错误的方法。检查资料有无错误的方法，一般有下列两种：

1. 逻辑检查：即查明资料项目之间有无矛盾。例如，性别栏内填写“男”，而在死亡原因栏内填写“宫颈癌”；又如，年令栏内填写“8岁”，而在身长栏内填写“175厘米”，象这样的错误很容易利用此法检查出来。实验数据中出现一些突出的大或突出的小，也是不难发现的。

2. 计算检查：即利用已经制好的表格中横行相加于右侧合计，纵列相加等于下面合计，纵横合计相加均等于总计

的方法进行验算。

对于错误和漏项的资料，应及时补正，如无法改正或补填的资料，则应作废。对观测值出现个别值能否舍弃，应按下述方法决定取舍。

例 1.1 用分光光度计测定血中转氨酶标准管的光密度，5 次测定值分别为：

0.15, 0.15, 0.17, 0.19, 0.28

其中 4 个值比较接近，均在 0.19 以下，唯有 0.28 显著过高，此值是否应舍弃，按下列步骤决定此值是否有效。

(1) 计算 5 个值的平均数 ($\bar{x}$)

$$\bar{x} = \frac{0.15 + \cdots + 0.28}{5} = 0.188$$

(2) 求极差 (R)

$$R = |0.15 - 0.28| = 0.13$$

(3) 求可疑值与平均值的差数绝对值，并除以极差。

$$\begin{aligned} t_i &= |x - \bar{x}| / R \\ &= |0.28 - 0.188| / 0.13 \\ &= 0.71. \end{aligned}$$

(4) 将算得的 t_i 值与附表 1 的 t_i 值比较，如 t_i 大于表值，则可舍弃，查表 $n=5$ 的 $t_i=0.86$ ，令算得的 $t_i=0.71$ ，小于表值，故 0.28 不应舍弃。

§2 资 料 分 组

按照研究现象的一定的标识将调查总体分成各组称作分组。分组能使原始资料在归纳的过程中更加系统化，更加便于揭明研究对象的成分和结构。如我们研究恶性肿瘤在人群

中发病情况。我们如果只求得一个发病率来说明发病情况，既不够全面又不够深刻，远远不能揭示它的发病规律。如果我们按地区分布，年令和性别分布，民族、生活习惯与癌的关系，职业与癌的关系等进行分组分析，那么就会发现许多有价值的、有规律性的东西。

分组在统计方法中占有重要地位。只有在同质的基础上进行分组，才能得出正确的结论。目前对许多多发病、常见病和职业病都有诊断标准，如矽肺、慢性气管炎、冠心病等。这样就为正确诊断和统一分期带来了良好基础，如果没有统一标准，就会出现病型、期别、程度的混乱情况，即所谓非同质集合的错误。所以，统计分组决不是单纯的统计数字问题，应当按照事物自身的本质分组，才能正确地反映客观的真实面目。

资料分组方法可分为品质分组和数量分组两种。

品质分组：是按被研究对象品质标识来分组，如性别、职业、疾病分类、死亡原因、文化程度、实验因素等级等等。

数量分组：按质的差异分组之后，在各组中又包含许多变量（观测值），为了便于计算又必须把这些变量按其大小顺序排列成变量数列。在进行数量分组时，组数不可过多，过多会使计算麻烦，但数也不可过少，过少会使计算结果误差变大。现举一年令分组为例：

0 ~	0
1 ~	1 ~ 4
5 ~	5 ~ 9
10 ~	10 ~ 14
15 ~	15 ~ 19
20 ~	20 ~ 29

30~	30~39
40~	40~49
50~	50~59
60~	60以上

上述两种年令分组表示法是相同的，但应注意归组问题，上例右侧分组0岁应包括未满一周岁的婴儿，满一周岁的直到不满5周岁的均归入“1~4”岁组，满5周岁到不满10岁的则归入“5~9”岁组内，余类推。左侧分组：0~表示出生到不满一周岁，“1~”表示满一周岁到不满5周岁，“5~”表示满5周岁到不满10周岁。“0~”下限为0，上限为1，“1~”下限为1，上限为5。统计上数量分组一般均采用左侧的分组法。

§3 归 纳 汇 总

归纳原始资料的方法有两类：手工归纳和机械归纳。手工归纳常用的有划记法和卡片分类法两种。

1. 划记法

此法是将记录簿或调查表中项目标识相同的事实，用划记法分别列入分组整理表内，然后合计。此法经济、简便，但容易划错，划错后又不易复查，必须重新开始，因此它只适用于简单的一览表。

例1.2 某地区勾虫病按年令分组的发病数。先分年令组，然后再按原始资料的发病数分别划在划记栏内，最后就可以归纳出各年令组的发病例数（或频数），这种表统计上称为频数表。

表 1.1 某地区勾虫病按年令分布频数表

年 令 (岁)	划 记	例数 (频数)
1 ~	正正正	14
7 ~	正正正正正正	31
15 ~	正正下	13
25 ~	正正丁	12
35 ~	正正丁	12
45 ~	正正	10
合 计		92

2. 卡片分类法

此法是将原始的单一调查表或卡片按分组把同类的标识归在一起, 再将计数结果填在整理表上。如果原始调查表或记录簿是将受检对象的均记在一览表式记录上, 或内容过多时, 可将一览表的每个观察单位过录到一种过录卡片上。过录卡片上分成若干个空白小格, 在过录时要统一各小格代表的标识, 如勾虫调查是用登记在一览表的形式, 要改用卡片分类法, 这里以第 1 格代表队别, 第 2 格代表性别, 第 3 格代表年令, 第 4 格代表职业, 第 5 格代表勾虫卵检查

编号: 001

1 7 队	2 男	3 12 岁	4 学生	5 勾 (-)
6 蛔 (+)	7 鞭 (-)	8	9	10
11	12	13	14	15
16	17	18	19	20

死亡率还包括年令别死亡率、死因别死亡率、婴儿死亡结果，第6格代表蛔虫卵检查结果，第7格代表鞭虫卵检查结果，其余各格子均代表一项内容。过录在过录卡片上，分组时即不必用划记法归纳，可用卡片记数分组了。

§4 编制统计表

统计表就是以表格的形式表示整理分组的统计数字，通过统计表可以更清晰地表明统计数字，并可以从多方面系统地比较分析现象的数量关系和数量变化的规律。

1. 统计表的基本结构

统计表是由标题、横标目、纵标目和表体四个部分组成。

标题，它是表的名称，要根据表的内容标出，标题置于表的上方。

横标目：位于表的左侧，这个位置一般做为主项。如表1.2的病名即表的主项。

纵标目：位于表的上端，这个位置一般做为宾项。如表1.2的例数，表1.3、表1.4的年令与性别。

表体：是记载数字的位置，统计表的表体应全部记载数字而不应参杂文字。

2. 统计表的种类

统计表通常有简单表、分组表和组合表三种。

简单表：它的结构简单，是一览表的形式，一个主项和一个宾项。表1.2即简单表，主项为病名，宾项为例数。