

实用临床流行病学

刘春望 主编

山东科学技术出版社

87
R18
16

BT271.0

实用临床流行病学

主编 方春望

编写 (以姓氏笔画为序)

李玉清 李绍忱

李 笠 张桂宁



山东科学技术出版社

一九八六年·济南

实用临床流行病学

方春望 主编

*

山东科学技术出版社出版

山东省新华书店发行

山东新华印刷厂临沂厂印刷

*

850×1168毫米32开本 11.125印张 2插页 238千字

1986年 12 月第1版 1986年 12 月第1次印刷

印数: 1—4,120

书号 14195·249 定价 2.95元

前 言

科学的进步，使人们的眼光开扩了，对于过去没有解决的问题，或解决不完全的问题，进行新的探索和研究，如此促使科学不断进步。过去临床工作者研究对象是以生物学的观点和方法研究个体，主要从微观方面进行研究。但人是社会的人，处在复杂的各种社会关系（包括自然环境）之中，受其影响和约束。所以单从生物学的观点和方法研究个体是远远不够了。临床流行病学的研究对象是人群，从宏观上对人群进行研究，丰富了临床医学。

临床流行病学是临床工作者需要掌握的一门基础科学，它指导临床工作者应用流行病学和医学统计学的基本原理和方法去发现和更好地解决临床问题，如病因、诊断、治疗、预后等，提高解决这些临床问题的能力。提高临床工作者的评价能力，也是临床流行病学的重要任务。评价包括对临床论据的评价，诊断和筛检试验的评价，治疗效果的评价，保健措施的经济效益的评价，医学文献的评价，等等。

本书的读者对象是一般临床工作者。为了使读者更好地理解临床流行病学，在前四章中介绍了医学统计学的基本知识和流行病学的基本知识，同时，对有关评价都有专门章节予以介绍。

随着临床流行病学的普及，必将有助于医疗质量的提高，
使临床医学更好地服务于人类。

方春望

1985年9月

于山东医科大学

目 录

第一章	医学统计基本知识和方法	(1)
第一节	几个基本概念	(1)
第二节	均数、标准差及标准误	(4)
第三节	显著性检验	(17)
第四节	机会	(33)
第五节	样本大小	(39)
第二章	医学研究常用指标	(54)
第一节	率	(54)
第二节	比	(70)
第三节	危险性	(75)
第三章	疾病病因	(85)
第一节	病因概念及范畴	(86)
第二节	病因模式及多病因概念	(95)
第三节	病因确立	(97)
第四章	流行病学研究方法	(107)
第一节	现况调查	(108)
第二节	病例对照研究	(124)
第三节	队列研究	(139)
第五章	异常的测量和判定	(156)
第一节	临床资料的种类和测量方法	(156)
第二节	正常范围的确定方法	(160)
第三节	测量的变异性	(168)

第四节	测量的质量控制·····	(174)
第五节	测量的质量评价·····	(178)
第六章	诊断试验·····	(184)
第一节	诊断试验常用指标·····	(185)
第二节	诊断试验评价·····	(208)
第三节	诊断试验中的偏倚·····	(215)
第七章	诊断策略·····	(217)
第一节	临床工作中的诊断策略·····	(217)
第二节	筛检·····	(227)
第八章	药物疗效和不良反应的流行病学研究·····	(242)
第一节	影响药物疗效的人群因素·····	(242)
第二节	药物不良反应的流行病学研究·····	(250)
第九章	临床试验·····	(264)
第一节	临床试验的意义和分期·····	(264)
第二节	临床试验设计·····	(266)
第三节	临床试验观察·····	(280)
第四节	临床试验结果分析·····	(284)
第十章	疾病预后的群体研究·····	(296)
第一节	临床经验判断预后的局限性·····	(296)
第二节	疾病的自然史·····	(298)
第三节	研究疾病预后的指标·····	(300)
第四节	应用流行病学方法研究预后·····	(310)
第十一章	医院内感染·····	(317)
第一节	医院内感染的危害和长期趋势·····	(317)
第二节	医院内感染的流行特征·····	(320)
第三节	医院内感染的调查与分析·····	(326)
第四节	医院内感染的预防措施·····	(328)

第十二章	怎样阅读医学文献	(333)
第一节	医学文献的类别和论著的结构	(333)
第二节	医学文献中的研究方法及其规范	(337)
第三节	提高阅读医学文献的效率	(340)

第一章 医学统计基本知识和方法

在科学研究中，恰当地应用数理统计的方法，可以帮助我们透过偶然性来揭示必然性。在医学研究中，比较和鉴别是常用的一种手段。无论从实验还是从观察获得的一宗数据，通过统计处理和检验，具体应用到研究中去，均有助于对结果作出比较正确的结论。因此，统计学知识是医学研究必不可少的工具。近年来，统计学方法应用于临床医学研究的范围越来越广。例如，临床试验设计、病因研究、诊断程序、疗效试验、实验室测定、疾病筛检的评价，等等。本章将只是选择临床医生最常遇到的、最基本的知识和方法作一扼要地阐述（在其他章节中将要讨论的内容，本章从略），每个公式都通过实例来介绍，以使阅者了然不惑，并作为开展临床医学科研工作的参考。至于较深的、复杂的统计学理论和方法，必要时请参阅有关数理统计学专著。

第一节 几个基本概念

一、总体与样本

同质的个体所组成的集体，称为总体（population）。例如所有冠心病病人，都具有患冠心病这个同质的特征，是一个总体。由于冠心病包括无症状、有非典型心绞痛、有典型心绞痛及心肌梗塞等各种临床类型病人，冠心病病人总体则是各种临

床类型冠心病病人的全体。冠心病病人的个体数是相当大的。任何一位临床医生都不可能也不需要把所有的冠心病病人进行研究。实际上，经常是对所抽取的数十例或数百例病人进行研究。从该病病人总体中随机抽取数十例乃至数百例，也只是总体中一个极小的部分。研究所用的这数十例或数百例病人，在统计上称为样本（sample）。同样，也不可能和不需要对所有的健康人进行研究，实践中经常是选择某一特定的人群进行研究，所研究的这个特定人群称为目标人群（target population）。目标人群一词是很灵活的。譬如，要研究某市公共汽车公司职工的冠心病，目标人群就是该公司的职工全体；如果确定以驾驶员和售票员为研究对象，则它就是该公司的驾驶员和售票员全体；假若定研究对象为35岁及以上的男驾驶员，那么，目标人群就是该公司35岁及以上的男驾驶员全体了。

医学研究的基本任务在于从样本推论总体。由于人类的疾病是复杂的，即使患同一种病，不同病人的病情、病程、接受治疗的情况、有无合并症等也各异，而且个体差异很大。因此，必须注意样本的选择。只有样本在本质特征上与总体相一致或近似，才能够代表总体。为了保证样本具有代表性，必须：

- ①按随机化原则抽取样本，就是说在抽取样本以前要使总体中每个单位有同等的被抽取的机会。同样，在分组时，也必须用随机的方法，使各个病人有同等的机会被分配到各组中去。
- ②有足够大的样本含量（详见本章第五节）。
- ③要确定总体的含义，例如病例必须严格地依据该病准确可靠的诊断标准来选取。
- ④病例组应当包含有与该病疾病谱相应比例的各种临床类型病人和各期病人。

二、对照与对照组

人们要想研究心肌梗塞的发病与吸烟是否有联系，既需要从心肌梗塞病人总体中抽取一个样本，与此同时，还需要从健康人（无心肌梗塞者）中抽取一个样本进行比较。否则，就无从得知心肌梗塞与吸烟之间是否存在联系。若要想研究一种新疗法是否比旧疗法或另一种疗法好，亦需进行比较。用来作比较的非心肌梗塞者，或采用旧疗法治疗的病例，称为对照。用作比较的一组非心肌梗塞者，或采用旧疗法治疗的病例组，称为对照组。

相互比较的两组或各组，除所研究的因素不同外，其他因素或特征，如年龄、性别、临床类型等，应当尽量使之相似，才有可比性。否则，由于某种混杂因素的存在，会使得所获的比较结果和结论是不可靠的。为了使相互比较的两组或各组均衡，病例组和对照组的选择都应当是随机抽取的；或者，按每一例病人的主要特征，即能影响所研究因素与疾病之间的联系的特征，如年龄、性别、职业、临床类型、病程等，选择对照者，此即所谓配比（matching）。

三、抽样误差

由于个体差异，从同一总体中随机抽取若干个含量相同的样本，这些样本与总体的均数（或率）之间，各个样本的均数（或率）相互之间，都会有或大或小的差异，这种差异是由于抽样造成的，称为抽样误差。抽样误差是不可避免的。一般说来，样本越大，抽样误差越小，所测得的数据越接近总体的情况。但在临床医学研究中往往不可能观察太多的病例，所研究的病例数经常较少。样本越小，抽样误差相应地越大。如果仅就表面的数值来看，不考虑机率，往往会导致判断错误。所以

需要用统计方法（如各种显著性检验）予以辨别，以便得到一个比较正确的结论。

四、概率

概率（亦称机率）这个词用来表示某个结果或事件出现的可能性大小。如果某个试验在相同的条件下重复进行多次，那么，该事件的概率就是此事件出现的次数与试验次数之商所趋向的那个值。例如我们说某药治疗某病的治愈率为90%（或0.9），就是说某药治愈某病的概率估计为0.9。概率通常用符号“ P ”来表示。必然发生的事件之概率为1，不可能发生的事件之概率为0。换句话说就是 $0 \leq P \leq 1$ 。概率越接近1，表明发生的可能性越大。概率越接近0，表示发生的可能性越小。 $P < 0.05$ ，意思是该事件出现的概率小于0.05。

第二节 均数、标准差及标准误

一、均数

均数有许多种，其中最常用的是算术均数和几何均数。

（一）算术均数：习惯上，以 μ 表示总体均数，以 $\bar{x}$ 表示样本均数。

1.小样本资料的均数算法：各个变量值的总和除以变量值的个数，即得均数。以公式表示即：

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x}{n} \quad (1-1)$$

式中 $\bar{x}$ 为均数； $x_1, x_2, x_3 \cdots x_n$ 为各变量值； $\sum x$ 为各变量值的总和； n 为变量值的个数。

例1-1：32名健康成年男性血红蛋白值（g/dl）分别为：

14.8, 15.4, 15.5, 13.7, 14.4, 14.1, 14.4, 14.4, 15.1, 15.3, 14.2, 14.8, 14.9, 14.3, 12.8, 13.4, 15.6, 14.6, 14.7, 14.4, 13.8, 15.4, 14.5, 16.4, 15.2, 12.5, 15.9, 15.5, 14.4, 17.0, 15.2, 14.4。

将各数值代入公式(1-1)，得：

$$\bar{x} = \frac{14.8 + 15.4 + 15.5 + \cdots + 14.4}{32} = \frac{471}{32} \\ = 14.72 \text{ (g/dl)}$$

2.大样本资料的均数算法：当变量值个数较多时，可先制成频数分布表，用简捷法计算均数。现举例说明。

例1-2：表1-1是根据110名男大学生血红蛋白的测定值，用画记法制成的频数分布表。

表 1 - 1 110名男大学生血红蛋白值
的频数分布表

血 红 蛋 白 (g / dl)	频 数 <i>f</i> (人数)
11.0~	2
11.5~	3
12.0~	5
12.5~	7
13.0~	13
13.5~	16
14.0~	19
14.5~	18
15.0~	14
15.5~	6
16.0~	4
16.5~16.9	3
合 计	110

简捷法计算均数步骤如下：

(1) 先列出计算表 (表1-2)。

表 1 - 2 均数的简捷法计算表

血 红 蛋 白 (g / dl)	<i>f</i> (频 数)	<i>x</i> (简 化 值)	<i>f x</i>
11.0~	2	- 6	- 12
11.5~	3	- 5	- 15
12.0~	5	- 4	- 20
12.5~	7	- 3	- 21
13.0~	13	- 2	- 26
13.5~	16	- 1	- 16
14.0~	19	0	0
14.5~	18	1	18
15.0~	14	2	28
15.5~	6	3	18
16.0~	4	4	16
16.5~16.9	3	5	15
合 计	110 (Σf)		- 15 ($\Sigma f x$)

表中 *x* 栏是以频数最大的一组之组中值作为假定均数 (x_0)。本例最大频数为19，该组的组中值为 14.25，即 $x_0 = 14.25$ (g / dl)。以 x_0 所在组的 *x* 为 0。当各组组距相等时，如变量值从上到下是由小到大排列，则从 0 向上依次为 - 1，- 2，- 3，…从 0 向下依次为 1，2，3 …

(2) 各行的 *f* 乘 *x* 得 *f x*，各行 *f x* 之和为 $\Sigma f x$ 。

(3) 按上表求算术均数的公式如下：

$$\bar{x} = x_0 + \frac{\Sigma f x}{\Sigma f} \quad (1 - 2)$$

式中*i*为组距，本例*i* = 0.5。将表1 - 2中数据代入公式（1 - 2），得：

$$\bar{x} = 14.25 + \frac{-15}{110} \times 0.5 = 14.18 \text{ (g/dl)}$$

（二）几何均数：当一组计量资料的各变量值之间呈倍数（等比）关系时，应按下列公式计算几何均数（ $\bar{x}_G$ ）：

$$\bar{x}_G = \lg^{-1} \frac{\sum f \lg x}{\sum f} \quad (1 - 3)$$

式中lg*x*为变量值的对数，

$\sum f \lg x$ 为同组频数与变量值对数的乘积之和，其余符号同前。

例1-3：测定济南市郊区444人血清标本甲₃型流感血凝抑制抗体的滴度，其分布如表1 - 3。

表 1 - 3 济南市郊区人群甲₃型流感血凝抑制抗体的滴度分布（1983年）

血凝抑制抗体的滴度	人 数
<1:5	71
1:5	122
1:10	84
1:20	54
1:40	47
1:80	40
1:160	26
合 计	444

几何均数计算步骤如下：

1. 将表1-3改列如表1-4。

表 1 - 4 几何均数计算表

稀释度 (x)	$\lg x$	f	$f \lg x$
1	0	71	0
5	0.6690	122	81.6180
10	1.0000	84	84.0000
20	1.3010	54	70.2540
40	1.6020	47	75.2940
80	1.9031	40	76.1240
160	2.2041	26	57.3066
合 计		444	444.5966

表1-4的第一栏为稀释度，第二栏为第一栏相应数值之对数，第四栏为第二栏、第三栏同组之乘积。

2. 将表1-4中有关数值代入公式 (1-3)，得：

$$\overline{x}_G = \lg^{-1} \frac{444.5966}{444} = \lg^{-1} 1.0013 = 10.03$$

该人群甲₁型流感血凝抑制抗体的几何平均滴度为 1 : 10。

二、标准差

标准差是表示一组变量值离散程度的指标。标准差越小，表示各变量值越接近均数；标准差愈大，表示各变量值离均数愈远。均数和标准差结合起来看，就可更全面地表示一组变量值的分布情况。

(一) 小样本资料的标准差算法：小样本的标准差可按下列公式计算。

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}} \quad (1-4)$$

式中 S 为标准差； $x - \bar{x}$ 为各变量值和它们的均数之差，称离均差； $\sum (x - \bar{x})^2$ 为离均差平方和。

现仍以例1-1资料为例，代入得：

$$S = \sqrt{\frac{6959.6 - \frac{471^2}{32}}{32 - 1}} = 0.9344(\text{g/dl})$$

(二) 大样本资料标准差算法：对大样本分组资料，可于表1-2内右边添列一个 fx^2 栏，即可按下列公式计算标准差。

$$S = (i) \sqrt{\frac{\sum fx^2 - \frac{(\sum fx)^2}{n}}{n - 1}} \quad (1-5)$$

公式(1-2)和本式中的 $x = \frac{x - x_0}{i}$ ，实际上是一个简化值。 $\sum fx^2$ 是各组频数与简化值平方乘积之和。现仍以例1-2的110名男大学生血红蛋白值资料为例，将表1-2改为表1-5。

表1-5 计算标准差用表

血红蛋白 (g/dl)	频 数 (f)	简 化 值 (x)	f x	f x ²
11.0~	2	- 6	- 12	72
11.5~	3	- 5	- 15	75
12.0~	5	- 4	- 20	80
12.5~	7	- 3	- 21	63
13.0~	13	- 2	- 26	52