

J.M. 英格兰 著

医学研究中统计学与流行病学方法

.1

人民卫生出版社

R195.1

4
3

医学研究中 统计学与流行病学方法

J. M. 英格兰 著

王 仁 安 译

田 凤 调 校



人 民 卫 生 出 版 社

A656207

内 容 简 介

本书为医学科学研究中应用统计学和流行病学方法的通俗著作。原书第一版于1975年出版。

书中对数理统计方面的最基本原理结合医学科学研究中的实际例子作了深入浅出的叙述。书中从临床试验与流行病学调查的研究设计到数据处理与结果解释都作了指导性的扼要介绍。

全书的重点是放在解决实际问题上，同时收集了一些很有实用价值的统计方法。

本书可供从事基础与临床医学、流行病学方面的研究人员、医学院校师生学习与参考。

J. M. England

Medical Research

A Statistical and Epidemiological Approach

CHURCHILL LIVINGSTONE

EDINBURGH LONDON AND NEW YORK 1975

医学研究中统计学与流行病学方法

王 仁 安 译

人民卫生出版社出版

北京通县印刷厂印刷

新华书店北京发行所发行

787×1092毫米32开本 5 $\frac{3}{4}$ 印张 123千字

1980年4月第1版第1次印刷

印数：1—11,600

统一书号：14048·3821 定价：0.55元

序 言

“统计学的教学会有助于对医学生的培养，这表现在以下两个主要方面：首先，这门学科是科学方法逻辑学的一个完整部分，并且会很自然地引出关于进行观察与解释观察结果以及实验的某些概念。其次，统计学包含对变异的测定与评价的一套方法，这些方法越来越广、越来越多地用于医学研究的许多方面，如用于关于诊断程序、疗效、新药的进展、致病因素、实验室测定及其他许多方面的课题。”

“掌握一些统计方法的基本知识是必要的，这样可以使医生们自己对发表在杂志与其他通讯上的关于医学进展的某些主张的有效性作出判断。”

皇家医学教育委员会 1965~68

皇家医学教育委员会与社会医学协会强调医学统计与流行病学对医学生与开业医生的重要性。上面的引语概括了主要理由，我在这里不再重复。

最近，联合王国皇家医师进修学院对许多明显缺乏统计基本原理知识的大学毕业生表示关心，现在对他们的考试包括了统计问题。医学研究人员也已受到批评。如《临床科学》与《分子医学》杂志编者在其“致作者”中已作过评论，“因为作者统计方法使用不当，文章经常被退回修改(并因此延误发表)。”

我希望本书对医学研究中统计与流行病学的最重要方面作一个概括介绍。本书前两章解释基本的统计名词，以后几章告诉读者如何作调查研究的设计(第三章)，需要多少研究

例数（第四章）与如何分析结果（第五～八章）。后几章为流行病学与临床试验，最后一章是关于评价一篇论文的几点意见。

由于大多数医生不喜爱数学，因此，我在书中力求减少数学，把大多数公式与计算放在本书最后的附录中。举一些计算的例子是为了使读者可以核对自己计算的正确性。当然，很多计算还是麻烦的，但可以用现有的台式电子计算机来代替笔算。

附录中还包括了一些较深入的材料与统计数值表。

目 录

第一章 重要名词介绍.....	1
总体、样本与变量	1
位置量数与离散度指标	2
无效假设与显著性检验	7
实际显著性与统计显著性	8
可信限	8
第二章 概率定理	9
贝努里分布	9
泊松分布	12
正态分布	14
正常范围	18
对数正态分布	20
“Student”t 分布.....	22
第三章 调查研究设计	25
检验何种假设?	25
研究哪些对象?	26
测定什么?	34
如何组织调查研究?	36
第四章 调查研究需要多少例数?	41
样本大小的确定	41
收集资料进程中对结果的分析	49
序贯分析或获得每个结果后的分析	51
似有显著性差异时的结果分析	56
总结	57
第五章 结果分析: 比例数.....	57

检验一个比例数是否等于一个特定值	57
两比例数的比较	60
第六章 结果分析：平均数	63
引言	63
检验样本均数是否等于一个特定值	66
检验两样本均数之差	67
第七章 结果分析：相关与回归	72
引言	72
相关与相关系数 r	77
回归线的配合与相关系数的计算	79
可信限与显著性检验	80
何时使用相关分析?	83
相关分析的结果意味着什么?	87
回归的特殊类型	91
第八章 几个显著性检验的结果分析	91
第九章 流行病学：一些方法	95
引言	95
患病率、发病率与死亡率	96
横断面研究	97
队列研究	100
前瞻性发病率研究	101
寿命表或生存率曲线	102
第十章 流行病学：一些结果	106
描述流行病学	106
分析流行病学	109
传染病的分析研究	109
新生物疾病与变性疾病的分析研究	112
筛选检验	117
第十一章 临床试验	118

对照试验	118
盲目对照试验	119
试验的组织	120
第十二章 怎样阅读文献	122
参考书目	123
参考文献	123
附录	131
1. 方差的和、积, 等等, 及误差的传播	131
2. 位置量数及其应用	132
3. 对数正态分布	134
4. 如何配合贝努里分布、泊松分布、正态分布、对数正 态分布、以及如何检验配合的适度	135
5. 比例数的显著性检验与可信限	141
6. 均数的显著性检验与可信限	144
7. 相关与回归	145
8. 显著性检验	152
9. 同位素计数	156
10. 标准化死亡比的计算	158
11. 官方统计与记录的来源	160
12. 以人为研究对象的责任	160
13. 剂量反应曲线	167
14. 正态分布	169
15. t 、 χ^2 与 r 表	169
16. 实用的统计数值表	171
17. 第二类标准化误差	173

第一章 重要名词介绍

总体、样本与变量

在研究工作中，不可能对某种疾病的所有病人，也不可能对每一个正常者都进行研究。这些数量上设想为无限大的群体，在统计上称为总体。研究者只得满足于从任一总体中选一小部分人进行研究，而被研究的这些人就组成一个所谓**样本**。

例如，人们要想研究动脉粥样硬化症的发病率或严重程度在糖尿病人中是否增高，则须从糖尿病人的总体中抽取一个样本，并与正常者的样本结果进行比较。这种为比较目的而研究一组正常者，是必不可少的，人们把这组正常者称为**对照**。

选择了糖尿病和正常者两个样本以后，必须作一些测量，以此作为动脉粥样硬化的指标。测量得到的参数称为**变量**。我们大概可以测定尸解标本中主动脉的胆固醇，并以每克主动脉中胆固醇的毫克数表示测定结果。如果测定值围绕最接近的毫克数或上或下，则测定的单位尺度是非连续的，这些测定结果就成为**非连续变量**。如果非常精确地表示测定结果，比如说，精确到最接近百万分之一毫克，这样实际上单位尺度就是连续的了，我们就把测定结果称为**连续变量**。

估计动脉粥样硬化严重度的另一种方法或许是简单地记录存在或不存在外围脉搏(例如腘窝)。这就是所谓**有或无变量**。

位置量数与离散度指标

变量的性质决定其分析结果的方法。我们现在来考虑一下血红蛋白值这样一个典型的连续变量。图 1·1 表示 32 个正常者的血红蛋白值。如果我们想要概括本图资料，则首先必须说出这些测量结果是否位于血红蛋白值可能的范围内。从统计上说，我们要一个**位置量数**。其次，我们必须知道测量结果有多大变异，任何估量变异大小的数值称为**离散度指标**。有了位置量数加上离散度指标就可以全面概括测量结果。

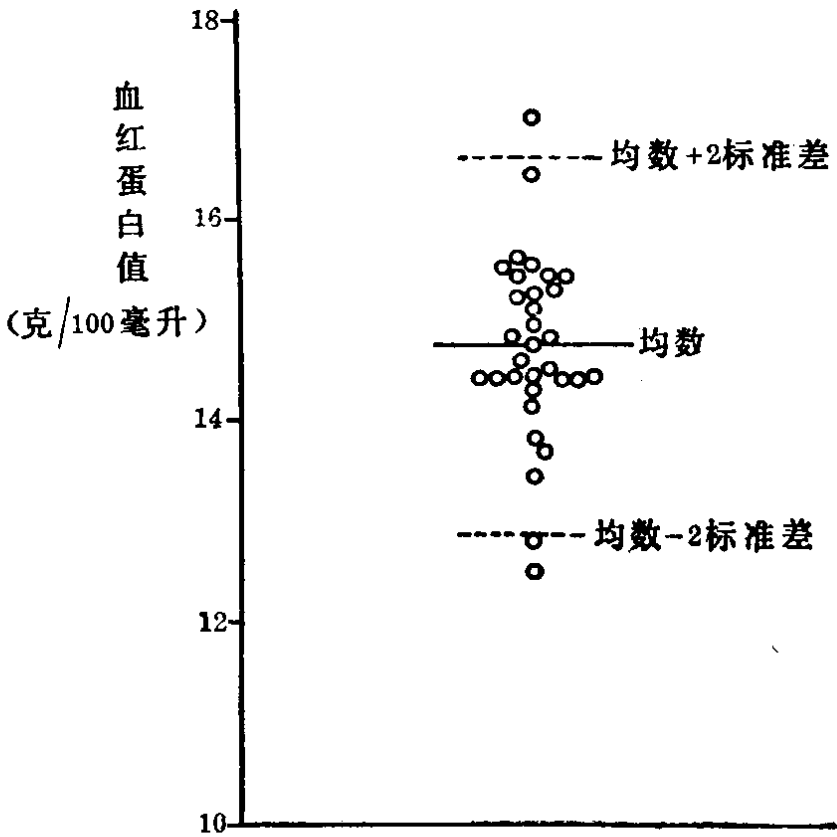


图 1·1 32名正常成年男性血红蛋白值
均值：14.7克/100毫升，标准差：0.9克/100毫升

平均数或算术**均数**是最重要的位置量数。定义是：

$$\text{均数} = \frac{\Sigma X}{n} = \bar{X}$$

希腊字母西格马， Σ ，意即总和， ΣX 表示所有 X 值的总和。 ΣX 被观察数值个数 n 除，得 X 的均值，符号 $\bar{X}$ 。

最常用的离散度指标是**标准差**，符号是 S 。标准差也称**根均差方**。它可以告诉我们：标准差是如何得出来的。

对每个 X 值，计算其与真实的总体均数 μ 的离差。注意：总体均数与样本均数 $\bar{X}$ 是不同的。小写希腊字母通常用来表示总体值，而普通字母表示样本值。离差是：

$$(X - \mu)$$

差方是：

$$(X - \mu)^2$$

均差方是：

$$\frac{\Sigma (X - \mu)^2}{n}$$

根均差方或标准差是：

$$\sqrt{\frac{\Sigma (X - \mu)^2}{n}}$$

由于我们不知道 μ 值，就不能用这个公式来计算标准差。我们对 μ 值所能有的最好估计值是样本均值，但用 $\bar{X}$ 代替 μ 常常会低估标准差。这是由于 $\bar{X}$ 与 μ 常有差异，计算 $\bar{X}$ 所包含的测定结果的离差会小于 μ 所包含的测定结果的离差。样本标准差就会小于总体标准差。考虑到样本标准差的偏小，因此，差方和就除以 $(n - 1)$ 以代替 n 。

$$\text{标准差 } S = \sqrt{\frac{\Sigma(X - \bar{X})^2}{(n-1)}}$$

这就是公式中用 $(n-1)$ 的理由, $(n-1)$ 也称为自由度(“自由度”是一个难理解的概念, 愿知者, 可参考附录 8 中的解释)。

求得 $\bar{X}$ 与 S 后, 可以计算其他三个指标。

$$\text{方差} = S^2$$

$$\text{变异系数} = C. V. = \frac{S}{\bar{X}} \times 100\%$$

$$\text{均数的标准误} = S. E. M. = \frac{S}{\sqrt{n}}$$

例: 图 1·1 32 名正常成年男性血红蛋白值 克/100 毫升: ——

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。

计算

$$\Sigma X = 14.8 + 15.4 + 15.5 \cdots + 17.0 + 15.2 + 14.4 = 471$$

$$\begin{aligned} \Sigma X^2 &= (14.8)^2 + (15.4)^2 + (15.5)^2 \cdots + (17.0)^2 \\ &\quad + (15.2)^2 + (14.4)^2 = 6959.6 \end{aligned}$$

$$n = 32$$

则

$$\text{平均数} = \bar{X} = \frac{\Sigma X}{n} = \frac{471}{32} = 14.7188$$

$$\begin{aligned}\text{标准差} &= \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}} \\ &= \sqrt{\frac{6959.6 - \frac{471^2}{32}}{32-1}} = 0.9344\end{aligned}$$

$$\text{方差} = S^2 = 0.8731$$

$$\text{变异系数} = C. V. = \frac{S}{\bar{X}} \times 100\%$$

$$= \frac{0.9344}{14.7188} \times 100\% = 6.3483\%$$

$$\text{均数的标准误} = S. E. M. = \frac{S}{\sqrt{n}} = \frac{0.9344}{\sqrt{32}} = 0.1652$$

方差是一个很有用的指标，常常要计算它，用到它。例如，我们想要把两组资料相加，看看合并后总的变异如何，只要把两个方差相加就可以知道。关于两组测量资料的差、和、积等的方差可参考附录 1。

变异系数（通常用百分数表示），一般都可以，而且也应该计算这个指标。如果变异系数（C. V.）超过 12%，分析数据时，很可能出现问题。（见第二章对数正态分布）。C. V. 也是重要的指标，因为它不受资料所用的单位的影响。血红蛋白值可以用克/100 毫升血也可以用克/升血表示，但它们的变异系数是一样的。

均数的标准误比标准差小。它来自标准差，把标准差除以计算均数时所用的观察数的平方根即得。可以想象：若对一定大小的样本进行重复抽样，并计算每个样本的均数，

S. E. M. 就是样本均数的标准差。S. E. M. 较小则意味着样本均数是它抽样的总体均数较精确的估计值。只有使观察数的平方根减小，才能降低 S.E.M.，认识到这一点是很重要的。表 1·1 所表明的道理，任何人在收集所需资料时都应记住这一点。

表 1·1 增加样本含量对均数的标准误的影响
从标准差为 100 的总体中抽样

样 本 含 量	均数的标准误
1	100.0
2	70.7
3	57.7
4	50.0
16	25.0
64	12.5
256	6.25

开始时样本含量稍有增加，样本均数的精确性就大大提高。样本含量从 1 增至 4，S. E. M. 就降低一半，但要再降低一半，样本含量就必须增加到 16。再要进一步降低标准误，就更困难了。若要把 S. E. M. 降至标准差的 6%，样本含量就要从 1 例增加到 256 例。这么大的样本是很少需用的，若要这么大的样本既浪费，又难于获得。

一切重要的离散度指标，我们都已叙述了。还有几个重要的位置量数指标没有讲到，不过已列入附录 2 中了。这些指标是中位数、众数、调和均数、几何均数。我们在附录 2 中叙述了它们的作用。

无效假设与显著性检验

我们从图 1.1 资料中得到血红蛋白均值以后，就要考虑所得到的数值与教科书中所见 14 克/100 毫升的正常值是否相同，为回答这个问题，我们作一个假设：我们的观察值是从均数为 14 克/100 毫升总体中得到的，并计算出我们所得的结果由机遇所致的次数。我们所作的：“该样本是从总体均数 = 14 的总体中抽取的”这个假设称为**无效假设**。而对我们所得结果由机遇所致的概率进行的计算称为**显著性检验**。我们在这里不必关心个体显著性检验数学上的精确性，这个问题将在后面详细讨论。在这里我们必须知道的是：如果无效假设是真实的，则显著性检验告诉我们：所得结果由机遇所致的概率是多少。假定我们发现，我们的结果由机遇所致的概率很大，就有理由说，无效假设是真实的（即样本均数和总体均数 14 之间无显著性差异）。如果我们的结果由机遇所致的概率很小，无效假设就不可能是真实的，样本均数很可能与总体均数 14 是有差别的。

很明显，我们对所得结果计算了由机遇所致的概率，就必须以某种方式来表示我们的结论。我们可以说，从总体均数为 14 的总体中抽取一个样本，得到我们这样的结果或更为极端的结果，由于机遇，比如每 100 次中可以出现一次。这样的表示方式就很确切。我们也可以换一种说法，出现我们这样的结果或更为极端的结果的概率是 $1/100$ 即 0.01。概率的符号是 P ，因此我们写 $p = 0.01$ 。用 P 表示概率的范围可以从 0（从未出现）至 1.0（总是出现）。

P 值并不是不可思议的，它只是表示如果所规定的无效假设是真实的，这样的实验结果出现的频率是多少。

大多数实验人员认为：如果 p 是 0.05 或小于 0.05，无效假设不可能是真实的。那就是说，本例的样本均数与总体均数 14 之差有显著性。至于选择 $p = 0.05$ 作为显著性水平完全是习惯，没有理由认为不能以 0.01 作为显著性水平。

实际显著性与统计显著性

格特鲁德·斯坦 (Gertrude Stein) 说：“一个差异，如果确实存在差异，一定不是一个无足轻重的差异。”统计学家总是说，不要因为实验结果达到了所要求的显著性水平，就认为所得结果有任何实际显著性。这似乎是不言而喻的。但人们只需看看医学文章，就可以找到许多这方面概念混乱的例子。一种治疗恶性肿瘤的新药可能平均延长存活期一天。如果研究了足够数量的病人，则可以显示出用药治疗与未用药治疗的病人存活期有显著差异，但这又有什么实际价值呢？

相反，一种治疗可以延长生命一或二年，但如果正常的存活时间变异很大，人们研究药物疗效时又没有足够的例数，则治疗与未治疗组之差异不可能有统计显著性。这样，一种重要的新药就可能被忽略。如何避免这种困难，方法将在后面(第六章)予以讨论。不过，计算一下可信限往往很有用处。

可 信 限

计算可信限就是从反方向作显著性检验。作显著性检验时，我们要问：如果所作的无效假设是真实的，某种结果由机遇所致的可能性有多大？当我们计算可信限时，要规定所取概率的水平，再决定在所要求的概率水平上接受无效假设的范围。通常用 $p = 0.05$ 作为显著性水平。例如，我们计算

了样本均数的可信限，就可以计算出大于平均数的例数，这样就可以刚好在 $p = 0.05$ 水平上拒绝无效假设。刚好在 $p = 0.05$ 水平上拒绝无效假设的这个界限称为可信限的上限。同样，我们可以计算出小于平均数的例数，这样刚好拒绝无效假设的这个界限称为可信限的下限。总体均数在上、下限之间的概率是 0.95(全部可能的概率之总和是 1.0)，通常把这样确定的上下限之间的范围称为 95% 可信限。

因此，我们可以说，有 95% 的把握可以断定我们的样本是从一个均数在某一范围内的总体中抽取的。

精确计算可信限的方法，在不同情况下是不同的，这将在后面章节中说明。可信限给我们提供了另一种十分有用的显著性检验方法。在前节里，我们提到变量值在显著性检验中可以掩盖其真正的差异。在这种情况下，如果实验者计算了可信限，就可以看到，他的结果与确实存在的真正差异并无矛盾。

第二章 概率定理

贝努里分布

每当我们作显著性检验时，经常要计算一下所得的结果由机遇所致的概率有多大？所以，概率的研究是理解显著性检验的基础。

不出所料，赌博是人们最初理解机遇的主要刺激。一些名人如伽里略 (Galileo)、牛顿 (Newton) 都曾被要求去解答赌台上的一些问题。〔戴维 (David), 1962〕例如，假定我们