

醫學統計方法

吉林醫科大學

1965. 11.

前 言

医学统计方法是医学研究的一种工具，如应用得当，则能起到相应的作用。随着医学科学实验的大力开展，感到需要使用这种工具的人日益增多。但是，由于它的概念比较抽象，特别是计算过程又很复杂、烦琐，使很多人想用而用不上或用不好，严重地阻碍着这一工具为广大的医学工作者服务。这种需要与使用之间的矛盾是十分突出的。为使医学统计方法便于被群众接受，使它真正成为人所共知的科学方法，我们把医学上最常用的一些基本的统计方法，尽量用通俗易懂的方式介绍出来，特别是力求避免原有的复杂计算过程，而通过查表、查图直接得出结果。这样，就可能使更多的人，特别是不谙计算者，容易学会它和应用它，使医学统计方法在医学研究中普及起来，从而有助于提高医学研究的质量。

近二年来，我们进行了简化医学统计方法，特别是简化计算过程的尝试，经过反复的探索 and 大量数据的计算，自制、改进和增补了共几十种工具表（图）。本书先介绍其中的一部分。利用这些表（图），有的不需任何计算，有的只需很简单的计算，就能得出与通过统计公式计算所能达到的同样效果。其中有些方法，如两个百分率的差异显著性测验用表（图），系根据精确概率分布制成，所得结果的精确程度，更非常用的 χ^2 测验方法所能及。同时，利用查表（图）的方法不但省事省时，而且可以大大减少计算上的误差，结果更为可靠。至于不同统计方法之间的小差别，若与实验研究数据本身的波动对比起来，更是无足轻重的了。

本书所介绍的方法虽经初步实践，在一定范围内试用，收到了较好效果，但由于我们的水平和时间所限，漏误之处在所难免。现在刊印出来是为了接受群众检验，以便继续改进与提高。我们认为，简化医学统计方法的途径尽管可能不同，我们的点滴工作更欠成熟；但是我们坚信，只要医学统计专业人员与广大医学工作者密切联系，从实际需要出发，勇于推陈出新，就能使医学统计方法同其它部门一样，终究会走出我国自己的道路来！

吉林医科大学保健组织教研室

1965年10月于长春

目 录

一 计数分析方法	1—12
1—1 百分数的用法	1
1—2 百分率的抽样误差	4
1—3 百分率的置信限	5
1—4 两个百分率的差异显著性测验	6
1—5 多个百分率的差异显著性测验	10
二 计量分析方法	13—27
2—1 平均数的用法	13
2—2 平均数的置信限	15
2—3 差数平均数的显著性测验	17
2—4 两个平均数的差异显著性测验	20
2—5 多个平均数的差异显著性测验	23
2—6 关于个体分布问题	25
三 相关分析方法	28—32
3—1 计数的相关分析	28
3—2 计量的相关分析	29
工具表 (图)	
表 A 1 百分率的 95% 置信限 (N : 1—50)	33—35
图 A 2 百分率的 99% 置信限 (N : 50—1,000)	36—39
表 A 3 百分率的 95% 置信限 (N : 1,000—100,000)	40—41
表 A 4 两个百分率差异显著性测验的显著界限 (N : 3—25)	42—71
图 A 5 两个百分率差异显著性测验的显著界限 (N : 10—1,000)	72—73
图 A 6 两个百分率差异显著性测验的显著界限 (N : 1,000—100,000)	74—75
图 A 7 双平方根纸 (二项概率纸)	附页
表 A 8 由全距计算平均数置信限用系数 (a)	76
表 A 9 差数平均数顺序测验的显著界限	77
表 A 10 两个平均数顺序测验的显著界限	78—85
表 A 11 多个平均数相互比较的系数 (b)	86
表 A 12 例数不等时全距的换算系数 (c)	87
表 A 13 由全距计算个体分布范围用系数 (d)	88
表 A 14 两个全距比值的显著界限	90—93
表 A 15 顺序相关测验的显著界限	94
关于工具表 (图) 的制法说明	95

一 計 数 分 析 方 法

如临床上用一种新疗法，治疗一批病人，其结果一般常用有效者若干人、无效者若干人的形式表示；其他如手术后的化脓和不化脓、化验结果的阳性和阴性、心脏杂音的有无、各型血型、……，等等；其结果都是用定性的形式表示的。所得到的数据必然是计个数的，这类数据在统计上用计数分析方法处理。这里所讨论的计数分析方法包括百分数的用法和百分率的置信限、百分率的差异显著性测验等。

1—1 百分数的用法

什么是百分数？为什么要用百分数？如何正确应用百分数？这是本节要讨论的问题。

如某地于1960年9月曾调查居民28,000人，发现其中有甲状腺肿大者1,200人，则患者占居民的4.3%：

$$\frac{1,200}{28,000} = \frac{4.3}{100} = 4.3\%$$

这就是百分数。它有两个作用：第一，表示平均每100名居民中有4.3名甲状腺肿患者，能够更直观、更明显地表示出此病在当地罹患的普遍程度；第二，把原来的基数（居民人数）化为100，统一了标准，便于同其他有关的百分数比较。

任何百分数都是由分母除分子计算出来的。因此，要正确地计算和应用百分数，必先弄明白分母与分子的关系。从根本上说，它们的关系有两种：一种是分母包括分子，分子是分母当中的一部分，由此种分数算得的百分数，称为百分率；另一种是分母与分子为独立的两个部分，互不包括，由此种分数算得的百分数，称为百分比。如用算式表示，则为

$$\text{百分率} = \frac{B}{A+B}; \quad \text{百分比} = \frac{B}{A}, \text{ 或 } \frac{A}{B}.$$

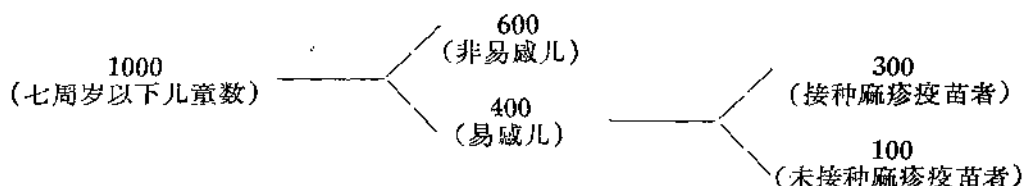
上面所举甲状腺肿患者占当地居民的4.3%这个百分数，是百分率的例子（此例称患病率）；下面再举一个百分比的例子。某市1950年发生流行性乙型脑炎，计男患者229名，女患者138名，男女患者之比为

$$\frac{229}{138} = 165.9\%.$$

这儿分母中并不包括分子，故165.9%是百分比（此例称患者性比）。

应用百分率时须注意以下四个问题。

① 关于分母的选择问题。如某地防疫站要计算“麻疹疫苗接种率”，设有关数据如下：



按接种率的意义，就是实际接种人数占全部应接种人数的百分率。分子是实际接种者，当然是 300 人；但用那个数作分母呢？因为只有麻疹易感儿才应该接种麻疹疫苗，故分母应该用 400，而不能用 1000。算得接种率应为

$$\frac{300}{400} = 75\%。$$

类似这样的例子如：某病复发率应以该病全愈（包括治愈和自愈）人数为分母，因为尚未全愈者就没有复发问题；牛痘复种率应以初种人数为分母，因为从未种过痘者根本谈不上复种；避孕率应以具备生育条件的妇女人数作分母，而以一定时期内采取各种避孕措施的妇女人数作分子；等等。诸如此类，不胜枚举。总之，一般地说，分母应该只包括有可能成为分子的单位（如上述接种率的分母以每个易感儿为一个单位），而不应包括不可能成为分子的单位（如上述接种率的分母不包括非易感儿）。这样选择的分母，才能与分子有恰当的对应关系，而算得的百分率也才能正确地反映出真实情况。

② 关于分子随时间而积累的问题。在百分率的计算中，如果分母不变（或基本上不变），而分子是随时间的延长而积累的结果，则算得的百分率亦必将随时间的延长而增高。举例说明如下。表1—1是某工地的痢疾发病率。从发病例数来看，每个月的发病例数都是当

表1—1 某工地工人的痢疾发病率，
1959年5—7月

	工 人 数 (分母)	发病例数 (分子)	发 病 率 (%)
五 月	6,500	350	5.4
六 月	6,600	977	14.8
七 月	6,800	615	9.0
合 计	6,633*	1,942	29.3

表1—2 某地克山病病死率，1959年
12月—1960年2月

	发病例数 (分母)	病死例数 (分子)	病 死 率 (%)
十二月	86	21	24.4
一 月	330	59	17.9
二 月	87	10	11.5
合 计	503	90	17.9

* 三个月人数的平均数。

月新发生的，故整个季度的发病人数应为350+977+615=1,942(人)。这是说，计算发病率的分子随时间的延长(由一个月延长到三个月)而积累了。再看该工地的工人数，虽然有些变动，但基本上还是那些人。整个季度的工人是三个月工人数之平均，而不能把三个月的人数相加，相加就重复了。这是计算发病率所用的分母，可见分母没有随时间而积累。这样算得的季度发病率，由于分子的积累，必定比各月发病率高。所以分子积累的期限不同，其百分率是不能直接互相比的。但是，分子的积累影响，是否一定会在百分率上反映出来呢？这就要看分母了。如在表1—2上，分母是每月新发生的病例数，同分子一样，也随时间的延长而

积累, 因为整个季度的发病人数就是各月新发生的病例数相加之和。这样, 分母与分子同时积累, 计算百分率时, 其影响互相抵消, 使算得的百分率便不随时间的延长而增高。如本例季度病死率并不高于各月病死率, 而只相当于各月的平均而已。此外, 如治愈率、手术化脓率等, 皆属此类情况。总之, 只有当分子是一定期限内积累的结果, 而分母是固定不变的或基本上不变的, 算得的百分率才受期限长短的影响。计算这种百分率时, 必须注明期限。

③ 关于百分率的表示形式问题。由于研究的目的不同, 百分率可以用两种不同形式表示出来。第一种是, 如果分母只需分为二部分表示, 则用一个百分率就足以同时说明全部结果了。如前面所说甲状腺肿的患病率 4.3%, 不言而喻, “未患病率” 一定是 95.7%; 另如用中药秦皮乙素治疗典型细菌痢疾 47 例, 治愈 26 例, 治愈率为 55.3%, 显然 “未治愈率” 为 44.7%。诸如此类的例子很多, 如阳性率 (相反是阴性率)、有效率 (相反是无效率)、化脓率 (相反是未化脓率), 等等。像这种知其一必知其二的情况, 只要把所研究的一个方面着重指出来就足够了, 无需同时用两个百分率把两方面都表示出来。第二种是, 如果必需把分母分成三个或三个以上的部分表示时, 则必须把每个部分占分母的百分率一一计算出来, 相应地表示出各部分所占的比重。如表 1—3 是各血型的人数占调查总人数的百分率, 这些百分率的合计必为 100%, 表示时亦必须写出来, 以示它们是一组完整的、不可分割的数据。如果只单独写出其中的一个百分率, 比如 O 型者占 31.0%, 对于表达 “某民族 ABO 血型的分布” 这一概念来说, 就会失去应有的完整意义。也正因为如此, 其合计项必定要用 100%, 不能因为计算上四舍五入的影响而写成 99% 或 101%。像表 1—3 就是这种百分率的具体表示形式。

表 1—3 某民族 ABO 血型的分布

	人 数	百 分 率 (%)
O 型	6,292	31.0
A 型	7,753	38.2
B 型	4,303	21.2
AB 型	1,949	9.6
合 计	20,297	100.0

④ 关于百分率的解释问题。对任何百分率的解释, 必须牢牢地把握住计算此百分率的分子与分母的关系。

关系弄错了, 解释也一定错误。如表 1—4 本是 106 例心房颤动病例之年令分布, 但原作者便据此得出结论谓 “21—30 岁阶段发病率最高, 占 26.4%”。这显然是把患者的年令分布误认为年令别发病率了。所谓 21—30 岁的发病率,

应是 21—30 岁的患者数占 21—30 岁总人口数的百分率, 即计算这种百分率的分母应该用 21—30 岁的总人口数。而 26.4% 乃是 21—30 岁的患者数占全部患者的百分率, 其分母是全部患者 106 例。两个百分率的含义迥然不同, 不能混淆。为具体说明此问题, 姑且假设上述 106 例患者是在 10,000 名居民中发生的, 如按一般人口的年令比例推算, 其中 21—30 岁者约 2,000 名左右, 而 51—60 岁者约 500 名左右, 算得两个年令组的发病率如下:

表 1—4 106 例心房颤动病例的年令

年令 (岁)	病 例 数	百 分 率 (%)
<20	2	1.9
21—30	28	26.4
31—40	19	17.9
41—50	16	15.1
51—60	21	19.8
61<	20	18.9
合 计	106	100.0

$$21-30\text{岁的发病率} = \frac{28}{2,000} = 1.4\%;$$

$$51-60\text{岁的发病率} = \frac{21}{500} = 4.2\%.$$

可见51—60岁的人当中，容易发生心房颤动的程度为21—30岁者的三倍，显然并非“21—30岁阶段为最高”，原文的结论是错误的。类似这样的例子很多，如以门诊病例中某病所占百分率最高，就误认为这种病最易发生，而不分析此病在一般人口中的发病率等等。产生这种错误的原因，主要是由于对百分率的含义不结合具体的分母去理解，像表1—4的分母本来只是病人总数，却当做一般的人口数了，这是解释或使用百分率时应切实注意的问题。

关于百分比的应用，在医学上远不如百分率广泛，而且也比较简单。只是在解释上应遵照其本来的含义，避免“误解”。例如前面举过的流行性乙型脑炎患者性比为165.9%的例子，表示男患者比女患者多，男患者约为女患者的一倍半；而假如这个城市的男性人口也达到女性人口的一倍半左右，则男女发病率就几乎相等，切不可误认为男性比女性易于发病。

至于有时用百分数，有时用千分数(‰)、万分数(‱)、十万分数(‱‱)等，则视表达的方便而选用。如计算某发病率，其分母是390,000，分子是229，则可写成下列各种形式：

百分数	0.059%
千分数	0.59 ‰
万分数	5.9 ‱
十万分数	59 ‱‱

其中万分数和十万分数带整数位，看起来比较方便。其实，它们都只是百分数的小数点移位而已，用法都是一样的，自不必赘述了。

1—2 百分率的抽样误差

要知道什么是抽样误差，必先明白什么是抽样；而欲知什么是抽样，又必先懂得什么是总体和样本。上节举过的一些百分率的例子，有的是总体，有的是样本。现在就用这些例子来说明百分率的抽样误差问题。

如调查居民28,000人，其中发现甲状腺肿大者1,200人，患病率为4.3%。因为28,000人就是当地的全部居民，而4.3%又只是用来说明该地区的甲状腺肿罹患的普遍程度，故这些人就是全部研究对象，构成了一个研究总体，4.3%便是这个总体的患病率。这里没有抽样问题，因而也没有抽样误差。但是，如某医生用中药秦皮乙素治疗典型细菌痢疾47例，治愈26例，治愈率为55.3%，情况就不同了。很明显，试用此药的目的，是要了解它对所有的典型细菌痢疾患者之疗效。如果这种疗法确实有效，则受治者中既可包括现有的患者，也可包括未来的患者。换言之，所有的典型细菌痢疾患者便构成这样一个研究总体，它所包括的人数是无限大的；而现在直接受治的47例患者只是这个总体中的一小部分，称为样本。55.3%是样本的百分率。样本既然属于总体的一部分，因而样本百分率必然与总体百分率接近；但因样本毕竟只是由总体内的一部分单位组成，所以样本百分率与总体百分率总还要有一定出入（差异）。这种差异由于从总体内抽取样本而产生，故称抽样误差。综上所述，可概括为以下两点：

① 直接观察的百分率如果只用来描述它本身所包括的单位，则无抽样问题，亦无抽样误差；百分率如果用来描述超过它本身所包括的单位，即用样本百分率描述总体百分率时，则必然有抽样误差。

② 在医学研究中，一般地说，对样本的研究是手段，对总体的研究才是目的。必须根据样本的结果，考虑到抽样误差的影响，然后再做出适合总体的结论。

在计数分析方面，根据样本的结果，做出适合于总体的结论，主要包括两方面内容：一是根据样本百分率估计总体百分率；二是根据样本百分率之间的差异，推测总体百分率之间是否也有一定差异。前者用百分率的置信限方法解决，后者用百分率的差异显著性测验方法解决。以下各节将讨论这两方面问题。

1—3 百分率的置信限

什么是置信限？先从下列具体问题谈起。

比如，我们不知道某人的确切年龄，只从外貌上加以估计，说：“他八成有三、四十岁了。”这话的意思是：虽然我们猜不出他的确切年龄，但是可以认为他不会小于30岁，也不会超过40岁，即用30—40岁这个范围去估计他的年龄。“30—40”就相当于置信限，30为其下限，40为其上限。这样估计有多大把握呢？有“八成”即80%的把握；当然，也还有20%猜错的危险。根据样本估计总体百分率的置信限，道理是同样的，只是可靠程度比较高，一般要求至少达到95%，甚至99%或更高一些，而估计的范围则是根据严密的统计方法计算出来的。

仍用上节治疗痢疾的例子加以说明。样本例数 $N=47$ 人，其中出现治愈者 $X=26$ 人，样本百分率 $p=55.3\%$ ，则其总体百分率（治愈率）的95%置信限为40%—70%。表示如果所有的典型细菌痢疾患者都用秦皮乙素来治疗，那末总体的治愈率将不会低于40%，也不会高于70%，作这样估计有95%的把握，而冒5%的危险。即平均每做100次估计时，其中将有95次正确，5次错误。

要用统计公式计算出百分率的置信限，是相当麻烦的。我们已把样本例数（ N ）1—100,000的置信限制成工具表和图，一查便知，不必计算。表和图的应用范围如下：

表 A1	共 3 个表	例数 (N) 从 1— 50
图 A2	共 4 个图	例数 (N) 从 50— 1,000
表 A3	共 2 个表	例数 (N) 从 1,000—100,000

下面举几个实例说明这些表和图的用法。

① 说明表 A1 用法的实例

例 1 用醋酸铈内服治疗头癣患者 38 例，其中 7 例有中毒反应，中毒反应率 $p=\frac{7}{38}=18.4\%$ 。求总体的中毒反应率之 95% 置信限。此例例数不超过 50，应查表 A1。在表 A1—2 上查得：样本例数 $N=38$ ，出现数 $X=7$ 时，对应之 95% 置信限为 8%—34%（表中细体字是 95% 置信限，粗体字是 99% 置信限）。意思是说：如果在大量的头癣患者中进行同样的治疗时，中毒反应率最低可至 8%，最高可达 34%。这样估计错误的机会将不超过 5%。又 99% 置信限为 6%—39%，结论的作法同上，只是估计的把握程度更大了。

例 2 用秦皮乙素治疗典型细菌痢疾 47 例，治愈 26 例，治愈率为 55.3%。求总体的治

癒率之95%置信限。仍查表A1, 因样本百分率 p 超过50% (或出现数 X 超过例数 N 之半数), 故查表A1—3上 $N=47$, $X=47-26=21$ 对应之95%置信限为30%—60%, 由100%内减去此二数, 即得所求之置信限为40%—70%。

② 说明图A2用法的实例

例3 解剖115具尸体, 发现其中30具尸体有肱二头肌额外头, 占26.1%。求总体百分率的95%置信限。例数在50以上, 查图A2。在图A2—1上, 横座标 $N=115$ 与图内曲线 $p=26.1\%$ (近似于26%) 交点之纵座标为18%, 此即下限; 同法在图A2—2上查得纵座标为35%, 此即上限。如果查图A2—3和图A2—4, 得99%置信限为16%—38%。如样本百分率大于50%, 则先由 $100\%-p$, 然后查图, 再将所得结果由100%内减去即得。

③ 说明表A3用法的实例

例4 1—1节表1—3血型调查例数 $N=20,297$, 其中O型者占31.0%。求总体百分率之95%置信限。例数超过1,000, 查表A3。在表A3—1上, 样本百分率 $p=31\%$ 的一列中, $N=20,297$ 在3,602—32,693 (下限项目下) 之间, 这两个数的下限分别为 $p-1.5\%$ 和 $p-0.5\%$, 而此例 N 靠近这两个数的中央, 故下限应为 $31.0\%-1.0\%=30.0\%$; 同法在表A3—1的上限项目下查得其上限为 $31.0\%+1.0\%=32.0\%$ 。表A3—2可查99%置信限。如样本百分率超过50%, 则以 $100\%-p$ 查表, 然后再将所得结果由100%内减去, 即为所求之总体置信限。

一般地说, 95%置信限是常用的范围, 但有时也用99%置信限。比如本节例1的结果:

95%置信限	8%—34%,
99%置信限	6%—39%。

如果从对总体百分率估计的把握大小来看, 当然99%的置信限把握要大些, 因为它只有1%估计错的危险, 但同时它的范围也比较大, 所以估计的精密程度就不如95%者; 反之, 95%置信限的把握稍为小些, 因有5%估计错的危险, 但精密程度却较高。可见估计总体百分率的把握大小与精密程度是相互制约的。实际上, 无论用那个置信限, 都是以一个明确的范围估计总体百分率, 比起把样本百分率就直接当作总体百分率, 当然要可靠得多, 严密得多。

置信限有下限和上限, 通常都是二者并用, 而把总体百分率置于此范围之内。但有时只用上限或下限一方面即足够了。例如本节例1, 我们可以说, 中毒反应率最高可达39% (99%置信限之上限); 又如本节例2, 我们也可以说, 治愈率至少在40%以上 (95%置信限之下限)。读者可根据需要选用。

1—4 两个百分率的差异显著性测验

前两节讨论过, 样本百分率是有抽样误差的; 因此, 两个样本百分率的差异也必定有抽样误差。要根据两个样本百分率的差异对总体作出结论, 须进行差异的显著性测验。例如, 癆型克山病人20例, 试服保心汤 (中药、西药合成), 其中男性15例, 女性15例, 服药后的疗效如表1—5所示。27%就是样本百分率的差异。根据这一差异, 能不能就作出女患者的有效率高于男患者的结论呢? 即能否根据表1—5所列之样本百分率有差异, 就遽然做出总体百分率亦有差异的结论呢? 显然是不能的。因为我们知道, 由于有抽样误差, 样本百分率之间的差异, 一般地不等于其总体百分率之间的差异。要对总体下结论, 必须进行差异的显著性测

验。测验的结果，得出概率 P ， P 值是表示做出“差异显著”的结论时，所冒“风险”的大小。在习惯上， P 值与结论的关系如下：

$P < 5\%$ 时，结论为差异显著，即认为两个总体百分率之间有一定差异。结论错了的危险程度在 5% 以下；

$P < 1\%$ 时，结论为差异非常显著，即认为两个总体百分率之间有一定差异。结论错了的危险程度在 1% 以下；

$P > 5\%$ 时，结论为差异不显著。

必须指出的是，我们这里所说的“差异不显著”或“无显著差异”，都意味着“尚不能认为总体之间有差异”，而并非肯定总体之间就没有差异了，因为当样本例数增加之后，还可能出现差异显著的情况。

我们已制成两个百分率差异显著性测验用的工具表和图，一查便知，不需计算。根据两个样本例数 N_1 和 N_2 的大小，各个表和图的使用范围如下：

表 A4	共 45 个表	用于 $N_1, N_2 \leq 25$
图 A5	共 2 个图	用于平均例数 ($\bar{N}$) 10—1,000
图 A6	共 4 个图	用于平均例数 ($\bar{N}$) 1,000—100,000

下面举几个实例分别说明这些工具表和图的使用法。

① 说明表 A4 用法的实例

例 1 见本节表 1—5。此例的研究目的，是要比较男女患者服用同样药剂后的疗效有无差异。这差异包括两种可能：一是男患者有效率高于女患者，一是男患者有效率低于女患者。在差异的显著性测验中，这类性质的问题属于“双侧测验”，查表时应取“双侧的界限”作标准。在表 1—5 上， N_1 和 N_2 均不超过 25，故宜用表 A4。查此表时，规定 $N_1 \geq N_2$ ， $p_1 > p_2$ 。本例符合这些规定，可直接查表。先在标题上找到 $N_1 = 15$ (双侧)，即表 A4—25，然后在 $X_1 = 10$ 和 $N_2 = 5$ 的交点找到细体字 $X_{2, .05} = 0$ ，表示当 $N_1 = 15$ ， $X_1 = 10$ ， $N_2 = 5$ ， $X_2 \leq 0$ 时， $P < 5\%$ ，即两个总体百分率之间有差异。查表的结果与结论之关系如下：

$X_{2, .05}$	$X_2 \leq X_{2, .05}$,	$P < 5\%$,	差异显著；
	$X_2 > X_{2, .05}$,	$P > 5\%$,	差异不显著。

本例 $X_2 = 2$ ，故 $X_2 > X_{2, .05}$ ， $P > 5\%$ ，表示尚不能认为保心汤的疗效有性别上的差异。

例 2 用大剂量鞣丸酮治疗慢性再生障碍性贫血患者 44 例，结果如表 1—6。问病程较短者有效率是否较高？本例

欲测验之差异只有一种：病程在三年以下者的有效率是否高于三年以上者，而不存在另一种相反的可能（指病程在三年以下者有效率反而较低）。这类性质的问题属于“单侧测验”，查表时应取“单侧的界限”作标准。本例 N_1 和 N_2 均不超过 25，应查

表 1—6 对不同病程的贫血患者之疗效

病 程	治疗人数	有效人数	无效人数	有效率 (%)	有效率之差异 (%)
> 3 年	$N_1: 23$	7	$X_1: 16$	$p_1: 30$	$p_2 - p_1: 32$
< 3 年	$N_2: 21$	13	$X_2: 8$	$p_2: 62$	

表A4。按 $N_1 > N_2$ ，先找到标题上为 $N_1 = 23$ （单侧）的表A4—40。但表1—6上 $p_1 < p_2$ ，故不能直接用“有效人数”作为出现数 X ，而须代之以“无效人数”：

$$X_1 = 23 - 7 = 16, \quad X_2 = 21 - 13 = 8.$$

在表A4—40上， $X_1 = 16$ ， $N_2 = 21$ 时， $X_{2, .05} = 8$ 。凡 $X_2 \leq X_{2, .05}$ 时均表示差异显著，故可认为病程在三年以下者之有效效率高干病程在三年以上者。

② 说明图A5用法的实例

例3 在175例小儿和成人痢疾患者中调查病原菌的种类，部分结果如表1—7。问小儿与成人痢疾中，福氏杆菌病原所占百分率有显著差异否？本例例数超过25，应查图A5。查此图时，规定 $p_1 < p_2$ ，对应之样本例数为 N_1 和 N_2 ，已如表1—7标出者。图上横座标为两组之“平均例数” $\bar{N}$ ，系按下式计算：

表 1—7 小儿和成人痢疾患者中福氏杆菌病原的百分率

	患者人数	病原为福氏杆菌的人数	百分率 (%)	百分率之差异 (%)
小 儿	$N_1: 104$	36	$p_1: 35$	$p_2 - p_1: 43$
成 人	$N_2: 72$	56	$p_2: 78$	

$$\bar{N} = \frac{2N_1 \cdot N_2}{N_1 + N_2} + \frac{N_1 - N_2}{100p_1 + a} \quad [\text{公式1—1}]$$

其中 $\frac{2N_1 \cdot N_2}{N_1 + N_2}$ 不必计算，可由图A5右边所附计算尺查出。查法是先在此尺上找到 N_1 和 N_2 的位置，然后取其中点即得（用有刻度的尺量或用一纸条对折均可）。本例在图A5—1右边计算尺上可查出104与72之中点在85（取日测的整数即可）。算式中右面一项是校正数，其中 a 是常数：

$$\text{当 } N_1 > N_2, \quad a = 4;$$

$$\text{当 } N_1 < N_2, \quad a = 8.$$

本例 $a = 4$ ，故校正数* 为

$$\frac{104 - 72}{35 + 4} \approx 1.$$

此例得 $\bar{N} \approx 85 + 1 = 86$ 。在图A5—1上，横座标 $\bar{N} = 86$ 与图内曲线 $p_1 = 35\%$ 交点之纵座标为 $p_{2, .05} = 51\%$ ，用法如下：

$$p_2 \geq p_{2, .05}, \quad P < 5\%, \quad \text{差异显著};$$

$$p_2 < p_{2, .05}, \quad P > 5\%, \quad \text{差异不显著}.$$

此例 $p_2 = 78\%$ ，故

$$p_2 > p_{2, .05} \quad P < 5\%.$$

得结论谓：小儿痢疾与成人痢疾患者中福氏杆菌病原所占百分率有显著差异，成人患者中福氏杆菌病原所占百分率高于小儿。

* 校正数：只要用心算按四舍五入估出近似的整数即可，如例数较小时，常为 +1，-1，+2，-2，……等。

在图 45 纵座标上直接查出者都是 p_2 的双侧 5% 界限值。如欲查出单侧 5% 和 1%、双侧 1% 的界限值, 须将算得之 $\bar{N}$ 乘以表 1—8 上相应的系数, 然后再按上述方法处理。看了例 4 自然明白。

例 4 某疟疾流行地区, 为观察二二三灭蚊的滞留效果, 在喷药的实验区与未喷药的对照区抽查了小学生的疟原虫感染情况, 结果如表 1—9。问实验区的疟疾感染率是否低于对照区? 也就是要根据实验区的样本感染率低于对照区, 以测验实验区的总体感染率是否也低于对照区。这里当然不存在实验区的感染率是否反会高于对照区的问题, 故应该用“单侧测验”。根据图 45—2 右边的计算尺、公式 1—1 和表 1—8 的系数, 可算出查图时用的 $\bar{N}$ 为

$$\bar{N} = \left[\frac{2 \times 113 \times 103}{113 + 103} + \frac{113 - 103}{3.5 + 4} \right] \times 1.35 \approx [108 + 1] \times 1.35 = 147.$$

表 1—9 二二三滞留实验区与对照区的小学生疟疾感染率

	受检人数	感染人数	感染率 (%)	感染率之差异 (%)
实验区	$N_1: 113$	4	$p_1: 3.5$	$p_2 - p_1: 9.1$
对照区	$N_2: 103$	13	$p_2: 12.6$	

由图 45—2 上横座标 $\bar{N} = 147$ 与图内曲线 $p_1 = 3.5\%$ 交点之纵座标, 得 $p_{2, .05} \approx 10\%$ (稍弱), 这是单侧的标准。此例 $p_2 = 12.6\%$, 故

$$p_2 > p_{2, .05}, \quad P < 5\%.$$

差异显著, 即实验区小学生的疟疾感染率确比对照区低, 说明二二三有滞留灭蚊抗疟的效果。

③ 说明图 46 用法的实例

例 5 某地调查七岁以下男女儿童的麻疹发病率, 如表 1—10。问男女儿童的麻疹发病

表 1—10 某地七岁以下男女儿童的麻疹发病率

	调查人数	发病人数	发病率 (%)	发病率之差异 (%)
男	$N_2: 2,023$	921	$p_2: 45.5$	$p_2 - p_1: 1.8$
女	$N_1: 1,946$	851	$p_1: 43.7$	

率有显著差异否? 既是男女之间的差异, 当然包括男发病率大于和小于女发病率两种可能, 必是双侧测验了。根据图 45 右边的计算尺* 及公式 1—1 可算出 $\bar{N}$ 如下:

* 如 N_1 和 N_2 超过图 45 右边计算尺上的刻度时, 可按原刻度的 10 倍、100 倍套用。

$$\bar{N} = \frac{2 \times 1946 \times 2023}{1946 + 2023} + \frac{1946 - 2023}{43.7 + 8} \approx 2000.$$

在图 A6—1 的左图上，横座标为 $\bar{N} = 2,000$ 和图内曲线为 $p_1 = 43.7\%$ (在 40%—50% 两条曲线之间估计即可) 交点之纵座标是 $(p_2 - p_1)_{.05} \approx 3.1\%$ (稍强)，表示只有当样本的 $p_2 - p_1 \geq 3.1\%$ 时，才得 $P < 5\%$ ，此例 $p_2 - p_1 = 1.8\%$ ，故 $P > 5\%$ ，即不能认为男女儿童的麻疹发病率有差异。

关于表 A4，图 A5，图 A6 的用法，须注意以下三点：①用表 A4 时，必须 $N_1 \geq N_2$ ，如果两个百分率也正好是 $p_1 > p_2$ ，则可直接以出现数 X_1 查表，得出 $X_{2,.05}$ ；如 $p_1 < p_2$ ，则须以不出出现数为 X_1 查表，得出 $X_{2,.05}$ ，后者为第二组的不出出现数之显著界限。②用图 A5 和图 A6 时，必须 $p_1 < p_2$ ，而 N_1 和 N_2 即随之而定，孰大孰小均可。只是当 $N_1 < N_2$ 时，应注意公式 1—1 上的校正数为负值了。③图 A5 和图 A6 在用法上的唯一不同点是：图 A5 纵座标为 $p_{2,.05}$ ，而图 A6 纵座标为 $(p_2 - p_1)_{.05}$ 。

关于单侧测验和双侧测验的选择问题，乃取决于进行差异显著性测验的目的和产生差异的可能情况。如果只欲测验两个总体百分率间有无差异，而且产生差异的可能情况包括：

甲总体百分率 > 乙总体百分率，

甲总体百分率 < 乙总体百分率，

则应该用双侧测验。如果只欲测验上述两种情况中的一种，而且产生差异的可能情况也只限于其中之一种，则应该用单侧测验。如本节例 1、3、5，须采取双侧测验，而例 2、4 须采取单侧测验，就是根据上述原则。在应该用单侧测验时正确地采用单侧的界限作标准，可以达到显著性测验应有的效果；反之，如果此时误用双侧标准，则会降低显著性测验的效果。

此外，也有一些问题须用一个样本百分率与某个总体百分率作差异显著性测验。测验时均可用 1—3 节的方法，以样本例数及其百分率查出置信限，如置信限范围不包括对比的总体百分率，则与此总体百分率差异显著；如包括在内，则差异不显著。举例说明如下：

例 6 某地在小黄鼠哺乳幼鼠时期，捕获该种鼠 495 只，其中雌鼠 295 只，占 59.6%。问在此时期雌雄鼠的多少（指地面活动者）有显著差异否？如果雌雄只数毫无差异，则雌鼠应占 50%，这就是总体百分率。根据样本 $N = 495$ ， $p = 59.6\%$ ，查图 A2—1 和图 A2—2，得 95% 置信限为 55%—64%。这个范围没有包括 50%，故雌雄小黄鼠的多少不是没有差异，而是雌多于雄。因为这一结论是根据 95% 置信限作出的，故冒了 5% 的危险。

1—5 多个百分率的差异显著性测验

上节讨论了两个百分率之间差异显著性测验问题，本节讨论三个及三个以上百分率之间差异显著性测验问题。这儿介绍的是用《双平方根纸》进行多个百分率的差异显著性测验的方法。双平方根纸系纵横轴皆用平方根的尺度制成的图纸，原图见图 A7。图上横轴由 0—100，纵轴由 0—50，50 以下每一小格代表 0.5，50 以上每一小格代表 1。图的左边附有 R 值（详下）的显著界限，图上边附有 R 值的标准尺。举下面实例说明其用法。

表 1—11 是肺结核各级病灶的涂片检验与培养检验之符合情况，要比较四个符合率之间的差异是否显著。步骤如下：

① 定纵横轴——在表 1—11 上“符合者”与“不符合者”两个方面，在作图时，一为横轴，一为纵轴。原则上可任意选用，但图上横轴的尺度范围较大，故常把总数较多者置于横轴上。此例“符合者”在横轴，“不符合者”在纵轴。纵横轴一经确定，即不能再行更换。

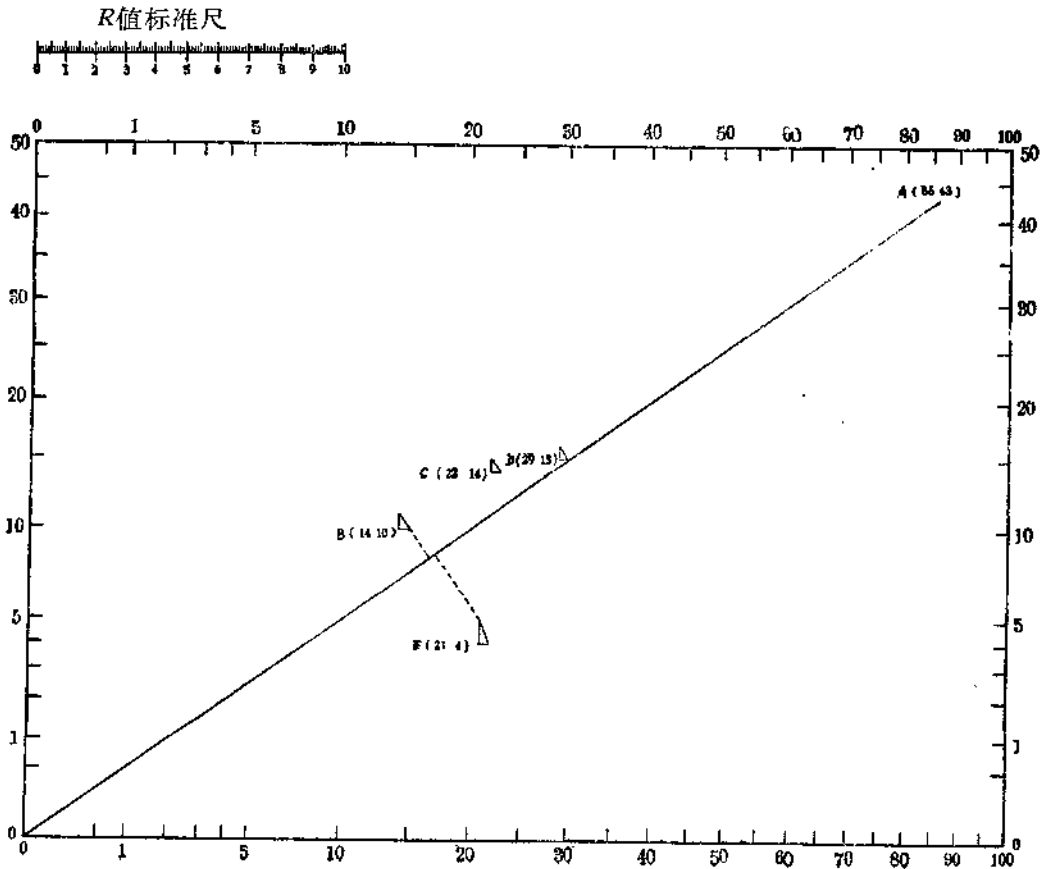
② 定斜线——以合计项的 86 为横座标，43 为纵座标得图 1—1 上的 A 点。将此点与原点 (0,0) 相连，为斜线。如果横座标超过 100 或纵座标超过 50，可同时以

表 1—11 肺结核各级病灶之涂片检验与培养检验结果符合率

病灶活动等级	涂片与培养符合情况		符合率 (%)
	符合者	不符合者	
0—I 级	14	10	58.3
Ⅱ 级	22	14	61.1
Ⅲ 级	29	15	65.9
Ⅳ—Ⅴ 级	21	4	84.0
合 计	86	43	—

横座标 纵座标

图 1—1 双平方根纸用法举例 (原图见图 A7)



原数之1/10或1/100等作图，斜线之位置是不变的，读者可在图1—1上用(8.6, 4.3)代替A点作斜线，结果相同。

③ 定座标点——有几个样本百分率，就要定几个座标点。此例可按B(14,10)，C(22,14)，D(29,15)和E(21,4)作四个座标点，见图1—1。然后在这四个点向右和向上作互相垂直的两个线段，长度各为1(注意：50以上时占1小格，50以下时占2小格)*，再分别连成四个小直角三角形。

④ 求R值——取各座标点的三角形与斜线最近之顶点向斜线作垂直线，然后在斜线之两侧各取一个最长的距离(注意：不论共有几个点，均只在斜线两侧各取一个距离)，见图1—1上虚线所示者。两个距离之和即R值，并在图上边的标准尺上度量，可得读数。此例 $R=3.7$ 。

⑤ 查表——在图A7左边有两个表，上表当组数为4时， $R_{.05}=5.45$ 。R与结论的关系如下：

$R > R_{.05}$, $p < 5\%$, 差异显著；

$R < R_{.05}$, $p > 5\%$, 差异不显著。

此例 $R < R_{.05}$ ，故 $P > 5\%$ ，即不能认为：肺结核患者的病灶活动性等级不同，会影响细菌的涂片检验与培养检验符合率的高低。如果例数较大，图上的尺度不够用时，可取原数的十分之一定座标点，这时就不必作三角形了(因为按十分之一作的三角形过小，可略去)，而最后须採用图A7左边下表的显著界限作标准，其他均同上例。

此外，本节方法亦可用于两个百分率的差异显著性测验。特别是当多个百分率的测验结果为差异显著时，如果还要进一步测验某两组之间的差异显著性，则仍可利用原来的座标点，只需加一条斜线并另量R值即可。这时所用R值的标准，当然是组数为2的 $R_{.05}$ 或 $R_{.01}$ 了。

【註】用 χ^2 測驗法作对比，所得結論一致：

$$\chi^2 = 4.641, \quad \text{自由度} = 3, \quad P > 5\%。$$

* 因图1—1系由图A7縮成，詳細刻度須見图A7。

二 計 量 分 析 方 法

如每个人的体重，都有他自己的数值，各不相同；其他如体温、血红素、心率、疾病的潜伏期、基础代谢率、药物剂量、……，等等，也都是每个人或每一次有一个数据。概括地说，就是各个研究对象（单位）之间都存在着数量上的差别，研究的结果是用定量的形式表示的。所得的数据必然是计量的，这类数据在统计上用计量分析方法处理。这里所讨论的计量分析方法包括平均数的用法和平均数的置信限、平均数的差异显著性测验，此外，也要分析关于个体分布的问题。

2—1 平均数的用法

什么是平均数？怎样计算它？先从具体例子谈起。下面是随机测得10名二十岁健康男子之体重（公斤）：

49 53 55 56 60 61 63 63 64 66

如果问：他们的平均体重是多少？则只有把这10个数——加起来之后用10除，即

$$\bar{x} = \frac{49+53+\cdots+66}{10} = 59 \text{ (公斤)}。$$

59公斤就代表它们的平均水平，在统计上称为算术平均数，或简称平均数，用符号 $\bar{x}$ 表示。

当例数不多时，如10例、20例，这样计算很方便。但例数过多，如有几十例、几百例、甚至几千例，显然，仍用这样逐个相加的算法既麻烦，又易发生错误，必须改用更简便的方法。

例如，测得500个八岁男童的身高，求它们的平均数时，先把这500个数据分成16个小组，分组的结果见表2—1第(1)栏。第一组“104—”表示从104厘米到106厘米，第二组“106—”表示从106厘米到108厘米；如身高恰好106厘米者则分到第二组“106—”去，余此类推。所以从104厘米开始，是因为这500个身高没有再低于它的了；同理，也没有再高于136厘米的了。这里每组都包括2厘米的范围。如使每组包括3厘米的范围时，则分组、计算都不方便；如使每组包括4厘米时，所分的组数又嫌太少（一般应在10—20组为宜）。每组分得的人数（统称次数）见表2—1第(2)栏。

表 2—1 用次数表计算平均数（长春市八岁男童身高，厘米，1957年）

分 组 (1)	次 数 (2)	中 点 (3)
104—	1	105
106—	2	107
108—	5	109
110—	10	111
112—	27	113
114—	28	115
116—	57	117
118—	69	119
120—	89	121
122—	72	123
124—	56	125
126—	38	127
128—	26	129
130—	6	131
132—	10	133
134—136	4	135
合 计	500	—

表 2—1 所列第 (1)、(2) 栏的形式称“次数表”。

在次数表中，每一组的次数（人数）必然分布在自己这个组的 2 厘米范围内。如第 4 组有 10 个人，他们的实测数据分布在 110—112 厘米之间：

110.0 110.5 110.6 110.8 111.0 111.2 111.4 111.5 111.6 111.9

它们的平均数是 111.05 厘米，与这个组的“中点” 111 厘米（即 111 在 110 与 112 的“中点”之意）十分接近。组内的次数越多，它们的分布也越趋于均匀，其平均数与该组的中点也必越趋于接近。既然组的中点与同组实测数据的平均数很接近，我们可否就用各组中点来代替这些实测数据呢？实践证明，是可以的。现在用表 2—1 第 (3) 栏上的 16 个中点分别代替原来的 500 个身高，就变成 1 个 105、2 个 107、……、89 个 121、……、4 个 135 厘米了。故算得平均数为

$$\bar{x} = \frac{105 + 2(107) + \cdots + 4(135)}{500} = \frac{60,586}{500} = 121.17 \text{ (厘米)}。$$

表 2—1 上的中点，都是大于 100 的数值，我们在计算时可以暂时先去掉 100，而在最后再加上一个 100，这样比较简便，其结果还是相同的。

$$\bar{x} = 100 + \frac{5 + 2(7) + \cdots + 4(35)}{500} = 100 + \frac{10,586}{500} = 100 + 21.17 = 121.17 \text{ (厘米)}。$$

这样计算准确吗？我们把 500 个原来的实测身高逐个加起来，用 500 除，结果是 $\bar{x} = 121.14$ 厘米。两者相差只有 0.03 厘米。这样微小的差别，对于平均身高来说，是完全可以忽略的了。

在医学中，虽然大部分资料都适合于用算术平均数表示，但也有一些资料，如用滴度表示的化验结果，算术平均数就不适用了。例如在研究某地正常人群中流行性感冒抗体的调查中，有一组 12 份血清对某型病毒之血凝抑制效价如下：

5 5 5 5 5 5 5 10 10 10 20 40

它们之间的关系有这样一个特点：要么相等，要么相差 1 倍、或 2 倍、或 4 倍、或 8 倍、……，在 5 与 10、10 与 20、20 与 40 之间都不是连线的，它们是按倍数“跳级”的。为什么这种资料不适于用算术平均数呢？比如要求 5 和 20 的平均，如果用算术平均数的方法则得

$$\bar{x} = \frac{5+20}{2} = 12.5,$$

显然有些偏高，如改用 5 乘 20，然后开平方，结果就得

$$G = \sqrt{5 \times 20} = \sqrt{100} = 10,$$

这样才符合实际情况。这种两个数相乘开平方，或三个数连乘开立方，或 N 个数连乘开 N 次方的方法也是一种“平均”，所得结果叫“几何平均数”，习惯上用 G 表示。仿此，现在我们来计算上面 12 个滴度的几何平均数。

$$G = \sqrt[12]{5 \times 5 \times 5 \times 5 \times 5 \times 5 \times 10 \times 10 \times 10 \times 20 \times 40} = (5^7 \times 10^3 \times 20 \times 40)^{\frac{1}{12}},$$