YIShI ShIYONG TONGJI

医师实用统计

孙瑞元 著

中国微循环与莨菪类药研究学会出版

一九八三年六月 宁波

前　言

随着微循环—莨菪类药研究工作的深入发展，课题选择范围广泛深入，科研设计和统计分析日益受到重视。鉴于广大会员大多是从事基层临床工作的医师，科研设计、统计知识欠缺，我会于去夏在四川举办了“科研设计与统计”学习班，深受广大会员的欢迎，但限于条件、人数，远远不能适应实验研究和临床工作之需要。广大会员迫切期望得到有关学习材料，以提高科研基本统计方法之技能。我会征得理事孙瑞元教授的同意，将他撰写的“医师实用统计”，奉献给广大会员及读者。

本书特点是从实际出发，通俗易懂。便于读者思考、记忆及掌握要领。页数虽不多，但无论从内容上，编排形式上都有其独到之处，对初学者不失为一本较为理想的入门书，对专业统计人员也是一册较为满意的参考书，是统计学书库中的不可多得的珍品。本书的问世，无疑会给学会今后的科研工作带来深远的影响。

本书亦为学会今后数年内分批、分地区举办科研与统计学习班的教材之一。赠送会员人手一册，为共同提高我会会员研究素质而努力！

学会秘书组

1983．6

目　录

医学统计分析的合理应用

一、概　说

（一）医学统计分析的意义：

统计分析在医学研究中的应用，已经愈益受到重视。这是因为：它可以帮助我们，根据一定数量的观察数据，归纳出具有规律性信息，进而以一定的把握程度，推断出带普遍意义的客观结论。概而言之，统计分析是使我们“分析样本，推断总体，透过偶然，揭示规律”的有力工具。

统计分析又是建立在①设计合理、②观察客观、③资料完整、④记录准确这四项基础上的。研究方法不合理或数据不准确时，任何统计分析都是无能为力的。

医学统计分析是为医学研究服务的一种手段。其中统计方法和计算公式的论证常涉及到高等数学，对于医学研究工作者来说，可以暂且不去追究公式的来源和理论证明，但应该对统计分析中的术语有着明确的概念，对各公式的适用范围和适用条件有充分的认识，能够正确地选择最恰当的统计方法。因此，本文不详细进行数理论证，但在如何合理应用医学统计分析方法上结合具体例子，进行较详细的讨论。

（二）医学分析中的一些基本概念：

1、总体与样本：理论上应观察的全部对象，也即研究结论所反映的对象，称之为“总体”。例如：我国所有初生儿的血糖值，某种高血压大鼠模型的平均血压值……等等。

显然总体的例数是相当多的，要对它进行全面的观察是不容易甚至是不可能的事。实际工作中，我们可以在总体中抽取一部分例数，作为直接研究观察的对象，这就是“样本”。

从全部观察对象中抽取一部分对象进行研究的过程称为“抽样过程”。样本的均数往往并不完全等于总体均数，其间的误差，称为“抽样误差”。

我们希望通过分析样本，达到推论总体的目的，也希望抽样误差尽可能小一些。这就要样本具有①一定数量②一定质量。

①样本例数：没有数量也就没有质量。样本例数越多，则从样本推断总体的能力也就越大，但是人力物力的耗费也越多。如何根据以往经验，估计样本例数的问题，是实验设计中的主要问题之一将在后面介绍。一般说来，对于测量数据为指标的研究，以20～50例为宜。误差较大或以百分率为指标者，样本可扩大到50～200例，调查性资料有时要用几百或几千例。

②样本质量：首先是样本在质量上应能够代表总体，也即样本与总体应有“同质性”。我们不能随便在门诊部找一批病人，就作出该地区某病发病率的结论，而应该根据该地区人口中年龄、性别、职业、居地等因素的比例情况，相应地按比例地抽取一些人作为该地人口的样本，这样才能较好地反映该地区的发病率。其次样本的质量还表现在资料的真实性和准确性上，否则就谈不上统计分析。

2、精密度与准确度：

多次重复测量时，各次数据的差别越小，说明研究工作的精密度越好。反之，表示精密度差，应当改进测量方

法，控制干扰因素，必要时也可增多测量次数，来加以补救。

对样本各例的测量误差中，还有一种偏性误差，如天平不准确，比色计有偏差等。每一测量数值都受该偏性误差的影响。尽管几次测量，数据很接近，精密度很好，但都不够准确，这时增加例数，也不能补救，但可按偏差理论原则进行校正处理。

精密度与准确度含义不同。精密者不一定准确，准确者不一定精密。我们的研究工作，对样本数据应力争又精密又准确。

3、偶然性与规律性。

世界上各种事物均有其规律性。在医学领域中，这种规律性常常被大量偶然因素所干扰或掩盖。但是通过调查和统计分析，我们可以在偶然因素的干扰中找出其规律性。例如钉螺的长度和重量各不相同，有的长一些，有的短一些，偶而也会碰到特别长或特别短的钉螺。我们虽然不能予言某湖沼地钉螺的重量，但是，我们通过对该地区钉螺重量的抽样分析，却可有95%的把握认为其重量在 77±28mg 范围内，有99%的把握度认为在77±37mg的范围内。再如，在雌雄各半的一大群动物中，随意抽取一只，我们虽不能予言其雌雄，但是根据统计学原理，可以推算出，取一只为雄性的发生率为50%（即0.5），连取二只均为雄性的发生率为0.25（=0.5×0.5），甚至还能得到其规律性：即连取n只均为雄性的发生率应为0.5^n。

4、概率与可信限。

概率又称机率，是总体中某事件的理论发生率。前述雌雄各半的一大群动物中，雄性的发生率（概率）为50%（0.5）。

一般情况下，总体中某事件的概率是未知的。因此我们可以通过多次重复观察样本，从样本中雄性发生率来推断总体的概率。所以概率也可以认为是无数次重复时，该事件的发生率。

在研究工作中观察到的百分率或均数，毕竟是由数量有限的样本中得到的，因此也仅仅是接近而不是等同于总体的概率或均数。根据统计学原理，我们可以在百分率或均数的上下扩大一定范围，使总体均数处于该范围内的可能性为95%，这就是“95%可信限”。

请注意，事实上我们并未测到总体的均数，但根据统计分析，却可以有95%甚至99%的把握，指出总体均数所在的范围。如果样本例数很大，重复次数很多，我们推断总体的准确性也就愈大。一般统计学上常用样本均数加减95%可信限或99%可信限来表达总体均数的预期范围。

5、显著性测验与显著性水平

医学研究中常需比较两组或各组数据的均数（或百分率）的差别在统计学上是否有显著意义。这种测验方法就是“显著性测验”。

由于生物差异及抽样差异的客观存在，我们比较两组均数时，就不能只看两组数值上的差别，而要分析这种差别是否基本上超过了生物差异及抽样差异的范围，是否反映着两组总体均数存在着差别。测验的目的是算一算，由于偶然因素的机遇影响，出现这种均数差别的可能性有多大。

由统计学术语来说，①先做出“无效假定”，即假定两组资料实际来自同一总体。目前的差别只不过是抽样误差所引起的；②根据两组样本的实测数据和例数，计算出上述

"无效假设"的可能性有多大？③如果这种可能性小于5％（或1％），就可以认为两组数据来自同一总体的可能性很小，可以否定"无效假设"，而承认两组均数的差异在统计学上有显著（或非常显著）的意义。

通常以"P"表示无效假设可以成立的概率。P值越小，无效假设成立的可能性越小，两组差别的统计意义越大，越不易批驳。因此，P值可以通俗地理解为"批（P）驳两组差别有统计意义的可能性"。

医学统计学传统上规定了，P＝0.01及P＝0.05为"显著性水平"作为判断显著意义的标准。

① P≤0.01，批（P）驳的可能性小于、等于1％，也即承认两组间差别有统计意义的可能性大于、等于99％。此时可作统计结论为："两组间差别有非常显著的意义"（过去定结论为"两组间差异非常显著"，易误解为"两组间差异非常大"，其实，"非常显著"是指统计意义很大，后者是指两组均数的差值很大，并非同一含义。）

② P≤0.05即0.01＜P≤0.05批驳的可能性小于、等于5％，说明有95％以上的机率可认为两组不是来自同一总体，下结论为："两组间差别有显著意义。"（过去定结论为"两组间差异显著"）。

③ P＞0.05批（P）驳的可能性大于5％，说明股组均数尽管不同，但该差别来自抽样误差的机率大于5％，已不容忽视。只能结论为："两组间差异无显著意义"（过去定结论为"两组间差异不显著"）。

近年来，也有人建议再加上两条"显著性水平"，作为辅助的判断标准，一是在药物筛选时，用P≤0.1作为筛选合格的标准。一是增加P≤0.001作为标准，表示两组不是来

自同一总体的可能性大于、等于99.9%。但是，这两条辅助标准尚未得到普遍承认。

6、统计结论与专业结论。

"有显著意义"是统计结论，表示着统计推断（否定无效假设）的可靠程度。只说明两组的总体很可能不同，并不说明差别的大小更不说明差别之间有何因果关系。"无显著意义"只是说明根据现有数据不足以否定无效假设，并不是说没有差异，更不是说两组基本相同。

在做专业结论时，除了要考虑统计结论外，还要结合专业知识全面考虑，不能做出脱离实际或脱离本题的结论。例如两组尿量每日只差100毫升，或血压只差5mmHg。这时即使统计结论差别有显著意义，实际上这细微差别并没有临床价值，不应做出有利尿作用或降压作用的专业结论。

二、计数资料（百分率）的显著性测验

凡是每个观察对象只能划分类别，在总数中能数出某种类别有几例，从而可以用百分率来表示数据的资料，称为计数资料或质反应资料，也可通俗地称为百分率资料。

（一）两组百分率资料的分析——$\chi^2(2\times2)$法。

χ^2（2×2）法又称χ^2四格法，是1900年K. Pearson提出，后经Yates改进校正，它是专用于两组百分率的最常用的显著性测验。χ是希腊字母，读音为"Kai"，故χ^2法又称"卡方法"。

$\chi^2(2\times2)$法公式为

$$\chi^2=\frac{(|ad-bc|-0.5N)^2\cdot N}{(a+b)(c+d)(a+c)(b+d)}\qquad \text{公式(1)}$$

判断标准：χ^2值<3.84，则P>0.05，差异无显著意义

χ^2值≥3.84，则P≤0.05，差异有显著意义

χ^2值≥6.63，则P≤0.01，差异有非常显著意义

1. $\chi^2(2\times2)$法的计算步骤

①先将资料中例数排成四格表，不用百分率；

②代入公式(1)，计算χ^2值，χ^2值是一种统计值，该值越大，统计意义也越大；

③根据χ^2值大小，按判断标准判断P值大小；

④做出统计结论。

2. $\chi^2(2\times2)$法的例算：

某次研究，甲药组50人，治愈率70%，乙药组50人，治愈率50%，问两组差别有无统计意义？

①先列四格表 甲药组50人，治愈率70%，则治愈35人，未愈15人，乙药组50人，治愈率50%，则治愈25人，未愈25人。

列表如下：

	治　愈	未　愈	共　计
甲　组	35(a)	15(b)	50(a+b)
乙　组	25(c)	25(d)	50(c+d)
共　计	60(a+c)	40(b+d)	100(N)

②将a、b、c、d、N代入公式(1)计算χ^2值：

$$\chi^2=\frac{(|35\times25-15\times25|-0.5\times100)^2\times100}{50\times50\times60\times40}$$

$=3.375$

③判断：

χ^2值$=3.375(<3.84)$，故$P>0.05$，差异无显著意义

④结论：

甲药痊愈率为70%，乙药痊愈率为50%，但两组间差异无显著意义，不能认为甲药疗效大于乙药。

3．$\chi^2(2\times2)$法的正确应用及注意事项：

两组百分率的显著性测验方法很多，其中以Fisher直接机率法最为准确，但计算过于繁复。χ^2法、t值法等均为近似计算法，其中以式（1）的χ^2法误差最小，最为常用。过去有种误解，似乎当例数小于30例时，或理论频数小于5时，χ^2原法才有误差，否则不必进行校正。事实上，我们已经证明了；只要是判断$P=0.05$或$P=0.01$的显著，χ^2校正的式（1）肯定优于χ^2原法，即使例数大于30，或理论频数大于5，采用式（1）也比χ^2原法的误差为小。χ^2原法所得χ_0^2值常常偏大，以致出现第一类错误，即假阳性错误，错误地否定了无效假设，把本质无显著意义的资料误判为有显著意义，当P值在0.05或0.01附近时，就可能由此做出错误的结论。

上例两组均为50例，χ^2原法计算时就得出了错误结论。现与其他几种两组百分率的显著性测验法，列表比较如下：

方法	计算结果	相应的P值	误差	结论	正确性
Fisher直接机率法	$P=0.0656$	0.0656		$P>0.05$ 不显著	正
χ^2法式(1)	$\chi^2=3.375$	0.0662	0.0006	$P>0.05$ 不显著	误差很小
χ^2原法	$\chi_0{}^2=4.167$	0.0413	0.0243	$P<0.05$ 显著	误差大，并有结论错误
t法	$t=2.04$	0.0452	0.0204	$P<0.05$ 显著	
t校正法	$t=2.03$	0.0459	0.0197	$P<0.05$ 显著	

χ^2法的四格表形式，可以直接代入数据进行演算，但要首先分析资料性质，不能盲目地套用公式，这也是易出现的错误，应予注意：

（1）注意χ^2法式（1）不能适用于四格表中出现0或1的资料，此时例数太少，χ_0^2原法误差极大，χ^2法虽然误差较小，但在$P=0.05$或0.01附近时，也可能做出错误结论。此时，应采用Fisher氏直接机率法或孙氏简化机率法（见后）。

（2）注意χ^2法中两组的内部构成是否一致。科研设计时应将两组内部各年龄的构成比，病情轻中重的构成比基本取得一致。如受条件限制，两组构成比不同，甲组重病比重大，乙组轻病比重大，这时就不能简单地用总人数及总阳性数来计算χ^2值。例如：

病情	甲	法		乙	法	
	人数	有效人数	有效率	人数	有效人数	有效率
重	38	8	21.05%	20	3	15.00%
中	40	35	87.50%	85	53	62.35%
轻	2	2	100%	15	12	80.00%
总计	80	45	56.25%	120	68	56.67%

本例中，从病情轻、中、重三行来看，甲法有效率均高于乙法，但总计以后，甲法有效率反而低于乙法。这一矛盾产生的原因就是两组构成比不同，我们不能用总计数据来计算 χ^2 值，而要采用“构成比不同的两组百分率的 χ^2 测验”（见后）。

（3）注意总计例数是否合乎实际。在比较两法检验阳性率或两疗程有效率时，有时是每一样本分别用甲法或乙法检验，或同一病人经历了一个或两个疗程，列成四格表以后，在形式上与两组百分率的四格表很相似，但不能用式(1)来计算 χ^2 值，而要采用“两种处理对比的 χ^2 检验”，（参考后节）。例如下面四格表就有错误，其错误就在于最后的总计例数（N）是不符合实际的。

两种检验方法对比

	阳性	阴性	总计
甲法	40	10	50
乙法	30	20	50
总计	70	30	100

错误，因实际只检查了50份标本

两个疗程对比

	有效	无效	总计
用一个疗程后	40	60	100
连用二个疗程后	80	20	100
总　　计	120	80	200

错误，因实际上只观察了100名病人

（4）注意在可以采用测量值均数时，最好不要简单地判分为阳性或阴性来计算百分率。例如在测量血压降低值时，最好用“两均数的t测验法”（见后），而不要以血压下降达到或超过某一值定为阳性，不足该值做为阴性。因为这样做，信息损失太大，有时会因划分界限不同，使同一资料算出两种不同的统计结论。试比较下面的例子：

血压降低值(mmHg)	甲组	乙组
5以下	10	20
5～9	30	30
10～19	30	40
20以上	30	10

以20以上为有效	甲组	乙组
有效	30	10
无效	70	90

$\chi^2 = 11.28 \quad P < 0.01$

以10以上为有效	甲组	乙组
有效	60	50
无效	40	50

$\chi^2 = 1.64 \quad P > 0.05$

（5）注意在多行多列资料中，$\chi^2(2\times2)$法的应用：

对于多行多列的资料，如无等级顺序关系时，一般可采用$\chi^2(R\times C)$法，或$\chi^2(2\times k)$法（见后），进行显著性测验。但这两种方法只能做出各组是否来自同一总体的判断，并不能逐对比较各组间的差别有无显著意义。实际工作中对于这类资料也常采取$\chi^2(2\times2)$法，将甲乙、甲丙、甲丁…丙丁…等逐组逐对比较，作出相应的统计结论。

（6）注意在等级顺序型资料中，不宜用$\chi^2(2\times2)$法。

对于多行多列而又有等级顺序关系的资料，应采用等级顺序法或Ridit法，不宜采用$\chi^2(2\times2)$法或$\chi^2(2\times K)$法，因为χ^2法只能分析各组来自同一总体的可能性，并不考虑各组间的等级顺序关系，盲目套用χ^2法，由于没有充分利用组间的等级型信息，往往会得出不同的统计结论。

（二）两组百分率（数据中有0或1时）的分析——孙氏简化机率法。

四格表数据中有0或1时，$\chi^2(2\times2)$法误差较大，已不适用；Fisher氏直接机率法计算又太繁，此时可用孙氏简化机率法。

$$P = 2\times\left(\frac{1+c+2d}{1+c+2d+2b}\right)^{c}\times\left(1+\frac{abc}{N}\right)$$

适用于a＝o或1　　　　式(2)

式中a、b、c、d、N意义同前，但要予先将四格表转换排列，使a＝0或1，而且b>c，现结合实例，计算如下：

	有效	无效	→		无效	有效
甲组	19	0	转换排列	甲组	0	19
乙组	6	3		乙组	3	6

$$P = 2\times\left(\frac{1+3+2\times6}{1+3+2\times6+2\times19}\right)^{3}\times\left(1+\frac{0\times19\times3}{28}\right)$$

$=0.0520$

结论：P>0.05，差异无显著意义

本例用Fisher直接机率法算得P＝0.0513，误差仅0.0007，结论一致。用χ^2原法，算得$\chi_c^2=6.75$，P＝0.0094，误差太大，结论判断为P<0.01，用χ^2(2×2)法，算得$\chi^2=4.04$，P＝0.0445，误差虽小于χ^2原法，但结论为P<0.05，也有误判。因此这种资料不宜用χ^2及χ_c^2原法，以用简化机率法为宜，有些统计专著列有“四格表显著性查对表”也可供查索。

（三）两组百分率构成比不同时的χ^2测验——权重χ^2法：

本法专用于构成比不同时，先采用两组合并例数作为标准化的依据，由此计算，权重百分率，作为各组应得的百分