



图1 胃双重对比造影检查,可清晰显示胃小区。图2 结肠双重对比检查,显示正常结肠袋和结肠沟。图3,4 腰椎椎管造影正侧位相,示腰3-4和腰4-5椎间盘突出,双侧神经根袖显示不清。图5 数字化血管造影,示腹主动脉 双侧髂动脉动脉硬化病变,左髂总动脉狭窄,右髂总动脉完全闭塞。图6 左侧椎动脉造DSA示正常基底动脉,双侧大脑后动脉及分支。

量图像最重要因素,图像后处理技术仅仅是改善图像质量的一个补偿。

该系统操作简单、方便,能进行全身普通X线检查、血管造影和介入操作,能代替专用机进行各种不同方法检查。

参 考 文 献

1 Kondoh H, Kozuka T. Generalization of digital radiography. Journal of

Japanese Medical Imaging, 1990; 10: 1164

- 2 Pfeiler M, Marhoff P, Hashimoto K. Recent development of image intensifier in digital radiography. Journal of Japanese Medical Imaging, 1990; 10: 1179
- 3 Fairbrother J, Romeo MA, Kramers K. Clinical evaluation of new X-ray fluoroscopy system for universal applications. Electromedica, 1991; 59: 134

(1995-09-26 收稿, 1996-02-26 修回)

(上接第 192 页)

5 SAS Institute Inc. SAS procedures guide, Release 6.03 edition. North Carolina: SAS Institute Inc, 1988; 405

6 SAS Institute Inc. SAS/STAT user's guide Release 6.03 edition. North Carolina: SAS Institute Inc, 1988; 711~726, 941

(1996-01-05 收稿, 1996-08-22 修回)

· 专题讲座 ·

【编者按】由于科学研究离不开统计学, 医学科研也不例外。因此, 越来越多的医学科研工作者自觉地运用统计学知识, 为提高科研的质量和速度, 付出了辛勤的努力, 作出了许多可喜的成绩。然而, 在运用统计学方面, 的确存在一些问题, 有些问题还相当严重, 且影响面较宽。为了真正发挥统计学在医学科研中应起的作用, 使在学术论文中运用统计学的水平得到较大提高, 本刊特组织撰写了“合理选用统计分析方法”专题讲座。该讲座明确指出了在运用统计学中易犯的错误, 扼要地阐明了合理选用统计方法的意义、方法, 并结合部分实例进行了分析和讲解。该讲座针对性强, 论述清楚, 非常实用。我们相信该讲座肯定会对广大读者有所帮助和提高。希望广大读者踊跃参加讨论, 并提出宝贵的意见或建议; 并欢迎读者将处理实际资料中遇到的问题提出来, 共同研究和讨论, 以便把统计学的运用提高到一个新水平。

185-192, 176

合理选用统计分析方法

胡良平¹ 王功鹏²

R31F

众所周知, 统计学是从事科学研究下可缺少的工具。从试验设计、资料收集与表达、数据处理和结果分析, 每一个环节都需要正确地运用统计知识, 才能真正发挥统计学在科学研究中应起的作用。然而, 在已出版和发表的一些学术专著和论文中, 通过评审的科研成果和答辩的学位论文中, 经常可以看到忽视、轻视和误用统计学的现象。下面将揭示一些在统计学运用中普遍存在的问题, 以便引起大家的重视, 使统计学能得到合理应用、促进科学研究向前发展^[1]。

1 统计学运用中存在的问题

1.1 在试验设计方面存在的问题^[2]

1.1.1 样本含量过小 相同试验条件下的试验次数称为样本含量, 用 n 表示。有些论文中 $n \leq 3$, 当指标的变异度较大时, 很难呈现出数据之间的内在规律性。

1.1.2 缺乏完善的试验计划 有些研究工作者对欲研究的课题缺乏细致地思考, 在试验研究开始之前, 不拟定切实可行的试验计划, 边做边看, 致使有些过失无法弥补。如遗漏了某些重要的影响因素或观测指标; 试验记录零乱, 甚至残缺不齐; 由于缺乏完善的试验计划和必要的预试验, 当试验过程中发生异常现象时措手不及。例如: 在测完一组鼠的白细胞(WBC)计数后发现其中有几只鼠的 WBC 值偏高, 于是, 对这几只鼠进行重测, 而对那些认为 WBC 值可以接受的鼠只测一次, 这就人为地造成了资料之间的不平衡性; 再如, 在研究毒物的半数有效量时, 盲目确定毒物剂量, 等做完全部试验后发现: 在所选定的几个剂量组内, 动物的死亡率要么都低于 50%, 要么都高于 50%。这样的资料当然就不适合计算毒物的半数有效量了。

1.1.3 所采用的试验设计类型过于简单 从统计学的角度看, 安排多因素试验的设计类型丰富多彩。例如: 配伍组设计、析因设计、正交设计、均匀设计、具有重复测量的设计和序贯设

计等。研究者应根据具体情况选用不同的设计类型。遗憾的是: 在很多实际是多因素的试验研究问题中, 不少人仍然按单因素的设计思想来设计试验, 使得很多本来只需半年或一年就能结束的科研课题, 实际上花了两、三年才完成; 不仅如此, 用单因素试验研究取代多因素试验研究, 当因素之间存在显著性的交互作用时, 容易得出错误的结论。总之, 这种以“单代多”的设计思想, 好似盲人摸象, 不仅难免其片面性, 更有少、慢、差、费之嫌。

1.1.4 设计类型辨别不清 一些人仅凭专业知识去设立试验组和对照组, 但试验中究竟涉及几个试验因素, 它们之间是怎样组合的以及所使用的设计类型是什么, 往往不予以重视或一概不知。因而, 犯了下述不应犯的错误。

1.1.4.1 拆散多因素的整体设计 不少人将两因素析因设计类型拆成多个单因素多水平设计类型; 而另有人, 将具有重复测量的多因素设计类型拆成几个单因素多水平设计类型。拆散多因素整体设计的后果是减少了检验统计量自由度, 降低了结果的可靠性; 不仅如此, 还无法分析因素之间的交互作用, 不能也不可能正确地揭示多因素之间的相互关系。

1.1.4.2 混淆了因素与水平之间的关系 研究者希望着重考察的试验条件称为试验因素, 如温度、剂量等; 同一个试验因素的不同取值称为该因素的不同水平, 如温度分为 (30、60、90)℃; 剂量分为小、中、大。在医学科研中, 常需进行如下的试验安排: 设立四个组, 即空白对照组、阳性药组、新药组、阳性药与新药合用组。不少人错误地把这四个组简单地看作是单因素四水平的设计, 其实是两因素的析因设计问题。设计类型弄

作者单位 1. 军事医学科学院情报所统计教研室, 北京 100850

2. 军事医学科学院放射医学研究所, 北京 100850

错了,统计分析方法也就随之选错了,其结论往往也就错了。这样的例子在不少论文中都可见到。

1.1.4.3 混淆了重复试验与重复测量之间的关系 在相同的试验条件下,对 n 个不同的受试对象(或样品)进行观测,称为 n 次独立的重复试验;而在不同的条件(通常是在不同的时间点)下,对同一个受试对象反复观测,称为重复测量。这是两种不同的重复,很多人将它们混为一谈,不加区分,影响了资料收集和数据处理合理性。

1.2 在收集和表达资料等方面存在的问题^[2]

1.2.1 收集资料方面存在的问题 在医学科研中,常需用到上述提及的具有重复测量的设计。一些人根据医学试验的客观要求,经常不自觉地运用此设计,但由于对此设计的本质缺乏了解,收集试验数据时不是以各受试者为观察单位,而是以同一种处理下的若干个受试者所形成的小组为观察单位。这样无法辨别出哪些数据是在不同时间点上从同一个受试者身上测得的,从而也就无法将受试者的个体差异从试验误差中分离出来,其结果是人为地扩大了试验误差,错误地改变了分析数据的数学模型。这样的例子在涉及类似试验研究的论文和科研成果中常常可见。

1.2.2 选取数据方面存在的问题 对于在设计和试验过程中就极力消除系统误差的影响后获得的试验数据,通过认真检查,如果确实没有过失误差,则不应人为舍弃或从中随意选用。沿用前述关于 WBC 指标的例子,有人在作新药评价时,遇某鼠有两个测定结果,则从两个数据中取前一个作为对照数据用于甲药,取后一个作为对照数据用于乙药;遇某鼠只有一个数据时,此数据同时被用于两药的比较中。这种随意选择数据的做法,是严重违背统计学原则的。另外,也有人习惯于用“均数 \pm 3 倍标准差”作为舍弃异常值的标准,在小样本的条件下,这一标准的合理性是值得怀疑的。用此标准时,一定要注意应用条件,即样本含量较大,且资料近似服从正态分布。文献[3]给出了舍弃异常值的正规方法,可供使用。

1.2.3 表达资料方面存在的问题 统计表和统计图是表达试验资料的有力工具,但不加区别地使用图、表的现象时有发生。本该用统计表表达的却用了统计图;本该用统计图表达的却用了统计表。在统计表设计中,栏目繁杂,线条过多,纵横交错,看不出欲说明的问题,甚至连表中的数字代表什么,都没有表达清楚。在设计统计图时,把随时间变化的几种病的发病率的变化趋势用条图而不是用本该适合的半对数线图来表示;统计图中坐标,特别是横坐标刻度表示不正确的现象,可从不少论文中看到。

1.3 在选用统计方法方面存在的问题^[2]

1.3.1 在分析定量资料中存在的问题 不少人把各种试验设计一律看成是单因素两水平设计,故一律采用 t 检验分析资料,这是对统计学理论的一种误解或歪曲。事实上,在 t 检验被使用的众多场合下,只有少数场合是符合用此方法的,而多数场合下则应选择方差分析。

1.3.1.1 在对均数作假设检验时适合选用 t 检验的场合 (1)用于单组设计资料,即检验一组定量资料所取自的总体均数与已知的某个总体均数之间是否有显著性差别。(2)用于配对定量设计资料,即检验配对设计下两组定量资料的差值所取自的总体均数与理论均数“0”之间是否有显著性差别。(3)用于单因素两水平设计(常称为成组设计)资料,即检验两组定量资料所取自的两个总体均数之间是否有显著性差别(其前提条件是资料满足正态性和方差齐性)。

1.3.1.2 在对均数作假设检验时适合选用方差分析的场合 首先,观测的指标是定量的,其次,取决于定性的影响因素的个数(设为 k)和因素的水平数(设为 m)。(1)当 $k=1, m \geq 3$ 时,应选单因素设计的方差分析方法;(2)当 $k \geq 2$ 时,无论各因素的水平数是多少,应根据具体的多因素设计类型选择相应的方差分析方法。资料经方差分析之后,通常还要根据具体情况,选择合适的方法进行多个均数之间的两两比较。

1.3.1.3 用 t 检验取代方差分析的不合理之处 (1)资料的利用率低;(2)误差项的自由度变小,结果的可靠性降低;(3)增大了犯假阳性错误的概率;(4)尺度不统一,缺乏可比性;(5)割裂了因素各水平之间的联系。

1.3.2 在进行直线回归和相关分析时存在的问题

1.3.2.1 盲目拟合直线回归方程和计算相关系数 不少人将测得的两组数据输入计算机,利用现成的统计分析软件很快获得了直线回归和相关分析的结果。一看显著性检验的 P 值都小于 0.05,便很有把握地认为所研究的两个变量之间呈直线关系是确信无疑的了。实际上,这样做是容易出问题的。参见本文“7.1.3”。

1.3.2.2 对两型回归问题不加区分 一些人并不了解直线回归分析中存在着两型回归问题,因而,不分资料的具体情况,只要作直线回归分析,一律采取同一个模式,这显然是不妥的。

1.3.2.3 对具有重复试验数据的回归问题简单化处理 在回归分析的资料中,若每个自变量 X 的取值对应着多个因变量 Y 的取值,则称为具有重复试验的回归问题。在处理这种资料时,有人常求出同一个 X 取值条件下多个 Y 值的均值,使具有重复试验的回归问题转化成简单的直线回归问题。这样做的不妥之处在于:人为降低了试验误差;更可惜的是损失了大量有用的信息,导致本来可以回答的一些问题无法回答,参见本文“7.3”。

1.3.3 在分析定性资料中存在的问题 当观测的两个指标都是定性变量时,需将其分类整理成 2×2 列联表或 $R \times C$ 列联表的形式后再作处理。不少人不管表中两个分组指标的性质如何,一律采用 χ^2 检验。这样做,在多数场合下是不妥的。正确的做法是:按表中两个分组指标是名义变量还是有序变量来划分列联表的种类,根据资料所具备的不同条件,选择相应的统计分析方法^[1,4],参见本文“6”。

2 合理选用统计分析方法的途径

2.1 了解统计学的全貌 统计学的内容十分丰富,要全面学

习和掌握它,并非易事。而科技工作者又有自己的专业主攻方向,所以,比较切合实际的做法是了解统计学的概貌,知道统计学究竟研究哪些内容,各种资料不该用什么方法分析,而应该用什么方法分析,才是正确的。只有了解了统计学的全貌,选择统计方法时才能做到胸中有数,有的放矢。实在拿不定主意时,应该向内行请教,千万不要盲目套用,更不应明知故犯。

2.2 弄清各种统计方法的应用范围和适用条件 要想做到合理选用统计分析方法,很重要的一条就是要弄清各种统计方法的应用范围和适用条件。比如说:在对均数作假设检验时,何时适合使用 t 检验?首先,应明确 t 检验的应用范围,即本文“1.3.1.1”中所述的3种场合;而适用条件则是正态性和方差齐性。所谓正态性,就是各组资料(配对设计资料的差值)要服从(或近似服从)正态分布;所谓方差齐性,就是各组资料所取自的总体方差应相等。对于给定的资料,欲知这两个前提条件是否满足,可通过假设检验来回答。在本文“1.3.1.2”中陈述了适合选用方差分析的场合,而方差分析时也要求资料具备正态性和方差齐性。同理,对于回归分析资料 and 列联表资料,应当如何正确地处理,在本文“7”中将作较详细的介绍。

2.3 学会运用现成的统计软件分析资料 在选择正确的分析方法后,复杂的统计计算常常令人生畏。要看懂各种复杂的计算公式,并自己编程计算,对多数实际工作者来说又是一个很难逾越的障碍。克服这一困难的有效途径是学会利用一个好的统计软件实现各种统计计算。不仅如此,还应重视结合专业和统计学知识,对计算结果作出正确的解释。在此推荐国际上著名的统计分析系统—SAS软件包^[5,6],因为它功能齐全,适用面宽,为全世界所公认;另外,我们已摸索出一条使用SAS的捷径,包括一本书^[1]和一张装有SAS引导程序的软盘,读者稍花气力学习,便可方便成功地运用SAS软件,并会解释计算结果。

3 表达统计资料时需注意的问题

3.1 统计表 运用统计表表达数量资料时,应遵循“简单明了”的原则。有些人编制统计表时只注重“简单”,但却忽视了“明了”。尤其是表中的数据代表什么,读者反复看也难看懂。在编制复合统计表时,一定要把表中的数据总称(如病例数、眼数、阳性率等)明确地写出来。参见本文“6”中表格的式样。

3.2 统计图 绘制统计图时,一定要根据资料的性质选用相应的统计图。例如,为了显示5所医院某年床位利用率时,可选用条图。因为医院名称之间没有数量大小或顺序之分,可以把它们任意地安排在横轴上;而为了显示5所医院10年来床位利用率的变化情况时,则适合选用半对数线图,因为此时横坐标上的刻度是连续变化的时间,若仍选用条图,就把连续变化的数量关系割裂开了,不能很好地发挥统计图应起的作用。绘制统计图时易犯的另一个错误就是横坐标上刻度标得不符合数学原则,例如:时间(天)分别为25,40,50,56,62,68,在横轴上所表示的方法既不是算术尺度,又不是对数尺度。

3.3 百分比 不少人在不同的条件下用 $n/3$ 个受试对象做

试验,也用百分比表示结果。例如当2个受试者($n=2$)中有1个阳性结果时,便用50%表示其阳性率,这是误用百分比最常见的实例。 n 达到何值时才适合求百分比,虽然没有统一的定论,但从“百分比”字面理解, $n \geq 100$ 时更为合理。考虑到百分比的资料常需用 χ^2 检验来处理,而分析四格表资料时,若两组总例数小于40,在统计学教科书中都要求采用Fisher的精确算法,而不宜使用 χ^2 检验。基于这一理论,在实践中,建议用 $n \geq 20$ 作为计算百分比的下限。当总例数小于20时,一般直接列出总例数和阳性数即可。

3.4 平均数 表达定量资料的平均水平需用平均指标,但并非所有的定量资料都适合用算术平均值 \bar{x} 来反映平均水平,因为 \bar{x} 只适合于呈正态分布(至少是近似正态分布或对称分布)的资料,而呈正偏态分布的资料(通常标准差 $S > \bar{x}$)宜选用几何均值;呈极严重正偏态分布的资料宜选用调和均值;信息不全的资料(如含有失访数据的生存资料)宜选用中位数;欲反映一组定量资料的峰值(如食物中毒的潜伏期)时则宜选用众数^[1]。应该指出,若在相同试验条件下某定量观测指标的变异度较大时,建议重复试验次数最好 $n \geq 5$ 为宜;若 $n < 5$,有时结果不够稳定。

4 三种简单设计下定量资料均值的假设检验^[1]

4.1 单组设计定量资料的统计分析

4.1.1 设计、试验与资料 首先,需按“估计总体均数时所需的样本含量”计算公式算出样本含量 n 的值。然后,随机抽取 n 个个体,并测试所关心的指标之值,此设计称为单组设计。

【例4.1.1】已知正常人乙酰胆碱脂酶的均值为1.44U,现测得13例慢性气管炎患者的乙酰胆碱脂酶分别为:1.50, 2.19, 2.32, 2.41, 2.11, 2.54, 2.20, 2.36, 1.42, 2.17, 1.84, 1.96, 2.39。问:慢性气管炎患者乙酰胆碱脂酶的总体均值与正常人有无显著性的差别?

【分析】已知的总体均值 $\mu_0 = 1.44$,13例慢性气管炎患者所代表的新总体均值 μ 未知,研究的目的是检验 μ 是否等于 μ_0 , μ_0 的信息靠13名患者乙酰胆碱脂酶样本的均值提供。

4.1.2 前提条件的检验对一组数据只需作正态性检验,可用SAS中UNIVARIATE过程^[5]来实现。检验结果为 $W = 0.8966$, $P = 0.1171$,说明这组资料服从正态分布,可运用单组设计资料的 t 检验来回答上述问题。

4.1.3 单组设计资料的 t 检验^[5]已知: $n = 13$, $\mu_0 = 1.44$ 。用SAS中UNIVARIATE过程可算得: $\bar{x} = 2.108$,标准误差 $S_x = 0.09546$, $t = 7.003$, $P = 0.0001$ 。

【结论】可以认为该病患者乙酰胆碱脂酶的总体均值高于正常人相应指标的取值。

4.1.4 单组设计资料的符号秩检验^[5]如果这组资料不服从正态分布,就需改用符号秩检验。

4.2 配对设计定量资料的统计分析

4.2.1 设计、试验与资料 首先,需按“定量资料的配对设计

时所需的样本含量”计算公式算出样本含量 n 的值。然后,进行随机抽样和试验。从形式上看,配对设计有下列三种情形。

4.2.1.1 自身配对设计

【例 4.2.1】 某医院用中药治疗 9 例再生障碍性贫血患者,测得各患者治疗前后血红蛋白(g/L)变化的数据(从略)。问:治疗前后是否有显著性差别?

4.2.1.2 同源配对设计

【例 4.2.2】 从 8 窝大鼠的每窝中选出同性别、体重相近的 2 只,分别喂以水解蛋白和酪蛋白饲料,4 周后测定其体重增加量(数据从略)。问:两种饲料对大白鼠体重的增加量有无显著性的影响?

4.2.1.3 条件相近者配对设计

【例 4.2.3】 为研究三棱莪术液的抑瘤效果,将 20 只小鼠按某些重要的非处理因素配成 10 对,然后把每对中的 2 只小鼠随机分到实验组和对照组中,两组小鼠都接种肿瘤,实验组在接种肿瘤 3d 后注射 30% 的三棱莪术液 0.5ml,对照组则不加任何处理,记录各对鼠瘤体改变量数据(从略)。问:两组瘤体大小的总体均值之间是否有显著性的差别?

4.2.2 前提条件的检验及其数据处理 上述三种配对设计定量资料的处理方法是相同的,即根据资料的条件,分别选用配对设计定量资料的 t 检验或符号秩检验。先将各对数据按相同的顺序相减,求出一组差值,把这组差值看作来自单组设计的资料,需要明确的一点是:此时的 $\mu_0 = 0$ 。以下的工作就与前面“4.1”节完全相同了,因篇幅所限,恕不赘述。

4.2.3 配对设计应用举例

【例 4.2.4】 在无震动和有一定震动的情况下,同时测定 10 份血样的血细胞沉降速度(mm/h),结果如下。问:该震动对血沉

编号	1	2	3	4	5	6
甲药:						
治疗前	6.78	7.38	6.62	6.15	6.00	7.10
治疗后	6.27	6.48	5.77	5.55	5.30	7.10
乙药:						
治疗前	6.80	6.50	7.34	7.15	6.10	6.35
治疗后	5.44	5.40	5.03	5.04	5.97	4.68

4.3.2 前提条件的检验 如果各药物组只有一组数据,通常可直接检验两个前提条件是否满足;若各组内部数据成倍增加,需对数据取对数变换后再检验前提条件;而本例本应属于具有重复测量的两因素(测定时间和药物)设计^[1,6],由于只有两个时间点,故可用相减的办法消除时间因素的影响,分别用甲药组、乙药组的差值作检验,将具有重复测量的设计简化为单因素两水平的成组设计。下面仅给出分析该资料的结果。

分别对两组差值作正态性检验。检验结果如下:甲药组, $W = 0.8608$, $P = 0.1248$;乙药组, $W = 0.9751$, $P = 0.9158$,说明两组资料都服从正态分布。再检验方差齐性是否满足, $F = 6.65$, $P = 0.0191$,在 $\alpha = 0.05$ 水平上拒绝两总体方差相等的零假设,故此资料需用近似的 t 检验(即 t' 检验)来处理。

测定有无影响?

样本号	1	2	3	4	5	6	7	8	9	10
无震动	3	3	1	90	2	7	67	27	18	32
有震动	6	5	1	123	3	11	74	33	21	38

【分析】 这个例子实际上可视为用两种方法(标准方法与新方法)对同一组样品进行检测,看看两种方法检测的结果是否一致。可用两种统计方法来分析此类资料,其一是用配对设计的 t 检验(若各对数据的差量满足正态性)或符号秩检验(若各对数据的差量不满足正态性);其二是用直线回归分析,请参见文献[3]

对差量进行正态性检验的结果为: $W = 0.5897$, $P = 0.0001$,说明此资料不服从正态分布,需用符号秩检验。检验结果为: S (即 $S_{gn} Rank$) = 22.5, $P = 0.0039$,说明震动对血沉测定有非常显著性的影响。具体地说,原血沉值越高,震动后的血沉值增加的幅度也就越大。值得注意的是:有些人用相关分析来回答上述问题,那是不妥的。因为即使两种方法的结果很不一致,只要一种方法测定的结果总高于另一种方法的测定结果(幅度接近相同)时,相关系数仍会很大。

4.3 成组设计定量资料的统计分析

4.3.1 设计、试验与资料 应按“成组设计时所需的样本含量”计算公式算得 n 的值。通常需将全部 n 个受试者随机地均分成两组,一组为对照组,另一组为试验组,观测一个或多个指标的取值。

【例 4.3.1】 用甲、乙两种药物治疗高血脂症,两种药物治疗前后血清胆固醇浓度(mmol/L)变化的测定结果如下。问:两种药物的疗效之间有无显著性的差别?

7	8	9	10	11	12

4.3.3 成组设计资料的方差齐性检验、 t 检验和 t' 检验^[1,6]

用 SAS 中 TTEST 过程可同时完成方差齐性检验、 t 检验和 t' 检验,只需在看计算结果时细心一点即可。本例 $t' = 1.309$, $P = 0.2097$ (Satterthwaite 法)。

【结论】 甲、乙两药对治疗血脂病的疗效之间无显著性差别。

4.3.4 成组设计资料的秩和检验 如果两组资料(或经过变量变换后)中至少有一组不服从正态分布,就需改用秩和检验。可用 SAS 中 NPAR1WAY 过程^[1,6]来实现计算。

5 方差分析与两两比较

5.1 方差分析的适用范围与前提条件 在本文“1.3.1.2”和“2.2”中分别介绍了方差分析的适用范围和前提条件,在此不

再赘述。

5.2 方差分析的种类 方差分析的种类很多,视设计类型不同而异。其本质区别在于两点:第一,全部数据构成的总离均差平方和被分解的项目不同,如单因素设计时,只分解出组内和组间(即误差)两项;配伍组设计时,可分解出处理因素、配伍组因素和误差三项来。第二,衡量各因素重要与否的F统计量的分母视设计类型不同而异。可分为误差项相对固定类型和误差项变动的设计类型^[1]。

不少人下仅不详细区分各种方差分析,相反,干脆用t检验取代一切方差分析。本刊第2卷第3期第174页上表2资料,实际上取自具有重复测量(有5个不同的时间点)的设计(无法简化成单因素两水平的成组设计),本应选用与该设计对应的方差分析方法,但原作者却误用了t检验。

5.3 两两比较^[1] 资料经方差分析后,若某些单个试验因素有显著性,往往需要对这些因素的各水平下的均值进行两两比较。这种比较通常分为两类:其一,全部均值中任何两个之间都要比较;其二,全部试验组分别与同一个对照组比较,而试验组之间不比较,可用DUNNETT法。根据所控制的误差类型,又可细分为许多具体的比较方法。人们较常用的Q检验(SAS中称为SNK法),用于第一种场合,且仅控制EERC(即在完全无效假设下,作完全部比较犯I型错误的概率),但多次比较的MEER(即在任何完全或部分无效假设下,作完全部比较所犯I型错误的概率之最大值)将趋近于1。此处建议用REGWQ法或REGWF法。若两个因素的交互作用项有显著性,则需将一个因素分别控制在不同水平上,对另一个因素各水平下的均值进行两两比较。

6 定性资料的统计分析^[1,4]

6.1 2×2表资料的假设检验

6.1.1 2×2表资料的设计类型

表 6.1.1 手术持续时间与切口感染关系

手术持续时间	手 术 例 数		
	感染情况:	感染	未感染
≤5h		13	229
>5h		7	56
合计		20	285

6.1.1.1 完全随机设计 如表 6.1.1 所示,就是从全体手术者中随机抽取 305 例,然后按手术持续时间(分为:≤5h 和 >5h)以及切口是否感染(分为感染与未感染)分成 4 格分别计数,得到上面的资料,任何一位患者,事前都不知道自己会出现在哪一个格内,故叫做完全随机设计。

6.1.1.2 配对设计 如表 6.1.2 所示,156 例患者分别用静脉

造影与 PPG 两种方法对交通静脉瓣膜功能进行检测,根据两种方法检测结构相同或不同,划分成 4 格。目的是看两法检测结果是否一致。

6.1.1.3 单侧固定的设计 (1)队列研究设计,如表 6.1.3。它是先将 565 名受试者按体检时血压是否正常分成两组(如偏高与正常编);再对每组中的个体进行追踪观察 6 年,并按是否患冠心病分成两组(如:患病与未患病),将全部 565 名受试者分到 4 个格子中去,得到相应的资料。(2)病例一对照研究设计,如表 6.1.4。它是先将 250 名孩子按染色体是否正常分为两组(异常与正常),再对每组孩子的母亲进行回顾性调查,并按其在孕前半年来服用避孕药与否分成两组(服药与未服药),将全部 250 名孩子分到 4 个格子中去,得到相应的资料。

表 6.1.2 两法对交通静脉瓣膜功能诊断结果

造影诊断结果	例 数		
	PPG 诊断结果:	正常	逆流
正常		73	27
逆流		11	45
合计		84	72

表 6.1.3 查体时血压值与随访 6 年后是否患冠心病之关系

查体时血压正常与否	(患冠心病与未患冠心病)例数		
	追踪结果:	患病	未患病
偏高		19	61
正常		20	465
合计		39	526

表 6.1.4 母亲孕前半年来服用避孕药与子代染色体状态之关系

孕前半年内服用避孕药情况	子 代 例 数		
	染色体状态:	异常	正常
服药		34	20
未服药		68	128
合计		102	148

6.1.2 2×2表资料的统计分析方法

6.1.2.1 完全随机设计定性资料的统计分析方法 取自完全随机设计的 2×2 表资料,需对两个分组变量进行独立性检验,可用多种统计方法,但需根据资料所具备的条件选择相应的计

算公式 设表中任何一个格内的理论频数为 T , 则

$$\left\{ \begin{array}{l} n \geq 40 \text{ 且 } T \geq 5, \text{ 可用一般的 } \chi^2 \text{ 检验公式;} \\ n \geq 40, \text{ 但至少有一个 } T \text{ 满足 } 1 \leq T < 5, \text{ 用校正的 } \chi^2 \\ \text{检验公式;} \\ n < 40 \text{ 或 } T < 1, \text{ 用 Fisher 的精确检验, 直接算出概率.} \end{array} \right.$$

对于表 6.1.1 资料, $n = 305 > 40$, 最小的理论频数 $T_{11} = 63 \times 20/305 = 4.13 < 5$, 故需选用校正的 χ^2 检验公式。当然, 用 Fisher 的精确检验更好。

本例, 校正的 $\chi^2 = 1.832$, $P = 0.176$; Fisher 的精确概率 $P = 0.147$ 。

【结论】 手术持续时间长短与切口是否感染无关。如果出现切口感染, 应认真寻找其他原因。

6.1.2.2 配对设计定性资料的统计分析方法 在一般统计教科书中, 都用配对设计的 χ^2 检验分析两种检测方法不一致部分差别是否具有显著性, 但此法的计算公式中并未利用总样本含量 n 所提供的信息。假定 n 无限增大, 两法不一致部分基本保持不变, 则两法的一致率将趋近于 1, 不一致部分相对于 n 来说, 微乎其微; 而配对设计的 χ^2 检验, 一旦出现显著性的结果, 无论前述的条件怎么改变, 结论都不会改变; 这说明此法有待进一步改进。

如果一定要看两种测定方法之间差异的大小, 可按单侧固定设计的分析方法(详见后)直接检验两法所对应的“阳性率”之间是否具有显著性。若研究的目的是看两种方法测定结果之间吻合程度的高低, 不妨直接用 Kappa 检验法(一般用于 3×3 以上的方表)检验两种测定方法的一致性。这两种统计分析方法都较好地克服了配对设计 χ^2 检验的不足之处。本例的三种分析结果如下:

(1) 用配对的 χ^2 检验

其结果为 $\chi^2 = 5.921$, $P = 0.015$, 说明两种测定方法不一致部分相差有显著性;

(2) 用 U 检验

$$X_1 = 100, n_1 = 156, P_1 = 100/156 = 0.641026;$$

$$X_2 = 84, n_2 = 156, P_2 = 84/156 = 0.538462$$

$$\begin{aligned} Z_1 &= \ln[X_1 + 0.5]/(n_1 + X_1 + 0.5) \\ &= \ln[(100 + 0.5)/(156 + 100 + 0.5)] \\ &= 0.575917 \end{aligned}$$

$$\begin{aligned} Z_2 &= \ln[X_2 + 0.5]/(n_2 + X_2 + 0.5) \\ &= \ln[(84 + 0.5)/(156 + 84 + 0.5)] \\ &= 0.153165 \end{aligned}$$

$$V_1 = 1/[n_1 \cdot P_1 \cdot (1 - P_1)] = 0.027857;$$

$$V_2 = 1/[n_2 \cdot P_2 \cdot (1 - P_2)] = 0.025794$$

$$U = (Z_1 - Z_2) / \sqrt{V_1 + V_2} = 1.825, P = 0.068$$

说明两法测定结果之间无显著性的差别, 即两种诊断方法

诊断的结果吻合程度较高

(3) 用 Kappa 检验

$$\text{观察的一致率 } P_o = (73 + 45)/156 = 0.756410$$

$$\begin{aligned} \text{期望的一致率 } P_e &= (100 \cdot 84/156 + 56 \cdot 72/156)/156 \\ &= 0.510848 \end{aligned}$$

$$\begin{aligned} K = \text{Kappa} &= (P_o - P_e)/(1 - P_e) \\ &= (0.756410 - 0.510848)/(1 - 0.510848) \\ &= 0.502016 \end{aligned}$$

$$\begin{aligned} \text{Kappa 的标准误差 } S_k &= 0.510848 - 0.510848^2 - [100 \cdot \\ &84 + (100 - 84) - 56 \cdot 72 + (56 + 72)] / \\ &156^3)^{1/2} / [(1 - 0.510858) \sqrt{156}] \\ &= 0.078284 \end{aligned}$$

$$U = K/S_k = 0.502016/0.078284$$

$$= 6.413, P < 0.001$$

说明两种方法测定结果的一致率(74.64%)较高。可用一种较为经济(或较为简便)的检测方法取代另一种较为耗费(或较繁琐)的检测方法。

说明: 从两种不同角度分析同一个配对设计定性资料, 有时配对设计的 χ^2 检验的结论与另两种方法的结论是矛盾的(如本例), 这是值得注意的。建议: 根据研究的目的, 选用后两种分析方法之一为宜。

6.1.2.3 单侧固定设计资料的统计分析方法 对于单侧固定设计的那两种情形, 若研究的目的是看血压偏高与正常者中患冠心病的比例有无显著性差别(见表 6.1.3)或子代染色体异常与正常者中母亲孕前服避孕药的比例有无显著性差别(见表 6.1.4), 在统计教科书中习惯选用一般的 χ^2 检验^[1,3], 但有些统计学专家建议先对资料作经验 logistic 变换, 然后用 U 检验分析^[4], 见前面“6.1.2.2”中的(2)。

对于表 6.1.3 资料, 若研究的目的是要看体检时受试者血压是否正常对未来若干年后他们患冠心病的危险性有多大, 则应计算相对危险度, 并作相应的检验; 对于表 6.1.4 资料, 若研究的目的是要看子代的染色体异常是否受其母亲在孕前半年内服用避孕药的影响, 则应计算比数比, 并作相应的检验^[1,3], 具体计算从略。

6.2 $R \times C$ 表资料的假设检验

6.2.1 $R \times C$ 表资料的分类 根据 2 维列联表中 2 个分组变量的类型(即名义的或有序的)以及分析的目的, 可对 $R \times C$ 表资料进行分类。因为不同类型的资料或不同的分析目的, 有不同的分析方法。为直观起见, 首先, 请看几组实际资料:

双向无序列联表, 见表 6.2.1 和表 6.2.2; 单向有序列联表, 见表 6.2.3; 双向有序且属性不同的列联表, 见表 6.2.4; 双向有序且属性相同的列联表, 见表 6.2.5 和表 6.2.6。

6.2.2 $R \times C$ 表资料统计分析方法的选用 对于上述的 6 个 2 维列联表资料, 不少人下加思索地就选择 χ^2 检验。事实上, 应

根据 R C 表中两个分组变量的类型, 分别选用不同的统计分析方法

表 6.2.1 某地 6094 人按 2 种血型划分的结果

ABO 血型	人 数			
	MN 血型:	M	N	MN
O		431	490	902
A		388	410	800
B		495	587	950
AB		137	179	325
合计		1451	1666	2977

表 6.2.2 心律失常种类与心肌梗塞部位关系

缓慢心律失常种类	梗 塞 例 数				合计
	部位:	下壁	前壁	真后壁	心内膜下
窦性过缓		8	7	2	1
被动心律		1	1	0	0
房室阻滞		6	3	1	1
束支阻滞		1	16	1	0
合计		16	27	4	2

注: 被动心律的全称为被动性交界性心律。

表 6.2.3 3 种药物疗效的观察结果

疗 效	例 数				合计
	药物:	A	B	C	
治愈		15	4	1	20
显效		49	9	15	73
好转		31	50	45	126
无效		5	22	24	51
合计		100	85	85	270

表 6.2.4 眼晶状体混浊度与年龄之关系

晶状体混浊程度	眼 数				合计
	年龄(岁):	20~29	30~39	≥40	
+		215	131	148	494
++		67	101	128	296
+++		44	63	132	239
合计		326	295	408	1029

表 6.2.1 和表 6.2.2 虽然都是双向无序的, 表 6.2.1 中各网格内的频数都很大, 无小于 5 的理论频数, 可以选用 Pearson 的独立性 χ^2 检验, $\chi^2 = 8.598$, $P = 0.198$, 说明两种血型系统之间是互相独立的。

表 6.2.5 CT 标准与 Ranson 标准记分的关系

CT 分级	患 者 例 数				合计
	Ranson 记分:	0~2	3~5	≥6	
I		15	2	0	17
II		11	2	0	13
III		7	6	0	13
IV		1	9	6	16
合计		34	19	6	59

表 6.2.6 静脉造影与 PPG 诊断对深静脉瓣膜功能检测的符合情况

造影诊断结果	肢 体 数				合计
	PPG 诊断:	正常	可疑	逆流	
正常		8	0	9	17
轻逆		6	3	1	10
重逆		1	0	22	23
合计		15	3	32	50

而表 6.2.2 中小频数出现的次数较多, 若选用此方法, 就极易得出假阳性的检验结果, 故应改用 Fisher 的精确检验方法, $P = 0.000026$, 说明两组变量之间不独立, 即窦性过缓患者的心肌梗塞部位常发生在下壁和前壁; 束支阻滞患者的心肌梗塞部位常发生在前壁。

表 6.2.3 中药物分组是无序的, 而疗效分组则是有序的。通常的 χ^2 检验与变量是否有序无关, 故应选择与指标分组的顺序有关的检验方法, 即秩和检验, Ridit 分析或有序变量的 logistic 回归分析等。其中 Ridit 分析的结果为: A 药疗效最好, B、C 两种药疗效接近, 且很差。具体的统计分析结果如下:

药物名称	Ridit 均值	U 值	P 值	指标分组的顺序	疗效判定
A 药	0.3325	6.2047	<0.01	治愈→无效	优
B 药	0.5981	-3.3516	<0.01	治愈→无效	劣
C 药	0.5989	-3.3784	<0.01	治愈→无效	劣

表 6.2.4 中两个分组变量都是有序的, 但它们的属性不

同;表6.2.5中两个分组变量也是有序的,但它们的属性基本相同。这两张表的共同特点是两个分组变量的最初取值都是定量或半定量的数据。对这种资料,研究者往往希望考察两个有序变量之间是否存在相关关系,此时,应选用相关分析。现以分析表6.2.5资料为例,等级相关分析的结果为: $r_s = 0.67591, P = 0.0001$;典型相关分析的结果为:最大典型相关系数 $r = 0.73921, P < 0.0001$ 。二者都说明:CT分级与Ranson记分之间有非常显著性的正相关关系,即患者的CT分级取值较大,Ranson记分取值也较大。若研究者希望考察两个有序变量之间是否具有线性变化趋势,此时,应选用线趋势检验,结果从略。

表6.2.6中两个分组变量都是有序的,且属性相同,但定量程度较表6.2.4和表6.2.5差。它是前面配对设计的扩展,此时很难考察两种测定方法之间的差异,实际工作者更关心两种测定方法之间的吻合程度的高低,故宜选用一致性检验或称Kappa检验。其结果和结论如下:

$$P_1 = 0.660, P_e = 0.4084,$$

$$K = \text{Kappa} = 0.425287, S_k = 0.102248,$$

$$U = 4.159, P = 0.000032$$

说明两种方法诊断结果的吻合程度很高。

7 回归与相关分析^[1]

7.1 基本概念

7.1.1 回归分析的种类 如果因变量是连续变量,设自变量的个数为 k ,当 $k=1$ 时,回归分析的种类有:(1)直线回归分析;(2)通过直线化实现的简单曲线回归分析(以下简称为曲线拟合);(3)非线性曲线拟合;(4)一般多项式曲线拟合;(5)正交多项式曲线拟合。当 $k \geq 2$ 时,称为多元回归分析。在多元回归分析中,简单而又实用的则是多元线性回归分析。

如果因变量是与时间有关的连续变量且未被离散化(如:生存时间、复发时间、死亡时间等),而自变量可以是定量的,也可以是定性的。此时需用生存分析中的半参数或参数回归分析方法。

如果因变量是名义或有序变量,无论它取二个离散值(如:死与活、复发与未复发等)还是多个离散值(自变量可以是定性和定量的)时,都可选用logistic回归分析。

7.1.2 直线回归与相关分析的异同点 从研究的目的来说,若仅仅为了了解两变量之间呈直线关系的密切程度和方向,宜选用线性相关分析;若仅仅为了建立由自变量推算因变量的直线回归方程,宜选用线性回归分析。

7.1.3 散布图在这两种分析中的作用 功能齐全的计算机和统计软件,会蒙骗盲目运用统计方法的人,进行直线相关和回归分析时,尤其要注意!因为统计方法只能帮助人们揭示数据之间内在的统计规律性,但它不能创造规律,也就是说,资料之间是否存在本质联系要靠专业知识来解释;另外,在专业上有

一定联系的两项指标之间的关系并非都是直线关系,事实上,如果两项指标之间呈一条弯曲度不大的“S”型或倒“S”型曲线趋势,错误地用一条直线回归方程来描述,在统计学上往往会得到较高的显著性,即该直线回归方程是成立的,但在生物学上是解释不通的(当因变量是某种率时最易发生这种现象)。正确的做法是:将 (X, Y) 的 n 对数值绘在直角坐标系内,得到 X 与 Y 变化趋势的散布图,如果 n 个点形成的散布图呈一条明显的曲线趋势时,宜拟合一条曲线回归方程;如果 n 个点在一条不太宽的长带内随机地分布着,且不存在明显的曲线趋势,可考虑进行直线相关和回归分析;如果 n 个点形成的散布图近似于一个圆盘,则说明 X 与 Y 之间无确定的变化趋势,几乎是互相独立的,不必硬把它们捏合在一起分析。

7.2 两型回归问题 I型回归问题:当 X 是一般变量, Y 是服从正态分布的随机变量时,只能拟合 Y 随 X 变化的直线回归方程,不能反过来做。II型回归问题:当 X, Y 是服从二维正态分布的随机变量时,根据需要,可拟合 Y 随 X 变化或 X 随 Y 变化的直线回归方程。

7.3 具有重复试验数据的直线回归分析

7.3.1 回归分析中进行重复试验的目的 目的在于弄清:影响因变量 Y 的因素除 X 外,是否还有1个或几个不可忽略的其他因素;以及 X 与 Y 的关系是否确是直线关系。如果除 X 的影响外,还有其他未加控制的、不可忽视的影响因素掺杂,则此直线的拟合效果就不能算是好的,称为失拟。即在无重复试验情况下所建立的直线回归方程,即使假设检验的结果是“回归方程显著”,仅仅说明 X 的一次项对 Y 的影响是不可忽视的,并不能表明这个回归方程是拟合得很好的。

7.3.2 重复试验数据的回归分析中应注意的问题 不少人通常用各自变量 X_i 件下 K 次重复观测所得的因变量 Y_{ik} 值求出算术均值 \bar{Y}_i ,然后,用 n 对 (X_i, \bar{Y}_i) 值拟合直线回归方程。这样不仅达不到进行重复试验的目的,当 X 取不同值的个数较少时,往往得不到有意义的结果。应该按重复试验数据的回归分析方法来处理,具体计算见文献[1]。

关键词 试验设计;统计描述;统计分析
中国图书资料分类法分类号 (O213)

参考文献

- 1 胡良平主编.现代统计学与SAS应用.北京:军事医学科学出版社,1996,1~264
- 2 胡良平.医学科研与新药评价等工作中一个不可忽视的问题——轻视和滥用统计学.军事医学科学院院刊,1996;20(3):202~205
- 3 杨树勤主编.中国医学百科全书医学统计学.上海:上海科学技术出版社,1985:74,92~193
- 4 张尧庭.定性资料的统计分析.广西:广西师范大学出版社,1991,111~156

(下转第176页)