

· 继续教育园地 ·

骨科科研实验中配对设计资料统计分析方法及 SAS 实现

胡良平, 王琪

(军事医学科学院生物医学统计学咨询中心, 北京 100850)

关键词 配对分析; 实验设计; 统计学, 医学; 数据说明, 统计**DOI:**10.3969/j.issn.1003-0034.2010.03.030

Zhongguo Gushang/China J Orthop & Trauma, 2010, 23(3):235-238 www.zggszz.com

当实验中仅涉及 1 个实验因素, 一切非实验因素对该因素各水平组的影响是均衡的, 选用单因素设计是合适的^[1]。在前面的统计讲座中, 我们详细介绍了单因素设计中的单组设计及其用 SAS 实现数据分析的方法, 本文将重点介绍配对设计资料的统计分析方法及 SAS 实现。配对设计是属于单因素还是两因素的实验设计类型呢? 从下面的定义和特点中可以获得答案。

1 配对设计的定义和特点

与同一个定量指标对应的两组数据成对出现, 每对数据有 3 种可能的来源: 其一, 来自同一个体(接受处理前后或两个对称部位), 称为自身配对设计; 其二, 来自母体(如双胞胎或多胞胎中的两个)相同的两个个体分别接受 A、B 处理, 称为同源配对设计; 其三, 来自条件接近(如性别、年龄、体重、病情等相同或相近)的两个个体分别接受 A、B 处理, 称为条件相近者配对设计^[2]。

上述 3 种形式的配对设计本质上都涉及 2 个因素, 一个称为实验因素(如处理前后或 A 与 B 两种处理), 另一个称为区组因素(即配对条件)。若观测结果为定量的, 则进行统计分析所用的指标为每对数据的差量, 此时, 就把一个原本为两因

素(一个实验因素、一个区组因素)的配对设计定量资料转化为一个单组设计定量资料了。

上述 3 种配对设计的配对条件的严格程度一样吗? 不一样! 在时间间隔不长且处理因素的 2 个水平中有 1 个为空白的前提下, 自身配对设计最为严格, 其结果的可信度最高; 若 A、B 两种处理都是真实的处理且观测的定量指标受遗传因素影响较大, 显然, 自身配对设计很不适用, 此时, 同源配对设计明显优于条件相近者配对设计。条件相近者配对设计不可轻易使用, 仅当以对定量观测指标有影响的全部重要非实验因素为配对条件时, 其结果和结论才是基本正确的, 否则, 宜采用单因素 2 水平设计(或简称为成组设计)取代条件相近者配对设计。

在配对设计中, 若观测结果为二值的(如阳性、阴性), 就称为配对设计定性资料, 即配对设计 2×2 表资料。它是按照两种处理方法或两位检测者对每一个受试者或样品检测的结果进行配对, 然后清点(阳性、阳性)、(阳性、阴性)、(阴性、阳性)和(阴性、阴性)这 4 种结果出现的频数。假定括号中写在第 1 个位置上的结果为 A 法检测的结果、第 2 个位置上的结果为 B 法检测的结果, 参见本文中的表 3。

- [7] 黄志伟, 谢华, 崔晓军, 等. 补阳还五汤对大鼠脊髓损伤后移植间充质干细胞分化的作用. 解剖学研究, 2008, 30 (3): 203-205.
- [8] 董晓先, 刘金保, 董燕湘, 等. 天麻诱导骨髓间质干细胞分化为神经元细胞的实验研究. 中国中西医结合杂志, 2004, 24(1): 51-54.
- [9] 张进, 徐志伟. 补肾法诱导间充质干细胞向神经方向分化研究. 现代医院, 2004, 4(9): 15-17.
- [10] 王勇, 陆长青, 王凡. 黄芪诱导大鼠骨髓间充质干细胞分化为神经样细胞的研究. 四川解剖学杂志, 2006, 14(1): 5-8.
- [11] 陈东风, 杜少辉, 李伊为, 等. 龟板含药血清体外诱导成年大鼠骨髓间充质干细胞分化为神经元. 广州中医药大学学报, 2003, 20(3): 224-226.
- [12] 陆长青, 王勇, 陈登榜, 等. 丹参诱导骨髓基质干细胞向神经元样细胞分化及相关基因的表达. 四川解剖学杂志, 2006, 14(1): 1-4.
- [13] 董燕湘, 董晓先, 何慧华, 等. 大鼠骨髓间质干细胞用中药绞股蓝诱导为神经细胞的研究. 中华神经科杂志, 2003, 36(5): 355-358.
- [14] 撒亚莲, 李海标. 川芎嗪诱导大鼠骨髓间质干细胞分化为神经元样细胞的研究. 解剖学报, 2003, 34(5): 515-518.
- [15] 余勤, 罗依, 鄂艳, 等. 丹参素定向诱导骨髓间质干细胞分化为神经元样细胞的研究. 中国中西医结合杂志, 2005, 25(1): 49-54.
- [16] 蔡光先, 林琳, 刘柏炎, 等. 地黄多糖诱导骨髓间充质干细胞分化为神经样细胞的效应. 中国临床康复, 2005, 9(17): 17-19.
- [17] 项平, 李海标. 黄连素诱导大鼠骨髓间质干细胞分化为神经元样细胞. 中国病理生理杂志, 2004, 20(1): 51-53.
- [18] 贾延勃, 杨于嘉, 周燕, 等. 黄芩甙体外诱导大鼠骨髓基质细胞成为神经细胞. 中华医学杂志, 2002, 82(19): 1337-1342.
- [19] 郑国庆, 王小同, 陈伟, 等. 人参总皂苷体外诱导大鼠骨髓间充质干细胞分化为神经元样细胞. 中华中医药学刊, 2008, 26(6): 1257-1259.

(收稿日期: 2009-12-04 本文编辑: 连智华)

当某定性资料在本质上属于配对设计定性资料,但其行数和列数都大于 2,此时,就称为“方表”。统计学上称其为双向有序且属性相同的列联表,参见本文中的表 4。与此表对应的主要研究目的是希望回答行上与列上所对应的检测方法的检测结果是否一致。它实际上是配对设计 2×2 列联表资料的“扩大”形式。只不过在处理配对设计 2×2 列联表资料时,人们更关心的是两种检测方法检测的结果不一致部分的数量之间的差别是否具有统计学意义,而在处理“方表”资料时,人们更关心的是两种检测方法检测的结果之间是否具有一致性,故常用的统计分析方法称为一致性检验或称为 Kappa 检验。

2 配对设计的资料性质分类和统计分析方法的合理选用

配对设计按资料性质可以分为配对设计定量资料和配对设计定性资料,配对设计定性资料又可分为配对设计 2×2 表资料和配对设计扩大形式的方表资料;当观测指标为定量指标时,又可分为配对设计一元定量资料和配对设计多元定量资料。常见分类与方法选择见表 1。

表 1 常用配对设计类型及统计分析方法

设计类型	可选用的统计分析方法
配对设计一元定量资料	配对设计 <i>t</i> 检验,或符号检验,或符号秩和检验
配对设计多元定量资料	多元方差分析
配对设计定性资料	McNemar χ^2 检验,对称性检验,或 Kappa 检验

3 配对设计应用举例

【例 1】采用 10 只正常成年大耳白兔下肢骨标本,共 10 对胫骨,行三点弯力学测试。从实验兔胫骨三点弯载荷挠度曲线得最大挠度,数据见表 2,试进行统计分析^[3]。

表 2 10 只家兔双侧胫骨最大挠度(mm)

序号	最大挠度(mm)	
	左侧	右侧
1	2.91	2.73
2	2.56	2.41
3	2.75	2.65
4	2.66	2.90
5	2.65	2.90
6	2.58	2.49
7	2.60	2.88
8	3.04	2.79
9	2.46	2.39
10	3.19	3.24

【分析与解答】该定量资料属于自身配对设计一元定量资料,每对中的两个定量数据的差值符合正态分布(具体正态性检验结果见 SAS 输出结果),故采用配对设计一元定量资料的 *t* 检验。

配对设计定量资料 *t* 检验与单组设计定量资料 *t* 检验不同之处在于:配对设计一元定量资料的 *t* 检验是检验差值所代表的总体均数与 0 之间的差别是否具有统计学意义;而单组设计定量资料的 *t* 检验是检验样本所代表的总体均数与理

```

DATA a;
  INPUT x1 x2;
  d=x2-x1;
CARDS;
2.91 2.73
2.56 2.41
2.75 2.65
2.66 2.90
2.65 2.90
2.58 2.49
2.60 2.88
3.04 2.79
2.46 2.39
3.19 3.24
;
RUN;
PROC UNIVARIATE NORMAL;
  VAR d;
RUN;
    
```

论值或标准值之间的差别是否具有统计学意义。

下面用 SAS 软件编程法进行分析:

【SAS 程序说明】d 为 x_2 与 x_1 的差值,读者处理自己的配对设计一元定量资料时只需更换本程序中的样例数据,每一对数据书写的先后顺序要一致。

【主要输出结果】如下所示:

UNIVARIATE 过程

变量: d

矩			
N	10	权重总和	10
均值	-0.002	观测总和	-0.02
标准偏差	0.194 753 41	方差	0.037 928 89
偏度	0.501 463 41	峰度	-1.421 263 7
未校平方和	0.341 4	校正平方和	0.341 36
变异系数	-9 737.670 3	标准误差均值	0.061 586 43

位置检验: Mu0=0

检验	统计量	P 值
学生 t	<i>t</i> -0.032 47	Pr > t 0.974 8
符号	<i>M</i> -1	Pr >= M 0.753 9
符号秩	<i>S</i> -1	Pr >= S 0.945 3

正态性检验

检验	统计量	P 值
Shapiro-Wilk	<i>W</i> 0.880 828	Pr < <i>W</i> 0.133 4

【统计和专业结论】首先查验正态性检验的结果。d 变量正态性检验的结果: $W=0.880 828, P=0.133 4 > 0.05$, 可看出差值服从正态分布,故选用 *t* 检验的结果, $t=-0.032 47, P=0.974 8$,故按 $\alpha=0.05$ 水准,认为平均来说实验兔胫骨左侧和右侧最大挠度之间的差别无统计学意义,可认为实验兔胫骨左侧和右侧最大挠度近似相等。

【例 2】161 名受试者在静态站立睁眼状态时用目前颈性眩晕的临床诊断方法(CDM 法)和重心测试诊断颈性眩晕的方法(SPG 法)分别进行诊断,结果见表 3。问诊断结果是否一

表 3 睁眼时 SPG 法与 CDM 法对两组的诊断结果比较

SPG 法	例数		
	CDM 法: 异常	正常	合计
异常	37	11	48
正常	38	75	113
合计	75	86	161

致^[4]。

【分析与解答】本例目的是评价两种方法的诊断结果的一致性,从设计上来看该资料属于配对设计 2x2 表定性资料。比较两种方法诊断结果不一致部分的差别是否有统计学意义,常用 McNemar χ^2 检验;比较两种方法诊断结果一致性部分(即 Kappa 值)与 0 之间的差别是否有统计学意义用 Kappa 检验。若两种诊断方法都存在假阳性和假阴性结果,则不值得分析;现假定 CDM 法为金标准,目的是考察可否用 SPG 法取代 CDM 法。

下面用 SAS 软件编程法进行分析:

DATA b;	PROC FREQ;
DO A=1 TO 2;	WEIGHT F;
DO B=1 TO 2;	TABLES A*B/AGREE;
INPUT F @@;	TEST KAPPA;
OUTPUT;	RUN;
END;	
END;	
CARDS;	
37 11	
38 75	
;	
RUN;	

【SAS 程序说明】数据步,建立名为 b 的数据集,建立数值变量 A、B、F,分别读入行号、列号、每格实际频数。过程步,调用 FREQ 过程,指定频数变量 F,用 TABLES A*B 语句表示二维列联表;AGREE 语句是配对四格表输出 McNemar χ^2 统计量的关键词句;TEST KAPPA 用于对 Kappa 值进行假设检验。

【主要输出结果】如下所示:

McNemar 检验	
统计量(S)	14.877 6
自由度	1
Pr > S	0.000 1

以上是 McNemar χ^2 检验计算的结果,此处使用的是未进行校正的计算公式。 $\chi^2=14.877 6, P=0.000 1<0.05$,说明两种方法的诊断结果不一致部分的差异有统计学意义。因 $38>11$,意味着 SPG 法的假阴性率高于其假阳性率。

简单 Kappa 系数	
Kappa	0.374 0
渐近标准误差	0.069 6
95% 置信下限	0.237 7
95% 置信上限	0.510 4

以上给出了 Kappa 值、标准误和 95%置信区间的上限和

下限。

H0 检验: Kappa = 0	
H0 下的渐近标准误差	0.074 0
Z	5.056 3
单侧 Pr > Z	<.000 1
双侧 Pr > Z	<.000 1

以上部分是对原假设“Kappa=0”的假设检验结果, $Z=5.056 3, P<0.000 1$,说明两种方法的诊断结果一致性部分与 0 之间的差异有统计学意义,也即一致性有统计学意义(Kappa 越接近于 1,说明一致性越好)。

【统计和专业结论】经 McNemar χ^2 检验,即差异性检验可知,两种方法的诊断结果不一致部分的差异有统计学意义。具体地说,假定 CDM 法为金标准,则 SPG 法的假阴性率高于其假阳性率;经 Kappa 检验,即一致性检验可知,两种方法的诊断结果一致性部分与 0 之间的差异有统计学意义,说明两种方法的诊断结果一致性较高。结合 Kappa 值的 95%置信区间(0.237 7, 0.510 4),可以认为 Kappa 值尚不够大,说明两种方法的一致性仅在统计学上有一定意义,但实际应用价值并不高(可根据专业知识对预期的一致率进行界定,当高于该值,如 85%或 90%以上,此时可以认为一致性具有专业上的意义)。最后的结论是:该两种方法诊断结果在统计学上具有较好的一致性,但一致率并不够高(本例实际一致率仅为 69.57%),实际意义不太大。在颈性眩晕的临床诊断中,若以 CDM 法为金标准,则尚不能用 SPG 法取代 CDM 法。

表 4 年资不同的两组医生对一批 X 线片判读的结果

低年资医生组	X 线片张数					合计	
	高年资医生组:	I	II	III	IV		V
I		23	11	3	2	0	39
II		10	32	3	3	0	48
III		1	29	10	7	3	50
IV		4	7	5	22	19	57
V		2	2	3	16	86	109
合计		40	81	24	50	108	303

【例 3】两组医生对一批 X 线片判读的结果如表 4 所示,试评价两组医生评判的结果是否一致。

【分析与解答】本例属于“双向有序且属性相同的方表”,由于各级评分存在较大的主观因素,因此可用加权 Kappa 检验进行分析。

下面用 SAS 软件编程法进行分析:

【SAS 程序说明】该程序关键词句是 TEST WTKAP,目的是对加权 Kappa 值进行假设检验。

【主要输出结果】如下所示:

对称性检验	
统计量(S)	29.029 8
自由度	10
Pr > S	0.001 2

以上是对称性检验结果, $S=29.029 8, P=0.001 2<0.05$,说明此表的频数分布不满足对称性假设,即此表中主对角线上方各格内的频数与其下方对应的各格内的频数是不对称的。

```

DATA c;
  DO A=1 TO 5;
    DO B=1 TO 5;
      INPUT F @@;
      OUTPUT;
    END;
  END;
CARDS;
23 11 3 2 0
10 32 3 3 0
1 29 10 7 3
4 7 5 22 19
2 2 3 16 86
;
RUN;
PROC FREQ;
  WEIGHT F;
  TABLES A*B;
  TEST WTKAP;
RUN;

```

简单 Kappa 系数

Kappa	0.441 6
渐近标准误差	0.034 6
95% 置信下限	0.373 8
95% 置信上限	0.509 4

加权的 Kappa 系数

加权的 Kappa	0.654 2
渐近标准误差	0.028 2
95% 置信下限	0.598 8
95% 置信上限	0.709 6

以上给出了简单 Kappa 和加权 Kappa 统计量和 95% 置信区间。

H0 检验: 加权的 Kappa = 0

H0 下的渐近标准误差	0.041 7
Z	15.673 4
单侧 Pr> Z	<.000 1
双侧 Pr> Z	<.000 1

以上部分是对加权 Kappa 作假设检验的结果。

【统计和专业结论】经加权 Kappa 检验可知, P<0.05, 两组

医生对 X 线片的判读结果一致性部分与 0 之间的差别有统计学意义, 说明两组医师对 X 线片的判读结果具有一致性。计算一下观测到的一致率为 57.10%, 说明两组医生判断的结果仅在统计学上具有一致性, 而实际的一致性尚不够满意, 从表 4 的主对角上方与下方对应的各格内频数可看出, 年资低的医生倾向于将患者判为较重的病情(如: 有 29 位患者, 高年资医生判为 II 型, 而低年资医生将其判为 III 型)。

4 配对设计需要注意的问题

配对条件必须是拟做的实验研究中对观测结果有重要影响的重要非实验因素的综合体现, 例如, 在某项实验研究中, 假定性别、年龄、体重和病情对观测结果都有一定的影响作用, 那么, 配对条件就应该在这 4 个重要非实验因素的复合结果, 即配对的两个个体必须性别相同、年龄相同或非常接近、体重相同或非常接近、病情轻重相当。配对的条件越严格, 对非实验因素的综合能力越强, 则配对设计的质量就越高, 结果的可信度就越高。

因篇幅所限, 本文未介绍配对设计定量资料的多元方差分析方法, 也未介绍与配对设计定量资料和定性资料的参数估计方法。需要时, 可参阅文献^[5-6]。

参考文献

- [1] 胡良平. 口腔医学科研设计与统计分析. 北京: 人民军医出版社, 2007. 125.
- [2] 胡良平. 统计学三型理论在实验设计中的应用. 北京: 人民军医出版社, 2006. 48-52.
- [3] 钟红刚, 刘卫华, 卜海滨, 等. 家兔胫骨的三点弯测试. 中国骨伤, 2008, 21(2): 103-105.
- [4] 黄阿勇, 周卫, 钟红刚, 等. 人体重心测试系统诊断颈性眩晕的评价. 中国骨伤, 2008, 21(5): 345-348.
- [5] 胡良平. 医学统计学——运用三型理论分析定量与定性资料. 北京: 人民军医出版社, 2009. 55-57.
- [6] Ronald N. Forthofer. Eun Sul Lee. Mike hernandez. Biostatistics: a guide to design, analysis, and discovery. Elsevier Academic Press, 2007. 169-212.

(收稿日期: 2010-01-05 本文编辑: 连智华)

· 读者 · 作者 · 编者 ·

本刊关于一稿两投和一稿两用等现象的处理声明

文稿的一稿两投、一稿两用、抄袭、假署名、弄虚作假等现象属于科技领域的不正之风, 我刊历来对此加以谴责和制止。为防止类似现象的发生, 我刊一直严把投稿时的审核关, 要求每篇文章必须经作者单位主管学术的机构审核, 附单位推荐信(并注明资料属实、无一稿两投等事项)。希望引起广大作者的重视。为维护我刊的声誉和广大读者的利益, 凡核实属于一稿两投和一稿两用等现象者, 我刊将择期在杂志上提出批评, 刊出其作者姓名和单位, 并对该文的第一作者所撰写的一切文稿 2 年内拒绝在本刊发表, 同时通知相关杂志。欢迎广大读者监督。

《中国骨伤》杂志社