

配对相关数据的统计学方法

献给方开泰教授 80 华诞

沈溪, 马长兴*

Department of Biostatistics, The State University of New York at Buffalo, Buffalo, NY 14214, USA

E-mail: xishen@buffalo.edu, cxma@buffalo.edu

收稿日期: 2019-04-24; 接受日期: 2019-08-27; 网络出版日期: 2020-04-28; * 通信作者

摘要 配对相关数据经常在医学研究中应用, 例如, 在眼科或耳鼻喉科研究中, 对配对器官中每个样本信息进行数据分析. 配对器官的测量数据通常具有高度相关性, 而大多数统计推断方法假定样本的观测值是相互独立的. 研究表明, 忽略配对数据的组内相关性会导致显著性水平的增加. 有很多统计方法来解配对数据的组内相关性. 此外, 忽略配对数据的相关性或混杂效应可导致结果发生偏差, 因此, 调整和控制统计推断中的混杂效应至关重要. 本文回顾讨论这些统计方法, 并提出统计推断方法的建议.

关键词 配对相关数据 组内相关系数 区间估计 分层配对数据 统计检验

MSC (2010) 主题分类 62F03, 62F05, 62F12

1 引言

在眼科或耳鼻喉科等临床研究中, 当研究人员收集配对器官 (或身体部位) 数据时, 会得到配对相关数据, 这种配对观测值通常具有高度相关性, 因此, 需要合适的方法对组内相关性进行数据分析. 若不考虑配对数据的组内相关性, 则可能会得出不可靠的结论. 基于这个原因, 在过去几十年中, 解决配对相关数据的统计模型成为一个热点问题, 其中包括新的分析方法和高效的计算算法. 本文主要对配对相关数据的统计方法进行综述, 重点介绍新方法.

1.1 配对相关数据的结构

假定对 g ($g \geq 2$) 个不同治疗方法的对比感兴趣, 其所有可能的治疗结果放在一个列联表中 (治疗结果为配对器官中两个全部治愈、只有一个治愈、两个都未治愈 3 种情况). 设患者总数为 N , 被随机地分配到 g 个治疗组中. 在表 1 中, 根据治疗结果, m_{li} 表示第 i ($i = 1, 2, \dots, g$) 组配对器官中 l ($l = 0, 1, 2$) 个器官被治愈的患者数. 对任意 $i = 1, 2, \dots, g$ 和 $l = 0, 1, 2$, 记 $m_{+i} = \sum_{l=0}^2 m_{li}$, $m_{l+} = \sum_{i=1}^g m_{li}$.

英文引用格式: Shen X, Ma C X. Statistical methods for the bilateral correlated data (in Chinese). Sci Sin Math, 2020, 50: 667–678, doi: 10.1360/N012019-00109

表 1 配对相关数据结构

眼睛治愈数 (l)	治疗方法组 (i)					总计
	1	2	3	...	g	
0	m_{01}	m_{02}	m_{03}	...	m_{0g}	$m_{0+}(S_0)$
1	m_{11}	m_{12}	m_{13}	...	m_{1g}	$m_{1+}(S_1)$
2	m_{21}	m_{22}	m_{23}	...	m_{2g}	$m_{2+}(S_2)$
总计	m_{+1}	m_{+2}	m_{+3}	...	m_{+g}	N

不失一般性, 取配对器官为两只眼睛. 若第 i ($i = 1, 2, \dots, g$) 组中第 j ($j = 1, 2, \dots, m_{+i}$) 个患者的第 k ($k = 1, 2$) 只眼睛治愈, 则 $Z_{ijk} = 1$, 否则为 0. 设 S_l ($l = 0, 1, 2$) 为 l 只眼睛治愈的患者总数, 第 i 组患者第 k 只眼睛的治愈概率为 $\Pr(Z_{ijk} = 1) = \pi_i$ ($0 \leq \pi_i \leq 1$).

1.2 统计模型

在上述框架下, 需要检验原假设 H_0 和备择假设 H_1 :

$$H_0: \pi_1 = \dots = \pi_g, \quad H_1: \text{至少一对 } \pi_r \neq \pi_s,$$

其中 $r, s \in \{1, 2, \dots, g\}$. 近年来, 一些特有的检验方法在配对数据分析中起着重要的作用, 如广义线性模型法、Rosner 方法、Dallal 方法和 Donner 方法.

1.2.1 广义线性模型法

对于配对相关数据研究, 我们主要通过广义线性模型法计算最大似然估计 (maximum likelihood estimation, MLE) 或进行统计分析. 对每位患者的每只眼睛进行观测, 可以用文献 [1] 中的广义线性模型去检验 g 组中的发病 (或治愈) 率是否相同. Liang 和 Zeger [2] 推广了该模型, 将每位患者的眼睛观测值视为纵向数据. 此外, 有很多统计软件 (如 SAS (statistical analysis system) 中的 GENMOD 和 GLIMMIX) 可以用来处理广义线性模型 (参见文献 [3]). 然而, Ma 等 [4] 指出广义线性模型法不能为 MLE 及统计检验提供具体表达式并且计算量较大, 而概率模型的统计推断可以提高计算效率, 特别适用于对小样本精确检验. 为了得到精确检验, 需要大量计算, 若使用广义线性模型法很难解决这个问题.

1.2.2 Rosner 方法

Rosner [5] 认为如果眼睛之间存在相关性, 将每只眼睛当作一个独立随机变量是不合理的. 因此, 假定一只眼睛治愈的条件下另一只眼睛治愈的概率与眼睛数据对应组的治愈率成正比的条件下, 提出了一个基于相关性的“常数 R 模型”, 即存在某个常数 R , 有

$$\Pr(Z_{ijk} = 1 \mid Z_{ij(3-k)} = 1) = R\pi_i, \quad i = 1, 2, \dots, g, \quad j = 1, 2, \dots, m_{+i}, \quad k = 1, 2,$$

其中常数 R 表示患者两只眼睛之间的相关关系, 若 $R = 1$, 则两只眼睛之间是独立的; 若 $R\pi_i = 1$, 则两只眼睛是完全相关的. 证明可知, 两只眼睛观测值之间的相关性为

$$\rho_i = \text{Corr}(Z_{ij1}, Z_{ij2}) = \frac{\pi_i}{1 - \pi_i}(R - 1), \quad i = 1, 2, \dots, g.$$

设 p_{li} ($l = 0, 1, 2; i = 1, 2, \dots, g$) 是第 i 组中恰有 l 只眼睛治愈的概率. 固定 i , 对 π_i 和 R , 有

$$\begin{cases} p_{0i} = R\pi_i^2 - 2\pi_i + 1, \\ p_{1i} = 2\pi_i(1 - R\pi_i), \\ p_{2i} = R\pi_i^2, \end{cases} \quad (1.1)$$

且 $p_{0i} + p_{1i} + p_{2i} = 1$. 因此, 对任意第 i 组, 随机向量 (m_{0i}, m_{1i}, m_{2i}) 服从三项分布, 样本容量为 m_i , 对应的概率分布为 (p_{0i}, p_{1i}, p_{2i}) . 令 $\mathbf{m}_i = (m_{0i}, m_{1i}, m_{2i})^T$, $\mathbf{p}_i = (p_{0i}, p_{1i}, p_{2i})^T$, 则联合概率分布函数为

$$f(\mathbf{m}_i | \mathbf{p}_i, m_{+i}) = \frac{m_{+i}!}{m_{0i}!m_{1i}!m_{2i}!} p_{0i}^{m_{0i}} p_{1i}^{m_{1i}} p_{2i}^{m_{2i}},$$

其中概率 p_{0i} 、 p_{1i} 和 p_{2i} 满足 (1.1).

Dallal^[6] 指出了 Rosner 模型的不足: 常数 R 模型的假设建立在给定一只眼睛治愈的条件下另一只眼睛也能治愈的条件概率与 π 成正比, 除了 π_i 几乎相等外, 所有第 j ($i \neq j$) 组的 R 都不能接近 1. 因此, 如果患者的眼睛都治愈, 并且不同组具有不同的治愈率, 该模型是不合适的.

1.2.3 Dallal 方法

Dallal^[6] 假定在一只眼睛治愈的条件下, 另一只眼睛也治愈的条件概率为常数, 与 π_i 无关, 即存在某个常数 γ , 使得

$$\Pr(Z_{ijk} = 1 | Z_{ij(3-k)} = 1) = \gamma, \quad i = 1, 2, \dots, g, \quad j = 1, 2, \dots, m_{+i}, \quad k = 1, 2.$$

在这种情况下, 两只眼睛观测值之间的相关系数为

$$\rho_i = \text{Corr}(Z_{ij1}, Z_{ij2}) = \frac{\gamma - \pi_i}{1 - \pi_i}, \quad i = 1, 2, \dots, g,$$

设 p_{li} ($l = 0, 1, 2; i = 1, 2, \dots, g$) 是第 i 组中恰有 l 只眼睛治愈的概率, 则

$$\begin{cases} p_{0i} = (\gamma - 2)\pi_i + 1, \\ p_{1i} = 2(1 - \gamma)\pi_i, \\ p_{2i} = \gamma\pi_i, \end{cases} \quad (1.2)$$

其中 $p_{0i} + p_{1i} + p_{2i} = 1$. 类似地, 随机向量 (m_{0i}, m_{1i}, m_{2i}) 服从样本容量为 m_i 和概率 (p_{0i}, p_{1i}, p_{2i}) 满足 (1.2) 的三项分布.

1.2.4 Donner 方法

Donner^[7] 假定所有治疗组共用一个组内相关系数 ρ , 并建立了一个替代模型— ρ 模型, Thompson^[8] 通过拟合验证了 ρ 模型对于配对数据具有稳健性.

假设 $\text{Corr}(Z_{ijk}, Z_{ij(3-k)}) = \rho$ ($0 \leq \rho \leq 1$) 是组内相关系数, 有

$$\begin{cases} p_{0i} = \rho(1 - \pi) + (1 - \rho)(1 - \pi_i)^2, \\ p_{1i} = 2\pi_i(1 - \rho)(1 - \pi_i), \\ p_{2i} = \rho\pi_i + (1 - \rho)\pi_i^2, \end{cases} \quad (1.3)$$

其中 $p_{0i} + p_{1i} + p_{2i} = 1$ ($i = 1, 2, \dots, g$). 因此, 随机向量 (m_{0i}, m_{1i}, m_{2i}) 服从样本容量为 m_i 和概率 (p_{0i}, p_{1i}, p_{2i}) 满足 (1.3) 的三项分布.

1.3 分层配对相关数据

在实践中, 治疗中心效应或混杂效应导致不同治疗组间的失衡, 在数据分析中需要调整分层或混杂因子. 在许多随机对照试验 (randomized controlled trial, RCT) 中, 研究者招募患者到多个治疗中心或医院, 而不是单一的治疗中心. 因此, 多治疗中心现象普遍存在, 对统计分析提出了挑战, 原因在于同一治疗中心的患者的结果具有相关性. 为了获得有效的统计推断, 研究表明在数据分析中应调整分层因子 (或中心效应), 参见文献 [9, 10].

此外, 忽略混杂因素的影响可能会得到具有误导性的结果. 在一些有大量混杂因素的随机对照试验中, 有些混杂因素可能在治疗组之间不平衡, 因此, 需要调整这些不平衡的混杂因子, 获得有效的推断. 适当地调整分析关键在于识别何时出现不可忽略的混杂因素. 基于此, 进行分层数据分析是有必要的.

1.4 分层配对数据结构

分层数据分析已经引起广泛关注, 并在临床试验中产生了很多统计方法. 例如, Nam^[11] 用两个得分统计量去检验分层数据的相对风险, Lui^[12] 研究稀疏数据风险差异的同质性检验, Zou 和 Donner^[13] 使用回归模型研究分层可变的数据.

令 m_{ij} 是第 j ($j = 1, 2, \dots, J$) 层的第 i ($i = 1, 2$) 个治疗组中的 l ($l = 0, 1, 2$) 只眼睛治愈的患者数, $m_{+ij} = \sum_{l=0}^2 m_{lij}$ 是第 j 层第 i 组中患者总数, 假定它是固定的. 相应地, 令 p_{lij} 是第 j 层第 i 组中有 l 只眼睛治愈的概率, Z_{hijk} 为第 h ($h = 1, 2, \dots, m_{+ij}$) 个患者的 k ($k = 1, 2$) 只眼睛的响应变量, 若有治愈, 则 $Z_{hijk} = 1$, 否则为 0. 表 2 给出第 j 层的配对数据结构, 每一列都服从多项分布.

假设来自同一层的两个治疗组中所有患者的两只眼睛治愈的相关性是相同的, 不同层的相关性不同. 令 $\Pr(Z_{hijk} = 1) = \pi_{ij}$ ($0 \leq \pi_{ij} \leq 1$) 是第 j 层第 i 组中第 h 个患者的第 k 只眼睛被治愈的概率, 那么上述公式和结论均可用于分层情形.

1.5 关联度量

目前, 主要有 3 种方法来描述和量化给定人群中两个比例之间关联的强度: 风险差异、风险比率和比值比 (odds ratio, OR) (参见文献 [14]). 比较风险差异是最直观的关联度量之一, 用来比较两个治疗组之间重要特征的相对风险比例, 比较易于解释. 在某些情形下, 两个风险的比例也能提供一些有价值的信息, 例如, 在比较两种治疗失败率的研究中, 0.010 和 0.001 之间的差异可能比 0.410 和 0.401 之间的差异更值得注意 (参见文献 [15]).

表 2 分层配对相关数据结构的第 j 层

眼睛治愈数 (l)	治疗组 (i)		总计
	1	2	
0	$m_{01j}(p_{01j})$	$m_{02j}(p_{02j})$	m_{0+j}
1	$m_{11j}(p_{11j})$	$m_{12j}(p_{12j})$	m_{1+j}
2	$m_{21j}(p_{21j})$	$m_{22j}(p_{22j})$	m_{2+j}
总计	m_{+1j}	m_{+2j}	m_{++j}

此外, OR 是在前瞻性、回顾性或横断面抽样设计中首选的关联度量, 而且它是多维列联表中数据分析的基本方法 (参见文献 [16]). 在分层配对数据中已经得到了风险方差和风险比率的各种方法, 但关于 OR 的渐近检验尚未研究.

本文目的主要有两个方面: 首先对上述模型分析数据的统计方法进行综合评价; 其次为医学研究人员提供建议, 针对配对相关数据, 重点说明 MLE 算法和不同统计推断及应用中的最佳方法. 这些统计方法有以下两种情形: (1) 无分层/混杂效应的配对相关数据; (2) 有分层效应的配对相关数据. 在下面, 本文将讨论几种可能的研究方向.

2 无分层效应的配对相关数据分析

2.1 风险比例的同质性检验

2.1.1 Rosner 方法

研究表明, Rosner 模型在某些情形下能得出较好的结论. 对小样本设计及稀疏数据结构, 文献 [17–19] 考虑了精确统计推断. 大样本统计推断方法在文献 [4, 20] 中已讨论.

设患者的观测数据 $\tilde{M} = (m_{01}, \dots, m_{0g}, m_{11}, \dots, m_{1g}, m_{21}, \dots, m_{2g})$, 由 (1.1), 得到其对数似然函数为

$$l(\pi_1, \dots, \pi_g; R) = \sum_{i=1}^g [m_{0i} \ln(R\pi_i^2 - 2\pi_i + 1) + m_{1i} \ln(2\pi_i(1 - R\pi_i)) + m_{2i} \ln(R\pi_i^2)],$$

在原假设 H_0 下, 计算出 π_i 和 R 的最大似然估计 (MLE) 为

$$\begin{cases} \hat{\pi}_{H_0} = \frac{S_1 + 2S_2}{2N}, \\ \hat{R}_{H_0} = \frac{4NS_2}{S_1 + 2S_2^2}. \end{cases} \quad (2.1)$$

此外, 可以用迭代方法求无约束 MLE.

Tang 等^[20] 提出 8 种方法, 检验了上述 MLE 下的 Rosner 模型中两个治疗组之间的治愈率是否相等, 即检验假设: $H_0: \pi_1 = \pi_2$ 与 $H_1: \pi_1 \neq \pi_2$. 结果表明, 得分统计量的效果较好, 因此建议使用.

Ma 等^[4] 将上述结果推广到 $g (\geq 2)$ 组的情形, 构造了 3 个大样本渐近统计量检验 Rosner 模型下多个治愈率的同质性, 即 $H_0: \pi_1 = \dots = \pi_g$ 与 $H_1: \text{一些 } \pi_i \text{ 不相等}$. 当参数很多时 (例如, g 很大), Fisher 评分算法收敛速度慢, 他们通过两步法估计得到无约束 MLE 的最大值. 首先化简 $\frac{\partial l}{\partial \pi_i} = 0$ ($i = 1, 2, \dots, g$), 得到一个简化的 3 次方程

$$\pi_i^3 - \frac{4m_{01} + 5m_{1i} + 6m_{2i}}{2Rm_i} \pi_i^2 + \frac{m_{0i} + (1+R)m_{1i} + (2+R)m_{2i}}{R^2m_i} \pi_i - \frac{m_{1i} + 2m_{2i}}{2R^2m_i} = 0,$$

通过上述等式的实根对 π_i ($i = 1, 2, \dots, g$) 进行 $t+1$ 次更新得到 MLE, R 通过 Fisher 评分算法得到.

此外, 他们提出了 $g (\geq 2)$ 组情形下的 3 个假设检验 (似然比检验、Wald 检验和得分检验). 结果表明, 不管组数多少、样本大小或参数配置, 得分检验具有高功率和令人满意的 I 类错误控制, 其检验统计量可简化为

$$T_{SC}^2 = \sum_{k=1}^g \frac{N(S_1^2 m_{0k} - S_0 S_1 (m_{1k} + 2m_{2k}) + 2S_0 S_2 m_{1k})^2}{S_0 S_1 (S_1^3 + S_0 S_1^2 + 4S_0 S_2^2) m_k}, \quad (2.2)$$

且 T_{SC}^2 服从自由度为 $g-1$ 的渐近 χ^2 分布.

2.1.2 Donner 方法

Donner 和 Klar^[21] 提出了 χ^2 检验统计量, 其有效性已在文献 [22, 23] 中得到验证. 在 Donner 模型下, 文献 [24, 25] 进一步研究了这些渐近检验方法.

基于观测数据 $\widetilde{M} = (m_{01}, \dots, m_{0g}, m_{11}, \dots, m_{1g}, m_{21}, \dots, m_{2g})$, 根据 (1.3), 得到对数似然函数为

$$l(\pi_1, \dots, \pi_g; \rho) = \sum_{i=1}^g [\ln(\pi_i^2 + \rho\pi_i(1 - \pi_i))m_{2i} + \ln(2\pi_i(1 - \rho)(1 - \pi_i))m_{1i} + \ln((1 - \pi_i)(\rho\pi_i - \pi_i + 1))m_{0i}].$$

针对同时存在单侧和双侧的联合数据, Pei 等^[24] 用 10 个统计量检验了两种治疗方法之间治愈率的同质性. 这些检验统计量要么依赖于独立模型, 要么依赖于单边模型和双边数据. 他们要检验假设 $H_0: \pi_1 = \pi_2$ 与 $H_1: \pi_1 \neq \pi_2$. 在独立模型下 (即 $\rho = 0$), 可以直接计算 MLE; 在相关模型下, 通过迭代方法得到 MLE.

更一般地, Ma 和 Liu^[25] 研究了 3 种渐近检验方法检验 $g (\geq 2)$ 组成治愈率的同质性, 即 $H_0: \pi_1 = \dots = \pi_g$ 与 $H_1: \text{一些 } \pi_i \text{ 不相等}$. 计算得到 H_0 成立下的最大似然估计为

$$\begin{cases} \widehat{\pi}_{H_0} = \frac{S_1 + 2S_2}{2N}, \\ \widehat{\rho}_{H_0} = \frac{4S_0S_2 - S_1^2}{(2S_0 + S_1)(S_1 + 2S_2)}. \end{cases} \quad (2.3)$$

在 H_1 成立下, MLE 没有显式解. 首先简化 $\frac{\partial l}{\partial \pi_i} = 0$ ($i = 1, 2, \dots, g$), 得到一个 3 次方程

$$\begin{aligned} & (4\rho - 2\rho^2 - 2)m_i\pi_i^3 + [3\rho^2m_i - \rho(5m_{0i} + 6m_{1i} + 7m_{2i}) + 2m_{0i} + 3m_{1i} + 4m_{2i}]\pi_i^2 \\ & + [(4\rho - \rho^2)m_i - 2\rho m_{0i} - m_{1i} - 2m_{2i}]\pi_i - \rho(m_{1i} + m_{2i}) \\ & = 0, \end{aligned}$$

然后通过 Fisher 评分算法更新 ρ 的估计值.

此外, 他们研究并推导出 3 个多种治疗中成功治愈率相等的统计检验. 结果显示, 得分统计量检验具有高功率和令人满意的 I 类错误误差控制, 因此建议使用. 通过推导, 得分统计量可简化为

$$T_{SC}^2 = \sum_{k=1}^g \frac{N[4S_0S_2(S_2 - S_0)m_{1k} + (S_2m_{0k} - S_0m_{2k})(S_1^2 + 4S_2S_0) + 4S_0S_1S_2(m_{0k} - m_{2k})]^2}{S_0S_2[(S_0 + S_2)(4S_0S_2 - S_1^2)^2 + 16S_0S_1S_2N^2]m_k},$$

且 T_{SC}^2 服从自由度为 $g - 1$ 的 χ^2 分布.

2.2 区间估计

2.2.1 治愈率差别的区间估计

Pei 等^[26] 构造了两组之间治愈率差异的渐近置信区间. Tang 等^[27] 研究 R 模型下治愈率差异的渐近 Wald 置信区间和似然比置信区间. 然而, Wald 统计量和相应的方差估计量在治愈率差异为零的假设下导出, 似然比置信区间在独立假设 (即 $R = 1$) 下得到. Yang 等¹⁾ 用两种替代方法改进了 Tang

1) Yang Z Y, Liu S, Liu X, et al. Confidence intervals for difference of proportions from correlated bilateral data in ophthalmologic studies. Unpublished manuscript, 2019

等^[27]的方法, 包括大样本设计中治愈率差异的渐近 Wald 和似然比置信区间. 文献 [28, 29] 在 Monte Carlo 模拟研究中, 与独立假设下的方法进行了比较.

下面介绍方差估计修正方法 (method of variance estimates recovery, MOVER). 设 θ_l ($l = 1, 2$) 为总体的比例, $\hat{\theta}_l$ 为估计值, 考虑 $\theta_1 - \theta_2$ 的置信区间 (confidence interval, CI) $CI(L, U)$. 设 (l_i, u_i) ($i = 1, 2$) 分别为 θ_i 的双边 $100(1 - \alpha)\%$ 置信区间. 为了得到 L , 在 $\theta_1 = l_1$ 和 $\theta_2 = u_2$ 条件下估计 $\text{var}(\hat{\theta}_1)$. 运用反演原理, 得到 $\widehat{\text{var}}(\hat{\theta}_1) = \frac{(\hat{\theta}_1 - l_1)^2}{z_{\alpha/2}^2}$ 和 $\widehat{\text{var}}(\hat{\theta}_1) = \frac{(u_1 - \hat{\theta}_1)^2}{z_{\alpha/2}^2}$, 其中 $z_{\alpha/2}$ 是标准正态分布的上 $\alpha/2$ 分位数. 类似地, $\widehat{\text{var}}(\hat{\theta}_2) = \frac{(\hat{\theta}_2 - l_2)^2}{z_{\alpha/2}^2}$ 和 $\widehat{\text{var}}(\hat{\theta}_2) = \frac{(u_2 - \hat{\theta}_2)^2}{z_{\alpha/2}^2}$. 所以, $CI(L, U)$ 由 $\hat{\theta}_1 - \hat{\theta}_2 \mp z_{\alpha/2} \sqrt{\widehat{\text{var}}(\hat{\theta}_1) + \widehat{\text{var}}(\hat{\theta}_2)}$ 给出, 插入 (L, U) 中的方差估计量, 得到

$$L = \hat{\theta}_1 - \hat{\theta}_2 - \sqrt{(\hat{\theta}_1 - l_1)^2 + (u_2 - \hat{\theta}_2)^2}, \quad U = \hat{\theta}_1 - \hat{\theta}_2 + \sqrt{(u_1 - \hat{\theta}_1)^2 + (\hat{\theta}_2 - l_2)^2},$$

上述表达式不依赖于 $\hat{\theta}_i$ 的分布.

结果表明, 该方法覆盖真值的概率接近目标置信水平, 且置信区间宽度比较合理. Wald 置信区间易于计算, 而似然估计置信区间没有显式表达, 需要迭代求解.

Wald 置信区间为 $[\max(-1, c\hat{\beta}^T - z_{1-\alpha/2}\sqrt{c\hat{I}^{-1}c^T}), \min(1, c\hat{\beta}^T + z_{1-\alpha/2}\sqrt{c\hat{I}^{-1}c^T})]$, 剖面似然置信区间满足

$$l_2(\hat{\Delta}, \hat{\pi}_1, \hat{R}) - l_2(\Delta_0, \hat{\pi}_1, \hat{R}) \leq \frac{\chi_{1-\alpha}^2(1)}{2},$$

其中 $\Delta \in [-1, 1]$.

此外, Yang 等²⁾构造渐近同时置信区间 (simultaneous confidence interval, SCI), 调整多重性及组内相关性, 用于多重治愈比率的比较, 重点调整多重性进行多对一的比较. 对于所有可能的成对比较, 有其他方法控制误报率, 如 Tukey 方法和 Schffe 方法. 在构造似然 SCI 时, 常使用 Bonferroni 方法, 因为它在计算上容易实现. 他们得到了 Wald 置信区间的简单估计 π_1 , 在 Monte Carlo 模拟中, 将相关性假设下的 Wald 置信区间和似然置信区间与 MOVER 进行了对比.

2.2.2 治愈率比值的区间估计

Peng 等^[30]针对配对相关数据的治愈率比值提出了 4 种置信区间 (基于 MLE 的 Wald 置信区间、似然置信区间、得分置信区间和 MOVRE), 应用对数变换形式的 MOVER, 通过除以“e 效应”调整方差估计. 基于 R 模型, Rosner^[5]给出了 $\text{var}(\hat{\pi}_i) = \text{var}(\tilde{\pi}_i)/e_i$, 其中 $\text{var}(\tilde{\pi}_i)$ 是配对测量值完全独立下的方差估计及 $e_i = \pi_i(1 - \pi_i)/(\pi_i(1 - \pi_i) + (R - 1)\pi_i^2)$. 由于 MOVER 中的方差估计是基于配对测量相互独立的假设, 因此可通过原始估计除去“e 效应”来得到调整后的方差估计, 将此调整用于 MOVER, 得到

$$L = \ln p_2 - \ln p_1 - \sqrt{\frac{[\ln p_2 - lp_{12}]^2}{e_1} + \frac{[lp_{u1} - \ln p_1]^2}{e_2}},$$

$$U = \ln p_2 - \ln p_1 + \sqrt{\frac{[lp_{u2} - \ln p_2]^2}{e_1} + \frac{[\ln p_2 - lp_{l1}]^2}{e_2}}.$$

Wald 置信区间和调整后的 MOVER 置信区间可以直接推导出, 而得分置信区间和似然置信区间没有显式形式, 需要通过迭代方法解决, 收敛速度令人满意. 结果表明, 得分置信区间在具有良好覆盖

2) Yang Z Y, Tian G L, Liu X, et al. Simultaneous confidence interval construction for many-to-one comparisons of proportion differences based on correlated paired data. Unpublished manuscript, 2019

范围和宽度情形下表现最佳, 因此建议使用, 该结果与在独立及配对相关情形下的二元数据推断结果一致. MOVER 置信区间也是一个很好的选择, 比 Wald 置信区间方法计算更简单, 性能更好.

3 有分层效应的配对相关数据分析

3.1 同质性检验

3.1.1 治愈率差别的同质性检验

同质性是分层配对数据统计分析中不可忽视的条件. 近年来, 其检验和计算方法均有很大改进, Tang 和 Qiu^[31] 研究了两个比例差异的同质性检验, 并提出了一种基于 Rosner 模型的修正得分法. 然而, 有约束的最大似然估计基于所有分层的差异均为零的特殊情形, 而不是所有分层的差异都相等. 在组内相关的假定下, 对分层配对相关数据的两个比例差异的同质性检验, Shen 和 Ma^[32] 根据检验方法 (似然比检验、Wald 检验和得分检验) 提出 3 个可替代的最大似然估计. 更一般地, 有约束的最大似然估计可利用分层之间相等的共同差异导出. 由于经典算法 (如 Fisher 评分算法和 Newton-Raphson 法) 在高维情况下很难实现, 他们根据 Ma 和 Liu^[25] 的方法计算了受约束的最大似然估计, 并提出了两步法获得无约束的估计, 效果明显.

结果表明, 得分检验在稳健地控制 I 类错误下具有较高功率, 因此, 在这 3 种方法中, 建议使用得分检验, 其得分检验统计量 T_{SC} 为

$$T_{SC} = \sum_{j=1}^J U_j I_j^{-1} (\pi_{1j}, \pi_{2j}, \rho_j) U_j^T \Big|_{\pi_{1j} = \hat{\pi}_{1jH_0}, \pi_{2j} = \hat{\pi}_{2jH_0}, \rho_j = \hat{\rho}_{jH_0}},$$

其中 $\hat{\pi}_{1jH_0}$ 、 $\hat{\pi}_{2jH_0}$ 和 $\hat{\rho}_{jH_0}$ 是原假设下的最大似然估计.

3.1.2 治愈率比值的同质性检验

Pei 等^[33] 分析了 Donner 模型中分层配对数据治愈率比值的同质性检验, Xue 和 Ma^[34] 针对常数 R 模型提出了 3 个假设检验 (似然比检验、Wald 检验和得分检验), 用分层随机临床试验中的配对相关二元数据来检验跨层治愈率比值的同质性. 结果表明, 得分检验优于其他统计检验, 高功率且能有效地控制 I 类错误, 因此建议使用. 得分检验统计量可以写成每个层次中 J 得分统计量的总和:

$$T_{SC} = \sum_{j=1}^J U_j I_j^{-1} U_j^T \Big|_{\delta_j = \delta_{jH_0}, \pi_{1j} = \hat{\pi}_{1jH_0}, R_j = \hat{R}_{jH_0}},$$

其中 $U_j = (\frac{\partial l_j}{\partial \delta}, 0, 0)$ 和 I_j 是 Fisher 得分函数的第 j 层信息矩阵.

3.1.3 治愈率比值比的同质性检验

Shen 等^[35] 提出了 3 种检验方法 (似然比检验、对数变换的 Wald 检验和得分检验), 在相关系数相等假设下检验分层配对相关数据的比值比同质性, 利用 Monte Carlo 模拟评估该检验方法的性能. 结果表明, 对数变换的 Wald 检验和得分检验表现良好, 高功率且能稳健控制 I 类错误, 因此建议使用. 令 $\delta = (\theta_1, \pi_{11}, \rho_1, \dots, \theta_J, \pi_{1J}, \rho_J)$, 得分函数 U 是一个行向量, 元素 $U_j = (\frac{\partial l}{\partial \theta_j}, \frac{\partial l}{\partial \pi_{1j}}, \frac{\partial l}{\partial \rho_j})$

是 U 中第 j 个分量, $I_{(\delta)}$ 表示 δ 的全局信息矩阵, 具有块对称结构, 在每个主对角线块中有 $I_j^{(\delta)}$, 即 $I_{(\delta)} = \text{diag}(I_1^{(\delta)}, \dots, I_j^{(\delta)})$, 其得分统计量为

$$T_{SC} = \sum_{j=1}^J \left(\frac{\partial l_j}{\partial \theta_j} \right)^2 [I_j^{(\delta)}]^{-1}(1, 1) \Big|_{\theta_j = \hat{\theta}_{jH_0}, \pi_{1j} = \pi_{1jH_0}, \rho_j = \hat{\rho}_{jH_0}},$$

其中 $[I_j^{(\delta)}]^{-1}(1, 1)$ 表示矩阵 $[I_j^{(\delta)}]^{-1}$ 的第 $(1, 1)$ 个元素.

3.2 共同参数的检验和区间估计

3.2.1 治愈率差异的共同参数检验和区间估计

Tang 和 Qiu^[31] 对 Rosner 模型进行两个比例差异的共同参数检验, 并讨论了相应的样本公式. 然而, 具有约束的最大似然估计是建立在所有分层的共同差异等于 0 的基础上的. Shen 等^[36] 提出了 3 种方法 (似然比检验、Wald 检验和得分检验) 检验在相关系数相等的条件下分层双边数据的共同风险差异. 此外, 他们得到了两个比值共同差异的 5 个置信区间, 其中包括两个权重修正方法 (全局 Wald 置信区间和替代 Wald 置信区间) 和 3 个基于测试的方法 (似然置信区间、完全 Wald 置信区间和得分置信区间). 结果表明, 得分检验统计量在某种程度上优于其他统计量, 前者高功率且稳健地控制 I 类错误. 根据得分检验统计量得到的得分置信区间具有合理间隔宽度的覆盖概率, 令人满意. 得分检验统计量可以简化为

$$T_{SC} = \left(\sum_{j=1}^J \frac{\partial l_j}{\partial d} \right)^2 I_n^{-1}(1, 1),$$

其中 $I_n^{-1}(1, 1)$ 表示矩阵 I_n^{-1} 的第 $(1, 1)$ 个元素.

由于得分检验统计量在原假设下服从自由度为 1 的 χ^2 分布, 对所有 $-1 \leq d_0 \leq 1$ 满足不等式

$$T_{SC} \leq \chi_{1, 1-\alpha}^2,$$

得到 $100(1-\alpha)\%$ 的置信区间, 其中 $\chi_{1, 1-\alpha}^2$ 是自由度为 1 的 χ^2 分布的上 $1-\alpha$ 分位数, 可用二分法搜索置信区间的上限和下限.

3.2.2 治愈率比值的共同参数检验和区间估计

假设每层的相关系数相等, Zhuang 等^[37] 提出了 4 个检验统计量 (似然比统计量 T_L 、Wald 统计量 T_W 、得分统计量 T_{SC}^a 和 T_{SC}^b) 来检验治愈率比值的共同参数. 结果表明, 得分统计量 T_{SC}^b 可以得到可控的 I 类错误, 功效令人满意, 其表达式为

$$T_{SC}^b = \left(\sum_{j=1}^J e_j \right)^2 I(1, 1),$$

其中 $I(1, 1)$ 是 Fisher 信息矩阵逆的第一个主对角元素,

$$e_j = -\frac{m_{02j}[2(1-\rho_j)(1-\pi_{1j})+\rho_j]}{\rho_j(1-\pi_{1j})+(1-\rho_j)(1-\pi_{1j})^2} + \frac{m_{12j}(1-2\pi_{1j})}{\pi_{1j}(1-\pi_{1j})} + \frac{m_{22j}[2\pi_{1j}(1-\rho_j)+\rho_j]}{\pi_{1j}\rho_j+(1-\rho_j)\pi_{1j}^2}.$$

进一步, 在每层的相关系数相等的条件下, Zhuang 等^[38] 给出了 6 个置信区间的方法 (基于 MLE 的平均加权 Wald 置信区间、基于样本大小加权的 Wald 置信区间、基于 MLE 的 Wald 置信区间、似

然置信区间、基于 MLE 的得分置信区间和基于全局 MLE 的 Wald 置信区间). Monte Carlo 模拟表明, 基于全局 MLE 的 Wald 置信区间可以得到最短的平均区间宽度 (mean interval width, MIW) 和满意的经验覆盖概率 (empirical coverage probability, ECP), 而且有显式解, 而似然置信区间和基于 MLE 的得分置信区间提供较优的非覆盖概率比率 (ratio of non coverage probability, RNCP) 和对称性.

似然比检验统计量可写为

$$T_L = 2[l(\hat{\delta}, \hat{\pi}_1, \hat{\rho}) - l(\delta_0, \tilde{\pi}_1, \tilde{\rho})],$$

治愈率比值 δ 的 $100(1 - \alpha)\%$ 似然置信区间满足

$$2[l(\hat{\delta}, \hat{\pi}_1, \hat{\rho}) - l(\delta_0, \tilde{\pi}_1, \tilde{\rho})] \leq \chi_{1,1-\alpha}^2, \quad (3.1)$$

类似地, $\delta \in (0, \infty)$ 的 $100(1 - \alpha)\%$ 得分置信区间满足

$$T_{SC} \leq \chi_{1,1-\alpha}^2, \quad (3.2)$$

可用二分法搜索置信区间的上限和下限. $100(1 - \alpha)\%$ 基于全局 MLE 的 Wald 置信区间的上限和下限分别为

$$\text{LCL}_C = \max\{0, \mathbf{C}\hat{\beta}^T - z_{1-\alpha/2}\sqrt{\mathbf{C}\mathbf{I}^{-1}\mathbf{C}^T}\}, \quad \text{UCL}_C = \mathbf{C}\hat{\beta}^T + z_{1-\alpha/2}\sqrt{\mathbf{C}\mathbf{I}^{-1}\mathbf{C}^T},$$

其中 $z_{1-\alpha/2}$ 是标准正态分布的上 $1 - \alpha/2$ 分位数.

4 讨论

本文综合评述用于分析配对相关数据及分层配对相关数据的统计模型. 迄今为止, 已经产生了大量统计方法解决配对相关数据的组内相关性, 尤其是检测分层因子 (或混杂效应) 的统计方法. 但是, 由于这些方法均具有很强的假设条件, 若偏离这些假设可能会导致错误的结果. 在配对相关数据分析中, 关键在于选择适合观察数据的统计模型. Tang 等^[39] 比较了两组相关数据的不同拟合优度, Liu 等³⁾ 用 g ($g \geq 2$) 组相关二进制数据进行综合调查. 结果显示, 通过功率和 I 类错误测量的拟合优度方法取决于模型, 样本量小或 (和) 高度相关数据结构时, 观察性能差异更加显著.

另一个重要的假设是 Donner 模型的假定条件: 组内相关性是相同的, 即 $\text{Corr}(Z_{ijk}, Z_{ij(3-k)}) = \rho$. 但是, 当两组之间的相关性显著不同时, 无法使用 Donner 模型进行数据分析. 在实践中, 首先验证相关系数相等来决定 Donner 模型是否合适. 为此, Pei 等^[40] 构建了用于评估每个成对身体部位的两组之间相关性相等的检验, 并在双臂随机临床试验中给出确定测试样本大小的计算方法. 他们提供了 4 种统计检验: 似然比检验、Wald 检验、Rosner 检验和得分检验. 然而, 他们仅考虑两组中的数据, 所提出的 3 种 Wald 检验均不是精确的 Wald 检验, 对于似然比和得分检验, 未能提供 Fisher 评分算法的详细过程. 更一般地, Liu 等^[41] 提供了可用于多个组的检验方法, 并得到了 Fisher 评分算法公式和 Wald 统计量的表达式.

参考文献

- 1 McCullagh P, Nelder J. Generalized Linear Models. Boca Raton: Chapman & Hall, 1983
- 3) Liu X, Liu S, Ma C X. Goodness-of-fit tests for correlated data from multiple groups. Unpublished manuscript, 2019

- 2 Liang K Y, Zeger S L. Longitudinal data analysis using generalized linear models. *Biometrika*, 1986, 73: 13–22
- 3 Hardin J, Hilbe J. *Generalized Estimating Equations*. Boca Raton: Chapman & Hall/CRC, 2013
- 4 Ma C X, Shan G, Liu S. Homogeneity test for correlated binary data. *PLoS ONE*, 2015, 10: e0124337
- 5 Rosner B. Statistical methods in ophthalmology: An adjustment for the intraclass correlation between eyes. *Biometrics*, 1982, 38: 105–114
- 6 Dallal G E. Paired bernoulli trials. *Biometrics*, 1988, 44: 253–257
- 7 Donner A. Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics*, 1989, 45: 605–611
- 8 Thompson J R. The χ^2 test for data collected on eyes. *British J Ophthalmology*, 1993, 77: 115–117
- 9 Kahan B C, Morris T P. Analysis of multicentre trials with continuous outcomes: When and how should we account for centre effects? *Stat Med*, 2013, 32: 1136–1149
- 10 Kahan B C. Accounting for centre-effects in multicentre trials with a binary outcome—When, why, and how? *BMC Med Res Methodol*, 2014, 14: 20
- 11 Nam J. Power and sample size for stratified prospective studies using the score method for testing relative risk. *Biometrics*, 1998, 54: 331–336
- 12 Lui K J. A simple test of the homogeneity of risk difference in sparse data: An application to a multicenter study. *Biom J*, 2005, 47: 654–661
- 13 Zou G Y, Donner A. Extension of the modified Poisson regression model to prospective studies with correlated binary data. *Stat Methods Med Res*, 2013, 22: 661–670
- 14 Edwards A W F. The measure of association in a 2×2 table. *J R Statist Soc Ser A (General)*, 1963, 126: 109–114
- 15 Agresti A. *Categorical Data Analysis*, Volume 482. New York: John Wiley & Sons, 2003
- 16 Walter S D. Choice of effect measure for epidemiological data. *J Clin Epidemiology*, 2000, 53: 931–939
- 17 Tang M L, Tang N S, Rosner B. Statistical inference for correlated data in ophthalmologic studies. *Stat Med*, 2006, 25: 2771–2783
- 18 Shan G, Ma C X. Exact methods for testing the equality of proportions for binary clustered data from otolaryngologic studies. *Stat Biopharm Res*, 2014, 6: 115–122
- 19 Liu X, Shan G, Tian L, et al. Exact methods for testing homogeneity of proportions for multiple groups of paired binary data. *Comm Statist Simulation Comput*, 2017, 46: 6074–6082
- 20 Tang N S, Tang M L, Qiu S F. Testing the equality of proportions for correlated otolaryngologic data. *Comput Statist Data Anal*, 2008, 52: 3719–3729
- 21 Donner A, Klar N. Methods for comparing event rates in intervention studies when the unit of allocation is a cluster. *Am J Epidemiol*, 1994, 140: 279–289
- 22 Ahn C, Jung S H, Donner A. Application of an adjusted χ^2 statistic to site-specific data in observational dental studies. *J Clin Periodontol*, 2002, 29: 79–82
- 23 Jung S H, Ahn C, Donner A. Evaluation of an adjusted chi-square statistic as applied to observational studies involving clustered binary data. *Stat Med*, 2001, 20: 2149–2161
- 24 Pei Y B, Tang M L, Guo J H. Testing the equality of two proportions for combined unilateral and bilateral data. *Comm Statist Simulation Comput*, 2008, 37: 1515–1529
- 25 Ma C X, Liu S. Testing equality of proportions for correlated binary data in ophthalmologic studies. *J Biopharm Statist*, 2017, 27: 611–619
- 26 Pei Y B, Tang M L, Wong W K, et al. Confidence intervals for correlated proportion differences from paired data in a two-arm randomised clinical trial. *Stat Methods Med Res*, 2012, 21: 167–187
- 27 Tang N S, Qiu S F, Tang M L, et al. Asymptotic confidence interval construction for proportion difference in medical studies with bilateral data. *Stat Methods Med Res*, 2011, 20: 233–259
- 28 Zou G Y, Donner A. Construction of confidence limits about effect measures: A general approach. *Stat Med*, 2008, 27: 1693–1702
- 29 Zou G Y. On the estimation of additive interaction by use of the four-by-two table and beyond. *Am J Epidemiol*, 2008, 168: 212–224
- 30 Peng X, Liu C, Liu S, et al. Asymptotic confidence interval construction for proportion ratio based on correlated paired data. *J Biopharm Stat*, 2019, 29: 1137–1152
- 31 Tang N S, Qiu S F. Homogeneity test, sample size determination and interval construction of difference of two proportions in stratified bilateral-sample designs. *J Statist Plann Inference*, 2012, 142: 1242–1251
- 32 Shen X, Ma C X. Testing homogeneity of difference of two proportions for stratified correlated paired binary data. *J*

- Appl Stat, 2018, 45: 1410–1425
- 33 Pei Y B, Tian G L, Tang M L. Testing homogeneity of proportion ratios for stratified correlated bilateral data in two-arm randomized clinical trials. *Stat Med*, 2014, 33: 4370–4386
 - 34 Xue Y Q, Ma C X. Interval estimation of proportion ratios for stratified bilateral correlated binary data. *Stat Methods Med Res*, 2019, 58: 096228021988204
 - 35 Shen X, Ma C X. Testing homogeneity of difference of two proportions for stratified correlated paired binary data. *J Appl Stat*, 2018, 45: 1410–1425
 - 36 Shen X, Ma C X, Yuen K C, et al. Common risk difference test and interval estimation of risk difference for stratified bilateral correlated data. *Stat Methods Med Res*, 2019, 28: 2418–2438
 - 37 Zhuang T, Tian G L, Ma C X. Homogeneity test of ratio of two proportions in stratified bilateral data. *Stat Biopharm Res*, 2019, 11: 200–209
 - 38 Zhuang T T, Tian G L, Ma C X. Confidence intervals for proportion ratios of stratified correlated bilateral data. *J Biopharm Statist*, 2008, 29: 1–23
 - 39 Tang M L, Pei Y B, Wong W K, et al. Goodness-of-fit tests for correlated paired binary data. *Stat Methods Med Res*, 2012, 21: 331–345
 - 40 Pei Y B, Tang M L, Wong W K, et al. Testing equality of correlations of two paired binary responses from two treated groups in a randomized trial. *J Biopharm Statist*, 2011, 21: 511–525
 - 41 Liu X, Liu S, Ma C X. Testing equality of correlation coefficients for paired binary data from multiple groups. *Comm Statist Simulation Comput*, 2015, 86: 1–11

Statistical methods for the bilateral correlated data

Xi Shen & Changxing Ma

Abstract Bilateral correlated data are often encountered in medical researches such as ophthalmologic (or otolaryngologic) studies, in which each unit contributes information from paired organs to the data analysis. Measurements obtained from paired organs of a subject are generally highly correlated, whereas many statistical tests assume observations in a sample are independent. Previous studies showed that the statistical inference for bilateral correlated data ignoring the presence of intra-class correlation could lead to inflated significance levels. Various statistical methods have been developed to tackle this intra-class correlation on the bilateral correlated data analysis. Furthermore, in some studies, it is very important to adjust the effect of confounder on statistical inferences, since either ignoring the intra-class correlation or confounding effect may lead to biased results. In this paper, we review these methods, discuss the applications and provide statistical advice.

Keywords bilateral correlated data, intra-class correlation coefficients, interval estimation, stratified bilateral data, statistical testing

MSC(2010) 62F03, 62F05, 62F12

doi: 10.1360/N012019-00109