

配对设计中风险差的置信区间构造

张丽平¹, 古丽斯坦·库尔班尼牙孜¹, 田茂再^{1,2}

(1. 新疆财经大学 统计与数据科学学院, 新疆乌鲁木齐 830012;

2. 中国人民大学 统计学院, 北京 100872)

摘要: 风险差在配对设计中常用来对比某种疾病治疗前后或某种政策措施实施前后的有效性. 文中针对风险差(Risk Difference)在配对设计中的置信区间估计问题, 利用Delta法, 改进Wald法, 反双曲正切变换法, 似然比检验法, 鞍点逼近法方法进行置信区间构造, 并通过Monte Carlo模拟计算区间覆盖率和区间长度, 以比较这5种方法的表现性能. 结果显示: 不同方法的区间表现均不相同. 其中鞍点逼近法明显优于其他4种方法; 改进Wald法的表现仅次于鞍点逼近法, 且随着样本量的增大, 改进Wald法和反双曲正切变换法的性能差距逐渐减小; 最后通过一个实例来验证这5种区间估计方法.

关键词: 配对设计; 鞍点逼近; Monte Carlo模拟; 置信区间

中图分类号: O212.2

文献标识码: A **文章编号:** 1000-4424(2024)01-0028-13

§1 引言

每当新型流行病到来时, 新药研制中两种药物疗效的差异度量, 治疗方案效果差异的度量, 政策措施的有效性评估等都成了必不可少的研究问题, 为了提高临床试验的效率和降低时间成本, 通常要先采用较小的样本量进行配对设计试验. 即先将病人按照性别, 年龄, 生活环境等分成若干个匹配对, 再让其随机分为实验组和对照组, 进行处理和试验. 配对设计试验的优势在于不仅能考虑到配对数据中组内相关性, 还能够避免混杂效应所导致的结果偏差^[1]. 另外, 某些配对设计试验研究中也常用自身对照, 如研究同一受试对象在某疾病治疗前后的某指标对比, 同一受试对象在施行政策措施前后的不同态度对比, 同一受试对象的不同检测方法对比等.

流行病学中有很多常用的评估指标, 如优势比, 风险差, 相对危险度等的一系列差值或比值类指标^[2]. Berkson^[3]曾指出当使用比率作为研究指标时, 其衡量的效果不是很好. 尤其是在研究疾病死亡率比率时, 百分之一与千分之一同时增加十倍, 虽然这十倍在数量上是相同的, 但后者增加的十倍明显比前者有着更严重的影响^[4]. 鉴于此, 本文将考虑了风险差(Risk

收稿日期: 2022-07-13 修回日期: 2023-04-11

通讯作者: 田茂再, E-mail: mztian@ruc.edu.cn

基金项目: 中国人民大学科学研究基金(中央高校基本科研业务费专项资金)项目(22XNL016)

Difference)指标. 在配对设计中, 风险差能够衡量对照组和实验组中两比率的差异, 从而确定某种药物或治疗方法的风险大小. 如在药物研制中, 风险差可以了解某新药与原有同类药疗效的差异程度. 当某疾病使用不同治疗方案时, 风险差可以度量两种治疗方案的成功率差异大小. 此外, 经文献梳理发现, 一方面, 目前大多数的风险差的区间构造都是在逆抽样, 二项抽样等条件下进行的(Lui^[5], Tang等^[6], Krishnamoorthy等^[7,8]), 相比之下, 有关配对设计中风险差置信区间构造研究非常有限. 另一方面, 从区间构造的方法来看, 对于置信区间构造的传统估计方法有很多, 本文之所以选取Delta法, 似然比检验法, 改进Wald法, 反双曲正切变换法这4种方法作为传统的区间估计方法, 是因为这些方法在流行病学指标的置信区间构造方法研究中, 是通用且传统的区间估计方法^[2,9-14]. 另外, 鞍点逼近法在流行病学中是较为新颖的指标区间构造方法, 近年来正逐渐在优势比, 总体成数, 需治疗数等指标的区间构造中流行. 例如罗玉波等^[15]早先在研究广义卡方型混合分布的过程中, 将正态近似, 卡方近似与鞍点逼近方法得到的逼近式进行了比较, 证明了鞍点逼近效果要优于其他两种近似方法. 后来, 孟令宾等^[16]考虑在二项抽样下, 使用Delta法, Woolf法, 似然比检验法, 鞍点逼近法来构造优势比的渐近置信区间, 并通过模拟比较发现鞍点逼近法表现着良好的区间特性. 夏丽丽等^[17]考虑了在二项分布下利用鞍点逼近法, 枢轴量法和正态近似法构造总体成数的置信区间, 经验证也发现鞍点逼近法相对较好. 当然还有其他文献^[10,14,18]也展现其良好的估计效果.

因此本文基于前期的相关研究^[9], 考虑在配对设计中风险差进行置信区间构造. 并将鞍点逼近的区间构造方法推广到风险差指标的置信区间构造当中, 以此来观察此方法和其他传统方法(改进Wald法, Delta法, 反双曲正切变换法, 似然比检验法)的估计效果对比情况. 并通过Monte Carlo模拟和实例分析来比较不同方法.

§2 风险差置信区间构造

现举例来说明配对设计试验数据, 假设感兴趣的是了解某药物治疗疾病的有效性. 在配对设计试验中, 将患者按照某些条件分为 n 对匹配对, 并随机分为病例组和对照组, 让病例组服用某药物, 对照组服用安慰剂, 一段时间后进行复查. 复查结果如表1所示, 病例组中未患病而对对照组中患病的匹配对为 n_b , 病例组中患病而对对照组中未患病的匹配对为 n_a , 病例组和对照组中均患病的匹配对为 n_c , 病例组和对照组中均未患病的匹配对为 n_d , 且相应结果发生的概率依次为 p_b , p_a , p_c , p_d .

表1 配对设计的数据结构

病例组	对照组		总数
	患病	未患病	
患病	$n_c(p_c)$	$n_a(p_a)$	$n_c + n_a(p_{1\cdot})$
未患病	$n_b(p_b)$	$n_d(p_d)$	$n_b + n_d(p_{2\cdot})$
总数	$n_c + n_b(p_{\cdot 1})$	$n_a + n_d(p_{\cdot 2})$	$n(1)$

由于病例组和对照组的受试对象都来自同一匹配对, 且疗效在同一匹配对很可能是相关的. 这时若使用独立二项抽样下进行风险差的区间估计是不合适的. 因此本文采用了三项式配对抽样设计, 假设随机变量 $(n_a, n_b, n - n_a - n_b)$ 服从概率为 $(p_a, p_b, 1 - p_a - p_b)$ 的三项式分布, 其对应的结局有三种结局. 根据风险差(Risk Difference)的定义^[1], 在配对设计中, 风险差表示对照组中未患病的概率 p_a 与病例组中未患病的概率 p_b 之差, 即 $R = p_a - p_b$, 且 $R \in [-1, 1]$. 那么随机变量 $(n_a, n_b, n - n_a - n_b)$ 的概率密度函数为

$$f(p_a, p_b) = \frac{n!}{n_a!n_b!(n - n_a - n_b)!} (p_a)^{n_a} (p_b)^{n_b} (1 - p_a - p_b)^{n - n_a - n_b}. \quad (1)$$

由于感兴趣参数为 R ,讨厌参数为 p_b ,所以其对数似然函数为

$$L(R, p_b) = n_a \log(R + p_b) + n_b \log(p_b) + (n - n_a - n_b) \log(1 - R - 2p_b) + C. \quad (2)$$

其中, C 是不依赖于参数 R 的常数. 进一步将似然函数对两个参数 (R, p_b) 分别求偏导为

$$\frac{\partial L(R, p_b)}{\partial R} = \frac{n_a}{R + p_b} + \frac{n_a + n_b - n}{1 - R - 2p_b} = 0,$$

$$\frac{\partial L(R, p_b)}{\partial p_b} = \frac{n_a}{R + p_b} + \frac{n_b}{p_b} + \frac{-2(n - n_a - n_b)}{1 - R - 2p_b} = 0.$$

由此得出各参数的极大似然估计为

$$\hat{p}_a = \frac{n_a}{n}, \hat{R} = \frac{n_a - n_b}{n}, \hat{p}_b = \frac{n_b}{n}. \quad (3)$$

2.1 基于Delta方法的置信区间的构造

Delta方法是一种计算随机变量函数渐近分布的一种方法, 它的思想是在知道某个随机变量的分布时, 使用Taylor展开式进行逼近得到复杂随机变量的均值, 方差等进行估计. 现考虑利用Delta法在配对设计三项式抽样下进行风险差置信区间估计. 首先根据下列定理计算 $R = p_a - p_b$ 的渐近方差, 即

定理2.1(二维Delta定理^[19]) 二维随机变量 $(\hat{p}_a, \hat{p}_b)^\top \xrightarrow{D} N((p_a, p_b)^\top, I)$, \xrightarrow{D} 表示依分布收敛, I 为 $(p_a, p_b)^\top$ 协方差矩阵, $(\hat{p}_a, \hat{p}_b)^\top$ 是 $(p_a, p_b)^\top$ 的极大似然估计值. 若 $g(\hat{p}_a, \hat{p}_b)^\top$ 为 $(\hat{p}_a, \hat{p}_b)^\top$ 映射到二维随机变量的连续函数, 且在 p_a, p_b 处可导, $g'(\hat{p}_a, \hat{p}_b) \neq 0$, 那么就有

$$g(\hat{p}_a, \hat{p}_b) \sim N(g(p_a, p_b), G^\top I G),$$

其中 $G = (\frac{\partial g(p_a, p_b)}{\partial p_a}, \frac{\partial g(p_a, p_b)}{\partial p_b})^\top$.

根据定理2.1, 令 $g(p_a, p_b) = R = p_a - p_b$, $G = (\frac{\partial g(p_a, p_b)}{\partial p_a}, \frac{\partial g(p_a, p_b)}{\partial p_b})^\top$, 若用 I_0 表示 $(\hat{p}_a, \hat{p}_b)^\top$ 的协方差矩阵, 就有

$$G = (1, -1)^\top, I_0 = \begin{pmatrix} \frac{p_a(1-p_a)}{n} & -\frac{p_a p_b}{n} \\ -\frac{p_a p_b}{n} & \frac{p_b(1-p_b)}{n} \end{pmatrix}.$$

由此可以得出 $g(\hat{p}_a, \hat{p}_b) = \hat{p}_a - \hat{p}_b = \hat{R}$ 的渐近方差为

$$\widehat{\text{Var}}(\hat{R}) = G^\top I_0 G = \frac{1}{n} [(\hat{p}_a + \hat{p}_b) - (\hat{p}_a - \hat{p}_b)]^2. \quad (4)$$

若 $Z_{\alpha/2}$ 为标准正态分布的上侧 $\alpha/2$ 分位数, 则构造出风险差 R 的 $100(1 - \alpha)\%$ 置信区间为

$$\left[\hat{R} - Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{R})}, \hat{R} + Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{R})} \right].$$

2.2 基于改进Wald法的置信区间的构造

May等^[20]依据Quesenberry和Hurst^[21]提出的二次不等式方法, 构造风险差的置信区间, 此方法弥补了Wald法在小样本下区间性能表现较差这一缺陷. 现考虑配对设计下风险差的置信区间, 根据Delta法中已经求得的风险差方差, 令 $\tau = p_a + p_b$, 由式(3)可得, $\hat{\tau} = \hat{p}_a + \hat{p}_b$ 是 τ 的相合估计量. 当用 $\hat{\tau}$ 代替 τ 时, 就有 $\widehat{\text{Var}}(\hat{R}) = (\hat{\tau} - \hat{R}^2)/n$. 若 n 取值很大, 则由中心极限定理^[19]可知

$$P\left(\frac{(\hat{R} - R)^2}{\widehat{\text{Var}}(\hat{R})} \leq Z_{\alpha/2}^2\right) = 1 - \alpha. \quad (5)$$

由此得出

$$AR^2 - 2BR + C \leq 0. \quad (6)$$

其中 $A = 1 + Z_{\alpha/2}^2/n$, $B = \hat{R}$, $C = \hat{R}^2 - Z_{\alpha/2}^2 \hat{\tau}/n$. 由于条件 $A > 0$, $B^2 - AC = (\hat{R} - \hat{R}^2) + (\hat{\tau} - R^2)Z_{\alpha/2}^2/n + Z_{\alpha/2}^4 \hat{\tau}/n^2 > 0$ 成立, 所以风险差 R 的 $100(1 - \alpha)\%$ 的置信区间为

$$\left[\max\left\{\left(\frac{B - \sqrt{B^2 - AC}}{A}, -1\right)\right\}, \min\left\{\left(\frac{B + \sqrt{B^2 - AC}}{A}, 1\right)\right\} \right].$$

2.3 基于反双曲正切变换法的置信区间构造

为了尝试改进风险差 \hat{R} 的正态近似, 考虑使用Edwardes^[22]中的 $\tanh^{-1}(x) = (1/2) \log((1+x)/(1-x))$ 变换, 由此在配对设计三项式抽样下, 很容易地证明估计量的渐近方差为

$$\widehat{\text{Var}}\left(\tanh^{-1}(\hat{R})\right) = \frac{\widehat{\text{Var}}(\hat{R})}{(1 - \hat{R}^2)^2}. \quad (7)$$

若 $Z_{\alpha/2}$ 是标准正态分布的上侧 $\alpha/2$ 分位数, 则风险差的 $100(1 - \alpha)\%$ 的渐近置信区间为

$$[\tanh(\tanh^{-1}(\hat{R}) - Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\tanh^{-1}(\hat{R}))}), \tanh(\tanh^{-1}(\hat{R}) + Z_{\alpha/2} \sqrt{\widehat{\text{Var}}(\tanh^{-1}(\hat{R}))})].$$

2.4 基于似然比检验法的置信区间的构造

Casella, Berger^[19]提到统计推断方法中假设检验和区间估计之间有着很强的联系, 可以通过假设检验的统计量来估计出参数的置信区间. 似然比检验就是这样一种经典的方法. 在这里, 对于配对设计的风险差指标, 考虑如下假设检验问题: $H_0 : R = R_0, H_1 : R \neq R_0$, 令参数的限制性似然估计为 $\tilde{\Delta} = (\tilde{R}_0, \tilde{p}_{b0}(R_0))$, 在原假设下, 就有

$$\begin{cases} \frac{\partial L(R, p_b)}{\partial R} = \frac{n_a}{R+p_b} + \frac{-n+n_a+n_b}{1-R-2p_b} = 0, \\ R = R_0. \end{cases} \quad (8)$$

方程组(8)的解为参数的限制性似然估计, 即 $\tilde{\Delta} = (\tilde{R}_0, \tilde{p}_{b0}(R_0)) = (R_0, \frac{n_a - R_0(n - n_b)}{n - n_b + n_a})$. 当样本量很大时, 似然比统计量满足

$$T_{LR} = 2 \left(\frac{L(\hat{R}, \hat{p}_b)}{L(R_0, \tilde{p}_{b0}(R_0))} \right) \rightarrow \chi_1^2. \quad (9)$$

计算时可通过二分法求得置信区间, 即风险差 R 的 $100(1 - a)\%$ 的置信区间为

$$\{R : T_{LR} \leq \chi_{1, 1-a}^2\}.$$

2.5 基于鞍点逼近法的置信区间构造

鞍点逼近方法最早在复变函数中应用, Fulks^[23]介绍了渐近分析方法中的Laplace方法, 让学者对鞍点逼近法有了最初步的了解. Daniels^[24]将这个概念引入到统计推断中, 并且给出了关于 n 个独立同分布随机变量均值的密度函数逼近式. 后来, Barndorff-Nielsen和Cox^[25]给出了极大似然估计的密度函数的鞍点渐近逼近式. 此后, 鞍点逼近理论的重要性在统计推断各方面才渐渐表现出来. 对于置信区间构造方面, 此方法具有计算准确的优点, 其理论能够对密度函数或分布函数的表达式进行逼近, 并且逼近效果较好. 因此, 现使用鞍点逼近方法对配对设计下风险差指标进行区间估计, 根据前文推导, 风险差 R 的极大似然估计为 $\hat{R} = \hat{p}_a - \hat{p}_b = \frac{n_a}{n} - \frac{n_b}{n}$, 令 $F_{\hat{R}}(r)$ 是 \hat{R} 的分布函数, 则有

$$F_{\hat{R}}(r) = P(\hat{R} \leq r) = P(n_a - n_b - rn \leq 0) = P(Z \leq 0). \quad (10)$$

其中 $Z = n_a - n_b - rn$.

定理2.2 三维随机变量 $(n_a, n_b, n - n_a - n_b)$ 服从多项分布 $M(n, p_a, p_b, 1 - p_a - p_b)$ 时, 则随机变量 n_a 边缘分布为二项分布 $B(n, p_a)$, 那么其矩母函数为

$$M_{n_a}(t) = (p_a e^t + 1 - p_a)^n.$$

接下来求解随机变量 Z 的矩母函数. 设随机变量 Z 的矩母函数为 $M_Z(\theta)$, 且 n_a 和 n_b 为随机变量. Z 是一个比较复杂的随机变量, 为了求 Z 的矩母函数, 这里采用一些统计技术来给出近似解^[26], 先使用双期望定理来简化 Z 自身的构造, 后用一元Delta法来给出一个近似解. 换言之, 根据双期望定理, 首先在 n_b 已知情况下求 n_a 的矩母函数. 其次求出随机变量 n_b 的数学期望, 最后根

据定理2.2, 便可得出随机变量 Z 的矩母函数为

$$\begin{aligned} M_Z(\theta) &= E(e^{\theta Z}) = E\left(e^{\theta(-n_b-rn)} E(e^{\theta n_a} | n_b)\right) \\ &= E\left(e^{\theta(-n_b-rn)} M_{n_a}(\theta)\right) = e^{-n\theta(p_b+r)} (p_a e^\theta + 1 - p_a)^n. \end{aligned}$$

进一步得出 Z 的累积生成函数为

$$K_Z(\theta) = \ln M_Z(\theta) = n \ln (p_a e^\theta - p_a + 1) - n\theta (p_b + r). \quad (11)$$

随机变量 Z 的一阶至四阶导数分别为

$$\begin{aligned} \mu_Z(\theta) &= K'_Z(\theta) = \frac{np_a e^\theta}{p_a e^\theta - p_a + 1} - n(p_b + r). \\ \sigma_Z^2(\theta) &= K''_Z(\theta) = \frac{np_a e^\theta}{p_a e^\theta - p_a + 1} - \frac{n(p_a e^\theta)^2}{(p_a e^\theta - p_a + 1)^2}. \\ K'''_Z(\theta) &= \frac{np_a e^\theta}{p_a e^\theta - p_a + 1} - \frac{3n(p_a e^\theta)^2}{(p_a e^\theta - p_a + 1)^2} + \frac{2n(p_a e^\theta)^3}{(p_a e^\theta - p_a + 1)^3}. \\ K^{(4)}_Z(\theta) &= \frac{np_a e^\theta}{p_a e^\theta - p_a + 1} - \frac{7n(p_a e^\theta)^2}{(p_a e^\theta - p_a + 1)^2} + \frac{12n(p_a e^\theta)^3}{(p_a e^\theta - p_a + 1)^3} - \frac{6n(p_a e^\theta)^4}{(p_a e^\theta - p_a + 1)^4}. \end{aligned}$$

根据鞍点逼近原理^[27], 求解鞍点逼近方程 $\mu_Z(\theta) = z$, 进而可以得到随机变量 Z 的鞍点 $\hat{\theta}$. 由于本文需知道 $F_{\hat{R}}(r) = P(Z \leq 0)$, 因此 $\mu_Z(\theta) = 0$ 的根即为随机变量 Z 的鞍点 $\hat{\theta}$. 使用 Φ 表示标准正态分布函数, 则运用Jensen^[27]中的公式(2.2.10)可以进一步得出 \hat{R} 的尾部概率为

$$\begin{aligned} P(\hat{R} \geq r) &= \frac{M_Z(\theta) \exp(-\theta r)}{\sigma_Z(\theta) \{1 - \exp(-|\theta|)\}} \left\{ B_0(\lambda) + \left[\frac{1}{\sigma_Z(\theta)} \left(\frac{1}{|\theta|} - \gamma_\theta \right) B_1(\gamma) + \frac{\zeta_3(\theta)}{6} \operatorname{sgn}(\theta) B_3(\lambda) \right] \right. \\ &\quad + \left[\left(\gamma_\theta \left(\frac{1}{2} - \frac{1}{|\theta|} \right) + \gamma_\theta^2 \right) \frac{B_2(\lambda)}{\sigma_Z(\theta)^2} + \frac{1}{\sigma_Z(\theta)} \left(\frac{1}{|\theta|} - \gamma_\theta \right) \frac{\zeta_3(\theta)}{6} \operatorname{sgn}(\theta) B_4(\lambda) \right. \\ &\quad \left. \left. + \frac{\zeta_4(\theta)}{24} B_5(\lambda) + \frac{\zeta_3(\theta)^2}{72} B_6(\lambda) \right] \right\}. \end{aligned}$$

其中 $\lambda = \sqrt{n}|\theta|\sigma_Z(\theta)$, $\sigma_Z(\theta)^2 = K''_Z(\theta)$,

$$\zeta_3(\theta) = K'''_Z(\theta) / \left(K''_Z(\theta) \right)^{3/2}, \zeta_4(\theta) = K^{(4)}_Z(\theta) / \left(K''_Z(\theta) \right)^2,$$

$$\gamma_\theta = \exp(-|\theta|) / \{1 - \exp(-|\theta|)\}, B_0(\lambda) = \lambda \exp(\lambda^2/2) \{1 - \Phi(\lambda)\},$$

$$B_1(\lambda) = -\lambda \{B_0(\lambda) - (2\pi)^{-1/2}\}, B_2(\lambda) = \lambda^2 \{B_0(\lambda) - (2\pi)^{-1/2}\},$$

$$B_3(\lambda) = -\{\lambda^3 B_0(\lambda) - (\lambda^3 - \lambda)(2\pi)^{-1/2}\}, B_4(\lambda) = \lambda^4 B_0(\lambda) - (\lambda^4 - \lambda^2)(2\pi)^{-1/2},$$

$$B_5(\lambda) = -\{\lambda^5 B_0(\lambda) - (\lambda^5 - \lambda^3 + 3\lambda)(2\pi)^{-1/2}\}, B_6(\lambda) = \lambda^6 B_0(\lambda) - (\lambda^6 - \lambda^4 + 3\lambda^2)(2\pi)^{-1/2}.$$

将鞍点 $\hat{\theta}$ 带入 $P(\hat{R} \geq r)$ 中, 求解出 \hat{R} 的分布函数为

$$F_{\hat{R}}(r) \approx 1 - P(\hat{R} \geq r). \quad (12)$$

通过鞍点逼近法得到风险差的极大似然估计量的渐近分布函数后, 可以根据如下定理得到配对设计风险差的置信区间.

定理2.3 设三维随机变量 $(n_a, n_b, n - n_a - n_b)$ 服从多项分布 $M(n, p_a, p_b, 1 - p_a - p_b)$, 风险差为 R , 且 R 的极大似然估计量为 $\hat{R} = \hat{p}_a - \hat{p}_b$, 其中 $\hat{p}_a = \frac{n_a}{n}$, $\hat{p}_b = \frac{n_b}{n}$. 假设风险差 \hat{R} 的分布函数为 $F_{\hat{R}}(r) = P(\hat{R} \leq r)$. F_{sp} 是 $F_{\hat{R}}(r)$ 的鞍点逼近函数. 设 $0 < \alpha < 1$ 为给定的名义水平, 对任意 r , \hat{R}_L 和 \hat{R}_U 满足

$$F_{sp}(r | \hat{R}_L) = \alpha/2, F_{sp}(r | \hat{R}_U) = 1 - \alpha/2.$$

则配对设计三项式抽样下风险差 R 的置信水平 $100(1 - \alpha)\%$ 的渐近置信区间为 $[\hat{R}_L, \hat{R}_U]$.

§3 Monte Carlo 模拟

通过模拟研究来比较不同方法的区间估计有效性, 根据Lui^[2]中的讨论, 如果一个置信区间的覆盖率接近于置信水平, 并且其区间长度较短, 那么这个方法区间估计的效果就较好. 所以本文通过计算置信区间的真值覆盖率和平均长度, 在保证覆盖率接近置信水平0.95的情形下, 选取区间长度较短的估计方法作为良好的估计方法这一标准, 从而比较上述5种置信区间构造方法(Delta方法(D), 改进Wald方法(MW), 反双曲正切变换法(IH), 似然比检验法(LR), 鞍点逼近法(SA)). 在Monte Carlo模拟中, 考虑如下参数设置: 样本量 $n = 100, 200, 300$, $p_b = 0.1, 0.2, 0.3, 0.4$, 风险差 $R = -0.03, -0.01, 0, 0.01, 0.03$, 置信水平 $1 - \alpha = 0.95$. 每种情况重复1000次. 得出5种不同的置信区间构造方法比较结果如图1-图4所示, 左图表示不同方法随风险差 R 变化的覆盖率情况, 右图表示不同方法随风险差 R 变化的区间长度情况.

通过图1至图4中模拟结果的覆盖率和区间长度情况进行比较, 得出不同方法在这两个指标下的结果表现各不相同. 在4种不同情形($p_b = 0.1, p_b = 0.2, p_b = 0.3, p_b = 0.4$)下, 可以看出随着样本量的增大, 不同方法的区间长度在逐渐减小, 其覆盖率也在逐渐稳定. 具体来看: (1) 根据覆盖率(左列图), D方法, MW法, IH法和SA法的真值覆盖率在多数情况下均接近置信水平0.95, 表现较为稳定. 而LR法的覆盖率在多数情况下都距离名义水平较远, 远不如其他4种方法. 尤其当 $p_b = 0.1, p_b = 0.2$ 时, LR法的覆盖率0.8左右. 对于其他情况($p_b = 0.3, p_b = 0.4$), 该方法的覆盖率也很难稳定在0.95附近, 说明该方法在此类情形下不是一个较好的估计方法. 换言之, 其区间估计结果相比较其他方法不包含真实值的概率较大, 区间估计效果与其他方法相比表现不稳定. (2) 根据区间长度(右列图), 可以看出, D法与MW法, IH法相比, 当样本量较小时, 区间长度性能相对较差. 但随着样本量的增大, 此方法的区间长度逐渐与其他方法差距缩小. 而MW法与其他4种方法相比, 多数情形下该方法的区间长度仅次于SA法和LR法, 尤其是在样本量较小的情况下此现象较为明显. 此外, IH法与MW法相比, 在多数情况下的区间长度较长. 但随着样本量的增大, 这两种方法的区间长度差距逐渐减小. 当然, 还可以看出, 虽然LR在极少数的情况下区间长度较短, 但此时的覆盖率远低于名义水平, 这显然不是良好的区间估计方法的表现. 因此, 在绝大多数情况下, SA法的区间长度要明显优于其他4种方法. (3) 整体来看, SA法的区间性能在这5种方法中表现着明显的优势. 回顾推导过程, 发现这是由于SA法在计算概率密度函数和分布函数时, 具有结果准确性高, 尾部分布表现较好的特点, 使得SA法在不同情况下, 区间整体性能始终呈现着最好的水平. 此外, 对于其他区间估计方法, 在不同情形下的表现各不相同. MW法与D法相比, 不论是区间长度还是覆盖率方面, 性能均有所提升, 尤其是在小样本情况下, MW法更加优于D法. 此外, 多数情况下, IH法比D法的区间性能要好, 而与LR法相比时, IH法在覆盖率方面始终表现较好. 这时虽然LR法的区间长度表现较好, 但是LR法的区间覆盖率在多数情况下比较低, 所以该方法在实际中区间估计中相比其他方法较不稳定. 另外, 还可以看出, 随着样本量和 p_b 增大时, IH法和MW法的区间长度差异逐渐减小.

最后综合考虑上述5种方法的模拟比较结果, 得出如下结论: 在保证覆盖率接近0.95, 且区间长度较短的这一标准下, SA法在这5种方法中表现最好, 是一个较为优秀的估计方法. 相比之下, 对于其他4种方法, 一方面, 虽然LR法在极少数情况下区间长度较短, 但其区间覆盖率较低, 且估计结果表现不是很稳定, 考虑到在实际中往往需要稳定的区间估计方法进行估计, 所以该方法的稳定性在实际中很有可能性不如其他方法. 另一方面, 其他方法的估计效果均没有SA法的表

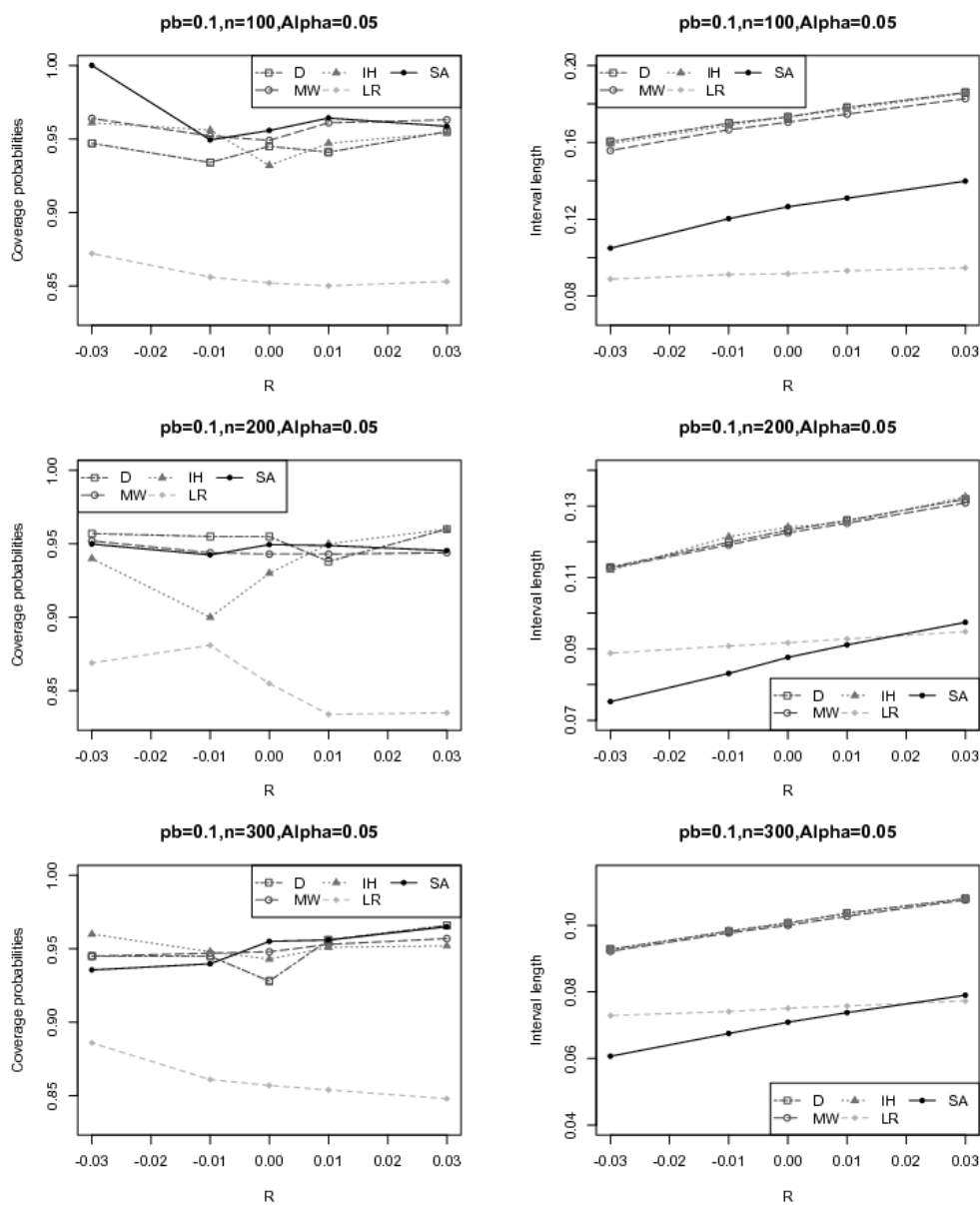


图 1 $p_b = 0.1, \alpha = 0.05, n = 100, 200, 300$ 下相应的覆盖率(左)及区间长度(右)

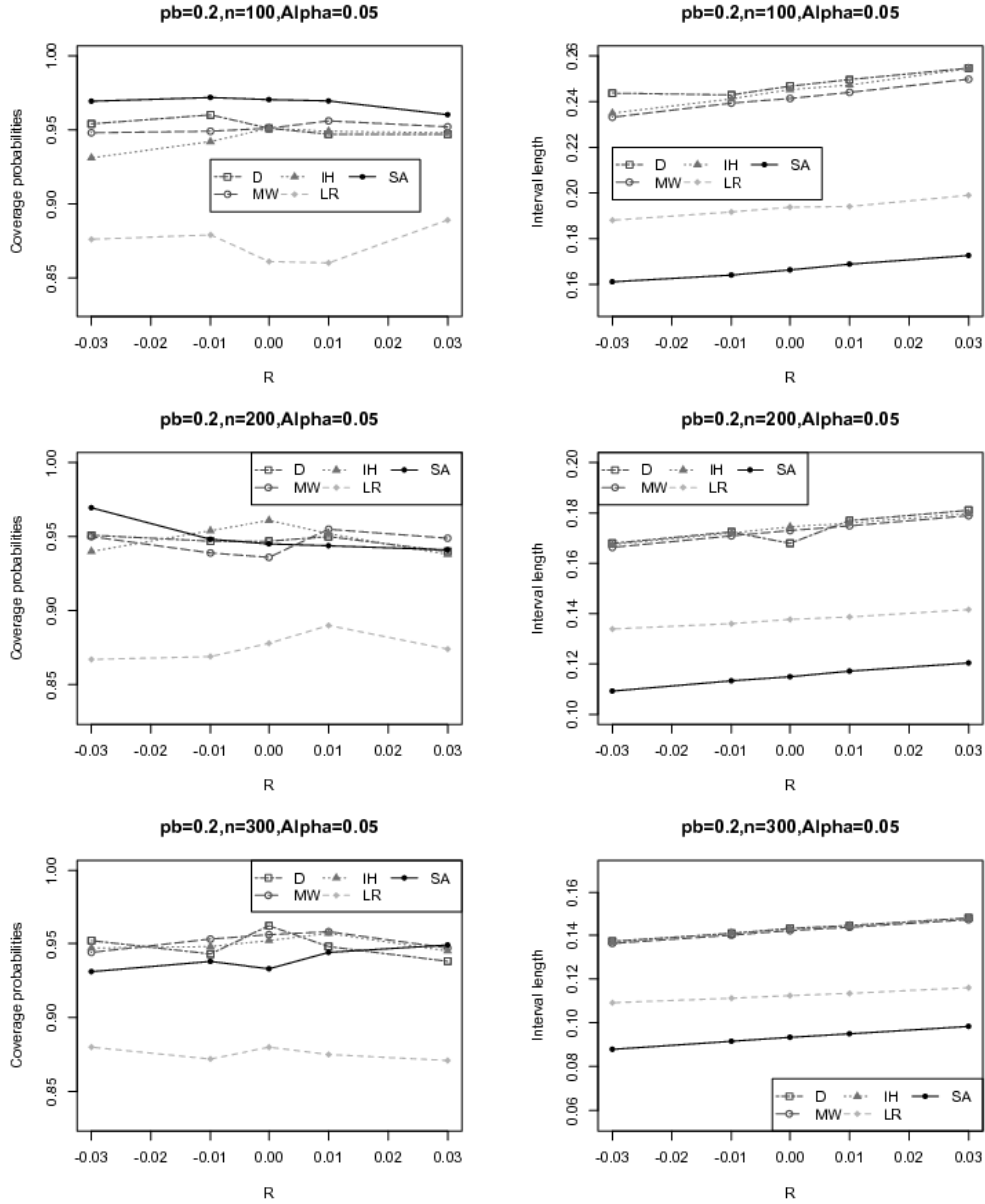


图2 $p_b = 0.2, \alpha = 0.05, n = 100, 200, 300$ 下相应的覆盖率(左)及区间长度(右)

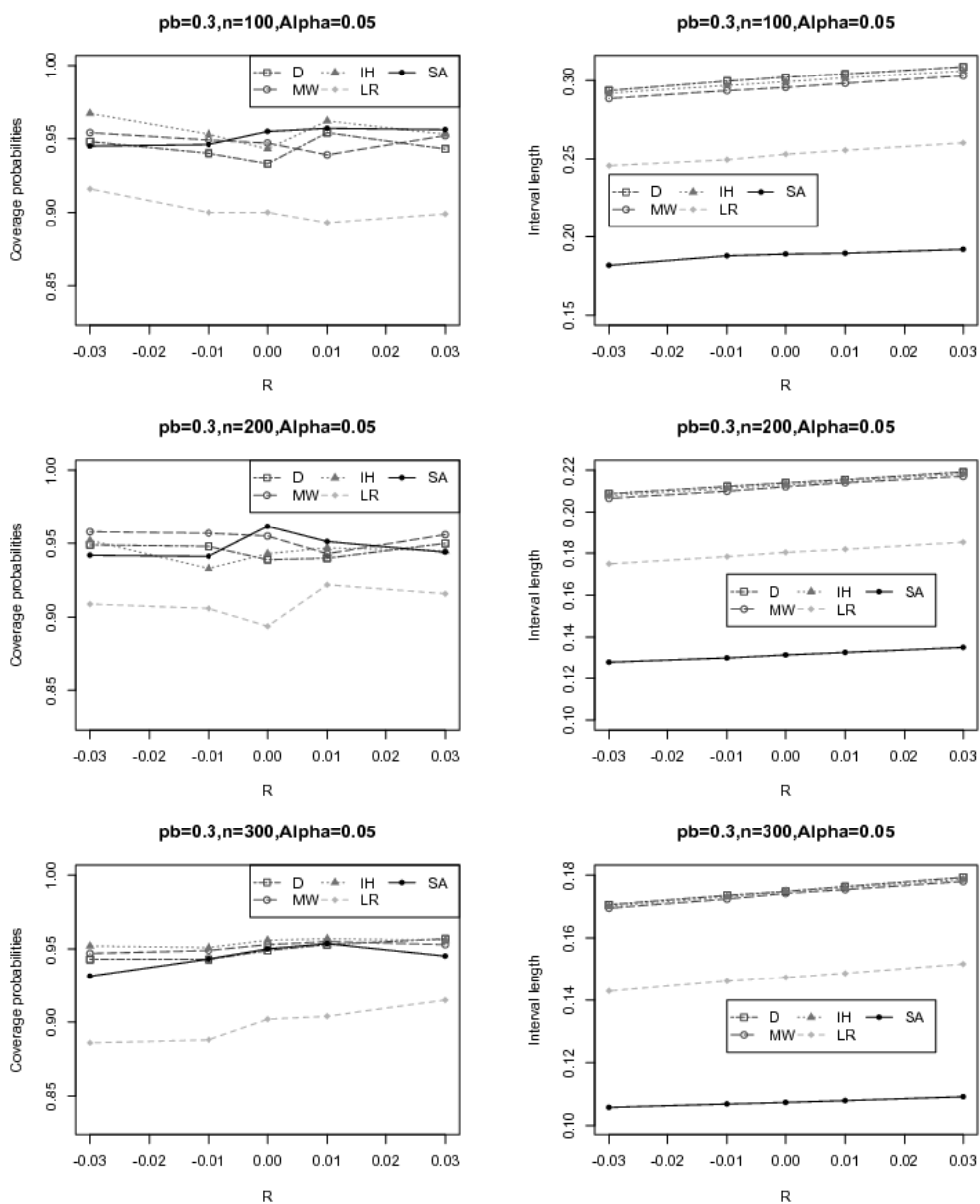


图3 $p_b = 0.3, \alpha = 0.05, n = 100, 200, 300$ 下相应的覆盖率(左)及区间长度(右)

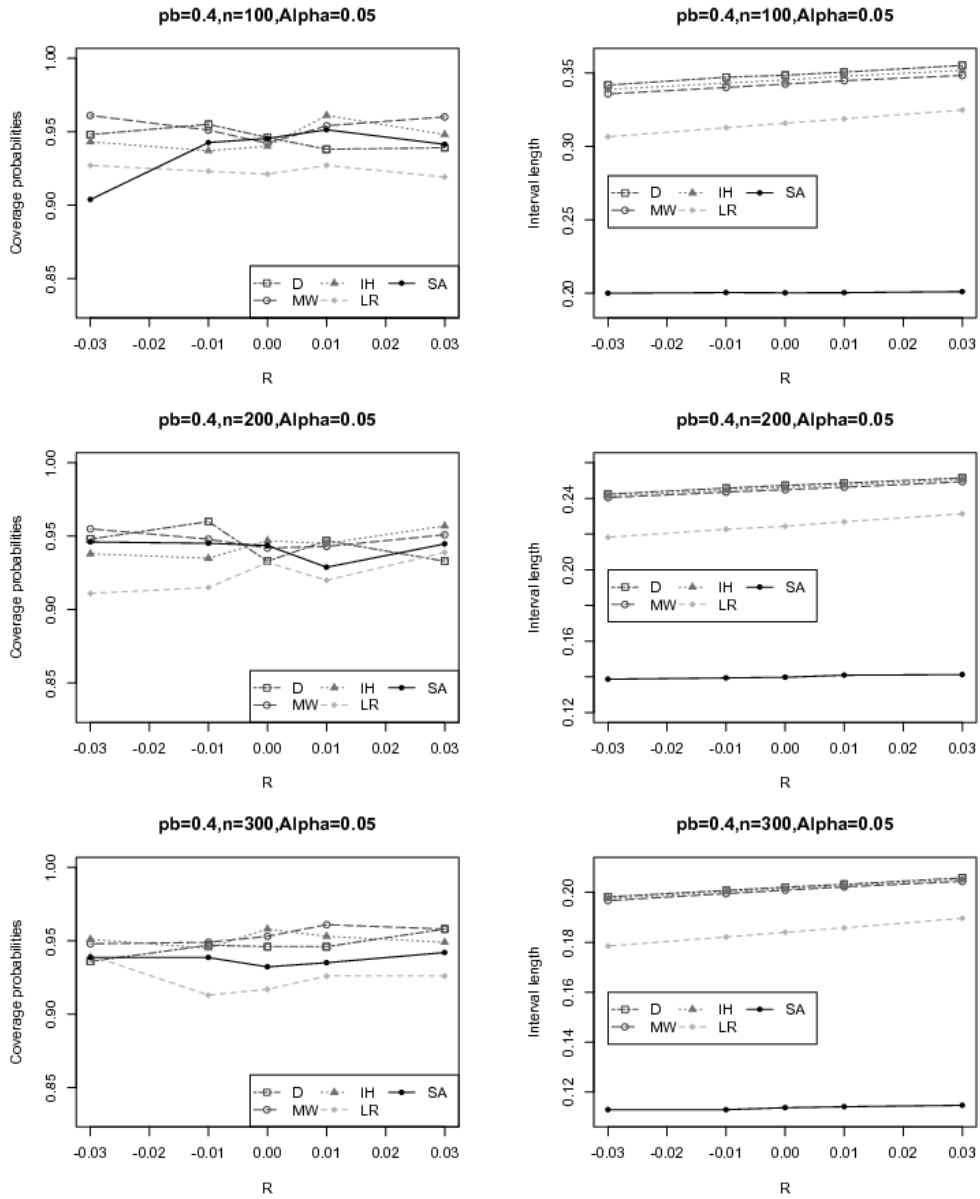


图 4 $p_b = 0.4$, $\alpha = 0.05$, $n = 100, 200, 300$ 下相应的覆盖率(左)及区间长度(右)

现好. 因此, 建议选用SA法作为配对设计中风险差指标的区间估计方法.

§4 实例分析

在研究患心脏病是否影响患糖尿病的过程中^[28], 将144名心脏病患者按年龄和性别与144名没有心脏病的人进行了匹配, 构成144对匹配对(心脏病患者组为病例组, 未患有心脏病组未对照组). 询问他们是否曾经被诊断为患有糖尿病. 询问结果如表2所示, 病例组曾患有糖尿病而对对照组不曾患糖尿病的有37位患者, 病例组未曾患糖尿病而对对照组曾患有糖尿病的有16位患者, 现研究心脏病对患糖尿病是否有影响及影响程度.

表2 患心脏病对患糖尿病的影响研究数据

病例组	对照组		总数
	患糖尿病	不患糖尿病	
患糖尿病	9	37	46
不患糖尿病	16	82	98
总数	25	119	144

通过以上数据可以得到风险差 R 的极大似然估计为 $\hat{R} = 0.15$, 对照组不患病概率为 $p_b = 0.11$, 使用文中的5种方法构建置信水平为0.95的置信区间, 分别计算出不同方法风险差的置信区间上, 下限以及区间长度, 结果如表3所示.

表3 患心脏病对患糖尿病的影响程度

	95%的置信区间下限	95%的置信区间上限	区间长度
Delta法	0.0536	0.2447	0.1912
改进Wald法	0.0508	0.2397	0.1889
反双曲正切变换法	0.0524	0.2431	0.1906
似然比检验法	0.0947	0.2346	0.1302
鞍点逼近法	0.0714	0.2187	0.1486

从表3的不同方法比较结果来看, 虽然似然比检验法的区间长度足够短, 但在模拟中发现其覆盖率离置信水平较远, 这反映出该方法的效果不是很稳定, 而鞍点逼近法在不同情形下覆盖率始终能够接近给定的名义水平, 表现着较好的稳定性. 因此在实际运用中, 建议使用鞍点逼近法. 此外, 这5种方法构造的置信区间都包含了风险差的极大似然估计值, 因此有95%的把握认为患心脏病对患糖尿病有影响, 即患心脏病与不患心脏病的人相比, 心脏病患者患糖尿病的风险更多.

§5 结论

风险差在医学上是一个用于评价治疗风险, 治疗效果的指标. 本文在配对设计中, 考虑了鞍点逼近法, Delta法, 改进Wald法, 反双曲正切变换法以及似然比检验法对风险差进行置信区间构造, 通过Monte Carlo模拟计算区间覆盖率和区间长度, 比较了不同方法的区间性能, 在这5种方法中, 鞍点逼近法要明显优于其他4种方法, 是一种良好的区间构造方法. 相比之下, 改进Wald法的整体性能仅次于鞍点逼近法. 改进Wald法与反双曲正切变化法相比, 随着样本量的增大, 这两种方法的区间长度差距逐渐减小. 此外, 本文所提出的方法还可用于其他分布或其他指标的置信区间构造当中, 例如相对危险度指标等.

参考文献:

- [1] 沈溪, 马长兴. 配对相关数据的统计学方法[J]. 中国科学A辑, 2020, 50(5): 667-678.
- [2] Lui K J. Statistical Estimation of Epidemiological Risk[M]. Chichester: Wiley, 2004.
- [3] Berkson J. Smoking and lung cancer: some observations on two recent reports[J]. Journal of the American Statistical Association, 1958, 53(281): 28-38.
- [4] Fleiss J L, Levin B, Paik M C. Statistical Methods for Rates and Proportions[M]. Hoboken: John Wiley and Sons Inc, 2003.
- [5] Lui K J. Interval estimation of simple difference under independent negative binomial sampling[J]. Biometrical Journal, 2015, 41(1): 83-92.
- [6] Tang Manlai, Tian Maozai. Asymptotic confidence interval construction for risk difference under inverse sampling[J]. Computational Statistics and Data Analysis, 2009, 53(3): 621-631.
- [7] Krishnamoorthy K, Zhang Dan. Approximate and fiducial confidence intervals for the difference between two binomial proportions[J]. Communications in Statistics-Theory and Methods, 2015, 44(8): 1745-1759.
- [8] Krishnamoorthy K, Lee M, Zhang Dan. Closed-form fiducial confidence intervals for some functions of independent binomial parameters with comparisons[J]. Statistical Methods in Medical Research, 2017, 26(1): 43-63.
- [9] 古丽斯坦·库尔班尼牙孜, 孟丽君, 田茂再. 配对设计中条件优势比的置信区间构造[J]. 系统科学与数学, 2021, 41(3): 824-836.
- [10] 王维贤, 田茂再. 鞍点逼近下相关差的置信区间构造[J]. 高校应用数学学报, 2019, 34(3): 264-272.
- [11] Tang Manlai, Qiu Shifang, Poon Waiyin. Confidence interval construction for disease prevalence based on partial validation series[J]. Computational Statistics and Data Analysis, 2012, 56(5): 1200-1220.
- [12] Tang Manlai, Qiu Shifang, Poon Waiyin, et al. Test procedures for disease prevalence with partially validated data[J]. Journal of Biopharmaceutical Statistics, 2012, 22(2): 368-386.
- [13] Qiu Shifang, Guo Lixuan, Zou G. Y, et al. Tests for homogeneity of risk differences in stratified design with correlated bilateral data[J]. Journal of Applied Statistics, 2019, 46(14): 2491-2513.
- [14] 白永昕, 田茂再. Poisson分布下基于鞍点逼近的慢性病风险差的置信区间构造[J]. 高校应用数学学报, 2017, 32(3): 253-266.
- [15] 罗玉波, 田茂再, 吴喜之. 广义卡方型混合分布的鞍点逼近[J]. 统计与信息论坛, 2008, 23(1): 29-31.
- [16] 孟令宾, 李二倩, 田茂再. 基于鞍点逼近的二项抽样下优势比的置信区间构造[J]. 数理统计与管理, 2017, 36(1): 85-102.
- [17] 夏丽丽, 田茂再, 朱钰. 二项分布下基于鞍点逼近的总体成数置信区间的构造[J]. 统计与信息论坛, 2019, 34(9): 3-9.

- [18] 王玉琢, 任耀峰. 一种构造指数分布族下鞍点逼近型置信区间的方法[J]. 统计与决策, 2015(14): 23-26.
- [19] Casella G, Berger R L. Statistical Inference[M]. New York: Springer, 2012.
- [20] May W L, Johnson W D. Confidence intervals for differences in correlated binary proportions[J]. Statistics in Medicine, 1997, 16(18): 2127-2136.
- [21] Quesenberry C P, Hurst D C. Large sample simultaneous confidence intervals for multinomial proportions[J]. Technometrics, 1964, 2(6): 191-195.
- [22] Edwardes M. A confidence interval for $\Pr(X < Y) - \Pr(X > Y)$ estimated from simple cluster samples[J]. Biometrics, 1995, 51(2): 571-578.
- [23] Fulks W. A generalization of Laplace's method[J]. Proceedings of the American Mathematical Society, 1951, 2(4): 349-354.
- [24] Daniels H E. Saddlepoint approximations in statistics[J]. The Annals of Mathematical Statistics, 1954, 25(4): 631-650.
- [25] Barndorff-Nielsen O E, Cox D R. Inference and Asymptotics[M]. London: Chapman and Hall, 1994.
- [26] Billingsley P. Probability and Measure 3rd[M]. New York: Wiley, 1995.
- [27] Jensen J L. Saddlepoint Approximations[M]. Oxford: Clarendon Press, 1995.
- [28] Agresti A. Categorical Data Analysis[M]. New York: Wiley, 2012.

Confidence interval construction of risk difference in paired design

ZHANG Li-Ping¹, GÜLISTAN • Kurbanayaz¹, TIAN Mao-Zai^{1,2}

(1. Xinjiang University of Finance and economics, College of Statistics, Urumqi 830012, China; 2. Renmin University of China, School of statistics, Beijing 100872, China)

Abstract: Risk difference is often used to compare the effectiveness of a disease before and after treatment or the implementation of a policy measure in paired design. Aiming at the problem of confidence interval estimation of risk difference in paired design, this paper uses delta method, improved Wald method, inverse hyperbolic tangent transformation method, likelihood ratio test method and saddle point approximation method to construct confidence interval, and calculates the interval coverage and interval length through Monte Carlo simulation to compare the performance of these five methods. The results show that the interval performance of different methods is different. Among them, the saddle point approximation method is significantly better than the other four methods; The performance of the improved Wald method is second only to the saddle point approximation method, and with the increase of the sample size, the performance gap between the improved Wald method and the inverse hyperbolic tangent transformation method gradually decreases; finally an example is given to verify the five interval estimation methods.

Keywords: matching design; saddle point approximation; Monte Carlo simulation; confidence interval

MR Subject Classification: 62F35; 63G35; 62J05