doi:10.16055/j.issn.1672-058X.2020.0006.007

基于分层部分核实数据对二项比例的齐性检验*

李天骄
(重庆理工大学 理学院,重庆 400054)

摘　要:针对分层情况下的疾病流行率问题,考虑了金标准存在时,在各层的敏感度和特异度不同时二项比例(疾病流行率)的齐性检验;提出了基于渐近检验过程的 7 种统计量和基于 Bootstrap 重抽样检验过程的 4 种统计量,并通过蒙特卡罗模拟研究来比较了各种检验的犯第一类错误的概率和检验功效;研究表明:score 统计量、似然比统计量和 4 种基于 bootstrap 重抽样的检验统计量具有良好的统计性质,推荐用于实际中;最后,实际数据进一步验证了方法的有效性。

关键词:金标准;齐性检验;Bootstrap 重抽样;分层部分核实数据

中图分类号:O212.1　　**文献标志码**:A　　**文章编号**:1672-058X(2020)06-0039-09

0 引 言

疾病患病率(二项比例)的估计是临床试验和医学研究中的一个重要课题,可以从特定的群体中抽取受试者进行诊断并分类为患有或不患有疾病,以此来估计疾病的患病率。筛检方法通常用于诊断过程的第一阶段,其优点是价格相对便宜、能够快速获得结果,并且对受试者通常是无害的。然而,筛检方法的结果往往是存在误判的,使用这些被错误分类的数据研究疾病流行率会导致估计有偏差[1]。另一方面,完全无误判的检测方法(金标准)往往价格昂贵且耗时,因而不能对所有个体使用。Tenenbein[2]为了解决这个问题,提出了二重抽样的方法。在二重抽样中,所有的受试者都将接受筛检方法的检测,但只有其中 n 个人接受了金标准的检测,因而有部分个体只接受了筛检方法的检测。Tang 等[3]将通过这样的方法得到的数据称之为部分核实数据。

二重抽样的设计和分析一直是统计研究中的重要课题。国内外已有大量学者进行了研究。例如,Espland 和 Odoroff[4]提出了对数线性模型,并通过 EM 算法的最大似然估计导出了参数估计和参数函数的协方差的表达式;Geng 和 Asano[5]提出了用于二重抽样下有误判的分类数据的贝叶斯估计方法;Alonzo[6]对各种疾病患病率的形式和属性进行了对比,并提出了半参数有效方法是二阶段研究中患病率估计的首选方法;Tang 等[7]提出了 12 种构建患病率置信区间的方法;Qiu 等[8]分别考虑了根据显著性检验和置信区间宽度的样本量确定;邱等[9]基于两种模型提出了在给定置信水平下的样本量的近似公式;邱和何[10]在无金标准情况下对 2 组疾病流行率进行了比较研究;Lui[11]提出了 3 个简单的统计量,用于在有患者不配合的情况下的分层随机临床试验中风险比的齐性检验。然而在大多数情况下,当年龄不同或生活环境等其他因素不相同时,人群中的患病率总是不相同的。这时就需要对受试者进行分层归类,探究分层情况下的疾病流行率问题。现研究在金标准存在时,基于分层设计下的部分核实数据,考虑了各层的敏感度和特异

收稿日期:2019-12-03;**修回日期**:2020-01-28.
* **基金项目**:重庆市基础研究与前沿探索项目(CSTC2018JCYJAX0241);重庆理工大学研究生创新项目(YCX20192082).
作者简介:李天骄(1986—),男,重庆巴南人,硕士研究生,从事生物医学统计研究.

度不同时二项比例(疾病流行率)的齐性检验过程。

1 统计模型及检验统计量

1.1 模型与参数估计

假设有 N_j 个个体是从第 j 个总体中随机抽取的,这 N_j 个个体每个都接受筛选检测,检测的结果为 T_j。设 $T_j=0$ 表示个体检测呈阴性,否则,$T_j=1$;从 N_j 个个体中随机抽取 $n_j(n_j<N_j)$ 个个体进行金标准检测,结果记为 D_j,$D_j=0$ 表示个体检测呈阴性,否则,$D_j=1$。数据结构如表 1 所示。

表 1 第 j 组的数据结构

Table 1 Data structure of Jth group

		筛检方法		
		$T_j=1$	$T_j=0$	合 计
金标准	$D_j=1$	$n_{11j}(\eta_j\pi_j)$	$n_{10j}(\pi_j(1-\eta_j))$	$n_{1+j}(\pi_j)$
	$D_j=0$	$n_{01j}((1-\theta_j)(1-\pi_j))$	$n_{00j}(\theta_j(1-\pi_j))$	$n_{0+j}(1-\pi_j)$
合 计		n_{+1j}	n_{+0j}	$n_j(1.0)$
只接受筛检的个体	$x_j(p_j)$	$y_j(1-p_j)$	N_j-n_j	
合 计	$n_{+1j}+x_j$	$n_{+0j}+y_j$	N_j	

令 $\pi_j=P(D_j=1)$ 表示第 j 个总体的疾病流行率,$\eta_j=P(T_j=1|D_j=1)$ 和 $\theta_j=P(T_j=0|D_j=0)$ 分别表示第 j 个总体的筛检检测的敏感度和特异度。p_j 表示第 j 个总体中用筛检方法判断为阳性的概率,显然,$p_j=\eta_j\pi_j+(1-\theta_j)(1-\pi_j)$,$1-p_j=\pi_j(1-\eta_j)+\theta_j(1-\pi_j)$。

对如下的假设检验感兴趣:

$$H_0:\pi_1=\pi_2=\cdots\pi_j=\pi\leftrightarrow H_1:\pi_1,\cdots,\pi_j$$

不全相等。令:

$$m=\{(n_{11j},n_{10j},n_{01j},n_{00j},x_j,y_j)',j=1,\cdots,J\},$$

$$\pi=(\pi_1,\cdots,\pi_J)',\eta=(\eta_1,\cdots,\eta_j)',\theta=(\theta_1,\cdots,\theta_j)',$$

则对数似然函数为

$$l_1=\sum_{j=1}^{J}[(n_{11j}+n_{10j})\log\pi_j+(n_{01}+n_{00j})\log(1+\pi_j)+n_{11j}\log\eta_j+n_{10j}\log(1-\eta_j)+n_{00j}\log\theta_j+n_{01j}\log(1-\theta_j)+x_j\log p_j+y_j\log(1-p_j)] \quad (1)$$

以上对数似然函数分别对 $\pi_j,\eta_j,\theta_j(j=1,\cdots,J)$ 求偏导,并令式子等于零,得到非限制性极大似然估计:

$$\hat{\pi}_j=\frac{n_{11j}}{n_{+1j}}\cdot\frac{x_j+n_{+1j}}{N_j}+\frac{n_{10j}}{n_{+0j}}\cdot\frac{y_j+n_{+0j}}{N_j},j=1,2,\cdots,J$$

$$\hat{\eta}_j=\frac{n_{11j}}{n_{+1j}}\cdot\frac{x_j+n_{+1j}}{N_j}/\hat{\pi}_j,j=1,2,\cdots,J$$

$$\hat{\theta}_j=\frac{n_{00j}}{n_{+0j}}\cdot\frac{y_j+n_{+0j}}{N_j}/(1-\hat{\pi}_j),j=1,2,\cdots,J \quad (2)$$

通过 Delta 方法可求得 $\hat{\pi}_j$ 的方差为

$$\mathrm{Var}(\hat{\pi}_j)=\frac{\pi_j(1-\pi_j)}{n_j}\left[1-\frac{\pi_j(1-\pi_j)}{p_j(1-p_j)}(\eta_j+\theta_j-1)^2\right]+\frac{\pi_j^2(1-\pi_j)}{N_jp_j(1-p_j)}(\eta_j+\theta_j-1)^2 \quad (3)$$

在原假设 $H_0:\pi_1=\pi_2=\cdots\pi_j=\pi$ 下,对数似然函数为

$$l_2=\sum_{j=1}^{J}[(n_{11j}+n_{10j})\log\pi+(n_{01j}+n_{00j})\log(1-\pi)+n_{11j}\log\eta_j+n_{10j}\mathrm{loh}(1-\eta_j)+n_{00j}\log\theta_j+n_{01j}\log(1-\theta_j)+x_j\log p_j^{\ 1}+y_j\log(1-p_j^{\ 1})]$$

其中,$p_j^{\ 1}=\eta_j\pi+(1-\theta_j)(1-\pi)$,以上对数似然函数分别对 $\pi,\eta_j,\theta_j(j=1,\cdots,J)$ 求偏导并令式子等于零,可得 π 的限制性极大似然估计为

$$\tilde{\pi}=\frac{n_{11+}}{n_{+1+}}\cdot\frac{x_++n_{+1+}}{N_+}+\frac{n_{10+}}{n_{+0+}}\cdot\frac{y_++n_{+0+}}{N_+}$$

其中,$n_{+k+}=\sum_{j=1}^{J}n_{0kj}+n_{1kj},n_{1k+}=\sum_{j=1}^{J}n_{1kj}(k=0,1)$,$x_+=\sum_{j=1}^{J}x_j,y_+=\sum_{j=1}^{J}y_j,N_+=\sum_{j=1}^{J}N_j$,$\eta_j,\theta_j(j=1,2,\cdots,J)$ 的限制性极大似然估计 $\tilde{\eta}_j,\tilde{\theta}_j(j=1,2,\cdots,J)$ 可通过解以下方程组得到:

$$\left.\frac{\partial L_2}{\partial\eta_j}\right|_{\pi_j=\tilde{\pi}}=0,\left.\frac{\partial L_2}{\partial\theta_j}\right|_{\pi_j=\tilde{\pi}}=0,(j=1,2,\cdots,J)$$

此方程组没有显式解,可通过迭代法如牛顿迭代法求解。

1.2　检验统计量

对于齐性检验 $H_0:\pi_1=\pi_2=\cdots\pi_j=\pi\leftrightarrow H_1:\pi_1,\cdots,\pi_j$ 至少有一个不相等。考虑了以下统计量:

1.2.1　加权最小二乘统计量

根据 Fleiss 等[12],可得到 π 的加权最小二乘估计量为:$\hat{\pi}_{\text{wls}}=\sum_{j=1}^{J}\hat{w}_j\hat{\pi}_j/\sum_{j=1}^{J}\hat{w}_j$,其中 $\hat{w}_j=1/\hat{\text{var}}(\hat{\pi}_j)$,$\hat{\text{var}}(\hat{\pi}_j)$ 是将式(2)中各参数的极大似然估计分别代入式(3)所得到的结果,对于所考虑的假设检验的加权最小二乘统计量为

$$T_{\text{wls}}=\sum_{j=1}^{J}\hat{w}_j(\hat{\pi}_j-\hat{\pi}_{\text{wls}})^2=\sum_{j=1}^{J}\hat{w}_j\hat{\pi}_j^{\,2}-\Big(\sum_{j=1}^{J}\hat{w}_j\hat{\pi}_j\Big)^2/\sum_{j=1}^{J}\hat{w}_j$$

根据 Fleiss 等[12]可得,当原假设成立,在 N_j 趋于无穷大时,T_{wls} 渐近服从自由度为 $J-1$ 的卡方分布。

1.2.2　T_{wls}的对数变换统计量

根据 Fisher 等[13],对数变换后的 T_{wls} 检验统计量将更加接近正态分布。因此,基于 T_{wls} 统计量考虑如下的检验统计量:

$$T_{\text{lwls}}=\{\log(T_{\text{wls}}/(J-1))/2+1/[2(J-1)]\}/\sqrt{1/[2(J-1)]}$$

当原假设成立,在 N_j 趋于无穷大时,统计量渐近服从标准正态分布。即当 $T_{\text{lwls}}\geqslant z_{1-\alpha}$,则拒绝原假设,其中 $z_{1-\alpha}$ 为标准正态分布的上 α 分位数。

1.2.3　基于对数变换的检验统计量

令 $b_j=\log(\pi_j)$,则 $\hat{b}_j=\log(\hat{\pi}_j)$ 为 b_j 的非限制性极大似然估计。由 delta 方法可得 $\hat{b}_j$ 的方差为 $\text{var}(\hat{b}_j)=\text{var}(\hat{\pi}_j)/\pi_j^2$,则基于对数变换的统计量为

$$T_{\log}=\sum_{j=1}^{J}\bar{w}_j(\hat{b}_j-\hat{b}_{wls})^2=\sum_{j=1}^{J}\bar{w}_j\hat{b}_j^{\,2}-\Big(\sum_{j=1}^{J}\bar{w}_j\hat{b}_j\Big)^2/\sum_{j=1}^{J}\bar{w}_j$$

其中,$\tilde{w}_j=1/\hat{\text{var}}(\hat{b}_j)=\hat{\pi}_j^2/\hat{\text{var}}(\hat{\pi}_j)$,$\hat{b}_{\text{wls}}=\sum_{j=1}^{J}\bar{w}_j\hat{b}_j/\sum_{j=1}^{J}\bar{w}_j$。当原假设成立,在 N_j 趋于无穷大时,$T_{\log}$ 渐近服从自由度为 $J-1$ 的卡方分布。

1.2.4　基于 logit 变换的检验统计量

令 $c_j=\text{log}it(\pi_j)=\log(\pi_j/(1-\pi_j))$,则 $\hat{c}_j=\text{log}it(\hat{\pi}_j)=\log(\hat{\pi}_j/(1-\hat{\pi}_j))$ 为 c_j 的非限制性极大似然估计。由 delta 方法可得 $\text{var}(\hat{c}_j)=\text{var}(\text{log}it(\hat{\pi}_j))=\text{var}(\hat{\pi}_j)/(\pi_j(1-\pi_j))^2$,则基于 logit 变换的统计量为

$$T_{\text{log}\,it}=\sum_{j=1}^{J}\tilde{w}_j(\hat{c}_j-\hat{c}_{\text{wls}})^2=\sum_{j=1}^{J}\tilde{w}_j\hat{c}_j^{\,2}-\Big(\sum_{j=1}^{J}\tilde{w}_j\hat{c}_j\Big)^2/\sum_{j=1}^{J}\tilde{w}_j$$

其中

$$\tilde{w}_j=1/\hat{\text{var}}(\hat{c}_j)=(\pi_j(1-\pi_j))^2/\text{var}(\hat{\pi}_j)$$

$$\hat{c}_{\text{wls}}=\sum_{j=1}^{J}\tilde{w}_j\hat{c}_j/\sum_{j=1}^{J}\tilde{w}_j$$

当原假设成立,在 N_j 趋于无穷大时,T_{logit} 渐近服从自由度为 $J-1$ 的卡方分布。

1.2.5　基于双对数变换的检验统计量

令 $d_j=\log(-\log(\pi_j))$,则 $\hat{d}_j=\log(-\log(\hat{\pi}_j))$ 为 d_j 的非限制性极大似然估计,由 delta 方法可得 $\text{var}(\hat{d}_j)=\text{var}(\log(-\log(\hat{\pi}_j)))=\text{var}(\hat{\pi}_j)/(\pi_j(\log(\pi_j)))^2$,则基于双对数变换的统计量为

$$T_{d\log}=\sum_{j=1}^{J}\widehat{w}_j(\hat{d}_j-\hat{d}_{wls})^2=\sum_{j=1}^{J}\widehat{w}_j\hat{d}_j^{\,2}-\Big(\sum_{j=1}^{J}\widehat{w}_j\hat{d}_j\Big)^2/\sum_{j=1}^{J}\widehat{w}_j$$

其中,

$$\widehat{w}_j=1/\hat{\text{var}}(\hat{d}_j)=(\pi_j(\log(\pi_j)))^2/\text{var}(\hat{\pi}_j)$$

$$\hat{d}_{\text{wls}}=\sum_{j=1}^{J}\hat{w}_j\hat{d}_j/\sum_{j=1}^{J}\hat{w}_j$$

当原假设成立,在 N_j 趋于无穷大时,$T_{d\log}$渐近服从自由度为 $J-1$ 的卡方分布。

1.2.6　Score 检验统计量

根据 Rao[14] 的 score 检验的一般理论,由对数似然函数 l_1 可得到原假设 H_0 下第 j 层的 Fisher 信息阵为

$$I_{11j}^0=\frac{n_j}{\pi(1-\pi)}+\frac{(N_j-n_j)(1-\eta_j-\theta_j)^2}{p_j^{\,1}(1-p_j^{\,1})}$$

$$I_{12j}^0=\frac{(N_j-n_j)\theta_j}{1-p_j^{\,1}}-\frac{(N_j-n_j)(1-\theta_j)}{p_j^{\,1}}$$

$$I_{13j}^0=\frac{(N_j-n_j)(1-\eta_j)}{1-p_j^{\,1}}-\frac{(N_j-n_j)\eta_j}{p_j^{\,1}}$$

$$I_{22j}^0=\frac{n_j\pi}{\eta_j(1-\eta_j)}+\frac{(N_j-n_j)\pi^2}{p_j^{\,1}(1-p_j^{\,1})}$$

$$I_{23j}^0=-\frac{(N_j-n_j)\pi(1-\pi)}{p_j^{\,1}(1-p_j^{\,1})}$$

$$I_{33j}^0=\frac{n_j(1-\pi)}{\theta_j(1-\theta_j)}+\frac{(N_j-n_j)(1-\pi)^2}{p_j^{\,1}(1-p_j^{\,1})}$$

其中,$p_j^{\,1}=\eta_j\pi+(1-\theta_j)(1-\pi)$,$1-p_j^{\,1}=\pi(1-\eta_j)+\theta_j(1-\pi)$

根据 Tang 等[3],第 j 层的 score 统计量为

$$T_{\text{sc}j}=\frac{(n_{1+j}-n_j\tilde{\pi})\tilde{p}_j(1-\tilde{p}_j)-(1-\tilde{\eta}_j-\tilde{\theta}_j)[x_j-(N_j-n_j)\tilde{p}_j]\tilde{\pi}(1-\tilde{\pi})}{\tilde{\pi}(1-\tilde{\pi})\tilde{p}_j(1-\tilde{p}_j)\sqrt{A-B/C}}$$

其中:$\tilde{p}_j=\tilde{\eta}_j\tilde{\pi}+(1-\tilde{\theta}_j)(1-\tilde{\pi})$

$$A=I_{11j}^0\big|_{\eta_j=\tilde{\eta}_j,\theta_j=\tilde{\theta}_j}$$

$$B=\{(I_{12j}^0)^2I_{33j}^0-2I_{12j}^0I_{13j}^0I_{23j}^0+I_{22j}^0(I_{13j}^0)^2\}\big|_{\eta_j=\tilde{\eta}_j,\theta_j=\tilde{\theta}_j}$$

$$C=\{I_{22j}^0I_{33j}^0-(I_{23j}^0)^2\}\big|_{\eta_j=\tilde{\eta}_j,\theta_j=\tilde{\theta}_j}$$

检验 $H_0:\pi_1=\pi_2=\cdots\pi_j$ 的 Score 统计量为:$T_{sc}=\sum_{j=1}^{J}T_{scj}^2(j=1,\cdots,J)$。当原假设成立,在 N_j 趋于无穷大时,T_{sc}渐近服从自由度为 $J-1$ 的卡方分布。

1.2.7 似然比检验统计量

根据式(1)和式(2),对于假设检验 $H_0:\pi_1=\pi_2=\cdots\pi_j=\pi\leftrightarrow H_1:\pi_1,\cdots,\pi_j$ 不全相等的似然比检验统计量为:$T_l=2[l_1(m;\pi,\eta,\theta)-l_2(m;\pi,\eta,\theta)]$。当原假设成立,在 N_j 趋于无穷大时,T_l 渐近服从自由度为 $J-1$ 的卡方分布。

1.3 检验过程

1.3.1 渐近的检验过程

设 t_i 是 $T_i(i=\text{wls},\text{lwls},\log,\text{log}it,d\log,\text{sc},l)$ 的观测值,在显著性水平 α 下,如果 $t_i(i=\text{wls},\log,\text{log}it,d\log,\text{sc},l)\geqslant\chi_{1-\alpha}^2(J-1)$,或者 $t_{\text{lwls}}\geqslant z_{1-\alpha}$,则拒绝原假设 H_0。其中,$\chi_{1-\alpha}^2(J-1)$表示自由度为 $J-1$ 的卡方分布的 $1-\alpha$ 分位数,$z_{1-\alpha}$表示标准正态分布的 $1-\alpha$ 分位数。

1.3.2 Bootstrap 重抽样检验过程

渐近的检验过程的模拟结果表明,小样本下除了 score 检验和似然比检验的其他检验统计量的表现并不是很好,此时,bootstrap 重抽样检验过程是一个有效的选择。因而,基于 T_{wls}, $T_{\log}$, $T_{\text{log}it}$ 和 $T_{d\log}$考虑了如下步骤的基于 bootstrap 重抽样的检验过程:

Step 1 计算观测数据 $m=\{(n_{11j},n_{10j},n_{01j},n_{00j},x_j,y_j)',j=1,\cdots,J\}$ 下参数 π_j,η_j,θ_j 的限制性极大似然估计 $\tilde{\pi},\tilde{\eta}_j,\tilde{\theta}_j(j=1,\cdots,J)$ 以及各检验统计量 T_q 的观测值 $t_q(q=\text{wls},\log,\text{log}it,d\log)$,其中 $n_j=n_{11j}+n_{10j}+n_{01j}+n_{00j}$, $N_j=n_j+x_j+y_j$。

Step 2 基于 $\tilde{\pi},\tilde{\eta}_j,\tilde{\theta}_j$,得到 $p_{11j}^*=\tilde{\pi}\tilde{\eta}_j$, $p_{10j}^*=\tilde{\pi}(1-\tilde{\eta}_j)$, $p_{01j}^*=(1-\tilde{\pi})(1-\tilde{\theta}_j)$, $p_{00j}^*=(1-\tilde{\pi})\tilde{\theta}_j$。产生 B 个 bootstrap 重抽样样本$\{m^*(b)\}\big|_{b=1}^B$,其中 $m^*(b)=\{(n_{11j}^*,n_{10j}^*,n_{01j}^*,n_{00j}^*,x_j^*,y_j^*)',j=1,\cdots,J\}^{(b)}$,$(n_{11j}^*,n_{10j}^*,n_{01j}^*,n_{00j}^*)$服从多项分布 $M(n_j;p_{11j}^*,p_{10j}^*,p_{01j}^*,p_{00j}^*)$,$(x_j^*,y_j^*)$服从二项分布 $B(N_j-n_j;p_{11j}^*+p_{01j}^*)$。

Step 3 计算 Bootstrap 重抽样样本下检验统计量的值 $t_q^{*(b)}(q=\text{wls},\log,\text{log}it,d\log)$,则

基于 Bootstrap 重抽样方法的检验 p 值为

$$p_q=\sum_{b=1}^{B}I(t_q^{*(b)}\geqslant t_q)/B$$

$(q=wls,\log,\log\ it,d\log)$,其中,$I(*)$为示性函数。在显著性水平 α 下,如果 $p_q<\alpha$,则拒绝原假设$(q=\text{wls},\log,\log\ it,d\log)$。

将基于 Bootstrap 重抽样的 4 种检验分别记做 T_{wls}^B, $T_{\log}^B$, $T_{\text{log}it}^B$ 和 $T_{d\log}^B$。

2 模拟研究

为了检验提出的各种检验的有效性,通过蒙特卡罗模拟方法对各种检验过程进行评价。考虑了$J=3$ 的模型和如下的样本量:平衡设计样本量(Ⅰ)小样本$(n_1,n_2,n_3,N_1,N_2,N_3)=(30,30,30,50,50,50)$;(Ⅱ)中等样本$(n_1,n_2,n_3,N_1,N_2,N_3)=(50,50,50,100,100,100)$;(Ⅲ)大样本$(n_1,n_2,n_3,N_1,N_2,N_3)=(200,200,200,500,500,500)$,(Ⅳ)非平衡样本量$(n_1,n_2,n_3,N_1,N_2,N_3)=(40,50,60,100,80,100)$以及如下的参数设置:$\pi=0.1,0.3,0.5$, $\eta_1=0.5,0.6$, $\theta_1=0.7,0.8$, $\eta_2=\eta_1+0.05$, $\eta_3=\eta_1+0.1$, $\theta_2=\theta_1+0.05$, $\theta_3=\theta_1+0.1$,即考虑了 3(π 的值)×2(η_j 的值)×2(θ_j 的值)= 12 种参数组合。对于每个检验统计量的经验功效,进行了如下的参数设置来考察:$\pi_1=0.1,0.2,0.3$, $\pi_2=\pi_1+0.1$, $\pi_3=\pi_1+0.2$, η_j,θ_j 的设置与考察犯第一类错误的概率时相同,即考虑了 3×2×2=12 种参数组合。对于各种参数组合和不同的样本量,随机产生 5 000 组样本$\{(n_{11j},n_{10j},n_{01j},n_{00j},x_j,y_j):j=1,2,3\}$来计算各种统计量的观测值,其中,$\{(n_{11j},n_{10j},n_{01j},n_{00j}):j=1,2,3\}$从$\{n_j:\pi_j\eta_j,\pi_j(1-\eta_j),(1-\theta_j)(1-\pi_j),\theta_j(1-\pi_j)\}$中产生,$\{(x_j,y_j):j=1,2,3\}$从$(N_j-n_j:p_j,1-p_j)$中产生。对于 Bootstrap 检验过程,产生了 $B=1\ 000$ 个 bootstrap 重抽样样本$\{m^*(b)\}\big|_{b=1}^B$。

在显著性水平 $\alpha=0.05$ 下,每个检验统计量犯第一类错误的概率为:检验统计量拒绝原假设的次数/5 000(H_0 成立下);经验功效的计算公式为:检

验统计量拒绝原假设的次数/5 000（H_1 成立下）。犯第一类错误的模拟结果见表 2—表 5。对于功效由于篇幅的限制只列出了中等样本（平衡与非平衡设计）的模拟结果，见表 6—表 7。

表 2　在小样本下各种检验犯第一类错误的概率（显著性水平 α=0.05）　%

Table 2　Actual type I error rates of the various procedures in small-sample performance（notability level α=0.05）　%

π	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\mathrm{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\mathrm{log}it}$	$T^B_{d\log}$
0.10	0.50	0.70	4.72	3.96	1.10	1.46	3.04	3.76	2.24	4.68	5.46	5.52	5.06
		0.80	5.62	4.54	1.18	1.66	3.38	4.30	2.54	5.60	6.18	5.98	5.78
	0.60	0.70	5.20	4.36	1.36	1.62	3.24	4.04	2.54	5.26	6.16	6.02	5.38
		0.80	5.70	4.90	1.40	1.84	3.38	4.08	2.38	5.42	5.58	5.68	5.62
0.30	0.50	0.70	7.22	6.60	4.36	5.26	6.28	5.22	5.56	4.90	5.16	5.06	4.98
		0.80	7.00	6.24	4.12	4.90	6.04	5.04	5.26	5.40	5.64	5.58	5.54
	0.60	0.70	7.52	6.72	4.56	5.34	6.50	5.36	5.68	4.94	5.18	5.36	5.10
		0.80	7.30	6.48	4.70	5.32	6.30	5.74	5.78	5.18	5.40	5.28	5.16
0.50	0.50	0.70	6.64	5.92	5.26	5.46	5.42	4.84	5.18	4.98	5.20	5.02	4.88
		0.80	7.32	6.58	5.98	6.24	6.08	5.54	5.86	4.20	3.90	4.24	4.44
	0.60	0.70	6.60	5.86	5.20	5.42	5.60	4.90	5.34	4.84	4.74	4.80	5.06
		0.80	6.40	5.62	4.80	4.94	5.10	4.84	4.82	4.88	4.76	5.02	4.94

表 3　在中等样本下各种检验犯第一类错误的概率（显著性水平 α=0.05）　%

Table 3　Actual type I error rates of the various procedures in moderate-sample performance（notability level α=0.05）%

π	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\mathrm{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\mathrm{log}it}$	$T^B_{d\log}$
0.10	0.50	0.70	7.48	6.54	2.42	2.72	5.36	5.08	4.18	5.64	5.38	5.38	5.72
		0.80	7.46	6.60	2.24	2.54	5.08	4.74	3.96	4.96	5.12	5.12	4.92
	0.60	0.70	7.44	6.60	2.42	2.60	4.96	5.00	3.96	5.26	4.88	4.82	5.00
		0.80	6.94	6.24	2.24	2.56	4.70	4.74	3.30	5.66	5.80	5.64	6.06
0.30	0.50	0.70	7.10	6.38	5.44	6.04	6.52	5.68	6.02	4.48	4.72	4.64	4.58
		0.80	5.66	5.06	4.40	4.80	5.22	4.70	5.02	4.74	4.54	4.86	4.74
	0.60	0.70	6.26	5.64	4.80	5.24	5.86	5.12	5.44	4.74	4.72	4.72	4.74
		0.80	6.14	5.54	4.66	5.16	5.62	4.92	5.30	4.68	5.08	5.20	4.80
0.50	0.50	0.70	6.02	5.18	5.04	5.40	5.18	4.72	5.12	4.66	4.64	4.54	4.70
		0.80	6.02	5.48	5.16	5.40	5.40	5.06	5.34	5.54	5.38	5.66	5.58
	0.60	0.70	6.32	5.56	5.44	5.58	5.68	5.26	5.60	5.20	5.14	5.08	5.12
		0.80	5.94	5.48	5.24	5.44	5.42	5.08	5.36	5.16	5.28	5.16	5.30

表 4 在大样本下各种检验犯第一类错误的概率(显著性水平 α=0.05) %

Table 4 Actual type I error rates of the various procedures in large-sample performance(notability level α=0.05) %

π	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\text{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\text{log}it}$	$T^B_{d\log}$
0.10	0.50	0.70	5.46	4.78	4.58	4.76	5.26	5.16	5.36	4.88	4.74	4.86	5.00
		0.80	5.88	5.32	4.58	4.84	5.56	5.34	5.74	4.62	4.78	4.80	4.58
	0.60	0.70	5.26	4.66	4.12	4.20	4.92	4.62	5.00	4.84	4.54	4.46	4.56
		0.80	5.66	4.84	4.50	4.70	4.96	5.02	5.16	4.64	4.58	4.54	4.56
0.30	0.50	0.70	5.76	5.10	5.28	5.52	5.64	5.50	5.62	4.80	4.66	4.74	4.88
		0.80	5.74	4.94	5.42	5.40	5.58	5.40	5.56	4.60	4.54	4.64	4.66
	0.60	0.70	5.48	4.84	5.30	5.38	5.36	5.26	5.30	5.96	5.66	5.78	5.98
		0.80	5.18	4.54	4.80	4.90	5.00	4.88	4.94	4.76	4.92	4.88	4.78
0.50	0.50	0.70	5.44	4.56	5.06	5.30	5.22	5.08	5.24	4.86	4.90	4.82	4.76
		0.80	5.56	4.94	5.42	5.44	5.42	5.32	5.48	5.14	5.04	5.14	5.06
	0.60	0.70	4.82	4.00	4.46	4.70	4.58	4.52	4.72	5.46	5.48	5.54	5.46
		0.80	5.94	5.30	5.98	5.84	5.80	5.74	5.82	5.12	4.96	5.06	5.00

表 5 在不平衡样本下各种检验犯第一类错误的概率(显著性水平 α=0.05) %

Table 5 Actual type I error rates of the various procedures in unbalanced-sample performance(notability level α=0.05) %

π	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\text{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\text{log}it}$	$T^B_{d\log}$
0.10	0.50	0.70	7.44	6.62	2.42	2.62	5.06	4.46	3.80	5.24	5.28	5.22	5.24
		0.80	7.90	7.20	2.76	2.88	5.42	5.12	4.16	6.34	5.66	5.68	6.00
	0.60	0.70	7.30	6.50	2.26	2.48	4.74	4.70	3.74	5.28	5.28	5.26	5.36
		0.80	7.64	6.78	2.18	2.54	4.88	4.68	3.68	6.00	5.06	5.16	5.56
0.30	0.50	0.70	6.74	5.98	5.04	5.44	6.12	5.26	5.78	4.66	5.20	5.10	4.84
		0.80	6.38	5.80	4.80	5.02	5.86	4.82	5.48	5.04	5.24	5.24	5.16
	0.60	0.70	6.94	6.08	5.26	5.60	6.14	5.46	5.78	5.02	5.16	5.08	5.10
		0.80	6.28	5.34	4.74	5.02	5.66	5.00	5.34	4.56	5.06	4.84	4.58
0.50	0.50	0.70	6.64	5.66	5.66	5.64	5.74	5.12	5.38	5.16	5.08	5.12	5.30
		0.80	6.00	5.32	5.14	5.34	5.46	4.92	5.22	5.08	4.88	5.10	5.02
	0.60	0.70	6.38	5.60	5.52	5.66	5.64	5.20	5.52	4.60	4.54	4.68	4.60
		0.80	6.42	5.76	5.62	5.62	5.66	5.26	5.50	4.82	4.72	5.00	5.12

表 6　在中等样本下各种检验的检验功效(显著性水平 $\alpha=0.05$)　%

Table 6　Actual test powers of the various procedures in moderate-sample performance(notability level $\alpha=0.05$)　%

π_1	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\text{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\text{log}it}$	$T^B_{d\log}$
0.10	0.50	0.706	9.68	67.70	59.26	63.04	67.76	63.70	65.94	64.32	62.42	63.48	64.48
		0.80	69.98	67.96	58.26	62.32	67.82	63.18	66.14	65.42	62.92	64.20	65.26
	0.60	0.70	69.96	67.92	58.16	62.30	68.00	63.36	65.94	63.92	62.58	63.46	64.28
		0.80	72.92	71.12	62.60	66.26	71.10	67.28	69.74	67.56	64.12	65.80	67.50
0.20	0.50	0.70	55.78	53.42	49.86	52.24	54.18	51.22	52.60	48.00	47.62	48.48	48.32
		0.80	58.06	55.70	51.10	54.02	56.24	53.22	55.02	51.66	49.88	51.18	51.80
	0.60	0.70	55.64	53.20	49.06	51.32	54.02	50.14	52.10	51.58	50.52	51.56	52.08
		0.80	57.58	55.18	51.04	53.58	55.80	52.96	54.50	54.54	52.60	54.14	54.82
0.30	0.50	0.70	49.20	47.00	45.12	46.92	47.78	45.36	46.44	44.32	43.46	44.34	44.38
		0.80	51.86	49.50	46.60	48.90	50.26	47.50	48.96	46.98	45.20	46.64	46.92
	0.60	0.70	49.36	47.18	44.78	46.84	48.04	45.50	46.64	45.82	44.78	45.48	46.08
		0.80	52.12	49.90	47.22	49.54	50.18	48.96	49.78	49.04	47.58	48.86	49.32

表 7　在不平衡样本下各种检验的检验功效(显著性水平 $\alpha=0.05$)　%

Table 7　Actual test powers of the various procedures in unbalanced-sample performance(notability level $\alpha=0.05$)　%

π_1	η_1	θ_1	T_{wls}	T_{lwls}	$T_{\log}$	$T_{\text{log}it}$	$T_{d\log}$	T_{sc}	T_l	T^B_{wls}	$T^B_{\log}$	$T^B_{\text{log}it}$	$T^B_{d\log}$
0.10	0.50	0.70	68.58	66.60	49.96	55.66	64.62	58.18	61.88	64.08	55.38	59.00	63.04
		0.80	71.32	69.50	52.76	58.66	67.14	61.10	64.40	66.86	56.92	61.34	65.22
	0.60	0.70	69.88	68.10	52.20	57.88	65.88	60.12	63.14	64.12	55.04	58.28	62.94
		0.80	71.92	70.10	53.82	60.16	68.48	63.14	65.92	65.30	55.84	59.56	63.76
0.20	0.50	0.70	53.62	51.40	42.80	47.58	51.62	46.64	49.14	49.54	43.64	47.28	49.28
		0.80	57.18	54.22	46.30	50.58	54.40	49.30	51.92	51.74	45.58	48.88	50.92
	0.60	0.70	55.72	53.32	44.28	48.98	53.40	48.00	50.56	51.62	45.64	48.92	51.10
		0.80	56.76	54.42	45.88	50.76	54.10	50.16	52.34	53.90	48.18	51.32	53.32
0.30	0.50	0.70	48.52	45.84	40.34	44.54	47.02	43.14	44.52	44.80	39.82	43.60	45.36
		0.80	50.82	48.10	42.46	46.50	49.18	45.14	46.84	46.74	42.28	45.60	47.20
	0.60	0.70	48.28	46.12	40.28	44.56	47.14	43.08	44.94	44.62	40.08	43.56	45.46
		0.80	51.68	49.20	43.20	47.52	50.00	46.00	48.26	47.36	42.96	46.66	48.08

模拟结果表明:

(1) 即使在小样本(如$(n_1,n_2,n_3,N_1,N_2,N_3)=(30,30,30,50,50,50)$)下,基于 score 统计量的渐近的检验过程犯第一类错误的概率都很接近给定的显著性水平且功效也较高。

(2) 除了在 $\pi=0.1$ 外,基于似然比统计量的渐近的检验过程也表现较好,其第一类的概率接近给定的显著性水平且具有较高的检验功效。

(3) 随着样本量的增大,基于各种统计量的渐近的检验过程犯第一类错误的概率越来越接近给定的显著性水平。

(4) 即使在小样本量下,4 种 bootstrap 重抽样检验 T^B_{wls},$T^B_{\log}$,$T^B_{\mathrm{log}it}$ 和 $T^B_{d\log}$ 犯第一类错误的概率都很接近显著性水平。因此,实际应用中推荐使用基于 T_{sc} 和 T_l 的渐近的检验过程和 bootstrap 重抽样检验 T^B_{wls},$T^B_{\log}$,$T^B_{\mathrm{log}it}$ 和 $T^B_{d\log}$,当样本量很大时,提出的所有检验统计量都可以使用。

3 实例分析

根据 Nedelman[15] 的文章,世界卫生组织和尼日利亚政府在 1969—1976 年间对尼日利亚的疟疾进行了大规模的流行病学和控制研究。按照受试者的年龄将调查结果分为了 7 个年龄段,选取了其中 3 个成年人组(即 19~28 岁,29~43 岁和 44 岁以上)的数据说明提出的方法。将高级显微镜学家的判断假定为金标准,数据结构如表 8:

表 8 尼日利亚疟疾数据

Table 8 Malaria data in Nigeria

		筛检方法					
		19~28		29~43		≥44	
		+	−	+	−	+	−
金标准	+	59	10	69	27	37	6
	−	1	42	2	91	2	69
未经金标准分类		483	544	715	995	370	560

注:+ 表示阳性,− 表示阴性

感兴趣的问题是:$H_0:\pi_1=\pi_2=\cdots\pi_j \leftrightarrow H_1:\pi_1,\cdots,\pi_j$ 不全相等。对于以上数据,可得到 $\pi_1,\pi_2,\pi_3,\eta_1,\eta_2,\eta_3,\theta_1,\theta_2,\theta_3$ 的非限制性极大似然估计分别为:$\hat{\pi}_1=0.569\ 4$,$\hat{\pi}_2=0.536\ 3$,$\hat{\pi}_3=0.420\ 3$,$\hat{\eta}_1=0.823\ 3$,$\hat{\eta}_2=0.75$,$\hat{\eta}_3=0.884\ 2$,$\hat{\theta}_1=0.981\ 5$,$\hat{\theta}_2=0.974\ 9$,$\hat{\theta}_3=0.965\ 3$;在 H_0 下,$\pi,\eta_1,\eta_2,\eta_3,\theta_1,\theta_2,\theta_3$ 的限制性极大似然估计为:$\tilde{\pi}=0.515\ 0$,$\tilde{\eta}_1=0.875$,$\tilde{\eta}_2=0.77$,$\tilde{\eta}_3=0.775$,$\tilde{\theta}_1=0.97$,$\tilde{\theta}_2=0.97$,$\tilde{\theta}_3=0.98$。所考虑的统计量 T_{wls},T_{lwls},$T_{\log}$,$T_{\mathrm{log}it}$,$T_{d\log}$,T_{sc},T_l,T^B_{wls},$T^B_{\log}$,$T^B_{\mathrm{log}it}$,$T^B_{d\log}$ 的检验 p 值分别为:$3.913\ 3\times10^{-4}$,0.005 2,0.001,$5.813\ 3\times10^{-4}$,$4.602\ 9\times10^{-4}$,0.001 4,$8.590\ 8\times10^{-4}$,6×10^{-4},1.2×10^{-3},5×10^{-4},4×10^{-4}。在显著性水平 $\alpha=0.05$ 下,由于所有检验的 p 值都远远小于显著性水平,所以拒绝原假设,即认为疾病流行率对于不同的年龄层有显著性的差异,这和 Nedelman[15] 的结论是一致的。

4 结 论

研究了基于金标准存在时的分层部分核实数据,在筛检方法对于不同的层敏感度和特异度非齐性时,二项比例(疾病流行率)是否相同的问题,提出了基于渐近检验过程的 7 种统计量和基于 Bootstrap 重抽样检验过程的 4 种统计量,并通过蒙特卡罗模拟研究来比较了各种检验的犯第一类错误的概率和检验功效。结果表明:score 统计量和似然比统计量能很好地控制犯第一类错误的概率且具有较高的经验功效,因而推荐在实际运用中使用。而 4 种 bootstrap 重抽样检验统计量 T^B_{wls},$T^B_{\log}$,$T^B_{\mathrm{log}it}$,$T^B_{d\log}$ 也具有良好的统计性质,因此同样推荐用于实际中。当样本量很大时,提出的所有检验统计量都可以使用。

参考文献(References):

[1] BROSS I. Misclassification in 2×2 Tables[J]. Biometrics, 1954, 10: 478—486

[2] TENENBEIN A. A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications[J]. Journal of the American Statistical Association, 1970, 65: 1350—1361

[3] TANG M L, QIU S F, POON W Y. Test Procedures for Disease Prevalence with Partially Validated Data[J]. Journal of Biopharmaceutical Statistics, 2012, 22(4): 368—386

[4] ESPLANDM A, ODOROFF C L. Log-linear Models for Doubly Sampled Categorical Data Fitted by the EM Algorithm[J]. Journal of the American Statistical Association, 1985, 80(11): 663—670

[5] GENG Z, ASANO C. Bayesian Estimation Methods for Categorical Data with Misclassification[J]. Communications in Statistics-Theory and Methods, 1989, 18(7): 1747—1766

[6] ALONZO T A. Estimating Disease Prevalence in Two-phase Studies[J]. Biostatistics, 2003(4): 313—326

[7] TANG M L, QIU S F, POON W Y. Confidence Interval

Construction for Disease Prevalence Based on Partial Validation Series[J]. Computational Statistics and Data Analysis,2012,56(2):1200—1220

[8] QIU S, POON W, TANG M L. Sample Size Determination for Disease Prevalence Studies with Partially Validated Data[J]. Stat Methods Med Res, 2016,25(4):37—63

[9] 邱世芳,曾小松. 不完全无误判金标准下二重抽样设计中样本量的确定[J]. 重庆理工大学学报(自然科学),2018(1):195—204

QIU S F, ZENG X S. Sample Size Determination in Double-Sampling Design with an Imperfect Gold Standard[J]. Journal of Chongqing University of Technology (Natural Science),2018(1):195—204(in Chinese)

[10] 邱世芳,何杰. 无金标准部分核实数据下基于风险差的等价性检验[J]. 重庆理工大学学报(自然科学),2019,33(10):191—202

QIU S F, HE J. Equivalence Test of Risk Difference for Partially Validated Series with Two Fallible Classifiers[J]. Journal of Chongqing University of Technology (Natural Science), 2019, 33 (10): 191—202 (in Chinese)

[11] LUI K J. Testing Homogeneity of the Risk Ratio in Stratified Noncompliance Randomized Trials[J]. Contemporary Clinical Trials,2007,28(10),614—625

[12] FLEISS J L. Statistical Methods for Rates and Proportions[M]. 2nd edition. New York: Wiley,1981

[13] FISHER R A. On A Distribution Yielding the Error Functions of Several Well-known Statistics[J]. Proceedings of the International Mathematical Congress, 1924(2):805—813

[14] RAO C R. Linear Statistical Inference and Its Applications[M]. 2nd ed. New York:Wiley,1985

[15] NEDELMAN J. The Prevalence of Malaria in Garki, Nigeria: Double Sampling with A Fallible Expert[J]. Biometric,1988,44(2):635—655

Homogeneity Test for Binomial Proportions Based on Stratified Partially Validated Data

LI Tian-jiao

(School of Science, Chongqing University of Technology, Chongqing 400054,China)

Abstract: Aiming at the problem of disease prevalence in stratified situation, the homogeneity test of binomial proportion (disease prevalence) is considered when the sensitivity and specificity of each strata are different in the presence of gold standard. Seven kinds of statistics based on the asymptotic test process and four kinds of statistics based on the bootstrap resampling test process are proposed, and the empirical type I error and the empirical power in various tests are compared by Monte Carlo simulation. The research shows that score statistics, likelihood ratio statistics and four kinds of test statistics based on bootstrap resampling have good statistical properties and are recommended to be used in the practice. Finally, a real data is further used to verify the proposed methods.

Key words: gold standard; homogeneity test; Bootstrap resampling method; stratified partially validated data

责任编辑:田 静

引用本文/Cite this paper:

李天骄. 基于分层部分核实数据对二项比例的齐性检验[J]. 重庆工商大学学报(自然科学版),2020,37(6):39—47

LI T J. HomogeneityTest for Binomial Proportions Based on Stratified Partially Validated Data[J]. Journal of Chongqing Technology and Business University (Natural Science Edition),2020,37(6):39—47