

# 如何正确运用 $\chi^2$ 检验——高维表资料优势比分析与 SAS 实现

胡纯严<sup>1</sup>, 胡良平<sup>1,2\*</sup>

(1. 军事科学院研究生院, 北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会, 北京 100029

\*通信作者: 胡良平, E-mail: lphu927@163.com)

**【摘要】** 本文目的是介绍  $g \times 2 \times 2$  表资料的优势比分析方法和基于 SAS 软件实现计算的方法。内容包括以下几个方面: 其一,  $g \times 2 \times 2$  表资料各层优势比的齐性检验; 其二, 当资料满足齐性要求时, 基于校正的方法对共同优势比进行点估计和置信区间估计; 其三, 当资料不满足齐性要求时, 基于“各层  $2 \times 2$  表中  $(1, 1)$  网格内的频数的条件分布”的方法对共同优势比进行精确点估计、置信区间估计和假设检验。通过两个实例, 并基于 SAS/STAT 中的 FREQ 过程实现全部计算, 对输出结果进行了解释, 并做出了统计和专业结论。

**【关键词】** 齐性检验; 共同优势比; 条件分布; 置信区间; 精确概率

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20210514005

## How to use $\chi^2$ test correctly——the analysis of common odds ratio for the data of a multiway table and the implementation of SAS software

Hu Chunyan<sup>1</sup>, Hu Liangping<sup>1,2\*</sup>

(1. Graduate School, Academy of Military Sciences PLA China, Beijing 100850, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China

\*Corresponding author: Hu Liangping, E-mail: lphu927@163.com)

**【Abstract】** The purpose of this article was to introduce the odds ratio analysis method of  $g \times 2 \times 2$  table data and the calculation method based on SAS software. The contents included the following aspects: firstly, the homogeneity test of the odds ratio of the data in the  $g \times 2 \times 2$  table was performed; secondly, when the data met the homogeneity requirements, point estimation and confidence interval estimation for the common odds ratio based on the correction method were implemented; thirdly, when the data did not meet the homogeneity requirements, based on the approach of "the conditional distribution of the frequency in the  $(1, 1)$  grid in each layer  $2 \times 2$  table", these things, such as the point estimation and confident interval estimation as well as hypothesis testing about the common odds ratio, were accurately estimated and performed. All calculations were realized based on the FREQ procedure in SAS/STAT by means of two real examples. Furthermore, the output results of SAS software were explained, and the statistical and professional conclusions were made.

**【Keywords】** Homogeneity test; Common odds ratio; Conditional distribution; Confidence interval; Exact probability

为了分析来自病例对照研究设计的  $g \times 2 \times 2$  表资料, 需要完成以下 4 项任务: 其一, 检验各层  $2 \times 2$  表资料的优势比是否满足齐性; 其二, 估计共同优势比的数值; 其三, 估计共同优势比的置信区间; 其四, 检验共同优势比是否等于 1。本文将对后三项任务有关内容进行介绍。

## 1 高维表资料共同优势比分析的基本概念

### 1.1 高维表 $g \times 2 \times 2$ 表的表达模式

设高维表  $g \times 2 \times 2$  表的表达模式如下, 见表 1。

表 1 病例对照研究设计下  $g \times 2 \times 2$  表的第  $h$  层  $2 \times 2$  表资料的表达模式

危险因素 接触与否	例数			
	是否患病:	病例组	对照组	合计
接触		$n_{h11}$	$n_{h12}$	$n_{h1\cdot}$
未接触		$n_{h21}$	$n_{h22}$	$n_{h2\cdot}$
合计		$n_{h\cdot 1}$	$n_{h\cdot 2}$	$n_h$

注:  $h=1, 2, \dots, g$

### 1.2 高维表资料共同优势比的含义

在分析“病例对照研究设计”的二维表资料时, 可以很方便地依据公式“ $OR=ad/bc$ ”计算出优势比  $OR$  的数值。然而, 对于“ $g \times 2 \times 2$  表资料”, 却无法直接计算出  $OR$  的数值。从概念上来说, 似乎可以采

取某种举措,将“ $g \times 2 \times 2$ 表资料”降维或压缩成一个“ $2 \times 2$ 表资料”。但事实上,这种理想的“ $2 \times 2$ 表资料”是无法直接呈现出来的。于是,统计学家通过统计学方法来体现出各层“ $2 \times 2$ 表资料”之间的“微小差别”,这就是求出各层“ $2 \times 2$ 表资料”的“权重系数  $w_h$ ”。通过它将各层“ $2 \times 2$ 表资料”进行加权平均,从而间接获得合并后的优势比  $OR$  的数值。在 SAS/STAT 的 FREQ 过程中,将其称为“common odds ratios”<sup>[1]</sup>,常译成“共同或普通或合并优势比”。

### 1.3 估计共同优势比的前提条件

基于“ $g \times 2 \times 2$ 表资料”计算共同优势比的前提条件是 高维表资料应满足齐性,针对“优势比”,检验高维表资料是否满足齐性的检验方法有 5 种,分别是“Breslow-Day 检验”“Breslow-Day-Tarone 检验”“ $Q$  检验”“ $I^2$  度量统计量及其不确定性限值”和“Zelen's 精确检验”。

### 1.4 估计共同优势比及其置信区间的方法概述

SAS/STAT 中的 FREQ 过程<sup>[1]</sup>采用 3 种方法估计共同优势比及置信区间,分别是:①校正的计算方法, Mantel-Haenszel 估计法;②校正的计算方法, logit 估计法;③基于条件分布的精确法。

## 2 高维表资料优势比分析及 SAS 实现

### 2.1 高维表资料优势比分析的具体算法

#### 2.1.1 高维表资料优势比分析的具体内容

高维表资料优势比分析的具体内容包括以下

$$B = \frac{\sum_h [(n_{h11} + n_{h22})(n_{h12}n_{h21}) + (n_{h12} + n_{h21})(n_{h11}n_{h22})] / n_h^2}{2(\sum_h n_{h11}n_{h22} / n_h)(\sum_h n_{h12}n_{h21} / n_h)} \quad (5)$$

$$C = \frac{\sum_h (n_{h12} + n_{h21})(n_{h12}n_{h21}) / n_h^2}{2(\sum_h n_{h12}n_{h21} / n_h)^2} \quad (6)$$

注意:当  $n_h$  较小时, MH 估计量不如 logit 估计量敏感。

#### 2.1.2.2 Logit 估计量

Woolf 于 1955 年提出了 logit 估计量<sup>[1]</sup>,见下式:

$$OR_L = \exp\left[\frac{\sum_h w_h \ln(OR_h)}{\sum_h w_h}\right] \quad (7)$$

式(7)中,  $OR_h$  是第  $h$  层的优势比,其  $100(1-\alpha)\%$  置信区间见下式:

4 项<sup>[1]</sup>:其一,检验资料是否满足齐性要求;其二,估计共同优势比;其三,估计共同优势比的置信区间;其四,检验共同优势比是否等于 1。

### 2.1.2 高维表资料共同优势比的点估计及置信区间估计

#### 2.1.2.1 Mantel-Haenszel 估计量

基于 Mantel-Haenszel 估计量(简称 MH 估计量)估计高维表资料共同优势比,见式(1):

$$OR_{MH} = \frac{\sum_{h=1}^g n_{h11}n_{h22} / n_h}{\sum_{h=1}^g n_{h12}n_{h21} / n_h} \quad (1)$$

在式(1)中,若设  $w_h = \frac{1}{n_h}$ ,则可以看出,式(1)的

分子与分母分别是各层  $2 \times 2$  表时的分子与分母的加权和。

设  $Z = Z_{(1-\frac{\alpha}{2})}$ ,于是,基于 Robins、Breslow 和

Greenland 于 1986 年提出的关于“ $\ln(OR_{MH})$ ”的方差估计量<sup>[1]</sup>,共同优势比的  $100(1-\alpha)\%$  的置信区间见式(2):

$$[OR_{MH} \times \exp(-Z\hat{\sigma}), OR_{MH} \times \exp(Z\hat{\sigma})] \quad (2)$$

在式(2)中,标准误差  $\hat{\sigma}$  的定义式见式(3):

$$\hat{\sigma}^2 = \widehat{Var}[\ln(OR_{MH})] = A + B + C \quad (3)$$

式(3)中的 A、B、C 分别见下式:

$$A = \frac{\sum_h (n_{h11} + n_{h22})(n_{h11}n_{h22}) / n_h^2}{2(\sum_h n_{h11}n_{h22} / n_h)^2} \quad (4)$$

$$\begin{aligned} & \left[ OR_L \times \exp\left(-Z / \sqrt{\sum_h w_h}\right), \right. \\ & \left. OR_L \times \exp\left(Z / \sqrt{\sum_h w_h}\right) \right] \quad (8) \end{aligned}$$

在式(7)和式(8)中,  $w_h$  是第  $h$  层的权重系数,其定义见下式:

$$w_h = 1 / \text{Var}[\ln(OR_h)] \quad (9)$$

在式(9)中,  $\text{Var}[\ln(OR_h)]$  的定义见下式:

$$\text{Var}[\ln(OR_h)] = \frac{1}{n_{h11}} + \frac{1}{n_{h12}} + \frac{1}{n_{h21}} + \frac{1}{n_{h22}} \quad (10)$$

当第  $h$  层的表格中出现 0 频数时,在计算  $OR_h$  和  $w_h$  之前,需要给该层的所有格加上  $0.5^{[1]}$ 。

### 2.1.3 高维表资料共同优势比的精确置信区间估计

假定所有各层 2x2 表的优势比是一个常数。精确置信限的构造原理:在各层 2x2 表的边际总数固定的条件下,基于  $S = \sum_h n_{h11}$  的分布来构造精确置信限。精确置信区间的精度:Agresti 于 1992 年指出,由于拟解决的问题是一个离散问题,所以,所构造的精确置信区间的“置信水平”不会恰好等于  $100(1-\alpha)\%$ ,但至少是  $100(1-\alpha)\%$ 。因此,所求得的置信限是保守的<sup>[1]</sup>。精确置信限算法的来源:SAS/STAT 中的 FREQ 过程计算共同优势比的置信限是依据 Vollset、Hirji 和 Elashoff 于 1991 年提出的算法,还可参考 Mehta、Patel 和 Gray 于 1985 年发表的有关文献<sup>[1]</sup>。

算法的详细描述:在第  $h$  层 2x2 表的边际总数固定的条件下,让随机变量  $S_h$  代表第  $h$  层 2x2 表中 (1,1) 网格内的频数。给定行合计  $n_{h1\cdot}$ 、 $n_{h2\cdot}$ ,列合计  $n_{\cdot 1}$ 、 $n_{\cdot 2}$ ,  $S_h$  的下限与上限是  $l_h$ 、 $u_h$ ,则它们的计算公式分别见式(11)和式(12):

$$l_h = \max(0, n_{h1\cdot} - n_{h2\cdot}) \tag{11}$$

$$u_h = \min(n_{h1\cdot}, n_{h1\cdot}) \tag{12}$$

让  $C_{S_h}$  代表超几何系数,计算公式见下式:

$$C_{S_h} = \binom{n_{h1\cdot}}{S_h} \binom{n_{h2\cdot}}{n_{h1\cdot} - S_h} \tag{13}$$

让  $\phi$  代表共同优势比,  $S_h$  的条件分布见下式:

$$P(S = s | n_{h1\cdot}, n_{h2\cdot}; h = 1, 2, \dots, q) = \frac{C_s \phi^s}{\sum_{x=l}^{x=u} C_x \phi^x} \tag{14}$$

在式(14)中,  $C_x$  的计算公式如下:

$$C_s = \sum_{s_1 + \dots + s_q = s} \left( \prod_h C_{s_h} \right) \tag{15}$$

让  $s_0$  代表所有  $q$  张表第 (1,1) 格上频数之和。通过迭代计算方法求解下列两个方程中的共同优势比置信限的下限值与上限值,  $\phi_1$  与  $\phi_2$ , 见式(16)、式(17):

$$\sum_{x=s_0}^{x=u} C_x \phi_1^x / \sum_{x=l}^{x=u} C_x \phi_1^x = \alpha / 2 \tag{16}$$

$$\sum_{x=l}^{x=s_0} C_x \phi_2^x / \sum_{x=l}^{x=u} C_x \phi_2^x = \alpha / 2 \tag{17}$$

当观测结果的和  $s_0$  等于下界  $l$  时, SAS/STAT 中的 FREQ 过程就将置信限的下限设置为 0, 并基于显著性水平  $\alpha$  来决定置信限的上限值; 同理, 当观测结果的和  $s_0$  等于上界  $u$  时, SAS/STAT 中的 FREQ 过程

就将置信限的上限设置为  $\infty$ , 并基于显著性水平  $\alpha$  来决定置信限的下限值。

### 2.1.4 高维表资料共同优势比是否等于 1 的精确检验

在运用 SAS/STAT 中的 FREQ 过程时, 若在 exact 语句中使用了选项“COMOR”, 该过程可以计算精确检验。设  $\phi=1$ , 在无效假设成立的条件下,  $S$  的条件分布变成如下形式:

$$P_0(S = s | n_{h1\cdot}, n_{h2\cdot}; h = 1, 2, \dots, q) = \frac{C_s}{\sum_{x=l}^{x=u} C_x} \tag{18}$$

在无效假设成立且满足分层 2x2 表的边际固定的条件下, 这个精确检验的点概率就是观测和  $s_0$  出现的概率, 这个概率可以用  $P_0(s_0)$  表示。在无效假设成立的条件下,  $S$  的期望值由下式定义:

$$E_0(S) = \sum_{x=l}^{x=u} x C_x / \sum_{x=l}^{x=u} C_x \tag{19}$$

单侧精确概率  $P$  值(记为  $P_1$ ) 可从条件分布  $P_0(S \geq s_0)$  或  $P_0(S \leq s_0)$  计算得到, 取决于观测结果的和  $s_0$  大于还是小于  $E_0(S)$ , 分别见下式:

$$P_1 = P_0(S \geq s_0) = \sum_{x=s_0}^{x=u} C_x / \sum_{x=l}^{x=u} C_x, \tag{20}$$

如果  $s_0 > E_0(S)$

$$P_1 = P_0(S \leq s_0) = \sum_{x=l}^{x=s_0} C_x / \sum_{x=l}^{x=u} C_x, \tag{21}$$

如果  $s_0 \leq E_0(S)$

基于下面 3 种定义, 可分别计算出双侧精确概率  $P$  值(记为  $P_2$ )。

定义 1: 双侧概率为单侧概率的 2 倍。若该值超过 1, 将其设置为 1。见下式:

$$P_2^a = 2 \times P_1 \tag{22}$$

定义 2: 双侧概率为所有小于等于观测结果的和  $s_0$  的点概率的概率之和, 求和范围是  $s$  的所有可能取值, 即  $l \leq s \leq u$ , 公式如下:

$$P_2^b = \sum_{l \leq s \leq u; P_0(s) \leq P_0(s_0)} P_0(s) \tag{23}$$

定义 3: 双侧概率为单侧  $P$  值与分布的对侧尾端(与期望值等距)相对应的面积之和, 计算公式如下:

$$P_2^c = P_0(|S - E_0(S)| \geq |s_0 - E_0(S)|) \tag{24}$$

## 2.2 高维表资料优势比分析的 SAS 实现

### 2.2.1 问题与数据

【例 1】文献[2]提供了如下资料, 试对 5 项研究的共同优势比进行分析。见表 2。

表2 吸烟与肝细胞癌关系的5项病例对照研究结果

研究编号	吸烟组例数		不吸烟组例数	
	病例组	对照组	病例组	对照组
1	49	566	67	557
2	44	714	64	707
3	27	290	32	277
4	102	730	126	724
5	85	725	52	354

【例2】文献[2]提供了如下资料,试对6项研究的共同优势比进行分析。见表3。

表3 鼻咽癌与EB病毒感染关系的6项病例对照研究结果

研究编号	EB阳性组例数		EB阴性组例数	
	病例组	对照组	病例组	对照组
1	31	2	320	120
2	20	3	72	205
3	31	3	79	51
4	62	10	57	30
5	43	12	60	32
6	55	18	49	17

2.2.2 共同优势比分析的实现

【例3】沿用例1中的“问题与数据”,试对5项研究的共同优势比进行分析。

【分析与解答】设所需要的SAS程序如下:

```
data a;
do k=1 to 5;
do a=1 to 2;
do b=1 to 2;
input f @@;
output;
end; end; end;
cards;
49 566 67 557
44 714 64 707
27 290 32 277
102 730 126 724
85 725 52 354
;
run;
proc freq data=a;
weight f;
tables k*a*b/cmh(I2 QOR);
exact eqor comor;
run;
```

【程序说明】“exact 语句”中的选项“eqor”要求对 各层优势比 OR 是否满足齐性进行 Zelen’s 精确检验;选项“comor”要求对共同优势比进行精确检验。

【SAS输出结果及解释】

普通优比和相对风险				
统计量	方法	值	95% 置信限	
优比	Mantel-Haenszel	0.7638	0.6475	0.9010
	logit	0.7642	0.6477	0.9016
相对风险 (第1列)	Mantel-Haenszel	0.7865	0.6788	0.9114
	logit	0.7882	0.6803	0.9133
相对风险 (第2列)	Mantel-Haenszel	1.0279	1.0106	1.0455
	logit	1.0278	1.0110	1.0448

以上输出的是“普通优比和相对风险”的计算结果,其中,“普通优比”也叫做“共同优比”。实际上,就是基于“Mantel-Haenszel 法”和“logit 法”计算出来的校正“共同优比”的估计值及其 95% 置信区间。在本例中,因 95% 置信区间不包含 1,说明共同优比与 1 之间的差别具有统计学意义。

此处原本是各层 2x2 表优比齐性检验结果,结果显示,此资料满足齐性(这部分计算结果在本期“科研方法专题”的《如何正确运用  $\chi^2$  检验——高维表资料齐性检验与 SAS 实现》中已经呈现了,限于篇幅,此处从略)。关于共同优比的估计和置信区间的计算,直接利用前面的“校正计算结果”即可。但在 SAS/STAT 的 FREQ 过程中,还可采取精确算法,以获得更加可靠的计算结果如下:

共同优比	
Mantel-Haenszel 估计	0.7638
精确置信限	
95% 置信下限	0.6456
95% 置信上限	0.9043

以上输出的是共同优比的点估计值及其精确置信区间,因 95% 置信区间不包含 1,说明共同优比与 1 之间的差别具有统计学意义。

H0:共同优比=1 的精确检验	
单元格(1,1)总和(S)	307.0000
H0 下 S 的均值	345.0348
单侧 Pr<=S	0.0008
点 Pr=S	0.0002
双侧 P 值	
2*单侧	0.0016
总和<=点	0.0016
Pr>= S-均值	0.0014

以上输出的是关于共同优比是否等于 1 的精确检验结果。其中,“S=307”是高维表资料中各层 2×2 表中第 (1, 1) 格上频数之和;“S<sub>no</sub>=345.0348”是 H<sub>0</sub> (即“共同优比=1”)成立条件下推导出各层 2×2 表中第 (1, 1) 格上频数之和;点概率 P(S=307)=0.0002, 与其对应的单侧概率为 0.0008;与其对应的双侧概率 P<sub>2</sub> 有 3 个,分别基于不同的定义而算得,即基于定义 1,得 P<sub>2</sub><sup>a</sup>=0.0016;基于定义 2,得 P<sub>2</sub><sup>b</sup>=0.0016;基于定义 3,得 P<sub>2</sub><sup>c</sup>=0.0014。

【结论】无论是基于单侧检验还是双侧检验,所得 P 值都小于 0.01,说明应拒绝 H<sub>0</sub> (即“共同优比=1”),接受 H<sub>1</sub> (即“共同优比≠1”)。因优比的点估计值为 0.7836 (MH 法),说明吸烟组的优势 (odd 值) 小于不吸烟组的优势 (odd 值);更明确的专业结论是:吸烟者患肝细胞癌的风险小于不吸烟者患肝细胞癌的风险 (注意:这个结论与临床专业知识不相符,具体原因可能是原始资料中存在过失误差,有待进一步核实)。

【例 4】沿用例 2 中的“问题与数据”,试对 6 项研究的共同优势比进行分析。

【分析与解答】设所需要的 SAS 程序如下:

```
data a;
do k=1 to 6;
do a=1 to 2;
do b=1 to 2;
input f @@;
output;
end; end; end;
cards;
31 2 320 120
20 3 72 205
31 3 79 51
62 10 57 30
43 12 60 32
55 18 49 17
;
run;
proc freq data=a;
weight f;
tables k*a*b/cmh (I2 QOR BDT);
exact eqor comor;
run;
```

【SAS 输出结果及解释】

普通优比和相对风险				
统计量	方法	值	95% 置信限	
优比	Mantel-Haenszel	3.2135	2.2348	4.6210
	logit	2.8415	1.9324	4.1784
相对风险 (第 1 列)	Mantel-Haenszel	1.3435	1.2472	1.4473
	logit	1.3632	1.2750	1.4574
相对风险 (第 2 列)	Mantel-Haenszel	0.4127	0.3018	0.5643
	logit	0.5052	0.3725	0.6852

以上输出的是“普通优比和相对风险”的计算结果,其中,“普通优比”也叫做“共同优比”。实际上,就是基于“Mantel-Haenszel 法”和“logit 法”计算出来的校正“共同优比”的估计值及其 95% 置信区间。在本例中,因 95% 置信区间不包含 1,说明共同优比与 1 之间的差别具有统计学意义。

此处原本是各层 2×2 表资料优比齐性检验结果,结果显示,此资料不满足齐性 (这部分计算结果在本期“科研方法专题”的《如何正确运用  $\chi^2$  检验——高维表资料齐性检验与 SAS 实现》中已经呈现了,限于篇幅,此处从略)。关于共同优比的估计和置信区间的估计,通常的做法是基于随机效应模型推导出的公式进行计算。而在 SAS/STAT 的 FREQ 过程中,采取精确计算法,输出结果如下:

共同优比	
Mantel-Haenszel 估计	3.2135
精确置信限	
95% 置信下限	2.2588
95% 置信上限	4.8242

以上输出的是共同优比的点估计值及其精确置信区间,因 95% 置信区间不包含 1,说明共同优比与 1 之间的差别具有统计学意义。

H0:共同优比=1 的精确检验	
单元格 (1,1) 总和 (S)	242.0000
H0 下 S 的均值	201.3895
单侧 Pr>=S	<0.0001
点 Pr=S	<0.0001
双侧 P 值	
2*单侧	<0.0001
总和 <= 点	<0.0001
Pr>= S-均值	<0.0001

以上输出的是关于共同优势比是否等于 1 的精确检验结果。其中,“ $S=242$ ”是高维表资料中各层  $2 \times 2$  表中第 (1, 1) 格上频数之和;“ $S_{H_0}=201.3895$ ”是  $H_0$  (即“共同优势比=1”)成立条件下推导出各层  $2 \times 2$  表中第 (1, 1) 格上频数之和;点概率  $P(S=242) < 0.0001$ , 与其对应的单侧概率  $< 0.0001$ ; 与其对应的双侧概率  $P_2$  有 3 个, 分别基于不同的定义而算得, 即基于定义 1, 得  $P_2^1 < 0.0001$ ; 基于定义 2, 得  $P_2^2 < 0.0001$ ; 基于定义 3, 得  $P_2^3 < 0.0001$ 。

【结论】无论是基于单侧检验还是双侧检验, 所得  $P$  值都小于 0.0001, 说明应拒绝  $H_0$  (即“共同优势比=1”), 接受  $H_1$  (即“共同优势比 $\neq 1$ ”)。因优势比的点估计值为 3.2135 (MH 法), 说明 EB 阳性组的优势 (odds 值) 大于 EB 阴性组的优势 (odds 值); 更明确的专业结论是: EB 阳性者患鼻咽癌的风险大于 EB 阴性者患鼻咽癌的风险。

### 3 讨论与小结

#### 3.1 讨论

欲基于  $g \times 2 \times 2$  表资料求共同优势比, 需要先检验资料是否满足齐性要求, 即检验各层  $2 \times 2$  表资料所对应的优势比是否相等。常规的做法<sup>[2-8]</sup>如下: 若资料满足齐性要求, 可基于固定效应模型推导出的公式估计优势比及其置信区间; 若资料不满足齐性要求, 可基于随机效应模型推导出的公式估计优势比及其置信区间。然而, 在 SAS/STAT 的 FREQ 过程中, 若资料满足齐性要求, 可通过 CMH  $\chi^2$  检验方法给出校正的计算结果, 包括“校正的共同优势比的点估计值”及其“校正的 95% 置信区间”; 若资料不满足齐性要求, 可通过各层  $2 \times 2$  表中 (1, 1) 网格内的

频数的条件分布来构造计算公式, 可求得“共同优势比的精确点估计值”“精确 95% 置信区间”以及“共同优势比是否等于 1”的精确单侧概率和精确双侧概率。

#### 3.2 小结

本文对  $g \times 2 \times 2$  表资料进行了优势比分析, 其全部内容包括“各层  $2 \times 2$  表资料齐性检验”“共同优势比的点估计和置信区间估计”和“共同优势比是否等于 1 的假设检验”。从计算的角度来看, 内容涉及“校正算法”和“精确算法”。通过两个实例, 演示了基于 SAS 软件实现优势比分析的内容, 并对结果进行了解释, 做出了统计和专业结论。

### 参考文献

- [1] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 1109-1204, 2997-3216, 6007-6303, 7991-8092.
- [2] 方积乾. 卫生统计学[M]. 7 版. 北京: 人民卫生出版社, 2012: 434-455.
- [3] 胡良平. 医学统计学运用三型理论分析定量与定性资料[M]. 北京: 人民军医出版社, 2009: 248-280.
- [4] 方积乾, 陆盈. 现代医学统计学[M]. 北京: 人民卫生出版社, 2002: 150-209.
- [5] 万崇华, 罗家洪. 高级医学统计学[M]. 北京: 科学出版社, 2014: 391-411.
- [6] 罗杰, 冷卫东. 系统评价/Meta 分析理论与实践[M]. 北京: 军事医学科学出版社, 2013: 248-286.
- [7] 赵仲堂. 流行病学研究方法与应用[M]. 2 版. 北京: 科学出版社, 2005: 545-565.
- [8] 曾宪涛. 应用 STATA 做 Meta 分析[M]. 北京: 军事医学科学出版社, 2014: 22-57.

(收稿日期: 2021-05-14)

(本文编辑: 戴浩然)