

· 科研方法专题 ·

如何正确运用 χ^2 检验——高维表资料危险率差分析与 SAS 实现

胡纯严¹, 胡良平^{1,2*}

(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 $g \times 2 \times 2$ 表资料的危险率差分析方法和基于 SAS 软件实现计算的方法。需要完成的任务包括以下两项: 其一, 共同危险率差的点估计及其置信区间估计; 其二, 检验“共同危险率差是否等于 0”。其中, 实现前述第一项任务的方法共有 6 种; 实现第二项任务的方法共有 3 种。本文基于 SAS 软件和一个实例, 分别实现了共同危险率差的点估计、置信区间估计和假设检验, 并对 SAS 输出结果进行了解释, 做出统计结论和专业结论。

【关键词】 共同危险率差; 置信区间; 假设检验; 标准正态分布

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20210719001

How to use χ^2 test correctly——the analysis of a common risk difference for the data of a multiway table and the implementation of SAS software

Hu Chunyan¹, Hu Liangping^{1,2*}

(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 risk rate difference analysis method of the $g \times 2 \times 2$ table data and the calculation method based on the SAS software. The tasks that needed to be completed included the following two terms: first, the point estimation of the common risk rate difference and its confidence interval estimation; second, the test "whether the common risk rate difference was equal to 0". Among them, there were a total of 6 approaches for achieving the aforementioned first task; and three approaches for achieving the aforementioned second task. Based on the SAS software and an example, the article realized the point estimation, the confidence interval estimation and the hypothesis test of the common risk rate difference, respectively. Next, the SAS output results were explained and the statistical and professional conclusions were made.

【Keywords】 Common risk rate difference; Confidence interval; Hypothesis test; Standard normal distribution

在分析来自队列研究设计的 $g \times 2 \times 2$ 表资料且选择危险率差作为效应指标时, 需要完成以下 3 项主要任务: 其一, 检验各层 2×2 表资料危险率差是否满足齐性^[1-4]; 其二, 估计共同危险率差及其 95% 置信区间^[5]; 其三, 检验共同危险率差是否等于 0^[5]。本文将介绍前述提及的后两项任务, 结合实例并基于 SAS 软件完成有关计算, 并对 SAS 输出结果进行解释, 做出统计结论和专业结论。

1 高维表资料共同危险率差分析的基本概念

1.1 高维表 $g \times 2 \times 2$ 表资料的表达模式

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

表 1 队列研究设计下 $g \times 2 \times 2$ 表资料的第 h 层 2×2 表资料的表达模式

危险因素 接触与否	例 数			合计
	是否患病:	患病	未患病	
接触		n_{h11}	n_{h12}	n_{h1}
未接触		n_{h21}	n_{h22}	n_{h2}
合计		$n_{h \cdot 1}$	$n_{h \cdot 2}$	n_h

注: $h=1, 2, \dots, g$

1.2 高维表资料共同危险率差的含义

在分析“队列研究设计”的二维表资料时, 可以很方便地依据公式“ $RD = P_1 - P_2 = \frac{n_{11}}{n_1} - \frac{n_{21}}{n_2}$ ”计算出危险率差 (RD) 的值。然而, 对于“ $g \times 2 \times 2$ 表资料”, 却无法直接计算出共同或合并 RD 的值。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“2×2 表资料”。但事实上,这种理想的“2×2 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“2×2 表资料”之间的“微小差别”,这就是求出各层“2×2 表资料”的“权重系数 w_h ”。通过权重系数将各层“2×2 表资料”进行加权平均,从而间接获得合并后的 RD 的数值。在 SAS/STAT 的 FREQ 过程中,将“合并后的结果”称为“common risk difference^[5]”,常译成“共同或普通或合并危险率差”。

【说明】基于“ $g \times 2 \times 2$ 表资料”对共同危险率差进行分析时,通常首先需要检验各层 2×2 表资料是否满足齐性,检验“危险率差”是否满足齐性的具体方法参见文献[1-4],此处从略。

1.3 SAS 中 FREQ 过程估计共同危险率差及其置信区间的方法概述

SAS/STAT 中的 FREQ 过程^[5]可采用 6 种方法估计共同危险率差及其置信区间,分别为:①Mantel-Haenszel 估计法(简称 MH 法);②最小风险估计法(简称 MR 法);③汇总得分估计法;④Klingenberg 估计法;⑤基于 MH 法估计权重的分层 Newcombe 估计法;⑥基于 MR 法估计权重的分层 Newcombe 估计法。其中,上述前三种方法还可以进行假设检验,即检验共同危险率差是否等于 0。

2 高维表资料危险率差分析及 SAS 实现

2.1 高维表资料危险率差分析的具体算法

2.1.1 高维表资料危险率差分析的具体内容

高维表资料危险率差分析的具体内容包括以下 4 项:其一,检验资料是否满足齐性要求;其二,估计共同危险率差;其三,估计共同危险率差的置信区间;其四,检验共同危险率差是否等于 0。其中,第四项任务可以通过第三项任务间接完成。也就是说,当共同危险率差的置信区间不包含 0 时,就等于接受“ H_1 :共同危险率差 $RD \neq 0$ ”这个备择假设。

2.1.2 高维表资料共同危险率差的点估计及置信区间估计

2.1.2.1 Mantel-Haenszel 估计法

Mantel 和 Haensael 于 1959 年提出了分层权重, Sato 于 1989 年提出了 Sato 方差估计量^[5],在此基础上,构建出 Mantel-Haenszel 估计法^[5]。

第 1 步,基于 Mantel-Haenszel 估计法估计高维表资料共同危险率差 \hat{d}_{MH} (简称 MH 估计量),见式(1):

$$\hat{d}_{MH} = \sum_{h=1}^g \hat{d}_h w_h \quad (1)$$

在式(1)中, \hat{d}_h 为第 h 层的风险率差,其计算见式(2):

$$\hat{d}_h = \hat{p}_{h1} - \hat{p}_{h2} = \frac{n_{h11}}{n_{h1\cdot}} - \frac{n_{h21}}{n_{h2\cdot}} \quad (2)$$

在式(1)中, w_h 为第 h 层的权重,其计算见式(3):

$$w_h = \frac{n_{h1\cdot} \cdot n_{h2\cdot}}{n_h} \bigg/ \sum_{h=1}^g \frac{n_{h1\cdot} \cdot n_{h2\cdot}}{n_h} \quad (3)$$

第 2 步,构建共同风险率差 \hat{d}_{MH} 的 $100(1-\alpha)\%$ 置信区间,见式(4):

$$\left[\hat{d}_{MH} - Z \times \hat{\sigma}(\hat{d}_{MH}), \hat{d}_{MH} + Z \times \hat{\sigma}(\hat{d}_{MH}) \right] \quad (4)$$

在式(4)中, $Z = Z_{\left(1-\frac{\alpha}{2}\right)}$ 代表标准正态分布曲线下左侧概率为 $\left(1-\frac{\alpha}{2}\right)$ 处横坐标轴上的分位数(下同);

$\hat{\sigma}(\hat{d}_{MH})$ 为 \hat{d}_{MH} 的标准误,其计算见式(5):

$$\hat{\sigma}(\hat{d}_{MH}) = \sqrt{\left(\hat{d}_{MH} \sum_{h=1}^g P_h + \sum_{h=1}^g Q_h \right) \bigg/ \left(\sum_{h=1}^g \frac{n_{h1\cdot} \cdot n_{h2\cdot}}{n_h} \right)} \quad (5)$$

在式(5)中, P_h 和 Q_h 的计算分别见式(6)、式(7):

$$P_h = \left[n_{h1\cdot}^2 \cdot n_{h21} - n_{h2\cdot}^2 \cdot n_{h11} + n_{h1\cdot} \cdot n_{h2\cdot} \cdot (n_{h2\cdot} - n_{h1\cdot}) / 2 \right] / n_h^2 \quad (6)$$

$$Q_h = \frac{n_{h11}(n_{h2\cdot} - n_{h21}) + n_{h21}(n_{h1\cdot} - n_{h11})}{2n_h} \quad (7)$$

第 3 步,对共同风险率差 \hat{d}_{MH} 进行假设检验,见式(8)。

检验假设: $H_0: \sigma_{MH} = 0, H_1: \sigma_{MH} \neq 0$, 检验的显著性水平为 α 。其检验统计量见式(8):

$$Z_{MH} = \frac{\hat{d}_{MH}}{\hat{\sigma}(\hat{d}_{MH})} \sim N(0, 1) \quad (8)$$

如果 $|Z_{MH}| \geq Z_{\left(1-\frac{\alpha}{2}\right)}$, 则 $P \leq \alpha$, 拒绝 H_0 , 接受 H_1 ; 反之亦然。

2.1.2.2 最小风险估计法

Mehrotra 和 Railkar 于 2000 年提出了最小风险估计法^[5]。

第 1 步,基于最小风险估计法估计高维表资料共同危险率差 \hat{d}_{MR} (简称 MR 估计量),见式(9):

$$\hat{d}_{MR} = \sum_{h=1}^g \hat{d}_h w_h^* \quad (9)$$

在式(9)中, \hat{d}_h 为第 h 层的风险率差,其计算见

上文式(2); w_h^* 为第 h 层的权重, 其计算见式(10):

$$w_h^* = \frac{\beta_h}{\sum_h \hat{V}_h^{-1}} \left(\frac{\alpha_h \hat{V}_h^{-1}}{\sum_h \hat{V}_h^{-1} + \sum_h \alpha_h \hat{d}_h \hat{V}_h^{-1}} \right) \left(\frac{\sum_h \hat{d}_h \beta_h}{\sum_h \hat{V}_h^{-1}} \right) \quad (10)$$

在式(10)中, α_h 和 β_h 的计算分别见式(11)、式(12):

$$\alpha_h = \sum_h \hat{V}_h^{-1} - \sum_h \hat{d}_h \hat{V}_h^{-1} \quad (11)$$

$$\beta_h = \hat{V}_h \left(1 + \alpha_h \sum_h f_h \hat{d}_h \right) \quad (12)$$

在式(12)中, f_h 是第 h 层的分数, 其计算见式(13):

$$f_h = \frac{n_h}{\sum_h n_h} \quad (13)$$

\hat{d}_{MR} 的方差计算公式见式(14):

$$\hat{V}(\hat{d}_{MR}) = \sum_h w_h^{*2} \hat{V}_h \quad (14)$$

在式(14)中, \hat{V}_h 的计算见式(15):

$$\hat{V}_h = \frac{\hat{P}_{h1}(1-\hat{P}_{h1})}{n_{h1.}} + \frac{\hat{P}_{h2}(1-\hat{P}_{h2})}{n_{h2.}} \quad (15)$$

第2步, 构建风险率差 \hat{d}_{MR} 的 $100(1-\alpha)\%$ 置信区间, 见式(16):

$$\left\{ \hat{d}_{MR} - \left[C + Z \sqrt{\hat{V}(\hat{d}_{MR})} \right], \hat{d}_{MR} + \left[C + Z \sqrt{\hat{V}(\hat{d}_{MR})} \right] \right\} \quad (16)$$

在式(16)中, C 为连续性校正数, 其计算见式(17):

$$C = 0.1875 / \sum_h (n_{h1.} n_{h2.} / n_h) \quad (17)$$

当 $C < |\hat{d}_{MR}|$ 时, 才使用校正数。

第3步, 对风险率差 \hat{d}_{MR} 进行假设检验, 见下式。

检验假设: $H_0: \sigma_{MR} = 0, H_1: \sigma_{MR} \neq 0$, 检验的显著性水平为 α 。其检验统计量见式(18)、式(19):

$$Z_{MR} = \frac{\hat{d}_{MR} - C}{\sqrt{\hat{V}_0(\hat{d}_{MR})}}, \text{ 当 } \hat{d}_{MR} > 0 \text{ 时} \quad (18)$$

$$Z_{MR} = \frac{\hat{d}_{MR} + C}{\sqrt{\hat{V}_0(\hat{d}_{MR})}}, \text{ 当 } \hat{d}_{MR} < 0 \text{ 时} \quad (19)$$

如果 $|Z_{MR}| \geq Z = Z_{(1-\frac{\alpha}{2})}$, 则 $P \leq \alpha$, 拒绝 H_0 , 接受 H_1 ; 反之亦然。

在式(18)和式(19)中, $\hat{V}_0(\hat{d}_{MR})$ 的计算见式(20):

$$\hat{V}_0(\hat{d}_{MR}) = \sum_h w_h^{*2} \hat{V}_0(\hat{d}_h) \quad (20)$$

在式(20)中, $\hat{V}_0(\hat{d}_h)$ 的计算见式(21):

$$\hat{V}_0(\hat{d}_h) = \bar{P}_h (1 - \bar{P}_h) \left(\frac{1}{n_{h1.}} + \frac{1}{n_{h2.}} \right) \quad (21)$$

在式(21)中, \bar{P}_h 的计算见式(22):

$$\bar{P}_h = \frac{n_{h1.} \hat{P}_{h1.} + n_{h2.} \hat{P}_{h2.}}{n_{h1.} + n_{h2.}} \quad (22)$$

2.1.2.3 汇总评分估计法

Agresti 于 2013 年提出了汇总评分估计法^[5]。

第1步, 基于汇总评分估计法估计高维表资料共同危险率差 \hat{d}_s , 见式(23):

$$\hat{d}_s = \sum_h \hat{d}'_h W'_h \quad (23)$$

在式(23)中, W'_h 的计算见式(24):

$$W'_h = \frac{1/S_h^2}{\sum_h (1/S_h^2)} \quad (24)$$

第 h 层风险率差评分的 $100(1-\alpha)\%$ 置信区间见式(25):

$$\left[\hat{d}'_h - Z \times S'_h, \hat{d}'_h + Z \times S'_h \right] \quad (25)$$

在式(23)、式(24)、式(25)中, \hat{d}'_h 是第 h 层评分置信区间的中点, S'_h 是第 h 层评分置信区间的宽度除以 $2Z$ 。

第2步, 构建风险率差的汇总评分的 $100(1-\alpha)\%$ 置信区间, 见式(26):

$$\left[\hat{d}_s - Z \times \hat{\sigma}(\hat{d}_s), \hat{d}_s + Z \times \hat{\sigma}(\hat{d}_s) \right] \quad (26)$$

在式(26)中, $\hat{\sigma}(\hat{d}_s)$ 的计算见式(27):

$$\hat{\sigma}(\hat{d}_s) = \sqrt{\frac{1}{\sum_h (1/S_h^2)}} \quad (27)$$

第3步, 对风险率差 \hat{d}_s 进行假设检验, 见式(28):

$$Z_s = \frac{\hat{d}_s}{\hat{\sigma}(\hat{d}_s)} \sim N(0, 1) \quad (28)$$

如果 $|Z_s| \geq Z = Z_{(1-\frac{\alpha}{2})}$, 则 $P \leq \alpha$, 拒绝 H_0 , 接受 H_1 ;

反之亦然。

2.1.2.4 Klingerberg 估计法

基于 Mantel 和 Haensael's 共同风险率差以及 Sato 方差估计量, Klingerberg 于 2014 年提出了 Klingerberg 估计法^[5]。

第1步, 基于 Klingerberg 估计法估计高维表资

料共同危险率差的置信区间中点 \hat{d}_{Mid} , 见式(29):

$$\hat{d}_{Mid} = \hat{d}_{MH} + 0.5Z^2(P/W^2) \quad (29)$$

第2步, 构建风险率差的Klingerberg's 100(1- α)% 置信区间, 见式(30):

$$\left[\hat{d}_{Mid} - ME_{(1-\frac{\alpha}{2})}, \hat{d}_{Mid} + ME_{(1-\frac{\alpha}{2})} \right] \quad (30)$$

在式(30)中, $ME_{(1-\frac{\alpha}{2})}$ 的计算见式(31):

$$ME_{(1-\frac{\alpha}{2})} = \sqrt{\hat{d}_{Mid} - \hat{d}_{MH}^2 + Z^2(Q/W^2)} \quad (31)$$

在式(29)、式(30)、式(31)中, 有关变量的定义如下:

$P = \sum_h P_h, Q = \sum_h Q_h, W = \sum_h w_h; w_h, P_h$ 和 Q_h 的计算分别见式(3)、式(6)和式(7)。

2.1.2.5 分层 Newcombe 估计法

Yan 和 Su 于 2010 年提出了分层 Newcombe 估计法^[5]。该法是基于分层的 Wilson 置信区间法经过改造后构建而成。设合并后的 2×2 表的第 1 行上比例的 Wilson 置信限的下限与上限分别为 L_1 与 U_1 ; 其第 2 行上比例的 Wilson 置信限的下限与上限分别为 L_2 与 U_2 。于是, 共同危险率(或比例)差的 100(1- α)% 分层 Newcombe 置信限的下限与上限分别为 L 与 U , 其计算见式(32)、式(33):

$$L = \hat{d} - Z \sqrt{\lambda_1 L_1(1-L_1) + \lambda_2 U_2(1-U_2)} \quad (32)$$

$$U = \hat{d} + Z \sqrt{\lambda_2 L_2(1-L_2) + \lambda_1 U_1(1-U_1)} \quad (33)$$

在上面两式中, \hat{d} 是共同风险率差的加权估计值, 其计算公式取决于所选用的权重。当权重为前面的式(3)时, 则 $\hat{d} = \hat{d}_{MH}$; 当权重为前面的式(10)时, 则 $\hat{d} = \hat{d}_{MR}$ 。

在式(32)、式(33)中, λ_1 与 λ_2 的计算分别见式(34)、式(35):

$$\lambda_1 = \sum_{h=1}^g \frac{w_h^2}{n_{h1}} \quad (34)$$

$$\lambda_2 = \sum_{h=1}^g \frac{w_h^2}{n_{h2}} \quad (35)$$

【说明】该法可分解成两种方法, 即“Newcombe 法($\hat{d} = \hat{d}_{MH}$)”和“NewcombeMR 法($\hat{d} = \hat{d}_{MR}$)”。

2.2 高维表资料共同危险率差分析的 SAS 实现

2.2.1 问题与数据

【例1】文献[4,6]提供了如下资料, 见表2。试对

7项研究的共同危险率差进行分析。

表2 阿司匹林预防心肌梗死后患者死亡与否的7项随机对照研究结果

研究编号	阿司匹林组例数		安慰剂组例数	
	死亡	存活	死亡	存活
1	49	566	67	557
2	44	714	64	707
...
7	1570	7008	1720	6880

注: 详细数据参见后文 SAS 程序

2.2.2 对例1资料共同危险率差分析的 SAS 实现

【例2】沿用例1中的“问题与数据”, 试对7项研究的共同危险率差进行分析。

【分析与解答】设所需要的 SAS 程序如下:

```
data a;
do k=1 to 7;
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
102 730 126 724
32 285 38 271
85 725 52 354
246 2021 219 2038
1570 7008 1720 6880
;
run;
proc freq data=a;
weight f;
tables k*a*b/cmh(I2 QOR BDT)
commonriskdiff(test=(mh mr score)
cl=(k mh mr score newcombe newcombemr));
run;
```

【程序说明】在“tables 语句”中有3个选项, 在选项“cmh”后指定了进行优比齐性检验的3种方法; 在选项“commonriskdiff”后指定了进行共同危险率差检验的3种方法; 在选项“cl”后指定了进行共同危险率差置信区间估计的6种方法。

【SAS 输出结果及解释】

普通风险率差值置信限

方法	值	标准误差	95%置信限	
Klingenberg	-0.0141		-0.0226	-0.0056
Mantel-Haenszel	-0.0141	0.0043	-0.0226	-0.0056
最小风险	-0.0141	0.0042	-0.0223	-0.0059
Newcombe	-0.0141		-0.0227	-0.0056
Newcombe(MR)	-0.0141		-0.0228	-0.0054
汇总评分	-0.0134	0.0042	-0.0216	-0.0052

列 1(b=1)

以上输出的是关于共同危险率差的 6 种置信区间估计结果,第 2 列为“阿斯匹林与安慰剂治疗后死亡率之差值的估计值”,差值为负值,表明阿司匹林组的死亡率低于安慰剂组的死亡率,6 种方法得到的共同危险率差的 95% 置信区间都没有包含 0,表明“共同死亡率差不等于 0”,即阿斯匹林与安慰剂治疗后死亡率之间的差别有统计学意义,在治疗心肌梗死时,阿斯匹林的治疗效果比安慰剂好。

普通风险率差值检验

方法	风险差值	Z	Pr> Z
Mantel-Haenszel	-0.0141	-3.2594	0.0011
最小风险	-0.0141	-3.3674	0.0008
汇总评分	-0.0134	-3.2010	0.0014

列 1(b=1)

以上输出的是关于共同危险率差是否等于 0 的 3 种检验结果,均得到 $P < 0.01$,说明阿斯匹林与安慰剂治疗后死亡率之间差异有统计学意义,在治疗心肌梗死时,阿斯匹林的治疗效果比安慰剂好。

Cochran-Mantel-Haenszel 统计量(基于表评分)

统计量	备择假设	自由度	值	概率
1	非零相关	1	10.6333	0.0011
2	行评分均值差异	1	10.6333	0.0011
3	一般关联	1	10.6333	0.0011

以上是采用 Cochran-Mantel-Haenszel 法(简称 CMH 法)对合并的 2x2 表资料进行 3 种假设检验的结果,表明“两种药物(阿斯匹林与安慰剂)”与“两种治疗结果(死亡与生存)”之间存在关联性(因第 3 行上的备择假设为“一般关联”),具体地说,即阿斯匹林治疗的死亡率低于安慰剂治疗的死亡率。

普通优比和相对风险

统计量	方法	值	95%置信限	
优比	Mantel-Haenszel	0.8977	0.8413	0.9579
	logit	0.8978	0.8413	0.9580
相对风险(第 1 列)	Mantel-Haenszel	0.9143	0.8664	0.9649
	logit	0.9145	0.8665	0.9651
相对风险(第 2 列)	Mantel-Haenszel	1.0169	1.0067	1.0273
	logit	1.0155	1.0058	1.0252

以上输出的是“普通优比和相对风险”的计算结果,其中,“普通”也叫做“共同或合并”。实际上,就是基于“Mantel-Haenszel 法”和“logit 法”计算出来的校正“共同相对危险度”的估计值及其 95% 置信区间。在本例中,应看“相对风险(第 1 列)”的两种方法计算的结果(因为第 1 列上的结果为“死亡”,是研究者特别关注的结局)。 $RR_{MH} = 0.9143$,其 95% 置信区间为 $[0.8664, 0.9649]$; $RR_L = 0.9145$,其 95% 置信区间为 $[0.8665, 0.9651]$ 。两种方法得到的置信区间都不包含 1,表明共同相对危险度小于 1,即阿斯匹林对应的死亡率小于安慰剂对应的死亡率。

优比齐性的 Breslow-Day 检验

卡方	9.9753
自由度	6
Pr>卡方	0.1257

优比齐性 Q 检验

卡方	自由度	Pr>卡方
9.9426	6	0.1271

I 平方异质测度

I 平方	自由度	95% 不确定性限值
39.65%	6	0.00% 74.62%

以上是基于“优比的齐性检验”结果,可作为基于“危险率差的齐性检验”结果的参考。结果显示,资料满足齐性要求。

【统计结论与专业结论】在本例中,因 95% 置信区间不包含 1,说明共同相对危险度与 1 之间的差别具有统计学意义。由于 $RR = 0.9145 < 1$,说明阿司匹林对应的死亡率低于安慰剂对应的死亡率。也就是说,相对于采用安慰剂而言,采用阿司匹林预防心肌梗死可明显降低心肌梗死患者的死亡率。这与前面基于“普通风险率差值置信限”计算结果所得结论是完全相同的。

3 讨论与小结

3.1 讨论

在 SAS/STAT 的 FREQ 过程^[5]中,只给出了针对“优比”的齐性检验,未给出针对“相对危险度”和“危险率差”的齐性检验(一般来说,对同一个资料而言,前述提及的几种齐性检验结果基本接近,故在要求不高的前提下,可以将“优比的齐性检验结果”作为“危险率差的齐性检验结果”的近似解);SAS 基于 6 种方法给出了关于共同危险率差及其置信区间的估计公式,其中,有 3 种方法还可以直接对“共同危险率差是否等于 0”进行假设检验。但 SAS 并没有明确交代前述提及的 6 种方法哪些仅适用于“资料满足齐性要求”,哪些方法适用于“资料不满足齐性要求”。当使用者需要基于“危险率差”进行“Meta 分析”时,解决的办法有两种:其一,基于相应的公式和 SAS 语言编程^[1,3-5,7-8]实现计算;其二,使用专门用于 Meta 分析的 RevMan 软件^[9-10]实现计算。

3.2 小结

本文对 $g \times 2 \times 2$ 表资料进行了危险率差分析,其内容包括 6 种“危险率差的点估计和置信区间估计方法”和 3 种关于“危险率差是否等于 0”的检验方

法。通过一个实例并基于 SAS 软件实现了危险率差分析,并对 SAS 输出结果进行了解释,做出了统计结论和专业结论。

参考文献

- [1] 胡纯严,胡良平.如何正确运用 χ^2 检验——高维表资料齐性检验与 SAS 实现[J].四川精神卫生,2021,34(3):202-207.
- [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] 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.
- [4] 万崇华,罗家洪.高级医学统计学[M].北京:科学出版社,2014:391-411.
- [5] 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.
- [6] 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.
- [7] 胡良平.面向问题的统计学——(2)多因素设计与线性模型分析[M].北京:人民卫生出版社,2012:389-405.
- [8] 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.
- [9] 曾先涛.应用 STATA 做 Meta 分析[M].北京:军事医学科学出版社,2014:22-57.
- [10] 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.

(收稿日期:2021-07-19)

(本文编辑:戴浩然)



科研方法专题策划人——胡良平教授简介

胡良平,男,1955年8月出生,教授,博士生导师,曾任军事医学科学院研究生部医学统计学教研室主任和生物医学统计学咨询中心主任、国际一般系统论研究会中国分会概率统计系统专业理事会常务理事、中国生物医学统计学学会副会长、北京大学口腔医学院客座教授和《中华医学杂志》等10余种杂志编委;现任世界中医药学会联合会临床科研统计学专业委员会会长、国家食品药品监督管理局评审专家和3种医学杂志编委;主编统计学专著48部、参编统计学专著10部;发表第一作者和通信作者学术论文300余篇、发表合作论文130余篇;获军

队科技成果和省部级科技成果多项;参加并完成三项国家标准的撰写工作、参加三项国家科技重大专项课题研究工作。在从事统计学工作的30年中,为几千名研究生、医学科研人员、临床医生和杂志编辑讲授生物医学统计学,在全国各地作统计学学术报告100余场,举办数十期全国统计学培训班,培养20多名统计学专业硕士和博士研究生。近几年来,参加国家级新药和医疗器械项目评审数十项、参加100多项全军重大重点课题的统计学检查工作。归纳并提炼出有利于透过现象看本质的“八性”和“八思维”的统计学思想,独创了逆向统计学教学法和三型理论。擅长于科研课题的研究设计、复杂科研资料的统计分析和 SAS 与 R 软件实现、各种层次的统计学教学培训和咨询工作。