

如何正确运用 χ^2 检验——三种 $R\times C$ 列联表资料的 CMH χ^2 检验

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】 本文目的是介绍三种 $R\times C$ 列联表资料的CMH χ^2 检验以及SAS软件实现的方法。第一种是“双向无序 $R\times C$ 列联表资料”,与这种资料对应的CMH χ^2 检验在本质上就是Pearson's χ^2 检验;第二种是“结果变量为有序变量的 $R\times C$ 列联表资料”,与这种资料对应的CMH χ^2 检验在本质上就是秩和检验;第三种是“双向有序且属性不同 $R\times C$ 列联表资料”,与这种资料对应的CMH χ^2 检验在本质上就是Pearson's相关分析或Spearman's秩相关分析。当 $R\times C$ 列联表资料中有1个或2个“有序变量”时,进行统计分析之前,需要给有序变量“赋值或评分”。在SAS/STAT的FREQ过程中,有四种评分方法。评分方法不同,CMH χ^2 检验统计量的表达形式和计算结果都会发生相应变化。

【关键词】 有序变量;CMH χ^2 检验;秩和检验;秩相关分析;SAS软件;R软件

中图分类号:R195.1

文献标识码:A

doi:10.11886/scjsws20210316005

How to use χ^2 test correctly——CMH χ^2 tests for the data collected from the three kinds of $R\times C$ contingency tables

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 paper was to introduce the CMH χ^2 test and SAS software implementation of the three kinds of $R\times C$ contingency table data. The first type was called “two-way unordered $R\times C$ contingency table data”. The CMH χ^2 test corresponding to this type of data was essentially the Pearson's χ^2 test. The second type was called “ $R\times C$ contingency table data with an ordinal outcome variable”. The CMH χ^2 test corresponding to this kind of data was essentially a rank sum test. The third type was called “ $R\times C$ contingency table data which was of two ordinal variables with different attributes”. The CMH χ^2 test corresponding to the data was essentially Pearson's correlation analysis or Spearman's rank correlation analysis. When there were 1 or 2 “ordinal variables” in the $R\times C$ contingency table data, it was necessary to “assign or score” the ordinal variables before performing statistical analysis. In the FREQ procedure of SAS/STAT, there were four scoring methods. With different scoring approach, both the expression form and the calculation results of CMH χ^2 test statistics could change accordingly.

【Keywords】 Ordinal variables; CMH χ^2 test; Rank sum test; Rank correlation analysis; SAS software; R software

在横断面设计的二维列联表资料或称为 $R\times C$ 列联表资料中,根据两变量是否为有序变量,可以归纳出三种不同的二维列联表资料:①双向无序 $R\times C$ 列联表资料;②结果变量为有序变量 $R\times C$ 列联表资料;③双向有序且属性不同 $R\times C$ 列联表资料。对其进行统计分析的方法分别为“独立性假设检验”“秩和检验”和“相关分析”,它们可以被概括成一种统计分析方法,即CMH χ^2 检验。本文介绍广义CMH χ^2 检验统计量及其三种变形,并基于SAS软件实现统计计算。

1 适合进行CMH χ^2 检验的三种 $R\times C$ 列联表资料的实例

1.1 双向无序 $R\times C$ 列联表资料的实例

【例1】文献[1]中有一个双向无序 $R\times C$ 表资料(说明:不考虑“时间”的有序性),见表1,试分析“条目”与“时间”之间是否存在关联性。

1.2 结果变量为有序变量 $R\times C$ 列联表资料的实例

【例2】文献[2]给出了如下临床资料:比较三种对肥厚性鼻炎(HR)治疗方法的效果。将162例HR

患者分为三组, A 组($n=56$)行下鼻甲骨黏骨膜下切除术, B 组($n=43$)行下鼻甲部分切除术, C 组($n=63$)行下鼻甲黏膜下微波热凝术。治疗效果见表 2。试对三组治疗效果进行比较。

表 1 基层精防医护人员 K6 评定结果

条 目	例 数				
	时间: 所有时间	大部分时间	少部分时间	偶尔	无
A	6	12	28	62	36
B	2	3	6	20	113
C	2	6	16	42	78
D	1	3	10	31	99
E	2	2	9	43	88
F	2	3	3	20	116

注: K6 代表“凯斯勒心理困扰量表”的英文缩写; A、B、C、D、E、F 分别代表“感到紧张”“感到没有希望”“感到烦躁不安”“感到太沮丧、无法愉快起来”“感到做每一件事情都很费力”和“感到无价值”

表 2 三组下鼻甲手术治疗的临床效果

组 别	例 数					
	治疗效果:	治愈	显著改善	改善	无效	合计
A 组		15	19	19	3	56
B 组		7	10	18	8	43
C 组		11	21	24	7	63
合计		33	50	61	18	162

1.3 双向有序 $R \times C$ 列联表资料的实例

【例 3】在文献[3]中, 为探讨新冠肺炎疫情期间居民急性应激障碍(ASD)症状的检出情况及影响因素, 得到有效问卷 16 048 份, 有效问卷回收率为 72.83%。在该研究中, 因素“媒体暴露情况(时间)”与结果变量(出现 ASD 症状的程度)之间的数量关系如表 3 所示。试分析“媒体暴露情况(时间)”与“ASD 症状的程度”之间是否存在线性相关关系。

表 3 媒体暴露情况(时间)与 ASD 症状检出情况的调查结果

媒体暴露情况(时间)	例 数			
	ASD 程度: 无	中度	重度	合计
<1 小时	2697	412	684	3793
1~3 小时	6432	857	1369	8658
3~5 小时	1402	196	343	1941
5~8 小时	624	91	180	895
>8 小时	519	74	168	761
合计	11674	1630	2744	16048

注: ASD, 急性应激障碍

2 CMH χ^2 检验的检验统计量及其三种变形

2.1 “CMH”的含义

“CMH”是“Cochran-Mantel-Haenszel”的缩写, 即采用三位作者姓名的第一个字母来命名该分析

方法, 实际上, Cochran、Mantel、Haenszel、Mantel、Birch、Landis、Heyman 和 Koch 都对该分析方法做出过贡献^[4]。

2.2 广义 CMH χ^2 检验统计量的定义

一般来说, CMH χ^2 检验统计量是为分析高维列联表资料而构造出来的。将高维列联表资料按某一个或多个因素进行分层, 每层都应该是规模相同的 $R \times C$ 列联表资料。下面以表 4 形式表示第 h 层的 $R \times C$ 表, $h=1, 2, \dots, q$ 。 q 为层数, R 为行数, C 为列数。

表 4 第 h 层 $R \times C$ 列联表的列表格式

原因变量	例 数					
	水平	结果变量水平: 1	2	...	C	合计
1		n_{h11}	n_{h12}	...	n_{h1C}	n_{h1+}
2		n_{h21}	n_{h22}	...	n_{h2C}	n_{h2+}
...	
R		n_{hR1}	n_{hR2}	...	n_{hRC}	n_{hR+}
合计		n_{h+1}	n_{h+2}	...	n_{h+C}	n_h

注: n_{hj} 表示第 h 层第 i 行第 j 列所对应的频数; n_{hi+} 为第 h 层第 i 行的合计数, $i=1, 2, \dots, R$; n_{h+j} 为第 h 层第 j 列的合计数, $j=1, 2, \dots, C$; n_h 为第 h 层的合计数

广义 CMH χ^2 检验统计量^[4]定义如下:

$$\chi^2 = Q_{CMH} = G'V_G^{-1}G \quad (1)$$

其中 $G = \sum_h B_h(n_h - m_h)$, G' 代表矩阵 G 的转置

矩阵, $V_G = \sum_h B_h [Var(n_h | H_0)] B_h'$, V_G^{-1} 代表矩阵 V_G 的逆矩阵。这里 $B_h = C_h \otimes R_h$, C_h 是列得分, R_h 是行得分, B_h 是基于列得分与行得分的常数矩阵。当原假设成立时, CMH χ^2 统计量渐近服从 χ^2 分布, 它的自由度等于 B_h 的秩。“ \otimes ”代表“叉积”或“直积”或“Kronecker 积”的符号, 它是两个矩阵之间的一种特定的乘积^[5]。

需注意的是, 当各层间效应方向不一致时, CMH χ^2 检验统计量的检验功效很低。

2.3 广义 CMH χ^2 检验统计量的三种变形

2.3.1 概述

式(1)是以矩阵形式呈现的计算公式, 很不直观。由于它试图达到高维列联表资料的多种不同分析目的(例如独立性分析、相关性分析、差异性分析), 故其隐含诸多前提条件, 包括“是否存在有序变量”“如何给有序变量赋值”。基于不同的前提条件, 就会产生出不同的统计计算公式。为了直观起见, 现假定只有一层, 即一个二维 $R \times C$ 列联表资料, 基于不同的前提条件, 将式(1)转变成三种具体的、直观的检验统计量。

2.3.2 与备择假设为“非零相关”对应的检验统计量

若拟分析的资料如本文表 3(即双向有序 $R \times C$ 列联表资料)所示,此时式(1)就简化成下式:

$$\chi_1^2 = (N - 1)r^2 \text{ 或 } \chi_1^2 = (N - 1)r_s^2 \quad (2)$$

在式(2)中, χ_1^2 代表该检验统计量服从自由度为 1 的 χ^2 分布;“ N ”代表 $R \times C$ 列联表中的总频数;“ r ”与“ r_s ”分别代表 Pearson’s 相关系数与 Spearman’s 秩相关系数;当给 $R \times C$ 列联表的行变量和列变量“赋值或评分”时,若按“表评分(具体方法详见下文)”方式“评分或赋值”,就采用 Pearson’s 相关分析;否则[包括“rank(秩)评分”“Ridit 评分”^[6]和“修正的 Ridit 评分”],就采用 Spearman’s 秩相关分析。

2.3.3 与备择假设为“行平均评分不同”对应的检验统计量

若拟分析的资料如本文表 2(即结果变量为有序变量 $R \times C$ 列联表资料)所示,此时式(1)就简化成下面两个式子之一:

$$\chi_{R-1}^2 \propto F = \frac{MS_{\text{组间}}}{MS_{\text{误差}}} \quad (3)$$

$$\begin{aligned} \chi_{R-1}^2 \approx H &= \frac{Q_{\text{组间}}}{Q_{\text{总}}/(N-1)} \\ &= \frac{\sum_{i=1}^R \frac{M_i^2}{n_i} - \frac{N(N+1)^2}{4}}{\frac{N(N^2-1)}{12(N-1)}} \end{aligned} \quad (4)$$

在式(3)中,“ \propto ”代表“呈正比”;当给 $R \times C$ 列联表的列变量(即结果变量)赋值或评分时,若按“表评分”方式评分,就采用式(3)(即单因素 R 水平设计一元定量资料方差分析)计算^[4,7];否则[包括“rank(秩)评分”“Ridit 评分”和“修正的 Ridit 评分”],就采用式(4)(即单因素 R 水平设计一元定量资料 Kruskal-Wallis’s H 秩和检验)计算^[4,7]。在式(4)中, n_i 与 M_i 分别代表第 i 组中的“样本含量”与“秩和”。

在式(3)中,可对 F 统计量进行变换,见下式:

$$F = \frac{MS_{\text{组间}}}{MS_{\text{误差}}} \approx \frac{SS_{\text{组间}}/df_{\text{组间}}}{\sigma^2} \quad (5)$$

通过对式(5)变形,可以得到下式^[8-9]:

$$df_{\text{组间}} \times F \approx \frac{SS_{\text{组间}}}{\sigma^2} \sim \chi_{df_{\text{组间}}}^2 \quad (6)$$

式(6)的含义是:采用单因素 R 水平设计一元定量资料方差分析得到的检验统计量 F 值后,乘以组

间自由度 $df_{\text{组间}}$ 所得到的结果,近似服从自由度为 $df_{\text{组间}}$ 的 χ^2 分布。

2.3.4 与备择假设为“一般关联性”对应的检验统计量

若拟分析的资料如本文表 1(即双向无序 $R \times C$ 列联表资料)所示,此时式(1)就简化成下式:

$$\chi_{CMH}^2 = \chi_P^2 \frac{N-1}{N}, df = (R-1)(C-1) \quad (7)$$

在上式中, χ_{CMH}^2 代表该检验统计量服从自由度为 $df=(R-1)(C-1)$ 的 χ^2 分布; χ_P^2 代表 Pearson’s χ^2 检验统计量,其计算公式见下式:

$$\chi_P^2 = N \left(\sum_{j=1}^C \sum_{i=1}^R \frac{n_{ij}^2}{n_i \cdot n_j} - 1 \right), df = (R-1)(C-1) \quad (8)$$

χ_P^2 为服从自由度为 $df=(R-1)(C-1)$ 的 χ^2 分布;“ N ”为列联表的总频数、“ n_{ij} ”为第 (i, j) 网格中的观测频数、“ n_i ”与“ n_j ”分别代表列联表中第 i 行与第 j 列上的“合计值”。

事实上,式(7)中的“ χ_P^2 ”就是检验 $R \times C$ 列联表资料中两属性变量之间是否独立的检验统计量,式(7)就是在“ χ_P^2 ”的基础上乘以一个校正数“ $\frac{N-1}{N}$ ”而已,当“ N ”足够大时,此校正数接近于 1。因此,式(7)与式(8)几乎是等价的。

3 CMH χ^2 检验的 SAS 实现

【例 4】沿用例 3 的资料,试将其分别视为“双向有序 $R \times C$ 列联表资料”“结果变量为有序变量 $R \times C$ 列联表资料”和“双向无序 $R \times C$ 列联表资料”,同时对其进行 CMH χ^2 检验。基于“表评分”算法所需要的 SAS 程序如下:

```
DATA a;
DO A=1 TO 5;
DO B=1 TO 3;
INPUT F @@; OUTPUT;
END;
END;
CARDS;
2697 412 684
6432 857 1369
1402 196 343
624 91 180
519 74 168
;
```

```

RUN;
ods select rowscores;
PROC FREQ data=a;
tables A*B/CMH scorout SCORES=table;
WEIGHT F;
RUN;
【SAS输出结果及解释】

```

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	6.6350	0.0100
2	行评分均值不同	4	33.7083	<0.0001
3	常规关联	8	36.6435	<0.0001

以上是基于“表评分”算法所得 CMH χ^2 检验的三种计算结果,分别对应三种不同的备择假设:第一种,“非零相关(即行、列两有序变量之间具有非零的相关关系)”, $\chi^2_{CMH} = 6.635, df=1, P=0.0100$;第二种,“行评分均值不同(即各行上有序结果变量的评分值的平均数不等)”, $\chi^2_{CMH} = 33.7083, df=4, P<0.0001$;第三种,“常规关联(即行、列两无序变量之间存在关联性)”, $\chi^2_{CMH} = 36.6435, df=8, P<0.0001$ 。

以上三种检验都得出 $P<0.01$, 故接受备择假设。对本例而言,可以认为:①“媒体暴露情况(时间)”与“ASD 程度”之间存在“非零相关”,具体地说,随着媒体暴露时间延长,出现重度 ASD 症状的比例呈非线性上升趋势;②五种不同媒体暴露时间所对应的平均 ASD 值是不相等的,除第 2 个时间段之外,其他 4 个时间段都随着“暴露时间延长”,平均 ASD 值变大(即 ASD 症状程度逐渐加重);③媒体暴露情况(时间)与 ASD 程度之间存在关联性(前面的“①”和“②”可以清楚体现出两变量之间是如何关联的)。

4 给有序变量评分的方法

4.1 表评分方法

在前面的 SAS 程序中,在“tables 语句”中出现了选项“SCORES=table”,其含义是用表格里规定的方式给有序变量评分。这里的“表格”,其实是指 SAS 程序中如何给“行变量”或“列变量”赋值。“DO A=1 TO 5;”就是让“行变量 A”依次取 1、2、3、4、5 分;同理,“DO B=1 TO 3;”就是让“列变量 B”依次取 1、2、3 分。当然,也可以给变量 A 或 B 赋任意的几个由小到大的数值,例如“DO A=0.1, 1.2, 13.5, 21.8, 34.6;”。

4.2 其他评分方法

在 SAS/STAT 的 FREQ 过程中,还有另外三种评分方法,分别为“rank 评分法”“ridit 评分法”和“修正的 ridit 评分法”,对应的选项依次为“SCORES=rank”“SCORES=ridit”和“SCORES=modridit”。

4.3 评分方法与检验统计量计算方法之间的关系

对“常规关联”的计算结果而言,四种评分方法所得计算结果是完全相同的,因为计算公式中不涉及“评分”;对“非零相关”的计算结果而言,“表评分法”(采用 Pearson's 相关分析)与另外三种评分法(其结果是一样的,采用 Spearman's 秩相关分析)计算结果可能相差很大、甚至结论相反;对“行评分均值不同”的计算结果而言,“表评分法”(采用单因素 R 水平设计一元定量资料方差分析)与另外三种评分法(其结果是一样的,采用单因素 R 水平设计一元定量资料 Kruskal-Wallis's H 秩和检验)计算结果稍有差别。例如,例 4“基于秩评分”的结果如下:

Cochran-Mantel-Haenszel 统计量(基于秩评分)

统计量	备择假设	自由度	值	概率
1	非零相关	1	0.5174	0.4720
2	行评分均值不同	4	31.2716	<0.0001
3	常规关联	8	36.6435	<0.0001

【说明】“ridit 评分法”和“修正的 ridit 评分法”输出结果与“基于秩评分”方法输出结果相同,此处从略。

注意:“非零相关”的结果与前面“基于表评分”的结果相反;四种评分法对应的“常规关联”结果都是 $\chi^2=36.6435, P<0.0001$;四种评分法对应的“行评分均值不同”有两个不同的计算结果,分别为“ $\chi^2=33.7083$ [采用方差分析计算再基于式(6)转换的结果]”与“ $\chi^2=31.2716$ (采用 Kruskal-Wallis's H 秩和检验计算的结果)”,但 P 值都小于 0.0001。

5 讨论与小结

5.1 讨论

在给“scores=”选项指定“table”或其他三种非参数评分法(即 rank、ridit 和 modridit)时,为什么“非零相关”有两个完全相反的结果?原因在于:前者采用的是“Pearson's 相关分析”,而后者采用的是“Spearman's 秩相关分析”。就例 3 资料而言,分别基于“Pearson's 相关分析”与“Spearman's 秩相关分

析”得到“ $r=0.02033$ ”与“ $r_s=0.00568$ ”。依据式(2)可算出对应的 CMH χ^2 值分别为 6.632368(与前面相应的计算结果 6.6350 很接近)和 0.517715(与前面相应的计算结果 0.5174 很接近)。

在“常规关联”的分析中,将有序变量视为无序变量或名义变量,故给“scores=”选项指定四种内容(即 table、rank、ridit 和 modridit)时,所得结果完全相同。就例 3 资料而言, $\chi_{CMH}^2=36.6435$ 。它在本质上就是 $\chi_p^2 = 36.6458$ (即 Pearson's 拟合优度 χ^2 值),由式(7)可知:

$$\frac{16048 - 1}{16048} \times 36.6458 = 36.6435 = \chi_{CMH}^2$$

也就是说, χ_{CMH}^2 等于 χ_p^2 乘以校正系数“(N-1)/N”。

在备择假设为“行评分均值不同”时,SAS/STAT 中 FREQ 过程将此时的资料视为“单因素 R 水平设计一元定量资料”。若采用“表评分(即 scores=table)”,则 SAS/STAT 中 FREQ 过程采用“单因素方差分析”;若采用其他三种评分,SAS/STAT 中 FREQ 过程采用 Kruskal-Wallis's 秩和检验。事实上,就是把结果变量 B 视为“定量变量”,因此,对结果变量 B 的赋值方法不同,将直接影响 R×C 列联表资料各行上结果变量平均值的计算结果。四种评分方法对应结果变量 B 的 3 个评分值如下:

B 的水平	table	rank	ridit	modridit
1	1	5837.5	0.363752	0.363730
2	2	12489.5	0.778259	0.778210
3	3	14676.5	0.914538	0.914481

若直接采用 SAS/STAT 中 ANOVA 过程计算,得到 $F=8.44$ 、 $df=4$,基于式(6)可算得 $\chi_4^2 \approx 8.44 \times 4 = 33.76$,与“scores=table”对应的“行评分均值不同”时的“ $\chi_4^2=33.7083$ ”比较接近;若直接采用 SAS/STAT 中 NPAR1WAY 过程计算,得到 $H=31.2716$,与“scores=rank”对应的“行评分均值不同”时的“ $\chi_4^2=31.2716$ ”

完全相同。

值得一提的是,在分析二维列联表资料时,三种非参数评分法输出的三种检验结果相同;但若是分析高维列联表资料,情况就不尽相同了。因篇幅所限,此处从略。

5.2 小结

本文呈现了三种 R×C 列联表资料的实例及其三种相应的统计分析方法,这些方法可概括成广义 CMH χ^2 检验;基于 SAS/STAT 中 FREQ 过程对 R×C 列联表资料进行分析时,可同时输出三种统计分析结果;本文还详细揭示了选取不同的“评分”方法,将会影响“非零相关”与“行评分均值不同”计算结果的真实原因。

参考文献

- [1] 范蓉馨,杨先梅,黄明金,等.新冠肺炎疫情防控中基层精防医护人员心理健康状况及需求调查[J].四川精神卫生,2020,33(3):207-210.
- [2] 胡良平.科研设计与统计分析[M].北京:军事医学科学出版社,2012:320-354.
- [3] 钟潇,袁东玲,王斌.新冠肺炎疫情期间居民急性应激障碍症状检出情况及影响因素[J].四川精神卫生,2020,33(5):398-402.
- [4] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 2997-3216.
- [5] 程云鹏.矩阵论[M].陕西:西北工业大学出版社,1989:276-288.
- [6] 金丕焕.医用统计方法[M].上海:上海医科大学出版社,1993:239-245.
- [7] 胡良平.面向问题的统计学——(1)科研设计与统计分析[M].北京:军事医学科学出版社,2012:379-394.
- [8] 陈希孺.数理统计引论[M].北京:科学出版社,1981:1-10.
- [9] 方开泰,许建伦.统计分布[M].北京:科学出版社,1987:136-211.

(收稿日期:2021-03-16)

(本文编辑:陈霞)