

## 卡方检验误用辨析

张平原 王开发

西南大学数学与统计学院, 重庆 400715

通信作者: 王开发, Email: kfwang72@163.com

**【摘要】** 计数资料在生物医学研究中十分常见, 此类资料常常被整理为列联表的形式。对计数资料, 常见的研究目的是检验 2 个因素之间是否独立, 因此常用  $\chi^2$  检验来进行统计分析, 却容易忽略  $\chi^2$  检验的适用条件和不同条件下对结果的校正。此外, 对计数资料, 还有检验 2 个分类变量之间是否存在线性变化趋势、结果是否一致等研究目的, 因此如何根据不同的研究目的选择合适的检验方法, 也是实际应用时值得关注的问题。该文通过系列示例比较系统地总结归纳了各类计数资料的常用统计分析方法, 并对常见的误用情况进行了辨析。

**【关键词】** 生物医学研究; 数据说明, 统计; 计数资料; 列联表;  $\chi^2$  检验

### Analysis on the misuse of chi-square test

Zhang Pingyuan, Wang Kaifa

School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Corresponding author: Wang Kaifa, Email: kfwang72@163.com

**【Abstract】** Count data are very common in biomedical research, and such kind of data is often organized in the form of contingency table. For count data, the common research purpose is to test whether two factors are independent, therefore chi-square test is often used for statistical analysis, but it is easy to overlook the applicable conditions of chi-square test and the correction of results under different conditions. In addition, for count data, there are also other research purposes, such as testing whether there is a linear trend between two categorical variables, whether the results are consistent, and so on. Therefore, how to choose appropriate testing methods based on different research purposes is also a problem worth paying attention to in practical applications. In this study, the commonly used statistical methods for various count data are systematically summarized through a series of examples, and the

common misuses of the chi-square test are analyzed.

**【Key words】** Biomedical research; Data interpretation, statistical; Count data; Contingency table; Chi-square test

计数资料在医学研究中十分常见, 此类资料常常被整理为列联表(如  $2 \times 2$  列联表、 $R \times C$  列联表与分层列联表等)的形式。对计数资料, 最常见的研究目的是确定 2 个计数变量之间是否独立或者数据是否符合某个理论分布或期望分布, 因而常用的统计方法是  $\chi^2$  检验<sup>[1-3]</sup>。但是在探讨这些研究目的时,  $\chi^2$  检验本身是有适用条件的<sup>[4]</sup>, 包括因素之间的独立性、样本量的大小、理论频数过小的格子数占总格子数的比例以及因素水平间有无顺序性等<sup>[5-6]</sup>, 所以不考虑该检验的适用条件可能会造成误用, 甚至在某些情形下还可能导致完全相反的统计结论。此外, 根据研究需要, 还有检验 2 个分类变量之间是否存在线性变化趋势<sup>[7-8]</sup>、结果是否一致等研究目的。显然, 不考虑研究目的而对所有计数资料简单地使用  $\chi^2$  检验也可能造成统计方法的误用。因此如何根据不同研究目的选择合适的统计方法, 也是实际应用时值得关注的问题。

本文通过系列示例系统地归纳了计数资料的常用统计检验方法, 并对常见的  $\chi^2$  检验误用情况进行了辨析。需要说明的是, 本文中的所有示例都是人工构建的, 并不代表真实的临床设计和数据, 主要目的是为医学科研工作者根据研究目的选择合适的统计分析方法提供直观示例, 并展示误用  $\chi^2$  检验带来的后果。

### 1 $2 \times 2$ 列联表资料的统计分析

当一个研究设计包括 2 个因素时, 这 2 个因素记为因素 A 和因素 B, 并且当它们都有 2 个水平时, 其数据格式常被整理为  $2 \times 2$  列联表形式。 $2 \times 2$  列联表的数据格式见表 1, 其中  $A_{ij}$  为实际频数,  $T_{ij}$  为理论频数,  $N_i = \sum_{j=1}^2 A_{ij}$ ,  $N_j = \sum_{i=1}^2 A_{ij}$ ,  $N = \sum_{i=1}^2 N_i = \sum_{j=1}^2 N_j$ ,  $i=1, 2, j=1, 2$ 。注意仅当

DOI: 10.3760/cma.j.cn501225-20240103-00004

本文引用格式: 张平原, 王开发. 卡方检验误用辨析[J]. 中华烧伤与创面修复杂志, 2024, 40(5): 482-488. DOI: 10.3760/cma.j.cn501225-20240103-00004.

Zhang PY, Wang KF. Analysis on the misuse of chi-square test[J]. Chin J Burns Wounds, 2024, 40(5): 482-488. DOI: 10.3760/cma.j.cn501225-20240103-00004.



2个因素之间独立时,才存在理论频数。

表1 2×2列联表的数据格式

因素B	因素A		合计
	水平1	水平2	
水平1	$A_{11}(T_{11}或\%)$	$A_{12}(T_{12}或\%)$	$N_{1.}$
水平2	$A_{21}(T_{21}或\%)$	$A_{22}(T_{22}或\%)$	$N_{2.}$
合计	$N_{.1}$	$N_{.2}$	$N$

注: $N$ 为样本总例数, $A_{ij}$ 表示第*i*行第*j*列格子的实际频数, $T_{ij}$ 表示相应第*i*行第*j*列格子的理论频数, $i=1,2,j=1,2$

$\chi^2$ 检验是以 $\chi^2$ 分布为基础的一种假设检验方法。对表1所示的2×2列联表进行统计分析时,应该按照研究设计的类型、样本总量*N*以及各格子的理论频数(若适用)大小来选用合适的统计方法,具体流程见图1。

1.1 完全随机设计2×2列联表

完全随机设计是指将*N*个研究样本随机分到2个或2个以上独立组别而进行的研究,也被称为成组设计。根据图1,针对完全随机设计,须特别注意根据*N*和 $T_{ij}$ 的大小来选用合适的统计方法,否则可能会导致相反的统计结论。

例1 某科研工作者欲探究德尔菲淋巴结阳性的甲状腺乳头状癌患者与德尔菲淋巴结阴性的甲状腺乳头状癌患者的中央淋巴结转移发病比例是否存在差异,收集整理了某一时期内41例甲状腺乳头状癌患者的资料,见表2。

表2 有德尔菲淋巴结与无德尔菲淋巴结的甲状腺乳头状癌患者的中央淋巴结转移结果[例(理论频数)]

德尔菲淋巴结	中央淋巴结转移		合计
	阳性	阴性	
阳性	20(16.82)	10(13.17)	30
阴性	3(6.17)	8(4.82)	11
合计	23	18	41

【辨析】本示例的研究设计类型为完全随机设计,总样本量*N*>40。如果直接使用 $\chi^2$ 检验,经软件计算可得 $\chi^2=5.07, P=0.024$ 。因此,在 $\alpha=0.05$ 的检验水准下,得到德尔菲淋巴结阳性组患者的中央淋巴结转移发病比例(20/30)与德尔菲淋巴结阴性组患者的中央淋巴结转移发病比例

(3/11)的差异有统计学意义( $P<0.05$ )。但是仔细观察表2中的理论频数,可以看到 $T_{22} = 4.82 < 5$ ,所以根据图1,本示例应选择更合适的统计方法—— $\chi^2$ 检验连续性校正。采用该方法,经软件计算可得 $\chi^2=3.60, P=0.058$ ,这意味着在 $\alpha=0.05$ 的检验水准下,尚不能认为德尔菲淋巴结阳性组与德尔菲淋巴结阴性组患者的中央淋巴结转移发病比例的差异有统计学意义( $P>0.05$ )。因此,不合适的统计方法可能会导致相反的统计结论。

例2 某医师使用2种药膏治疗患者皮肤软组织感染,见表3。问这2种药膏的疗效差异是否有统计学意义?

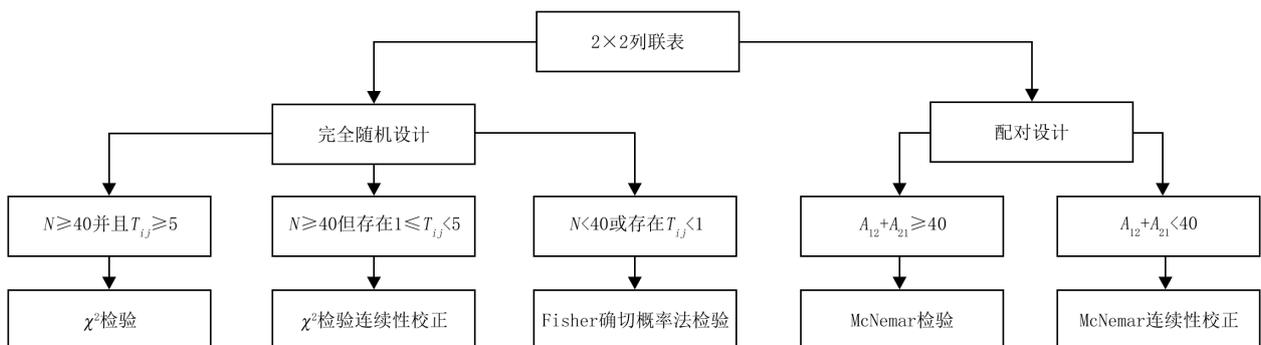
表3 药膏A和药膏B治疗患者皮肤软组织感染的疗效结果[例(理论频数)]

药膏	有效	无效	合计
药膏A	38(36.61)	1(2.39)	39
药膏B	8(9.39)	2(0.61)	10
合计	46	3	49

【辨析】本示例同样属于完全随机设计,总样本量*N*>40。如果直接使用 $\chi^2$ 检验,经软件计算可得 $\chi^2=4.21, P=0.040$ 。因此,在 $\alpha=0.05$ 的检验水准下,可以认为2种药膏对患者皮肤软组织感染的治疗有效比例的差异有统计学意义( $P<0.05$ )。但是因为 $T_{22} = 0.61 < 1$ ,因此相较于直接使用 $\chi^2$ 检验或者 $\chi^2$ 检验连续性校正,更适合的统计方法是进行Fisher确切概率法检验。采用该方法,经软件计算可得 $P=0.102$ ,因此在 $\alpha=0.05$ 的检验水准下,尚不能认为2种药膏对患者皮肤软组织感染的治疗有效比例的差异有统计学意义( $P>0.05$ )。

1.2 配对设计2×2列联表

在医学研究中,为了控制可能存在的非处理因素影响,从而增加组间的可比性,有时会采用配对设计。配对设计主要包括以下4种类型的数据:(1)同一受试对象处理前后的数据;(2)同一受试对象2个部位的数据;(3)同一样品用2种方法检验的数据;(4)配对的2个受试对象分别接受2种处理后的数据。针对配对设计,如果不辨析研究设计类型,直接将配对数据看成独立数据来进行组间比较,可能会导致相反的统计结论。另根据图1,还需要考虑副对角线频数



注: $N$ 为样本总例数, $A_{ij}$ 表示第*i*行第*j*列格子的实际频数, $T_{ij}$ 表示相应第*i*行第*j*列格子的理论频数, $i=1,2,j=1,2$

图1 2×2列联表资料统计检验方法的选择流程图

之和的大小,否则同样可能会导致不同的统计结论。

**例 3** 某科研人员研发了一种新的针对某标志物的检测方法 A,欲探究新的检测方法 A 与传统的检测方法 B 的检测结果是否有差异,采用 2 种检测方法分别对 120 个样本进行检测,结果整理见表 4。

表 4 检测方法 A 与检测方法 B 检测样本的结果(例)

检测方法 A	检测方法 B		合计
	阳性	阴性	
阳性	31	14	45
阴性	40	35	75
合计	71	49	120

**【辨析】** 本示例是对同一个样本分别采用 2 种检测方法进行检测,属于配对设计范畴。如果直接使用随机设计的  $\chi^2$  检验,经软件计算可得  $\chi^2=2.82, P=0.093$ 。因此,在  $\alpha=0.05$  的检验水准下,不能认为 2 种检测方法的检测效果差异有统计学意义 ( $P>0.05$ )。由于本示例属于配对设计,应该选择 McNemar 检验。经过 R 软件计算可得  $\chi^2=12.52, P<0.001$ 。值得注意的是,在 SPSS 统计软件中,尽管副对角线频数  $40+14>40$ ,仍会对 McNemar 检验进行校正<sup>[9]</sup>,SPSS 统计软件运行结果为  $\chi^2=11.57, P=0.001$ 。因此,在  $\alpha=0.05$  的检验水准下,这 2 种检测方法对该标志物的检测效果比较,差异应该是有统计学意义的 ( $P<0.05$ )。

**例 4** 有学者采集患者末梢血与静脉血检测乙型肝炎抗原,欲比较 2 种血液样本的检测总体阳性率有无差异。检测结果见表 5。

表 5 患者末梢血与静脉血乙型肝炎抗原检测结果(例)

静脉血	末梢血		合计
	阳性	阴性	
阳性	25	9	34
阴性	20	83	103
合计	45	92	137

**【辨析】** 本示例资料是对每例患者分别采取静脉血与末梢血进行乙型肝炎抗原检测的结果,属于配对设计范畴。但是如果直接使用 McNemar 检验,经 R 软件计算可得  $\chi^2=4.17, P=0.041$ 。因此,在  $\alpha=0.05$  的检验水准下,2 种血液样本的检测总体阳性率差异有统计学意义 ( $P<0.05$ )。然而观

察表 5 可知其副对角线频数  $9+20<40$ ,因此更合适的统计方法是进行 McNemar 连续性校正。经软件计算可得  $\chi^2=3.45, P=0.063$ 。因此,在  $\alpha=0.05$  的检验水准下,尚不能认为 2 种血液样本的检测总体阳性率差异有统计学意义 ( $P>0.05$ )。因此,针对配对设计数据资料,如果不注意副对角线频数之和的大小,可能会导致相反的统计结论。

### 2 R×C 列联表资料的统计分析

相对于 2×2 列联表资料,R×C 列联表资料指因素 A 与因素 B 中至少有 1 个因素的水平数≥3,数据格式如表 6 所示,其中  $A_{ij}$  为实际频数,  $T_{ij}$  为理论频数,  $N_{i.} = \sum_{j=1}^m A_{ij}, N_{.j} = \sum_{i=1}^n A_{ij}, N = \sum_{i=1}^n N_{i.} = \sum_{j=1}^m N_{.j}, i = 1, 2, \dots, n, j = 1, 2, \dots, m$ 。注意仅当 2 个因素之间独立时,才存在理论频数。

此时,根据指标变量与分组变量的类型,可以将 R×C 列联表划分为 4 种资料类型,分别为双向无序列联表、单向有序列联表、双向有序属性不同列联表以及双向有序属性相同列联表。对于 R×C 列联表的统计分析,需要根据因素之间是否独立、因素水平是否有序、理论频数的大小以及分析的目的来选择合适的检验方法,具体流程如图 2 所示。

#### 2.1 双向无序列联表

双向无序列联表指表 6 中的 2 个因素皆为无序分类变量。在进行多个样本率的比较或者因素间关联性分析时,根据图 2,此时还需要考虑是否有 <1 的理论频数或者理论频数>1 而<5 的格子数占总格子数的比例大小,从而选择合适的统计方法。

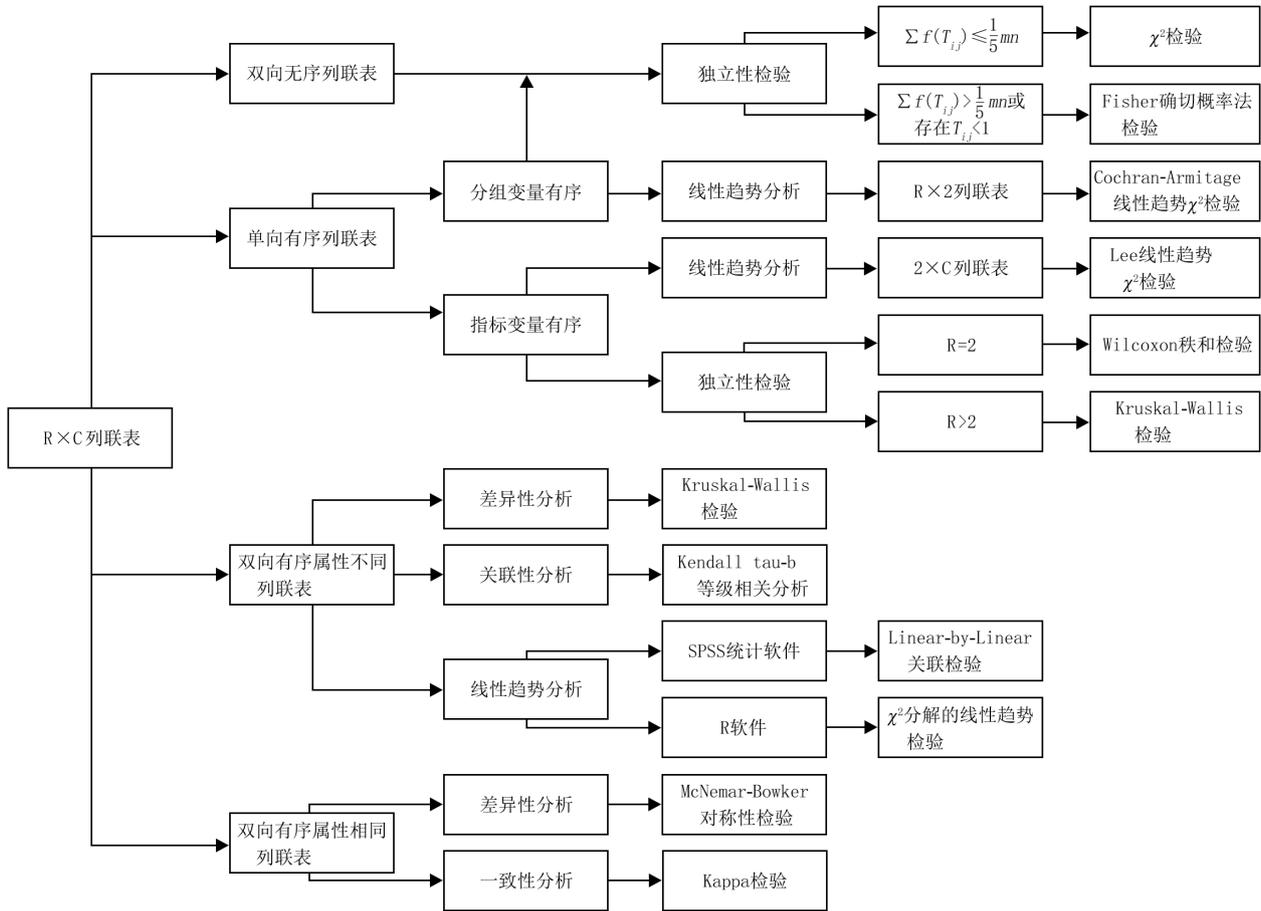
**例 5** 研究者欲研究某癌症 3 个亚种患者的地区分布是否相同,按照 3 个亚种分别抽样,结果见表 7。

**【辨析】** 本示例中癌症类型为分组变量,地区为指标变量,两者都是无序分类变量,因此属于双向无序列联表范畴。如果直接使用  $\chi^2$  检验,经软件计算可得  $\chi^2=10.04, P=0.039$ 。因此,在  $\alpha=0.05$  的检验水准下,可认为 3 个癌症亚种患者的地区分布差异有统计学意义 ( $P<0.05$ )。但观察该列联表每个格子的理论频数,有 4 个格子的理论频数<5,这已经超过了总格子数的 1/5,因此本示例更为恰当的统计方法是进行 Fisher 确切概率法检验。经软件计算可得校正后  $P=0.065$ 。因此,在  $\alpha=0.05$  的检验水准下,尚不能认为 3 个癌症亚种患者的地区分布差异有统计学意义 ( $P>0.05$ )。说明如

表 6 R×C 列联表的数据格式

因素 B (分组变量)	因素 A(指标变量)				合计
	水平 1	水平 2	……	水平 m	
水平 1	$A_{11}(T_{11} \text{ 或 } \%)$	$A_{12}(T_{12} \text{ 或 } \%)$	……	$A_{1m}(T_{1m} \text{ 或 } \%)$	$N_{1.}$
水平 2	$A_{21}(T_{21} \text{ 或 } \%)$	$A_{22}(T_{22} \text{ 或 } \%)$	……	$A_{2m}(T_{2m} \text{ 或 } \%)$	$N_{2.}$
……	……	……	……	……	……
水平 n	$A_{n1}(T_{n1} \text{ 或 } \%)$	$A_{n2}(T_{n2} \text{ 或 } \%)$	……	$A_{nm}(T_{nm} \text{ 或 } \%)$	$N_{n.}$
合计	$N_{.1}$	$N_{.2}$	……	$N_{.m}$	$N$

注: $N$ 为样本总例数, $A_{ij}$ 表示第*i*行第*j*列格子的实际频数, $T_{ij}$ 表示相应第*i*行第*j*列格子的理论频数,  $i=1, 2, \dots, n, j=1, 2, \dots, m$



注:RxC里的R和C分别指的是因素A和因素B的水平数; $T_{ij}$ 为理论频数;如果 $T_{ij} < 5$ ,则 $f(T_{ij}) = 0$ ,否则 $f(T_{ij}) = 1$

图2 RxC列联表资料统计检验方法的选择流程图

表7 某癌症3个亚种患者的地区分布[例(理论频数)]

癌症类型	地区1	地区2	地区3	合计
亚种1	2(3.44)	3(3.93)	10(7.62)	15
亚种2	8(3.67)	3(4.20)	5(8.13)	16
亚种3	4(6.89)	10(7.87)	16(15.25)	30
合计	14	16	31	61

果不注意每个格子理论频数的大小,就可能会得到相反的统计结论。

### 2.2 单向有序列联表

单向有序列联表是指表6中的因素A和因素B中仅有1个因素为有序变量。如果指标变量无序,则可以将其看成双向无序列联表,直接利用 $\chi^2$ 检验来分析2个因素之间的独立性,或者用Cochran-Armitage线性趋势 $\chi^2$ 检验来研究分组变量各水平上的线性变化趋势。如果指标变量有序,并且研究目的是进行独立性检验,此时不能直接使用双向无序列联表的 $\chi^2$ 检验方法,因为这样会损失指标变量顺序的有效信息。当分组变量为2个时,此时可以选择Wilcoxon秩和检验;当分组变量超过2个时,可以选择Kruskal-Wallis检验。另外如果目的是研究线性变化趋势,则宜选用Lee线性趋势 $\chi^2$ 检验。

例6 为了观察3种药物治疗高血压的疗效,即检验3种药物的总体疗效差异是否有统计学意义。研究者采用随机数字表法将患者分为3组,分别使用3种药物,再记录各组患者的疗效,结果见表8。

表8 3种药物治疗患者高血压的疗效情况(例)

药物类型	无效	好转	显效	治愈	合计
药物A	6	7	20	9	42
药物B	5	9	11	22	47
药物C	10	5	8	15	38
合计	21	21	39	46	127

【辨析】 本示例中分组变量为药物类型,有3个水平且无序,指标变量为疗效,分为无效、好转、显效和治愈,为有序变量,因此本示例为单向有序列联表。根据差异性分析目的,本示例宜选用Kruskal-Wallis检验。但由于相持较多,本示例使用校正后的 $H_c$ 值,经软件计算可得 $H_c = 2.83$ , $P = 0.243$ 。因此在 $\alpha = 0.05$ 的检验水准下,尚不能认为3种药物的总体疗效差异有统计学意义( $P > 0.05$ )。如果对本示例直接使用双向无序列联表 $\chi^2$ 检验,经软件计算可得 $\chi^2 = 13.76$ , $P = 0.032$ ,将得到完全相反的统计结论。

**例 7** 某研究者欲探究随着体检者年龄的增长其高血压发病率是否呈现线性增长,整理了某医院近半年的体检资料,得到表 9 所示的数据资料。

**表 9** 不同年龄段体检者的高血压分布(例)

年龄段	有	无	合计
<18 岁	15	85	100
≥18 岁且<35 岁	27	76	103
≥35 岁且<60 岁	75	115	190
≥60 岁	48	56	104
合计	165	332	497

**【辨析】** 本示例中指标变量为是否有高血压,为无序变量。年龄为分组变量且有序,并且研究目的是探究分组变量各水平上的高血压发病率是否存在线性趋势关系。根据图 2,此时宜使用 Cochran-Armitage 线性趋势  $\chi^2$  检验。值得注意的是,SPSS 统计软件中尚未提供该检验,但是提供了可替代的 Linear-by-Linear 关联检验。经过 SPSS 统计软件计算可得  $\chi^2=27.79, P<0.001$ 。采用 R 软件行 Cochran-Armitage 线性趋势  $\chi^2$  检验计算可得,  $Z=5.28, P<0.001$ 。因此,虽然 2 种统计软件提供的计算方法不尽相同,但结果都显示随着年龄的增长,体检者的高血压发病率也随之增长,呈现出线性变化的趋势。

### 2.3 双向有序列联表

双向有序列联表是指表 6 中的分组变量和指标变量都为有序变量,根据两者属性是否相同又可进一步划分为属性相同列联表和属性不同列联表。当  $R \times C$  列联表为双向有序属性相同列联表时,它实际上是配对设计  $2 \times 2$  列联表的扩展。例如,当用 2 种检测方法对同一批样本的测定结果进行比较时,根据研究目的可归纳得相应的检验方法主要有:(1)评价 2 种检测方法是否存在差异,可使用 McNemar-Bowker 对称性检验;(2)评价 2 种方法检测结果的一致性,可使用 Kappa 检验。当  $R \times C$  列联表为双向有序属性不同列联表时,此时根据研究目的可归纳得到相应的检验方法如下:(1)若研究目的为分析不同分组变量的指标变量之间有无差别时,可视为单向有序  $R \times C$  列联表资料,选用 Kruskal-Wallis 检验进行分析;(2)若研究目的为分析 2 个有序分类变量间是否存在相关关系,宜用 Kendall tau-b 等级相关分析;(3)若研究目的为分析 2 个有序分类变量间是否存在线性变化趋势,宜用 Linear-by-Linear 关联检验(SPSS 统计软件提供)或  $\chi^2$  分解的线性趋势检验(R 软件提供)。

**例 8** 某医师使用 2 种检测方法对同一批样本进行检测,结果如表 10 所示。

**【辨析】** 本示例属于配对设计,并且检测结果分为-、+和++,是有顺序的,因此该表属于双向有序属性相同  $3 \times 3$  列联表。(1)如果研究目的是检验这 2 种检测方法是否存在差异,那么宜选用 McNemar-Bowker 对称性检验。经软件计算可得  $\chi^2=8.88, P=0.031$ ,故在  $\alpha=0.05$  的检验水准下,可以认为 2 种检测方法比较,差异有统计学意义( $P<0.05$ )。(2)如果

**表 10** 2 种检测方法的检测结果(例)

检测方法 B	检测方法 A			合计
	-	+	++	
-	19	10	8	37
+	4	16	3	23
++	9	13	21	43
合计	32	39	32	103

研究目的是检验这 2 种检测方法结果的一致性,则宜选用 Kappa 检验。经软件计算可得  $Kappa=0.32, P<0.001$ ,故在  $\alpha=0.05$  的检验水准下,可以认为 2 种方法的检测结果具有一致性。

**例 9** 某医院整理了 1 年内不同年龄段某疾病患者的疗效情况,如表 11 所示。

**表 11** 不同年龄段的某疾病患者疗效情况(例)

年龄段	无效	好转	治愈	合计
<18 岁	8	32	45	85
≥18 岁且<45 岁	4	25	37	66
≥45 岁	11	18	26	55
合计	23	75	108	206

**【辨析】** 本示例中分组变量是年龄段,有 3 个水平,为有序变量;指标变量是疗效,也有 3 个水平,仍为有序变量。但由于这 2 个变量的属性不同,因此属于双向有序属性不同  $3 \times 3$  列联表。(1)如果本示例的研究目的是分析不同年龄之间的疗效构成比差异是否显著,那么本示例与例 6 相同,宜选用 Kruskal-Wallis 检验。经软件计算可得  $\chi^2=2.60, P=0.272$ ,故在  $\alpha=0.05$  的检验水准下,尚不能认为不同年龄之间的疗效差异有统计学意义( $P>0.05$ )。(2)如果本示例的研究目的是检验年龄与疗效之间是否存在相关关系,则宜使用 Kendall tau-b 等级相关分析。经软件计算可得  $r=-0.06, P=0.368$ 。因此,在  $\alpha=0.05$  的检验水准下,尚不能认为年龄与疗效之间的相关关系有统计学意义( $P>0.05$ )。(3)如果本示例的研究目的是分析年龄与疗效之间是否存在线性趋势,则在 SPSS 统计软件中宜使用 Linear-by-Linear 关联检验。经过 SPSS 统计软件计算可得  $\chi^2=1.46, P=0.227$ 。因此,在  $\alpha=0.05$  的检验水准下,尚不能认为年龄与疗效之间的线性趋势有统计学意义( $P>0.05$ )。在 R 软件中,由于指标变量的组别数量  $>2$ ,故 Cochran-Armitage 线性趋势  $\chi^2$  检验不再适用,此时可以采用  $\chi^2$  分解的方法,将列联表  $\chi^2$  检验值分解成线性回归分量与偏离线性回归分量<sup>[2]</sup>,再分别检验它们的显著性并作出统计推断。

### 2.4 分割比较时检验水准的校正

对  $R \times C$  列联表资料,如果差异性分析的结果表明多组 ( $>3$  组)间比较,差异有统计学意义( $P<0.05$ ),那么要探究具体是哪些组间的差异有统计学意义时,需要进行组间的多重比较。当需进行  $k$  次组间的多重比较时,如果每次比较时犯第一类错误的概率都被控制在给定的检验水准  $\alpha$  上,那

么当这些检验独立进行时,理论上可得 $k$ 次多重比较后犯第一类错误的总概率将为 $1 - (1 - \alpha)^k \xrightarrow{k \rightarrow \infty} 1$ 。因此,需要对检验水准 $\alpha$ 进行校正。最常用的校正方法是 Bonferroni 校正,即设置新的比较检验水准为 $\alpha' = \alpha \div k$ 。Bonferroni 校正是两两比较方法中最为保守的,适用于所有的两两比较,包括多个均数比较和多个频率比较。当比较的次数 $k$ 不大时,该校正方法效果较好。但是,当比较次数 $k$ 较大时,例如 $k > 10$ ,由于此时选择的新检验水准 $\alpha'$ 很小,结论将偏于保守。

**例 10** 有学者欲比较 3 种中药方剂对腰椎间盘突出患者的疗效是否有差异,将患者分为 3 组,分别使用 3 种中药方剂进行治疗,结果见表 12。

**表 12** 3 种中药方剂对腰椎间盘突出患者的疗效  
[例(理论频数)]

中药方剂类型	疗效		合计
	有效	无效	
方剂 A	24(28.56)	20(15.44)	44
方剂 B	20(22.07)	14(11.93)	34
方剂 C	30(23.37)	6(12.63)	36
合计	74	40	114

**【辨析】** 本示例中分组变量为中药方剂,为无序变量;指标变量为疗效,分别为有效和无效,仍然为无序变量;因此本示例的数据资料属于双向无序列联表。根据图 2,直接使用 $\chi^2$ 检验。经软件计算可得 $\chi^2=7.99, P=0.018$ 。因此,在 $\alpha=0.05$ 的检验水准下,认为 3 种中药方剂的疗效差异有统计学意义( $P < 0.05$ )。进一步如果想知道具体是哪些组之间的差异有统计学意义,就需要进行 3 次两两比较,比较结果见表 13。

**表 13** 3 种中药方剂治疗腰椎间盘突出患者的总体有效率间的两两比较[例(%)]

对比组	2×2 列联表		$\chi^2$ 值	$P$ 值
方剂 A 组与方剂 B 组	24(54.55)	20(45.45)	0.14	0.706
	20(58.82)	14(41.18)		
方剂 A 组与方剂 C 组	24(44.44)	20(55.56)	7.48	0.006
	30(83.33)	6(16.67)		
方剂 B 组与方剂 C 组	20(58.82)	14(41.18)	5.15	0.023
	30(83.33)	6(16.67)		

如果直接采用原始的检验水准 $\alpha=0.05$ ,由表 13,可得出方剂 C 组与方剂 A 组和方剂 B 组患者的疗效比较,差异均有统计学意义( $P < 0.05$ )。但是,若采用 Bonferroni 校正对检验水准进行校正,可得校正后的检验水准为 $\alpha'=0.05 \div 3=0.017$ 。表 13 中结果仅显示方剂 A 组与方剂 C 组患者的疗效差异有统计学意义( $P < 0.017$ )。因此,如果不进行检验水准的校正,将会得到方剂 B 组与方剂 C 组患者的疗效差异也有统计学意义( $P < 0.05$ )的虚假统计结论。

### 3 分层数据的相关性检验

无论是 2×2 列联表还是 R×C 列联表,它们都只有 2 个因素变量。但在某些研究设计中,除了需考虑的因素变量之外,可能还存在混杂变量。此时该计数资料就需要被整理成分层列联表的形式,对其进行检验通常需使用 Cochran-Mantel-Haenszel  $\chi^2$  检验(CMH  $\chi^2$  检验)。

**例 11** 有学者欲探究高血压对患者心肌梗死的影响,采用病例对照研究方法调查了纳入 586 例心肌梗死患者的病例组与纳入 530 例未发生过心肌梗死患者的对照组的高血压情况。考虑到年龄是混杂因素,将年龄按照 35 岁进行分层后,整理得到如表 14 的分层 2×2 列联表。拟分析排除了年龄因素的影响后,病例组和对照组患者的总体高血压发病率比较,差异是否有统计学意义。

**表 14** 按年龄分层的心肌梗死与高血压关系的病例对照研究结果(例)

年龄段与分组	高血压		合计
	是	否	
<b>&lt;35 岁</b>			
病例组	155	53	208
对照组	88	43	131
小计	243	96	339
<b>≥35 岁</b>			
病例组	213	165	378
对照组	145	254	399
小计	358	419	777
合计	601	515	1 116

**【辨析】** 本示例的研究目的是在排除年龄的混杂因素之后,探究心肌梗死是否与高血压有关。利用软件首先进行了 Breslow-Day 检验对各层的比值比进行齐性检验,结果得到 $P=0.107$ 。因此,在 $\alpha=0.05$ 的检验水准下,2 个年龄组患者中高血压对心肌梗死的总体比值比同质,即年龄分层因素与发生心肌梗死之间不存在交互作用,说明此时运用 CMH  $\chi^2$  检验结果是可靠的。若 Breslow-Day 检验拒绝齐性假设,则说明年龄分层因素与发生心肌梗死之间存在交互作用,此时 CMH  $\chi^2$  检验的结果不能说明问题,需要选用其他统计方法,比如多因素 logistic 回归分析。值得注意的是,虽然 SPSS 统计软件与 R 软件均提供了分层  $\chi^2$  检验,但它们执行的具体计算公式有稍许差异。本示例的计算结果分别为 Cochran  $\chi^2=31.01$  (来自 SPSS 统计软件), CMH  $\chi^2=30.96$  (来自 R 软件),但都有 $P < 0.001$ 。这都表明在控制了年龄的影响后,病例组与对照组患者的总体高血压概率是不同的,即心肌梗死与高血压有关。

### 4 结语

总之,在对列联表资料进行检验之前,首先要明确研究设计类型、研究目的和样本量的大小,然后再参考图 1 和图 2 选择合适的检验方法。希望本文的示例可以为生物医学

科论文中的计数资料统计分析提供帮助和参考。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 方积乾. 生物医学研究的统计方法[M]. 2版. 北京: 高等教育出版社, 2019.

[2] 孙振球, 徐勇勇. 医学统计学[M]. 4版. 北京: 人民卫生出版社, 2014.

[3] 贾俊平, 何晓琴, 金勇进. 统计学[M]. 7版. 北京: 中国人民大学出版社, 2018.

[4] 冯国双. 分类资料组间比较的思路及误区分析[J]. 中华全科医师杂志, 2017, 16(6): 490-492. DOI: 10.3760/cma.j.issn.1671-7368.2017.06.022.

[5] Yates F. Contingency tables involving small numbers and

the  $\chi^2$  test[J]. Supp J R STAT SOC, 1934, 1(2): 217-235. DOI: 10.2307/2983604.

[6] Fisher RA. The logic of inductive inference[J]. J R STAT SOC, 1935, 98(1): 39-82. DOI: 10.2307/2342435.

[7] 吴学森, 王洁贞. 双向有序分类资料线性趋势分析方法的前提条件[J]. 中国卫生统计, 2003, 20(2): 79-81. DOI: 10.3969/j.issn.1002-3674.2003.02.005.

[8] 胡纯严, 胡良平. 如何正确运用 $\chi^2$ 检验——两种单向有序二维列联表资料线性趋势 $\chi^2$ 检验[J]. 四川精神卫生, 2021, 34(2): 116-120. DOI: 10.11886/scjsws20210316004.

[9] 金英良, 赵华硕, 黄水平, 等. 配对四格表资料的SPSS软件实现及相关结果解读[J]. 预防医学论坛, 2017, 23(3): 171-173, 179. DOI: 10.16406/j.pmt.issn.1672-9153.2017.03.004.

(收稿日期: 2024-01-03)

· 消息 ·

中华医学会杂志社学术期刊出版服务平台已经启用

为适应期刊网络采编技术的不断发展,更好地为广大作者、读者提供高质量服务,中华医学会杂志社学术期刊出版服务平台(简称新版采编平台)已于2023年7月18日正式启用,网址:<https://medpress.yiigle.com>。敬请广大作者登录新版采编平台进行投稿。中华医学期刊网、中华医学期刊APP、中华医学全文数据库、各刊官网及中华医学会杂志社远程稿件管理系统用户等可直接登录新版采编平台,无须重新注册。

敬请广大作者尽快在新版采编平台进行登录测试,确认您的手机号码、Email等联系方式是否正确有效。根据网络信息安全相关管理规定,各刊编辑部不再提供协助修改Email等服务。

旧版采编系统(中华医学会杂志社远程稿件管理系统)已关闭投稿通道。所有新投稿敬请在新版采编平台投递。

您可在新版采编平台首页左下方下载操作指南,或通过入口页右下角网页客服系统进行在线咨询,详情请见《中华烧伤与创面修复杂志》微信公众号发布的相关通知。



中华医学会杂志社

广告目次

辽宁味邦生物制药有限公司	封二
南海朗肽制药有限公司	对中文目次1
保赫曼(上海)贸易有限公司	对中文目次2
上海腾瑞制药股份有限公司	对英文目次1
浙江医学科技开发有限公司	对英文目次2
珠海亿胜生物制药有限公司	封三
武汉维斯第医用科技股份有限公司	封底