

如何正确运用 χ^2 检验——队列设计 四格表资料的 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】 本文目的是介绍队列设计四格表资料的 χ^2 检验、相对危险度的假设检验和区间估计、SAS 以及 R 软件的实现。针对队列设计的特点, 重点介绍相对危险度的概念、假设检验和区间估计方法。对 SAS 与 R 软件计算的结果作出解释, 并给出统计和专业结论。

【关键词】 队列设计; χ^2 检验; 相对危险度; 置信区间; SAS 软件; R 软件

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20210115005

How to use χ^2 test correctly—— χ^2 test for the data of four-fold tables collected from the cohort design

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 χ^2 test for the data of four-fold tables collected from the cohort design, the hypothesis test and interval estimation of relative risk, and the implementation of SAS and R software. In view of the characteristics of cohort design, the concept of relative risk, hypothesis testing and interval estimation were emphasized. It was given that the interpretation of the results calculated by SAS and R software, as well as the statistical and professional conclusions.

【Keywords】 Cohort design; χ^2 test; Relative risk; Confidence interval; SAS software; R software

在流行病学研究中, 为了获得精准的研究结果, 研究者首选“队列研究设计方法”。因为通过对大样本的队列人群进行长期观察和随访, 不仅可以获得暴露组与非暴露组人群结局变量的最终取值, 而且可以获得两组人群各种非处理因素及其取值。所以, 队列设计属于“由因果果”的前瞻性研究方法, 其结果的可靠性和准确性高。若假定所有非处理因素在“暴露组”与“非暴露组”之间处于“均衡状态”, 而且结果仅分为“患某种病”与“未患某种病”, 此时, 基于队列设计所收集的资料就可以简化成“队列设计四格表资料(参见下文表1和表2)”。对其进行统计分析时, 包括两个方面: 其一, 检验“暴露与否”与“患病与否”之间是否存在关联性(所需要的统计分析方法与处理横断面设计四格表资料的统计分析方法相同); 其二, 检验“相对危险度(简称为RR)是否等于1”。本文着重介绍相对危险度RR的概念、对总体RR的Mantel-Haenszel's χ^2 检验(简称MH χ^2 检验)、对总体RR的区间估计方法以及用SAS和R软件实现计算的方法。

1 队列设计的概念及其四格表资料的实例

1.1 概念

所谓“队列设计”, 就是依据专业知识, 通过对不同暴露水平的对象进行追踪观察, 确定其疾病发生情况, 从而分析暴露因素与疾病发生之间的因果关系, 它是由因溯果的前瞻性研究设计, 是论证疾病因果关系的一个重要研究方法。

通常, 依据专业知识, 将符合研究目的的总体按某个原因变量的两个水平(例如暴露、未暴露)划分成两组(即“形成两个队列”), 研究者对两组受试对象进行长时间的“追踪观察”, 然后, 再依据定性的结果变量的两个水平(例如发病、未发病)将各组受试对象进一步划分成两部分。一般放在横向的两行代表可疑的两个“原因”(即原因变量的两个水平), 而放在纵向的两列代表研究者所关心的两个“结果”(即结果变量的两个取值), 于是, 就将全部观察对象分成了4个小组, 清点出各小组的个体数

目(即频数);先观察到“原因”,后观察到“结果”,其间隔较长(往往需要一年、甚至几年时间)。队列设计四格表资料的表达格式见表 1。

表 1 n 个受试对象队列研究的结果

原因变量 A	例数			
	结果变量 B:	发病	未发病	合计
暴露		a	b	e=a+b
未暴露		c	d	f=c+d
合计		g=a+c	h=b+d	n=a+b+c+d

注:a、b、c、d 分别代表“暴露组发病人数”“暴露组未发病人数”“未暴露组发病人数”和“未暴露组未发病人数”;e、f 分别代表“第 1 行合计人数”与“第 2 行合计人数”;g、h 分别代表“第 1 列合计人数”与“第 2 列合计人数”;n 代表四格表中的“总人数”

1.2 实例

【例 1】在文献[1]中,作者根据 MINI 5.0 中文版自杀模块判断未治疗抑郁障碍患者有无自杀风险,共 6 个条目,评分范围 0~33 分。评分 <6 分为无自杀风险,评分 ≥6 分为有自杀风险。再将两组受试者分别按“性别”“婚姻状况”“有无精神障碍家族史”和“是否伴有精神病性症状”进行划分,可得 4 个“四格表资料”,现将其中一个列在表 2 中。问:(1)未治疗抑郁障碍患者“有无自杀风险”与“有无精神病性症状”之间是否存在关联性?(2)“有自杀风险者的精神病性症状发生率”是“无自杀风险者的精神病性症状发生率”的多少倍?

表 2 未治疗抑郁障碍患者“有无自杀风险”与“有无精神病性症状”之间关系的调查结果

有无自杀风险	例数			
	有无精神病性症状:	有	无	合计
有		11	41	52
无		4	61	65
合计		15	102	117

注:假定收集资料时,先将全部受试对象按“有无自杀风险”分为“有,52 例”“无,65 例”,追踪观察后,再将每组受试对象分成“有精神病性症状”与“无精神病性症状”,分别计数

【统计分析方法的选择】回答第 1 个问题,可选用“Pearson’s χ^2 检验”“校正的 Pearson’s χ^2 检验”“似然比 χ^2 检验”或“Fisher’s 精确检验”。因为这些检验方法所对应的检验假设均为“两属性变量互相独立或无关联性”。回答第 2 个问题,需要先计算“相对危险度 RR”,然后,选用“Mantel-Haenszel’s χ^2 检验”,该检验的检验假设为:“ $H_0: RR = 1, H_1: RR \neq 1$ ”。因为 MH χ^2 检验所对应的检验假设为“相对危险度是否等于 1”。

【例 2】在文献[2]中,研究者对拟出院的患者采取两种不同的方式来决定出院与否。对照组由医

生评估后下达医嘱办理出院;研究组由医生评估后下达医嘱,并接受出院准备度量表评定,评估达标后再办理出院。记录两组患者在出院后 12 个月时“有无复发且再次入院”情况,现将结果呈现在表 3 中。问:(1)“组别”与“是否复发且再次入院”之间是否存在关联性?(2)“对照组复发且再次入院的发生率”是“研究组复发且再次入院的发生率”的多少倍?

表 3 两组患者出院 12 个月时“有无复发且再次入院”的调查结果

组别	例数			
	*	有	无	合计
对照组		14	16	30
研究组		5	25	30
合计		19	41	60

注:*代表有无复发且再次入院

【统计分析方法的选择】参见表 2 之后“统计分析方法的选择”,此处从略。

1.3 独立性或关联性的检验方法

虽然队列设计四格表资料中的两个变量有“原因变量”与“结果变量”之分,但在回答两变量之间是否存在“关联性”或“独立性”时,仍可将其视为“两属性变量”之间关系的研究问题,故可采用的统计分析方法与横断面设计四格表资料是完全相同的,具体方法此处不再赘述。

2 相对危险度的概念、假设检验与区间估计

2.1 概念

相对危险度(RR)是两个人群发病率之比值^[3-4],这两个人群分别受到一个可疑危险因素两个水平(例如暴露与非暴露)的影响。具体地说,RR 是可疑危险因素处于暴露水平时的发病率为其处于非暴露水平时的发病率的一个倍数,其数值的大小反映的是因素的暴露水平相对于其非暴露水平对发病影响的方向与大小。在 $RR > 1$ 的情况下,RR 越大说明暴露水平对发病的促进作用越大;在 $RR < 1$ 的情况下,RR 越小说明暴露水平对发病的保护作用越大。RR 越接近于 1 说明暴露水平对发病的作用越不明显。

$RR = (\text{暴露者发病的概率}) / (\text{非暴露者发病的概率})$,基于表 1 中所设定的符号,可按下式计算:

$$RR = \frac{a/(a+b)}{c/(c+d)} = \frac{a(c+d)}{c(a+b)} \quad (1)$$

【前提条件】两个人群分别处于可疑危险因素“暴露”与“非暴露”水平之下；每个人群都有足够大的样本含量（两个人群的样本含量之差越小越好）；应尽可能保证在追踪观察的整个时期内，两个人群所受到其他一切非试验因素的影响是几乎相等的（两个人群之间在一切重要非试验因素方面越均衡越好）。

2.2 假设检验

RR 是一个一般的样本统计量，存在抽样误差，要想知道总体中的 RR 是否为 1，就需要对其进行假设检验，即

$$H_0: RR = 1; H_1: RR \neq 1 \quad (2)$$

对式(2)进行假设检验所需要的检验统计量为 Mantel-Haenszel's χ^2 ，一般记为 χ_{MH}^2 ，见下式：

$$\chi_{MH}^2 = \frac{(n-1)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (3)$$

2.3 区间估计

2.3.1 概述

由于 RR 是一个一般的样本统计量，通常，人们需要知道与其对应的总体参数所在的范围，这就是总体相对危险度的区间估计问题。在 SAS/STAT 的 FREQ 过程中^[5]，给出了 3 种置信区间的计算方法，即“渐近置信区间法（简称 WALD 法）”“评分置信区间法”和“精确非条件置信区间法”。因篇幅所限，以下仅对“渐近置信区间法”进行介绍^[4-5]。

2.3.2 渐近置信区间法

RR 的渐近 $100(1-\alpha)\%$ 置信区间的计算公式如下：

$$\left[RR_1 \times \exp(-z\sqrt{v}), RR_1 \times \exp(z\sqrt{v}) \right] \quad (4)$$

在式(4)中， RR_1 是基于队列设计四格表资料中第 1 列数据计算的相对危险度； z 是标准正态分布曲线下左侧面积为 $100(1-\alpha/2)\%$ 处横坐标轴上的分位数；而 v 为对数相对危险度 RR_1 的方差，见下式：

$$v = \text{Var}(\ln RR_1) = \frac{1-p_1}{a} + \frac{1-p_2}{c} \quad (5)$$

在式(5)中， a 与 c 分别是表 1 中第 (1,1) 格与第 (2,1) 格中的观察频数；而 p_1 与 p_2 分别是表 1 中第 1 行与第 2 行上的发病率（注意：第 1 列为“发病”、第 2 列为“未发病”）。

3 队列设计四格表资料统计分析的软件实现

3.1 问题与数据

【例 3】沿用例 1 中的“问题与数据”，试完成下列 4 项任务：(1) 检验两变量（即“有无自杀风险”与“有无精神病性症状”）之间的关联性；(2) 计算相对危险度 (RR)；(3) 对 RR 进行假设检验，即“ $H_0: RR=1$ ； $H_1: RR \neq 1$ ”；(4) 对 RR 进行区间估计。

3.2 基于 SAS 实现关联性或独立性分析

设所需要的 SAS 程序如下^[5-6]：

```
data a;
do a=1 to 2;
do b=1 to 2;
input f @@;
output;
end;
end;
cards;
11 41
4 61
;
run;
proc freq; /*过程步 1*/
weight f;
tables a*b / chisq relrisk;
run;
proc freq; /*过程步 2*/
weight f;
tables a*b / chisq relrisk;
exact chisq relrisk;
run;
proc freq; /*过程步 3*/
weight f;
tables a*b / relrisk(cl=wald);
run;
proc freq; /*过程步 4*/
weight f;
tables a*b / relrisk(cl=score);
run;
proc freq; /*过程步 5*/
weight f;
tables a*b / relrisk(cl=exact);
```

exact relrisk;

run;

【程序说明】第 1 个过程步对四格表资料进行“关联性或独立性”检验(包括 χ^2_P 检验、校正 χ^2_{CP} 检验、 χ^2_{LR} 检验和 Fisher's 精确检验);计算 RR , 还包括对 RR 是否等于 1 的 χ^2_{MH} 检验。第 2 个过程步对四格表资料进行“关联性或独立性”检验(包括 χ^2_P 检验及其精确检验、校正 χ^2_P 检验、 χ^2_{LR} 检验及其精确检验和 Fisher's 精确检验);计算 RR , 还包括对 RR 是否等于 1 的 χ^2_{MH} 检验及其精确检验。第 3 个过程步对四格表资料进行“ RR 计算”, 并基于“渐近置信区间法(即 WALT 法)”求“ RR 的 95% 置信区间”。第 4 个过程步对四格表资料进行“ RR 计算”, 并基于“评分法”求“ RR 的 95% 置信区间”。第 5 个过程步对四格表资料进行“ RR 计算”, 并基于“精确非条件置信区间法(简称精确法)”求“ RR 的 95% 置信区间”。

【SAS 主要输出结果及解释】

第 1 部分, “关联性或独立性”检验的结果, 见表 4。

表 4 例 1 中队列设计四格表资料关联性或独立性检验结果

检验统计量或方法	值	渐近概率	精确概率
χ^2_P	5.8156	0.0159	0.0242
χ^2_{CP}	4.5510	0.0329	-
χ^2_{LR}	5.8966	0.0152	0.0242
Fisher's 法	-	-	0.0242

第 2 部分, 相对危险度的计算结果及其假设检验的结果如下:

$RR=3.4375$, $\chi^2_{MH}=5.7659$, 渐近概率=0.0163、精确概率=0.0242。

第 3 部分, 基于 3 种方法估计总体相对危险度 RR 的 95% 置信区间的结果如下:

渐近法: [1.1619, 10.1702]; 评分法: [1.2280, 9.8313]; 精确法: [0.0885, 69248.3206]。

【结论】由表 4 结果可知, “有无自杀风险”与“有无精神病性症状”之间是不独立的, 具体地说, “有自杀风险”的患者更容易出现“精神病性症状”; “有自杀风险”者是“无自杀风险”者出现“精神病性症状”的 3.4375 倍, 总体相对危险度 RR 的 95% 置信区间为“[1.1619, 10.1702] (渐近法)”“[1.2280, 9.8313] (评分法)”和“[0.0885, 69248.3206] (精确法)”。

3.3 基于 R 实现关联性或独立性分析以及 RR 的计算、假设检验与区间估计

设所需要的 R 程序如下^[6-7]:

```
>rownum<- c(11,41)
>colnum<- c(4,61)
>chisq.test(rbind(rownum,colnum))
>relativerisk<-function(a,b,c,d,alpha)
{
  n=a+b+c+d
  e=a+b
  f=c+d
  g=a+c
  h=b+d
  p1=a/e
  p2=c/f
  rr=p1/p2
  r2=(a*d-b*c)^2/(e*f*g*h)
  chisqmh=(n-1)*r2
  prob_mh=1-pchisq(chisqmh,1,nep=0)
  q=1-alpha/2
  z1_alpha=qnorm(q,mean=0,sd=1)
  v1=(1-p1)/a
  v2=(1-p2)/c
  rv=sqrt(v1+v2)
  i1=exp(-z1_alpha*rv)
  i2=exp(z1_alpha*rv)
  lowerci=rr*i1
  upperci=rr*i2
  return(list(rr=rr,lowerci=lowerci,upperci=upperci,
  chisqmh=chisqmh,prob_mh=prob_mh))
}
>relativerisk(11,41,4,61,0.05)
```

【程序说明】“>”代表 R 软件运行环境中的“提示符”, 上面的 R 程序中共有 5 个提示符, 说明共有 5 个 R 语句; 第一句组织第一个“行向量”, 代表四格表资料的第 1 行数据; 第二句组织第二个“行向量”, 代表四格表资料的第 2 行数据; 第三句调用函数“chisq.test()”进行 χ^2 检验, 没有选项“correct=FALSE”, 系统默认值为“校正 Pearson's χ^2 检验”, 等价于写“correct=TRUE”; 而选项“rbind()”是将两个行向量合并在一起, 即构成两行两列的四格表数据。第四句目的是创建一个自定义函数 function(a,b,c,d,alpha), 并将其赋值给变量 relativerisk。自定义函数有五个参数, 分别为四格表中的四个频数, 第五个参数为双侧假设检验标准正态分布曲线下两尾端的概率之和为 alpha; 第五句是调用 relativerisk 所代表的自定义函数, 计

算相对危险度并对其进行假设检验和区间估计。

【R 主要输出结果及解释】

Pearson's Chi-squared test with Yates' continuity correction

data: rbind(rownum, colnum)

X-squared=4.551, df=1, p-value=0.0329

以上为进行了连续性校正的 Pearson's χ^2 检验结果, 即 $\chi_{cp}^2=4.551, P=0.0329$ 。

其他结果摘要汇总如下:

求得相对危险度 $RR=3.4375$ 、其 95% 置信区间为 $[1.1619, 10.1702]$; 检验总体 RR 是否等于 1 对应的 $\chi_{mu}^2=5.765882, P=0.016340$ 。

以上 R 输出结果与 SAS 输出的“渐近法”的结果是相同的。

【结论】参见前文 SAS 输出结果及结论, 此处从略。

4 讨论与小结

4.1 讨论

在 SAS/STAT 的 FREQ 过程中^[5], 增加了一些新的方法求总体相对危险度的置信区间, 其中, 精确非条件置信限的计算方法计算出来的上限值特别大, 其原因有待进一步查证。建议读者在实际应用中选用渐近法或评分法为宜。

4.2 小结

本文交代了队列设计的概念、呈现了队列设计四格表资料的实例和模式, 给出了相对危险度的概念、计算公式和区间估计方法; 基于 SAS 和 R 软件实现了对队列设计四格表资料的多种假设检验, 对软件输出的结果做出了解释, 并给出了统计和专业结论。

参考文献

- [1] 徐海婷, 刘嫣然, 吕婧, 等. 未治疗抑郁障碍患者自杀风险与认知情绪调节策略的关系 [J]. 四川精神卫生, 2020, 33(1): 44-48.
- [2] 范利芳, 陈岚, 李姝. 出院准备度评估在首发精神分裂症患者中的应用 [J]. 四川精神卫生, 2020, 33(3): 241-243.
- [3] 颜虹. 医学统计学 [M]. 北京: 人民卫生出版社, 2008: 95-100, 163-166.
- [4] 伯纳德·罗斯纳. 生物统计学基础 [M]. 孙尚拱, 译. 北京: 科学出版社, 2004: 551-579.
- [5] SAS Institute Inc. SAS/STAT®15.1 user's guide [M]. Cary, NC: SAS Institute Inc, 2018: 2997-3216.
- [6] 胡良平. 现代医学统计学 [M]. 北京: 科学出版社, 2020: 244-257.
- [7] 约瑟夫·阿德勒. R 语言核心技术手册 [M]. 2 版. 刘思喆, 李舰, 陈钢, 等译. 北京: 电子工业出版社, 2014: 410-416.

(收稿日期: 2021-01-15)

(本文编辑: 戴浩然)