

医学论文中 χ^2 检验的误用及案例分析*

韩宏志 官鑫 李欣欣 姜瑾秋 王丽†

《吉林大学学报(医学版)》编辑部,130021,长春

摘要 医学论文中错误应用 χ^2 检验分析定性资料时,主要是误用 χ^2 检验处理单向有序(分组变量无序、指标变量有序)列联表资料和双向有序属性不同 $R \times C$ 列联表资料。选取公开发表的 2 篇 χ^2 检验误用的案例,辨识存在的错误,指出正确用法,为医学编辑处理类似稿件提供参考和借鉴。

关键词 χ^2 检验;定性资料;列联表资料;单向有序;双向有序

Misuse of chi-square test in medical papers and case analysis//

HAN Hongzhi, GUAN Xin, LI Xinxin, JIANG Jinqiu, WANG Li

Abstract The misuses of chi-square test in dealing with the qualitative data in medical papers mainly include unidirectional ordering contingency table data (grouping variables were disordered, while and target variables were ordered), and bidirectional ordering $R \times C$ contingency table data. By taking two published articles for example, the misuses are analyzed and the correct use of chi-square test in dealing with the qualitative data is suggested, in order to provide the reference for medical editors in dealing with similar problems.

Keywords chi-square test; qualitative data; contingency table data; unidirectional ordering; bidirectional ordering

Authors' address Editorial Board of Journal of Jilin University (Medicine Edition), 130021, Changchun, China

DOI:10.16811/j.cnki.1001-4314.2018.03.012

医学论文中统计资料通常可按性质划分为定量资料和定性资料。定性资料通常又根据其取值的个数及取值之间的关系划分为二分类定性资料(如成功与否,是否患有某种疾病)、多分类名义定性资料(如血型、民族和季节等)和多分类有序定性资料(也常被称为等级资料,如癌症分期为“早、中和晚期”;疗效分为“治愈、显效、好转、无效和死亡”)。在统计学处理中,定性资料统计方法的选择主要考虑 2 个方面因素:一

是定性变量的性质,有序分类变量还是无序分类变量;二是研究的目的。有序资料可以根据结果变量或者原因变量有序与否分为单向有序、双向有序但属性不同及双向有序但属性相同等 3 类。 χ^2 检验是定性资料统计分析中最常用方法,但 χ^2 检验被误用情况也是最常见的,常误将 χ^2 检验视为分析列联表资料的“万能工具”,而不考虑 χ^2 检验使用的前提条件及应用范围^[1-2]。本研究通过列举医学期刊公开发表的 2 篇 χ^2 检验误用的案例,说明误用 χ^2 检验处理单向有序(分组变量无序、指标变量有序)列联表资料和双向有序属性不同列联表资料辨识方法,指出 χ^2 检验正确使用方法,供医学编辑同人借鉴。

1 误用 χ^2 检验处理单向有序(分组变量无序、指标变量有序)列联表资料

例 1 探讨自适应生物反馈训练对出口梗阻型便秘患者临床症状疗效、心理状况和生命质量的改善程度。206 例 OCC 患者分成自适应生物反馈训练(ABF)组和固定式生物反馈训练(FBF)组,采用 χ^2 检验比较 2 组患者临床疗效及满意情况(表 1)^[3]。

辨析 该数据是 $2 \times C$ 列联表资料,结果(疗效)变量“临床疗效和满意情况”为有序变量的单项有序列联表资料,采用 χ^2 检验分析 2 组患者治疗效果和临床满意情况,只能推断 2 组患者疗效不同等级的频数分布(构成比)是否有差别,而频数分布有差别不等同于 2 组患者疗效总体水平有差别,不能得出 2 组疗效的差别是否有统计学意义的结论,并且损失了有序指

表 1 2 组临床疗效与满意情况比较

例(率)

分组	例数	临床疗效				总有效率	满意情况				总满意
		显著	较好	一般	差		非常满意	满意	一般	不满意	
ABF 组	76	39 (51.3%)	23 (30.3%)	10 (13.2%)	4 (5.3%)	72 (94.7%)	35 (46.1%)	22 (28.9%)	9 (11.8%)	10 (13.1%)	66 (86.8%)
FBF 组	62	29 (46.8%)	17 (27.4%)	7 (11.3%)	9 (14.5%)	53 (85.5%)	26 (41.9%)	17 (27.4%)	8 (12.9%)	11 (17.7%)	51 (82.2%)
χ^2 值	—	0.282	0.134	0.110	3.426	3.426	0.235	0.039	0.036	0.556	0.556
P 值	—	0.360	0.431	0.474	0.048	0.048	0.378	0.498	0.526	0.305	0.305

注:ABF 为自适应生物反馈训练;FBF 为固定式生物反馈训练。“—”为无此值。

* 吉林省卫生计生委 2017 年科技能力提升项目(2017G014,2017G015)

† 通信作者

标提供的“等级”信息。该例中在分析过程中将原始资料拆成多个四格表,进行了多次 χ^2 检验,但未对检验水准进行校正,增加了犯I类错误的概率。根据研究目的,该资料不适合选择 χ^2 检验,应采用Wilcoxon秩和检验或Ridit分析计算2组患者临床疗效及满意情

况是否一致。

笔者依据例1研究目的采用Wilcoxon秩和检验对数据重新进行统计学分析,结果显示:2组患者临床疗效和满意情况比较差异均无统计学意义($Z = -0.941, P = 0.347; Z = -0.700, P = 0.484$)(表2)。

表2 2组临床疗效与满意情况比较

例(率/%)

分组	例数	临床疗效				总有效率	满意情况				
		显著	较好	一般	差		非常满意	满意	一般	不满意	总满意
ABF组	76	39(51.3)	23(30.3)	10(13.2)	4(5.3)	72(94.7)	35(46.1)	22(28.9)	9(11.8)	10(13.1)	66(86.8)
FBF组	62	29(46.8)	17(27.4)	7(11.3)	9(14.5)	53(85.5)	26(41.9)	17(27.4)	8(12.9)	11(17.7)	51(82.2)
Z值	—	-0.941					-0.700				
P值	—	0.347					0.484				

注:ABF为自适应生物反馈训练;FBF为固定式生物反馈训练。“—”为无此值。

2 误用 χ^2 检验检测双向有序且属性不同的 $R \times C$ 列联表资料

例2 探讨睑缘炎相关角结膜病变的临床特征及其治疗效果。原文采用 χ^2 检验分析不同年龄组患者睑缘炎相关角膜病变患者病变程度分布,结果显示不同年龄组患者睑缘炎相关角膜病变程度分布差异无统计学意义($\chi^2 = 10.967, P = 0.089$)(表3)^[4]。

表3 96例不同年龄组睑缘炎相关角膜病变患者病变程度分析

分组	轻度		中度		重度		合计	
	例数	百分比/%	例数	百分比/%	例数	百分比/%	例数	百分比/%
儿童	16	16.7	4	4.2	7	7.3	27	28.1
青年	14	14.6	11	11.5	9	9.4	34	35.4
成年	9	9.4	7	7.3	14	14.6	30	31.2
老年	0	0.0	2	2.1	3	3.1	5	3.1
合计	39	40.6	24	25.0	33	34.4	96	100.0

辨析 该数据中分组变量“不同年龄”和分析变量“患者病变程度”均是有序分类变量,但属性不同,称为双向有序且属性不同的 $R \times C$ 列联表资料。本文作者采用 χ^2 检验分析不同年龄分组患者睑缘炎相关角膜病变患者病变程度差异是否有统计学意义,可选用基于秩的非参数检验。仅用 χ^2 检验检测率或者构成比的差异有无统计学意义,只能说明分类变量间是否独立,无法阐明各组间所观察变量的效果优劣和程度高低等信息,降低了资料的利用率和结论的可信度,掩盖了有序变量的信息,有可能得出错误的结论。此类资料应根据研究目的不同选用不同的统计学方法:1)若只关心各年龄组患者睑缘炎相关角膜病变程度是否有差别,此时可视为“结果变量为单向有序 $R \times C$ 列联表资料”,可选用秩和检验和Ridit分析;2)若研

究目的是分析不同年龄组分布与睑缘炎相关角膜病变患者病变程度是否存在线性相关关系,宜采用Spearman秩相关分析或典型相关分析;3)若研究目的是分析不同年龄组分布与睑缘炎相关角膜病变患者病变程度是否存在线性变化趋势,宜采用有序资料的线性趋势检验;4)若研究者仅希望考察各行中的频数分布是否相同,此时可将该定性资料视为双向无序 $R \times C$ 列联表资料,可根据定性资料所得前提条件选用 χ^2 检验或Fisher确切概率法。

笔者依据例2研究目的对数据重新进行统计学分析,采用Kruskal-Wallis H检验、Nemenyi检验和Bonferroni校正法计算不同年龄组患者睑缘炎相关角膜病变程度分布是否存在统计学差异,Kruskal-Wallis H检验结果显示:不同年龄组患者睑缘炎相关角膜病变程度分布差异有统计学意义($\chi^2 = 8.051, P = 0.045$)。采用Nemenyi检验方法进行组间两两比较,因多次进行组间两两比较会增加犯I类错误的概率,故采用Bonferroni校正法对原检验水准 $\alpha = 0.05$ 进行校正,共4组数据,需进行6次两两比较,检验水准调整为 $\alpha = 0.05/6 = 0.0083$ 。Nemenyi检验分析结果显示:不同年龄组间睑缘炎相关角膜病变患者病变程度两两比较差异均无统计学意义($P > 0.05$)。见表4和5。

表4 96例不同年龄组睑缘炎相关角膜病变患者病变程度分析

分组	轻度		中度		重度		合计	
	例数	率/%	例数	率/%	例数	率/%	例数	率/%
儿童	16	16.7	4	4.2	7	7.3	27	28.1
青年	14	14.6	11	11.5	9	9.4	34	35.4
成年	9	9.4	7	7.3	14	14.6	30	31.2
老年	0	0	2	2.1	3	3.1	5	3.1
合计	39	40.6	24	25.0	33	34.4	96	100.0

表5 不同年龄组睑缘炎相关角膜病变患者病变程度

组别	Nemenyi 检验分析结果		两样本平均秩和之差	χ^2	P
	样本例数 n_a	n_b			
儿童与青年	27	34	5.85	0.757	0.860
儿童与成年	27	30	15.13	4.780	0.189
儿童与老年	27	5	28.38	4.993	0.172
青年与成年	34	30	9.28	2.016	0.569
青年与老年	34	5	22.53	3.251	0.355
成年与老年	30	5	13.25	1.106	0.776

3 讨论

列联表资料可以分为配对设计 2×2 列联表资料、单项有序 $R \times C$ 列联表资料、双向有序属性不同的 $R \times C$ 列联表资料和双向有序属性相同的 $R \times C$ 列联表资料,其中单项有序 $R \times C$ 列联表资料又包括原因或分组变量为有序的,而指标变量是无序的资料。宋岩等^[5]研究晚期肾细胞癌患者不同T分期和不同MSKCC分级条件下手足皮肤反应发生率情况一文中,分组变量“T分期和MSKCC分级”为有序分类变量,指标变量(结果变量)“有无手足皮肤反应”是二分类变量, $R \times 2$ 列联表资料,该例资料中,原因变量有序,而结果变量无序,统计分析时应按照无序资料采用 χ^2 检验。但在实际操作中一般会误认为该数据处理方法同表1,而两者有本质的区别。首先表1中结果(疗效)变量“临床疗效和满意情况”为有序变量,而分组变量为无序,结果变量中包含等级资料信息,研究目的是比较不同方法的疗效,如用 χ^2 检验仅能推断2组疗效的构成比是否有差别,不能比较2组的临床效果的差异,应采用非参数检验方法的秩和检验或Ridit分析;而文献^[5]分组变量“T分期和MSKCC分级”为有序分类变量,指标变量(结果变量)“有无手足皮肤反应”是无序变量,旨在分析不同分期或不同分级样本中手足皮肤反应的构成比的差别,可视为双向无序,可以采用 χ^2 检验,但要观察数据是否符合 χ^2 检验的应用条件。编辑在稿件审阅过程中应仔细辨别上述2种数据类型,有效区分 $R \times 2$ 列联表资料与 $2 \times C$ 列联表

资料在研究目的和统计学方法中的差异,指导作者正确运用统计学方法进行分析。

χ^2 检验是定性资料中应用最为广泛的一种统计学方法,但也不应作为检测列联表资料的“万能工具”而盲目应用,在实际操作中应正确辨别资料设计类型(定性资料或定量资料),正确认识统计学分析方法的前提条件和研究目的,才能有的放矢地正确处理列联表资料,最大限度地挖掘统计数据中的价值。

作为医学期刊编辑,受限于所学专业,对稿件中出现的统计学方法应用的辨析能力不尽相同,因此在工作中应加强医学统计学方面知识的积累,编辑部应该注重收集稿件中比较典型的统计学误用错用的实例,在编辑部内部定期组织关于实验研究设计和统计分析方法应用的集体讨论,提高编辑处理统计学实验设计和分析方面的能力,对于涉及多因素实验设计类型,并且统计学分析方法比较复杂的稿件,编辑部会邀请统计学专家进行同行评议,更好地对统计学处理方面进行把关^[6],提高稿件的编校质量和学术水平。

4 参考文献

- [1] 胡良平,赵铁牛,李长平,等.医学综合统计设计与数据分析[M].北京:电子工业出版社,2014:140
- [2] 金永勤,王维.医学期刊编辑应重视统计学中的几个问题[J].编辑学报,2009,21(1):33
- [3] 李小平,史久煜,CHEN Jiande D Z,等.自适应性生物反馈训练对出口梗阻型便秘患者临床疗效、心理状况和生命质量的影响[J].中华消化杂志,2016,36(5):325
- [4] 张晓玉,王智群,张阳,等.睑缘炎相关角膜病变172例临床分析[J].中华眼科杂志,2016,52(3):174
- [5] 宋岩,杜春霞,张雯,等.抗血管生成酪氨酸激酶抑制剂治疗转移性肾细胞癌的手足皮肤反应与疗效的关系[J].中华医学杂志,2016,96(34):2717
- [6] 颜廷梅,任延刚.科技期刊编辑从事科研活动对提升期刊质量及编辑成长的作用:以《中国实用内科杂志》编辑科研实践为例[J].编辑学报,2017,29(1):87

(2017-10-24 收稿;2018-02-12 修回)

数学中的集合符号使用什么字体?

答 一般而言,集合符号采用大写白斜体字母,如集合A、集合B(可简称为集A、集B)。但集合中的5个特殊集合,即自然数集(非负整数集)、整数集、有理数集、实数集、复数集,按GB 3102.11—1993《物理科学和技术中使用的数学符号》的规定,其符号应使用大写空心正体或黑正体字母。因大写空心正体字母不易录入,所以在实践中通常都使用大写黑正体字母,如自然数集N、整数集Z、有理数集Q、实数集R、复数集C,都

是规范、正确的。

顺便通报如下新信息:ISO 80000-2:2009《Mathematical signs and symbols to be used in the natural sciences and technology》规定的特殊集合的符号为大写黑正体字母,删去了空心正体;给出了6个特殊集合,即增加了1个质数集,其符号为P。

(陈浩元)