

医学稿件中常见统计学方法误用辨析

章新生 姚仁斌 刘璐 刘畅 马启

[摘要]目的:提高医学科研工作者对医学资料常用统计学方法的认识和应用能力。方法:对医学原始稿件中常见的一般 χ^2 替代四格表确切概率法、配对 χ^2 检验和校正 χ^2 检验,用 χ^2 检验替代秩和检验(u_c 、 H_c)和处理相关性分析,两两 t 检验替代方差分析和 q 检验, t 检验替代配对 t 检验的误用进行辨析。结果:误用 χ^2 检验主要是未结合实际情况和统计分析的目的来正确使用统计学方法;误用 t 检验主要是未充分理解研究资料是否满足参数检验的条件,以及不能正确判断计量资料所对应的实验设计类型。结论:医学科研工作者应进一步学习和掌握最基本统计学分析方法和经典统计学方法的使用要求。

[关键词] 医学统计方法; 医学论文; χ^2 检验; t 检验

[中国图书资料分类法分类号] R-32

[文献标识码] A

医学期刊是宣传和传播生物医学科研成果重任的载体,正确评价期刊来源稿件的好坏是保证期刊质量的重要一环。考量稿件质量主要是从 2 个方面来评价,即专业水平和统计学分析是否恰当;具有创新性、实用性和先进性是高质量论文的前提,而经准确统计学分析得出的经得起时间考验的结论是保证^[1]。在科研成果公布于众的最后一步中,对资料的正确统计分析显得格外重要。我们调查《蚌埠医学院学报》2009 年作者的原始稿件发现,统计学方法应用错误达 20.40%^[2]。提高医学期刊中统计学应用的质量是一项长期而艰巨的工作,它涉及到作者、编者、审者及读者等多个方面,需要大家共同的努力,才能逐步减少以至消除统计学误用现象,提高医学论文的科学性^[3];为此,我们就近几年编辑稿件中几种常见的统计学分析方法错误使用案例进行辨析,以期对读者有所帮助。

1 χ^2 检验的常见误用

1.1 误用 χ^2 检验替代秩和检验

例 1 稿件题目《皖北矿区婴幼儿气质特点调查分析》,表 1 为作者调查不同性别婴幼儿气质类型分布比较的原表。

表 1 不同性别婴幼儿气质类型分布

性别	<i>n</i>	易养型	中间偏易养型	中间偏难养型	难养型	发动缓慢型
男	120	48(40.00)	49(40.83)	8(6.67)	13(10.83)	2(1.67)
女	110	39(35.45)	45(40.91)	5(4.55)	14(12.73)	7(6.36)
χ^2				4.18		
<i>P</i>				>0.05		

该例作者将不同性别的婴幼儿气质分为易养型、中间偏易养型、中间偏难养型、难养型、发动缓慢型,这 5 种类型实质上是一个等级分类,从易养型到难养型,再到发动缓慢型,是结果变量(气质类型)为有序变量的单向有序列联表资料。要观察此 2 组分布的差异,不能使用 χ^2 检验,应使用成组设计两样本比较的秩和检验(Wilcoxon 两样本比较法),得 $u_c = 1.05$, $P > 0.05$ 。但是,有学者会认为作者使用的方法虽然不恰当,但是结果都是一样的,不必小题大做。实际上 $2 \times C$ 列联表的 χ^2 检验,与结果变量的是否有序没有关系,用 χ^2 检验分析得出的结果是男女不同气质类型的频数是否相同,而得不出男女不同气质类型的差异是否有统计学意义的结论。

例 2 稿件题目《无创机械通气联合硝普钠治疗急性心源性肺水肿》,作者分别采用无创机械通气联合硝普钠(A组)、单纯硝普钠治疗(B组)和常规治疗(C组)急性心源性肺水肿患者,并统计分析 3 组的治疗效果,见表 2。

表 2 3 组患者疗效比较(*n*)

分组	<i>n</i>	显效	有效	无效	χ^2	<i>P</i>
A 组	11	10	0	1	8.05	>0.05
B 组	11	8	1	2		
C 组	11	4	3	4		

由于例 2 是多组单向有序资料的分析,应采用成组设计多个样本比较的秩和检验(Kruskal-Wallis 法),得 $H_c = 6.50$, $P < 0.05$,3 组疗效差异有统计学意义。

1.2 误用 χ^2 检验替代四格表确切概率法

例 3 作者观察白细胞介素-6(IL-6)在正常卵巢组织和卵巢癌组织中的表达情况,见表 3。

[收稿日期] 2011-03-06

[作者单位] 蚌埠医学院学报编辑部,安徽蚌埠 233030

[作者简介] 章新生(1976-),男,编辑。

[通讯作者] 马启,研究员。

表3 正常卵巢组织和卵巢癌组织中 IL-6 mRNA 表达定性比较

分组	n	IL-6 mRNA		χ^2	P
		表达降低	表达增强		
正常卵巢组织	10	2	8		
卵巢癌组织	20	18	2	14.70	<0.01
合计	30	20	10		

当四格表最小理论频数 < 1 或 $n < 40$ 宜用四格表确切概率法分析; 由于本例 $n = 30$ 故不能采用一般四格表 χ^2 检验, 应使用四格表确切概率法分析, 得 $P = 0.000$ 。

1.3 误用 χ^2 检验替代配对 χ^2 检验

例4 稿件题目《超声弹性成像及阴道超声检查在子宫颈占位性病变诊断中的对比研究》, 其中一项观察是作者应用阴道超声联合纵、横切面弹性成像诊断子宫颈占位性病变, 并与病理学诊断比较得出表4结果。

表4 阴道超声联合弹性成像与病理诊断结果比较

阴道超声	病理学诊断		合计	χ^2	P
	恶性	良性			
联合横切面弹性图像					
恶性	12	6	18	50.18	<0.01
良性	2	89	91		
联合纵切面弹性图像					
恶性	12	7	19	46.74	<0.01
良性	2	88	90		

作者在作统计学分析时采用一般四格表资料 χ^2 检验进行分析, 得出差异均有统计学意义, 说明阴道超声联合弹性成像诊断的假阳性率高。正确统计学方法是采用配对四格表 χ^2 检验分析, 要注意是: 当 $b + c < 40$ 时用校正配对 χ^2 检验, 正如表5 得出差异均无统计学意义, 说明阴道超声联合弹性成像诊断与病理诊断一致。

表5 阴道超声联合弹性成像与病理结果对照比较

阴道超声	病理学诊断		合计	χ^2	P
	恶性	良性			
联合横切面弹性图像					
恶性	12	6	18	1.13	>0.05
良性	2	89	91		
联合纵切面弹性图像					
恶性	12	7	19	0.56	>0.05
良性	2	88	90		

1.4 误用 χ^2 检验处理相关性分析

例5 作者为观察凋亡抑制基因 Livin 及 PTEN 在卵巢上皮性癌组织中表达的相关性, 作了表6 统计学分析, 得出卵巢上皮性癌组织中 Livin 与 PTEN 表达无相关性。

表6 Livin 与 PTEN 在卵巢上皮性癌组织中表达的相关性

PTEN 表达	Livin 表达		合计	$\chi^2_{校正}$	P
	+	-			
+	10	6	16	2.82	>0.05
-	38	6	44		
合计	48	12	60		

作者误用 χ^2 检验分析双向有序且属性不同列联表资料; 而作者的意图是分析 2 种基因的蛋白表达是否有相关关系, 所以应用 Spearman 等级相关分析, 得 $r_s = -0.336$, $P < 0.01$, 卵巢上皮性癌组织中 Livin 与 PTEN 的表达呈负相关关系的结论。

2 t 检验的常见误用

2.2 误用两两 t 检验替代方差分析和 q 检验

例6 《缺血性脑血管病患者同型半胱氨酸和基质金属蛋白酶-9 与颈动脉粥样硬化的关系》一文, 作者分别观察比较了正常对照组和缺血性脑血管病患者有斑块组及无斑块组的同型半胱氨酸 (Hcy) 和基质金属蛋白酶-9 (MMP-9) 水平, 原稿统计学方法中采用两两 t 检验, 见表7。

表7 3组 Hcy 及 MMP-9 水平的比较 ($\bar{x} \pm s$)

分组	n	Hcy ($\mu\text{mol/L}$)	MMP-9 (ng/ml)
对照组	236	8.15 \pm 7.30	2.88 \pm 1.91
有斑块组	123	13.98 \pm 6.23 ^{ab}	5.91 \pm 4.36 ^{ac}
无斑块组	117	11.81 \pm 5.78	4.11 \pm 3.54

注: 与对照组比较 a $P < 0.001$, 与无斑块组比较 b $P < 0.01$, c $P < 0.001$

该例研究实际上是检验 3 个相互独立样本均数差异有无统计学意义, 属于单因素方差分析; 而 t 检验只能用于 2 个成组样本均数差异有无统计学意义的检验, 故应采用单因素方差分析进行检验。若差异有统计学意义, 则可进行多个样本的两两比较, 如 q 检验等。本例若直接通过 t 检验作两两比较, 会增加犯第一类错误的概率 α , 如本例犯第一类错误的概率 α 为 0.142 6, 比 0.05 大多了, 所以应在方差分析的基础上进行 q 检验, 改正后见表8, 还要有标记

标注各组间差异有无统计学意义。

表 8 3 组 Hcy 及 MMP-9 水平的比较($\bar{x} \pm s$)

分组	n	Hcy($\mu\text{mol/L}$)	MMP-9(ng/ml)
对照组	236	8.15 \pm 7.30	2.88 \pm 1.91
有斑块组	123	13.98 \pm 6.23**	5.91 \pm 4.36**
无斑块组	117	11.81 \pm 5.78** Δ	4.11 \pm 3.54** $\Delta\Delta$
F	—	33.48	38.08
P	—	<0.01	<0.01
MS _{组内}	—	44.680	9.789

q 检验: 与对照组比较 **P < 0.01; 与有斑块组比较 $\Delta P < 0.05$, $\Delta\Delta P < 0.01$

2.2 误用 t 检验替代配对 t 检验

例 7 稿件题目《单束重建联合胫骨嵌入重建后交叉韧带》,作者采用 Lysholm 膝关节功能评分评定手术前后的膝关节功能,见表 9。

表 9 治疗前后 Lysholm 膝关节功能评分比较

观察时间	n	Lysholm 评分	t	P
治疗前	10	58.6 \pm 2.9	24.69	<0.01
治疗后	10	88.5 \pm 2.5		

本例作者采用自身对照研究,比较治疗前后变化观察疗效,而错误地采用随机 t 检验进行统计学处理。随机 t 检验的应用是完全随机设计的计量资料,是研究者观察 2 样本 $\bar{x} \pm s$ 差异有无统计学意义。配对 t 检验是适用于配对设计的计量资料,包括自身对照和配对对照研究。由于配对 t 检验最大程度地减少了个体变异对处理间变异的可能影响,用较少样本可得较多的信息,提高了检验效能;但其数据处理主要是研究者对观察对象处理前后观察指标的差值平均数和差值标准差的处理,如 $t = \bar{d} / (s_d / \sqrt{n})$ 。正确处理见表 10。

表 10 治疗前后 Lysholm 膝关节功能评分比较(n = 10)

观察时间	Lysholm 评分
治疗前($\bar{x} \pm s$)	58.6 \pm 2.9
治疗后($\bar{x} \pm s$)	88.5 \pm 2.5
$\bar{d} \pm s_d$	29.9 \pm 4.84
t	19.54
P	<0.01

例 8 《醒脑静治疗急性脑出血的临床观察》一文,作者分对照组和治疗组分别采用常规治疗和常规治疗基础上加用醒脑静治疗,于治疗 14 天后评价

2 组的临床神经功能缺损程度评分。作者在同组和组间治疗前后比较中均采用了成组设计定量资料的 t 检验,见表 11。

表 11 2 组治疗前后临床神经功能缺损程度评分比较($\bar{x} \pm s$)

分组	n	治疗前	治疗后	t	P
治疗组	80	22.12 \pm 2.13	11.14 \pm 1.95	34.00	<0.01
对照组	80	21.56 \pm 1.97	14.92 \pm 1.761	22.48	<0.01
t	—	1.73	12.87		
P	—	>0.05	<0.01		

从临床应用角度来讲,该例中的对照组为临床常用的有效治疗方案,而作者研究的目的是观察 2 组治疗后的疗效谁更好。本例作者对组内治疗前后指标变化比较直接采用原始数据计算各自 $\bar{x} \pm s$ 并作成组比较的 t 检验分析,此时扩大了样本例数,如治疗组由 80 例扩大到 160 例,易得出假阳性结果;另外,表 11 对 2 组治疗后 $\bar{x} \pm s$ 作成组 t 检验只能反映治疗组治疗后的水平较对照组治疗后低,并未能真实反映疗效,因为 2 组治疗前的基本数据不完全一样。欲真实反映疗效需作 2 步分析:第一,每组作治疗前后的自身对照分析,即用 $\bar{d} / (s_d / \sqrt{n})$ 计算 t 值、P 值,评价每组治疗是否有效,此时各组样本数仍各为 80 例;第二,在 2 组自身对照差异均有统计学意义的情况下,欲知道治疗组的效果是否真正优于对照组,需比较 2 组的 $\bar{d} \pm s_d$ 差异是否有统计学意义。通过计算得出 2 组治疗均有效,但治疗组效果更好于对照组,见表 12。

表 12 2 组治疗前后临床神经功能缺损程度评分比较($\bar{x} \pm s$)

分组	n	治疗前	治疗后($\bar{d} \pm s_d$)	t*	P
治疗组	80	22.12 \pm 2.13	-10.98 \pm 0.18	545.60	<0.01
对照组	80	21.56 \pm 1.97	-6.64 \pm 0.21	282.81	<0.01
t	—	1.73	140.35		
P	—	>0.05	<0.01		

* 示配对 t 检验

医学计数资料和等级资料中常用的统计学分析方法为 χ^2 检验和秩和检验,计量资料常用方法为方差分析、q 检验和 t 检验,但是论文中 t 检验和 χ^2 检验既是应用最多的,也是误用最多的^[5]。出现以上问题主要还是医学科研工作者对最基本的统计学方法使用范围掌握得不够,未能结合实际情况和统计分析的目的来正确选用统计学方法,未能充分理解研究资料是否满足参数检验的条件,以及不能正确判断计量资料所对应的实验设计类型。所以,医学科研工作者应进一步学习和掌握最基本统计学分析

[文章编号] 1000-2200(2011)06-0656-03

· 精神卫生 ·

大学生师源性焦虑现状调查

黄中岩¹, 唐启寿², 刘锦秀¹, 蒋玉敏¹, 陈善³

[摘要]目的:调查大学生师源性焦虑状况。方法:采用 Zung 编制的焦虑自评量表(SAS)和自编调查问卷对4所高校1 079名在校大学生和100名教师进行调查。结果:大学生焦虑标准得分高于全国正常人常模得分($P < 0.05$);134名学生处于焦虑状态,总体焦虑发生率为12.42%,其中轻微至轻度焦虑126名,中重度焦虑8名。有约40%的学生和教师认为,大学生的焦虑发生与教师不恰当的教育观念和言行,以及教师心理健康教育知识的缺乏有关。结论:当前大学生总体焦虑发生率较高,主要是由于教师不恰当的教育观念和言行,以及教师心理健康教育知识的缺乏导致,高校应重视大学生师源性焦虑问题。

[关键词] 焦虑; 大学生; 心理健康; 教师行为

[中国图书资料分类法分类号] R 749.72 **[文献标识码]** A

Surveys of teacher-oriented anxiety among college students

HUANG Zhong-yan¹, TANG Qi-shou², LIU Jin-xiu¹, JIANG Yu-min¹, CHEN Shan³

(1. Department of Nursing 2. Department of Society Science 3. Department of Clinical Medical Bengbu Medical College Bengbu Anhui 233030, China)

[Abstract] Objective: To investigate the teacher-oriented anxiety among college students. **Methods:** A total of 1 079 college students and 100 teachers selected from four universities were inquired by Self-Rating Anxiety Scale(SAS) and a self-designed questionnaire.

Results: The anxiety score of the college students was higher than that of normal people($P < 0.05$). The general anxiety rate of the students was 12.42%. One hundred and thirty-four students demonstrated anxiety, 126 of whom were in a mild state and 8 were in a moderate to severe state. About students and teachers considered that the incidence of students' anxiety was closely correlated with teachers' improper teaching concept and performance as well as their lack of mental health knowledge. **Conclusions:** The prevalence of anxiety is rather high among college students, which is mainly induced by teachers' improper teaching concept and behavior as well as their lack of mental health knowledge. It is suggested that more attention should be paid to the teacher-oriented anxiety in college students'.

[Key words] anxiety; college students; mental health; teachers' behavior

焦虑是自主神经系统受到干扰以后产生的紧张、不安和恐惧情绪^[1]。在教育过程中,教师稍有不慎就有可能对学生的心理产生不良影响,如焦虑。从影响源的角度,这类焦虑现象称师源性焦虑,是由于教师不恰当的教育观念和言行,以及教师心理健康教育知识的缺乏,导致学生产生焦虑。师源性焦虑因其普遍性、隐蔽性和与教育的主体同一性而具有特殊的研究价值。国内外专家、学者在学生师源性焦虑状况研究方面已经开展了大量工作,取得许

[收稿日期] 2011-02-10

[基金项目] 教育部人文社会科学研究资助项目(09JDSZ3049);安徽省教育厅人文社会科学研究资助项目(2009SK394)

[作者单位] 蚌埠医学院 1. 护理学系 2. 社会学部 3. 临床医学一系 安徽 蚌埠 233030

[作者简介] 黄中岩(1977-),男,硕士,讲师。

[通讯作者] 唐启寿,硕士,讲师。

方法的概念和经典统计学方法使用的要求。医学期刊也应加大对稿件中统计学分析方面的审查,不仅要聘请统计学专家审稿,而且还要提高期刊编辑人员的医学统计学知识,以便于在编辑加工稿件中及时发现统计学应用方面的问题;只有这样,才能使医学论文中统计学方法应用上的缺陷和错误得以弥补和改正,保证论文的科学性和结论的可靠性。多年来,我刊一直严把统计学审稿关,保证了刊出后论文的学术质量。

本文承蒙卫生统计学教研室周纯先教授悉心指导,特此致谢!

[参 考 文 献]

- [1] 胡良平. 科研课题的研究设计与统计分析——错误案例辨析与释疑(第一集) [M]. 北京: 军事医学科学出版社, 2008: 1.
- [2] 章新生, 刘璐, 姚仁斌, 等. 医学稿件中统计学分析使用调查[J]. 中国科技期刊研究, 2011, 22(3): 393-395.
- [3] 王玖, 徐天和, 祁爱琴, 等. 医学论文统计学误用及其防范对策[J]. 编辑学报, 2002, 14(6): 417-418.
- [4] 方积乾, 徐勇勇, 余松林, 等. 医学统计学与电脑实验[M]. 上海: 上海科学技术出版社, 1997: 3-4.
- [5] 毛宗福, 陈捷. 临床论著中常见的统计学缺陷[J]. 中华儿科杂志, 1998, 36(7): 444-445.

(本文编辑 姚仁斌)