

医学期刊常见 t 检验应用错误及案例分析

陈文娟¹⁾ 汤雷²⁾ 马莉¹⁾

1) 山东卫生报刊社山东医药编辑部; 2) 济南日报集团人口健康报社: 250014, 济南

摘要 统计学错误一直是医学论文中常见的错误, 其中尤以 t 检验的误用最常见。文中就医学期刊的常见 t 检验误用情况进行分类, 并进行案例分析, 从而避免此类错误发生。

关键词 t 检验; 常见错误; 案例分析

Case analysis of common errors of t test // CHEN Wenjuan, TANG Lei, MA Li

Abstract Statistical errors, especially t tests, always exist in medical papers. This paper discusses and classifies the common errors in medical sci-tech journals and analyzes some cases for avoiding such mistakes.

Keywords t test; common error; case analysis

First-author's address Editorial Board of Shandong Medical Journal affiliated to health newspapers and periodicals of Shandong, 250014, Jinan, China

DOI: 10.16811/j.cnki.1001-4314.2016.03.010

一篇优秀的医学论文不仅需要研究设计有创造性, 符合“随机、均衡、对照、重复”等基本原则, 更重要的是对实验结果做出统计学分析。笔者随机查阅中华医学会系列杂志及中文核心期刊的论著类文章共 30 篇, 其中出现统计学错误者 15 篇, 占 50%。在统计学错误中最常见的错误是统计描述错误以及 t 检验的误用。现将其错误类型进行分类并做案例分析。

1 统计学方法描述不当

目前, 医学期刊常见的统计学方法描述格式一般是介绍统计软件名称及其型号, 然后描述本文所涉及的资料属于定量资料还是定性资料, 定量资料是否符合正态分布, 符合正态分布的资料以均数 \pm 标准差表示, 不符合正态分布的资料以中位数或百分位数表示。定性资料常以相对数率或构成比表示。数据之间的比较用的是什么统计学方法表示。最后注明统计学检验标准。

常见的统计学方法描述错误是缺少统计软件及其型号, 非正态分布资料用均数 \pm 标准差表示, 不区分计数及计量资料, 无统计学检验标准, 统计学方法描述含糊不清, 未具体说明什么数据之间比较用某种具体的统计学方法, 等等。使用已经废弃的“ P 小于 0.05 为有统计学差异”, 正确描述是“ $P < 0.05$ 为差异有统计学意义”。

如文献[1]表 1 中 CRP 的数据 ISI、BSI、TSI 分别为 79.3 ± 44.1 、 21.7 ± 22.2 、 41.6 ± 41.2 , 标准误差明

显大于均数的 1/2, 提示此组资料为非正态分布资料, 应该用中位数或者是百分位数表述。

又如文献[2]中的统计学方法描述为: 应用 SPSS 16.0 软件, 正态分布资料以 $\bar{x} \pm s$ 表示, 2 组资料比较用二项分布检验、独立样本 t 检验、 χ^2 检验或 Fisher 确切概率法。 $P < 0.05$ 为差异有统计学意义。这里描述统计学方法比较笼统, 而且容易让人误解。2 组什么资料比较用二项分布检验, 什么资料比较用独立样本 t 检验等。文章中, 血肌酐的数值 (309.5 ± 222.3) $\mu\text{mol/L}$, 明显是偏态分布资料, 不该用此表达方式, 而且在统计学方法描述中并未出现偏态分布资料的描述。

2 误用 t 检验

t 检验主要分为单样本 t 检验、配对样本 t 检验及完全随机设计 2 样本比较的 t 检验^[3]。 t 检验的应用条件是: 当样本量较小时, 理论上要求样本为来自正态分布总体的随机样本; 当 2 小样本均数比较时, 要求 2 总体方差相等。

单样本 t 检验用于样本均数与已知总体均数的比较。在实际应用中, 此检验较少见。配对 t 检验又称成对 t 检验, 指的是受试对象相同或基本相同的实验设计, 其优点是抽样误差较小, 实验效率高, 所需样本量相对较少。配对设计分为前后配对设计、左右配对设计、异体配对设计。其中的前后及异体配对设计采用配对 t 检验。

在医学研究中, 配对设计主要有以下情形: 1) 2 同质受试对象接受 2 种不同的处理; 2) 同一受试对象分别接受 2 种不同的处理; 3) 同一受试对象(1 种) 处理前后(自身前后), 要求前后观察条件基本相同, 如动物的急性试验或者慢性病短期处理后的疗效评价试验等。凡是资料呈正态分布或经转换后呈正态分布者可采用; 因此, 配对设计实验计量资料在符合正态分布时, 统计分析应当按配对 t 检验进行, 否则可能降低统计意义。

由于左右配对比较实验的变异来源包括处理与对照、左右、个体间及误差 4 项, 实际上属于三因素实验, 对其资料应进行方差分析。在此不再赘述。

异体配对设计的实验效率多数情况下高于完全随机设计, 其可靠性大于自身前后配对设计。在前瞻性随机对照实验中, 通常采用 1:1 配对。完全随机设计

2 样本比较的 t 检验用于实验研究中完全随机设计 2 样本均数的比较,或观察性研究中受试对象按某种特征分组的 2 组间样本均数比较。

2.1 有些文章仅仅笼统叙述计量资料之间比较用 t 检验,未给出统计量及 P 值

例 1 见文献[4]中的表 2(即本文的表 1)。

表 1 各实验组豚鼠耳蜗外毛细胞残存计数比较($\bar{x} \pm s$)

组别	耳数 n	毛细胞计数
正常对照组	8	1 042.650 \pm 33.690
顺铂加 DHA 组	8	725.029 \pm 116.640 * Δ
顺铂组	8	576.560 \pm 2.960 *

注: * $P < 0.01$,与正常对照组比较; $\Delta P < 0.05$,与顺铂组比较。

此表中应该是先采用方差分析进行多组间的比较,然后采用 LSD- t 检验进行组间两两比较。可是统计学方法中没有统计量及 P 值,仅仅有个结果,导致其结果的可信度降低。

3 应该用 χ^2 检验而误用 t 检验

例 2 如文献[5]巴德-吉亚利综合征合并肝细胞癌的危险因素分析一文中的表 1(即本文的表 2)。

表 2 合并与不合并肝细胞癌的 B-CS 类型分布比较

组别	例数	B-CS 类型[例(%)]		
		下腔静脉型	肝静脉型	混合型
肝细胞癌组	30	0(0)	4(13.3)	26(86.7)
非肝细胞癌组	106	5(4.7)	18(17.0)	83(78.3)
统计值		$t = 2.894$		
P 值		0.235		

表 2 是 2 组构成比的比较,采用 t 检验显然错误,而且此组数据中,有 1 个理论频数 $1 < T < 5$ 。正确的处理方法是应该让作者补充病例数或者是根据专业知识合并相邻行或者列,或者是删除理论频数较小的行或列。本文作者的目的是比较 B-CS 的类型,显然合并项或者删除项都不是很合适的,最好的处理办法是建议作者再积累更多病例数后再发表。

4 应该用重复测量方差分析而用 t 检验

例 3 见文献[6]中的表 1(即本文的表 3)。

表 3 3 组实验动物耳蜗实验前后 ABR 检测比较($n = 16, \bar{x} \pm s$)

组别	ABR 阈值	
	顺铂给药前 1 d	顺铂给药后第 9 天
正常对照组	21.80 \pm 3.42	21.200 \pm 1.427
顺铂加 DHA 组	20.58 \pm 3.39	39.634 \pm 3.122
顺铂组	22.32 \pm 3.16	47.690 \pm 5.34

表 3 是 3 组资料治疗前后 ABR 阈值的比较,作者在统计学方法描述中采用了配对 t 检验,显然是不正

确的。通常 2 组之间的均数比较采用 t 检验,而表 3 是 3 组资料比较,如 3 组独立指标比较,应该用方差分析;但此表是非独立的计量资料,为避免出现 I 类错误的概率,应该用重复测量数据的方差分析。

5 应该用方差分析而用 t 检验

例 4 见文献[7]中的表 1(即本文的表 4)。

表 4 4 组研究对象血清 CysC、Hey、Cr 检测结果比较($\bar{x} \pm s$)

组别	CysC/(mg/L)	Hey/(μ mol/L)	Cr/(μ mol/L)
A 组	0.89 \pm 0.20	11.90 \pm 3.46	53.45 \pm 7.65
B 组	1.36 \pm 0.44	16.85 \pm 5.29	78.93 \pm 25.45
C 组	3.83 \pm 1.91	23.21 \pm 13.03	501.16 \pm 369.97
健康对照组	0.87 \pm 0.16	9.17 \pm 2.59	52.28 \pm 4.67

表 4 中既无统计量值又无 P 值,明显表格不完整,从其统计学描述看是直接进行的 t 检验;而表 4 中是 4 组数据的各指标比较,首先应该采用方差分析。如果有统计学意义,再进行组内的两两比较。常用方法为 LSD- t 检验及 SNK 检验,可根据情况选择某一种方法。如果直接多次采用 2 样本均数比较的 t 检验,则会加大 I 类错误的概率。

6 应该用非参数检测方法而用 t 检验

例 5 文献[8]中的表 1(即本文的表 5)

表 5 不同材料干预前后菌落数量比较($\bar{x} \pm s$)

材料	菌落数量		t	P
	干预前	干预后		
聚乙烯	6.08 \pm 1.58	2.77 \pm 1.70	4.46	<0.05
钴铬钼	2.16 \pm 1.24	0.93 \pm 0.56	2.85	<0.05
钛合金	1.57 \pm 0.52	0.55 \pm 0.41	4.86	<0.05

表 5 中钴铬钼、钛合金干预前后的菌落数量明显不呈正态分布,应该用非参数法检验。

7 配对 t 检验与成组 t 检验误用

例 6 文献[9]中的表 1(即本文中的表 6)。

此文统计学描述中,组间比较采用配对 t 检验,显然有误。表中的 2 组各指标比较应该采用完全随机设计 2 样本比较的 t 检验。通常情况下治疗前后各指标的比较才应该采用配对 t 检验。

8 结论

本文所列举的 t 检验误用案例仅仅只是统计学方法误用中的少部分,希望借此能够抛砖引玉。医学期刊编辑不仅要重视文章的文字编校,而且应该扩充研究设计、研究统计学方法方面的知识,向书本学习,向