



如何正确运用 t 检验

胡良平, 高辉

军事医学科学院生物医学统计学咨询中心, 北京 100850

关键词: 统计学; 医学; 统计分析; t 检验

中图分类号: O212; **文献标识码:** A; **文章编号:** 1672-1977(2008)02-0209-04

Hu LP, Gao H. *J Chin Integr Med/Zhong Xi Yi Jie He Xue Bao.* 2008; 6(2): 209-212.

Received January 14, 2008; published online February 15, 2008.

Free full text (PDF) is available at www.jcimjournal.com.

Indexed/abstracted in and full text link-out at PubMed.

Forward linking and reference linking via CrossRef.

DOI: 10.3736/jcim20080221

Open Access

How to use t -test correctly

Liang-ping HU, Hui GAO

Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China

Keywords: Statistics; medicine; statistical analysis; t test

1 引言

t 检验因其节省样本、计算简单及检验功效较高而成为广大科研工作者最为熟悉且应用最多的统计分析方法之一。但有许多科研人员却盲目相信 t 检验, 将其视为分析定量资料的万能工具, 甚至还有人用它来处理定性资料。究其原因, 主要是对 t 检验的使用条件、资料性质和所能处理的试验设计类型认识不清。本文以部分误用 t 检验的例子为基础, 讨论如何正确应用 t 检验进行定量资料平均值的比较。

2 正确表达 t 检验的名称

一些科研人员在表达所用的统计分析方法时, 凭个人感觉和理解, 给统计分析方法命名。就 t 检验而言, 就有“两样本均数 t 检验”、“两独立样本 t 检验”、“非配对 t 检验”和“团体比较 t 检验”等多种不准确的表达。实际上, 根据拟处理定量资料所对应

的试验设计类型不同, t 检验可分为三种, 即“单组设计定量资料 t 检验”、“配对设计定量资料 t 检验”和“成组设计定量资料 t 检验或单因素两水平设计定量资料 t 检验”。

3 熟练掌握 t 检验的使用条件

3.1 结果指标的性质应为定量资料

例 1 某研究者欲研究某种口服液对厌食大鼠食欲中枢神经元兴奋性的影响, 选择日龄 35~40 d 的 SD 大鼠 34 只, 平均体质量 (60 ± 10) g, 随机等分成 2 组: 模型组和治疗组。观察下丘脑外侧区 (lateral hypothalamic area, LHA) 神经元对刺激胃迷走神经的反应, 共记录了 34 只大鼠的 138 个 LHA 神经元。见表 1。采用 t 检验分析两组间 LHA 神经元对刺激胃迷走神经的反应分布是否存在差异。请问, 用 t 检验分析此试验数据正确吗?

表 1 LHA 神经元对刺激胃迷走神经的反应

组别	只数	LHA 神经元个数			
		反应情况:	兴奋性	抑制性	无反应
模型组	17		10	22	34
治疗组	17		9	29	34

释疑 表 1 中的数据并非实际测量的数值,而是发生某种反应的 LHA 神经元的个数(从每只大鼠身上观测多少个 LHA 神经元也未明确交代),目的是比较两组 LHA 神经元对刺激胃迷走神经的反应的频数分布是否相同,并非是从每组每只大鼠身上观测到一个准确的数据(若果真如此,可能要考虑是否要对数据进行平方根变换,因为此时各组数据很可能接近 Poisson 分布),因而不属于对两组均值进行比较的问题。该资料在本质上应属于定性资料,不应采用 t 检验进行分析。此研究者不分资料的性质和分析目的,盲目套用 t 检验,将模型组的 3 个数据与治疗组的 3 个数据代入 t 检验计算公式进行计算是错误的。

正确做法 若反映情况的 3 个水平在专业上可视为有序(应将抑制性与无反应两列互换一下顺序),可选用秩和检验处理这个“结果变量为有序变量的单向有序 2×3 列联表资料”;若反映情况的 3 个水平在专业上不可视为有序,则应选用一般 χ^2 检验或 Fisher 精确检验处理这个“双向无序 2×3 列联表资料”。

3.2 资料应满足正态性要求

例 2 某研究者欲研究健脾益肾方对小鼠移植性前胃癌术后复发的影响。选择清洁级近交系 6~7 周 615 小鼠 15 只,制作前胃癌术后肺高转移肿瘤模型,手术切除原发瘤后,将小鼠随机等分成 3 组,分别接受安慰剂、化疗和健脾益肾方治疗 15 d。见表 2。若采用 t 检验分析各组间复发瘤质量是否存在显著性差异,可以吗?

表 2 不同治疗方法对复发瘤质量的影响

($\bar{x} \pm s$)		
组别	只数	瘤质量(g)
安慰剂组	5	1.01 \pm 0.93
化疗组	5	0.26 \pm 0.52
健脾益肾方治疗组	5	0.35 \pm 0.35

释疑 表 2 中有 2 个单元格的标准差等于或大于均值,此定量资料很可能服从偏态分布,不满足正态性的要求。此时,就不宜采用 t 检验进行统计分析了,可对原始数据进行变量变换。如转换后的数据满足正态性和方差齐性要求,则可用 t 检验对转换后的数据进行统计分析;如转换后的数据仍不满足正态性或方差齐性要求,则可选用非参数检验法,如秩和检验法。

对于配对设计定量资料,不要求两组原始数据一定要符合正态分布,而是先求出各配对组的差量,

只要两组差量符合正态分布即可。

从试验设计类型上判断,此定量资料应为单因素三水平设计定量资料,用 t 检验分析是错误的。因为 t 检验一般仅适用于分析单组、配对及成组设计定量资料。

正确做法 先检查 3 组定量资料是否满足独立性、正态性和方差齐性,若满足应选用单因素三水平设计定量资料方差分析;否则,要么寻找到合适的变量变换方法,使其满足前述的 3 个前提条件,再对变换后的定量数据使用单因素三水平设计定量资料方差分析;要么直接采用单因素三水平设计定量资料秩和检验。

3.3 资料应满足方差齐性要求

例 3 某科研人员欲研究苦参碱对人肾癌细胞系 GRC-1 细胞株 Bcl-2 和 Bax 蛋白表达的影响,将对数生长状态良好的 GRC-1 细胞株等分成两组:空白对照组和苦参碱处理组。见表 3。采用成组设计定量资料 t 检验分析试验数据。请问,这样做 t 检验对吗?

表 3 苦参碱对人肾癌细胞系 GRC-1 细胞株 Bcl-2 和 Bax 蛋白表达的影响

($\bar{x} \pm s$)			
组别	例数	Bcl-2 蛋白表达量	Bax 蛋白表达量
空白对照组	10	2.765 \pm 0.020	2.467 \pm 0.060
苦参碱处理组	10	2.655 \pm 0.005	2.699 \pm 0.020

释疑 表 3 中不同组间标准差相差较大,方差齐性检验发现,此资料不满足方差齐性的要求。此时,选用成组设计定量资料 t 检验分析此定量资料是错误的。

正确做法 对方差不齐的成组设计定量资料可采用近似 t 检验(或称 t' 检验),也可直接选用秩和检验。

3.4 资料应满足独立性要求 独立性是指一批试验数据之间是相互独立的,若从某受试者身上通过重复测量获得同一个指标的多个观测值,则这些观测值之间就不是独立的^[1]。具体而言,单组设计和成组设计定量资料要求每个个体的测量值要相互独立;配对设计定量资料要求不同对子间的测量值要相互独立^[2]。

3.5 资料通常为总体标准差未知的小样本资料

t 检验是英国统计师 Gosset 为解决总体标准差未知的小样本抽样研究而提出的。所以,当样本含量 n 较小($n < 60$)且总体均数未知时,可采用 t 检验进行统计分析;若样本含量 n 较大,或 n 虽小但总

体标准差已知时,则应采用 u 检验^[3], u 检验现在多写成 Z 检验。当然,当总体标准差未知且为大样本定量资料时,若定量资料满足参数检验的前提条件,一般也应选用相应设计定量资料 t 检验处理。

4 根据试验设计的类型选择合适的 t 检验方法

例 4 某研究者研究补肾活血法治疗多囊卵巢

表 4 补肾活血法治疗前后血清性激素的变化

($\bar{x} \pm s$)

时间	例数	促黄体生成素(U/L)	促卵泡成熟激素(U/L)	雌二醇(ng/L)	促黄体生成素/促卵泡成熟激素
治疗前	45	20.21±12.36	6.32±2.81	40.31±17.58	3.56±1.08
治疗后	45	13.18±6.23	4.23±1.72	72.14±25.59	2.53±0.73

释疑 表 4 为多囊卵巢综合征患者采用补肾活血法治疗前、后 4 个定量指标的观测结果,对每一个定量指标来讲,治疗前、后的两个数据来自于同一个个体,所以此定量资料应为配对设计定量资料。而此研究者在进行统计分析时,仅说明采用 t 检验进行处理,表达不清晰,没有明确说出 t 检验所对应的试验设计类型。

若 4 个定量指标之间在专业上存在一定联系,则宜采用配对设计定量资料 Hotelling T^2 检验;若 4 个定量指标之间在专业上无联系,则应在核查定量资料满足参数检验的前提条件后,对每个定量指标进行一次配对设计定量资料 t 检验。

既然此研究者没有明确说明采用配对设计定量资料 t 检验分析数据,那么其所用的分析方法就有可能是成组设计定量资料 t 检验。事实上,很多科研工作者都曾经用成组设计定量资料 t 检验去分析配对设计定量资料。这样做对吗?

值得注意的是,虽然有人对“配对设计 t 检验的检验效能高于成组设计 t 检验”提出异议,认为配对设计 t 检验和成组设计 t 检验在处理配对设计定量资料时检验效能的高低与两组数据的相关性有关,当配对资料之正相关关系达到一定程度时,配对设计 t 检验检验效能较成组设计 t 检验好;当配对资料之负相关关系达到一定程度时,配对设计 t 检验检验效能较成组设计 t 检验低;当配对资料之相关系数接近于 0 时,配对设计的效能趋于消失^[4]。

正确做法 当研究者采用配对设计的依据(找出全部对观测结果有影响的重要非试验因素,且配对条件由这些重要非试验因素复合而成)较为充分并排除了实施过程中其他混杂因素的干扰和影响时,配对设计定量资料 t 检验检验效能一般高于成组设计定量资料 t 检验;况且,试验设计类型一旦确

综合征的临床疗效,随机选取某医院门诊就诊的多囊卵巢综合征 45 例,给以补肾活血法治疗。测得治疗前后患者血清性激素的数据。见表 4。以 t 检验方法分析数据。请问,这样表达统计分析方法是否存在歧义?

定,其所能采用的统计分析方法也就确定了。所以,一般不要采用成组设计定量资料 t 检验去分析配对设计定量资料。除非在收集试验数据时,没有严格遵循配对设计的要求,具体表现在:其一,配对条件并非是由影响观测结果的所有重要非试验因素组成;其二,同一对数据并非来自配对的两个受试者或同一受试对象处理前后的两次观测结果;其三,自身配对设计中时间间隔过长(自身条件已发生很大变化);其四,其他一切导致“配对”失败的场合。在上述各种情况下,“配对设计”已名存实亡,应按成组设计定量资料去检查其前提条件,并选用相应的统计分析方法(成组设计定量资料的 t 检验或秩和检验)。

5 其他注意事项

5.1 以专业知识为依据,选择单侧检验或双侧检验

当科研人员依据其专业知识可认定某个总体的均数不小于(或不大于)某个标准值或另一总体的均数时,可进行单侧检验;反之,则应进行双侧检验。

5.2 切勿随意应用 t 检验处理单因素 k ($k \geq 3$) 水平设计定量资料或多因素设计定量资料

许多科研人员在分析试验数据时,一律采用 t 检验处理定量资料。例如,用多次 t 检验处理单因素多水平设计定量资料、析因设计定量资料或具有重复测量设计定量资料,这是很不科学的。会导致以下 4 方面的问题。其一,资料的利用率很低,统计量所对应的自由度小,结论不可靠;其二,多次采用 t 检验,会使犯假阳性错误的概率大大增加;其三,割裂了原先的整体设计,只能断章取义,而不能给出概括性的结论;其四,无法分析多个因素之间可能存在的交互作用^[5]。正确的做法是:在进行统计分析时,一定要清楚资料中涉及几个试验因素和每个试验因素有几个

水平,进而正确判定定量资料所对应的试验设计类型;然后再严格检查定量资料是否满足参数检验的前提条件。依据上述正确分析定量资料的两个关键点,方能做到合理选用统计分析方法处理定量资料。

REFERENCES

- 1 Hu LP, Li ZJ. Fundamental of medical statistics and discrimination of typical misuse. Beijing: Press of Military Medical Sciences. 2003; 148-149. Chinese.
胡良平, 李子建. 医学统计学基础与典型错误辨析. 北京: 军事医学科学出版社. 2003; 148-149.
- 2 Yu SL. Medical statistics. Beijing: People's Medical Publishing House. 2002; 47-51. Chinese.
余松林. 医学统计学. 北京: 人民卫生出版社. 2002; 47-51.
- 3 Sun ZQ. Medical statistics. Beijing: People's Medical Publishing House. 2002; 30-31. Chinese.
孙振球. 医学统计学. 北京: 人民卫生出版社. 2002; 30-31.
- 4 Chen PZ, Lu SZ. Research on the power of paired t -test. Zhongguo Wei Sheng Tong Ji. 1996; 13(6): 19-22. Chinese.
陈佩珍, 陆守曾. 关于配对 t 检验效能的研究. 中国卫生统计. 1996; 13(6): 19-22.
- 5 Hu LP. Applied course of statistical analysis by Windows SAS 6.12 & 8.0. Beijing: Press of Military Medical Sciences. 2001; 152. Chinese.
胡良平. Windows SAS 6.12 & 8.0 实用统计分析教程. 北京: 军事医学科学出版社. 2001; 152.

Cochrane 系统综述摘要

刺激内关穴(P6)预防术后恶心与呕吐

薛慧敏¹ 翻译, 刘建平² 审校

1. 北京中医药大学人文学院, 北京 100029

2. 北京中医药大学循证医学中心, 北京 100029

背景:术后恶心与呕吐(postoperative nausea and vomiting, PONV)是手术以及麻醉后的常见并发症。应用药物进行预防只对部分患者有效。另一种方法就是刺激手腕部的内关穴(穴位代码 P6),尽管已有许多相关的临床试验研究了该技术的效果,但获得的结果并不一致。

目的:评价刺激 P6 预防术后恶心与呕吐的有效性及其安全性。

检索策略:我们对 Cochrane 对照试验中心注册库(Cochrane Central Register of Controlled Trials, CENTRAL)(Cochrane 图书馆 2003 第 1 期)、MEDLINE(1966 年 1 月~2003 年 1 月)、EMBASE(1988 年 1 月~2003 年 1 月)以及国家医学图书馆关于针灸研究的出版目录(更新至 2003 年 1 月)进行检索,同时查找所检索论文的参考文献目录以获得更多的参考文献信息。

纳入标准:纳入所有刺激 P6 与假治疗或药物疗法对照预防 PONV 的随机临床试验。纳入的试验涉及的干预包括针刺、电针、经皮肤神经刺激、激光刺激、针刺仪治疗以及指压疗法。

数据收集与分析:两名评估员独立地评估方法学质量并提取数据。主要结局为恶心与呕吐的发生率,次要结局为病人需要急救性止吐治疗的情况和不良反应的发生。我们采用随机效应模型,并报告相对危险度(relative risk, RR)及其 95%可信区间(95% confidence intervals, 95% CI),同时应用 Egger 方法(Egger's test)测量漏斗图不对称性。

主要结果:我们共纳入 26 个试验($n = 3\ 347$),其中无一报

告了恰当的随机分配隐藏。尽管许多试验都存在不一致性,但与假治疗比较,刺激 P6 组的恶心($RR = 0.72$, 95% CI 0.59~0.89)和呕吐($RR = 0.71$, 95% CI 0.56~0.91)明显缓解,对止呕治疗的需求($RR = 0.76$, 95% CI 0.58~1.00)亦明显降低。尚没有关于刺激 P6 与单个止吐药物对比对于恶心与呕吐危险度影响差别的研究证据,然而将不同止呕药混合应用时,刺激 P6 组($RR = 0.70$, 95% CI 0.50~0.98)相较于止吐药组($RR = 0.92$, 95% CI 0.65~1.29)恶心危险度明显减低,呕吐的危险度的差异则无统计学意义。刺激 P6 的副作用较小。有一些证据表明漏斗图存在不对称性。

结论:本系统综述支持刺激 P6 的应用,从而避免止呕药物预防措施的使用。与预防性止吐药物相比,刺激 P6 可能减少恶心的风险,但不能减少呕吐的风险。

声明:本文翻译的准确性由译校者负责,不代表 Cochrane 系统综述原作者的责任。资料来源:Lee A, Done ML. Stimulation of the wrist acupuncture point P6 for preventing postoperative nausea and vomiting. Cochrane Database Syst Rev. 2004; (3): CD003281. Copyright © 2007 The Cochrane Collaboration. Published by John Wiley and Sons, Ltd. DOI: 10.1002/14651858.CD003281.pub2. This version first published online: July 19, 2004.