

医学期刊中容易误用的统计学方法辨析

张 维 邓强庭 冷怀明[†]

《第三军医大学学报》编辑部,400038,重庆

摘 要 对医学期刊中容易误用的几种统计学方法进行辨析,包括重复测量方差分析的误用、非参数检验的误用,并对多元线性回归与 Logistic 回归进行了差异性分析,使编辑能够分辨论文统计方法的正误,并作出正确处理,以提高论文的科学性和期刊质量。

关键词 医学期刊;统计学;重复测量;非参数;回归分析

Misuse of statistical methods in medical journals // ZHANG Wei, DENG Qiangting, LENG Huaiming

Abstract This paper analyzes some statistical methods which are easy to be misused in medical journals, including the misuse of ANOVA for repeated design data, and the misuse of nonparametric statistics. Then we analyze the difference between multiple linear regression analysis and Logistic regression analysis, attempting to improve the judgment ability of editors and the journal quality.

Keywords medical journal; statistics; repeated design data; nonparametric statistics; regression analysis

Authors' address Editorial Office of Journal of Third military Medical University, 400038, Chongqing, China

医学统计学是医学科学研究的重要工具,统计学方法运用得正确与否直接影响医学科研质量和科研论文水平。医学期刊编辑在处理来稿时,并非将每篇论文都送统计学专家评审,这就要求编辑应该具有一定的统计学知识,即使没有统计学专家的帮助,也能通过已掌握的知识和一些简单、有效的方法来进行判断。

笔者查阅了国内一些医学期刊,发现有 3 类情况作者特别容易出错,而编辑也往往容易忽略,导致登出来的文章统计方法错误,降低了期刊的质量。本文针对这 3 种情况并结合本刊文献进行分析,以期使编辑能够正确选择统计方法。

1 重复测量数据误用 *t* 检验或方差分析来处理

重复测量数据在医学研究中十分普遍,据对临床医学类杂志研究论著的统计,约占 1/4^[1]。重复测量设计(repeated measurement design)是指在给予 1 种或多种处理后,在多个时间点上从同一受试对象重复获得指标的观察值。其研究目的是探讨同一研究对象在不同时间点上某指标的变化情况。在审稿和统计咨询的过程中我们发现,在对重复测量资料进行分析时,作者常常采用 *t* 检验或者随机区组方差分析。重复测量的数据与 *t* 检验或者随机区组方差分析的区别在于,重复观测数据间存在相关性,而 *t* 检验或者随机区组方差分析的数据彼此独立^[2],因而若采用 *t* 检验或者随机区组方差分析,就有可能得出错误的结论。

本文对医学论文中重复测量数据误用情况的案例辨析如下。

如文献[3]中表 1 所示,作者对术后各时相点裸鼠埋植真皮替代物中 VEGF165 蛋白含量进行比较。

表 1 术后各时相点鼠埋植真皮替代物中 VEGF 165 蛋白含量比较($\bar{x} \pm s$)

| 样本数 | 术后 3 d | 术后 7 d | 术后 14 d | 术后 21 d | F_2 | P_2 |
|-------|---------------|----------------------------|------------------------------|------------------------------|-------|-------|
| 9 | 1 | 1.111 ± 0.018 | 1.299 ± 0.027 | 1.362 ± 0.103 | 3.34 | >0.05 |
| 9 | 1.106 ± 0.040 | 1.167 ± 0.080 | 1.311 ± 0.044 | 1.402 ± 0.023 ^c | 4.28 | <0.05 |
| 9 | 1.113 ± 0.020 | 1.357 ± 0.070 | 1.652 ± 0.086 ^a | 2.152 ± 0.062 ^{ac} | 6.42 | <0.01 |
| 9 | 1.303 ± 0.040 | 1.593 ± 0.032 ^a | 2.403 ± 0.091 ^{abc} | 2.879 ± 0.047 ^{abc} | 8.67 | <0.01 |
| F_1 | 1.64 | 4.23 | 6.44 | 7.11 | | |
| P_1 | >0.05 | <0.05 | <0.01 | <0.01 | | |

注:表中数据为各组各时相点真皮替代物中血管内皮生长因子 165(VEGF 165)蛋白相对表达量与生理盐水组术后 3 d 真皮替代物中蛋白相对表达量的比值; F_1 、 P_1 为组间同一时相点比较所得, F_2 、 P_2 为组内各时相点比较所得;a 与生理盐水组比较, $P < 0.05$;b 与未转染细胞组比较, $P < 0.05$;c 与组内术后 3 d 比较, $P < 0.05$ 。误用成单因素方差分析。

分析本例,要判断是重复测量设计还是随机单位组设计,主要看重复测量的时间是不能随机分配给实验单位的伴随因素,还是可以随机分配的处理因素。显然,该文重复测量的时间是不能随机分配给实验单位的伴随因素,而非处理因素,因此属于重复测量数据的两因

素多水平设计(两因素指干预和测量时间,多水平指干预和测量时间均有 2 个以上水平),应用重复测量的方差分析;而该文采用单因素方差分析,显然不妥。

2 非参数检验误用 *t* 检验或方差分析来处理

在医学论文中,我们经常会发现这样一种情况,就

[†] 通信作者

是其研究结果的数据标准差很大,有的甚至超过了均值;但有很多作者还是采用 $\bar{x} \pm s$ 的形式来表示数据的集中趋势和离散程度,并且统计方法选用的是 t 检验或者方差分析。而编辑在遇到这种情况时可能由于统计知识缺乏,往往无法判断其采用统计方法的正误。那么,这种表示形式以及统计方法是否正确?编辑在遇到这种情况时该审查些什么?我们以本刊一篇初投稿为例来进行辨析。

首先,我们发现在表2所示的案例中,有相当一部分的数据其标准差是平均数的1/2以上,甚至超过均值。这就提示数据有可能偏离了正态分布。查看作者采用的统计学方法,是独立样本的 t 检验。我们知道, t 检验及方差分析是由正态分布数据以及各组方差相同推导出来的,因而要求数据为正态分布(至少接近正态分布)且方差齐性^[4];所以,此时编辑应予以高度重视,因为一旦证实该表数据不服从正态分布,就表明该文的统计方法是错误的,应让作者重新计算。通过对该表原始数据进行正态性检验,大部分指标 $P < 0.05$,表明不符合正态分布,并且方差不齐 ($P < 0.05$)。事实上,当分布偏离正态分布不大时, t 检验和方差分析对其结果的影响并不大;但是,如果正

态性检验结果认为数据不服从正态分布,就必须进行数据转换,使其转化为正态分布。这样才能使用 t 检验或者方差分析,但比较麻烦。简单的方法是用非参数统计,因为非参数统计方法适用于总体分布非正态或分布形式未知的计量资料。过去一般认为非参数统计效率较差,但这是指当数据为正态分布时;如果数据是非正态分布时其效率比参数法高,甚至可以高出很多^[4]。

其次,我们知道,只有属于正态分布(至少是近似正态分布或对称分布)的数据才可以用算术平均数和标准差来表示其集中趋势和离散程度($\bar{x} \pm s$);若不属于正态分布,此时宜用中位数和四分位间距[即 $Md(P_{25}, P_{75})$]描述集中趋势和离散程度^[5-7]。如果数据是偏态分布,仍用算术平均数和标准差来描述其分布特征,将导致资料表达不准确,也无法全面描述数据的分布特征。

在采用非参数统计方法(Kruskal-Wallis H 检验)重新进行计算,并用 $Md(P_{25}, P_{75})$ 正确表示其数据集中趋势和离散程度后,该文予以发表,见文献[8]表3。由表3可知,两两比较结果与作者初投稿时分析的结果存在不同。

表2 哮喘患者诱导痰细胞学分类(% , $\bar{x} \pm s$)

| 组别 | <i>n</i> | Eos | Nen | Lym | Mon | Total |
|-----|----------|--------------------------|-------------|-----------|-------------|--------------------------|
| Eos | 119 | 19.7 ± 16.3 ^a | 42.9 ± 14.2 | 2.0 ± 1.3 | 35.1 ± 15.0 | 5.97 ± 3.24 ^a |
| Nen | 79 | 0.1 ± 0.2 | 79.3 ± 8.2 | 2.3 ± 2.0 | 18.3 ± 8.7 | 5.64 ± 3.25 ^a |
| Pau | 32 | 0.1 ± 0.2 | 45.9 ± 14.3 | 3.4 ± 7.3 | 49.6 ± 16.4 | 4.68 ± 1.52 ^a |
| Mix | 57 | 5.2 ± 5.7 | 76.6 ± 8.7 | 2.3 ± 3.8 | 15.3 ± 8.8 | 8.01 ± 6.05 |

注:a $P < 0.01$,与 Mix 组比较。误用成独立样本 t 检验。

表3 各组哮喘患者诱导痰细胞学分类(% , $Md(P_{25}, P_{75})$)

| 组别 | <i>n</i> | Eos | Nen | Lym | Mon | Total |
|-----|----------|-----------------------------|------------------------------|--------------|-------------|------------------------------|
| Eos | 119 | 15.4(6.8,28.0) ^a | 42.6(33.3,55.1) ^a | 1.9(1.2,2.6) | 35.1 ± 15.0 | 5.37(4.12,7.14) ^a |
| Nen | 79 | 0.0(0.0,0.0) ^a | 78.9(72.3,87.4) ^a | 2.0(1.3,2.7) | 18.3 ± 8.7 | 4.86(3.50,6.76) ^a |
| Pau | 32 | 0.0(0.0,0.0) ^a | 51.2(32.7,58.4) ^a | 2.1(1.3,2.6) | 49.6 ± 16.4 | 4.39(3.76,5.75) ^a |
| Mix | 57 | 3.5(1.8,6.0) ^a | 75.9(69.2,82.3) ^a | 1.6(1.1,2.1) | 15.3 ± 8.8 | 5.88(4.66,9.68) ^a |

注:采用非参数统计方法重新计算。a $P < 0.01$,与 Mix 组比较。

3 多元线性回归与 Logistic 回归的差异性辨析

大多数编辑对于一些简单的统计学方法的判断和应用已经掌握得比较好了,但是对于论文中出现的一些高级的统计学处理方法的使用原则还比较陌生。比如多元线性回归、Logistic 回归在论文中经常出现,但很多编辑对于其差异性还不是很了解,也难以判断作者所用的统计方法是否正确。其实,对这2种回归分析有简单、有效的判断方法。下面我们将两者的区别辨析如下。

多元线性回归分析通过作出多元线性回归方程,

来描述1个因变量与多个自变量间的线性依存关系^[9]。Logistic 回归分析比较常用的情形是探索某疾病的危险因素,根据危险因素预测某疾病发生的概率等。多元线性回归与 Logistic 回归实际上有很多相同之处;正是因为如此,这2种回归可以归于同一个家族,即广义线性模型(generalized linear model)。它们最大的区别是多元线性回归分析要求应变量 Y 为连续型变量,而 Logistic 回归分析要求应变量 Y 为某现象发生与不发生的二分类变量或多分类变量;但是二分类变量更为常用,如发病与未发病、阳性与阴性、死亡与生存、治愈与未治愈、暴露与未暴露等。