

科技期刊编辑应注意统计描述中的错误

邵晓军

(韶关学院 学报编辑部, 广东 韶关 512005)

摘要:针对科技论文中统计方法应用的广泛性和重要性,从统计描述的基本概念、统计指标的选择及统计图表的设计三个方面分析了统计描述中常见的错误类型和形式,阐述了科技期刊编辑在日常处理稿件时如何发现、判断和避免统计描述中的错误。

关键词:科技期刊;编辑;统计描述

中图分类号: G255.2;C812 **文献标识码:** A **文章编号:** 1673-260X(2011)08-0208-03

随着科学技术的高速发展,统计作为一门科学,已广泛应用到自然科学和社会科学的各个领域,也是科学研究不可或缺的手段和工具。但在统计方法的实际应用中,由于许多作者对统计学的基本理论掌握不充分,盲目参考、照本宣科甚至滥用别人的统计分析方法,结果得到的是荒谬或与事实相反的结论。更有甚者,为使实验结果符合“预期的结论”,置科研要求中的科学性和严谨性于不顾,毫无根据地选择统计方法或修改实验数据,致使数理统计方法的可信度降低,科学的权威性遭到破坏。编辑人员做为学术论文的把关人,如果能在处理文稿时发现论文中统计错误,就能让这些数据造假者无处遁形。基于此,本文从统计描述的基本概念、统计指标的选择和统计图表的设计三个方面总结了科技论文中常见的统计学错误形式以及编辑在审稿过程中如何发现、判断和避免统计描述中的错误。

1 统计描述中概念的错误

统计描述是对数据的直接处理和分析,目的在于通过一定方式的描述与整理,计算统计数据的特征值,进而发现其数量的规律性,为用样本统计量推断未知总体的参数提供充分的依据^[1]。它是统计推断和预测的基础,在一定程度上决定了统计工作的成败^[2]。调查问卷是学术论文中常见的研究方法之一,它是通过有针对性的设计一系列问卷问题,直接对特定的调查对象进行调查的一种方法。论文中的调查问卷是否准确、合理而又全面,很大程度上取决于作者设计的调查问卷表的质量,因此,编辑在审稿时必须注意调查问卷中的常见错误。

1.1 要注意调查问卷的定义是否准确

调查问卷的问题对于所有调查对象,应该只有一种解释,代表同一主题。如果调查的问题不严谨会产生很多歧义,使调查对象无所适从,当然也不可能搜集到准确的数据资料。因此,在审稿过程中对问卷设计的问题要注意定义是否明确,即在问题中是否有明确的人物、时间、地点、做什么、为什么做、怎么做;要注意问题中是否使用了含糊的形容词、副词,特别是在描述时间、数量、价格等情况的时候,还要注意问题中是否含有隐藏的选择和选择后果,对于不明确的定义要用定量描述代替^[3]。

1.2 要注意统计描述的概念是否明确

统计描述中有不少反映数据特征的概念,尤其是有些概念可以用不同的测度值来表示。如集中趋势的测度值可以用平均数、众数、中位数等不同的值来表示,但每个值所反映的侧重点不同。所以,在统计描述中必须注意这些基本概念是否明确。如果作者在一组数据中用了中位数,但编辑对中位数的概念不太了解,换成了平均数,那么这组数据所反映的问题就完全不同,甚至是错误的了。统计描述中有几类基本概念是编辑需要了解的:集中趋势的测度、离散程度的测度、偏态与峰度的测度以及分类资料的统计描述。这几类概念比较常用,具体可以参考相关专业的书籍^[4-6]。

2 统计指标的误用

2.1 应用相对性指标的注意事项

(1)正确区分率和构成比。率和构成比所说明的问题不同,绝不能以构成比代替率。构成比只能说明各组成部分的比重或分布,而不能说明某现象发生的频率或强度。例如:以男性各年龄组高血压分布为例,50~60岁年龄组的高血压病占52.24%,

所占比重大,60 岁以上年龄组只占到 6.74%.这是因为 60 以上受检人数少,造成患病数低于 50~60 岁组,因而构成比相对较低.但不能认为年龄在 50~60 岁组的高血压患病率最严重,而 60 岁以上反而有所减轻.若要比较高血压的患病率,应该计算患病率指标.

(2)计数资料相对数的分母不宜过小.一般来说,样本数量较多,计算的相对数可靠性也较大,否则相对数的计算没有说服力.

(3)用率或构成比进行组间比较时,要注意资料间是否有可比性.

(4)分组资料计算合并率时,不能用各个率简单相加,而应当用有关的合计数进行计算.

2.2 平均数指标的应用范围和条件

(1)算术平均数适合于分布均匀的小样本数据或近似正态分布的大样本数据;几何平均数适合于等比级数的数据,尤其是平均增长速度、传染言不由衷发病的平均潜伏期等习惯上都用几何平均数表示;中位数适合于各种类型的数据,但尤其适合于大样本偏态分布的数据.

(2)对于同一资料,可能同时选用几个平均数指标,如对于某些偏态分布的数据,可能会选用算术平均数、几何平均数和中位数.对于这种情况,统计上的处理原则是,如果算术平均数或几何平均数与中位数接近,应采用算术平均数或几何平均数作

为平均数指标,否则用中位数作为平均数指标.

(3)在医学方面,要了解各专业平均数的习惯用法.如儿童龋齿个数虽然呈偏态分布,但在口腔预防保健系统中,仍习惯上计算算术平均数.

2.3 应用变异指标的注意事项

(1)方差与标准差属于同类指标,但标准差与均数的单位相同,作为变异指标多采用标准差.

(2)变异(离散)系数主要用于不同类型观察指标或同类型观察指标,但均数差悬殊时变异程度的比较.也常用于评价仪器测量精度和稳定性方面.

(3)四分位间距适合于任何分布数据,计算结果比极差稳定,对于大样本偏态数据尤其适用.

3 统计图表中的常见错误

统计表与统计图中的错误非常多见,本文根据编辑规范和各种类型错误出现的频率,列举如下.

3.1 统计表编制的常见错误

(1)不符合列表的原则和编制结构要求.这方面的错误其实是违背了编排规范的一种结果,大多是由于作者对编排规范掌握不足导致的,也有是理解上的错误.表 1 是对 119 例宫颈糜烂冷冻治疗效果的列表,可以看出该表的主要目的在于表达冷冻治疗宫颈糜烂的近期疗效.存在的问题是:标题未突出“近期疗效”这一主要内容;主谓词安排不当且标目重复,如例数和%多处出现;总计意义不明确;线条过多,以致数据隔离,不便比较.应改为表 2 的

表 1 119 例宫颈糜烂冷冻治疗结果

效果	轻度糜烂		中度糜烂		重度糜烂		总计	
	例数	%	例数	%	例数	%	例数	%
治愈	39	32.77	11	9.24	2	1.68	52	43.70
好转	2	1.68	19	15.97	14	11.76	35	29.41
无效	8	6.72	7	5.88	17	14.29	32	26.89
合计	49		37		33		119	

表 2 冷冻治疗宫颈糜烂患者的近期疗效

糜烂程度	例数	疗效			疗效构成比 /%		
		治愈	好转	无效	治愈	好转	无效
轻度	49	39	2	8	79.6	4.1	16.3
中度	37	11	19	7	29.7	51.4	18.9
重度	83	2	14	17	6.1	42.4	51.5
合计	119	52	35	32	43.7	29.4	26.9

编排.

(2)主辞和宾辞倒置,横、纵标目不明确.此类错误也比较多见,多数由作者自己的主观偏好或列表习惯所致.也有部分作者为使表格排列美观,

将主辞和宾辞倒置或将部分纵列转排至横行,致使横、纵标目不明确.表 3 是作者来稿中的一个表格,看起来简单明了,规范整齐.但结合论文及表题可知,表 3 的设置,完全不能表达本来的意义.正确的

表格如表 4 所示.

表 3 FDP 对幼兔离体心脏模型再灌注期肌酸肌酶的影响($\bar{x} \pm s$, U/L, n=10)

测定时间(min)	FDP 组	对照组
缺血前	6.73 $\pm$ 1.42	6.41 $\pm$ 1.26
5	12.50 $\pm$ 1.42bc	39.33 $\pm$ 1.99c
10	15.53 $\pm$ 1.09bc	44.17 $\pm$ 1.28c
30	20.20 $\pm$ 1.96bc	48.10 $\pm$ 2.04c

^bP<0.01vs 对照组; ^cP<0.01vs 缺血前.

表 4 FDP 对幼兔离体心脏模型再灌注期肌酸肌酶的影响

分组	例数	肌酸激酶含量($\bar{x} \pm s$)/(U·L ⁻¹)			
		缺血前	再灌注期 5 min	再灌注期 10 min	再灌注期 30 min
FDP 组	10	6.73 $\pm$ 1.42	12.50 $\pm$ 1.42bc	15.53 $\pm$ 1.09bc	20.20 $\pm$ 1.96bc
对照组	10	6.41 $\pm$ 1.26	39.33 $\pm$ 1.99c	44.17 $\pm$ 1.28 c	48.10 $\pm$ 2.04c

^bP<0.01vs 对照组; ^cP<0.01vs 缺血前.

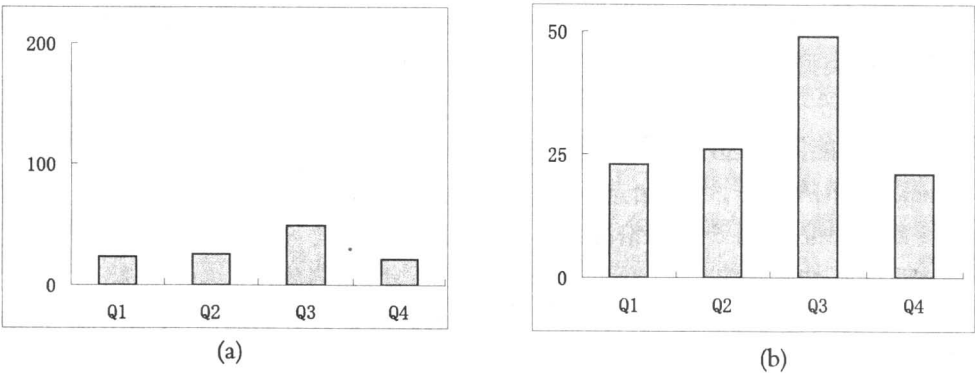


图 1 某行业季度销售额(单位:百万元)

现错误的形式也有很多，尤其是在指标的选择、规范性要求以及数据的引用等方面更为普遍.本文仅以科技期刊编辑在日常处理稿件时遇到的统计描述中常见的问题进行整理 ,供年青编辑人员参考.

参考文献:

[1]张梅琳.新编统计学[M].北京:立信会计出版社,2001.
[2]董时富.生物统计学[M].北京:科学出版社,2002.

3.2 统计图的误用

统计图中也经常会出现一些编排上的错误 ,主要是坐标单位选择不当、图例形式选择不当和坐标轴的缺失等.如图 1(a)所示 ,纵坐标的取值范围为20~50 ,但纵坐标单位 100 时 ,看上去四个季度销售额差不多 ;但如果以 10 为纵坐标单位 (如图 1(b)所示) ,同样的数据就会充分反映出四个季度销售额的不同.

科技论文中运用统计方法的形式多种多样 ,出

[3]付志远.市场调查问卷设计的几类常见错误及纠正[J].企业经济,2000(1).
[4]谢启南,韩兆洲.统计学原理[M].广州:暨南大学出版社,2002.
[5]袁卫,庞皓,曾五一,等.统计学[M].北京:高等教育出版社,2005.
[6]章昕.概率统计教材辅导[M].北京:科学技术文献出版社,2005.