

标准差与标准误

郝拉娣¹⁾ 于化东²⁾

(1)大连水产学院学报编辑部,116023,2)数学研究与评论杂志编辑部,116024;辽宁大连)

摘要 对容易引起混淆的统计量“标准差”和“标准误”从意义、特征、计算公式、符号表示等方面作了准确描述与区分,并对统计学结果表示中“平均数 ±标准差”“平均数 ±标准误”的符号表示进行了统计分析,指出了存在问题。通过原因分析,提出了避免二者混淆和不规范的符号表示的一些应对措施。

关键词 科技论文;算术平均数;标准差;标准误

中图分类号 G237.5

Standard deviation and standard error of arithmetic mean Hao Ladi, Yu Huadong

Abstract The “standard deviation” and “standard error of arithmetic mean” that being easy to cause confusion are accurately described and distinguished from the meaning, characteristic, formula of calculation and symbolization etc. The symbolization of “mean ±standard deviation” and “mean ±standard error of mean” in the expression of statistics result are analyzed. Then, some countermeasures to prevent the two cases from being obscure and being expressed with abnormal symbols are put forward.

Key words sci-tech paper; arithmetic mean; standard deviation; standard error of mean

First author's address Editorial Office of Journal of Dalian Fisheries College, 116023, Dalian, China

在科学实验和工程实践中,常遇到实验结果中包含的随机误差,一般都需要在假定系统误差得到消除的情况下,计算出实验结果可能达到的准确范围,因此,在科技论文中常有“平均数 ±标准差”与“平均数 ±标准误”(本文中“平均数”均指“算术平均数”)的统计学结果表达。虽然“标准差”与“标准误”均用来反映随机误差,但一字之差,如果分不清它们的实质含义,很容易混淆这2种表达。

1 标准差与标准误

1.1 总体标准差与样本标准差 标准差(standard deviation)作为随机误差(或真差)的代表,是随机误差绝对值的统计均值。在国家计量技术规范中,标准差的正式名称是标准偏差,简称标准差^[1],用符号 σ 表示。标准差的名称有10余种,如总体标准差、母体标准差、均方根误差、均方根偏差、均方误差、均方差、单次测量标准差和理论标准差等^[2]。

标准差的定义式为

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \quad (1)$$

式中 x_i 为一组样本变量(从总体中抽取的一部分个体的集合)。由于式(1)中含有的参数——总体算术平均数 μ (亦称数学期望,或称真值)和总体数 N ,是不能进行实际计算的,因此,式(1)只有理论上的意义,无法求出;而经常采用的方法是用样本参数来估计总体的参数,即用样本标准差 s 的值作为总体标准差 σ 的估计值。样本标准差的计算公式为

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2)$$

式中 \bar{x} 为样本算术平均数(以下简称平均数); n 为观测样本数。 s 反映了整个样本变量的分散程度。样本标准差小,说明样本变量的分布比较密集在平均数附近,否则,表明样本的分布比较离散。当 $n \rightarrow \infty$ 时, s 趋向于 σ 。

在表达有随机误差的实验结果数据中,我们常见到“平均数 ±标准差”中的“标准差”,实际上是指总体标准差的估计值 s ,在以下的讨论中如无特别指明,标准差均指样本标准差 s 。

1.2 样本平均数的标准误及其估计 在抽样试验(或重复的等精度测量)中,常用到样本平均数的标准差,亦称样本平均数的标准误或简称标准误(standard error of mean)^[3]。因为样本标准差 s 不能直接反映样本平均数 \bar{x} 与总体平均数 μ 究竟误差多少,所以,平均数的误差实质上是样本平均数与总体平均数之间的相对误差^[4]。可推出样本平均数的标准误为

$$s_{\bar{x}} = (1/\sqrt{n}) s. \quad (3)$$

与总体标准差 σ 类似,样本平均数的标准误 $s_{\bar{x}}$ 也无法求出,只能估计。为了区别是用样本标准差 s 的值来估计总体标准差 σ 的值,样本平均数的标准误 $s_{\bar{x}}$ 的估计值用 $s_{\bar{x}}$ 表示(也称平均数的标准偏差^[5]),即

$$s_{\bar{x}} = (1/\sqrt{n}) s. \quad (4)$$

$s_{\bar{x}}$ 反映了样本平均数的离散程度。标准误越小,说明样本平均数与总体平均数越接近,否则,表明样本平均数比较离散。当 $n \rightarrow \infty$ 时, $s_{\bar{x}}$ 趋向于 $\sigma_{\bar{x}}$ 。

在表达有随机误差实验结果的数据中,我们常见到“平均数 ±标准误”中的“标准误”,实际上是指样本平均数标准误的估计值 $s_{\bar{x}}$,在以下的讨论中如无特别指明,标准误均指 $s_{\bar{x}}$ 。

1.3 标准差与标准误的区别 标准差与标准误的意

义、作用和使用范围均不同。标准差(亦称单数标准差^[4])一般用 s 表示,是表示个体间变异大小的指标,反映了整个样本对样本平均数的离散程度,是数据精密度的衡量指标;而标准误一般用 $s_{\bar{x}}$ 表示,反映样本平均数对总体平均数的变异程度,从而反映抽样误差的大小^[6],是量度结果精密度的指标。

随着样本数(或测量次数) n 的增大,标准差趋向某个稳定值,即样本标准差 s 越接近总体标准差 σ ,而标准误则随着样本数(或测量次数) n 的增大逐渐减小,即样本平均数越接近总体平均数 μ ;故在实验中也经常采用适当增加样本数(或测量次数) n 减小 $s_{\bar{x}}$ 的方法来减小实验误差,但样本数太大意义也不大。

标准差是最常用的统计量,一般用于表示一组样本变量的分散程度;标准误一般用于统计推断中,主要包括假设检验和参数估计,如样本平均数的假设检验、参数的区间估计与点估计等^[3]。

标准差与标准误既有明显区别,又密切相关:标准误是标准差的 $1/\sqrt{n}$;二者都是衡量样本变量(观测值)随机性的指标,只是从不同角度来反映误差;二者在统计推断和误差分析中都有重要的应用^[7]。

2 标准差与标准误的表达现状

2.1 各种字母符号的表示 目前各期刊对“平均数 \pm 标准差”和“平均数 \pm 标准误”的符号表示,在不区分大小写、正斜体等的情况下,共有 30 种之多(表 1),比较混乱,有些表示也不确切。同一期刊对同一种误差的符号表示也有多种不同的形式,基本上按照作者所写的给出,没有再加工统一成标准的或约定俗成的符号表示。

从表 1 可以看出,“平均数 \pm 标准差”和“平均数 \pm 标准误”基本上各形成 2 大类表示方式,一类是用统计量的变量符号表示,另一类是用统计量的英文词缩写表示。如 \bar{x} 、 s 、 $s_{\bar{x}}$ 分别是平均数、标准差、标准误的符号;而 M 、 $S. D.$ 、 $S. E.$ 分别是平均数 $mean$ 、标准差 $standard deviation$ 、标准误 $standard error$ 的英文缩写。一个统计学结果表达式中,如果全用变量符号或全用中、英文名称表示都可,但不能用变量符号与量名称(中、英文名称)相加减。

此外,作者利用《中国期刊网》进一步对主要的几种表示统计量及统计学结果(列表表示者)的符号的使用频率进行了统计(均不区分大小写,不区分 x 或 \bar{x} ,不区分正斜体及上下标),结果如下:

1) 对于“ \pm 标准差”,2003 年有 34 984 篇文章用“ $\pm s$ ”表示,有 1 448 篇用“ $\pm SD$ ”(或 $\pm S. D.$);而对于“平均数 \pm 标准差”,有 29 792 篇文章用“ $\bar{x} \pm s$ ”表示,

763 篇用“ $\bar{x} \pm SD$ ”(或 $\bar{x} \pm S. D.$)。

表 1 平均数 \pm 标准差、平均数 \pm 标准误的常见符号表示(2003 年)

序号	表示方式	杂志名称	卷(期):页
1	$\bar{x} \pm s$	暨南大学学报(医学版)	24(4):34
2	$x \pm s$	现代农业	2(1):22
3	Mean $\pm s$	右江民族医学院学报	25(4):497
4	$M \pm s$	福建医药杂志	25(4):147
5	变量名 $\pm s$	沈阳药科大学学报	20(4):300
6	$\bar{x} \pm SD$	上海水产大学学报	12(3):240
7	$x \pm SD$	中国野生植物资源	22(6):52
8	mean $\pm SD$	植物遗传资源学报	4(4):327
9	mean $\pm S. D.$	应用生态学报	14(7):1098
10	$M \pm SD$	中国农业科学	36(12):1556
11	平均数 $\pm SD$	色谱	21(3):298
12	变量名 $\pm SD$	淡水渔业	33(2):55
13	mean $\pm SE$	动物学报	49(2):174
14	mean $\pm SEM$	中国医学运动杂志	22(5):455
1	$\bar{x} \pm s_{\bar{x}}$	中国危重病急救医学	15(12):723
2	$x \pm s_{\bar{x}}$	中国媒介生物学及控制杂志	14(2):89
3	$\bar{x} \pm s_{\bar{x}}$	中国矫形外科杂志	11(13):905
4	$x \pm s_{\bar{x}}$	癌症	22(3):315
5	$X \pm S_{\bar{x}}$	中华医学丛刊	3(5):38
6	mean $\pm S_{\bar{x}}$	食品科学	24(7):122
7	$M \pm s_{\bar{x}}$	华夏医学	16(2):235
8	$\bar{x} \pm SE$	南京农业大学学报	26(4):128
9	$x \pm SE$	动物学研究	24(5):393
10	mean $\pm SE$	植物营养与肥料学报	9(3):382
11	mean $\pm S. E.$	水生生物学报	27(4):435
12	$M \pm SE$	生态学报	23(11):2495
13	平均数 $\pm SE$	植物生理学通讯	39(6):617
14	变量名 $\pm SE$	南京林业大学学报	27(6):61
15	mean $\pm SD$	作物学报	29(4):612
16	mean $\pm SEM$	畜牧兽医学报	34(6):606

2) 对于“ \pm 标准误”,2003 年有 120 篇文章用“ $\pm s_{\bar{x}}$ ”表示,有 218 篇用“ $\pm SE$ ”(或 $\pm S. E.$);而对于“平均数 \pm 标准误”,有 91 篇用“ $\bar{x} \pm s_{\bar{x}}$ ”表示,55 篇用“ $\bar{x} \pm SE$ ”(或 $\bar{x} \pm S. E.$)。

3) 对于“平均数”,绝大部分用“ \bar{x} ”表示,也有部分用“ x ”“mean”“ M ”“均值”“均数”等表示。

GB 3358—82《统计学名词及符号的规定》已对一些常用统计学符号作了明确规定:统计学符号一般用斜体,但有大小写之分,如 n (样本大小)、 \bar{x} (样本的算术平均数)、 s (标准差)、 $s_{\bar{x}}$ (标准误)、 $\bar{x} \pm s$ (平均数 \pm 标准差)、 $\bar{x} \pm s_{\bar{x}}$ (平均数 \pm 标准误)等。因此,正确的表示是:“平均数 \pm 标准差”为“ $\bar{x} \pm s$ ”;“平均数 \pm 标准误”为“ $\bar{x} \pm s_{\bar{x}}$ ”。而英文“Mean $\pm S. D.$ ”“Mean $\pm S. E.$ ”和用中文“平均数 \pm 标准差”“平均数 \pm 标准误”都属于用量名称表示,除此之外,均属于不正确的表示。

2.2 多符号表示以及意义混淆的原因分析 由表 1 可见,标准差与标准误的符号表示混乱,甚至有些作者将标准差与标准误混淆,造成这种情况的原因有多种。

1) 统计量的名称混乱。标准差、标准误、平均数等

统计术语的名称,多则十几个,少也有几个,另外还有简称,非常混乱。如总体标准差的名称就有十几种,而总体平均数也有几种,如称为数学期望、真值等。

2)同一种统计量的多符号表示。如有的用 μ 代表总体平均数(期望),也有的用 x_0 代表真值,而真值也就是总体平均数,只是使用场合不同而已;样本标准差 s 也有用 \hat{s} 表示的^[8];因此,免不了对 \bar{x} 、 s 、 s_x 、 $s_{\bar{x}}$ 、 μ 等常见符号所代表变量的意义不清楚。

3)标准差、标准误差的称呼混乱。有的根据统计量的计算中是用误差还是用偏差,把式(1)~(4)分别称为测量列的标准误差(σ)、测量列的标准偏差(s)、平均数的标准误差($\sigma_{\bar{x}}$)、平均数的标准偏差($s_{\bar{x}}$)计算公式^[5];也有的不分误差、偏差,统一简称为标准差,将式(1)~(4)分别称为单次测量标准差、实验标准差、算术平均数标准差及其估计^[9]。

4)各种数理统计参考书,对统计量的称呼及字母表示不尽相同,对一些基本统计量,如误差、偏差、标准误差、标准差等基本概念没有统一详尽的解释。

5)使用统计分析软件对实验数据作统计分析后,所得各统计量均用英文或字母缩写表示。如平均数、标准差、标准误分别用 Mean、Std. Dev(或 SD)、S. E. Mean 表示^[10],因此,有很多作者也用英文名称的缩写作为变量符号,但忽视了不能用变量符号与英文名称的缩写混在一起作表达式。

6)有些作者使用变量符号很随意,甚至误用,而同一学科期刊的读者、编者互相参照,不免以讹传讹。

3 结束语

目前对实验结果用计量数据表示是科技论文写作中的一大进步^[11],但不少期刊对统计学结果的符号表达非常混乱,有的仅有统计学数据,文中却没有任何文字或符号解释,也有的甚至将标准差与标准误混为一谈。为了避免这些问题,可采取如下相应的措施。

首先,编辑应明确一些基本统计量的意义及特征,对统计学结果表达式有一个准确的理解,尤其对一些

统计量的简称,如统计学结果表达式“平均数 \pm 标准差”和“平均数 \pm 标准误”中的简称“标准差”和“标准误”,实际上是指“样本标准差”和“平均数标准误”,它们的英文表达为“mean \pm S. D.”“mean \pm S. E.”,变量符号为 $\bar{x} \pm s$ 、 $\bar{x} \pm s_{\bar{x}}$ 。

第二,可在《稿约》中约定一些常用统计学变量的符号表示,尤其统计学数据表达式的意义及变量符号。

第三,凡是包含有统计学方法的文章中,均应当在“材料与方法”中给出所用的统计学方法,并标明统计学数据的变量名称及符号表达式;同时在列表(图)中也应当给出统计学数据的变量表达式及其样本数(n)。

4 参考文献

- [1] 李谦. 纵谈标准差[J]. 广东电力, 2001, 14(2): 76-78
- [2] 朱洪海. 随机误差的标准差及其应用[J]. 机械设计与制造工程, 2001, 30(6): 34, 49
- [3] 李春喜, 王志和, 王文林. 生物统计学[M]. 北京: 科学出版社, 2002: 13-40
- [4] 邬实光. 昆虫生态学的常用数学分析方法[M]. 北京: 农业出版社, 1985: 107
- [5] 张敏, 袁辉. 关于标准差应用问题的讨论[J]. 郑州工业大学学报, 1996, 17(3): 95-100
- [6] 刘钢. 撰写医学期刊稿件的统计学参考[J]. 白求恩医科大学学报, 2001(1): 108-110
- [7] 凌树森. 试验数据的统计处理和误差分析: 第6讲 误差分析(续)[J]. 理化检验: 物理分册, 2001, 37(9): 410-414
- [8] 凌树森. 试验数据的统计处理和误差分析: 第3讲 参数估计和假设检验[J]. 理化检验: 物理分册, 2001, 37(3): 133-137
- [9] 李谦. 试验数据的统计处理方法: 总体均值和总体标准差的估计[J]. 西北电力技术, 2000(2): 62-64
- [10] 卢纹岱, 朱一力, 沙捷, 等. SPSS for Windows 从入门到精通[M]. 北京: 电子工业出版社, 1998
- [11] 孙长生, 何扬举. “ $\bar{x} \pm s$ ”使用中应注意的问题[J]. 编辑学报, 2003, 15(2): 92

(2004-08-16 收稿; 2004-10-20 修回)

中文论著中外国人名内的圆点用哪一种?

问: 在中文论著中, 外国人名内的圆点用中圆点还是下圆点?

答: GB/T 15834—1995《标点符号用法》明确指出: “外国人和某些少数民族人名内各部分的分界, 用间隔号标示。”这里的间隔号就是人们常说的中圆点。

例如: 赫伯特·乔治·韦尔斯;

赫·乔·韦尔斯;

赫·G·韦尔斯;

H·G·韦尔斯。

但也有人认为, 国家标准说“用间隔号”是针对人名全译成中文而言的, 如果名字采用外文缩写, 则应在缩写字母后采用下圆点。例如: 赫·G·韦尔斯, H. G. 韦尔斯。

空间隔号也好, 用下圆点也成, 重要的是全刊或全书用法应保持一致。 (郝欣)