

农业和生物学期刊 应重视对表格中数据统计分析结果的审校

王 音

中国农业科学院植物保护研究所《植物保护》编辑部,100193,北京

摘要 抽样调查部分农业和生物学期刊表格中数据统计分析结果的标注情况,针对发现的问题总结出在数据统计分析标注方面的常见错误,提出了相应的对策和建议。

关键词 农业和生物学期刊;数据统计分析;标注

Emphasis on review and revision of results of data statistical analyses in tables in agricultural and biological periodicals // WANG Yin

Abstract Twenty-four periodicals were sampled to investigate the marking status for the data statistical analysis results in the tables. Mistakes in the markings of data statistics are summarized according to the problems found in the samples, and countermeasures and suggestions are proposed.

Key words agricultural and biological periodical; data statistical analysis; marking

Author's address Editorial Department of Plant Protection, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, 100193, Beijing, China

生物科学和农业科学是以试验为基础的学科,常常根据大量的试验数据来得出相应的结论,试验数据的准确性和可靠性关系到结论的正确性和科学性;因此,试验数据在这类论文中占有重要地位。这类论文中除要有合理的试验设计外,必须对试验数据进行科学的整理与分析,几乎所有试验的结果,都需要用适当的统计分析方法^[1]来处理,以找出规律性并作出正确的判断,由此得出所揭示的原理及其普遍性^[2]。生物统计能判断处理间的差异究竟是本质的,还是偶然误差所引起的,从而使试验结果更加准确和可靠^[3-4]。

笔者所编刊物来稿中涉及大量数据统计分析结果的表达,最常见的是在表格中用大小写字母来表示差异显著性分析结果。在稿件审读和编辑过程中发现,无论作者、审稿专家还是编辑都对论文中数据统计分析结果的核对不重视,表现在作者来稿中存在统计分析错误,专家审稿时在这方面内容不进行审查,编辑退修时也没有核查;因此,稿件到了定稿阶段仍然存在严重的数据统计分析问题,使论文的科学性大大降低。

为了弄清这种现象是否普遍存在,笔者共抽查了13个省市出版的24种生物学和农业期刊,其中20种为中文科技核心期刊,影响因子为0.1~0.8^[5]。每种

刊物均随机选取2009年的某一期,对其中涉及差异显著性分析的所有文章及表格进行检查,共核对了133篇文章的275个表格,其中有标注错误的表格54个,占19.6%,共发现错误77处。24种刊物中只有5种在所抽查这一期中未发现错误,其余19种均存在各种问题,错误率在30%以上的有6种。从此结果可以看出,目前生物学和农业期刊论文中数据分析结果存在着严重的问题,必须引起相关期刊编辑的重视。

1 数据统计分析常见问题和编辑核查重点

考虑到数理统计方法并非大多数编辑之所长,本文不讨论具体的数据统计处理方法和原理,而是将出现数据统计问题的表面现象进行归纳总结,使编辑能够通过这些表象发现有问题的数据,从而将信息反馈给作者,及时对数据处理结果进行核对、调整、补充和修改。

1.1 未进行数据统计分析 在处理作者来稿时发现,许多作者在试验方法中已写明对不同处理设置了3次或3次以上重复;但在结果部分仅将数据列出却没有进行统计分析,这样的数据不能完全反映隐藏在数据中的本质规律,即无法确定不同处理间的差异是由于偶然因素的影响还是本质上的差异。在这种情况下,应该对数据进行差异显著性测验,常用的有 t 测验、卡方测验和 F 测验。编辑若发现存在此问题,应及时要求作者补充数据统计分析结果。

1.2 统计分析方法不正确 目前作者一般使用现成的统计分析软件,如SAS、Statistics、SPSS、DPS、Excel等进行数据分析,分析数据时作者只需将试验数据按一定要求输入计算机中,然后由统计软件进行计算后给出分析结果。虽然不需要作者亲自动手进行相应计算,但必须能够正确选择所用的分析程序和正确输入需要处理的数据。最常见的分析方法方面的错误是没有在同一条件下对数据进行差异显著性比较,而是将不同条件下的试验数据进行比较,这些数据之间本没有可比性却被放在一起进行了统计分析,导致得出错误的结论。编辑应注意核查产生数据的条件是否一致,进而判断数据之间是否具有可比性。

1.3 表示差异显著性的字母标注错误 表示差异显著性的字母标注的一般规律为:用小写英文字母表示 0.05 水平的差异显著性,用大写英文字母表示 0.01 水平的差异显著性。按照数据从大到小或从小到大依次标注,数据后具有不同小写字母者表示差异显著($P < 0.05$),具有不同大写字母者表示差异极显著($P < 0.01$)。为便于说明和理解,特从本次抽查刊物出现的错误中摘录部分数据拼接组成常见错误虚拟表(表 1)。若出现下述现象,则说明数据统计分析结果有错误。

表 1 常见数据统计分析结果标注错误虚拟表

处理	平均产量/kg	糖含量	糖含量/%	致死率/%	SOD 活性/ ($U \cdot g^{-1} \cdot min^{-1}$)
1	8484aA	6.0b	8.17aA	14.8 ± 0.8cde	476.22aA
2	8955bcA	4.0c	7.70bB	36.6 ± 2.8	339.40dCD
3	9159bA	6.0ab	7.23dC	23.4 ± 2.1bc	369.14cC
4	8914cA	7.0a	7.18dC	25.1 ± 4.5b	435.53bB
5	3.0D	6.23dD	18.0 ± 2.9bcd		306.38eEd

1)表示差异显著性的字母未按数据从大到小或从小到大顺序依次排列,而是出现了跳跃。如表 1 的第 2 列中,8 484 < 8 914 < 8 955 < 9 159,而 0.05 水平的标注却出现了 a—c—bc—b 的顺序,假设现有标注反映的数据间的差异显著性是对的,正确的标注顺序应为 a—b—bc—c。

2)同列(行)测验对象相同的数据后标注的字母却不一样。如表 1 的第 3 列中 2 个为 6.0 的数据后分别标注了 b 和 ab。

3)缺少某个字母,即字母标注出现跳跃。如表 1 中第 4 列 0.05 水平测验结果没有 c,由 b 直接跳到 d。

4)0.01 水平的标注与 0.05 水平的标注出现了矛盾,即 0.05 水平差异不显著,而 0.01 水平却显著。如表 1 第 4 列中 7.23、7.18 与 6.23 这 3 个数据之间在 0.05 水平差异不显著,但 7.23 和 7.18 却与 6.23 在 0.01 水平上差异显著,这违背了统计学的基本原理。

5)个别数据标注缺失,如表 1 第 5 列中 36.6 ± 2.8 后面未标注。

6)未按顺序标注字母,如表 1 中第 5 列 18.0 ± 2.9 后面的标注为 bdc,应为 bcd。

7)大小写字母交叉。如表 1 第 6 列数据 306.38 后面标注为 eEd,出现了字母交叉,由于大、小写字母代表不同的显著水平,因此,这种交叉是错误的。从书写形式看,应为 deE 或 eDE,但结合其他数据一起看,若为 306.38deE,则与 339.40dCD 有矛盾,即两者在 0.05 水平差异不显著却在 0.01 水平显著;因此,可以排除第 1 种可能。若为 306.38eDE,与其他数据的标

注不矛盾,理论上是成立的,但不能排除原标注中 d 为误标字母,其实际标注应为 eE 的可能:所以,遇到这种情况,编辑最好与作者联系,请作者自行核对,不要擅自修改。

8)同一列(或行)测验中个别数据后面标注的字母出现了与其他数据后面字母大小写不同的情况。如表 1 中第 3 列最后一个数据 3.0 后面标注的为大写字母 D,其他数据后均为小写字母。这种情况为误标或录入排版问题,编辑可以代为修改,将 D 改为 d。

1.4 表注标注有误 从作者来稿和已发表文章看,大多数表格是对同列数据作差异显著性分析,但也有一些情况下需要对同行数据进行差异显著性测定,在表注中一定要正确标注分析对象。表 2 就出现了标注错误:本应进行同行比较,即比较取食不同食物条件下伪钝绥螨某一虫态发育历期的差异,因标注错误变为比较取食某种食物时螨不同虫态历期的差异,根据常识可知不同的虫态历期没有可比性。

表 2 伪钝绥螨取食花粉和二斑叶螨各螨态的发育历期

发育阶段	发育历期/h	
	花粉	二斑叶螨
卵	45.94a	50.88b
幼螨	26.17a	22.08b
前若螨	44.23a	39.36b
后若螨	46.40a	36.24b
未成熟期	162.74a	148.56b

注:历期后面的英文字母为 Duncan 新复极差测验法,同一列中小写字母不同,表示在 0.05 水平差异显著,否则不显著。

1.5 格式方面的问题

1)差异显著性分析结果指代不明。例如,某表从左到右依次列出了不同密度条件下所得出的平均产量、增产数量、增产率、单株生产力、投资、收益等 6 项数据指标,并将 LSR 测验结果放在表格的最后 2 列。一般来讲,这个测验结果应该是针对倒数第 3 列数据“收益”的,但仔细阅读全文才发现,这个测验结果实际是来自于对“平均产量”数据的测验。由于该数据位置与表示和其相关的差异显著性测验结果的字母相距甚远,又未在表注中加以说明,使读者很难直接看出两者的关系,易对读者造成误导,同时违反了“表格应具有自明性”的规定^[6]。

2)同一个表有些数据标注大写字母,有些标注小写字母,且无表注。表 3 中有 2 组数据应引起编辑注意:一组是吐絮期的载铃量,2 个数据一个标注大写字母一个标注小写字母,两者未在同一水平下进行比较,肯定有错误;另一组是盛铃期的根冠比数据,与其他数据用大写字母标注不同,该组数据均以小写字母标注。这种情况有 2 种可能:一是本应与其他数据一样也标

注大写字母而在打字时误标;二是该组数据仅在 0.05 水平显著,而其他数据都是在 0.01 水平显著。由于该表没有表注,所以很难判断属于哪种情况。编辑应与作者联系,明确作者原意。此次抽查发现有些期刊这种标注情况较多,因此建议:如果作了 0.05 和 0.01 这 2 个水平的测验,最好将分别代表不同水平的大、小写字母都标出,同时在表注中说明大小写字母所代表的显著水平。这样表达的结果既完整又易于读者理解。

表 3 不同时期单位叶面积载铃量与根冠比差异

品种	载铃量(g/m ²)			根冠比		
	盛花期	盛铃期	吐絮期	盛花期	盛铃期	吐絮期
甲	121A	303A	873A	0.04B	0.13b	0.03B
乙	43B	158B	438b	0.09A	0.20a	0.09A

3)同一表格中标注顺序不一致。如有的数据按从大到小的顺序标注,还有的按从小到大的顺序标注。同一个表格甚至同一篇文章最好能按同一顺序标注。

4)不标注测验的显著水平。虽然大多数情况下作者都是用小写字母代表 0.05 水平,大写字母代表 0.01 水平,但有时根据需要可以用小写字母表示对同列数据进行测验,用大写字母表示对同行数据的测验结果;因此,最好每个表格进行相应的标注,使表格具有自明性。

5)字母顺序排列错误。如本应为 80.89bcd,却标注为 80.89bdc。虽然 3 个字母都在,但按照统计分析过程中表示差异显著性字母的生成原理,这样排列是不对的,应予改正。

6)同一种期刊标注风格不一致。有的表中以主体字母标注,有的表用上标来标注。虽然不算作错误,但对于同一种期刊应有统一的规定,笔者推荐用主体字母标注。

2 对策和建议

由于编辑缺乏数据统计分析的基础和意识,不重视对统计结果的检查核对,导致刊出的许多文章在数据方面存在严重的错误,使得论文的结论失去了科学价值。考虑到编辑是整个编辑出版工作的调控者^[7],因此,要避免带有数据分析方面错误的文章刊出,编辑必须从作者、审稿人和编辑自身 3 个方面做好工作。

1)强化编辑对数据的审查意识。数据是科技论文的核心,结论来源于对数据的分析;因此,编辑审读文章时应将审查数据作为重点之一,通过一些表象来检查数据处理方面是否存在问题。只有把好数据审查关,才能保证论文结论的科学性。

2)将审查数据的要求传达给审稿人。审稿人都是各个领域的专家,在数据分析方面的知识或许比编辑更丰富,如果能明确要求他们审稿时对数据处理结果进行认真审核,将避免许多错误。

3)对作者提出相应要求。可以在征稿简则中明确要求作者对数据进行统计分析,并认真核对分析结果。

4)编辑把好初审和复审关。编辑在对稿件进行初审时就可以对数据的处理情况作初步检查,发现存在明显错误,可与作者联系,让其修改后再投稿。对已经通过外审的稿件,若审稿人没有提出统计分析方面的问题,编辑可以再核查,以确定审稿人是否对这方面进行了审查。

5)编辑应学习并掌握本学科所涉及的数据处理方法,以便能及时发现问题。

3 参考文献

- [1] 杨志伟. 实验数据的统计分析与科研论文的撰写[J]. 中国比较医学杂志, 2009, 19(1): 76-78
- [2] 谭云. 论科技期刊来稿中常见问题[J]. 粮油食品科技, 2009, 17(3): 68-69
- [3] 陈年春. 农药生物测定技术[M]. 北京: 北京农业大学出版社, 1991: 95
- [4] 北京林学院. 数理统计[M]. 北京: 中国林业出版社, 1985: 147
- [5] 中国科学技术信息研究所. 2009 版中国科技期刊引证报告: 核心版[M]. 北京: 科学技术文献出版社, 2009
- [6] GB/T 7713—1987 科学技术报告、学位论文和学术论文的编写格式[S] // 中国科学技术期刊编辑学会. 科学技术期刊编辑教程[M]. 2 版. 北京: 人民军医出版社, 2007: 479
- [7] 张琦. 谈如何快速高效编辑加工研究生论文[J]. 山东教育学院学报, 2009(1): 105-107

(2010-02-08 收稿; 2010-03-17 修回)

更 正

本刊 2009 年第 4 期第 334 ~ 336 页发表的王晴、杜冰的《高校医学期刊集群发展的探讨》一文中,第二作者杜冰应为通信作者,虽然在篇首页地脚处加注了“† 通信作者”,但在“杜冰”的右上角未标注“†”。特此更正,并向作者表示歉意。