

科技期刊论文中频数(或频率)分布直方图的编辑加工

黄 鹂

长江大学期刊社,434023,湖北荆州

摘要 针对科技期刊论文中频数(或频率)分布直方图坐标轴不清,频率、频数不分,分组区间界定不明,图题笼统,图的形状不规范等问题,根据 GB/T 3358.1—2009《统计学词汇及符号 第1部分:一般统计术语与用于概率的术语》以及有关文献中的内容,提出了规范化的直方图编辑加工方法并给出了应用实例。

关键词 频数;频率;直方图;编辑加工

Editing the sample number (or frequency) distribution histogram in Chinese sci-tech journals // HUANG Li

Abstract There are many problems in sample number (or frequency) distribution histograms in Chinese sci-tech journals, such as unclear coordinates, non-distinguishing frequency and sample number and unclear sorting and diagram titles as well as nonstandard diagram shapes. Based on GB/T 3358.1—2009 "Statistics vocabulary and symbols—Part 1: General statistical terms and terms used in probability" and relevant documented information, some cases and methods of editing such histograms are provided in the paper.

Keywords sample number; frequency; histogram; editing

Author's address Periodical Press of Yangtze University, 434023, Jingzhou, Hubei, China

在编辑工作中,经常会碰到带有频数(或频率)分布直方图(以下简称“直方图”)的稿件。由于作者提供的图自明性不好,需要编辑加工时修改补充,于是希望通过学习其他期刊对这种图的加工方法来指导自己

编辑学会肿瘤专业委员会制订更明确、具体的伦理规则,以适应科学技术与时代发展的需求。

5 参考文献

- [1] 吴泰相,李幼平,郑尚维,等. 推动我国医学期刊发表伦理发展的联合声明[J]. 中国全科医学,2011,14(34):4001-4002
- [2] 李幼平,吴泰相,李静,等. 创建中国临床试验注册和发表机制的联合宣言[J]. 中国循证医学杂志,2006,6(6):393-394
- [3] 靳光华. 科技期刊论文多作者署名引发的伦理问题思考[J]. 中国科技期刊研究,2006,17(1):109-111
- [4] 关键. 生物医学研究涉及的伦理问题及其审核要求: II. 医学动物实验[J]. 协和医学杂志,2011,2(4):393-395
- [5] 关键. 生物医学研究涉及的伦理问题及其审核要求: III. 人体试验[J]. 协和医学杂志,2012,3(1):124-126

的工作实践;因此,笔者收集了一些期刊上发表的直方图。在整理这些直方图的过程中,笔者发现其中出现了不少条形图,而且直方图的表达形式也比较混乱,不便于读者阅读理解。为了使直方图的编辑加工有规范可循,笔者学习了 GB/T 3358.1—2009《统计学词汇及符号 第1部分:一般统计术语与用于概率的术语》^[1]以及文献[2-3]中的有关内容,指出了目前科技期刊论文中频数(或频率)分布直方图表现形式存在的问题,给出了解决办法,并用实例作了具体说明。

1 直方图和条形图比较

GB/T 3358.1—2009^[1]对“直方图”的定义是:频数分布的一种图形表示,由一些相邻的长方形组成,每个长方形的底宽等于组距,面积与组的频数成比例。对“条形图”的定义是:由一组宽度相同、高度与频数成比例的长方形组成的,表示名义特性频数分布的图形(注:条形图中的长方形并不需要相邻)。

根据 GB/T 3358.1—2009 的定义以及文献[2-3]的内容,对直方图与条形图进行了对比,结果如下:

1) 直方图横轴上的数据是连续的,是一个范围。条形图横轴上的数据是孤立的,是具体的数据。

2) 直方图用长方形的面积表示频数,长方形的面积越大,表示这组数据的频数越大;只有当长方形的底

- [6] 张浩,黄瑾,杨放. 加强伦理委员会建设,促进临床试验健康发展[J]. 药学实践杂志,2009,27(2):137-139
- [7] 唐天驷. 对科技期刊论文伪造现象的反思[J]. 中国脊柱脊髓杂志,2007,17(10):726-727
- [8] 中国科技期刊编辑学会医学委员会肿瘤专业委员会在穗成立[J]. 肿瘤,2001,21(1):50
- [9] 冯远景,陈希宁,于长谋. 科技期刊审稿人的权利与义务[J]. 中国科技期刊研究,2001,12(4):246-249
- [10] Hames I. 科技期刊的同行评议与稿件管理:良好实践指南[M]. 张向谊,译. 北京:清华大学出版社,2012: VII
- [11] 汪挺,阮星星,刘昱菡,等. 医学期刊编辑的伦理要求:对国际编辑伦理标准的思考与补充[J]. 编辑学报,2012,24(1):100-102
- [12] 荣梅. 论科技期刊编辑的伦理原则[J]. 编辑学报,2006,18(5):328-330

(2013-04-03 收稿;2013-05-26 修回)

宽都相等即组距相等时,才可以用长方形的高表示频数的大小。条形图用条形的高度表示频数的大小。

3)直方图中各长方形对应的是一个范围,由于每2个相邻范围之间不重叠、不遗漏,因此直方图中的长方形之间没有空隙;而条形图中各个数据之间是相对独立的,各个条形之间是有空隙的,并不需要相邻。

2 直方图表达形式存在的问题

通过对收集的科技期刊论文中的频数(或频率)分布直方图进行分析,发现它们主要存在以下问题。

1)将直方图画成了条形图。

2)直方图横轴的坐标标值线不能明确地界定分组区间,有的分组区间不是半开区间。

3)直方图纵轴坐标的名称(即标目)形式多样,如分布频率/%,频率/%,频率,频数,频数/%,频数/个,样品/个,样品数/块,样品数(个),百分比/%,百分数(%),含量(%),数量(%),油气单元(个数)。频率与频数用法混乱,如该用“频数”的用成了“频率”,该用“频率”的用成了“频数”。

4)图题笼统,如“……均一化温度直方图”“……孔隙度-渗透率频率直方图”“……储集层物性分布直方图”“……孔隙度频数分布”“……包裹体测温统计图”“……孔隙度渗透率统计直方图”“沙山迎风坡不同粒级含量”“……油气水平运移距离统计”“……有机碳分布直方图”“……碳同位素对比”“……同位素分布直方图”“……孔隙类型特征”。

3 规范化编辑加工方法

3.1 横轴坐标 直方图横轴坐标反映考察对象的类别,从横轴坐标的名称(即标目)可以了解统计的是考察对象的定性特征还是定量特征,如果统计的是定量特征还需给出对应的量和单位。

如果统计的是定性特征,那么要求横轴坐标的标值线应能清楚地反映统计对象的分组情况:分组的组数(把全体样本分成的组的数量称为组数),每个分组的特征名称。

如果统计的是定量特征,那么要求横轴坐标的标值线应能清楚地反映统计对象的分组情况:分组的组数,每组的组距,分组区间的开闭情况(分组区间必须是半开区间,这样才能保证每个数据都能落入且只能落入某一个区间)。

3.2 纵轴坐标 直方图的纵轴坐标反映的是考察对象的频率与组距之比^[4],只有当组距相同时,才可以用长方形的高即纵坐标的数值(即标值)表示频率(频

数)的大小。由于科技期刊论文中的直方图多数都采用相同的组距,所以本文仅讨论等组距的情况。

纵轴坐标名称采用频数(落在不同小组中的数据数量称为该组的频数)或频率(频数与样本总数的比称为该考察对象的频率)来表示。各分组的频数之和等于这组数据的样本总数。 $0 < f_i \leq 100$ 且 $\sum f_i = 100$ 。其中: f 为频率,用百分数表示; i 为分组的数量, $i = 1, 2, \dots, m$ 。频率大小反映各组频数在数据样本总数中所占的比例。

如果是频率分布直方图,纵轴坐标标目采用“频率/%”,如果是频数分布直方图,则采用“频数”。

纵轴坐标标目是“频率/%”,那么 $\sum f_i = 100$ 。如果是“频数”,那么各统计对象的频数之和($\sum n_i = n$)必须等于样本数据总数 n 。通过这种方法来初步判定作者给出的是频率还是频数分布直方图。

3.3 图的形状 从文献[1]对直方图的定义中知道,直方图是由相邻的长方形组成的图形。

利用 Excel 绘制直方图时,要先通过绘制柱形图,将各柱形图之间的分类间距设置为0后得到,具体步骤是:选中某一个数据系列,单击右键,在弹出的浮动选单中选择“数据系列格式”,单击“选项”选项卡,将“分类间距”设置为“0”,同时勾选“依据数据点分色(V)”复选框,按“确定”键后,柱形图之间的间隔即被取消,成为长方形相连的符合标准要求的直方图。

3.4 图题 图题应能反映考察对象的类别名称及图形的样式名称,而不是用笼统的图题表示。建议增加明确指示图形类型的“频数(或频率)分布直方图”的字样,这样可以与条形图明确区别开来,也便于读者检索。例如可将第2章4)中示例的图题改为“……均一化温度频数分布直方图”“……孔隙度、渗透率频率分布直方图”“……储集层孔隙度、渗透率频数分布直方图”“……孔隙度频率分布直方图”等。

3.5 其他 由于频数(或频率)分布直方图是一种统计图,所以要求图中应给出样本总数。当考察对象不止1个时,即横向指标反映的是多个对象的特征时,须用图例指明。

4 实例应用

图1为摘录自文献[5]中的图件。文献[5]对此的描述是:“对松辽盆地十屋油田营城组34口取心井192条裂缝进行了观察和统计分析后,将裂缝长度分成了7个区间(图1a),前3个长度区间是主力裂缝区,即裂缝以短缝者居多,且以长度为11~15cm的裂缝数量最多。”

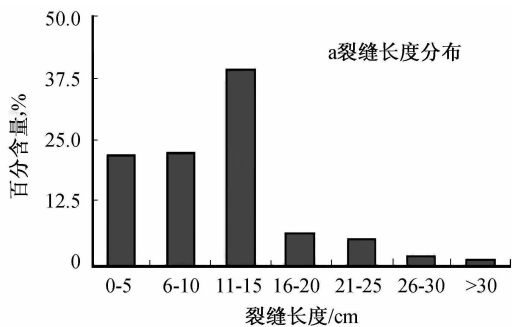
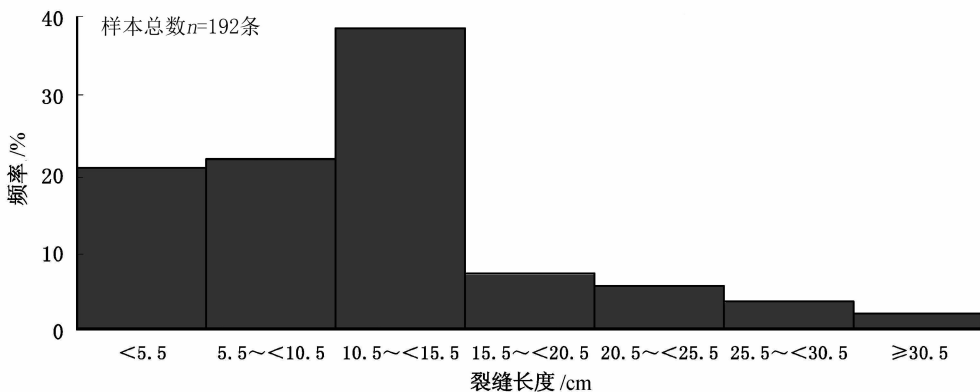


图1 取自文献[5]的图1片段

通过文中的描述,可以知道图1实际上是一张反映裂缝长度的频率分布直方图。它存在如下问题:

- 1) 横轴坐标分组区间不连续,存在遗漏数据的情况(例如,对裂缝长度处于5~6、10~11、15~16 cm等区间的数据如何统计);
- 2) 纵轴标目使用“百分含量,%”,不规范;
- 3) 没有标注进行裂缝长度统计时所用的样本总数 n ;



十屋油田营城组裂缝长度频率分布直方图

图2 对图1进行规范化编辑加工后的图

5 结束语

对于直方图和条形图,应注意区分二者的不同。根据直方图编辑加工规范,可以要求作者按规范修改图件,补充必要的信息,然后编辑再作加工。如此得到的图具有自明性,方便读者阅读理解。

6 参考文献

[1] GB/T 3358.1—2009 统计学词汇及符号 第1部分:一般统计术语与用于概率的术语[S].北京:中国标准出版社,2009

数 n ;

4) 用条形图表示直方图。

为了确认作者给出的图是频率分布直方图,联系到文献[5]的作者,就图中横轴坐标及纵轴坐标的含义作了咨询,最后根据直方图的规范化编辑加工要求重新对图1作了整理,利用 Excel 软件^[6]绘制出如图2所示的频率分布直方图:

- 1) 根据与作者的交流沟通,将横轴坐标分组区间作了调整,使2个相邻的裂缝长度范围不重叠,也没有遗漏。
- 2) 将纵轴坐标标目改为“频率/%”,因为采用的是等组距,所以纵轴坐标标目可以用频数与样本总数之比乘以100%来表示频率。
- 3) 根据文中描述,在图2中补充了用于统计的样本总数192。
- 4) 根据原始数据,利用 Excel 软件制作出长方形相连的直方图。

[2] 杨耀南. 频数分布直方图与条形图岂容混淆[J]. 中小学数学:中学版,2010(3):18-19

[3] 胡春洪,张莹. 这样的统计图是频数分布直方图吗? [J]. 中学数学:初中版,2009(7):43-44

[4] 姜琳. 一表三图话统计[J]. 中学生数理化:高中版,2012(1):8

[5] 许君玉. 松辽盆地十屋油田营城组裂缝识别[J]. 新疆石油地质,2012,33(1):65-68

[6] 人力资源和社会保障部人事考试中心. Excel 2003 中文电子表格[M]. 北京:中国人事出版社,2010:164-166 (2012-11-01 收稿;2013-04-23 修回)

“80 万元以上”不等于“至少 80 万元”

一本编校人员业务培训讲义在分析“销售国债至少 80 万元以上”的语病时说:“80 万元以上”与“至少 80 万元”实际上两者意思是相同的,选择其中一种说法即可”。科学地分析这 2 种说法,意思不完全相同:“80 万元以上”是“> 80 万元”,而

“至少 80 万元”是“≥80 万元”,后者多包含了“80 万元”这一确定数。此例提示我们,在科技书刊中遇到“以上”“以下”等方位词和“至多”“至少”等副词时,一定要科学地加以辨析。

(焦 贞)