

科学评价体系中同行评议与文献计量的实证分析

邓荔萍

(广州华夏职业学院, 广州 510935)

摘要: 通过对比F1000因子与被引频次、F1000因子与期刊评价指标, 并对主要指标进行相关性分析, 来验证同行评议与引文分析间的相关性。结果表明, F1000因子与被引频次呈正相关性, 即专家打分与被引频次变动方向相同; 但也存在专家打分高的论文被引频次低, 而专家打分低的论文被引频次高的事实。相关性分析的结果表明: 在特征因子、SNIP等主要指标中, SJR、IF与专家评议值相关度最大。

关键词: 科学评价; 同行评议; 文献计量

分类号: G250

DOI: 10.3772/j.issn.1673—2286.2014.06.007

1 研究背景与问题提出

继影响因子(IF)之后, 新型的文献计量指标层出不穷, 如特征因子、h指数、SJR、SNIP。这些指标从不同的角度评价科研机构、学术期刊及衡量研究人员的科研绩效, 是对传统文献计量指标的修正和扩展, 也标志着文献计量的深化与成熟。

关于指标的关系, 国内外都做了不少研究。Saad通过特征因子、论文影响分值、h指数皮尔逊相关性分析, 指出两两指标呈高度相关, 相关系数为0.9^[1]。Rousseau等人通过对10多个领域77种期刊的SJR、特征因子、论文影响分值、h指数与IF的分析, 指出这些指标与IF高度相关, 且两两相关^[2]。宋丽萍则通过Wos与Scopus中主要指标相关性分析, 得出IF与SJR、SNIP高度相关^[3]。虽然这些指标高度相关, 但却不能完全替代其他指标。美国国家科学基金会“科学与创新政策计划”主任Julia Lane曾说过“偏颇、狭隘的科学计量会产生偏颇的科学”^[4]。目前, 科学评价中对于如何把同行评议与引文分析结合起来没有唯一的定论。

REF注重科研成果的质量思想将定性评价方法与定量评价方法相结合, 以同行评议为主, 以文献计量指标进行佐证^[5]。本文以生态学及心理学这两门学科的数据为例, 分别以文章、期刊两个研究对象对同行评议与文献计量指标进行相关性分析, 从而探讨文献计量指标是否可以作为同行评议的参考依据。

2 数据来源及处理方法

本文的数据分别来源于F1000、WOS、Scopus。F1000是同行评议的代表, 能够及时准确地评选出学术价值高的文章, 不考虑论文的来源期刊。Wos、Scopus数据库是文献计量的代表。

笔者选取了生态学、心理学两门学科, 以2005年为回溯年限进行统计。通过F1000数据库的高级检索界面, 分别检索到725条生态学的文章记录和703条心理学的文章记录, 共计1428篇文献。检索式为: article publication data=2005 and subject="ecology" or subject="psychology"。对于相同的篇名, 可以利用作者来区分, 统计区间为2005年—2011年7月。采用计算机编程语言Java抓取了文章记录的篇名、期刊名、F1000因子、作者等信息, 并导入Excel表格。

以篇名为检索字段, 利用Scopus引文数据库检索出每篇论文自2005年发表以来的被引频次; 以期刊名为检索字段, 从www.scimagojr.com/journalrank.php网站采集了每种期刊2008年的SJR、h指数。通过荷兰莱顿大学科技研究中心(CWTS)网站(info.scopus.com/journalmetrics/sjr)获得期刊的SNIP值。五年期影响因子、特征因子、均采集自JCR(2010年版), IF来自ISI的JCR(2007年版)。最后, 通过Excel筛选、整理、汇总出F1000、Scopus、Wos三个数据库共同收录的生态学文献、心理学文献以及相应的期刊评价指标信息, 统计

结果如下:

725篇生态学文献中,723篇被Scopus收录,来自176种期刊,例如《科学》、《自然》、《自然遗传学》等;F1000因子最高值的论文刊登在《自然》杂志上,为62分。703条心理学的文献全部被Scopus收录,源自66种期刊,例如《科学》、《自然》;F1000因子最高值为15,该文刊登在《自然》上,具体分布如图1。

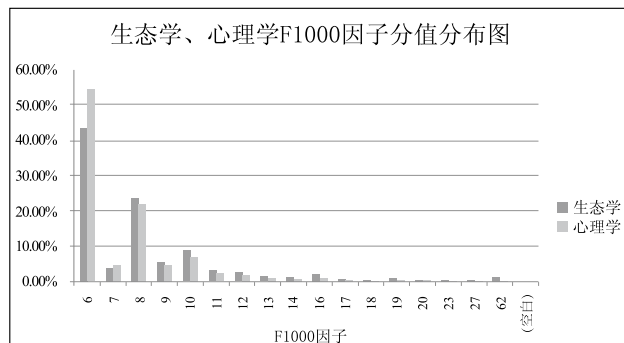


图1 生态学、心理学F1000因子分值分布图

文章采用SPSS18.0分别对生态学和心理学的F1000因子与被引频次、F1000因子与期刊评价指标,如IF、特征因子、SJR、SNIP等六个指标的相关性进行检验。剔除统计中的缺失值,有722篇生态学文献、703篇心理学文献参与该部分的相关性分析。

3 数据的相关性分析

3.1 正态分布检验

采用Kolmogorov-Smirnov Test (K-S检验方法)对生态学的样本数据进行验证,结果表明显著性水平小于0.05,说明拒绝 H_0 。可见,F1000因子、Scopus被引频次都不符合正态分布,所以应该采用Spearman相关分析方法比较适宜。

3.2 F1000因子与Scopus中被引频次相关性分析

722篇生态学文献的F1000因子与Scopus被引频次的斯皮尔曼检验结果表明,两者在0.01的显著性水平上呈正相关,相关系数分别为0.360。703篇心理学文献的斯皮尔曼检验结果表明,F1000因子与Scopus中的被引频次的相关系数为0.403。可见,在生态学和心理学中

两个变量都显著相关,呈现正相关。

3.3 F1000与被引频次的结果分析

为了分析两个变量低度相关的原因,本文首先分析出F1000因子与被引频次差异最大的50篇文献,然后按照同样的方法找出F1000因子与被引频次差异最小的50篇文献,再分别对F1000因子与被引频次进行相关性分析。

(1) F1000与被引频次差异最大的50篇文献

按照F1000因子的分值、被引频次分别从高到低进行排序,同时与相对应的被引频次与F1000因子的位次进行比较,分别筛选出生态学与心理学中F1000因子与被引频次差异最大的50篇文献。由于两个变量的数据具有不可比性,所以先进行标准化处理。采用SPSS描述性中的Z分数和T分数将两个变量的数据转化成为方差为10、平均分50的标准值,进而把两组数据相减,按照相减后绝对值的大小进行排序,筛选出前50组数据,具体信息略去。这50组数据的相关性分析显示,在0.05的显著性水平上,生态学文献50组数据中F1000因子与Scopus被引频次的相关系数为-0.291,呈现负相关。在0.01的显著性水平上,50篇心理学文献中F1000因子与Scopus被引频次的相关系数为-0.527。可知,一些由F1000专家遴选出的高学术价值的文章被引次数低,而被引次数高的论文专家打分却较低。

(2) F1000与被引频次差异最小的50篇文献

用同样的方法获取了F1000与被引频次差异最小的50组数据,具体信息略去。对差异最小的50组数据进行相关性分析。在0.01显著性水平上,生态学、心理学50组数据的相关系数分别为0.673、0.528。因此,F1000因子与被引频次两个变量具有相同方向的变化趋势,即呈现正相关。这说明一篇文章在同行评议中得分越高,其被引次数也越高,学术水平也被广泛认可。相反,一些专家打分低的文章,其被引次数也较低。

(3) F1000与被引频次的分析结果

上述F1000因子与被引频次的相关性分析说明同行评议与文献计量之间具有正相关性。由于F1000因子与被引频次差异最大的50组数据成负相关,而与差异最小的50组数据成正相关,因此最终导致整体上F1000因子与被引频次低度相关。

F1000因子评价的是论文的质量,其分值是固定的。一些具有潜在价值、预测科研未来发展趋势的论

文在一开始得不到认可,但随着时间的推移,其价值逐渐显现出来,研究成果广泛地运用在学术界中,其被引次数也就不断增加。此外,一些专家打分高的论文,其被引受到期刊出版周期等因素的影响有滞后性,一般在论文发表的3年后会出现被引高峰期。论文的被引次数会随着引文峰值的变化而变化。上述原因导致同行评议与文献计量指标之间有相反方向上变化的趋势。

50组差异最小的数据也反应了一个事实:专家打分高的文章,被引频次也高。有些经典的具有学术价值的论文不但得到专家的高度认可,而且其被引频次受到马太效应的影响,会不断地增加;一些专家打分低的论文,其学术价值也不被认可,所以论文被引用的次数也较低。这些原因解释了F1000因子与被引频次差异最小的50组数据,也正好说明了同行评议与文献计量指标在科学评价中的一致性。

3.4 F1000因子与期刊评价指标相关性分析

F1000因子的评价对象是论文,并不考虑论文所发表期刊的影响力。期刊评价指标是以期刊为评价对象,无法评价论文的科研价值。那么F1000因子与期刊评价指标之间有什么关系? IF、SNIP值、SJR值、特征因子、h指数在期刊评价中都各有优势,以F1000数据库作为定性评价的工具时,哪些文献计量指标可以作为专家评议的“参照系”?

3.4.1 生态学 F1000因子与期刊评价指标相关性分析

对生态学176种期刊的影响因子从大到小进行排序,排名前3名的期刊是《自然遗传学》、《自然》、《科学》。表1只简单地罗列了影响因子排在前3位的F1000因子为6的文章记录。

可见,高影响力、声望高的期刊上刊载的并非都是专家打分高的论文。在评估一项科研成果的价值时,用

其他的方式来代替阅读文章本身是不可取的。因此,以是否发表在核心期刊上来判定文章的质量是有失公允的。

在生态学文献中,以IF、h指数、SJR值、SNIP值、特征因子与F1000因子分别进行非参数Spearman相关性分析。自然科学引文周期较短,所以这里不采用五年期影响因子进行相关性分析。

生态学文献的F1000因子与Scopus中被引频次差异最大与最小的50篇中F1000因子与IF、h指数、SJR值、SNIP值、特征因子的斯皮尔曼检验结果如表2所示。

表2 50篇被引与F1000因子差异大/最小的生态学文献中F1000与主要指标的Spearman相关性分析结果

类型	50篇差异大的文献		50篇差异最小的文献	
	相关系数	P值	相关系数	P值
F1000因子 VS SJR	0.351*	0.012	0.496**	0.000
F1000因子 VS SNIP	0.290*	0.041	0.556**	0.000
F1000因子 VS IF	0.357*	0.011	0.507**	0.000
F1000因子 VS h指数	0.311*	0.028	0.339*	0.016
F1000因子 VS 特征因子	0.254	0.075	0.353*	0.012

注: **表示在0.01水平上相关性是显著的(双尾检验)
*表示在0.05水平上相关性是显著的(双尾检验)

表2可见, F1000因子与被引频次差异最大的50篇生态学文献中, 根据相关系数从大到小对指标进行排序: IF>SJR值>h指数>SNIP值>特征因子。差异最小的50篇生态学文献中, 其相关性排序从大到小依次为SNIP值>IF>SJR值>特征因子>h指数。

3.4.2 心理学 F1000因子与期刊评价指标相关性分析

在心理学中, 同样采用非参数Spearman相关分析法验证指标与专家评价价值之间的关系。由于心理学属于社会科学的分支学科, 人文社会科学的学术影响力具有滞后性, 一般在论文发表或著作出版若干年后才

表1 生态学F1000因子为6的文章发表在影响因子TOP3的期刊上

篇名	期刊名	F1000因子	影响因子
Natural variation in cardiac metabolism and gene expression in <i>Fundulus heteroclitus</i>	Nature Genetics	6	36.377
Insect communication: 'no entry' signal in ant foraging	Nature	6	36.104
Transoceanic migration, spatial dynamics, and population linkages of white sharks	Science	6	31.377

能体现出真正的学术价值。所以数据分析过程中不采用影响因子。

心理学文献的F1000因子与Scopus中被引频次差异最大与最小的50篇中F1000因子与五年期影响因子、h指数、SJR值、SNIP值、特征因子的斯皮尔曼检验结果如表3所示。

表3 50篇被引与F1000因子差异大/最小的心理学文献中F1000与主要指标的Spearman相关性分析结果

类型	50篇差异大的文献		50篇差异最小的文献	
	相关系数	P值	相关系数	P值
F1000因子 VS SJR	0.348*	0.011	0.502**	0.000
F1000因子 VS SNIP	0.312*	0.028	0.480**	0.000
F1000因子 VS IF	0.360*	0.009	0.552**	0.000
F1000因子 VS h指数	0.311*	0.035	0.340*	0.016
F1000因子 VS 特征因子	0.268	0.075	0.560*	0.000

注: **表示在0.01水平上相关性是显著的(双尾检验)

*表示在0.05水平上相关性是显著的(双尾检验)

表3中, F1000因子与被引频次差异最大的50篇生态学文献中, 根据相关系数从大到小对指标进行排序: 五年期影响因子>SJR值>SNIP值>h指数>特征因子。差异最小的50篇心理学文献中, 相关性排序从大到小依次为特征因子>五年期影响因子>SJR值>SNIP值>h指数。

3.4.3 F1000因子与期刊评价指标的结果分析

从上述相关系数排序中, 可以看出每组F1000因子与各指标的排序顺序都不相同。对于生态学而言, F1000因子与被引频次差异最大的50篇文章中, IF、SJR值与F1000因子之间的非参数关系最大, 表明在对专家打分与文章被引次数差异较大的情况下, 用F1000对文章质量定性评价时, 可以IF、SJR值作为同行评议的参照指标; F1000因子与被引频次差异最小的50篇文献中, 相比其他指标, SNIP值、IF与F1000之间的相关度最大, 表明在专家打分与文章被引频次相一致的情况下, 在定性分析的基础上, SNIP值、IF这两个指标完全可以对专家评议的结果进行佐证。对于心理学来说, F1000因子与被引频次差异最大的50篇文章中, 五年期影响因子、SJR值与F1000的相关系数最大, 表明参照这两个指标可以为同行评议提供依据; 反之, 则以特征因子与

五年期影响因子作为同行评议的参照系。

总之, IF或五年期影响因子完全可以对专家评议的结果进行佐证。同时, 在不同情况下, 生态学可分别结合SJR值、SNIP值作为同行评议的辅助指标; 心理学则分别结合SJR值、特征因子作为辅助即可。

4 结语

从以上统计数据分析以及相关分析, 得出以下结论:

(1) 同行评议与文献计量指标具有正向相关性, 专家认可的文章, 其被引次数高; 反之, 专家评分低的文章, 其被引率低。所以在科学评价中结合同行评议与文献计量是有必要的。

(2) 从同行评议结果与文献计量指标的负相关性可以看出, 一些专家认可为重要成果的优秀论文, 由于期刊出版滞后性等各种原因, 被引次数低。所以, 单独以文献计量指标作为评价标准会导致过于看重从外在特征反映成果价值, 必然不足以全面评价科研成果的质量。而同行评议由于其主观性强, 过程复杂, 也有许多局限性。

(3) F1000因子与期刊评价指标的相关性分析结果表明, 在对论文的质量评价之后, 可以适当的期刊评价指标作为同行评议结果的依据。因此, 对学术质量、学术影响力两方面的评价都应将同行评议结果、被引频次与论文所在期刊的影响力三者结合起来, 才可以做到科学、公正的评估。文献计量指标是对同行评议结果的补充, 不能完全取代同行评议。同时, Harnad认为同行评价被认为是验证计量指标有效性的自然标准^[6]。可见, 文献计量与同行评议相结合的方法是未来科学评价方法中不可阻挡的发展趋势。

参考文献

- [1] SAAD G. Convergent validity between metrics of journal prestige: the Eigen factor, article influence, and h-index scores. Preprint.
- [2] Ronald Rousseau, the Stimulate 8 Group. On the relation between the Wos impact factor, the Eigen factor, the SCImago Journal Rank, the Article Influence Score and the journal h-index. Preprint.
- [3] 宋丽萍. 相关性视角下的Wos与Scopus之比较[J]. 图书情报工

作,2012,56(4):22-26.

http://www.hefce.ac.uk/pubs/hefce/2009/09_39/09_39.pdf.

[4] LANE J. Let's make science metrics more scientific [J]. Nature,2010, 464(25): 488-489.

[6] HARNAD S. Validating research performance metrics against peer rankings [J]. Ethics in Science and Environmental Politics, 2008, 8(11): 103-107.

[5] Report on the pilot exercise to develop bibliometric indicators for the Research Excellence Framework [EB/OL]. [2013-10-02].

作者简介

邓荔萍, 女, 1986年生, 硕士生, 广州华夏职业学院助教。E-mail: quietywinter@yeah.net。

Empirical Study on the Association between Expert Review and Bibliometrics in Research Assessment

DENG LiPing

(Guangzhou Hua Xia Technical College, Guangzhou 510935, China)

Abstract: In order to verify the association between expert assessment and bibliometric indicators, this paper compares the F1000 factor with cited times of articles. The result shows that there is a significant positive correlation between expert assessment of importance and impact as measured by number of citations overall. However, analysis indicates that there are some exceptions that some papers highly rated by experts were not highly cited, and vice versa. Meanwhile, the correlations analysis shows that the SJR and IF factor are more relevant with peer review.

Keywords: Research performance assessment; Peer review; Bibliometrics

(收稿日期: 2014-03-02)

■ 书 讯 ■

《中国高被引分析报告2012》

在您所关注的研究领域, 哪些论文最受同行关注? 哪些研究主题最为热门? 哪些学者最具学术影响力? 哪些期刊最获同行认可? 又有哪些机构占据领域研究的制高点? 上述问题请您参考近期出版的《中国高被引分析报告2012》。

该书将理、工、农、医、人文、社科等领域划分为51个学科, 综合分析了各个学科的研究热点与前沿、高影响力论文、高影响力作者、高影响力期刊和高影响力科研机构, 并以关联图谱的方式展现各种学术关系, 有助于科研人员及时发现并跟踪研究热点, 可为科研管理机构评估科研能力提供依据, 还有利于期刊编辑部监测本刊学术影响力, 是高等院校、科研院所及期刊编辑部等相关单位和人员参考工具书。

该书以“中国知识链接数据库”为依托, 数据样本覆盖我国6000余种期刊的论文及引文, 分学科揭示高影响力的学者、研究机构(大学、研究所、医院等)、地区(省/自治区/直辖市)、学术期刊、图书、外文期刊和会议录, 并采用共词分析、共被引分析和合著分析方法绘制出各学科的前沿主题分布以及作者、机构和期刊间关联的知识图谱。

《中国高被引分析报告2012》由中国科学技术信息研究所编著, 科学技术文献出版社出版, 全书86万字, 定价298.00元。