

基于“h指数族”的高校图书馆 科研影响力实证研究

□ 胡玉宁 / 滨州医学院图书馆 烟台 264003

丁明 / 滨州医学院卫生管理学院 烟台 264003

张士靖 / 华中科技大学医药信息系 武汉 430030

摘要: 为了揭示“h指数族”各项指数的相关性特点,对山东省内37所本科高校图书馆科研影响力进行综合实证研究,为图书馆工作者及相关机构进行馆际科研评价、交流和培训等提供决策依据。文章以《中国引文数据库》为数据源,用“h指数族”各项指数对评价机构在1996—2010年的论文数量和总被引频次进行统计分析。结论如下:鲁东大学图书馆h指数族取值特性异常显著;烟台大学图书馆等h指数族排名位次靠前;泰山医学院图书馆具备相当数量的潜在高被引论文,预期其科研影响力将会有显著提升。h指数族与论文数量和被引频次之间呈幂律关系,h指数族两两之间呈线性关系。图书馆h指数族取值随着图书馆论文数量和总被引频次的增长存在“规模效益递减”现象;图书馆应该密切关注能提升其g指数的潜在高被引论文,从而实现提升h指数和S指数、不断提高科研影响力的愿望。

关键词: h指数族, h指数, g指数, R指数, S指数, 高校图书馆, 科研影响力

DOI: 10.3772/j.issn.1673-2286.2011.09.010

1 研究背景

高校图书馆除了履行文献信息资源建设、文化传播和教育等职能外,更是一所高校乃至某个地区重要的科研机构。学术论文是科研成果的主要呈现形式,因此对学术论文进行统计和计量分析是进行科研评价的常用方式。2005年Hirsch提出了一种新的文献计量学指标——h指数, h指数体现和实践了“质”与“量”并重^[1],在讲质量的基础上求数量的新评价理念,已被扩展用于学术期刊影响力、科研集体学术绩效、成果预测和热点课题分析^[2]等不同领域。随着研究的深入,人们发现h指数对低被引和高被引论文均不敏感,存在局限和不足。因此,国内外许多学者在研究基础上进一步优化h指数,提出了多种新的衍生指数,如《科学观察》主编、国家科学图书馆金碧辉^[3]研究员提出的R指数、国外学者Egghe^[4]提出的g指数等。鉴于g指数和R指数等都由h指数衍生而来,本文将h指数及相关衍生指数统一命名为“h指数族”。文献回顾表明,利用h指数对图书馆进行科研能力测度的探讨已成为热点,赵星^[5]等对省级图书馆的科研能力进行了h指数相关探

讨;刘颖^[6]对省级图书馆和作者学术影响力进行评价研究,均取得了有意义的结论。笔者最近研究发现,省内关于高校图书馆范围的科研影响力研究较少,而立足于省内高校图书馆科研评价,研究层面客观实际,有利于找准本馆定位,进而有针对性地开展馆际交流和培训活动,具有深刻的理论价值和现实指导意义。本文利用“h指数族”各项指数对样本数据进行统计分析,揭示各项指数相关性特点,对山东省内37所本科高校图书馆科研影响力进行综合实证研究,为图书馆工作者及相关机构进行馆际科研评价、交流和培训等提供决策依据。

2 样本与方法

2.1 样本

选择CNKI中国引文数据库为数据源,利用中国大学网(www.cunet.com.cn)汇总的山东省内37所本科院校的图书馆为样本。数据采集时选择“引文检索”入口,引文文献类型选择“期刊类型引文”,检索项

选择“被引单位”，关键词为各高校图书馆的名称，检索时间为2011年1月。为了保证检索结果的全面性，在处理单位名称时，尽量考虑到高校合并、更名等因素。时段选择1996年至2010年，资源范围选择“全部期刊”，检索结果按照“被引频次”排序，对检索结果通过邮政编码等进行核对。对检索到的各图书馆发表论文的被引情况进行统计分析，计算得到各图书馆的各项指数，利用SPSS 15.0软件对各指数进行统计分析。

2.2 定义描述及计算方法

h指数定义可描述为：在统计时段内，某图书馆的所有馆员发表的总计N篇论文中，有h篇论文的被引频次至少为h次，而其余的N-h篇论文的被引频次均 $\leq h$ 次，则h即为该图书馆的h指数。

g指数定义可描述为：在统计时段内，某图书馆的所有馆员发表的N篇论文按被引频次降序排列，将每篇论文的序号取平方，每篇论文的被引次数逐一累加，当前g篇论文累计被引频次 $\geq$ 第g篇论文序号平方，前g+1篇论文累计被引频次 $\leq$ g+1篇论文序号平方时，则g即为该图书馆的g指数。

R指数定义可描述为：在统计时段内，确定某图书馆h指数后，h篇论文的被引频次总和的平方根即为该图书馆的R指数。

标准化(h,g,R)指数的计算方法：标准化(h,g,R) $_i = h_i / \sum h_i + g_i / \sum g_i + R_i / \sum R_i$ 。

3 结果分析与讨论

利用《中国引文数据库》检索后计算得出的统计结果包括了山东省37所本科高校图书馆在统计时段（1996-2010年）内的科研影响力各项指标数据，包括各图书馆发表论文数量、总被引频次、h指数、g指数、R指数和标准化(h,g,R)指数，结果如表1所示。下面就各项指数的分布、相关性等作统计分析和讨论。

3.1 高校图书馆科研影响力相关指标数据分析

h指数在衡量产出规模的同时兼顾论文质量。由表1可见，随着h指数的增大，评价机构的论文数量和

总被引频次大体呈现增长趋势，即h指数与论文数量和总被引频次之间存在正相关关系。但评价机构论文数量多，并不一定具有较高的h指数，如论文数量在145篇左右的评价对象有烟台大学图书馆、聊城大学图书馆、山东农业大学图书馆，三者论文产出规模相当，但由于论文被引频次的区别，h指数差别较大，分别为14、11、8；评价对象h指数为11的有山东工商学院图书馆、聊城大学图书馆、山东理工大学图书馆和临沂师范学院图书馆，论文数量分别为92、145、188、206，总被引频次分别为412、504、687、572，差别较大。因此，以论文数量作为计量指标评价图书馆科研影响力时，排名相同或接近的评价对象得到了不同的h指数排名，而且h指数与论文数量并不存在严格的正比关系，即论文数量大，h指数不一定就高，这正表明了h指数考量评价对象科研影响力的严格性，体现了科研产出规模和论文质量制衡兼顾的优势。

在获取每个评价对象h指数的过程中发现，排序 $\geq h$ 的论文被引频次只要不小于h，就不会影响到h值的变动，无论具体被引频次是多少；某些排序小于h的论文的被引频次却接近h，一旦这些论文的被引频次稍有增长，将对h值的变动具有较大影响。论文的被引频次直接反映了同行和业界科研工作者对该论文的肯定程度，高被引论文具有普遍认可的学术影响力，对科研影响力的贡献最大，在考量评价机构科研影响力时应该体现高被引论文的价值。排序接近h的论文存在向高被引论文转化的可能性，在体现评价对象科研影响力方面有潜在推动力，应当考虑到这部分论文对科研影响力的贡献。

g指数较好地兼顾了潜在高被引论文对评价对象科研影响力的贡献，R指数强化了高被引论文对评价对象科研影响力的贡献，而且由于计算方式不同，h值、g值为整数，R值可为非整数，h指数、g指数和R指数三者结合可增强各评价指数的区分度和灵敏度，在一定程度上解决了评价对象因h值和g值相等无法具体判别的难题。为了进一步提高评价效果的客观性，计算了各评价对象的标准化(h,g,R)指数，弱化三个指数的各自侧重，兼顾论文数量、高被引论文被引频次和潜在高被引论文被引频次对评价对象科研影响力的贡献。因此，将h指数、g指数、R指数和标准化(h,g,R)指数作为一族统计指数相结合，笔者称之为“h指数族”，用来考量评价对象的科研影响力，具有更强的可操作性、更佳的适用范围和更客观的评价效果。

表1 山东省37所本科高校图书馆科研影响力相关指标数据统计表

图书馆名称	论文数量	总被引频次	h指数	g指数	R指数	标准化(h,g,R)
鲁东大学图书馆	361	1577	18	24	22.38	0.1840
烟台大学图书馆	146	692	14	21	19.26	0.1539
山东大学图书馆	253	811	13	18	16.79	0.1363
曲阜师范大学图书馆	258	923	12	20	17.78	0.1403
山东工商学院图书馆	92	412	11	14	13.53	0.1104
聊城大学图书馆	145	504	11	16	15.00	0.1193
山东理工大学图书馆	188	687	11	16	14.32	0.1175
临沂师范学院图书馆	206	572	11	12	11.71	0.1005
青岛理工大学图书馆	66	296	10	14	13.23	0.1060
青岛大学图书馆	111	359	10	15	13.56	0.1094
济南大学图书馆	136	459	10	14	13.27	0.1062
山东师范大学图书馆	165	516	10	14	13.26	0.1061
中国海洋大学图书馆	43	227	9	12	11.57	0.0930
山东财政学院图书馆	58	269	9	15	14.39	0.1081
青岛农业大学图书馆	113	304	9	10	10.20	0.0844
山东科技大学图书馆	88	310	9	14	13.04	0.1020
潍坊学院图书馆	95	311	9	12	11.71	0.0934
山东经济学院图书馆	58	249	8	12	11.23	0.0886
泰山医学院图书馆	71	334	8	16	14.94	0.1085
山东农业大学图书馆	146	336	8	10	9.22	0.0782
德州学院图书馆	94	249	7	9	8.54	0.0703
山东轻工业学院图书馆	40	119	6	9	8.06	0.0655
青岛科技大学图书馆	78	205	6	8	7.87	0.0624
山东建筑大学图书馆	46	105	6	8	6.86	0.0598
泰山学院图书馆	44	89	5	7	6.71	0.0533
滨州医学院图书馆	52	148	5	9	8.66	0.0635
菏泽学院图书馆	60	122	5	5	5.29	0.0445
枣庄学院图书馆	17	60	4	6	6.33	0.0462
潍坊医学院图书馆	29	69	4	5	5.29	0.0409
滨州学院图书馆	46	86	4	4	5.39	0.0387
山东体育学院图书馆	14	42	3	6	5.91	0.0415
山东警察学院图书馆	13	50	3	7	6.56	0.0457
济宁医学院图书馆	39	47	3	3	3.00	0.0262
山东艺术学院图书馆	8	10	2	3	2.68	0.0218
山东中医药大学图书馆	8	7	2	2	2.24	0.0181
山东交通学院图书馆	5	6	2	2	2.00	0.0175
山东工艺美术学院图书馆	6	11	2	3	2.65	0.0217

3.2 “h指数族”的相关统计学分析

为了更直观和科学地考量评价对象的科研影响力,对“h指数族”相关数据进行统计学量化分析,并结合评价对象进行实证判别,为实际工作和相关决策提供理论依据和参考。

为了直观显示h指数族各项指数之间的相关关系,利用SPSS 15.0绘制了矩阵散点图,具体分布如图1所示。通过散点图分析可以总结出以下几点规律:

① 各矩阵散点分布共同点:两两变量之间呈现出不同程度的正相关趋势;

② 有突出表现的数据点:每个子矩阵右上象限都有一个远离群体的数据点。经过表1数据的实证分析,该数据点代表鲁东大学图书馆,表明该图书馆的各项指标虽然基本符合整体规律,但各指标取值与数据集合主题有较大区别,这正体现了鲁东大学图书馆在科研论文产出规模和论文质量两方面的突出优势。

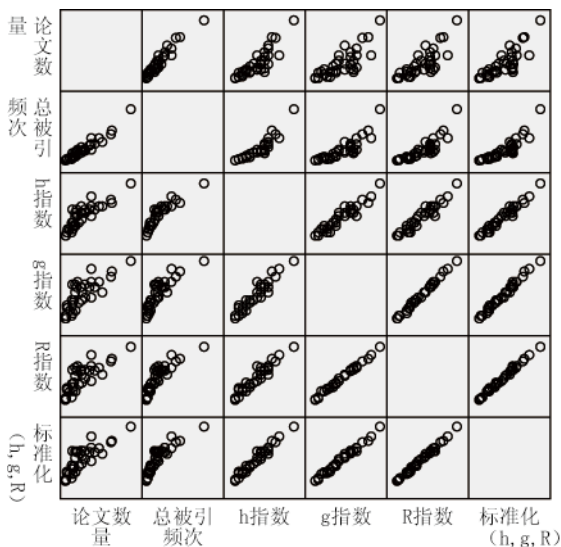


图1 h指数族的矩阵散点图

③ 变量间相关关系的表现形式不同:论文数量与总被引频次和h指数族之间呈现出幂函数曲线趋势,h指数族两两变量之间呈现出近似直线的相关趋势。

④ 散点图分布密度和形状:从散点矩阵中不难发现,出现了由较宽基底和指向坐标的尖端构成的类似“扇形”区域,区域内有浓密的核心。这说明,图书馆依靠增加论文数量来提升科研影响力水平是有一定效果的,而且在论文数量增长的初期,效果明显;随着论文数量的增加,h指数族数值的增长呈现出“边际效益递减”的趋势。因此,图书馆欲通过提高论文数量来提升科研影响力的愿望可行,但存在“长期效应”,即图书馆需要足够长的时间来体现“边际效益”。比如山东农业大学图书馆在统计时段内论文数量为146篇,但h指数族得分都远远低于论文数量同为146篇的烟台大学图书馆。浓密的核心部分代表了山东省高校图书馆中存在科研影响力相当的一部分图书馆,这部分图书馆h指数族取值接近,如山东工商学院图书馆、山东理工大学图书馆、青岛大学图书馆、济南大学图书馆等,这部分图书馆可能具备规模和水平相当的科研资源。分散稀疏的区域代表了科研影响力较强的部分图书馆,如以鲁东大学图书馆、烟台大学图书馆等为代表的强势科研团体。

为了进一步分析各变量之间的相关关系,用Q-Q概率图验证各变量样本数据的分布。检验结果中h指数族各项指数散点都分布在45度对角线的附近,可以认为h指数族的样本数据总体上服从正态分布,涉及h指数族的样本数据服从Pearson相关系数检验的理论分布。论文数量和总被引频次的正态Q-Q图中的散点分布表明两个变量的样本数据总体上不满足正态分布假设,对涉及这两个变量的相关分析可选择Spearman等级相关系数进行检验。利用SPSS 15.0软件进行相关分析,得出六个变量的相关性矩阵(如表2所示)。

表2 各指数相关性矩阵

	论文数量	总被引频次	h指数	g指数	R指数	标准化(h,g,R)
论文数量	1.000	0.965**	0.910**	0.831**	0.837**	0.855**
总被引频次		1.000	0.970**	0.931**	0.934**	0.949**
h指数			1.000	0.957**	0.964**	0.983**
g指数				1.000	0.997**	0.994**
R指数					1.000	0.996**
标准化(h,g,R)						1.000

注:此相关性矩阵为阴影部分是Spearman相关系数,其余为Pearson相关系数。**表示在0.01水平显著。

由表2可知,图书馆h指数族与总被引频次的相关性强于与论文数量的相关性,说明图书馆h指数族更侧重于表现图书馆科研产出的质量。利用h指数族进行评价对象科研影响力的考量是科学、合理的。在h指数族的四个指数中,h指数与论文数量和总被引频次的相关性最强;g指数和R指数与论文数量和被引频次的相关性相对较弱;而标准化(h,g,R)指数介于二者之间,因此标准化(h,g,R)在考量评价对象科研影响力时,进一步修正了h指数、g指数和R指数的不足,评价标准更加公正合理。

通过回归分析进行模型的假设和检验,发现图书馆各项指数[以N表示论文数量、C表示总被引频次、S代表标准化(h,g,R)指数]之间的相关关系可通过以下方程式进行数学描述:

$$h=0.773N^{0.521} (R^2=0.876, \text{Sig}<<0.01)$$

$$h=0.717C^{0.430} (R^2=0.954, \text{Sig}<<0.01) \quad (1)$$

$$g=1.088N^{0.518} (R^2=0.735, \text{Sig}<<0.01)$$

$$g=0.844C^{0.453} (R^2=0.900, \text{Sig}<<0.01) \quad (2)$$

$$R=1.121N^{0.502} (R^2=0.762, \text{Sig}<<0.01)$$

$$R=0.927C^{0.436} (R^2=0.924, \text{Sig}<<0.01) \quad (3)$$

$$S=0.009N^{0.511} (R^2=0.804, \text{Sig}<<0.01)$$

$$S=0.007C^{0.438} (R^2=0.944, \text{Sig}<<0.01) \quad (4)$$

$$g=0.119+1.400h (R^2=0.917, \text{Sig}<<0.01) \quad (5)$$

$$R=0.520+1.273h (R^2=0.929, \text{Sig}<<0.01) \quad (6)$$

$$S=0.002+0.010h (R^2=0.966, \text{Sig}<<0.01) \quad (7)$$

$$S=0.003+0.007g (R^2=0.988, \text{Sig}<<0.01) \quad (8)$$

$$S=-0.001+0.008R, R \neq 0 (R^2=0.992, \text{Sig}<<0.01) \quad (9)$$

经过不同模型拟合和回归分析发现:h指数族与论文数量和被引频次之间的关系与幂函数有较高的拟合度;h指数族两两之间的关系呈线性关系。通过模型的F检验结果来看,Sig值都远远小于0.01,说明模型成立的统计学意义都非常显著。从R²统计量看,所选取的模型拟合效果较好。(1)、(2)、(3)、(4)方程式中的幂指数大于0小于1,说明图书馆h指数族取值随着图书馆论文数量和总被引频次的增长而增长,但同时存在“规模效益递减”现象,这与前文中根据矩阵散点图的分析是一致的,说明随着图书馆论文数量和总被引频次的增长,h指数族取值上升难度愈发加大。这就意味着总被引频次的增长若主要来自机构新近发表的论文所产生的被引次数的累积,对该机构h指数的上升效果并不显著。(5)、(6)、(7)、(8)、(9)方程式显示h指数族各项指数之间呈线

性相关,表明指数之间关系密切而且稳定。从长远来看,g指数增长带给h指数增长的效果较为明显,h指数增长带给S指数增长的效果较为明显。这就说明,评价机构应该密切关注能提升其g指数的潜在高被引论文,提升这部分论文的影响力,从而实现提升h指数和S指数、不断提高科研影响力的愿望。

4 结语

本文利用中国引文数据库为数据源,利用h指数族各项指标对山东省内37所高校图书馆的科研影响力进行了综合分析和评价。研究表明:

鲁东大学图书馆h指数族取值特性异常显著,排名第一,表明现阶段该图书馆在科研论文产出规模和论文质量两方面具备突出优势,在评价对象群体中科研影响力最强。烟台大学图书馆、曲阜师范大学图书馆、山东大学图书馆、聊城大学图书馆等h指数族排名位次靠前,表明各图书馆具备较强的科研影响力。这为山东省内各高校图书馆开展馆际交流和学术研讨活动提供了决策支持,笔者认为与以上图书馆开展对口交流和培训活动预期收效较为显著。

h指数族各项指数各有侧重,h指数同时结合了评价机构的论文数量和总被引频次;g指数较好地兼顾了潜在高被引论文对评价对象科研影响力的贡献;R指数强化了高被引论文对评价对象科研影响力的贡献;S指数[即标准化(h,g,R)指数]弱化三个指数的各自侧重,兼顾论文数量、高被引论文被引频次和潜在高被引论文被引频次对评价对象科研影响力的贡献。泰山医学院图书馆按照h指数排序位次靠后,但具有较高的g指数、R指数和S指数排名,说明该图书馆具备相当数量的潜在高被引论文,这部分论文的学术影响力是相当可观的,预期在不远的统计时段内,泰山医学院图书馆h指数将会有显著增长,科研影响力会有显著提升。综合来看,用h指数族来考量评价对象的科研影响力,具有更强的可操作性、更佳的适用范围和更客观的评价效果。

h指数族与论文数量和被引频次之间的关系密切,与幂函数有较高的拟合度,其中幂指数大于0小于1,说明图书馆h指数族取值随着图书馆论文数量和总被引频次的增长而增长,但同时存在“规模效益递减”现象。因此要提升图书馆的科研影响力,更重要的是要开拓创新,进行更有创新性、更有科研价值和影响力

的学术研究。h指数族两两之间呈线性关系,评价机构应该密切关注能提升其g指数的潜在高被引论文,提升这部分论文的影响力,从而实现提升h指数和S指数、不断提高科研影响力的愿望。

科研影响力评价所用到的各项指标的属性都是

在不断发展与完善的,如各种指数计算权重选择、数据统计分析方法处理等。随着科学的发展和技术的进步,h指数族的队伍会不断壮大,为更加科学、公平、系统和完整地进行科研影响力评价作出更大贡献。

参考文献

- [1] HIRSCH J E. An index to quantify an individual's scientific output [C]// Proceedings of the National Academy of Sciences of the United States of America. USA:2005, 102(46):16569-16572.
- [2] 赵基明,邱均平,等.一种新的科学计量指标——h指数及其应用述评[J].中国科学基金,2008(1):23-32.
- [3] 金碧辉.科学家为自己设计了一项评价指标:h指数[J].科学观察,2006,1(1):8-9.
- [4] EGGHE L. Theory and Practice of the g-index [J]. Scientometrics,2006(1):131-152.
- [5] 赵星,黄娟,彭晓东.我国省级图书馆科研产出的h指数测度[J].中国图书馆学报,2009(11):103-107.
- [6] 刘颖.省级图书馆学术影响力评价研究[J].图书情报工作,2010(3):46-50,71.

作者简介

胡玉宁(1984-),女,硕士,助理馆员,主要研究领域:信息评价与信息服务等。E-mail: yuning1688@yahoo.com.cn

丁明(1980-),男,硕士,讲师,主要研究领域:信息检索与信息分析等。

张士靖(1958-),女,教授,硕士生导师,主要研究领域:医学信息检索与评价、情报检索语言等。

Empirical Study on the Research Influence of University Library Based on the Evaluation of "H-Index Family"

Hu Yuning / Library of Binzhou Medical University, Yantai, 264003

Ding Ming / Department of Health Management, Binzhou Medical University, Yantai, 264003

Zhang Shijing / Department of Medical Informatics, Huazhong University of Science & Technology, Wuhan, 430030

Abstract: Objective: To reveal the correlation characteristics of h-index family, and make empirical study on the research influence of 37 university libraries in Shandong province, and provide basis for librarians and relevant institutions when they make decisions about interlibrary research evaluation, communication and training. Methods: Using Chinese Citations Database as data source, the paper made statistical analysis and the total number of papers and the times cited with h-index family in statistical period from 1996 to 2010. Results: The results indicate that library of Ludong University Library had a significant value and some university libraries perform well on h-index family such as Yantai University Library. Library of Taishan Medical University is expected to improve research influence significantly with a considerable number of potentially highly cited papers. The results also show that there is a power-law relation between h-index and sum of the times cited, the same as the total number of papers, and there is a linear relationship between any two. Conclusion: The increment of the total times cited and the total number of papers will contribute as extensive descending phenomenon to that of h-index family. Libraries should pay more attention to those potentially highly cited papers which will contribute to its g-index in order to enhance h-index and S-index to achieve the desire of improving research influence.

Keywords: H-index family, H-index, G-index, R-index, S-index, University library, Research influence

(收稿日期: 2011-06-01)