

质量视角下中国区域人文社科 研究水平评价研究

陈玉卿¹ 俞立平²

(1. 贵州财经大学 贵州绿色发展战略高端智库 贵阳 550025;

2. 浙江工商大学 统计与数学学院 杭州 310018)

摘要: [研究目的]在高质量发展背景下,开展地区人文社科水平评价具有重要意义。[研究方法]在理论分析的基础上建立地区人文社科水平评价指标体系,并提出区域h指数指标,采用sigmoid函数改进了传统的指标标准化方法,并采用主流的线性加权法进行评价。[研究结论]在研究质量视角下开展区域人文社科评价非常独特,宏观评价方法主要依靠指标体系,一些指标不能用于微观但可以用在宏观,宏观评价应注重数量和质量;我国地区人文社科总体发展态势良好;传统线性加权汇总评价中标准化方法有待改进;基于sigmoid函数标准化的线性加权是一种优秀的评价方法。

关键词: 人文社科;线性加权;评价方法;评价体系;评价指标

中图分类号: G302

文献标识码: A

文章编号: 1002-1965(2021)11-0147-08

引用格式: 陈玉卿,俞立平.质量视角下中国区域人文社科研究水平评价研究[J].情报杂志,2021,40(11):147-154.

Research on the Evaluation of Regional Humanities and Social Sciences Level in China from the Perspective of Research Quality

Chen Yuqing¹ Yu Liping²

(1. Guizhou Green Development Strategy High End Think Tank, Guizhou University of Finance and Economics,

Guiyang 550025;

2. School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018)

Abstract: [Research purpose] In the context of high-quality development, it is of great significance to carry out regional humanities and social sciences level evaluation. [Research method] Based on the theoretical analysis, this paper establishes a regional humanities and social sciences level evaluation index system, and proposes the regional h index. The sigmoid function is used to improve the traditional index standardization method, and the mainstream linear weighting method is used for evaluation. [Research conclusion] From the perspective of research quality to study the regional humanities and social sciences evaluation is very unique; the macro evaluation method mainly relies on the indicator system; some indicators cannot be used in the micro evaluation but can be used in the macro evaluation; the macro evaluation should focus on quantity and quality; the standardization method in traditional linear weighted summary evaluation needs to be improved; linear weighting based on sigmoid function standardization is an excellent evaluation method.

Key words: humanities and social sciences; evaluation method of linear weighting; evaluation system; evaluation indicator

0 引言

人文社会科学是人类的精神家园,对人类社会的

发展具有深远的影响。人文科学主要研究人类社会的本质和发展规律,社会科学研究社会整体物质客体及其内在联系,以及人的主体意识及其内在联系,广义的人文社会科学一般也称为哲学社会科学。一个没有发

达的自然科学的国家不可能走在世界前列,一个没有繁荣的哲学社会科学的国家也不可能走在世界前列,这一点已经取得人们的普遍共识。向世界展示中国特色、中国风格、中国气派,是中国人文社会科学发展的历史使命,构建中国特色哲学社会科学体系已上升为国家战略。

提高人文社会科学研究质量是人文社科发展的重要目标。《高等学校哲学社会科学繁荣计划(2011—2020年)》强调要完善以创新和质量为导向的科研评价制度。我国人文社会科学论文质量有待提高,根据科睿唯安数据,2018年全球发表SSCI论文33.20万篇,其中高水平论文2177篇,中国SSCI论文2.50万篇,其中高水平论文241篇。中国SSCI论文总量占7.54%,高水平SSCI论文占11.07%。据中国知网的统计数据,2018年国内期刊共发表大陆人文社科论文113.58万篇,是全球SSCI论文的3.42倍,其中核心期刊论文13.73万篇。中国人文社科论文总量太多,质量不高,这是制约我国人文社科发展的重要问题。

地区人文社科研究水平对区域经济社会发展具有重要意义。中国地大物博,人口众多,不同地区形成了不同的文化和特色,这一切给区域经济社会发展也打上了深深的烙印。加强人文社科研究,对地方政府的决策会产生重要影响,进而会影响到经济社会发展,从而彰显区域经济社会特色,弘扬传统文化,展现传统特色,促进经济高质量发展。

在高质量发展背景下,开展区域人文社科研究评价具有重要意义。首先,相关领域的研究总体较少,现有研究只涉及人文社科研究效率、人文社科全要素生产率等,对区域人文社科研究水平缺乏研究。人文社科学术期刊,因学科门类庞杂、研究范式多元,稿件的质量判断始终是一个让人头疼的问题^[1]。其次,研究区域人文社科研究的基本原则和理论架构,有助于丰富科学计量学与文献计量学理论。再次,在实践角度,也有利于地方政府相关部门明确差距,从而加大投入,促进人文社科健康快速发展。

1 文献综述

关于人文社科的研究质量的界定,Talib^[2]提出科研评估必须权衡数量与质量。Mazlish等^[3]指出人文社会科学的质量包括学术界内部质量和社会外部质量,内部质量主要指同行认可的学术价值,外部质量是科研成果的外部贡献。Moed^[4]研究发现英国的科技评价工作经历数量—质量—数量三个阶段。

关于人文社科研究质量评价的相关研究,蒋玲等^[5]提出人文社会科学成果具有主观性和广泛性,标新立异和百家争鸣才能推动人文社会学向前发展,科

研成果的独创性是价值基础,大数据可以为同行评议提供一定的补充。高自龙等^[6]认为对于人文社科成果,必须坚持质量为本原则,不断完善评估指标体系的设置,改进实施策略,寻求较为科学可行的评估方式和手段。李贞等^[7]强调协同创新和质量标准是科研团队水平的重要标志,但具体质量评价指标选取尚有待研究。潘启亮^[8]认为人文社科评价包括同行评议和科学计量,同行评议主要是质量的评价,科学计量主要是数量的评价。

关于人文社科评价对象,视角也是多元的。第一是从学科角度,高自龙等^[9]研究发现,在人文地理学、社会学、管理学、经济学等学科中,某些指标合作论文水平优于独著论文,在哲学、马克思主义理论、外国语言文学以及图情档、法学学科中,合作论文的个别指标水平低于独著论文。第二是从大学评价角度,蒋玉梅^[10]发现国外大学对人文社科研究评价包括研究成果的数量、质量、影响,以及产生的社会效益方面的途径。第三是从学术期刊角度,Prathap^[11]在数量和质量的基础上,提出引文分布一致性系数,并构建出z指数进行评价。刘仲翔^[12]认为2018年是中国社科学术期刊发展的重要转折年份,学术期刊进入了高质量发展的关键阶段。第四是从博士论文角度,Ives等^[13]研究发现,项目参与对人文社科博士论文的综合质量、前沿涉猎、创新性均有积极正向影响,课堂讨论、跨学科学习和导师交流对博士论文的综合质量和成果创新性均没有影响。李霞等^[14]将人文社科博士学位论文质量的要素分为形式要素和实质要素,后者就是指论文质量,并对博士论文质量进行评价。

关于宏观区域视角的人文社科评价,Kocher等^[15]对国际经济合作组织21个国家经济学科研究效率进行了比较分析。姜彤彤^[16]采用基于DEA的Malmquist分析方法,对我国大陆省际高校人文社科研究的全要素生产率进行评价和分解研究。同时还基于DEA分析方法,对我国省际高校人文社科科研效率及其区域差异情况进行分析^[17]。

从现有的研究看,关于人文社科研究成果的评价研究,已经有不少研究涉猎,对其研究质量的认识和界定已经取得共识。关于人文社科质量评价方法总体上是零碎的,相关研究并不充分,涉及同行评议、文献计量、大数据、指标体系方法等。学术界普遍认识到对人文社科质量的评价标准要强调多元化和创新。单纯对人文社科评价的研究,研究视角包括学科评价、大学评价、学术期刊评价、博士论文评价等,区域全要素生产率与效率评价等,研究视角有待进一步拓展。总体上在以下方面有待进一步深入研究:

第一,关于质量视角下区域人文社科研究的评价

理论,现有研究很少有涉及,需要进一步开展研究。相关内容包括:区域人文社科评价的基本原则、数量与质量的关系、宏观评价与微观评价的关系等。

第二,关于质量视角下区域人文社科研究的评价方法,也有待进一步探索。相关研究包括:同行评议的地位、文献计量学的运用、多属性评价方法的选择等等。其中,基于传统线性加权评价方法的改进研究也需要深入。

2 质量视角下区域人文社科评价理论与方法

2.1 评价规模与评价方法的关系 评价规模与评价方法的关系如图 1 所示。学术评价包括宏观评价、中观评价与微观评价,宏观评价方法主要采用多属性评价方法,中观评价采用多属性评价与同行评议,以多属性评价为主,微观评价主要采用同行评议方法。

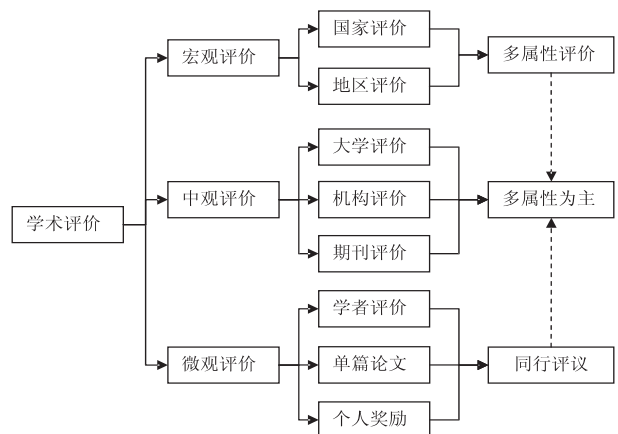


图 1 评价规模与评价方法的关系

宏观评价包括国家或地区层面的人文社科评价,其评价视角可以采用单个指标,也可以采用多个指标。单个指标就是文献计量指标或科技统计指标,如论文数量、研发经费投入总额等等,可以直接进行比较。如果采用多个指标,一般要采用某种多属性评价方法将其转化为一个汇总的评价结果,比如国家创新指数评价就是类似的评价。多属性评价方法目前有几十种,如线性加权汇总、主成分分析、灰色关联、粗糙集、TOPSIS、CRITIC 等等,目前已广泛应用在学术评价中。

中观评价包括大学评价、学科评价、团队评价、院系评价、学术期刊评价等等,由于评价对象范围相对较大,评价成果众多,所以一般采用多属性评价方法为主,同行评议作为必要的补充,比如教育部第四轮学科评价就采用这种方法。在同行评议中,主要针对高水平成果以及那些难以用具体指标展示的成果进行评价,比如社会影响力等。

微观评价主要包括学者评价、单篇论文评价、个人奖励评价等等。在学术期刊中采用的双向匿名审稿采用的就是同行评议的方法,职称评审中的论文送审类

似。但是一些微观评价有时接近中观评价,比如对于成果较多的学者进行评价。现在推行的代表作制度就是强调用同行评议代替评价指标,对论文本身的创新进行评价。

2.2 微观评价指标在宏观评价中的异化 所谓微观评价指标在宏观评价中的异化,就是人们公认的一些存在问题的微观评价指标,在宏观评价中却总体上是适用的,这是一个十分有意思的问题。

比如单篇论文的被引次数,在微观评价中,我们永远不能采用被引次数来判定一篇论文的质量,对单篇论文质量的认定,唯一的方法就是采用同行评议,这一点几乎已经得到公认。但是在宏观评价中,一个国家或地区平均每篇论文的被引次数一定程度却能说明论文质量。这是因为对于单篇论文而言,永远可以将其作为统计学上的异常点,但对于一个国家或地区的论文总体而言,从统计学意义上,这些文献计量指标是能说明一定问题的。

还有一个典型的指标是核心期刊论文,同时涉及到以刊评文问题。在微观视角,学术界对简单用核心期刊论文来判断论文水平高低的现象诟病很深。曹骏等^[18]认为核心期刊不是优秀论文专刊,不能保证刊出的每篇文章都是精品。周兴旺^[19]指出用刊物影响力的高低来推断论文影响力的高低,这在逻辑上是不对的,期刊影响力并不等同于论文影响力。但是如果换一个视角,从地区层面,每个省市的核心期刊论文数量一定程度上就可以代表其人文社科的水平,这其实并没有太大的争议。

所以对于微观层面的文献计量指标,往往可以适用于中观和宏观层面,对于一些中观层面的文献计量指标,往往适用于宏观层面,但是如果这个次序颠倒则不一定可以,宏观层面的指标不一定能用于中观或微观,中观层面的指标也不一定能用于微观。在学术评价中,微观、中观和宏观并不是简单的加总关系。当然具体指标要看其内涵并进行进一步分析,这个原则并不是简单适用于任何指标。

2.3 省际人文社科评价中数量与质量的关系 人文社科研究的数量与质量问题是一对矛盾,并且与评价规模紧密相关。评价数量质量与评价规模的关系如图 2 所示。

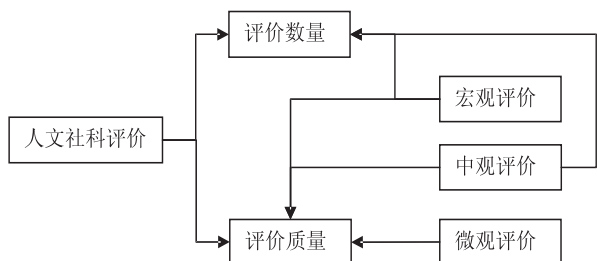


图 2 评价数量质量与评价规模的关系

第一,在微观评价层面,一定要强调研究质量。尽管人文社科研究具有自己的特点,与自然科学研究存在较大差异,但是在学术研究质量上,还是要坚持质量至上的原则。2018 年科技部、教育部等联合发文,要求开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动。随后不久,两办印发《关于深化项目评审、人才评价、机构评估改革的意见》,推行代表作评价制度,注重标志性成果的质量水平、贡献高低、影响力大小。2020 年教育部、科技部研究制定了《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》,破除论文“SCI 至上”。这些举措的根源就是为了解决研究质量问题。固然不同学者的研究水平存在差异,但只要树立质量至上的原则,就会鼓励广大学者进行高水平的研究,而不是仅仅追求论文数量。

第二,在中观和宏观层面,不仅要追求数量,也要追求质量。质量代表了人文社科科研成果的高度,而数量代表了人文社科科研成果的广度,两者不可偏废。如果一个机构一共只有 10 篇公认的高质量成果,而另外一个机构拥有 50 篇公认的高质量成果。两个机构孰强孰弱一目了然,所以对于机构评价,采用代表作制度是要慎重的。

第三,微观质量要求是宏观兼顾数量与质量评价的基础。只有在微观学者层面强调科研质量,那么中观和宏观层面结合数量和质量的评才有意义,如果微观层面科研质量得不到保证,急功近利行为并不是个别现象,那么中观和宏观评价就会失去质量至上的基础,从而导致评价异化。科技部、教育部“反四唯”,推进论文代表作制度的根本动因,就是为了保证微观层面的质量的基础。

2.4 地区人文社科评价体系 地区人文社科评价体系如图 3 所示,一级指标包括 3 个,分别是科研投入、成果数量、成果质量,成果数量和成果质量均为产出指标,因此地区人文社科评价也反映了科研成果的投入产出关系。

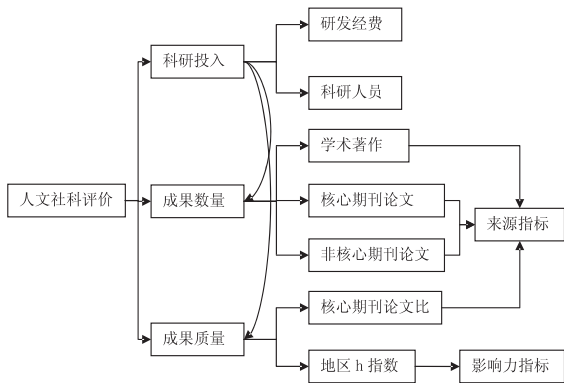


图 3 地区人文社科评价体系

第一,科研投入包括两个指标,分别是研发经费和研发人员。从可见度看,研发经费和研发人员是人文

社科研究的两大重要投入,当然其他因素还包括国家或地方政府的相关政策、外部环境等等,由于相关数据难以获得,因此没有选取。对于人文社科而言,科研人员的重要性相对重要,这和人文社科研究的特点有关。相对自然科学而言,人文社科的大多数学科对研发经费的依赖程度相对较弱,当然一般情况下,人文社科科研经费的资助力度也远远小于自然科学。

第二,成果数量指标包括学术著作、核心期刊论文、非核心期刊论文。之所以将学术论文分为核心期刊论文和非核心期刊论文,主要目的也是为了加强质量视角下的评价。对于人文社科而言,学术著作的研究更加全面系统,近两年国家新闻出版广电总局加大了学术专著刊号的管理,学术专著质量得到大幅提高。所以学术著作和核心期刊论文数量既是数量指标,也是质量指标,代表着较高水平的研究成果,而非核心期刊论文总体上属于数量指标。

第三,成果质量指标包括核心期刊论文比、地区 h 指数两个指标。核心期刊论文比就是地区核心期刊论文数量占论文总量的比重。地区 h 指数是仿照 h 指数提出的一个新的指标,其含义是:某地区过去两年发表的论文在统计当年至少有 h 篇论文被引用了 h 次。地区论文 h 指数某种程度上是一种更高水平论文的影响指标,而核心期刊论文比是一种高水平论文的指标,两者表示的质量等级有所区别。

如果将地区人文社科的评价指标分为来源指标与影响力指标,那么学术著作、核心期刊论文数量、非核心期刊论文数量、核心期刊论文比均属于来源指标,只有地区论文 h 指数为影响力指标。

总体上,本文建立的评价体系兼顾了人文社科研究的投入与产出、数量与质量、来源指标与影响力指标、绝对指标和相对指标,是一个相对完善系统的评价体系。

2.5 评价方法与指标权重

2.5.1 指标权重的确定 评价指标的权重采用专家会议法确定,本文之所以不用客观赋权法的原因,是没有哪种客观赋权法能够保证评价质量的权重最大。邀请相关领域的 10 名专家,采用背对背远程打分形式,每次将平均值再反馈给各位专家,如有需要请他们进行必要的调整,如此经过专家三轮讨论,权重设定如表 1 所示。关于投入产出的权重,投入指标的权重为 0.40,产出指标的权重为 0.60。产出指标中,成果数量指标为 0.25,成果质量指标为 0.35。实际上,由于成果数量指标中,学术著作和核心期刊论文数量均是高质量成果,所以实际上质量指标的权重更高,符合质量视角下的评价原则。

在科研投入指标中,研发经费和研发人员采用等

权重处理。在成果数量指标中,学术著作和核心期刊论文数量的权重均为 0.40,非核心期刊论文的权重为 0.20。成果质量指标中,核心期刊论文比的权重为 0.40,地区 h 指数为 0.60,因为后者是高水平的研究成果。

表 1 评价指标与权重

一级指标	二级指标	权重
科研投入 0.40	研发经费	0.50
	研发人员数	0.50
成果数量 0.25	学术著作数	0.40
	核心期刊论文数	0.40
	非核心期刊论文数	0.20
成果质量 0.35	核心期刊论文比	0.40
	地区 h 指数	0.60

2.5.2 评价方法的选择 本文采用线性加权方法汇总,这种评价方法是一种线性评价方法,能够更好地尊重专家权重。在目前主流的学术评价中,比如英国泰晤士报大学排名、软科大学排名、南京大学 CSSCI 期刊评价、北大核心期刊评价等应用中,均采用这种方法。

在线性加权中,指标标准化一般采用极大值法,标准化公式为:

$$Y = \frac{X}{\max(X)} \times 100 \tag{1}$$

其中, X 为指标原始值, Y 为标准化后的指标值。

在线性加权中,有一个关键问题被忽略了,就是由于评价指标标准化后指标均值不等,这会导致实际权重偏离专家设定的权重,俞立平^[20]将这种现象称为自然权重问题。比如小学生考试,语文的平均成绩为 72,数学的平均成绩为 87,在计算总分时,语文的实际权重就比较低,这就是自然权重导致的。所以采用线性加权汇总,最重要的是通过某种标准化方法,使得不同指标的均值接近相等。为了解决这个问题,可以采用 sigmoid 函数进行标准化。

采用 sigmoid 函数进行标准化,第一步是计算 z 值:

$$z = \frac{X - \mu}{\sigma} \tag{2}$$

其中, X 为指标原始值, μ 为指标均值, σ 为指标标准差。

第二步是根据 z 值计算标准化后的指标 Y ,即:

$$Y = \frac{1}{1 + e^{-z}} \tag{3}$$

Sigmoid 函数图像如图 4 所示,可以明显看出,sigmoid 函数和成长曲线非常相似,这符合人文社科发展的规律和成长特征。标准化后其极大值为 1,极小值为 0,标准化值具有绝对打分功能,比较适合用于评

价。通常情况下,标准化后的均值相差不大,可以有效解决自然权重问题。同时可以通过对评价结果进一步标准化处理,分析各个省份人文社科发展阶段。在 sigmoid 函数中,拐点位置为(0,0.5),进入成长期的坐标为(-1.317,0.211),进入成熟期的坐标是(1.317,0.789)。

需要说明的是,本文采用截面数据进行评价,难以体现成长性,但是对于每个地区的人文社科发展而言,从时间序列数据角度,是具有成长曲线特征的。此外,即使对于截面数据而言,本着发达地区今天就是欠发达地区未来的思想,一定程度上也具有成长特征。

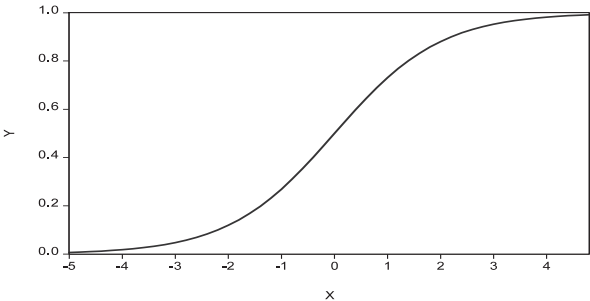


图 4 sigmoid 函数

3 数据来源与实证结果

3.1 数据来源 研发经费、研发人员、学术著作:这 3 个指标采用教育部人文社科网站公布的省际高校数据。人文社科研究的主要阵地是广大高校,另外还有各地哲学社科院所、各级政府的研究机构等部门,但相关数据难以获取,加上高校研发投入所占比重较大,因此采用高校数据作为人文社科评价的替代变量,本文各指标数据采用的均为 2018 年的数据。

其他指标:其他所有指标均来自中国知网的引文数据库,学科包括所有的人文社科学科,即哲学与人文科学、社会科学 I 辑、社会科学 II 辑、经济与管理科学 4 大类,并不包括基础科学、工程科技 I 辑、工程科技 II 辑、农业科技、医药卫生科技、信息科技。在自然科学期刊中发表的少量人文社科论文,或与自然科学存在学科交叉的人文社科论文,并不在统计之列,但这些论文总体数量不多,并不影响宏观统计结果。

机构 h 指数、论文总数、核心期刊论文数为直接查询数据,在此基础上计算出非核心期刊论文数和核心期刊论文比。对于来源指标,所有论文数据均为 2018 年数据,对于机构 h 指数的计算,考虑到时效性,参照影响因子计算的时间跨度,载文时间为 2016-2017 年,被引数据为 2018 年。

3.2 评价结果分析 评价结果如表 2 所示。根据评价结果,通过 sigmoid 函数计算 z 值,以判定地区人文社科的发展阶段,当 z 值小于 -1.317 时,处于起步期,当 z 值在[-1.317,1.317]时,处于成长期,当 z 值

大于 1.317 时,处于成熟期。

表 2 地区人文社科评价结果

地区	评价总分	排序	评价结果 z 值	所处阶段
北京市	90.79	1	2.431	成熟期
广东省	76.93	2	1.630	成熟期
上海市	75.63	3	1.555	成熟期
江苏省	73.54	4	1.434	成熟期
湖北省	69.36	5	1.193	成长期
浙江省	67.86	6	1.107	成长期
山东省	61.86	7	0.760	成长期
四川省	59.97	8	0.651	成长期
河南省	57.24	9	0.493	成长期
重庆市	54.30	10	0.324	成长期
湖南省	54.16	11	0.315	成长期
陕西省	53.81	12	0.295	成长期
辽宁省	51.67	13	0.172	成长期
天津市	48.86	14	0.010	成长期
福建省	48.22	15	-0.028	成长期
吉林省	47.68	16	-0.059	成长期
河北省	46.77	17	-0.111	成长期
安徽省	45.28	18	-0.197	成长期
江西省	42.77	19	-0.342	成长期
广西壮族自治区	39.12	20	-0.553	成长期
云南省	38.88	21	-0.567	成长期
黑龙江省	38.20	22	-0.606	成长期
山西省	37.05	23	-0.673	成长期
贵州省	35.01	24	-0.790	成长期
甘肃省	33.40	25	-0.883	成长期
内蒙古自治区	31.03	26	-1.020	成长期
新疆维吾尔自治区	30.54	27	-1.048	成长期
海南省	27.78	28	-1.208	成长期
宁夏回族自治区	25.09	29	-1.363	起步期
西藏自治区	23.93	30	-1.430	起步期
青海省	22.89	31	-1.490	起步期
平均值	48.70	--	0.00	--

从评价结果看,北京、广东、上海、江苏处于第一方阵,均处于成熟期,北京得分最高,为 90.79 分,是唯一超过 90 分的地区。宁夏、西藏、青海排序最低,属于第三方阵,均处于起步期。其他地区均在第二方阵,处在成长期。

3.3 一级指标评价结果比较分析 一级指标评价结果如表 3 所示。从人文社科科研投入看,广东、北京、江苏、浙江排在第四位。人文社科发展,离不开研发经费与研发劳动力的投入支持,广东省由于经济发达,在研发投入方面具有优势,近年来发展较快。浙江省人文社科投入发展势头良好。

从人文社科成果数量看,北京、江苏、上海、湖北排在前四位。北京作为首都,其人文社科数量具有明显优势,湖北省作为人文社科大省,其人文社科成果数量发展态势较好。

从人文社科成果质量看,北京、上海、湖北、天津排在前四位,北京、上海研究质量高是正常的,比较有特

色的是天津,其研发投入排序 22 位,成果数量排序 20 位,而成果质量排在第 4 位,天津规模较小,但研究质量较高,有两点启示:第一,在人文社科研究质量方面,天津可圈可点,值得学习。第二,从评价的角度,单纯评价质量是有问题的,必须兼顾数量进行评价,光有质量没有规模不能算是人文社科强省。

表 3 一级指标分析

地区	评价总分	排序	研发投入	排序	成果数量	排序	成果质量	排序
北京市	90.79	1	83.30	2	96.09	1	95.58	1
广东省	76.93	2	90.77	1	67.45	5	67.90	5
上海市	75.63	3	68.09	6	71.82	3	86.95	2
江苏省	73.54	4	77.88	3	82.03	2	62.50	7
湖北省	69.36	5	62.75	9	68.46	4	77.55	3
浙江省	67.86	6	76.53	4	62.48	7	61.80	8
山东省	61.86	7	68.85	5	64.36	6	52.08	12
四川省	59.97	8	64.62	7	54.09	11	58.85	9
河南省	57.24	9	63.08	8	57.70	8	50.24	14
重庆市	54.30	10	50.86	14	47.21	13	63.31	6
湖南省	54.16	11	58.07	10	54.21	10	49.67	15
陕西省	53.81	12	51.82	12	55.41	9	54.94	11
辽宁省	51.67	13	51.55	13	53.08	12	50.80	13
天津市	48.86	14	36.37	22	40.25	20	69.29	4
福建省	48.22	15	50.43	15	43.81	16	48.85	16
吉林省	47.68	16	42.40	19	44.40	15	56.05	10
河北省	46.77	17	54.62	11	45.46	14	38.75	20
安徽省	45.28	18	50.10	16	42.59	17	41.70	18
江西省	42.77	19	42.82	18	40.60	18	44.26	17
广西壮族自治区	39.12	20	44.11	17	38.07	23	34.18	24
云南省	38.88	21	38.74	20	39.21	21	38.79	19
黑龙江省	38.20	22	38.55	21	40.29	19	36.31	22
山西省	37.05	23	35.74	24	38.87	22	37.25	21
贵州省	35.01	24	35.86	23	36.58	24	32.92	26
甘肃省	33.40	25	31.16	25	36.49	25	33.74	25
内蒙古自治区	31.03	26	30.76	26	34.62	26	28.77	28
新疆维吾尔自治区	30.54	27	27.18	27	29.83	27	34.89	23
海南省	27.78	28	25.04	28	29.57	28	29.64	27
宁夏回族自治区	25.09	29	23.32	29	28.22	29	24.87	30
西藏自治区	23.93	30	21.52	31	26.30	31	25.00	29
青海省	22.89	31	21.59	30	26.93	30	21.50	31
均值	48.70		48.98		48.27		48.67	

3.4 评价结果与传统标准化评价结果比较 采用本文方法得到的评价总分与采用传统标准化方法得到的传统得分及一级指标评价结果如表 4 所示。采用传统标准化方法进行评价,其问题主要表现在以下几个方面:

第一,评价得分偏倚。采用极大值标准化,意味着每项指标的极大值就是满分,但是这没有任何依据,而采用 sigmoid 函数标准化,极大值很难是满分,俞立平等^[21]讨论了这个问题,认为传统标准化将极大值认定为满分理论依据不足。

由于评价得分偏倚,传统评价得分超过 60 分的只

有 3 个省市,这明显与人们的感受不符。而采用 sigmoid 函数标准化后,评价得分超过 60 分的有 8 个省市,明显更加合理。

第二,传统标准化的评价得分以及各一级指标评价得分的均值并不相等,也就是说存在自然权重问题,

比如期刊质量的均分为 38.87 分,而期刊数量的均分为 21.39 分,那么期刊质量就显得更加重要,问题是在专家赋权时,这个因素已经被考虑进去了,最终结果就是期刊质量的权重被进一步加大了。

表 4 最终结果与传统标准化结果比较

地区	评价 总分	排 序	传统 总分	排 序	研发 投入	排 序	创新 数量	排 序	创新 质量	排 序
北京市	90.79	1	91.27	1	84.35	2	92.42	1	98.37	1
广东省	76.93	2	65.19	2	93.74	1	36.13	5	53.33	5
上海市	75.63	3	62.09	3	59.74	5	41.31	3	79.62	2
江苏省	73.54	4	58.86	4	68.68	3	56.81	2	49.12	7
湖北省	69.36	5	50.89	6	48.94	9	37.53	4	62.66	3
浙江省	67.86	6	51.54	5	66.83	4	32.03	7	48.01	8
山东省	61.86	7	46.00	7	58.87	6	33.54	6	40.18	12
四川省	59.97	8	42.64	8	51.11	8	24.74	11	45.74	9
河南省	57.24	9	41.19	9	51.17	7	28.04	8	39.18	13
重庆市	54.30	10	37.14	11	36.60	14	19.33	13	50.48	6
湖南省	54.16	11	37.33	10	44.03	10	25.00	10	38.48	15
陕西省	53.81	12	36.45	12	37.55	12	25.83	9	42.79	11
辽宁省	51.67	13	34.65	13	37.35	13	24.05	12	39.14	14
天津市	48.86	14	31.89	14	21.51	22	13.47	20	56.91	4
福建省	48.22	15	31.60	15	36.14	15	16.17	16	37.45	16
吉林省	47.68	16	30.63	17	27.95	19	16.87	15	43.52	10
河北省	46.77	17	31.29	16	40.70	11	17.99	14	30.04	20
安徽省	45.28	18	29.39	18	35.81	16	15.45	17	32.01	18
江西省	42.77	19	26.82	19	28.23	18	13.92	18	34.43	17
广西壮族自治区	39.12	20	23.95	20	29.68	17	11.72	23	26.13	24
云南省	38.88	21	23.43	21	24.09	20	12.93	21	30.18	19
黑龙江省	38.20	22	22.75	22	23.73	21	13.65	19	28.12	22
山西省	37.05	23	21.56	23	20.77	24	12.49	22	28.94	21
贵州省	35.01	24	19.86	24	20.93	23	10.63	24	25.23	26
甘肃省	33.40	25	17.65	25	15.48	25	10.30	25	25.38	25
内蒙古自治区	31.03	26	15.81	26	15.12	26	8.98	26	21.48	28
新疆维吾尔自治区	30.54	27	14.36	27	10.70	27	2.59	29	26.96	23
海南省	27.78	28	11.89	28	7.62	28	4.23	27	22.25	27
宁夏回族自治区	25.09	29	8.88	29	5.05	29	2.74	28	17.63	29
西藏自治区	23.93	30	7.08	30	2.13	31	0.77	31	17.25	30
青海省	22.89	31	6.20	31	2.34	30	1.46	30	14.01	31
均值	48.70		33.24		35.71		21.39		38.87	

第三,排序结果有差异。虽然评价对象不多,只有 31 个省市,但对比评价结果,还是存在较大差异。说明评价方法由于自身的瑕疵会导致评价结果产生较大变化,进而影响排序。由于目前主流学术评价均采用传统评价方法,因此有必要进行改进。

分布如图 5 所示,Jarque-Bera 检验结果为 1.488,相伴概率为 0.475,不能拒绝正态分布原假设,说明采用 sigmoid 函数标准化进行评价后,最终评价结果服从正态分布,这对评价是非常有益的。

传统标准化方法评价结果的数据分布如图 6 所示,Jarque-Bera 检验结果为 0.351,相伴概率为 0.042,拒绝正态分布原假设,说明其并不服从正态分布,作为区域人文社科水平评价,应尽可能服从正态分布。

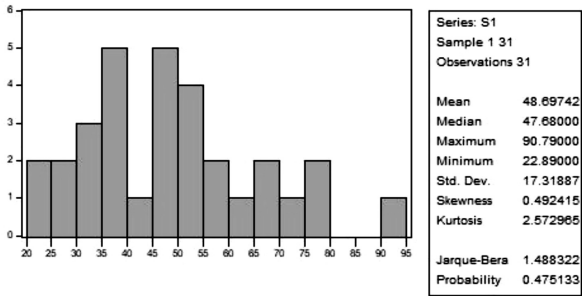
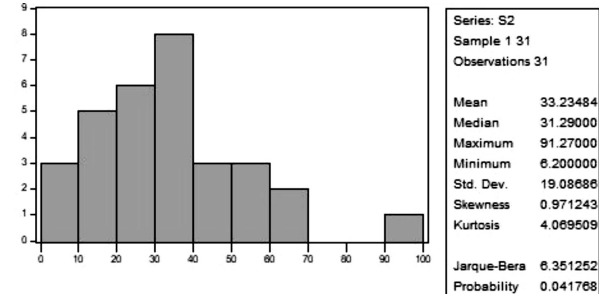


图 5 最终结果的数据分布



第四,评价结果数据分布差异。最终结果的数据

图6 传统结果的数据分布

4 研究结论

a. 在研究质量视角下进行区域人文社科评价非常独特。在高质量发展背景下,对区域人文社科水平进行评价意义重大。本文通过理论分析认为,在微观评价中,应更多采用同行评议,而在中观与宏观评价中,应该更多采用指标体系进行多属性评价。一些科学计量指标或文献计量指标,尽管不能用在微观评价中,但可以用在宏观评价中。微观评价应该以质量评价为主,而宏观评价必须兼顾数量和质量。

b. 我国地区人文社科总体发展态势良好。研究发现,人文社科水平较高的省市主要是北京、广东、上海、江苏等地,相对较弱的地区是宁夏、西藏、青海等地。31个省市自治区中,处在成熟期的省市有4个,处在起步期的省市有3个,其他均处于成长期,这和我地区人文社科发展情况总体上是相符的。北京、上海、江苏属于传统人文社科水平较高地区,而广东由于经济发展迅速,人文社科投入规模较大,近年来发展较快。单纯从成果质量看,北京、上海、湖北、天津位居前列。天津规模较小,但人文社科研究质量较高,当然这是值得鼓励的。但是从另外一个角度,地区人文社科发展如果仅仅追求质量,不追求人文社科的规模和数量,那也不利于人文社科的发展。

c. 传统线性加权汇总评价中标准化方法有待改进。传统线性加权汇总评价方法作为一种主流的学术评价方法,采用极大值进行指标标准化是值得商榷的。第一是存在自然权重问题,即标准化后指标均值不相等导致实际权重偏倚。第二是打分缺乏绝对标准,某个评价对象指标为极大值并不代表其就是满分。第三是标准化方法不同也会影响评价结果。

d. 基于 sigmoid 函数标准化的线性加权是一种优秀的评价方法。本文的实证研究表明,采用 sigmoid 函数标准化可以有效解决传统线性加权汇总中的几个问题,并且标准化后指标均值接近相等,这样会直接导致一级指标评价结果的均值也接近相等,彻底解决了自然权重问题。此外,sigmoid 函数接近成长曲线,符合事物发展规律,拥有绝对的打分标准。并且评价结果更加接近正态分布,这对宏观地区人文社科评价而言更有应用价值。

参考文献

- [1] 刘庆昌. 人文社科学术期刊稿件质量的判断[J]. 编辑之友, 2015(5): 75-80.
- [2] Talib, Ali A. Simulations of the submission decision in the re-

search assessment exercise; The "who" and "where" decision [J]. Education Economics, 1999, 7(1): 39-51.

- [3] Mazlish B. The quality of "The Quality of Science": An evaluation[J]. Science Technology & Human Values, 1982, 7(1): 42-52.
- [4] Moed H F. UK research assessment exercises: Informed judgments on research quality or quantity? [J]. Scientometrics, 2008, 74(1): 153-161.
- [5] 蒋玲,施立红,苗林. 大数据对人文社科成果评价的价值研究[J]. 新世纪图书馆, 2016(3): 90-93.
- [6] 高自龙,杨红艳. 学术评价:理想与现实之间的优化选择——人文社科论文评估指标体系的完善与实施新探[J]. 江汉论坛, 2011(11): 124-129.
- [7] 李贞,张体勤,吴彬. 高校人文社科学术创新团队绩效评价体系的优化探析[J]. 东岳论丛, 2013, 34(12): 188-191.
- [8] 潘启亮. 人文社科科研量化评价研究[J]. 暨南学报(哲学社会科学版), 2011, 33(3): 31-37.
- [9] 高自龙,范晓莉. 合作研究与人文社科论文质量的相关性探析——以人大《复印报刊资料》2010年转载论文评估数据为例[J]. 中州学刊, 2011(6): 246-250.
- [10] 蒋玉梅. 全球科研评价体系的演进与发展[J]. 国家教育行政学院学报, 2013(9): 81-86.
- [11] Prathap G. Measures for bibliometric size, impact, and concentration[J]. Journal of the Association for Information Science and Technology, 2015, 66(8): 1740-1741.
- [12] 刘仲翔. 高质量发展与学术期刊转型:2018年国内社科学术期刊动态盘点[J]. 科技与出版, 2019(3): 33-39.
- [13] Ives G, Rowley G. Supervisor selection or allocation and continuity of supervision: Ph. D. students' progress and outcomes [J]. Studies in Higher Education, 2005, 30(5): 535-555.
- [14] 李霞,宋俊波. 人文社科博士学位论文质量要素特征评价——以中国人民大学人文社科类博士学位论文为例[J]. 学位与研究生教育, 2017(11): 13-17.
- [15] Martin G. Kocher, Mikulas Luptacik, Matthias Sutter. Measuring productivity of research in economics: A cross-country study using DEA[J]. Socio Economic Planning Sciences, 2006, 40(4): 314-332.
- [16] 姜彤彤. 高校人文社科研究全要素生产率评价及分析[J]. 研究与发展管理, 2013, 25(5): 90-97.
- [17] 姜彤彤. 我国各省高校人文社科科研效率评价及区域差异研究[J]. 科技管理研究, 2014, 34(15): 92-96, 123.
- [18] 曹骏,宋丽萍. "以刊评文"的利弊分析与建议[J]. 陕西青年管理干部学院学报, 2004(4): 40-41.
- [19] 周兴旺. 从逻辑学的角度看期刊影响力和论文影响力的关系[J]. 攀枝花学院学报, 2009, 26(4): 55-56, 31.
- [20] 俞立平. 线性科技评价中自然权重问题及修正研究——动态最大均值逼近标准化方法[J]. 统计与信息论坛, 2018, 33(10): 27-33.
- [21] 俞立平,阮先鹏,吴贤豪,等. 基于 Sigmoid 函数的文献计量指标评价标准研究[J]. 情报杂志, 2020, 39(9): 176-182.

(责编/校对:王平军)