

基于K-均值聚类分析的河北省区域社会经济持续发展等类研究

张丽君

(河北工程大学 财务处,河北 邯郸 056038)

[摘要] 由于多年来的不合理经济发展方式,造成了现在的各种社会经济问题,政府提出可持续发展的理念来解决这些问题,其中有关社会经济的可持续发展是关键。区域经济的发展越来越重要,在整体经济发展中占有十分重要的地位,社会经济的发展依赖于多个方面,本文主要根据经济与社会关系来分析社会经济持续发展的发展。通过快速聚类方法对河北省区域社会经济可持续发展的类别及等级进行处理,在理论上对区域社会经济方面持续发展政策制定提供依据。

[关键词] 社会经济;持续发展;聚类
[中图分类号] F127 [文献标识码] A [文章编号] 1673-9477(2010)02-0037-02

河北省是资源型发展突出的省份,伴随着工业化进程的加速,社会物质生活日益丰富的同时,生态破坏、环境污染、能源短缺等一系列问题也随之而来。一直在致力于寻找一条生态、资源、环境、人口、科技和经济相互协调的可持续发展之路,而国内外对可持续发展的研究大都是从整个国家的角度出发,从宏观方面研究多,很少从区域经济的角度研究。事实上,区域经济与国家的可持续发展不同。同时,区域经济发挥着越来越大的作用,但是在区域社会经济持续发展等类研究上存在不足,以至于在服务政府政策制定时缺乏强有力支持。本文从区域社会经济的角度研究可持续发展的理论,借助聚类分析方法处理建立的指标体系获得结果,从而来确定区域社会经济持续发展的等类。

一、区域社会经济持续发展

(一) 区域社会经济

区域可以按行政管理标准划分,也可以按流域分水岭划分,还可以按其他研究目的来划分。各个地区独特的自然、社会和经济条件,导致了各地区经济发展水平、结构和布局的差异。区域与国家存在的不同,导致了区域经济与国家经济的不同。区域与国家有较大区别,虽然国家也是一种区域,但这里所说的区域,则主要指国家内部的区域,二者有三个主要区别。第一,区域较开放,而国家较封闭。第二,区域政策工具较少,缺乏国家政府所拥有的如关税、进口配额、移民、外汇、金融、财政等政策工具。第三,区域发展受国家政策及该区域的地缘条件影响较大。区域经济发展的影响因素有人口因素、自然资源、科学技术水平、生态环境因素,本文研究是区域社会经济持续发展问题,故而在对区域社会经济的影响因素中选择人力、科技力、社会、经济本身,这些因素与区域经济构成了社会经济方面研究关系。

(二) 可持续发展理论

广义的定义认为可持续发展是一种新的发展思想和战略,目标是保证社会具有长期的持续性发展能力,确保自然资源、生态环境的安全和稳定,避免社会、经济大起大落的波动。将可持续发展的侧重角度摆放不一会有不同的解释,本文侧重社会经济方面,根据已有研究成果,侧重于社会方面的定义,可持续发展定义为“在生存于不超出维持生态系统涵容能力之情况下,改善人类的生活品质”。侧重于经济方面的定义,可持续发展定义为“在保持自然资源的质量及其所提供服务的条件下,使经济发展的净利益增加到最大限度”。

(三) 区域社会经济可持续发展

区域社会经济可持续发展,就是指在现有社会环境约束条件下,充分发挥区域优势,实现区域经济的持续有效增长。将可持续发展理论应用于区域社会经济的发展研究中,可以从以下4个方面综合进行。区域内人力资源。人力资源是区域经济可持续发展的主体,人作为生产和消费的主体,是经济发展中最活跃的因素,离开人力资源的支持,区域经济的可持续发展将无从谈起;区域内科学技术。科学技术是第一生产力,科学技术为经

济的可持续发展提供不可或缺的智慧支持,经济可持续发展要通过利用科学技术来发展集约型经济、循环经济;区域内社会。社会有序发展是经济持续增长的重要保证,经济与社会协调发展的最重要表现形式就是促进人的全面发展。贯彻“以人为本”的发展原则,就应该在经济与社会协调发展的指标体系中,更注重“人文指标”,如把生活满意度或幸福感作为经济与社会协调发展状况的重要衡量指标;经济本身。区域经济的发展首先是发展,只有经济发展,才能使人类摆脱贫困,才能从根本上解决资源、环境、生态、社会等存在的问题。

二、等类分析理论及实证分析

(一) K-均值聚类

K-均值聚类是J. B. MacQueen在1967年提出的K-means算法到目前为止用于科学和工业应用的诸多聚类算法中一种极有影响的技术。它是聚类方法中一种基本的划分方法,常常采用误差平方和准则函数作为聚类准则函数。K-means算法的工作原理:算法首先随机从数据集中选取K个点作为初始聚类中心,然后计算各个样本到聚类中的距离,把样本归到离它最近的那个聚类中心所在的类。计算新形成的每一个聚类的数据对象的平均值来得到新的聚类中心,如果相邻两次的聚类中心没有任何变化,说明样本调整结束,聚类准则函数已经收敛。本算法的一个特点是在每次迭代中都要考察每个样本的分类是否正确。若不正确,就要调整,在全部样本调整完后,再修改聚类中心,进入下一次迭代。如果在一次迭代算法中,所有的样本被正确分类,则不会有调整,聚类中心也不会有任何变化,这标志着已经收敛,因此算法结束。

(二) 河北省区域社会经济持续发展等类分析

区域社会经济可持续发展等类分析借助K-均值聚类需要一个指标体系,评价指标体系作为一个复杂的系统。根据本文对区域社会经济可持续发展的内涵引用,遵循建立指标体系的原则及实际数据的获得,本文建立了一个区域社会经济可持续发展的评价指标体系,指标分为三级,一级为评价体系,二级为四个分项,人力资源(人口数、人口自然增长率)、科技能力(人才密度指数、专业技术人才数(万人))、区域内社会(医生数、消费价格指数)、经济自身(区域GDP、GDP增长率)(表1)。

表1 区域社会经济可持续发展的评价指标体系

一级指标	二级指标	三级指标
区域	人力资源	人口数X ₁ 人口自然增长率X ₂
	科技能力	人才密度指数X ₃ 专业技术人才数(万人)X ₄
社会	区域内社会	医生数X ₅ 消费价格指数X ₆
经济		区域GDPX ₇ GDP增长率X ₈
持续	经济本身	
发展		

表2 河北省区域社会经济持续发展评价数据

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
石家庄	9780000	5.82	11.60	52300	10628	104.3	96983000000	13
承德	3390000	7.84	7.70	14900	2212	107.5	16156000000	17.6
张家口	4200000	5.93	8.20	21000	2697	105.2	22789000000	14.5
秦皇岛	7930000	6.06	11.30	33300	3493	106.5	26313000000	12.6
唐山	7390000	5.43	10.50	60000	8202	104	91505000000	14.2
廊坊	4060000	5.77	9.60	16600	2160	104.8	32891000000	16.1
保定	10850000	7.44	7.20	21000	4100	105.5	59222000000	13.5
沧州	6930000	6.94	9.70	21800	2211	105.3	45134000000	10.8
衡水	4260000	5.68	7.70	14800	2277	104.9	27722000000	16.3
邢台	6880000	6.96	6.60	18600	2955	105.4	36628000000	13.2
邯郸	8760000	7.29	8.30	29800	5215	104.7	54235000000	14.5

表3 初始聚类中心

	聚类		
	1	2	3
人口总数	9780000	3390000	8760000
人口自然增长率	6	8	7
人才密度指数	12	8	8
专业技术人才数(万人)	52300	14900	29800
医生数	10628	2212	5215
消费价格指数	104	108	105
区域 GDP	96983000000	16156000000	54235000000
GDP 增长率	13	18	15

表4 迭代历史记录

迭代	聚类中心内的更改		
	1	2	3
1	2.739E9	9.018E9	5.430E9
2	.000	1.909E9	4.059E9
3	.000	.000	.000

表5 最终聚类中心

	聚类		
	1	2	3
人口总数	8585000	5120000	8846667
人口自然增长率	6	6	7
人才密度指数	11	9	8
专业技术人才数(万人)	56150	19867	24200
医生数	9415	2632	3842
消费价格指数	104	106	105
区域 GDP	94244000000	27083166667	52863666667
GDP 增长率	14	15	13

注:a. 由于聚类中心内没有改动或改动较小而达到收敛。任何中心的最大绝对坐标更改为 .000。当前迭代3。初始中心间的最小距离为 38079000378.649。

表6 每个聚类中的案例数

聚类	1	2.000
	2	6.000
	3	3.000
有效		11.000
缺失		1.000

处理结果在原始数据中会有 QCL - 1 的对应值,它们对应各地区依次为 1、2、2、1、2、3、3、2、2、3。表3-3 显示了3个类的初始类中心情况,可以看出,第一类的各类指标值总体上使最优的,往下依次为第二类、第三类。表3-4 展示了3个类中心点每次迭代的偏移情况,可知第一次迭代3个类的中心点分别偏移了 2.739E9、9.018E9、5.430E9,直到第三次迭代3个类的中心点偏移才达到指定判定标准(0)。表3-5 展示了3个类的最终类中心情况,总体来看,第一类各指标值仍是最优的。表3-6 给出了各类中的样品数目,第一类包括2个地区,第二类包括6个地区,第三类包括3个地区。从 QCL - 1 值对应地区来看,第一类是石家庄、唐山;第二类是承德、张家口、秦皇岛、廊坊、衡水、邢台;第三类是保定、沧州、邯郸。

三、结论

经济的可持续发展并不是孤立存在的,本文建立了以社会经济本身为中心,同时与人力资源、科学技术、社会相互作用的区域经济可持续发展评价体系,提出了新的想法,并在数据选择、算法选择上提出一些比较适用的方法。但是由于区域经济可持续发展评价体系系统的复杂性,在评价指标和评价方法的选择上还需要进一步研究与完善,获得的结果在政策制定研究上提供的是理论支持。

[参考文献]

[1]何晓群.多元统计分析[M].中国人民大学出版社,2009.
[2]吴殿延.区域经济学[M].科学出版社,2003.
[3]孙曰瑶.区域经济持续发展理论与应用[M].山东大学出版,1996.
[4]袁建林.区域经济可持续发展评价指标体系的构建[J].中国市场,2008.

[责任编辑:陶爱新]

Classification study of socio - economic sustainable development in Hebei province based on K - means clustering analysis

ZHANG Li - jun

(Hebei University of Engineering, Handan 056038, China)

Abstract: The years of irrational economic development mode results in the current socio - economic problems, and the government introduces the concept of sustainable development to address these issues, in which socio - economic sustainable development is the key. Regional economic development becomes more and more important, and plays an important role in the overall economic development. The socio - economic development depends on many aspects. This article primarily analyzes social and economic sustainable development on the basis of economic and social relations. Process the classification and types of socio - economic sustainable development in Hebei Province by fast clustering method and in theory, provide the basis for the region socio - economic sustainable development policy.

Key words: socio - economy; sustainable development; clustering