

城市宜居程度的一个评价模型

卢天秀,唐浩,刘美君,刘玲,杨勇

(四川理工学院数学与统计学院,四川 自贡 643000)

摘要:针对宜居城市评价指标问题,在合理的假设条件下,采用理论分析法和频度统计法筛选出主要指标(包括5个一级指标和8个二级指标),利用熵值法建立了评价宜居城市的数学模型。然后以8个城市为例,计算出一级指标和二级指标的权重,以及每个二级指标的熵值,从而给出8个城市综合得分排名。最后,得出生态环境和文化教育这两个指标的变化会对宜居城市排名产生显著的影响。

关键词:频度统计法;熵值法;主成分分析法;KMO值(Kaiser-Meyer-Olkin值);模糊物元模型

中图分类号:O29;O24;O213

文献标志码:A

引言

城市宜居性是当前城市科学研究领域的热点议题之一,也是政府和城市居民密切关注的焦点,对提升城市居民生活质量、完善城市功能和提高城市运行效率具有重要意义^[1]。

我国宜居城市的排名每年都是热门话题,评价指标体系不同,宜居城市排名结果也会发生变化。2016年,中科院发布了《中国宜居城市研究报告》^[2],在被调查的40个城市中,排名前十的城市分别为:青岛、昆明、三亚、大连、威海、苏州、珠海、厦门、深圳、重庆。而美世人力资源咨询公司(William Mercer)公布的2016年全球宜居城市排行中大陆前十名分别为上海、北京、广州、成都、南京和深圳(并列)、西安、重庆、青岛、沈阳、吉林。

人们选择留在某个城市,不单是为了生存,更是寄托了自己的梦想与希望。对大多数人,衡量是否宜居或许就是八个字:衣食住行、安居乐业。

本文筛选出评价宜居城市的5个一级指标,8个二级指标,建立评价宜居城市的数学模型,对淮海经济区内的8个城市(宿迁、连云港、宿州、商丘、济宁、枣庄、徐州、淮北)进行合理性研究,给出宜居城市排名。

1 模型的建立

首先,假设获得的数据全部真实可靠,且取自同一个正态总体。评价宜居城市的各个指标之间的相互作用关系可以忽略不计。

根据文献[3]以及2007年5月30日正式发布的《宜居城市科学评价标准》(该文件是由科学研究会研究

收稿日期:2017-08-11

基金项目:国家自然科学基金(11501391);人工智能四川省重点实验室开放基金项目(2015RZJ01);桥梁无损检测与工程计算四川省高校重点实验室开放基金项目(2014QZJ02);四川理工学院教学改革研究项目(JG-1708;JG-1102)

作者简介:卢天秀(1976-),女,四川自贡人,教授,博士,硕士生导师,主要从事拓扑动力系统,混沌理论及其应用方面的研究,(E-mail) lubeeltx@163.com

编订,并通过建设部科技司组织评审验收,具有一定的权威性),选取的主要指标见表 1。

表 1 主要指标

第一级指标	编号	第二级指标	编号
经济发展	H_1	人均 GDP(元)	x_1
		城镇居民家庭人均可支配收入(元)	x_2
生态环境	H_2	生活垃圾无害化处理率(%)	x_3
		建成区绿化覆盖面积(公顷)	x_4
居住条件	H_3	人口密度(人/平方公里)	x_5
文化教育	H_4	教育业单位从业人员数(万人)	x_6
社会保障	H_5	城镇居民最低生活保障人数(人)	x_7
		城镇登记失业率(%)	x_8

表 1 中的五个一级指标基本概括了城市宜居的条件^[4]。人均 GDP 是衡量经济指标的总量,反映了一个城市的经济发展情况,城镇居民可支配收入反映了人们的生活水平,因此选取“ x_1 ”和“ x_2 ”两个二级指标反映经济发展;绿化和卫生直接体现了城市生态改善强度,所以选取“ x_3 ”和“ x_4 ”两个二级指标反映生态环境^[5]; x_5 是大多数国家和城市反映居住条件的指标;一个城市整体的文明程度,可以用教育发展的状况体现,所以选择“ x_6 ”直观体现文化教育这一指标;最低生活保障、失业率和广大基层人民的生活息息相关,因此选取“ x_7 ”和“ x_8 ”反映社会保障情况^[6]。

应用熵值法建立宜居城市评分模型,熵值法能够深刻反映出指标信息熵值的效用价值,适合对多元指标进行综合评价。运用熵值法可克服多指标变量间的信息重叠和人为确定权重的主观性^[7,9]。

主要步骤:

(1) 选取 n 个城市, m 个指标,则 X_{ij} 为第 i 个城市的第 j 个指标的值 ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)。

(2) 指标的标准化处理:异质指标同质化。由于各项指标的计量单位并不统一,因此在计算综合指标前,先要对其进行标准化处理,即把指标的绝对值转化为相对值,并令 $X_{ij} = |X_{ij}|$, 从而解决各项不同指标值的同质化问题。而且,由于正向指标和负向指标数值代表含义不同(正向指标数值越高越好,负向指标数值越低越好),因此,对于高低指标用不同的算法进行数据标准化处理。

正向指标:

$$X_{ij}^* = \frac{X_{ij}}{X_{j,\max}} \quad (1)$$

负向指标:

$$X_{ij}^* = \frac{X_{j,\min}}{X_{ij}} \quad (2)$$

则 X_{ij}^* 为第 i 个城市的第 j 个指标的数值 ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$)。

(3) 计算第 j 项指标下第 i 个城市占该指标的比重

P_{ij} :

$$P_{ij} = \frac{X_{ij}^*}{\sum_{i=1}^n X_{ij}^*} \quad (3)$$

其中, $i = 1, 2, \dots, n; j = 1, 2, \dots, m$ 。

(4) 计算第 j 项指标的熵值 e_j :

$$e_j = -k \sum_{i=1}^n P_{ij} \ln(P_{ij}) \quad (4)$$

其中, $k = \frac{1}{\ln(n)} > 0$, 满足 $e_j \geq 0$ 。

(5) 计算第 j 项指标的差异性系数 d_j :

$$d_j = 1 - e_j \quad (5)$$

d_j 越大,第 j 项指标越重要。

(6) 计算各项指标的权重 w_j :

$$w_j = \frac{d_j}{\sum_{j=1}^m d_j} \quad (6)$$

(7) 构建指标综合评价模型 $f(i)$:

$$f(i) = \sum_{j=1}^m w_j \cdot P_{ij} \quad (7)$$

由此模型,可以得到第 i 个城市的综合得分。

2 八个宜居城市排名

用 y_1, y_2, \dots, y_8 分别代表宿迁、连云港、宿州、商丘、济宁、枣庄、徐州和淮北 8 个市,根据《中国城市统计年鉴》、《中国民政统计年鉴》、以及江苏、安徽、山东、河南省的统计年鉴,2015 年 8 个城市关于 8 个二级指标 x_1, x_2, \dots, x_8 的数据矩阵为:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
35484	18846	78.2	8156	565	5.86	31326	2.38
40416	22985	82.4	10348	582	5.07	14399	2.3
18768	21712	87.1	2671	656	5.69	42428	3.7
21073	20214	82	2821	880.6	9.38	114849	3.7
42796	27956	100	6310	749.5	8.65	44521	2.99
48346	25238	100	5870	867.8	4.59	45361	2.45
51714	23770	99.9	17806	763	10.5	35320	2.14
32996	22460	100	3552	782.6	2.15	57947	4.13

可计算出每列的平均数为:

$$\bar{X}_{.1} = 36449.18, \bar{X}_{.2} = 22897.77$$

$$\bar{X}_{.3} = 91.2, \bar{X}_{.4} = 7236.74$$

$$\bar{X}_{.5} = 730.8, \bar{X}_{.6} = 6.48$$

$$\bar{X}_{.7} = 48268, \bar{X}_{.8} = 2.97$$

原始数据矩阵可知 y_2 (连云港市) 和 y_7 (徐州市) 的大部分正向指标高于平均数, 负向指标均低于平均数。商丘的正向指标明显偏低, 负向指标偏高, 特别是城镇居民低保人数远高于平均数。

按照熵值法建立的模型确定 6 个正向二级指标: $x_1, x_2, x_3, x_4, x_5, x_6$; 2 个负向二级指标: x_7, x_8 , 按照正向指标的标准化公式(1)与负向指标的标准化公式(2)进行处理, 并计算:

(1) 对标准化后的数据, 按照熵值法步骤中式(3)、式(4)和式(5)得到熵值 e_j 、差异性系数 d_j :

	e_j	d_j
x_1	0.976002298	0.023997702
x_2	0.996758385	0.003241615
x_3	0.997612176	0.002387824
x_4	0.907152309	0.092847691
x_5	0.994248622	0.005751378
x_6	0.958834828	0.041165172
x_7	0.924643276	0.075356724
x_8	0.987034236	0.012965764

(2) 按式(6)计算一级指标的权重为:

$$w(H_1) = 0.105695968$$

$$w(H_2) = 0.369539735$$

$$w(H_3) = 0.0223169130$$

$$w(H_4) = 0.159732078$$

$$w(H_5) = 0.0342715307$$

二级指标的权重为:

$$w_1 = 0.09311762$$

$$w_2 = 0.012578348$$

$$w_3 = 0.009265406$$

$$w_4 = 0.360274329$$

$$w_5 = 0.022316913$$

$$w_6 = 0.159732078$$

$$w_7 = 0.292404612$$

$$w_8 = 0.050310695$$

(3) 运用综合评价模型式(7), 得到 8 个城市的一级指标分数矩阵为:

	H_1	H_2	H_3	H_4	H_5
y_1	0.072	0.180	0.014	0.089	0.180
y_2	0.083	0.217	0.015	0.077	0.339
y_3	0.044	0.062	0.017	0.087	0.128
y_4	0.047	0.065	0.022	0.070	0.137
y_5	0.090	0.137	0.019	0.132	0.131
y_6	0.098	0.128	0.022	0.070	0.137
y_7	0.104	0.370	0.019	0.160	0.170
y_8	0.070	0.081	0.020	0.033	0.099

从而, 淮海经济区 8 个城市综合得分和排名为:

$$\text{Score}(y_7) = 0.822; \text{Score}(y_2) = 0.732$$

$$\text{Score}(y_1) = 0.535; \text{Score}(y_5) = 0.508$$

$$\text{Score}(y_6) = 0.455; \text{Score}(y_4) = 0.343$$

$$\text{Score}(y_3) = 0.338; \text{Score}(y_8) = 0.302$$

3 指标显著性分析

在建立的评价模型中, 不同指标对于权重的影响不同, 需要分析其中哪些是重要指标, 这里选择主成分分析法, 通过保留低阶主成分, 忽略高阶主成分达到减少数据集的维数, 同时保持数据中对方差贡献最大的那些特征^[10-11]。

将原始数据在 SPSS 软件中进行因子分析, 得到 $x_1,$

$x_2, x_3, x_4, x_5, x_6, x_7, x_8$ 的相关矩阵:

1	0.594	0.564	0.738	0.065	0.135	-0.536	-0.761
0.594	1	0.777	0.151	0.320	0.134	-0.248	-0.183
0.564	0.777	1	0.159	0.555	-0.021	-0.077	0.047
0.738	0.151	0.159	1	0.247	0.448	0.531	0.795
0.065	0.320	0.555	0.247	1	0.211	0.717	0.332
0.135	0.134	-0.021	0.448	0.211	1	0.276	0.296
-0.536	-0.248	-0.077	0.531	0.717	0.276	1	0.626
-0.761	-0.183	0.047	0.795	0.332	0.296	0.626	1

此矩阵不是正定矩阵, 而且没有产生 KMO 检验值。原因主要可能集中在两点: (1) 样本量太少, 而指标过多; (2) 某些变量间的相关性太强; 没有产生 KMO 值的情况可能是两个以上指标为纯线性关系。

这里不可能增加样本, 故只能从指标间的相关性考

虑。通过查看相关矩阵,大部分变量存在相关性,故需要删除高度相关的指标。利用逐一淘汰法找到了 5 个指标: x_1, x_4, x_6, x_7, x_8 , 得出相关矩阵:

$$\begin{pmatrix} 1 & 0.738 & 0.135 & -0.761 & -0.536 \\ 0.738 & 1 & 0.448 & -0.795 & -0.531 \\ 0.135 & 0.448 & 1 & -0.296 & 0.276 \\ -0.761 & -0.795 & -0.296 & 1 & 0.626 \\ -0.536 & -0.531 & 0.276 & 0.626 & 1 \end{pmatrix}$$

KMO 检验和 Bartlett 检验结果见表 2。

表 2 KMO 检验和 Bartlett 检验

检验方式	数值	
取样足够度的 Kaiser - Meyer - Olkin 度量。	0.618	
Bartlett 的球形度检验	近似卡方	15.794
	df	10
	Sig.	0.106

KMO 值大于 0.6,说明可以进行因子分析。碎石图(图 1)横轴表示指标的序号,纵轴代表特征根。

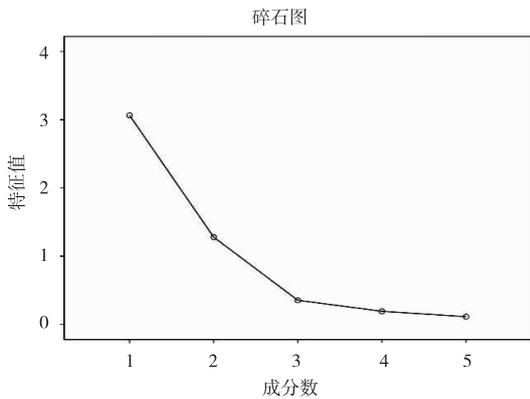


图 1 碎石图

由图 1 可知:随着指标序号增大,特征根迅速降低,后两个指标的特征根变化非常小,所以认为得到两个因子比较合适。从而得到成份矩阵 R_1 :

$$R_1 = \begin{pmatrix} -0.933 & -0.026 \\ 0.918 & 0.212 \\ 0.875 & -0.077 \\ -0.706 & 0.604 \\ 0.296 & 0.929 \end{pmatrix} \begin{matrix} x_8 \\ x_4 \\ x_1 \\ x_7 \\ x_6 \end{matrix}$$

具有 Kaiser 标准化的斜交旋转法,在 5 次迭代后收

敛,模式矩阵 R_2 :

$$R_2 = \begin{pmatrix} -0.921 & -0.107 \\ 0.881 & 0.291 \\ 0.878 & -0.001 \\ -0.784 & 0.542 \\ 0.164 & 0.952 \end{pmatrix} \begin{matrix} x_8 \\ x_4 \\ x_1 \\ x_7 \\ x_6 \end{matrix}$$

最后得到结构矩阵 R_3 :

$$R_3 = \begin{pmatrix} -0.927 & -0.155 \\ 0.896 & 0.337 \\ 0.878 & -0.045 \\ -0.756 & 0.5 \\ 0.215 & 0.961 \end{pmatrix} \begin{matrix} x_8 \\ x_4 \\ x_1 \\ x_7 \\ x_6 \end{matrix}$$

由成分矩阵 R_1 、模式矩阵 R_2 和结构矩阵 R_3 可得:第一公因子在指标 x_4 数据上具有最大的载荷,第二公因子在 x_6 数据上具有最大的载荷。 x_4 属于第一指标 H_2 , x_6 属于第一指标 H_4 , 所以,得出生态环境和文化教育的变化会对宜居城市的排名产生显著的影响。

4 结果检验

本文得出的城市宜居程度排名:徐州市、连云港市、宿迁市、济宁市、枣庄市、商丘市、宿州市、淮北市。根据 2015 年 8 个城市二级指标值原始数据矩阵,徐州市的全部正向指标数据均高于平均数,负向指标数据均低于平均数,连云港市的大部分正向指标高于平均数,且其差值不大,负向指标均低于平均数;而淮北市的各项指标在 8 个城市中都偏低,且大部分正向指标都低于平均数,负向指标均高于平均数,所以说徐州排名第一,连云港市排名第二,淮北排名靠后是合理的,由此确定本文的排名是基本合理的。当然,使用模糊物元模型解决实际问题时,由于选取的指标和使用的数学方法不同,或者原始数据的准确性等因素的影响,可能会影响模型的可靠性^[12-15]。在本文讨论的问题中,人们对评价宜居城市的标准具有不同的认识,而且选取的指标不一定非常全面,有的代表性指标因为数据限制而没有选取,所以排名结果并不绝对。

参考文献:

- [1] 石永林,王要武.建设可持续发展生态城市的研究[J].中国软科学,2003(8):122-126.
- [2] 中科院.中国宜居城市研究报告[M].北京:社会科学文献出版社,2016.
- [3] 周维,张小斌,李新.我国人居环境评价方法的研究进展[J].安全与环境工程,2013,20(2):14-18.
- [4] 尹东旭.内蒙古城市人居环境质量空间差异评价研究[D].内蒙古:内蒙古师范大学,2016.
- [5] 张晓,戴菲.基于POE模式的街旁绿地规划设计研究:以武汉市为例[J].规划师论丛,2010(1):1-4.
- [6] 毛通.城乡社会保障一体化评价体系研究及实证分析[J].社会保障研究,2012(1):15-27.
- [7] 薛丽芳,欧向军,谭海樵.基于熵值法的淮海经济区城市中心性评价[J].地理与地理信息科学,2009,25(3):63-66.
- [8] 李杰,陈璐.基于熵值法的上市公司绩效评价模型构建:以有色金属行业为例[J].财会通讯,2014(14):25-27.
- [9] 刘娜,艾南山,方艳,等.基于熵权的模糊物元模型在城市生态系统健康评价中的作用[J].成都理工大学学报:自然科学版,2017(5):589-595.
- [10] 李莉娟,张娜.关于主成分分析的几种常用改进方法[J].统计与咨询,2007(21):167-168.
- [11] 黄俐,梁鹏.基于因子分析的地铁盾构施工沉降风险辨识[J].内蒙古大学学报:自然科学版,2016(2):209-216.
- [12] JUN J F, LING Y Z, YU S C, et al. Health evaluation of a regional logistics industrial ecosystem in China based on fuzzy matter-element analysis method[J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(4): 2195-2202.
- [13] 宁宝权,陕振沛.基于改进熵和灰关联分析的模糊物元分析模型及应用[J].数学的实践与认识,2016,46(20):280-284.
- [14] GUO Q X, TIAN T L, MING X S. Evaluation analysis for sortie generation of carrier aircrafts based on nonlinear fuzzy matter-element method[J]. Journal of Intelligent & Fuzzy Systems, 2016, 31(6): 3055-3066.
- [15] ZHENG X G, JIAN G Y. Fuzzy Matter-Element Evaluation Method for Reliability Analysis of an Existing Highway Tunnel [J]. Advanced Materials Research, 2011, 1067(163): 3110-3113.

An Evaluation Model on the Degree of Urban Livability

LU Tianxiu, TANG Hao, LIU Meijun, LIU Ling, YANG Yong

(School of Mathematics and Statistics, Sichuan University of Science & Engineering, Zigong 643000, China)

Abstract: Aiming at the evaluation index of urban livability, based on the reasonable assumptions, by theoretical analysis and frequency statistics, the main indexes (including five first-level indicators and eight second-level indexes) were selected. Using the entropy method, mathematical model of urban livability is carried out. Then, taking the eight cities as examples, the weight of the primary indexes and the secondary indexes, the entropy of secondary indexes are calculated. So the overall score ranking of the eight cities is given. Finally, it is obtained that the changes in the eco-environment and cultural education have significant impact on the degree of urban livability.

Key words: frequency statistics; entropy method; principal component analysis; Kaiser-Meyer-Olkin value; fuzzy matter-element model