

Study on the Impact of Urbanization on Ecological Environment Based on Regression Analysis

—Taking Z City as an Example

Jiaquan Lv, Xiaoyu Li, Yuxiang Zong, Wenyu Zhang, Xinyu Gao, Feng Hu

School of Statistics, Qufu Normal University, Qufu Shandong
Email: tjlvjiaquan@163.com

Received: Aug. 2nd, 2019; accepted: Aug. 14th, 2019; published: Aug. 23rd, 2019

Abstract

The development of urbanization is an important factor to measure the comprehensive strength of a region. At the same time, the ecological environment is inextricably linked with people's production and life. In order to study the impact of urbanization development on the ecological environment, this paper selects the relevant data of Z city, constructs the urbanization and ecological environment index system, takes the relevant indicators of urbanization development as the independent variables, and integrates multiple indicators related to the ecological environment through the entropy method. For the dependent variable, R software was used to analyze the sample data obtained, and the multiple regression equation was established by stepwise regression method. The analysis results show that factors such as per capita GDP have a positive impact on the ecological environment, and the amount of civilian vehicles has a negative impact on the ecological environment. Based on the above analysis results, this paper proposes rationalization opinions on urbanization development from the three aspects of population, economy and society, and promotes the harmonious development of ecological environment and urbanization.

Keywords

Urbanization, Ecological Environment, Entropy Method, Regression Analysis

基于回归分析的城镇化发展对生态环境影响的研究

——以Z市为例

吕嘉全, 李晓宇, 宗宇翔, 张文宇, 高鑫雨, 胡 锋

文章引用: 吕嘉全, 李晓宇, 宗宇翔, 张文宇, 高鑫雨, 胡锋. 基于回归分析的城镇化发展对生态环境影响的研究[J]. 统计学与应用, 2019, 8(4): 689-695. DOI: [10.12677/sa.2019.84078](https://doi.org/10.12677/sa.2019.84078)

曲阜师范大学统计学院, 山东 曲阜
Email: tjlvjiaquan@163.com

收稿日期: 2019年8月2日; 录用日期: 2019年8月14日; 发布日期: 2019年8月23日

摘要

城镇化发展是衡量一个地区综合实力的重要因素,同时生态环境与人们的生产生活有着密不可分的关系。为研究城镇化发展对生态环境的影响,本文选取Z市相关数据,构建城镇化、生态环境指标体系,以城镇化发展相关指标为自变量,通过熵权法将生态环境有关的多个指标综合为因变量,运用R软件分析所获得的样本数据,通过逐步回归法建立多元回归方程。分析结果表明:人均生产总值等因素对生态环境正向影响显著,民用汽车拥有量对生态环境负向显著影响。基于上述分析结果,本文从人口、经济、社会等三方面对城镇化发展提出合理化意见,进而促进生态环境与城镇化和谐发展。

关键词

城镇化, 生态环境, 熵权法, 回归分析

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

对于生态环境指标体系构建,经济合作与发展组织(OECD)和联合国环境规划署(UNWFP)共同提出“压力-状态-响应”(PSR)模型,用于评价人类活动对生态环境的影响程度,是较为成熟的构建生态环境指标体系的模型,王国惠[1]根据此模型从生态环境水平、压力和治理三个方面选取了7个指标作为衡量生态环境的因子。郭海红[2]引入向量自回归模型结合脉冲响应函数及方差分解法,解析新型城镇化与生态环境之间的动态响应关系,得到中国的西部、中部、东部地区城镇化的确对于生态环境造成不同程度的压力,提出重视区域差异探索特色发展道路的建议。王国惠[1]对山西省的城镇化与生态环境关系进行研究,采用熵值法对城镇化指标、生态环境指标赋予权重,引入“协调度”指标,通过计算城镇化和生态环境的协调度,得到山西省新型城镇化与生态环境处于中低度协调状态的结论。基于此提出促进产业转型升级,发展生态工业、改善人民生活等建议。

“城镇化”是指农村人口不断向城镇转移,第二、三产业不断向城镇聚集,从而使城镇数量增加、规模扩大的一种历史过程。城镇化问题是我国经济社会发展的重大问题,是“十二五”和“十三五”期间的重要任务。城镇经济建设与发展,必须注意合理利用资源保护生态环境,坚持可持续发展的方针,使得城镇的经济、社会和环境协调发展[3]。近年来,城镇化建设推动了我国经济社会的持续发展,促进了城乡居民生活水平的全面提升。但是,在城镇化取得卓越成就的同时,许多生态环境问题接踵而至,遏制了城镇化进程[3]。本文旨在通过多元统计分析、建立回归模型,初步研究城镇化建设对生态环境的影响,在不影响城镇化进程的前提下,提出改善农村生态环境的合理建议。

2. 理论依据[4]

2.1. 多元回归模型

设随机变量 y 与自变量 x_1, x_2, \dots, x_p 的回归模型为:

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

式中 $\beta_0, \beta_1, \dots, \beta_p$ 是 $p+1$ 个未知参数, β_0 称之为回归常数, β_1, \dots, β_p 称之为回归系数, ε 是随机误差项。

2.2. 逐步回归

逐步回归的基本思想是有进有出, 具体做法是将变量一个一个的引入, 每引入一个变量后都要进行 F 检验, 并对已选入的变量要进行逐个的 t 检验, 当原引入的变量由于后面变量的引入而变得不再显著时, 要将其删除, 以确保每次引入新的变量之前回归方程只包含显著的变量。反复执行此过程, 直至既无显著的自变量选入回归方程, 也无不显著的自变量被剔除回归方程为止。

2.3. 熵权法的基本原理

熵权法的基本思想是根据指标变异性的确定来确定客观权重。一般来说, 若某个指标的信息熵越小, 表明指标值得变异程度越大, 提供的信息量越多, 其权重也就越大。相反, 某个指标的信息熵越大, 表明指标值得变异程度越小, 提供的信息量也越少, 其权重也就越小。

3. 模型构建与结果分析

3.1. 指标体系的选取

考虑数据的可获得性和真实性, 本文选取了《山东省 2005~2018 年统计年鉴》中的相关数据作为样本。通过阅读文献得知仅考虑城镇化率这一指标不能全面衡量济宁市的城镇化水平, 基于此本文选取了城镇化率、人均生产总值等八个指标来衡量城镇化水平的高低。我们根据联合国环境规划署提出的 PSR 模型以及王国惠[2]选取衡量生态环境的因子, 选取了耕种面积、工业废水排放总量等八个指标来衡量济宁市的生态环境水平。具体内容如表 1 所示。

3.2. 数据预处理

3.2.1. 数据标准化

数据的标准化是指对数据进行按比例缩放, 使其落入一个小的特定区间。在多指标评价体系中, 由于各评价指标的量级不同, 当各指标间的水平相差很大时, 若直接利用原始指标值进行分析, 就会突出数值水平较高的指标在综合分析中的作用, 降低数值水平较低指标的作用。因此, 本文为了保证结果的可靠性, 分别对自变量和因变量的原始数据进行标准化处理。

3.2.2. 数据差分处理

对本文中 x_1, x_2, \dots, x_8 是一组时间序列数据, 在绝大多数情况下, 彼此之间不是独立的, 而是相关的。本文采取差分法降低自变量 x_1, x_2, \dots, x_8 之间的相关性。由图 1 所示, 在差分前, x_1, x_2, \dots, x_8 之间具有较强的相关性, 且自变量和因变量 y 之间的相关性较低; 在差分后, x_1, x_2, \dots, x_8 之间的相关性显著降低, 且一定程度上提高了 y 与 x_1, x_2, \dots, x_8 的相关性。

Table 1. Urbanization and ecological environment index system
表 1. 城镇化与生态环境的指标体系

城镇化	x_1	城镇化率(%)
	x_2	人均生产总值(元/人)
	x_3	城镇居民人均可支配收入(元/人)
	x_4	第一产业占 GDP 比重(%)
	x_5	第三产业占 GDP 比重(%)
	x_6	执业医师(人)
	x_7	民用汽车拥有量(辆)
	x_8	人均财政教育支出(元)
生态环境	y_1	人均水资源拥有量(立方米/人) (+)
	y_2	耕种面积(平方米/人) (+)
	y_3	森林覆盖率(%) (+)
	y_4	实际用水量(亿立方米) (+)
	y_5	人均工业废水排放总量(吨/人) (-)
	y_6	人均二氧化硫(吨/人) (-)
	y_7	人均工业烟尘(吨/人) (-)
	y_8	人均用电量(千瓦时/人) (+)

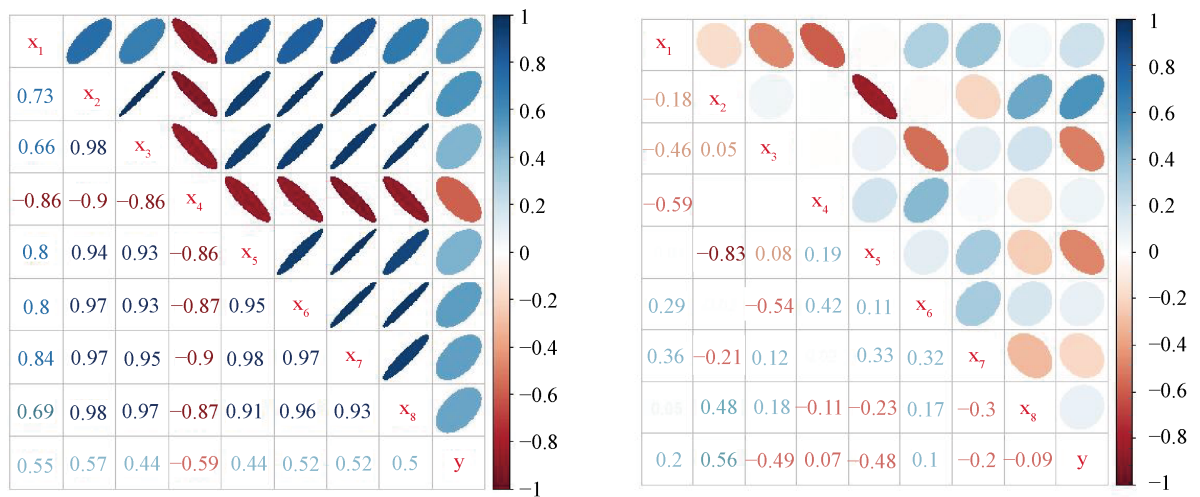


Figure 1. Comparison of correlation coefficients before and after differential data

图 1. 数据差分之前与差分之后的相关系数对比

3.3. 因变量综合指标的提取

本文通过熵权法对生态环境各指标数据进行综合指标提取，提取步骤如下所示：

第一步：将各个指标的数据进行标准化处理。

$$y_{ij} = \frac{x_{ij} - \min\{x_{1j}, \dots, x_{nj}\}}{\max\{x_{1j}, \dots, x_{nj}\} - \min\{x_{1j}, \dots, x_{nj}\}}$$

第二步：计算第 j 项指标下第 i 个样本值占该指标的比重。

$$p_{ij} = \frac{y_{ij}}{\sum_{i=1}^n y_{ij}}, i = 1, \dots, n, j = 1, \dots, m$$

第三步：计算第 j 项指标的熵值。

$$E_j = -\ln(n)^{-1} \sum_{i=1}^n p_{ij} \ln(p_{ij}), j = 1, \dots, m$$

第四步：计算信息熵冗余度。

$$d_j = 1 - E_j, j = 1, \dots, m$$

第五步：计算各项指标的权重。

$$\omega_j = \frac{d_j}{\sum_{j=1}^m d_j}, j = 1, \dots, m$$

权重计算结果如表2所示：

Table 2. Ecological environment indicators weight
表2. 生态环境各指标权重数据

因变量	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
权重	0.157	0.324	0.003	0.158	0.144	0.164	0.033	0.018

由表 2 可知， y_2 的权重最大，说明 y_2 对生态环境整体水平的影响最大， y_1, y_4, y_5, y_6 的变动对生态环境整体水平影响较大，其他指标影响较小。

3.4. 全模型

经过数据预处理，对所得综合指标因变量 y 和自变量 $\Delta x_1, \Delta x_2, \dots, \Delta x_8$ 进行多元回归分析，所得各自变量系数如下表所示：

Table 3. Full model coefficient and Variance expansion factor
表 3. 全模型的系数和方差扩大因子

	Estimate	Std. Error	t value	Pr(> t)	方差扩大因子
(Intercept)	-3.911242211	1.440375578	-2.71543219	0.113076009	
Δx_1	1.364738478	0.424629585	3.213950529	0.084692876	12.704138
Δx_2	17.09360859	5.402977195	3.163738801	0.087058812	23.600796
Δx_3	0.611596059	1.171364779	0.52212263	0.653654248	14.552557
Δx_4	0.553017396	0.228246041	2.422900275	0.136353784	4.664023
Δx_5	1.960100152	0.847352868	2.313204129	0.146814812	16.545492
Δx_6	-0.252335	0.9051865	-0.278765757	0.806604251	11.98105
Δx_7	-3.472180043	1.856520162	-1.870262501	0.202363781	15.989522
Δx_8	-2.508926833	1.430263046	-1.754171612	0.22149134	16.120982

根据表 3 可以得到标准化后全模型多元线性回归方程为:

$$\Delta y = -3.911 + 1.365\Delta x_1 + 17.094\Delta x_2 + 0.612\Delta x_3 + 0.553\Delta x_4 + 1.960\Delta x_5 - 0.252\Delta x_6 - 3.472\Delta x_7 - 2.509\Delta x_8。$$

由表 3 中回归系数的显著性检验结果,可以看出该回归方程并不是十分理想。取显著性水平 $\alpha = 0.05$ 时,我们无法认为因变量 y 与自变量 $\Delta x_1, \Delta x_2, \dots, \Delta x_8$ 之间有显著的线性关系。因此,考虑采用逐步回归的方式逐步筛选变量,进行多元线性模型的构建。

3.5. 逐步回归

利用 R 软件对经过预处理的因变量 y 和自变量 $\Delta x_1, \Delta x_2, \dots, \Delta x_8$ 进行逐步回归,所得各自变量系数如下表所示:

Table 4. Select model coefficient and variance expansion factor

表 4. 选模型的系数和方差扩大因子

	Estimate	Std. Error	t value	Pr(> t)	方差扩大因子
(Intercept)	-3.456873363	0.881190579	-3.922957697	0.017204743	
Δx_1	0.960781781	0.200601532	4.789503696	0.008714723	2.687793
Δx_2	15.07549123	3.306996401	4.55866575	0.010349307	8.381667
Δx_4	0.333707354	0.143297555	2.328772141	0.080360119	1.742747
Δx_5	1.680025167	0.571728796	2.938500173	0.042452513	7.140585
Δx_7	-2.734745521	0.709338868	-3.855344244	0.01821938	2.212821
Δx_8	-2.092409822	0.570347141	-3.668660142	0.021416376	2.430193

基于表 4 得到多元回归方程为:

$$\Delta y = -3.457 + 0.961\Delta x_1 + 15.075\Delta x_2 + 0.334\Delta x_4 + 1.680\Delta x_5 - 2.735\Delta x_7 - 2.092\Delta x_8。$$

首先,在给定显著性水平 $\alpha = 0.1$ 的条件下,所有自变量均通过显著性检验,表明 $\Delta x_1, \Delta x_2, \Delta x_4, \Delta x_5, \Delta x_7, \Delta x_8$ 对 y 均有显著影响。从回归方程可以看出, $\Delta x_1, \Delta x_2, \Delta x_4, \Delta x_5$ 的回归系数都为正,即城镇化率越高,也就是人口向城市集中的比率越高,污染扩散率降低,致使生态环境状况越好;人均生产总值、第一产业和第三产业占 GDP 比重对生态环境均有正向的影响。其次, R^2 是 0.9096,调整后 R^2 为 0.7741,回归方程的拟合效果较好。最后,依次计算选模型中各个自变量的方差扩大因子,可得所有 VIF 均小于 10,表明该选模型的自变量之间不受多重共线性的影响。

4. 相关建议

表从人口方面来讲。首先要控制农村人口向城镇的转移,合理安排城镇人口结构,城镇化率提高可以改善生态环境情况,但若一味地追求城镇化速率,忽视城镇所能承受的最大人口数,容易产生环境污染、交通堵塞等一系列问题。其次,城镇化不是一蹴而就,政府要采取有效措施,合理指引农业人口有序变为非农业人口,逐步改善农民生活条件,提高农民的文化素养,一步步实现城镇化。

从经济方面来讲。应促进社会经济发展,提高人均可支配收入,增大对生态环境保护工作的资金投入。着重发展第三产业,提高第三产业占 GDP 比重,对第一产业及第二产业进行合理规划与布局,减少对生态环境的负面影响。将生态文明建设融入到城镇化进程中,实现生态环境与经济的和谐发展。

从社会方面来讲。首先,政府应大力推广市民使用环保型、新能源等车种,同时大力倡导市民使用

公共交通工具出行,最大程度减少机动车尾气污染。其次,在道路两侧建立生态隔离带和绿化带,以充分发挥植物对大气的净化作用。最后,人的环保意识是生态环境保护中最最重要的因素,政府和相关部门应结合当地经济和社会发展实际,大力加强教育、宣传和行为引导力度,让市民正确认识到保护环境的意义和作用。

致 谢

论文最后感谢山东省自然科学基金和曲阜师范大学大学生创新训练计划项目的资助,感谢胡锋老师的指导和审稿人提出的建议。

基金项目

山东省自然科学基金(批准号:ZR2017MA012);曲阜师范大学大学生创新训练计划项目(批准号:2018A152)。

参考文献

- [1] 王国惠,赵新燕,黄永胜. 新型城镇化与生态环境协调发展关系探究[J]. 经济问题, 2018(3): 112-117.
- [2] 郭海红,张在旭. 新型城镇化与生态环境响应度的区域差异研究[J]. 宁夏社会科学, 2018(6): 86-96.
- [3] 史亚楠,佟光霁. 浅析城镇化建设中生态环境保护的问题及对策[J]. 农业经济, 2013(8): 45-47.
- [4] 何晓群,刘文卿. 应用回归分析[M]. 第4版. 北京:中国人民大学出版社, 2015: 132-157.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询。
2. 通过知网首页 <http://cnki.net/> 顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org