

2019 数维杯大学生数学建模竞赛论文

论文题目：中国省际生态环境与经济交互状况综合评价研究

队伍编号：A20198001

中国省际生态环境与经济交互状况综合评价研究

摘 要：

本文讨论了我国各省份环境与经济交互状况的影响因素问题，选取了影响环境与经济发展具有代表性的指标，并且通过建立指标模型，来分析和研究经济与生态环境的影响因素及提高生态质量的有效措施。

问题一中，进行数据归一化处理，得出影响经济与生态环境的主要因素有生产总值GDP、人口、废气、废水、废渣等，根据各影响因素的内容和表现形式构造模糊综合评价指标，以便更好地对经济和生态环境的影响指标进行综合评价，从而验证指标的准确性，进而完成各省综合排序。

问题二中，基于问题一的研究，采用主成分分析法得出对经济发展具有影响力的关键指标，接着引用多元线性回归方法对经济和生态环境与各主要影响因素进行拟合，验证其结果的准确度，根据模型结论，我们给予了一些合理性建议。

问题三中，通过查询中国统计年鉴，对相关指标数据重新进行量化处理，采用因子分析法对目标因素进行综合评价，得出综合得分，并对近几年影响经济与生态环境发展的问题提出合理化措施。

关键词：归一化处理；模糊综合评价；主成分分析；多元线性回归；因子分析

一、问题重述

1.1 背景分析

近年来，国民经济迅速发展，并取得了趋势瞩目的成就。经济的快速发展伴随着工业化和城镇化进程的快速推进。伴随城镇规模体系扩大、产业趋于集聚发展、形成交通运输网络、基础设施建设加快和公共服务水平提高的为城市群区域人口的过度集聚、地域空间的无序扩展、经济扩张的不断加速、文明影响的范围不断延伸，使得城市群资源环境问题高度集中且激化，资源环境瓶颈制约日益加剧。在长期粗放式的发展模式下，生态脆弱性增加、生态恶化和环境污染问题掣肘现代化进程、制约经济社会发展、影响人民生活质量。推进生态文明建设是破解资源高效利用，减少环境污染，提升国家长远可持续发展能力。在推进生态文明建设这项庞大的系统工程中，我国当前最突出的短板是缺乏系统性的评价与整体协协作治理机制，这将会严重影响我国生态文明建设的成效。

1.2 问题重述

问题一：站在自然的角度去观察题目给出的数据，可以判定废水、废气、废渣的数值越小越好。其次，对于数据进行标准化处理，无法统计的或者是不完全的数据或者指标进行有效的删除，接着对标准化处理后的数据进行 A、B、C、D、E、F 的等级划分，从总体把握各个城市的情况。进行合理的评价时算法选用模糊评价法。

问题二：通过建立有效合理的数学模型，评价我国 2017 年各省经济处于持续发展中，同时对各省得经济与生态环境问题得异同做出分析比较，采取针对性的措施扬长避短来减少对经济和生态环境的影响，达到各省生态环境协调治理的目的。

问题三：国家统计局年鉴能够全面的反应我国经济和社会环境交互发展情况，对年鉴数据重新进行指标的选取和模型的建立和分析，并更深一步地解决问题一和问题二。

二、问题分析

对于问题一，首先我们以数据为基础，对数据进行了量化处理，反映出各影响因素的影响情况，以图表的形式表达作用规律，然后运用分类处理的方法对影响生态环境与经济交互状况的各个因素进行具体评价分析，最后建立模糊评价法模型验证其精确性，从而完成综合排序。

对于问题二，以模型一为基础，首先在明确主要影响因素的情况下，使用主成分分析法选择出贡献经济发展的最为关键的指标，同时将结果与问题一中得到的各主要指标进行比较，可检验得到的主要指标是否一致，从而对提取出的各主要指标进行验证。然后利用多元线性回归方法对经济和生态环境与各主要影响因素进行拟合，验证结果的准确度，最后论述各省之间生态环境协作治理的具体措施。

对于问题三：首先查询中国统计局官网得到近 9 年来影响经济与生态环境发展的 12 项重要指标的具体数据，通过数据的无量纲化对数据进行预处理，然后建立利用因子分析模型选取主要指标，对经济与生态环境协调发展的主要影响因子进行的综合评价，得出综合得分，进而对问题一和问题二重新进行了解答，最后对影响经济与生态环境发展提出合理有效的建议

三、模型假设

1. 假设从官方获取的数据信息来源可靠，真实准确；
2. 假设数据具有代表性；

- 3. 假设影响生态环境与经济交互的各因素已全部给出；
- 4. 假设在仿真模型中，未来几年没有重大自然突变以及国际政策变革；

四、符号说明

为了便于问题的求解，我们给出以下符号说明：

max	最大值
min	最小值
U	评判因素集合
F	主成分
A	正交矩阵
E	单位矩阵
$X=\{X_1, X_2, ..., X_i\}$	条件属性集
D	决策属性集
e_i	单位特征向量
P	原始指标个数
X_i	第 i 个原始指标
F_i	第 i 个主成分
K	所提取主成分个数
R	X 的相关系数矩阵
λ_i	R 的特征值
α_i	λ_i 对应的特征向量
ε	随机误差
Cov	协方差
F	因子变量
A	因子载荷矩阵
a_{ij}	因子载荷
β	特殊因子

五、模型的建立与求解

5.1 各省生态环境与经济交互状况的评价与排序

5.1.1 数据预处理

不同评价指标往往具有不同的量纲和量纲单位，这样的情况会影响到数据分析的结果，为了消除指标之间的量纲影响，需要进行数据标准化处理，以解决数据指标之间的可比性。原始数据经过数据标准化处理后，各指标处于同一数量级，适合进行综合对比评价。归一化化就是要把你需要处理的数据经过处理后（通过某种算法）限制在你需要的一定范围内。首先归一化是为了后面数据处理的方便，其次是保证程序运行时收敛加快。

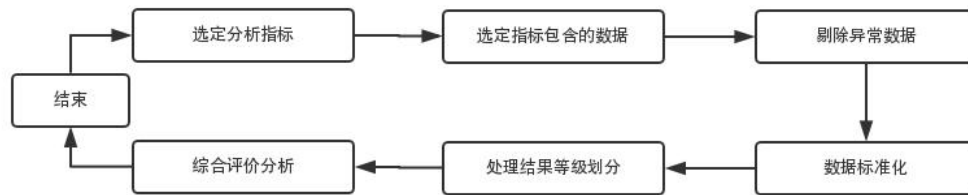


图 1 数据处理流程

以提供数据为样本进行建模，在进行数据预处理，对提供的海量数据进行判断、整理、挖掘与分析，剔除部分异常数据，定量与定性分析相结合。然后进行数据标准化，也叫离差标准化，对预处理的数据进行线性变换，使结果落到[0, 1]区间，原始数据均转换为无量纲化，数据无量纲化的处理主要解决数据的可比性，即不同量级的数据通过标准化变为同一量级的数据，各指标值都处于同一个数量级别上，能够进行综合测评分析。

转换函数如下：

def Normalization(x):

return [(float(i)-min(x))/float(max(x)-min(x)) for i in x]

其中 max 为样本数据的最大值，min 为样本数据的最小值，x 表示原始数据。这种方法有一个缺陷就是当有新数据加入时，可能导致 max 和 min 的变化，需要重新定义。其原理为，将某一问项的原始值 x 通过标准化映射成在区间[0, 1]中的值，其公式为：新数据=（原数据-极小值）/（极大值-极小值），也称为离差标准化，是对原始数据的线性变换，使结果值映射到[0 - 1]之间。

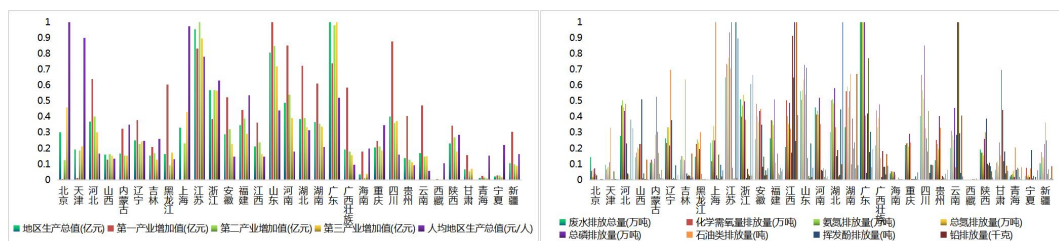
数据标准化处理的新数映射到[0 - 1]之间，但是每个作用因素在各个城市层面的反映不够直观，所以，接着对标准化处理后的数据进行等级划分。

表 1 等级划分明细

因素	等级					
	A	B	C	D	E	F
人口	(0.4-0.6)	(0.3-0.4]、[0.6-0.7)	(0.2-0.3]、[0.7-0.8)	(0.1-0.2]、[0.8-0.9)	(0-0.11]、[0.9-1)	0、1
GDP	1	[0.8-1)	[0.5-0.8)	[0.2-0.5)	(0-0.2)	0
废水	0	(0-0.2)	[0.2-0.5)	[0.5-0.8)	[0.8-1)	1
废气	0	(0-0.2)	[0.2-0.5)	[0.5-0.8)	[0.8-1)	1
废渣	0	(0-0.2)	[0.2-0.5)	[0.5-0.8)	[0.8-1)	1
财政收入	1	[0.8-1)	[0.5-0.8)	[0.2-0.5)	(0-0.2)	0
财政支出	1	[0.8-1)	[0.5-0.8)	[0.2-0.5)	(0-0.2)	0
居民收支	1	[0.8-1)	[0.5-0.8)	[0.2-0.5)	(0-0.2)	0

5.1.2 数据预处理结果

对 2017 年各省生态环境与经济交互状况的影响因素废气、废水、废渣、GDP、人口、财政收入、财政支出、全体及城乡居民收支基本情况预处理，作用因素数据标准化处理后的效果图如下，数据标准化处理值及等级划分结果见附表。



(1) GDP

(2) 废水

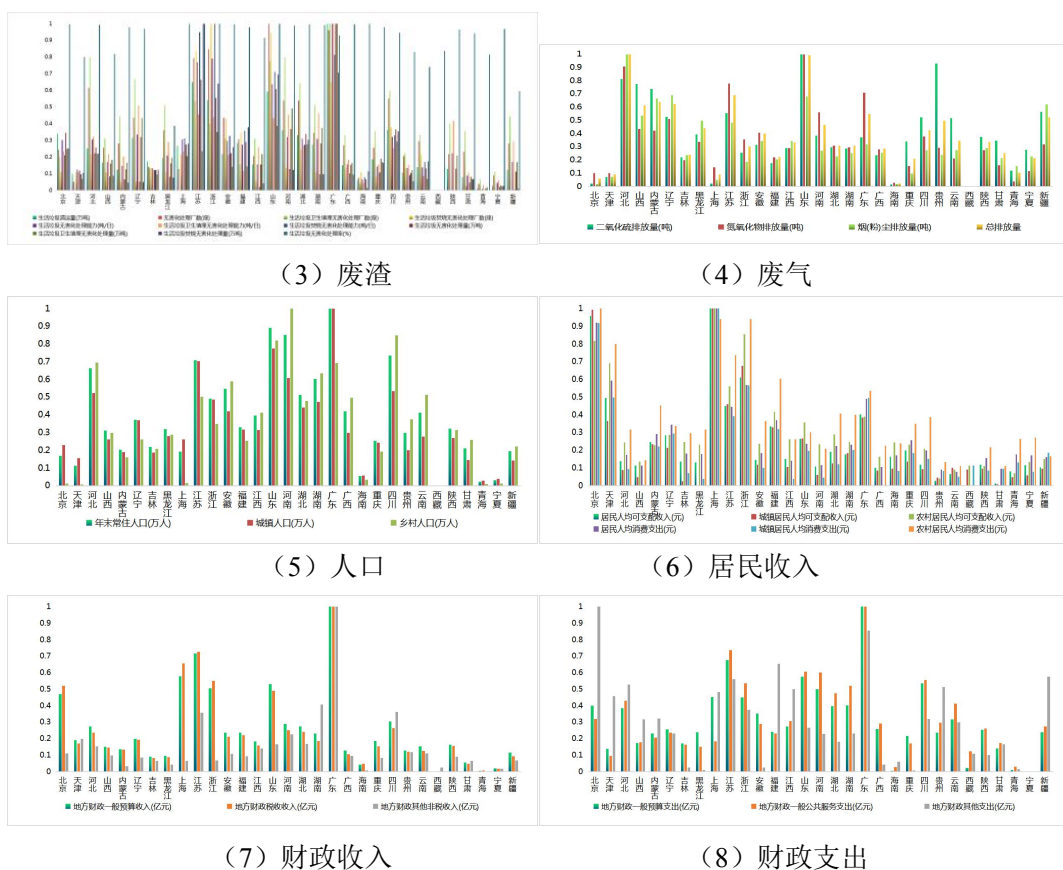


图 2 相关数值标准化处理结果

5.1.3 模糊综合评价

在复杂大系统中，需要考虑的因素往往是很多的，而且因素之间还存在着不同的层次。这时，应用单层次模糊综合评判模型就很难得出正确的评判结果^[1]。所以，在这种情况下，就需要将评判因素集合按照某种属性分成几类，先对每一类进行综合评判，然后再对各类评判结果进行类之间的高层次综合评判。这样，就产生了多层次模糊综合评判问题。

多层次模糊综合评判模型的建立^[2]，可按以下步骤进行：

第一步：对评判因素集合 U ，按某个属性，将其划分成 m 个子集，使它们满足：

$$\begin{cases} \sum_{i=1}^m U_i = U \\ U_i \cap U_j = \Phi (i \neq j) \end{cases} \quad (1)$$

这样，就得到了第二级评判因素集合：

$$U = \{U_1, U_2, \dots, U_m\} \quad (2)$$

在上式中， $U_i = \{U_{ik}\} (i=1, 2, \dots, m; k=1, 2, \dots, nk)$ 表示子集 U_i 中含有 nk 个评判因素。

第二步：对于每一个子集 U_i 中的 nk 个评判因素，按单层次模糊综合评判模型进行评判，如果 U_i 中的诸因数的权数分配为 A_i ，其评判决策矩阵为 R_i ，则得到第 i 个子集 U_i 的综合评判结果：

$$B_i = A_i \times R_i = [b_{i1}, b_{i2}, \dots, b_{im}] \quad (3)$$

第三步：对 U 中的 m 个评判因素子集 $U_i (i=1, 2, \dots, m)$ ，进行综合评判，其评判决策矩阵为：

$$R = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_m \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} \quad (4)$$

如果 U 中的各因子子集的权数分配为 A，则可得综合评判结果：

$$B^* = A \times R$$

在式中 B^* 既是 U 的综合评判结果，也是 U 中的所有评判因数的综合评判结果。若 U 中仍含有很多因素，则可以对它再进行划分，得到三级以至更多层次的模糊综合评判模型。多层次的模糊综合评判模型，不仅可以反映评判因素的不同层次，而且避免了由于因素过多而难于分配权重的问题。在进行模糊综合评价方法结果分析时，由于各指标评价因子的原始数据来源不尽相同，因此需要对原始数据进行规范化处理，采用的方法是原始数据除以对应指标的平均值。

当指标值越大越好时

$$P_i = 1 - (S_i - P_i)/(S_i - P_{min}) \quad (5)$$

当指标值越小越好时

$$P_i = 1 - (P_i - S_i)/(P_{max} - S_i) \quad (6)$$

上海市交通出行方式综合得分(F_i)是根据它所包含的因子层在经始数据规范化处理后的数值乘以各自的权重后进行加总，计算公式为：

$$F_i = \sum_{i=1}^N a_i V_i \quad (7)$$

其中， F_i 为综合得分， a_i 为某因子经层次总排序后的权重， V_i 为经标准化处理后的指标数据， i 表示某个城市， n 为评价指标的总数。综合评价过后，进行赋值计算综合得分，令A=6、B=5、C=4、D=3、E=3、F=1，得到最终结果如下。

表 2 综合评价得分结果

城市	北京	天津	河北	山西	内蒙古	辽宁	吉林	黑龙江
人口	9	8	16	13	10	14	11	13
GDP	16	14	15	10	12	15	12	12
废水	52	50	44	46	46	45	48	48
废气	12	12	6	13	13	12	15	16
废渣	43	47	39	49	49	43	50	49
财政收入	9	6	8	6	6	6	6	6
财政支出	12	7	9	7	9	9	6	7
居民收支	31	21	14	12	18	18	14	13
城市	上海	江苏	浙江	安徽	福建	江西	山东	河南
人口	9	14	17	18	14	16	10	9
GDP	15	25	19	15	16	13	23	17
废水	44	32	42	46	47	32	41	46
废气	12	13	15	16	15	16	7	15
废渣	38	29	29	43	45	48	30	42
财政收入	10	11	10	8	8	6	9	9
财政支出	8	11	10	8	9	9	11	11
居民收支	35	20	26	14	19	14	17	14
城市	湖北	湖南	广东	广西	海南	重庆	四川	贵州
人口	18	16	7	16	6	11	13	13
GDP	16	16	25	12	10	14	16	11
废水	41	41	37	48	52	47	42	48
废气	16	16	14	16	12	14	15	14
废渣	40	44	18	50	51	50	40	50
财政收入	8	8	18	6	6	6	9	6
财政支出	8	10	15	8	6	7	11	9
居民收支	15	16	19	12	14	15	14	12

城市	云南	西藏	陕西	甘肃	青海	宁夏	新疆	
人口	16	3	14	11	6	6	10	
GDP	11	7	14	9	10	11	11	
废水	37	44	47	46	50	50	47	
废气	15	24	16	15	12	15	13	
废渣	50	2	32	50	37	47	45	
财政收入	6	4	6	6	5	6	6	
财政支出	9	6	8	6	6	3	9	
居民收支	12	9	13	11	13	13	12	

5.2 生态环境协作治理

5.2.1 主成分分析法原理的阐述

根据题意要求和相应的分析，对于问题一本文采用主成分分析方法。数学上的处理就是将原来 p 个指标作线性组合，作为新的综合指标，如果将选取的第一个线性组合即第一个综合指标记为 F_1 ，这里的“信息”用 F_1 的方差来表达， $\text{Var}(F_1)$ 越大，表示 F_1 包含的信息越多。因此在所有的线性组合中所选取的 F_1 应该是方差最大的，故称为第一主成分。如果第一主成分不足以代表原来 p 个指标的信息，再考虑选取即选第二个线性组合，为了有效反映原来信息^[4]， F_1 已有的信息就不需要出现在 F_1 中，用数学语言表达就是要求 $\text{Cov}(F_1, F_2)$ ，称 F_2 为第二主成分，依此类推可以构造出第三，四，...，第 p 个主成分。

通过题目所给的数据，得到如下表所示的影响经济与生态环境的指标体系（见表5-1-1）。并对各指标的统计数据进行实证分析。

表3 影响经济与生态环境的指标体系

指标	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
意义	GDP	人口	废气	废水	废渣	地方财政支出	地方财政收入	全体及城乡居民收支基本情况

5.2.2 主成分分析法的数学模型及实证分析结果

设有样本容量为 n 的 p 个变量，通过变换将原变量 X_i 转换成主成分（用 F 表示），主成分是原变量的线性组合，且具有正交特征，即将 X_1, X_2, \dots, X_p 综合成 $k(k \leq p)$ 个变量 F_1, F_2, \dots, F_k ，可用多项式表示：

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots\dots\dots \\ F_k = a_{1k}X_1 + a_{2k}X_2 + \dots + a_{pk}X_p \end{cases} \text{或 } F = AX \quad (8)$$

矩阵 A 满足 $AA^T = E$ ，即 A 为正交矩阵，其中 E 为单位阵，且 a_{ij} 由下列原则决定：

- (1) F_i 与 F_j ($i \neq j$) 不相关；
- (2) $\text{Var}(F_1) > \text{Var}(F_2) > \dots > \text{Var}(F_k)$

主成分分析法步骤如下：

(1) 假定输入一个决策表 $T=(U, C \cup D, F)$ ，其中 U 为论域， $X=\{X_1, X_2, \dots, X_i\}$ 为条件属性集， D 为决策属性集。需输出条件属性的主成分 $F=\{F_1, F_2, \dots, F_k\}$ 。

(2) 按 $X_{ij} = \frac{X_{ij} - \bar{X}_j}{\sqrt{\text{Var}(X_j)}}$ 对原始数据 $X=\{X_1, X_2, \dots, X_p\}$ 进行标准化处理，使

每个属性均值为0，方差为1。在SPSS16.0软件运行下得到描述性统计分析。

表4 各指标的描述性统计分析

	Mean	Std. Deviation	N
废气（吨）	1.01527	.019847	22055

废水	7625.40	3400.748	22055
废渣	37.554	1.1961	22055
GDP（亿元）	12445.361	6275.8115	22055
人口（万人）	243.7262	58.71080	22055
地方财政支出（亿元）	.8146	.02935	22055
地方财政收入（亿元）	6.18917	.477327	22055

(3) 根据概率论中的相关系数公式： $\text{Cov}(X_i, X_j) = E[(X_i - E(X_i)) - (X_j - E(X_j))]$ 计算原始数据集 $X=\{X_1, X_2, \dots, X_p\}$ 的相关系数矩阵 R 。在SPSS16.0软件运行下得到 X 的相关系数矩阵 R （见表5-1-3）。

表 5 X 的相关系数矩阵 R

	人口	GDP	废气	废水	废渣
人口	1.000	.074	.398	.132	-.365
GDP	.074	1.000	-.019	.909	-.121
废气	.398	-.019	1.000	.258	-.085
废水	-.132	.909	.258	1.000	.062
废渣	-.365	-.121	-.085	.062	1.000

(4) 计算相关系数矩阵 R 的特征值 λ_i 及其对应的单位特征向量 e_i , $i=1,2,\dots,k$,并将特征值按由大到小的顺序排列, 即 $\lambda_1 > \lambda_2 > \dots > \lambda_k$ 。

(5) 计算主成分的方差贡献率和累计方差贡献率。第 k 个主成分方差为 $\alpha_i = \lambda_i / \sum_{i=1}^k \lambda_i$, 主成分 F_1, F_2, \dots, F_k 的累计方差贡献率为

$$\frac{\left(\sum_{i=1}^k \lambda_i\right)}{\left(\sum_{j=1}^p \lambda_j\right)} \quad (5-1-2) \quad (9)$$

其中 α_1 的值最大, 则说明 F_1 综合 X_1, X_2, \dots, X_p 信息的能力最强, 主成分 k 值得选取一般为使得累计方差贡献率 $\geq 80\%$ （或特征值大于1）的前 k 个特征值。但本文中为了使信息尽量不流失, 选取累计方差贡献率达到 $\geq 99\%$ 的前 k 个特征值。

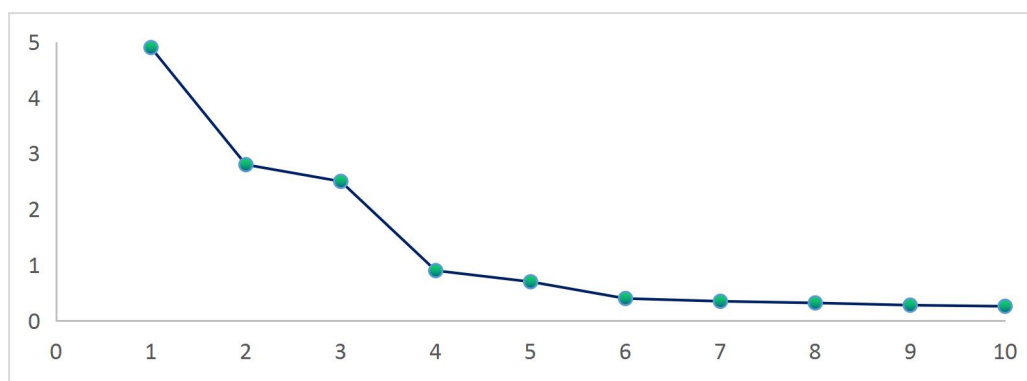


图3 主成分散点图

(6) 利用前 k 个特征值对应的单位向量 $e_1 = (e_{11}, e_{12}, \dots, e_{1p})'$, $e_2 = (e_{21}, e_{22}, \dots, e_{2p})'$, \dots , $e_p = (e_{k1}, e_{k2}, \dots, e_{kp})'$, 按 $F = AX$ 计算原始数据的主成分 F_1, \dots, F_k 。

在SPSS16.0软件运行下, 我们最终得到特征值与方差贡献率（见表5-1-4）, 以及旋

转前的因子载荷矩阵(见表5-1-5)

表 6 特征值与方差贡献表

成分	初始化自我评估			平方载荷的提取和		
	总计	% 变化	累积 %	总计	方差 %	累积 %
1	5.550	55.500	55.500	5550	55.500	55.500
2	1.350	13.498	68.998	1.350	13.498	68.998
3	1.219	12.186	84.185	1.619	16.186	84.185
4	.835	8.352	92.536	.835	8.352	92.536
5	.583	5.828	98.364	.583	5.828	98.364
6	.095	.953	99.318			
7	.042	.423	99.741			
8	.000	.003	100.000			

如上表所示，本文最终提取到影响经济与生态发展的5个主成分，并且使累计方差贡献率达到了 $\geq 99\%$ ，从而保证影响房价的有效信息不流失，为更好地解决后面的问题做准备。

表 7 旋转前的因子载荷矩阵

	成分					
	1	2	3	4	5	6
人口（万人）	-.065	.708	-.448	.479	.365	-.131
GDP（亿元）	.874	.183	-.034	.076	-.072	.035
废气	-.251	.876	.369	.006	-.075	.148
废水	.854	.400	.300	.110	-.067	-.032
废渣	.016	-.367	.746	.092	.550	.022

表 8 各指标对生态环境与经济的总贡献率

	1	2	3	4	5	6	各指标的总贡献率
人口（万人）	.065	.708	-.448	.479	.365	-.131	0.52
GDP（亿元）	.874	.183	-.034	.076	-.072	.035	0.28
废气	-.251	.876	.369	.006	-.075	.148	0.18
废水	-.854	.400	.300	.110	-.067	-.032	0.12
废渣	-.016	-.367	.746	.092	.550	.022	0.10

得到载荷矩阵后，本文根据各指标分别对各主成分的贡献率加权（权指个主成分所表示的信息量）平均之后我们可以得到各指标对经济和生态环境的总贡献率（见表 5-1-6），对经济和生态环境影响最主要的五个因素依次是： X_1 GDP， X_2 人口， X_3 废气， X_4 废水， X_5 废渣。

5.2.3 利用逐步回归模型，验证所提取主成分的正确性

1.多元线性回归数学模型

多元线性回归模型是指含有多个解释变量的线性回归模型，用于解释被解释变量与其他多个变量之间的线性关系，其数学模型为：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (10)$$

式(10)表示一个p元线性回归模型,其中有p个解释变量。表明被解释变量y的变化可由两部分组成:第一,由p个解释变量x的变化引起的y的线性变化部分,即 $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$;第二,由其他随机因素引起的y的变化部分,即 $\varepsilon \sim N(0, \sigma^2)$ 部分,叫随机误差。 $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ 都是模型中的未知参数,分别为回归常数和偏回归系数^[3]。

对y和 $x_1, x_2, \dots, x_{p-1}, x_p$,分别进行n次独立观测,取得n组数据样本 $y_i, x_{i1}, x_{i2}, \dots, x_{i(p-1)}$ ($i = 1, 2, 3, \dots, n$) 则有:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1(p-1)} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2(p-1)} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{n(p-1)} + \varepsilon_n \end{cases} \quad (11)$$

其中 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ 相互独立,且服从 $N(0, \sigma^2)$ 分布。令:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad (12)$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & x_{22} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n(p-1)} \end{bmatrix} \quad (13)$$

则式(5-2-2)用矩阵形式表示为:

$$Y = X\beta + \varepsilon \quad \text{其中 } \varepsilon \sim N(0, \sigma^2 I_n) \quad (14)$$

模型参 β 的最小二乘法估计与误差方差 σ^2 的估计 β 的最小二乘法估计即选择 β 使误差项的平方和为最小值 这时 β 的值 $\hat{\beta}$ 作为 β 的点估计。

$$S(\beta) = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \quad (15)$$

为了求 β ,由(7)式将 $S(\beta)$ 对 β 求导,并令其为零,得:

$$\frac{ds}{d\beta} = \frac{d(Y - X\beta)^T (Y - X\beta)}{d\beta} = \frac{d(Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta)}{d\beta} = 0 \quad (16)$$

由(5-2-5)、(5-2-6)式可解出 $\hat{\beta}$:

$$\hat{\beta} = ((X^T X)^{-1} (X^T Y)) \quad (17)$$

对残差向量:

$$\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta} = [I - X(X^T X)^{-1} X^T] \quad (18)$$

对残差平方和:

$$\hat{\varepsilon}^T \hat{\varepsilon} = \hat{\varepsilon}^T [I - X(X^T X)^{-1} X^T] y = y^T y - \hat{\beta}^T X^T Y \quad (19)$$

$$E(y) = X\beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (20)$$

$$E\left(\hat{\varepsilon}^T \hat{\varepsilon}\right) = \sigma^2(n-p) \Rightarrow \sigma^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n-p} \quad (21)$$

多元线性回归数学模型建立后,是否与实际数据有较好的拟合度,其模型线性关系的显著性如何等,还需通过数理统计进行检验。

$$R = \sqrt{1 - \frac{\sum_{i=1}^n \left(y_i - \hat{y}_i \right)^2}{\sum_{i=1}^n \left(y_i - \bar{y}_i \right)^2}} \quad (22)$$

R 是复相关系数,用于测定回归模型的拟合优度,R 越大,说明y 与 $x_1, x_2, \dots, x_{p-1}, x_p$ 的线性关系越显著,为 \bar{y}_i 的平均值,R 取值范围为 $0 < |R| \leq 1$ 。

通回归系数得出拟合线性关系为:

$$Y = 5325.453 + 0.843X_1 + 1.719X_2 - 0.028X_3 - 0.822X_4 - 0.564X_5 \quad (23)$$

表 9 回归系数表

模型	非标准化系数		标准化系数	t	相关性		
	B	误差	测试		零阶	部分	部分
常数	5325.453	42.650		122.231			
人口（万人）	.843	.011	.129	75.310	.738	.452	.054
GDP（亿元）	1.719	.008	1.780	215.917	.927	.824	.155
废气	-.028	.002	.054	16.057	.903	.108	.011
废水	-.822	2.324	-.425	-119.990	.697	-.629	-.086
废渣	-.564	.039	-.645	-90.468	.828	-.520	-.065

表 10 数据量化表

省份	人口	GDP	废气	废水	废渣	省份	人口	GDP	废气	废水	废渣
北京	0.1693	0.3021	0.0225	0.1440	0.3380	湖北	0.5137	0.3865	0.2869	0.3035	0.3316
天津	0.1126	0.1950	0.0709	0.0955	0.1002	湖南	0.6021	0.3687	0.2869	0.3353	0.2765
河北	0.6631	0.1950	0.8141	0.2817	0.2513	广东	1	1	0.3715	1	1
山西	0.3106	0.1608	0.7742	0.1461	0.1665	广西	0.4198	0.1947	0.2363	0.2182	0.1508
内蒙古	0.2023	0.1672	0.7378	0.1109	0.1242	海南	0.0543	0.0356	0.0146	0.0421	0.0641
辽宁	0.3722	0.2499	0.5250	0.2638	0.3148	重庆	0.2527	0.2049	0.3397	0.2211	0.1859
吉林	0.2197	0.1542	0.2210	0.1306	0.1726	四川	0.7353	0.4035	0.5242	0.4060	0.3631
黑龙江	0.3186	0.1650	0.3944	0.1496	0.1950	贵州	0.2993	0.1383	0.9297	0.1266	0.1066
上海	0.1921	0.3317	0.0204	0.2340	0.2681	云南	0.4121	0.1704	0.5178	0	0.1395
江苏	0.7101	0.9566	0.5535	0.6492	0.6498	西藏	0	0	0	0.1929	0
浙江	0.4911	0.5708	0.2542	0.5106	0.5419	陕西	0.3229	0.2329	0.3750	0.0655	0.1280
安徽	0.5463	0.2908	0.3153	0.2590	0.2177	甘肃	0.2113	0.0695	0.3471	0.0227	0.0801
福建	0.3299	0.3492	0.1772	0.2641	0.2847	青海	0.0240	0.0148	0.1209	0.1075	0.0120
江西	0.3955	0.2114	0.2881	0.2082	0.1558	宁夏	0.0318	0.0241	0.2773	0.0269	0.0279
山东	0.8926	0.8068	1	0.5631	0.5946	新疆	0.1946	0.1082	0.5637	0.1075	0.1250
河南	0.8513	0.4891	0.3844	0.4594	0.3614						

沿用上文中多元线性回归数学模型，通过 SPSS16.0 软件运行得到经济与生态环境与各影响因素的关系(见表)，由 R^2 值为 0.989 可看出拟合效果和精度都是比较高的，显著性非常明显，能够很好的反映经济与生态环境发展与其主要因素之间的线性关系。

5.2.4 主成分综合分析结果

分析评价表见附表，分析图如下：



(1) 人口



(2) GDP



(3) 废气



(4) 废水



(5) 废渣

通过以上分析，证明了在保证各省经济发展平稳进行时，各省的经济增长数据的拟合效果是否符合平稳增长。

5.3 问题三：中国经济与生态协调发展的综合评价

通过查阅中华人民共和国国家统计局官网^[5]，我们得到近9年生态与经济等相关统计数据，并在问题一和问题二的基础上，得到如表1所示的影响经济与生态协调发展的主要指标体系。并选取从2010年至2018年期间各指标的统计数据进行分析。

5.3.1 指标选取及原始数据

选取影响中国经济发展与生态环境协调发展的 14 个指标，指标体系构建如下表：

表 11 经济发展与生态环境协调发展指标

经济发展与生态环境协	经济发展指标	GDP	地区生产总值 GDP(亿元)
			第三产业占 GDP 的比重

协调发展指标			财政收入占 GDP 的比重	
			人均地区生产总值 GDP	
		人口	常住人口	
			城市恩格尔系数	
	生态环境指标	废气	SO ₂ 排放量	
			烟尘排放量	
		废水	工业废水排放达标率	
		废渣	工业固体废物综合处理率	
		环境治理	能源消耗量	
			环境治理投资总量	

表 12 2010-2018 年原始数据表

指标	2010	2011	2012	2013	2014	2015	2016	2017	2018
地区生产总值 GDP(亿元)	412119	487940	538580	592963	641280	685992	740060	820754	917980
第三产业占 GDP 的比重	0.431	0.441	0.581	0.627	0.689	0.745	0.787	0.806	0.845
财政收入占 GDP 的比重	0.0518	0.0522	0.0571	0.086	0.0899	0.093	0.106	0.117	0.147
人均地区生产总值 GDP	28497	38102	39267	44047	44651	48430	50765	60855	61543
常住人口 (万人)	4189	4257	4278	4349	4356	4479	4534	4578	4657
城市恩格尔系数	0.46	0.45	0.45	0.45	0.42	0.42	0.37	0.34	0.32
SO ₂ 排放量	279864	280164	280519	280775	281018	281154	281371	282386	28321
烟尘排放量	390823	378901	378523	355786	362567	332469	289524	256859	25586
工业废水排放达标率	354854	326556	268747	235676	247842	246445	213564	225697	247312
工业固体废物综合处理率 (%)	85.2	85.2	85.8	86.2	86.3	86.6	86.6	86.7	87.3
能源消耗量	15.4	15.9	16.1	16.3	16.8	16.9	17.2	17.5	17.8
环境治理投资总量	165.4	165.8	152.7	122.8	116.6	134.4	132.0	121.5	126.6

考虑数据的可获取性及指标的代表性，下面用因子分析法对中国经济发展与生态环境协调发展状况进行实证分析，以达到降维、简化问题的目的；同时，利用该法消除指标间的相互影响，减少指标选择的工作量。

5.3.2 因子分析法

1.数据的无量纲化

由于多个变量具有不同的属性和单位，既有定性指标又有定量指标，各指标间无统一的标准，难以进行比较，而直接用因子分析法进行处理，结果易出偏差。为此，对数据进行标准化处理，计算步骤如下：

(1) 确定评价指标集

选取影响经济发展与生态环境协调发展的主要评价指标，确定指标集 $\{X_1, X_2, \dots, X_p\}$ ， p 为指标数目，并通过实际调查取得原始指标数据 X_{ij} 。

(2) 将指标原始数据标准化

为了消除不同指标间量纲不同的影响，将原始数据 X_{ij} 标准化，标准化后的指标值为 Y_{ij} 。

$$\text{对正向指标: } y_{ij} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, i = 1, 2, \dots, p;$$

$$\text{对负向指标: } y_{ij} = \frac{X_{\max} - X_i}{X_{\max} - X_{\min}}, i = 1, 2, \dots, p$$

2.因子分析

(1) 求标准化后的指标数据的相关系数矩阵

$$R = (r_{ij})_{n \times m}, r_{ij} = \frac{1}{n} \sum_{k=1}^n y_{ki} y_{kj}, \{i, j = 1, 2, \dots, p\} \quad (24)$$

(2) 求相关系数矩阵 R 的特征值与特征向量

假设特征值按序排列为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ ，相应的特征向量为 U_1, U_2, \dots, U_m ，它们标准正交。经过因子分析，可用较少的几个主要因子来代替原有变量的大部分信息，用下面的数学表达式表示为：

$$\begin{cases} y_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + a_1\varepsilon_1 \\ y_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + a_2\varepsilon_2 \\ \dots \\ y_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + a_p\varepsilon_p \end{cases} \quad (25)$$

其中， y_1, y_2, \dots, y_p 为 P 个原有的变量，是均值为 0、标准差为 1 的标准化变量， F_1, F_2, \dots, F_m 为 m 个因子变量， m 小于 p ，表示成矩阵为：

$$X = AF + \alpha\beta \quad (26)$$

其中 F 为因子变量或公共因子； A 为因子载荷矩阵； a_{ij} 为因子载荷，是第 i 个原有变量在第 j 个因子上的载荷； β 为特殊因子，表示原有变量中不能被因子变量所解释的部分。

(3) 按因子的累积贡献率选择 m 个主要因子

第一主因子的贡献率为 $\lambda_1 / \sum_{i=1}^m \lambda_i$ ，它是第一主因子的方差在全部方差中的比值，这个比值越

大，表明第一主因子综合原指标 $\{X_1, X_2, \dots, X_p\}$ ，信息的能力越强。前 m 个主因子的累积贡献率为

$\sum_{i=1}^m \lambda_i / \sum_{i=1}^p \lambda_i$ 。如果前 m 个主因子的累计贡献率 $> 85\%$ ，表明前 m 个主因子基本包含了全部测量

指标的大部分信息，这样既减少了指标的个数，又有利于对实际问题进行分析和研究。

(4) 将 m 个主因子指标综合成单指标

将累积贡献率达到 85% 的前 m 个主因子 F_1, F_2, \dots, F_m 的因子得分乘以主因子 F_i

($i=1,2,\dots,m$)的方差贡献率 $\alpha_i = \lambda_i / \sum_{i=1}^p \lambda_i$ 作为权数构造一个综合评价各指标的主因子函数：

$G = \alpha_1 F_1 + \alpha_2 F_2 + \dots + \alpha_m F_m$ 。通过因子分析法，可将原来的 p 个指标 X_1, X_2, \dots, X_p 减为现在的 m 个主因子，方便了研究。

(5) 综合评价公式为：

$$U_i = \sum_{j=1}^n G_{ij} W_{ij} (i=1,2,3,\dots,n) \tag{27}$$

U_i 为所求的系统值； G_{ij} 为 i 系统提取的第 j 个因子与指标的相关系数； W_{ij} 为 i 系统提取第 j 个因子的贡献率； n 为 i 系统提取因子的个数。将各系统值进行加权求和，得到经济发展与生态环境协调发展系统的协调。发展指数。U 为经济发展与生态环境协调发展的综合评价指数，指数越大，相对越协调；指数越小，相对越不协调。依据不同时期经济发展与生态环境的协调发展指数，可以判断经济发展与生态环境协调发展的相对程度。

5. 3. 3 数据处理及结论

表 13 特征值累积贡献率因子载荷矩阵表

成分	初始化评估			平方载荷			方形载荷		
	总计	变化率	累积率	总计	变化率	累积率	总计	变化率	累积率
1	12.358	49.500	64.495	12.432	64.928	64.495	12.205	61.443	61.443
2	3.029	18.498	79.595	3.209	15.498	79.595	3.432	16.753	75.332
3	2.008	16.186	89.739	2.439	10.429	89.739	2.345	13.729	89.739
4	.774	8.352	93.500						
5	.212	5.828	95.918						
6	.320	.953	97.337						
7	.328	.423	98.250						
8	.129	.227	99.329						
9	7.581E-02	.379	99.989						
10	2.943E-02	.144	100.000						
11	8.977E-16	4.488E-15	100.000						
12	4.355E-16	2.178E-15	100.000						
13	3.258E-16	1.629E-15	100.000						
14	3.108E-16	1.554E-15	100.000						

通过上述方法对原始数据进行处理，可得到影响经济发展与生态环境协调发展的各因子累积贡献率表 13。由表 13 可以看出，在影响经济与生态环境协调发展的指标中，前三个因子的累积贡献率为 89.635%，大于 85%，基本包含了变量中的大部分信息，可用这三个因子来反映中国经济发展与生态环境协调发展的状况。公因子 1 在 GDP 总量、第三产业占 GDP 的比重、常住人口、城市恩格尔系数、环境治理投资、工业废水排放达标率、工业固体废物处理率、能源消耗量、等变量上都有大于 0.8 的载荷，其贡献率最大，为 64.495%，主要反映经济综合实力与环境治理的能力。公因子 2 与人均 GDP、SO₂ 排放量、烟尘排放量有较大的关系，其方差贡献率为 13.10%，主要反映工

业化程度与大气污染状况。公因子 3 与财政收入占 GDP 的比重有较大的关系，其贡献率为 10.144%，主要反映体制因素和公众参与度。

通过回归法计算主因子得分，并以各因子的方差贡献率为权重进行加权求和，得 2010~2018 年中国经济发展与生态环境协调发展的综合得分，计算公式如下：

$$F = 0.5322F_1 + 0.2321F_2 + 0.16722F_3 \tag{28}$$

表 14 主因子得分及综合指数得分

年份	主因子 1	主因子 2	主因子 3	得分
2010	-1.85575	0.24747	1.45095	-1.01385
2011	-1.37392	-0.08951	0.48962	-0.85048
2012	-0.75486	-0.86675	-0.69289	-0.78826
2013	-0.42677	-0.6369	-1.07576	-0.47941
2014	0.11722	-0.46822	-1.62645	-0.15837
2015	0.30589	2.44499	-0.56116	0.510197
2016	0.40331	1.25446	-0.19931	0.465479
2017	0.44597	-0.63443	-0.10335	0.181397
2018	1.27911	-0.72459	0.58818	0.774576
权重	0.64495	0.15102	0.10038	0.89635

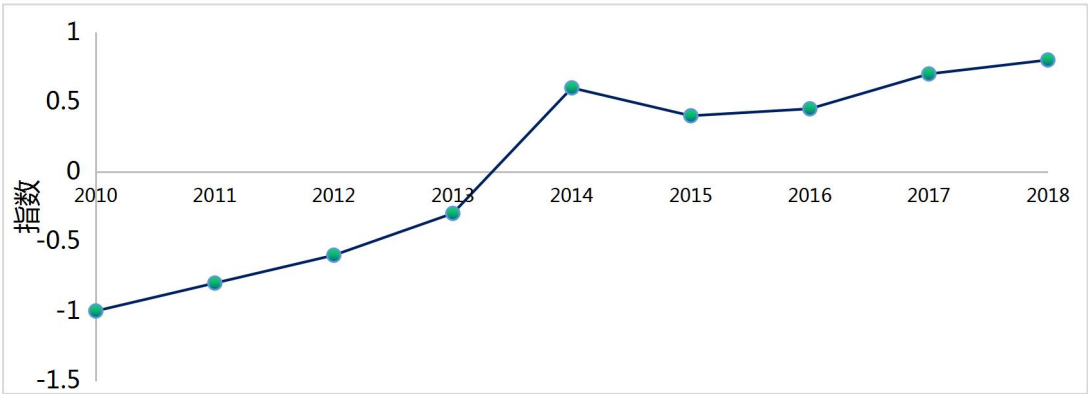


图 4 中国经济与生态环境协调发展综合评价指数

从图中可以看出，2010~2014 年，中国经济与生态环境协调发展的指数呈上升趋势；2014 年呈下降趋势，但仍高于前几年的水平；2015~2018 年经济发展与生态环境协调发展指数呈上升趋势。总之，2010~2018 年中国经济与生态环境协调发展指数总体上呈上升趋势。

5.3.4 中国经济与生态协调发展的建议及措施

1.坚持以绿色发展理念为指导，加强绿色治理

要牢固树立并践行“绿水青山就是金山银山”的绿色发展理念，走以生态保护优化经济增长之路。党的十八大以来，我国大力推动生态文明建设，践行绿色发展理念，生态环境得到了显著改善。相关部门要继续严格做好生态环境治理工作，坚持环境治理、生态修复与资源节约利用齐抓共管、齐头并进，做到治理理念创新、治理手段创新、治理管理创新。在涉及人民群众生产生活的大气、土壤、水环境、工农业、生活垃圾处理等方面加大治理力度，建立生态环境治理目标考核机制，做好治理监管工作。

2.转变经济增长方式，创新绿色技术

进入新时代，我国经济已由高速增长阶段转向高质量发展阶段，正处在转变发展方式、优化经济结构、转换增长动力的攻关期，建设现代化经济体系是跨越关口的迫切要求和我国发展的战略目

标。经济增长方式的转变，核心在科技。在新能源开发、清洁能源使用、循环产业发展等方面培养一批人才，加快建立一批绿色技术重大创新平台。要大力推动清洁能源代替传统能源，提高能源集约效率，用绿色技术治理生态环境，推动经济增长方式转变，实现生态保护与经济建设同步协调发展^[6]。

3.加快产业结构优化升级，发展绿色产业

随着我国改革进入深水区，优化产业结构已经摆在了突出位置。民族地区也正在逐步整一、二、三产业在国民经济中的比重，加快发展先进制造业、高新技术产业和服务业，深化供给侧结构性改革，淘汰化解落后产能，集中布局特色优势产业，严禁浪费资源和破坏自然生态。

4.增强全民环保意识，践行绿色生活

民族地区各级政府要进一步拓宽公众参与环境管理的渠道，加大环境信息公开力度，集民智、聚民意。要善用新闻媒体，大力宣传绿色环保理念，倡导全民践行绿色生活方式。社会基层组织要多开展生态环保活动，立足社区、家庭，推进以家庭为单位的节水、节电行动，降低能耗、物耗。消费方式和价值观念的转变，需要政府的大力倡导，更需要全社会的通力配合。要凝聚共识，汇聚生态保护力量，形成生态保护与绿色生活的强大合力，加快推进生产生活方式绿色化，引导绿色生产、绿色消费，从多源头保护生态环境，控制污染排放量。

5.加强教育及国家政策的宣传力度

要围绕经济发展与生态环境之间的关系，加大宣传教育力度，充分运用各种手段如广播、电视、报刊、标语及各种公益活动，开展全社会生态环境保护的宣传教育活动，普及公众生态环境保护与人类生存发展关系的知识，不断增强全社会的生态环境保护意识和公德意识，促进生产方式、生活方式和消费观念的转变，大力营造经济发展与生态环境协调发展的良好社会环境，尽快形成符合经济发展与生态环境协调发展原则的新的价值观和道德观，使人们的思想道德修养、科学文化水平、民主与法制观念得以显著提高，形成生态建设的精神动力。

6.建立长期稳定的生态环境投入机制

环境建设是一项长期性的系统工程，需要投入大量的资金，且其经济效益在短期内不显著。为保障生态建设的顺利实施，各级政府要把生态建设资金列入本级预算，并逐年按一定比例增长，充分发挥公共财政在生态环境保护 and 建设方面的导向作用。

六、模型评价

6.1 模型的评价

6.1.1 模型的优点

- 1.模糊综合评价法具有结果清晰，系统性强的特点，能较好地解决模糊的、难以量化的问题。
- 2.模型能比较细致地反应各指标对经济和生态环境交互状况的影响程度，且模型内容简易，构架清晰，通俗易懂，便于理解和运用到现实生活。

6.1.2 模型的缺点：

- 1 影响经济与生态环境的因素指标不是很全面，还有待更进一步的探讨和研究。
- 2 反映该问题的指标很多，对指标的选择和指标权重设置也各有侧重，往往这些差异会产生不同的评价结果。

6.2.模型的改进

(1) 为了使模型更加的丰富和完善，选取尽可能多的影响指标，而不是局限于题目中所指定的指标，使整个评价体系更加全面，更加具有信服力，防止因某些指标的影响带来不同的结果。

(2) 每个指标数据运用上存在差异，会使得一些熟知的比较存在一定的误差，尽量保证数据的真实有效，需要对更多的数据进行修正完善。

(3) 模型三考虑的是年度数据，考虑更细致的月度或季度数据，模型会更精确。

七、参考文献

- [1]叶珍, 基于 AHP 的模糊综合评价方法研究与应用[D].华南理工大学,2010.6
- [2]李智锋.基于改进回归和主成份分析的商品房定价决策模型[J].理论新探, 2011,(2)31-33.
- [3]陈博, 土建国.多元线性回归组合模型算法研究[J].西安文理学院学报, 2012,15(1):82-85.
- [4]<http://www.stats.gov.cn/>
- [5]孙秋鹏. 经济高质量发展对环境保护和生态文明建设的推动作用[J].当代经济管理,2019.05.30
- [6]蔡平.经济发展与生态环境的协调发展研究[D].新疆大学,2004.6