# Study on Application of Factor Analysis in Regional Environmental Assessment

Ping Xiao

Hunan University of Humanities, Science and Technology, China

applefly13@126.com

## Abstract

Environmental problems have become a research hotspot, and environmental quality evaluation is the premise and foundation of environmental research. In this paper, panel data of all provinces and regions in China in 2017 were selected, and seven major environmental pollution detection indexes were selected. Comprehensive indexes of environmental pollution were constructed through factor analysis, and three factors were extracted from the seven indicators. Then, the total score of factors of all provinces and regions was obtained, and the environmental pollution degree of all provinces and regions was comprehensively ranked. The results show that the provinces with less pollution are hainan, Tibet, ningxia, yunnan and qinghai. The most polluted areas are shandong, hebei, jiangsu and henan. This conclusion is helpful for China's environmental management departments to take effective measures to improve environmental pollution situation in different regions.

## Keywords

Factor analysis, enviromental quality, environmental pollution.

## 1. Introduction

Since the reform and opening up, China's economy has experienced rapid growth in the past 40 years. In the take-off stage of the economy, the labor-intensive and resource-intensive economic growth model has led to the excessive use of resources, causes the increase of pollution emissions and the decline of environmental quality. The economic development in the eastern and central and western regions of China is unbalanced, and the environmental pollution levels in various provinces and regions are not the same. Environmental pollution detection and evaluation is the foundation of environmental protection. The environmental pollution detection indexes include air pollution, water pollution and solid waste pollution. These indicators provide abundant information for environmental pollution situation in different regions. Environmental quality assessment is the premise and foundation of environmental research. Quantitative study of environmental quality by using statistical methods is conducive to the authenticity of evaluation results.

Multivariate large sample studies provide abundant information, but in most cases, due to the possible correlation between many variables, the complexity of the problem analysis is increased and the analysis is inconvenient. If each indicator is analyzed separately, the analysis may be isolated and cannot be comprehensively analyzed. Blindly reducing indicators will lose a lot of information and easily lead to wrong conclusions. Therefore, it is necessary to find a reasonable method to reduce the analysis index while minimizing the loss of information contained in the original index and make a comprehensive analysis of the collected data. Since there is a certain correlation between variables, less comprehensive indicators can be used to synthesize all kinds of information in each variable, which will make the research easier. In this paper, panel data of all provinces and regions in China in

2017 were selected, and seven major environmental pollution detection indexes were selected. Comprehensive indexes of environmental pollution were constructed through factor analysis, and three factors were extracted from the seven indicators. Then, the total score of factors of all provinces and regions was obtained, and the environmental pollution degree of all provinces and regions was comprehensively ranked.

## 2. Empirical analysis

Data construction factor analysis of 31 provinces and regions in China in 2017 was selected. Levels of air pollution as measured by indicators of sulfur dioxide, nitrogen oxides (N02), smoke dust and PM2.5. Solid waste (solid) and desertification (desert) measures the geological environment, wastewater emissions (wastewater) to measure the degree of water pollution.

Table 1. Descriptive statistics

| | The average | The standard deviation | Numbers |
|---|---|---|---|
| wastewater | 229385.5806 | 194225.72374 | 31 |
| S02 | 35.5765 | 25.08926 | 31 |
| NO2 | 44.9774 | 30.07389 | 31 |
| dust | 32.6023 | 26.65016 | 31 |
| solid | 9974.5274 | 8589.50206 | 31 |
| desert | 5095435.7732 | 15343836.65831 | 31 |
| PM2.5 | 52.8626 | 17.22187 | 31 |

Table 2. KMO and Bartley test

| | | |
|---|---|---|
| KMO statistics | | .664 |
| Bartlett test | chi-square | 181.917 |
| | df | 21 |
| | significant | .000 |

It can be seen from table 2.2 that KMO statistic =0.664>0.5, chi-square statistic of spherical test =181.917, and unilateral p=0.00<0.01, which is suitable for factor analysis.

The first component of the characteristic is 3.964, the second component of the characteristic value is 1.253, the eigenvalues of the former two factors are greater than 1. The percentage of variance components are listed the percentage of the total variance explained variance, which is the sum of the various factors of eigenvalue percentage, the first factor characteristic value accounts for 56.63% of the total characteristic value, the second factor characteristic value accounted for 17.903% of total eigenvalues. "Accumulative %" from top to bottom is the cumulative percentage of factor variance as a percentage of total variance. The sum of the eigenvalues of the first two factors accounted for 74.534% of the total variance. The first three factors explained 74.534% variation of the original 7 variables.

Table 3. Total variance decomposition

| component | Initial eigenvalue | | | Extract the sum of squares of loads | | | Sum of squares of rotational loads | | |
|---|---|---|---|---|---|---|---|---|---|
| | tatal | Percent variance | The cumulative % | tatal | Percent variance | The cumulative % | tatal | Percent variance | The cumulative % |
| 1 | 3.964 | 56.630 | 56.630 | 3.964 | 56.630 | 56.630 | 3.834 | 54.772 | 54.772 |
| 2 | 1.253 | 17.903 | 74.534 | 1.253 | 17.903 | 74.534 | 1.383 | 19.762 | 74.534 |
| 3 | .882 | 12.596 | 87.130 | | | | | | |
| 4 | .583 | 8.327 | 95.457 | | | | | | |
| 5 | .198 | 2.824 | 98.281 | | | | | | |
| 6 | .097 | 1.388 | 99.669 | | | | | | |
| 7 | .023 | .331 | 100.000 | | | | | | |

Extraction method: principal component analysis.

Table 4. Factor loading matrix for principal component analysis

Rotation matrix

| | component | |
|---|---|---|
| | F1 | F2 |
| wastewater | .297 | .794 |
| S02 | .903 | .119 |
| NO2 | .891 | .385 |
| dust | .962 | .055 |
| solid | .908 | -.078 |
| desert | .127 | -.763 |
| PM2.5 | .609 | .013 |

Extraction method: principal component analysis.

Rotation method: Kaiser standardized maximum variance method.

A. Rotation converges after 3 iterations.

The correlation between the three factors and each variable is fuzzy, and it is difficult to name the two factors. Therefore, in order to name the factors, rotation can be carried out to make the coefficients to divide into 0 and 1. The first factor F1 is highly correlated with variables such as sulfur dioxide, nitrogen hydride, smoke dust, and solid pollution, and is less correlated with desertification and water pollution. The second factor, F2, was highly correlated with water pollution and desertification, and negatively correlated with solid pollution and desertification. The higher the score of the first factor

(F1), the more serious the pollution of sulfur dioxide, nitrogen oxides and soot and solids. The higher the second factor F2 score, the more seriously affected by water pollution.

Table 5. Estimate the covariance matrix of the regression factor fraction

Component scoring matrix

| | component | |
|---|---|---|
| | 1 | 2 |
| wastewater | -.010 | .578 |
| SO2 | .238 | -.014 |
| NO2 | .203 | .193 |
| dust | .261 | -.070 |
| solid | .262 | -.166 |
| desert | .124 | -.603 |
| PM2.5 | .168 | -.061 |

Extraction method: principal component analysis.

Rotation method: Kaiser standardized maximum variance method.

Component scoring.

According to the factor score coefficient and the standardized value of the original variable, the score of each factor of each observation can be calculated, and the observation can be further analyzed accordingly.

F1=-0.01*wastewater+0.238SO2+0.203NO2+0.261dust+0.262solid+0.124desert+0.168PM2.5

F2=0.578*wastewater-0.014SO2+0.193NO2-0.070dust-0.166solid-0.6031desert-0.0616PM2.5

According to the factor formula calculating the total score of inter-provincial environmental quality, the higher the score of environment pollution is more serious, from the table, the shandong, hebei, jiangsu, henan and other places of factor, total score is higher, this is because the provinces as part of the country of the old industrial base, have many industrial polluting enterprises, some provinces of township enterprises has developed rapidly, because of the township enterprise's environmental protection consciousness is not strong, lack of environmental pollution control measures, these provinces scored higher on the industrial air pollution factor. The top 10 provinces with less pollution are hainan, Tibet, ningxia, yunnan, qinghai, guizhou, fujian, guangxi, jilin and gansu. Some regions are the undeveloped western regions, while some are economically developed regions that mainly focus on financial services, shipping, communications and light industrial manufacturing. Industrial air pollution is relatively light.

Table 6. Factor score table of each province

| region | F1 | F2 | The total score |
|---|---|---|---|
| Hai nan | -1.17632 | -1.02672 | -2.20304 |
| Tibet | -1.382 | -0.7291 | -2.1111 |
| Ningxia | -1.07994 | -0.60992 | -1.68986 |
| Yunnan | 0.74257 | -2.2314 | -1.48883 |
| Qinghai | -1.08343 | -0.24206 | -1.32549 |

| | | | |
|---|---|---|---|
| Guizhou | 0.37223 | -1.62317 | -1.25094 |
| Fujian | 0.11716 | -1.21487 | -1.09771 |
| Guangxi | -0.07758 | -0.89014 | -0.96772 |
| Jilin | -0.72838 | -0.1994 | -0.92778 |
| Gansu | -0.82168 | 0.00718 | -0.8145 |
| Chongqing | -0.61743 | -0.13336 | -0.75079 |
| Shanghai | -0.8308 | 0.20903 | -0.62177 |
| Jiangxi | 0.13235 | -0.6647 | -0.53235 |
| Tianjing | -1.57463 | 1.17805 | -0.39658 |
| Heilongjiang | -0.29545 | -0.0151 | -0.31055 |
| Neimenggu | 1.40696 | -1.68384 | -0.27688 |
| Hubei | -0.12347 | 0.29087 | 0.1674 |
| Hunan | 0.5387 | -0.33293 | 0.20577 |
| Shanxi | -0.58074 | 0.79197 | 0.21123 |
| Beijing | -1.40453 | 1.66394 | 0.25941 |
| Sichuan | 0.44367 | -0.06277 | 0.3809 |
| Anhui | -0.02474 | 0.46208 | 0.43734 |
| Xingjiang | -0.37556 | 0.84944 | 0.47388 |
| Liaoling | 0.45954 | 0.14994 | 0.60948 |
| Zhejiang | 0.12191 | 0.4988 | 0.62071 |
| Shanxi | 0.47061 | 0.32718 | 0.79779 |
| Guangdong | 1.85094 | -0.40036 | 1.45058 |
| Henan | 0.11321 | 1.70932 | 1.82253 |
| Jiangsu | 1.83467 | 0.93376 | 2.76843 |
| Hebei | 1.44752 | 1.43257 | 2.88009 |
| Shandong | 2.12463 | 1.5557 | 3.68033 |

## 3. Conclusion

In this paper, environmental measurement indexes of 31 provinces and regions in 2017 were selected to build a statistical model to evaluate environmental quality, and two common factors were extracted by factor analysis to represent seven environmental factors, including air, water, solid waste, garbage and desertification. The results show that the environmental quality is the best in the west, the second best in the east and the worst in the middle. The top 10 provinces with less pollution are hainan, Tibet, ningxia, yunnan, qinghai, guizhou, fujian, guangxi, jilin and gansu. Some areas are undeveloped western regions, while others are in financial services, shipping, communications and light industrial manufacturing.

## References

[1] Li zhengda, yuan xiaoling, Yang wan. Status quo, confusion and prospect of environmental quality evaluation research. Resource science.2014 (1).

[2] Yuan xiaoling, li zhengda, liu bolong. Dynamic comprehensive evaluation of regional environmental quality in China -- based on the perspective of pollution discharge. Resources and environment of the Yangtze river basin.2013 (1).

[3] Huang tongcheng, li shoude, gu mengdi. Application of factor analysis in environmental pollution detection and evaluation. Industrial engineering and management.2007 (5).