

# Multivariate Statistical Analysis of the Difference of Regional Economic Development of our Country

Tianci Zhai

99 ShangDa Road, BaoShan District, ShangHai, China

18817363937@163.com

**Keywords:** Differences in economic development; Factor analysis; Cluster analysis; Policy suggestions

**Abstract.** China has a large population and abundant resources. The differences between all the regions and nations are also very evident in many aspects. Especially the regional economic difference, it is the key to China's regional unbalanced development. In addition, the unbalanced development of regional human resources affected the balanced development of regional economy directly. Therefore, the research of the regional differences in economic development in China and the coordinated development of countermeasures has a guiding significance to the western development strategy and help to realize the sustainable development of China. This paper use the factor analysis, cluster analysis and some other statistical methods to analyze the economic development differences between the 31 provinces, municipalities directly under the central government and autonomous regions in China, and the coordinated development policy suggestion is given.

## Introduction

Since 1978, our country implements the strategy of taking economic construction as the center of the national economy fast development, the comprehensive national strength increasing, our country's economic status in the world also is increasing day by day. At present, our country is in the transferring period from populous country to economic giant country. However, because of too much emphasis on economic benefits, there has appeared many problems in regional development, the regional economic difference is one of the hotspots and key problems. Regional economic difference is a universal economic phenomenon, and it also has the objectivity.

Moderate regional economic differences have a positive role in promoting the development of the whole macro economy. However, the regional economic differences' expanding can make money, population and technology constantly concentrated to developed area, causing the local city swelling, infrastructure overtaxed, worsening environmental pollution, the deterioration of social public security problem, the urban poverty population increase.

A regional economic development research, formed by multiple indicator comprehensive value calculation, is a multi-index comprehensive decision problem in statistics. There are a variety of ways in multi-index comprehensive decision-making, here we will use factor analysis method firstly, reduce the dimension of index, drop into a few comprehensive indices which have fairly information. Then use clustering analysis to obtain several different economic regions according to the result of factor analysis.

## Factor Analysis

**The Model of Factor Analysis.** The observed random variable  $X$  has  $p$  ingredients, it's mean is  $\mu$ , the covariance matrix is  $\Sigma$ . Factor model ask for  $X$  to be linear and dependent on a few unobservable random variables  $F_1, F_2, \dots, F_m$  which we call them common factors and  $p$  additional

variation source  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$  which we call them error or special factor. In particular, the factor analysis model is:

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2$$

$$X_3 - \mu_3 = l_{31}F_1 + l_{32}F_2 + \dots + l_{3m}F_m + \varepsilon_3$$

...

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p$$

**The Steps of Factor Analysis.** Firstly, extract the initial factor by the correlation matrix.

Secondly, determine the initial factor. The principle is: the extracted factor's accumulated contribution rate is greater than or equal to 80%, factor loads' variable results show the same, the total amount of information without loss.

Thirdly, the rotation of the factor. When the initial factor couldn't represent the meaning of the variables typically, rotate the factor loading matrix to get new ideal factor loading matrix and the factor score matrix.

Fourthly, the factor score model.

$$G_i = \sum d_{ik}x_i (i = 1, 2, \dots, m)$$

Lastly, the synthesis score. Sum of the product of all factors and their contribution rate can get each factor's synthesis score.

**Data Preparation.** In this paper, according to the scientific, practical, completeness and feasibility principle, we select 10 indicators from two aspects of economic and human resources respectively for 31 provinces, cities, autonomous regions and municipality directly under the central government.

Table 1 Ten Indicators

Indicator	Meaning
X1	GDP
X2	fixed investment in the whole society
X3	GNPP
X4	final consumption
X5	added value of the tertiary industry
X6	total number of households
X7	natural population growth rate
X8	number of students which study in common colleges and universities
X9	application for patents
X10	faulty members in common colleges and universities

## **Rlanguage Modeling and Analysis of the Results**

### **Economic Subsystem.**

According to the characteristic root is greater than 1, selecting two common factor, the cumulative variance contribution rate is 89.9%, see Table1.

Table2 Factor scores

	Factor1	Factor2
SS loadings	3.380	1.116
Proportion Var	0.676	0.223
Cumulative Var	0.676	0.899

These variables have very strong powers of persuasion on regional economic development in our country. The factor loading matrix after rotating is:

$$F1=0.952X1+0.912X2+0.173X3+0.925X4+0.87X5$$

$$F2=0.298X1+0.822X3+0.344X4+0.482X5$$

The first factor has 67.6% amount of explanation, represents the total economic development level. The second factor can explain 22.3% information of the variable, on behalf of the relative levels of economic development among regions. Guangdong and Jiangsu ranked the first and second in the comprehensive score ranking, the score is 2.12 and 1.9874. So we can see that China's eastern coastal areas and the Pearl River Delta Regions have the highest degree in region economy development.

### **Human Resources Subsystem.**

Also select two factors from the five indexes, the cumulative contribution rate is 82.7%, see Table2.

Table3 Factor scores

	Factor1	Factor2
SS loadings	3.022	1.113
Proportion Var	0.604	0.223
Cumulative Var	0.604	0.827

These two factors have very strong powers of persuasion on regional human resources development in our country. The factor loading matrix after rotating is:

$$F1=0.911X1-0.166X2+0.937X3+0.669X4+0.917X5$$

$$F2=0.984X2-0.264X3-0.118X4-0.243X5$$

The first factor has 60.4% amount of explanation, represents the regional population quality in our country. Among them, Shanghai, Beijing, Tianjin three municipalities occupy the position of the top three Tibet has the lowest score -2.49588. The second factor on behalf of our country's population growth, it can explain 22.3% information of the intelligence. From the point of factor score, Guangdong and Shandong ranked as the top two. This phenomenon is inseparable with that these two provinces are both populous provinces. In the composite scores and rankings, ranked the top three is Guangdong, Jiangsu and Shandong, still has a comprehensive human resources power.

## Clustering Analysis

**Clustering Analysis Theory.** Based on the shortest distance method to calculate the Euclidean distance between 31 provinces, municipalities and autonomous regions to classify the ultimate factor score which come from the clustering analysis before. The shortest distance refers to the distance between the class and class is equal to the distance between two recentsamples. This method use  $D_k(p, q) = \min \{d_{ij} | i \in G_p, j \in G_q\}$  to describe the two closest samples' distance between  $G_p$  and  $G_q$ . If  $G_p$  and  $G_q$  combined into  $G_r$ , then the distance between  $G_r$  and other  $G_s$  is:

$$D_k(r, s) = \min\{D_k(p, s), D_k(q, s)\}$$

**Regional Spatial Difference of Economic Development and Human Resources in our Country.**

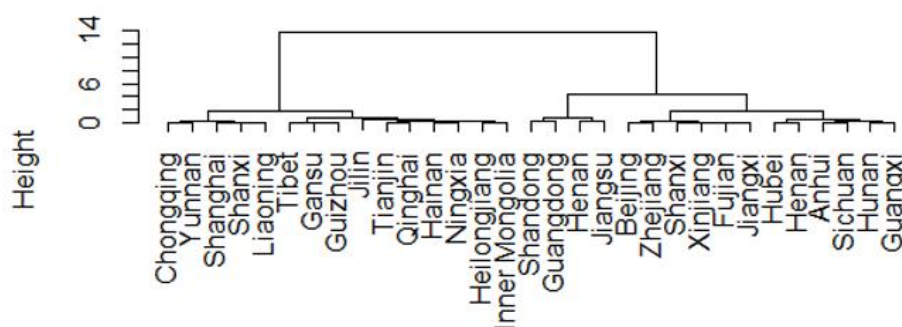


figure 1. Cluster Tree

According to fig. 1, we can divide our country into three kinds of different economic development and human resources area (see Table 3).

Table 4 Spatial difference in the regional economic development in our country

First kind	Shanghai, Beijing, Shandong, Jiangsu, Guangdong, Zhejiang, Henan, Anhui, Fujian, Hubei, Hunan, Inner Mongolia, Hebei, Liaoning, Sichuan, 15 developed regions
Second kind	Jilin, Shanxi, Yunnan, Shanxi, Jiangxi, Guangxi, Chongqing, Heilongjiang, Tianjin, 9 economic less-developed region
Third kind	Tibet, Hainan, Qinghai, Ningxia, Gansu, Guizhou, Xinjiang, 7 economic backward regions

## Policy Suggestions

Strengthen the macroeconomic regulation and control, make the necessary government intervention.

Reform and optimize government functions, constraint and regulate the behavior of local government from the aspectssuch as law, organization and risk. Strengthen regional economic functions.

Overall plan the regional development, establish regional interests adjustment mechanism.

Straighten out the resource price system.

Backward economic regions should to play its local advantages and outstanding characteristic.

Enhance the cooperation among eastern, central and western enterprises.

## References

- [1] J Shang, Y Yang, XH Li. Journal of China West Normal University (Natural Sciences) Vol.27(2006), No.3,p1
- [2] National Statistics Bureau: China Statistical Yearbook (China Statistics Press, 2014)
- [3] Richard A. Johnson Dean W .Wichem: Practical Multivariate Statistical Analysis (Tsinghua university press, 2001)
- [4] Z Y Zhang: statistical analysis of the differences and changes of the regional economic development level among the capital economic circle under the fiscal policy (MS. Capital University of Economics and Business,China,2013),p36
- [5] Z F Yuan, J Y Zhou: Multivariate statistical analysis.(Science press,2002)
- [6] B H Wang: Multivariate Statistical Analysis and R Language Modeling(Jinan University Press, 2010)
- [7] W Xue: SPSS Statistical Analysis Method and Application(Electronic Industry Press, 2004)
- [8] H Qiao. SCI-TECH INFORMATION DEVELOPMENT & ECONOMY, Vol.21(2011) No.1,p2-4
- [9] Y N Shi. Journal of Hubei University of Economics (Humanities and Social Sciences), Vol.11 (2014) No.8, p1
- [10] Wolfgang Hardle, Leopold Simar: Applied Multivariate Statistical Analysis (Peking University Press, 2011)