

# 基于因子分析和聚类分析研究云南省区域经济发展差异

姚 英

云南财经大学, 云南 昆明  
Email: 1220361023@qq.com

收稿日期: 2021年5月9日; 录用日期: 2021年5月23日; 发布日期: 2021年6月4日

---

## 摘 要

为了科学的评价云南省各地州市的经济发展水平, 本文选取2019年16个城市的统计数据中的10个指标, 对云南省各地区的经济发展水平进行分析。采用因子分析方法, 得到每个地州市的综合得分, 从得分上看云南省各地州市的经济发展水平存在一定的差距, 经济发展水平最好的是昆明市, 怒江州是最差的城市, 进而采用系统聚类法将其分为6个类, 最后对缩短这一差距提出了相应的建议。

## 关键词

因子分析, 聚类分析, 区域经济

---

# Research on the Regional Economic Development Differences in Yunnan Province Based on Factor Analysis and Cluster Analysis

Ying Yao

Yunnan University of Finance and Economics, Kunming Yunnan  
Email: 1220361023@qq.com

Received: May 9<sup>th</sup>, 2021; accepted: May 23<sup>rd</sup>, 2021; published: Jun. 4<sup>th</sup>, 2021

---

## Abstract

In order to scientifically evaluate the economic development level of various prefectures and ci-

ties in Yunnan Province, this paper selects 10 indicators from the statistical data of 16 cities in 2019 to analyze the economic development level of various regions in Yunnan Province. The factor analysis method is used to obtain the comprehensive score of each prefecture and city. From the score, there is a certain gap in the economic development level of the prefectures and cities in Yunnan Province. Kunming is the city with the best economic development level, and Nujiang Prefecture is the worst city. And then divide it into 6 categories by using hierarchical clustering method, and finally put forward corresponding suggestions for shortening this gap.

## Keywords

Factor Analysis, Cluster Analysis, Regional Economy

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

## 1. 引言

我国经济一直存在非均衡发展,不同地区的经济发展差别一般都存在,且存在差异原因多种多样,不仅仅是在东部、中部和西部地区之间,在省与省之间,经济的发展情况也存在着显著的不同,甚至在省内的各地州之间存在的差异也日益突出,云南各地州市之间也不例外。云南省旅游业的繁荣发展得益于它地处中国的西南地区,因为拥有特殊的地理条件,再加上独特的自然风光,一年四季适宜人居住和生活的气候以及由于民族聚居产生的多民族文化资源;同时云南省的植物资源、生物医药和烟草以及云南省的其他资源,比如矿产、生物、水资源等也十分的丰富。但同时,云南省地形复杂,以山地居多,由于这种地势,造成了各地州之间通行的不便,导致各地州之间的经济交流不频繁,从这个方面来说,区域经济差异还是比较明显的。适度的差异在一定程度上能够提高经济发展效率,但是不断扩大差距将阻碍国民经济的稳定发展,缩小区域之间的差距,才能更好的交流和共同发展。将国家的层面缩小化就映射到一个省的发展,省内发展水平的差距也会影响整个省的发展情况,云南省地处中国的边境,以世界的视角来看,它处在东亚、东南亚和南亚的交汇点,随着“一带一路”的发展,与对外的交流必不可少,而云南就是一个打开“三亚”地区的突破口,发展云南的经济成了一个必要的问题。然而云南各地区经济发展存在一定的不平衡,研究云南各州市的经济发展状况,对保持经济整体持续、快速、健康发展具有重要的指导意义,也能促进“三亚”地区的交流和发展。

## 2. 文献综述

吴波[1]用因子分析研究了珠三角地区的经济差异,将珠三角地区分为了最发达、次发达和相对落后地区,同时针对这一现象提出相应的措施。

张自敏等[2]对广西区域经济发展差异,运用因子分析和聚类分析两种方法进行研究,将这一区域的14个地区分为四种区域经济。

石学刚等[3]采用了2005年的云南省统计数据,选取12个经济指标,用因子分析的方法对云南省各地州市的经济发展情况进行研究,并分析了云南省每个州市的经济发展水平存在差异的原因。

综上所述,在这一文章中从经济综合实力及人民生活水平和产业结构两个方面使用因子分析和聚类分析两种统计方法分析云南省的区域经济差异。

### 3. 相关理论

#### 3.1. 因子分析

因子分析这一方法的基本思想最初是 Charles Spearman 和 Karl Pearson 等人在进行智力的定义和测量工作中使用的[4]。它可以看做是主成分分析方法的一种引申，它主要是根据变量间的相关性大小对变量进行分组，使组内变量的相关性较强，而组间变量的相关性较弱[4]。随着学者们的不断研究，在各种各样的领域，比如医学和经济学，都有使用因子分析的理论进行的研究，因子分析的理论和方法得以被发展和补充[5]。

#### 3.2. 聚类分析

聚类分析是根据数据间的亲疏关系将数据进行分类，它是研究如何将同类聚集在一起的一种统计分析方法[4]。聚类方法可以分为系统聚类法和分解聚类法两种，本文主要采用系统聚类方法。

### 4. 数据的分析

#### 4.1. 数据的来源

文章中所使用的数据均出自于《云南省统计年鉴 2020》，区域经济的涵盖面比较广，指标选取的范围也很广，本文根据之前学者研究的经验以及变量选取的科学性、系统性、综合性以及可行性原则，选取的指标为旅游总收入( $x_1$ )，地方一般公共预算收入( $x_2$ )，社会消费品零售总额( $x_3$ )，进出口总额( $x_4$ )，第三产业比重( $x_5$ )，第二产业比重( $x_6$ )，人均生产总值( $x_7$ )，就业人员( $x_8$ )，科技人员数( $x_9$ )，企业单位数( $x_{10}$ )，一共有 10 个指标，收集了云南省的 16 个城市的相关数据，将这些指标分为两个方面，如下表 1：

Table 1. Index classification

表 1. 指标分类

经济综合实力及人民生活水平	旅游总收入(亿元) ( $x_1$ )
	地方一般公共预算支出(亿元) ( $x_2$ )
	社会消费品零售总额(亿元) ( $x_3$ )
	进出口总额(亿美元) ( $x_4$ )
	人均生产总值(元/每人) ( $x_7$ )
	就业人员(万人) ( $x_8$ )
	科技人员数(人) ( $x_9$ )
	企业单位数(个) ( $x_{10}$ )
	第三产业比重(%) ( $x_5$ )
	第二产业比重(%) ( $x_6$ )
产业结构	

#### 4.2. 数据的描述性分析

近年来，云南省借助自己特别的地理优势，愈来愈多的外地游客慕名来到云南游玩，旅游业由此不断发展，在旅游业方面的收入不断增加，人们的生活质量也被其带动着提高，从图 1 可以看出从 1978 年到 2019 年云南省生产总值逐年增加，在 2002 年的时候迅速增长。2019 年，全省的生产总值达到了 23223.75 亿元。

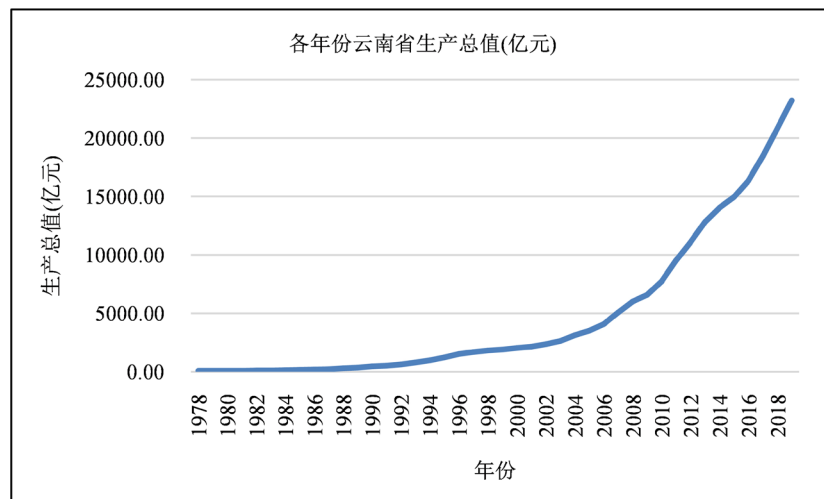


Figure 1. Yunnan's GDP in each year

图 1. 各年份的云南生产总值

尽管云南省总体经济发展蒸蒸日上,但不同地方的经济仍然存在着明显的差距,在图 2 中很明显的体现了这一点,云南省 16 个城市中昆明的人均生产总值是最高的,这是因为昆明市作为云南省的省会,是云南省的经济发展中心,同时,它还是中国与南亚、东南亚经济交流往来的重要通道。其次是玉溪,它因为气候的原因,且物产资源富饶,成为世界品质优良烤烟的最佳种植地。位于第三的是迪庆州,它的位置刚好是云南、西藏和四川三个省的交界点,是云南省唯一的藏族自治州,少数民族居多,景点丰富,虽然迪庆州的生产总值居于全云南省的第 15 位,但由于迪庆州的人口密度小,总人口数只有 40 万人,由此迪庆州的人均生产总值在 16 个地区中较靠前。其中人均生产总值最低的是昭通市,昭通市虽然区位优势、资源丰富,但是由于存在交通不便、产业弱小、生态脆弱等问题导致其经济发展受阻。昆明市的人均生产总值约是昭通市人均生产总值的 4.42 倍,很明显可以看出云南省各地区的经济差异非常大。

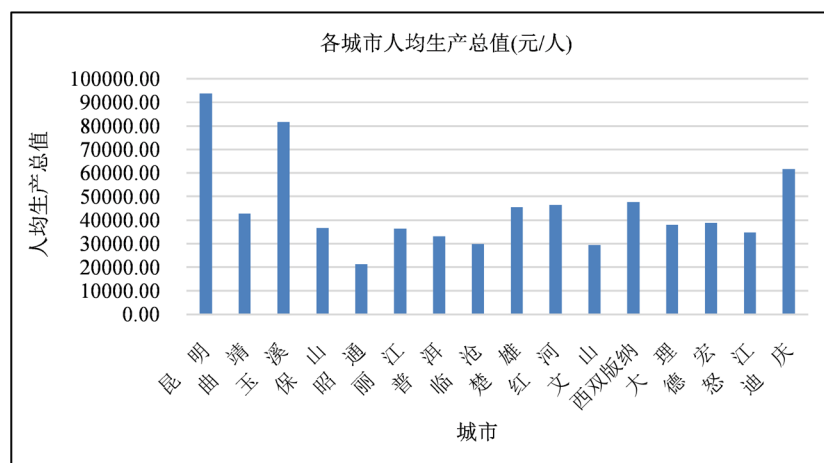


Figure 2. Per capita GDP of each city

图 2. 各城市的人均生产总值

产业结构在一定程度上代表着地区经济的发展水平,每个城市的产业结构往往由该地区的资源类型决定,这样就导致了发展受限制,虽然充分利用了该地区的优势,但也造成了产业结构呈现单一化的特

点,使不同地区的产业结构差别较大,从而也就导致经济差异[6]。从图3可以看出昆明市、德宏和西双版纳的第三产业占比远高于第二产业的占比。玉溪、保山、楚雄这三个城市的第二产业和第三产业的占比差距较小。

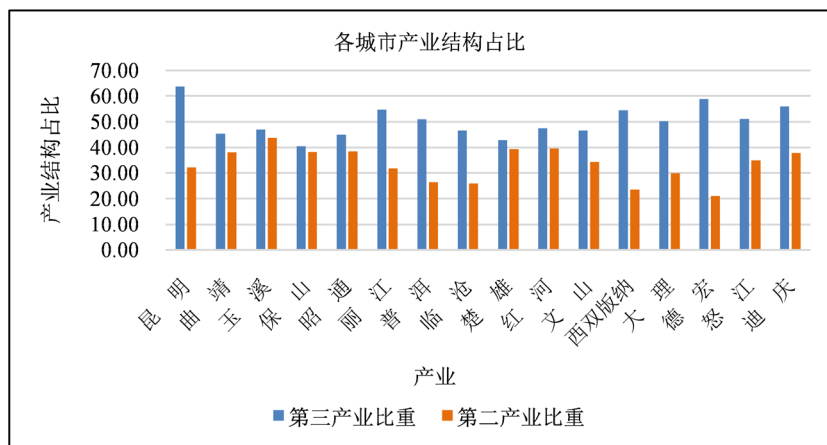


Figure 3. Proportion of the industrial structure of each city  
图3. 各城市产业结构占比

就业人口从某一方面来说可以反映一个地区的经济发展水平。一个地区只有经济发展得好,就业人数才会多,就业人数多意味着人们可支配的收入就相对多,人们的生活相对就好一点。从图4中可以看出,云南省中昆明市的就业人数最高,这是昆明市经济发展水平高的一种表现,其次是曲靖、昭通,就业人数相对较少的是西双版纳、怒江和迪庆。

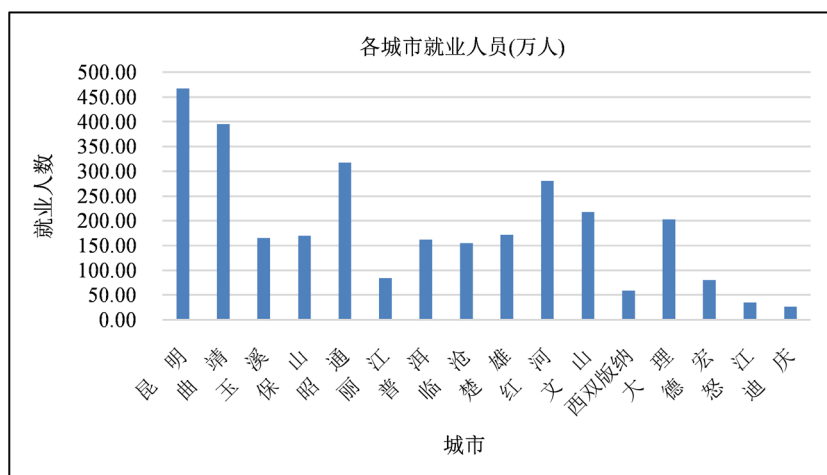


Figure 4. Number of employees in each city  
图4. 各城市就业人数

科技的进步同样可以促进经济的发展,所以各城市科技人员数也是导致区域经济发展产生差异的重要原因。从图5可以看出,昆明市的科技人员数是最多的,昆明市作为云南省的经济中心,会使大部分人才聚集在这里,且主要的企业也主要集中在昆明市,所以科技人员主要向昆明靠拢,从而也就带动经济的发展。其次是西双版纳,这是因为西双版纳独特的气候,促成了生物的多样性,所以吸引了很多技术人员前来研究。

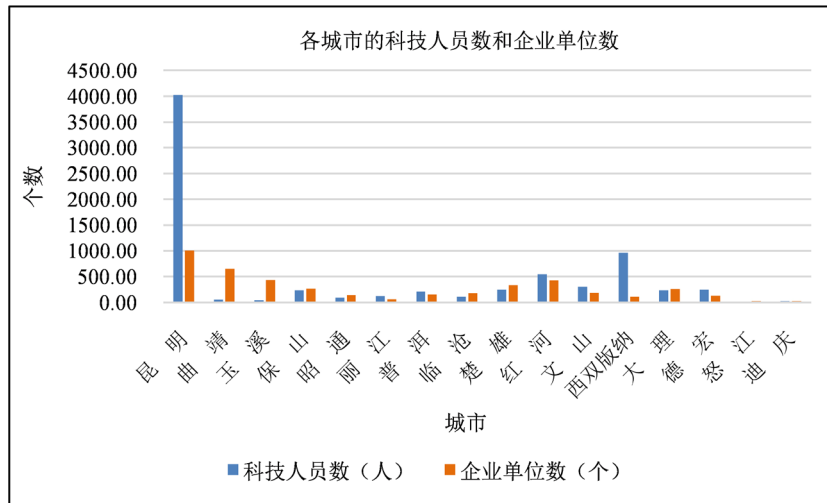


Figure 5. Number of scientific and technological personnel and enterprise units in each city

图 5. 各城市科技人员数和企业单位数

### 4.3. 因子分析

#### 4.3.1. 各指标间的相关矩阵

由表 2 可知，10 个变量之间的相关性较高，所以判断这一数据初步可以做因子分析。

Table 2. Correlation matrix between indicators

表 2. 各指标间的相关矩阵

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
$x_1$	1.000	0.904	0.893	0.837	0.594	-0.129	0.623	0.581	0.924	0.742
$x_2$	0.904	1.000	0.994	0.884	0.436	0.106	0.710	0.761	0.927	0.902
$x_3$	0.893	0.994	1.000	0.878	0.396	0.108	0.694	0.791	0.915	0.920
$x_4$	0.837	0.884	0.878	1.000	0.606	-0.122	0.715	0.538	0.886	0.768
$x_5$	0.595	0.436	0.396	0.606	1.000	-0.522	0.483	-0.068	0.602	0.148
$x_6$	-0.129	0.106	0.108	-0.122	-0.522	1.000	0.256	0.293	-0.144	0.267
$x_7$	0.623	0.710	0.694	0.715	0.483	0.256	1.000	0.273	0.659	0.664
$x_8$	0.581	0.761	0.791	0.538	-0.068	0.293	0.273	1.000	0.557	0.851
$x_9$	0.924	0.927	0.915	0.886	0.602	-0.144	0.659	0.557	1.000	0.731
$x_{10}$	0.742	0.902	0.920	0.768	0.148	0.267	0.664	0.851	0.731	1.000

#### 4.3.2. 因子分析的可行性检验

使用因子分析方法的条件，除了要求数据之间具有较强的相关关系，还要对待使用的样本数据进行检验，在这一文章中主要使用 KMO 检验和 Bartlett 的球形检验这两种常见的方法进行因子分析的可行性检验[7]。

KMO 检验统计量通常被用来衡量变量之间的相关系数和偏相关系数[8]。它的取值范围通常介于 0 和 1 之间。当 KMO 的值越趋近于 1 时,说明变量间的相关性就越强,因子分析对于这一数据而言就越适合;反之,当 KMO 的值越趋近于 0 时,说明变量间的相关性就越弱,因子分析对于这一数据而言就越不适合。

Bartlett 的球形检验可以用来查验变量之间是否独立。如果变量之间独立(原假设),则该数据不适合做因子分析,而如果变量之间不独立(备择假设),则该数据适合做因子分析。

两种检验方法的结果如表 3 所示:

**Table 3.** Results of two inspection methods

**表 3.** 两种检验方法的结果

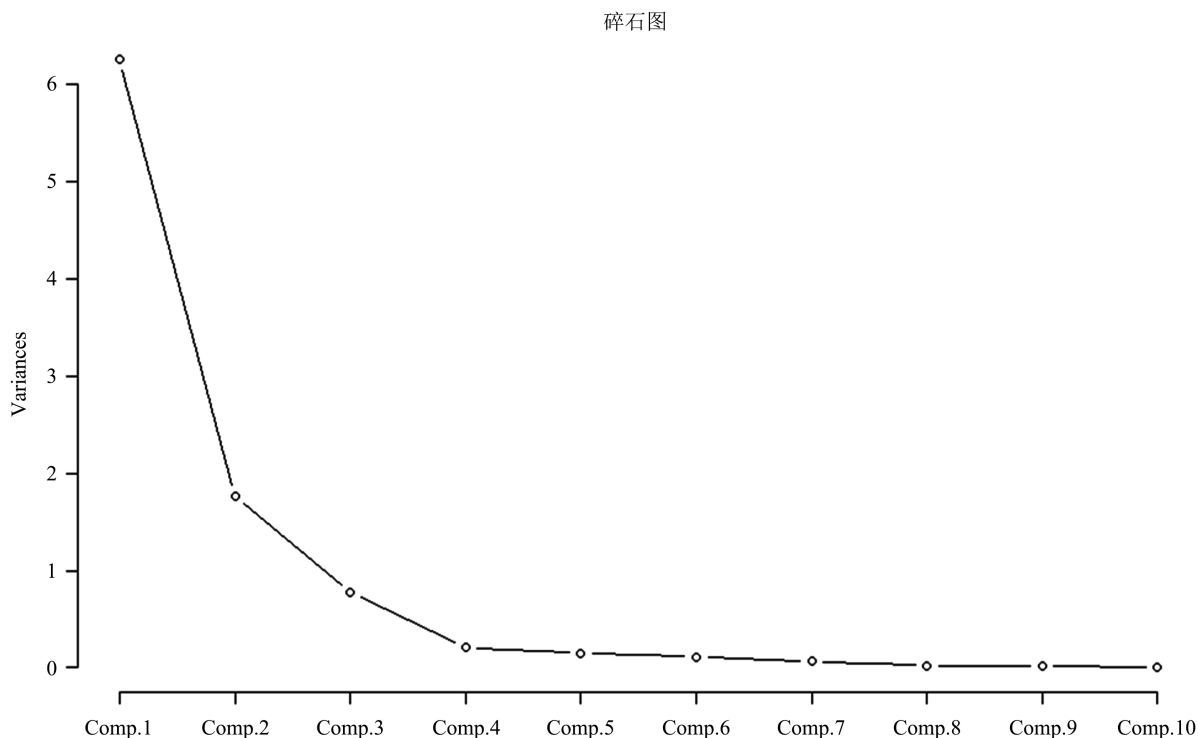
KMO 测量取样适当性	Bartlett 的球形检验		
	近似卡方	自由度	$p$ 值
0.768	209.953	45	0

从表 3 可知, KMO 检验值为 0.768, 明显大于 0.5, 同时, Bartlett 的球形检验的  $p$  值为 0, 小于 0.05 (显著性水平为 0.05), 所以变量之间不是独立的, 拒绝原假设。这一数据可以做因子分析。

#### 4.3.3. 基于主成分法的因子分析

首先用碎石图进行初步判断公共因子的个数。

由碎石图(图 6)可以初步判断出该数据可以取两个公共因子。



**Figure 6.** Gravel map

**图 6.** 碎石图

基于主成分法作因子分析，同时采用方差最大化作因子正交旋转，取两个公共因子的方差贡献率结果如表 4 所示：

**Table 4.** Variance contribution rate after rotation of each common factor  
**表 4.** 各公共因子旋转后的方差贡献率

因子	因子 1 ( $f_1$ )	因子 2 ( $f_2$ )
方差贡献率	66.3%	19.32%

**Table 5.** Rotated factor loading matrix  
**表 5.** 旋转后的因子载荷矩阵

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$
公共因子 $f_1$	0.902	0.992	0.992	0.901	0.431	0.127	0.750	0.760	0.914	0.926
公共因子 $f_2$	0.276	0.015	-0.017	0.283	0.830	-0.848	0.085	-0.445	0.300	-0.269

从表 4 可以看出前两个因子的方差累积贡献率为 85.62%，大于 80%，这说明总体的 85.62% 的信息可以由这两个因子解释，原始数据中的绝大信息被保留下来。

表 5 为采用方差最大化作因子正交旋转后的因子载荷矩阵，通过结果可以看出旋转后的因子载荷矩阵更容易解释。从表中可以知道：公共因子  $f_1$  在旅游总收入( $x_1$ )，地方一般公共预算收入( $x_2$ )，社会消费品零售总额( $x_3$ )，进出口总额( $x_4$ )，人均生产总值( $x_7$ )，就业人员( $x_8$ )，科技人员数( $x_9$ )，企业单位数( $x_{10}$ )上的载荷较大，所以公共因子  $f_1$  主要反映的是云南省的经济综合实力及人民生活水平；公共因子  $f_2$  在第三产业比重( $x_5$ )，第二产业比重( $x_6$ )上的载荷较大，所以公共因子  $f_2$  主要反映云南省的产业结构。

根据正交旋转后的因子载荷矩阵，可以得出因子分析的模型为：

$$\begin{aligned}
 x_1 &= 0.902f_1 + 0.276f_2 & x_2 &= 0.992f_1 + 0.015f_2 \\
 x_3 &= 0.992f_1 - 0.017f_2 & x_4 &= 0.901f_1 + 0.283f_2 \\
 x_5 &= 0.431f_1 + 0.830f_2 & x_6 &= 0.127f_1 - 0.848f_2 \\
 x_7 &= 0.750f_1 + 0.085f_2 & x_8 &= 0.760f_1 - 0.445f_2 \\
 x_9 &= 0.914f_1 + 0.300f_2 & x_{10} &= 0.926f_1 - 0.269f_2
 \end{aligned} \tag{1}$$

通过采用回归估计可以计算出每个城市的因子得分，同时根据表 4 中各公共因子旋转后的方差贡献率，可以计算出综合因子得分为：

$$f = 66.3\% f_1 + 19.32\% f_2 \tag{2}$$

公共因子  $f_1$ ，公共因子  $f_2$  以及综合因子的得分及排名情况见表 6。

由表 6 可知，昆明的综合因子  $f$  的得分是最高的，这是由于它的公共因子  $f_1$  的得分最高，也就是说昆明人民生活水平和经济实力相较于其他城市是较好的，这也符合昆明作为云南省省会，经济发展中心，旅游业也比较成熟的情况。同样的，综合因子  $f$  的得分为正的城市还有红河、德宏、西双版纳、玉溪、曲靖这四个城市，其中红河、玉溪和曲靖三个城市的综合因子  $f$  的得分为正，主要是由于它们的公共因子  $f_1$  的得分居于前五，这些城市在云南省都是经济发展水平较好的。而德宏和西双版纳的综合因子  $f$  的得分为正，主要取决于它们在公共因子  $f_2$  的得分排名居于前二，虽然它们的公共因子  $f_1$  的得分为负，但德宏和西双版纳的经济发展主要靠的是旅游业的发展。德宏作为云南省的自由贸易试验区片区之一，打造了“一带一路”和长江经济带互联网互通的重要通道，成为连接南亚和东南亚的纽带[9]，这给



德宏的经济贸易带来了新的发展机会，带动了德宏经济的快速发展。同时，伴随着旅游宣传力度的日渐增强，德宏的旅游业也迅速发展起来，带动着第三产业的发展。同样的，在“一带一路”以及孟中印缅经济走廊建设等国家发展战略的贯彻实施下[10]，也为西双版纳的经济发展带来了很大的机遇，带动第二、三产业迅速发展。所以德宏和西双版纳的公共因子  $f_2$  的得分较高。怒江虽然公共因子  $f_2$  的得分不算低，但由于公共因子  $f_1$  的得分为全云南省最低，这是因为怒江州的地理位置限制了发展，所以导致它的综合因子  $f$  的得分是最低的。

**Table 6.** Factor score and ranking

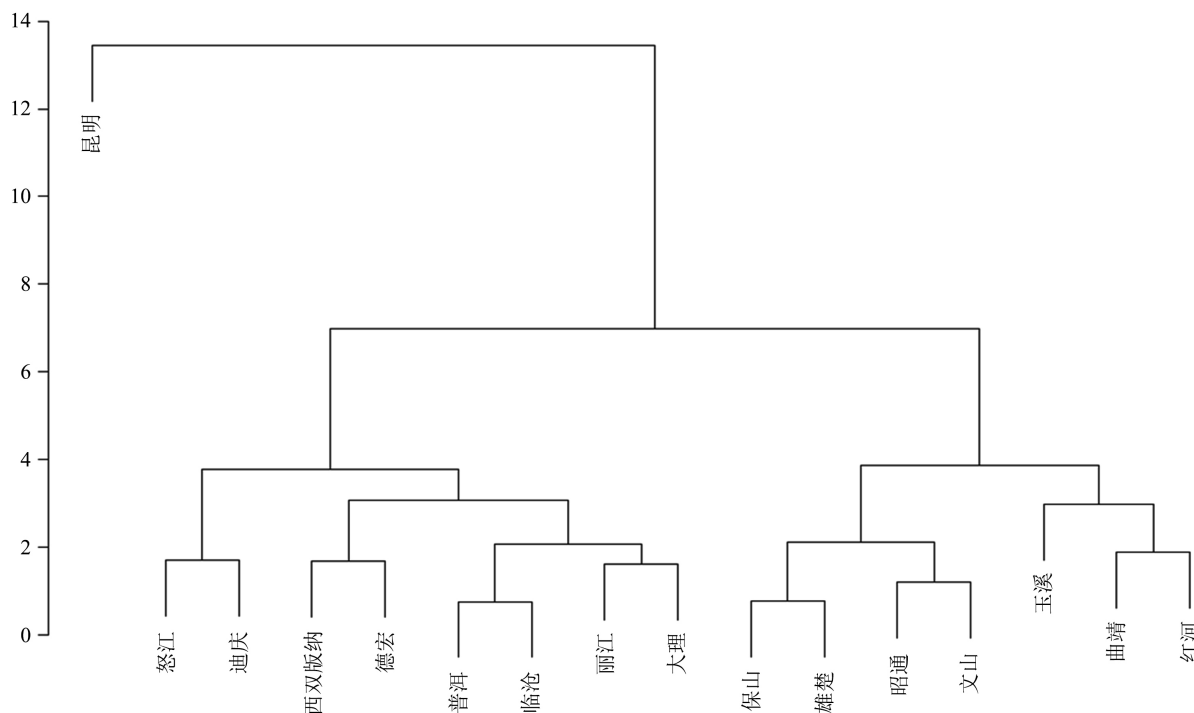
**表 6.** 因子得分及排名

城市	公共因子 $f_1$		公共因子 $f_2$		综合因子 $f$	
	得分	排名	得分	排名	得分	排名
昆明	3.478	1	0.862	3	2.4724524	1
曲靖	0.433	3	-1.479	16	0.0013362	6
玉溪	0.35	4	-1.034	13	0.0322812	5
保山	-0.305	7	-1.045	14	-0.404109	14
昭通	-0.319	9	-1.052	15	-0.4147434	15
丽江	-0.46	12	0.791	4	-0.1521588	8
普洱	-0.468	13	0.589	5	-0.1964892	9
临沧	-0.549	14	0.29	7	-0.307959	13
楚雄	-0.099	5	-0.98	12	-0.254973	10
红河	0.493	2	-0.785	11	0.175197	2
文山	-0.31	8	-0.426	10	-0.2878332	11
西双版纳	-0.336	10	1.471	2	0.0614292	4
大理	-0.113	6	0.178	9	-0.0405294	7
德宏	-0.356	11	1.869	1	0.1250628	3
怒江	-0.842	16	0.281	8	-0.5039568	16
迪庆	-0.596	15	0.472	6	-0.3039576	12

#### 4.4. 聚类分析

在这一文章中使用离差平方和法(Ward 方法)进行聚类分析[4]。

根据图 7，取合并距离为 4，我们可以将 16 个地区分为六类，玉溪、曲靖和红河为一类，这三个城市的公共因子  $f_1$  的得分都为正，且排在前五，人民生活水平 and 经济实力都较好。保山、楚雄、昭通和文山为一类，该类城市的公共因子  $f_1$  和公共因子  $f_2$  的得分都不高，公共因子  $f_1$  的得分稍高。西双版纳、德宏为一类，这一类城市的公共因子  $f_2$  的得分排名为前二，即它们的第二、三产业占比较重。普洱、临沧、丽江、大理为一类，这类城市公共因子  $f_1$  和公共因子  $f_2$  的得分都不高，公共因子  $f_2$  的得分稍高。怒江和迪庆为一类，这类城市公共因子  $f_1$  得分最低，相较于其他城市人民生活水平 and 经济实力都较差。昆明为单独一类，该类城市的公共因子  $f_1$  和公共因子  $f_2$  的得分都较高，昆明作为云南省的省会，是经济发展中心，经济实力最雄厚。



**Figure 7.** Dendrogram of the economic development of 16 cities by the sum of squared deviation method and the clustering dendrogram

**图 7.** 16 个城市的经济发展离差平方和法系统聚类树状图

## 5. 结论

从以上的结果分析可以得出，云南省的各区域由于历史、地理位置等要素而导致各城市经济发展存在着差异，经济发展水平最好的是云南省的省会昆明，怒江州是云南省经济发展水平最差的城市，且昆明市和怒江州的经济差距十分显著。云南省要缩减区域经济差异，需要各城市克服地域限制，因地制宜，发展适合且具有自己特色的经济，逐渐缩小各地区的经济发展差异和人民生活水平的差距，从而促进云南省整体经济的繁荣发展。比如，对于西双版纳和德宏这一类的城市，虽然旅游业迅速发展起来，旅游总收入有所增加，但它们的人均生产总值仍然较低，所以可以采取一些新兴技术来提高人民的经济收入。比如这两个城市的气候属于热带和亚热带季风气候，盛产热带水果和特色产品，可以适应时代的发展，借助网络平台宣传产品，利用电商平台推广产品，可以尽量减少信息不流通的问题。也可以通过各种直播平台，打破消费者和商家之间的信息不对称的现状，获得更多消费者的信任，从而提高经济水平。普洱、临沧、丽江、大理这些城市也主要是发展旅游业，建设具有自己本地区文化特色的景点和打造立体化的旅游产品，才能够与其他地区具有竞争的优势。

## 参考文献

- [1] 吴波. 基于因子分析的珠三角地区经济差异实证研究[J]. 中国市场, 2013(20): 19-20.
- [2] 张自敏, 樊艳英, 陈冠萍, 等. 基于因子和聚类分析的广西区域经济发展差异研究[J]. 科技和产业, 2013, 13(2): 15-18, 43.
- [3] 石学刚, 汪戎, 吴春芳. 基于因子分析法的云南省各市(州)经济发展水平综合评价[J]. 现代商业, 2008(8): 226-227.
- [4] 费宇. 多元统计分析——基于 R [M]. 北京: 中国人民大学出版社, 2014.
- [5] 黄春松. 长三角经济区、珠三角经济区与海西经济区的城市竞争力综合比较——基于因子分析模型的研究[J].

---

经济社会体制比较, 2013(1): 205-216.

- [6] 张艳萍. 浅谈我国工业经济区域差异及协调发展[J]. 现代工业经济和信息化, 2014, 4(8): 11-12.
- [7] 薛薇. 基于 SPSS 的数据分析[M]. 北京: 人民大学出版社, 2006.
- [8] 吴海. 基于因子分析和熵权法的全国科技发展水平评价[J]. 经济师, 2016(1): 44-46.
- [9] 李翔. 加快云南连接东南亚, 南亚国际大通道公路建设步伐的探讨[J]. 云南现代交通, 2004, 1(1): 23-29.
- [10] 刘宪立, 杨蔚. 基于数据包络分析法云南低碳物流效率评价研究[J]. 现代商业, 2020(18): 124-125.