

Research on Urban Economic Development Based on Multivariate Statistical Analysis

—Take Wuhan as an Example

Haiyan Wu

Qufu Normal University, Qufu Shandong
Email: qfnuwhy@163.com

Received: Jul. 16th, 2020; accepted: Jul. 29th, 2020; published: Aug. 6th, 2020

Abstract

Since the “rise of central China” plan, the development of Wuhan city has been one of the key national concerns. After years of development, Wuhan’s economy has made remarkable achievements, but there are still some differences in the economic development of various urban areas. Because of the large-scale outbreak of covid-19, the economic development of Wuhan has been impacted. In view of this, this paper studies the economic development of each urban area in Wuhan. Firstly, it selects appropriate index variables, and then objectively compares the ranking of urban economic development in Wuhan based on the scores of principal component analysis and factor analysis, and gives a comprehensive evaluation model. Then cluster analysis and discriminant analysis are used to study the structure of economic development, and indicate the current situation of economic development in Wuhan city. Finally, it provides optimization measures and suggestions for the economic development of Wuhan.

Keywords

Urban Economic Development, Multivariate Statistical Analysis, Comprehensive Evaluation

基于多元统计分析的城区经济发展研究

——以武汉市为例

吴海燕

曲阜师范大学, 山东 曲阜
Email: qfnuwhy@163.com

收稿日期: 2020年7月16日; 录用日期: 2020年7月29日; 发布日期: 2020年8月6日

摘要

自“中部崛起”计划以来，武汉市的发展一直是国家重点关注对象之一。经过多年发展，武汉市经济总体取得了骄人的成绩，但是各城区经济发展仍存在一定差异。又由于COVID-19的大规模爆发，给武汉市的经济发展带来一定的冲击。有鉴于此，本文对武汉市各城区经济发展进行研究，首先选取恰当的指标变量，并在此基础上客观的比较了基于主成分分析和因子分析得分的武汉市城区经济发展排名，给出综合评价模型。再利用聚类分析和判别分析研究其经济发展结构，并指示出武汉市城区经济发展的现状。最后为武汉市经济发展提供优化措施和建议。

关键词

城区经济发展，多元统计分析，综合评价

Copyright © 2020 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

区域经济发展问题是统计学研究的重要领域之一，本文参考了大量的相关文献，发现之前在研究区域经济问题时大多以省份或地级市为研究对象，发现并解决国家宏观经济中的有关问题[1]，而具体研究地级市管辖的城区经济寥寥无几。2020年是全面建成小康社会的关键一年，城区经济发展地位日趋重要和受瞩目。对市政府而言，若想推动当地经济发展，必须充分了解当地城区经济的发展状况。因此，本文在充分运用有关数据、资料的基础上，结合多种统计软件以及统计分析方法，从不同角度对武汉市城区经济发展指标进行综合分析，对多种方法分析的结果进行综合评价，得出一些合理的结论，从而为制定武汉市城区经济发展的规划提供更客观、全面的参考依据，也可供其他城市制定相关的城区发展策略进行借鉴[2]。

2. 材料与方法

2.1. 城区经济发展研究变量体系的设立原则

武汉市城区经济发展选取的变量是用于对各城区的经济发展状况进行全面的评价，其变量应能够反映经济发展状况且既相互联系又彼此独立，因此，本文选取变量的基本原则是[1]：

1) 代表性。对城区经济发展进行评价一定要体现代表性，选取的变量要与经济中能够区分不同城区经济发展水平，并且可以反映问题具有代表性的变量。

2) 可行性。选取指标时，要充分考虑所选取的变量是否适合用来定量分析，以确保分析的可行性和分析结果的客观性。

3) 可获得性。选取变量时要考虑所需数据的来源，在遵循代表性和可行性的基础上，选择容易通过统计年鉴、单位部门等有关途径获得准确数据的变量。

4) 数据正确性。数据最好源自国家或各级政府部门公开出版的报告、发布的统计资料，以确保数据的准确性。

2.2. 武汉市城区经济发展研究变量的选取

在以上选择变量的基础上, 本文以 2017 年武汉市 13 个城区的经济发展为对象(数据来源于 2018 年武汉市统计年鉴), 选择了九项能够反映城区经济发展的主要经济指标[4]。

- 1) 全社会固定资产投资额: 投资额越高, 表示用来满足人民生活生产的投资活动越多。
- 2) 社会消费品零售总额: 用来反映一定时期内人民物质文化生活水平的提高情况, 是反映社会商品购买力、零售市场规模和经济活跃度的重要指标。
- 3) 从业人员平均数: 指报告期内平均每天拥有的从业人员数, 包括全部职工、再就业的离退休人员、个体从业人员等, 它是一定时期内全部劳动力资源的实际利用情况的一个反应。
- 4) 分区生产总值: 反应的是国内生产总值在城区范围内一定时期里生产活动的最终成果, 从不同的侧面反映某地区的经济规模和经济总量情况。
- 5) 城镇居民人均可支配收入: 反映居民可以用来安排家庭日常生活的全部现金收入, 是反映居民收入状况的一个重要指标。
- 6) 分区公共财政总收入: 地区财政参与社会产品分配所取得的收入, 是地区政府行使相关职能的财力保障。
- 7) 主营业务收入: 通常指企业的经常性的、主要业务所产生的基本收入, 是反映企业盈利能力的一个重要指标。
- 8) 城镇居民人均消费支出: 指城镇居民在购买商品和劳务活动两方面的支出, 往往用来衡量居民的消费支出情况。
- 9) 工业总产值: 指以货币形式表现的工业企业在报告期内生产的工业产品总量, 是衡量一定时间内工业生产水平和规模状况的重要指标。

为了文中表述方便, 以上指标分别用以下记号表示, 见表 1:

Table 1. Variable list

表 1. 变量表

记号	变量	记号	变量	记号	变量
X_1	全社会固定资产投资	X_2	社会消费品零售额	X_3	从业人员平均人数
X_4	分区生产总值	X_5	城镇居民人均可支配收入	X_6	分区公共财政总收入
X_7	主营业务收入	X_8	城镇居民人均消费支出	X_9	工业总产值

2.3. 城区经济发展研究的方法

本文首先选取 9 个影响城区经济发展的主要指标, 构建了武汉市城区经济发展评价变量体系, 借助 R、SPSS 统计软件, 运用主成分分析、因子分析、聚类分析和判别分析对数据进行分析。前两种方法用于对原始数据进行统计分析, 并根据分析的结果对各城区经济发展情况进行综合评价和排名, 并采用 Spearman 相关性检验两种结果的等价性。之后运用聚类分析对 13 个城区进行分类, 以便指示出各城区发展的共性和个性。最后引入判别分析法, 以检验聚类分析的结果是否合理。

2.3.1. 主成分分析法

① 数学模型

设 $X_1, X_2, X_3, \dots, X_p$ 表示 P 个原始指标, 用这 p 个指标对 n 个样品进行评价, 那么共有 np 个数据, 在实际问题的研究中, 通常这 p 个原始变量会具有不同的量纲, 为了消除量纲的影响, 进行主成分分析之前, 需要将数据进行标准化, 设标准化后的变量为 x_1, x_2, \dots, x_p 。

若存在某个线性变换:

$$\begin{cases} y_1 = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p \\ y_2 = w_{21}x_1 + w_{22}x_2 + \dots + w_{2p}x_p \\ \vdots \\ y_p = w_{p1}x_1 + w_{p2}x_2 + \dots + w_{pp}x_p \end{cases}$$

此线性变换将 p 个标准指标变量 x_1, x_2, \dots, x_p , 变成 p 个新变量, 并且该线性变换满足以下三个条件:

- 1) y_i 和 y_j 相互独立
- 2) $\text{var}(y_1) \geq \text{var}(y_2) \geq \dots \geq \text{var}(y_p)$
- 3) $w_{i1}^2 + w_{i2}^2 + \dots + w_{ip}^2 = 1, i=1, 2, \dots, p$

那么称上述线性变换的式子为 p 个标准指标变量的主成分分析, 其中 y_1, \dots, y_p 就是由上述线性变换式子所确定的 x_1, x_2, \dots, x_p 的 p 个主成分, 其分析流程图见图 1。

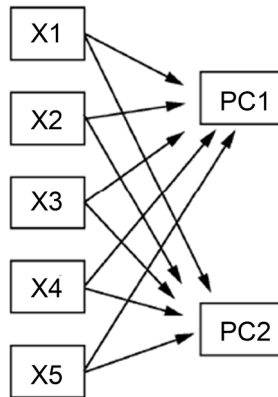


Figure 1. PCA model (variable x1-x5 maps the PC1, PC2)
图 1. 主成分分析模型(变量 x1-x5 映射为主成分 PC1, PC2)

② 主成分的选取

定义 $\alpha_k = \frac{\lambda_k}{\sum_{i=1}^p \lambda_i} (i=1, 2, \dots, p)$ 为第 k 个主成分 y_k 的方差贡献率

则 $\sum_{i=1}^m \alpha_i = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} (m < p)$ 表示前 m 个主成分的累计方差贡献率

实际意义来确定提取的主成分个数, 通常提取主成分的方法有两个: 一个是根据主成分的累计贡献率的大小, 当其达到一定程度(如 85%以上)来确定 m , 另一个是根据特征值来决定, 提取那些绝对值大于 1 的主成分。

③ 主成分得分

上述线性变换的 w_{ij} 成为得分系数, 其具体含义为第 i 个主成分对应在第 j 个标准指标变量 x_j 上的得分系数, 把每一个样品的 p 个标准化数值代入上述线性变化式子中计算其结果, 可以得到每一个样品的 k 个主成分得分, 进而可以提取各主成分得分计算其综合得分, 从而根据综合得分大小可以对各样品进行综合排名, 方便第各个样品进行综合评价。

2.3.2. 因子分析

因子分析是通过研究原始变量内部相关的依赖关系出发, 分析这些原始变量的基本结构, 使得每个变量可以用少数几个不可观测的公因子变量来表述, 它属于数据降维的统计方法, 可以利用公因子来解释众多变量所代表的主要信息, 达到简化数据结构, 方便研究的目的, 流程图见图 2。

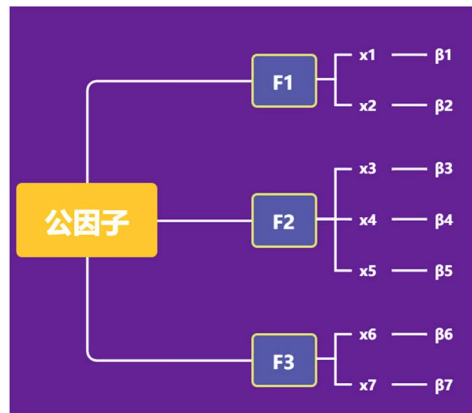


Figure 2. Factor analysis
图 2. 因子分析

因子分析的数学表达式可概括为为: $X = AF + B$, 即:

$$\begin{cases} x_1 = \alpha_{11}f_1 + \alpha_{12}f_2 + \cdots + \alpha_{1k}f_k + \beta_1 \\ x_2 = \alpha_{21}f_1 + \alpha_{22}f_2 + \cdots + \alpha_{2k}f_k + \beta_2 \\ \vdots \\ x_p = \alpha_{p1}f_1 + \alpha_{p2}f_2 + \cdots + \alpha_{pk}f_k + \beta_p \end{cases}$$

模型中, 向量 $X = (x_1, x_2, \dots, x_p)$ 是可观测随机向量, 称之为原始观测变量。 $F = (f_1, f_2, \dots, f_k)$ 是 X 的公共因子, 他们是均值为 0, 方差为 1 的相互独立的不可观测向量, 即各因子是相互独立的, 公共因子所代表的实际含义必须结合实际问题来界定。其中 $A = (a_{ij})$ 成为因子载荷, 它反映第 I 个变量在第 j 个因子上的负荷大小, 在因子载荷矩阵 A 中, 公共因子的方差贡献率对因子分析结果的解释十分重要。

2.3.3. 聚类分析

聚类分析是依据研究对象的特征, 对其进行分类, 已达到减少研究对象的数量的目的。聚类分析的思想是将一组数据分到不同的类别, 因此, 属于同一个类别中的对象有很大的相似性, 而不同类别中对象倾向于不相似。分类过程中无需进行事先分类, 聚类分析能够从样本数据出发, 自动生成不同类别。

2.3.4. 判别分析

判别分析要解决的问题是, 在已知历史上用某些方法已经把研究对象分成若干组(亦称类或总体)的情况下, 来判定新的观测样品应归属的组别。判别分析是多元数据分析的重要方法之一, 通常被解释变量是非数值变量, 解释变量是非数值变量。

3. 结果与分析

3.1. 主成分分析

利用 R 软件将数据进行标准化, 并作出相关系数矩阵。所得相关系数矩阵见如下表 2:

Table 2. Correlation matrix
表 2. 相关系数矩阵

	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1	-0.292	0.435	0.375	-0.492	0.453	0.464	-0.518	-0.216
X2	-0.292	1	-0.596	0.48	0.698	0.297	-0.678	0.601	-0.563
X3	0.435	-0.596	1	0.008	-0.431	0.183	0.912	-0.339	0.227
X4	0.375	0.48	0.008	1	0.526	0.916	0.008	0.429	-0.59
X5	-0.492	0.698	-0.431	0.526	1	0.402	-0.477	0.956	-0.359
X6	0.453	0.297	0.183	0.916	0.402	1	0.158	0.35	-0.535
X7	0.464	-0.678	0.912	0.008	-0.477	0.158	1	-0.405	0.323
X8	-0.518	0.601	-0.339	0.429	0.956	0.35	-0.405	1	-0.33
X9	-0.216	-0.563	0.227	-0.59	-0.359	-0.535	0.323	-0.33	1

从表 2 中可以看出九个原始变量指标存在较高的相关性，因此可以采用较少的综合指标进行主成分分析，对各变量中各类信息加以解释[3]。然后，用 R 软件进行主成分分析，将数据降维，提取出主要的因素。结果整理如表 3、表 4，图 3：

Table 3. Operation results
表 3. 运行结果

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
Standard deviation	2.3044	1.1700	1.0588	0.8127	0.5012	0.3824	0.2901	0.2235	0.0884
Proportion of Variance	0.5900	0.1521	0.1246	0.0734	0.0279	0.0162	0.0094	0.0055	0.0009
Cumulative Proportion	0.5900	0.7421	0.8667	0.9300	0.9678	0.9842	0.9936	0.9991	1.0000

Table 4. Load list
表 4. 载荷表

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8	Comp.9
X1	0.201	0.698	0.114	0.294	0.000	0.533	0.225	0.133	0.135
X2	-0.373	0.153	0.000	-0.257	-0.842	0.142	0.137	-0.124	0.000
X3	0.324	0.193	-0.534	0.000	-0.250	-0.516	0.325	0.345	0.128
X4	-0.387	0.143	0.000	0.446	-0.112	-0.152	-0.522	0.537	-0.352
X5	-0.381	0.000	-0.436	0.000	0.000	0.180	-0.170	0.000	0.747
X6	-0.371	0.193	0.000	0.508	0.125	-0.352	0.308	0.411	0.108
X7	0.353	0.142	-0.482	0.009	-0.163	0.159	-0.542	0.452	0.000
X8	-0.341	-0.104	-0.523	-0.135	0.259	0.368	0.303	0.24	-0.53
X9	0.206	-0.600	0.000	0.602	-0.312	0.299	0.208	0.000	0.000

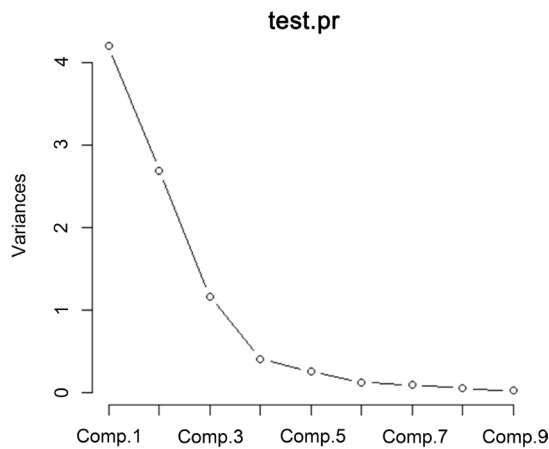


Figure 3. Scree plot
图 3. 碎石图

由图显示，前三个结果的累计贡献率接近 87%，因此可以舍去另外六个主成分，达到降维的目的。由载荷图可以得到各主成分表达式：

$$F1 = 0.201x_1 - 0.373x_2 + 0.324x_3 - 0.387x_4 - 0.381x_5 - 0.371x_6 + 0.353x_7 - 0.341x_8 + 0.206x_9;$$

$$F2 = 0.698x_1 + 0.153x_3 + 0.193x_3 + 0.143x_4 + 0.193x_6 + 0.142x_7 - 0.104x_8 - 0.600x_9;$$

$$F3 = 0.114x_1 - 0.534x_3 - 0.436x_5 - 0.482x_7 - 0.534x_8$$

将原始数据标准化后的变量值 x_1, x_2, \dots, x_9 带入主成分表达式，可以计算出每个城区在经济发展研究中各主成分的得分，充分利用各主成分的方差贡献率，以得到更加准确的综合评价公式[4]，再以各主成分的加权方差贡献率为权重，得到武汉区经济发展的综合评价公式为：

$$Y = (F1 * 0.5900 + F2 * 0.1521 + F3 * 0.1246) / 0.8667$$

根据综合得分对各城区经济发展进行排名，见下表 5：

Table 5. Scores of principal component and comprehensive ranking
表 5. 各主成分得分及综合排名

	F1	第一主成分 得分排名	F2	第二主成分 得分排名	F3	第三主成分 得分排名	Y	综合得分 排名
江汉区	1.7014	1	-0.32154	4	-0.57124	11	0.96	1
江岸区	1.3046	2	0.07245	5	0.34561	7	0.83	2
洪山区	0.9628	4	-0.02361	7	0.65232	4	0.68	3
武昌区	1.0986	3	0.0568	6	-0.56442	10	0.57	4
汉阳区	0.3254	6	0.42063	3	0.96001	3	0.43	5
青山区	-1.3212	13	2.88634	1	-0.52883	9	0.24	6
东西湖区	0.23538	7	0.44827	2	0.52346	5	0.12	7
硚口区	0.6213	5	-0.29367	8	-1.12517	12	0.08	8
江夏区	-0.76684	9	-0.41712	9	1.41523	1	-0.23	9
黄浦区	-0.59846	8	-0.84126	12	1.29634	2	-0.38	10
新洲区	-0.85956	10	-0.80123	11	0.34999	6	-0.64	11
蔡甸区	-0.96077	11	-0.6321	10	0.02436	8	-0.75	12
汉南区	-1.06345	12	-1.14257	13	-2.2834	13	-1.32	13

从表中综合得分排名来看,可以看出各中心城区排名比远郊城区排名靠前,表中不仅给出了综合得分排名,也给出了每一个主成分排名,从中可以发现,第一主成分能最大限度的反映样品间的差异,是影响各城区经济发展排名的最关键因素。

3.2. 因子分析

3.2.1. 利用 SPSS 进行因子分析

因子分析法是从原始变量的相关矩阵 R 出发,建立因子模型,将一些具有错综复杂关系的变量综合为少数几个新的不可观测而又彼此独立的综合指标(称为“因子”),这几个相互无关系的因子能够结实原始变量的绝大部分信息,并且相关性较低,基于这种思想,本文用因子分析来研究武汉个城区经济发展情况。

利用 SPSS 25 进行因子分析(结果见表 6、表 7、表 8、表 9、表 10、图 4),首先将原始数据标准化,得到标准化矩阵和相关系数矩阵,下表是 KMO 和 Barlett 的检验结果,从中可以看出 KMO 的值为 0.611,大于最低标准 0.5,表明该数据适合做因子分析。Barlett 的球形度检验的显著性水平小于 0.01,表明标准化原始变量值的相关系数矩阵间有共同元素存在,适合做因子分析。

Table 6. KMO and Barlett test

表 6. KMO 和 Barlett 的检验

KMO 和巴特利特检验		
KMO 取样适切性量数。		0.611
	近似卡方	97.554
巴特利特球形度检验	自由度	36
	显著性	0.000

Table 7. Common factor variance

表 7. 公因子方差

公因子方差		
	初始	提取
全社会固定资产投资	1.000	0.896
社会消费品零售额	1.000	0.773
从业人员平均人数	1.000	0.929
分区生产总值	1.000	0.825
城镇居民人均可支配收入	1.000	0.986
公共财政总收入	1.000	0.784
主营业务收入	1.000	0.950
城镇居民人均消费支出	1.000	0.939
工业总产值	1.000	0.720

提取方法:主成分分析法。

由表 7、表 8 公因子方差值、总解释方差和图 4 碎石图可知,前三个因子的累计贡献率已达 86.668%,即选取这三个因子作为主因子就可以解释原始变量 86.668%的信息,也就是说,通过选择三个因子进行分析,就能充分反应武汉市城区经济发展的综合水平。

Table 8. Total variance interpretation
表 8. 总方差解释

成分	总方差解释								
	初始特征值			提取载荷平方和			旋转载荷平方和		
	总计	方差百分比	累积%	总计	方差百分比	累积%	总计	方差百分比	累积%
1	5.310	59.004	59.004	5.310	59.004	59.004	3.124	34.710	34.710
2	1.369	15.209	74.213	1.369	15.209	74.213	2.956	32.839	67.550
3	1.121	12.455	86.668	1.121	12.455	86.668	1.721	19.118	86.668
4	0.661	7.339	94.007						
5	0.251	2.791	96.798						
6	0.146	1.625	98.423						
7	0.084	0.935	99.358						
8	0.050	0.555	99.913						
9	0.008	0.087	100.000						

提取方法：主成分分析法。

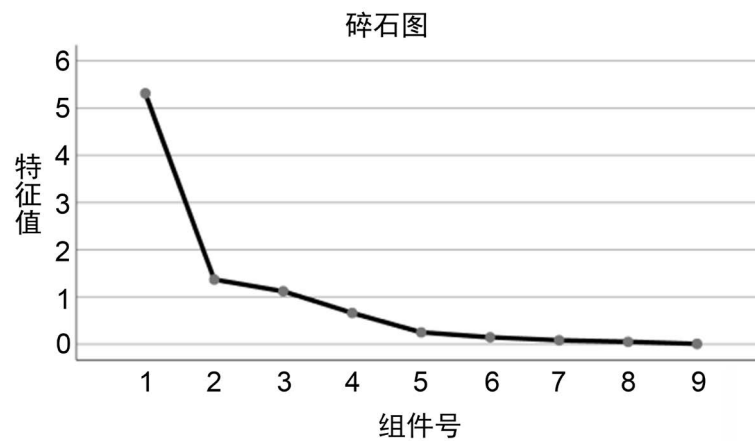


Figure 4. Scree plot
图 4. 碎石图

为了更加明确各因子的经济含义，下面对因子载荷矩阵进行正交旋转，使得旋转后的因子载荷值向 0 和 1 靠近，得到的正交因子旋转成分矩阵见表 9。

Table 9. Rotating component matrix
表 9. 旋转成分矩阵

	旋转后的成分矩阵 ^a		
	成分 1	成分 2	成分 3
x1	-0.435	0.348	0.765
x2	0.588	-0.589	0.283
x3	-0.166	0.946	0.078

Continued

x4	0.622	-0.602	0.274
x5	0.958	-0.262	-0.002
x6	0.651	-0.508	0.319
x7	-0.251	0.941	0.017
x8	0.955	-0.145	-0.085
x9	-0.325	0.223	-0.752

提取方法：主成分分析法；旋转方法：凯撒正态化最大方差法；a. 旋转在 5 次迭代后已收敛。

通过因子旋转，容易看出旋转之后的三个主因子均有比较典型的代表变量。第一个主因子在指标 X8：城镇居民人均消费支出、X5：城镇居民人均可支配收入、X6：公共财政总收入具有较高载荷和解释能力，该类指标主要反映了经济发展的总量以及人民生活水平的状况，故将其命名为综合经济实力因子，第二主因子在 X2：社会消费品零售额、X3：从业人员平均人数、X7：主营业务收入上具有较高载荷和解释能力，该指标主要反映地区的收入和劳动力水平，可以将其命名为收入和劳动力水平发展因子。第三主因子在 X1：全社会固定资产投资、X9：工业总产值上具有较高载荷和解释能力，该指标主要反映了地区投资规模水平和工业发展水平状况，可以将其命名为投资规模以及工业发展水平因子。

在 SPSS 中采用主成分方法提取主因子，下面输出因子得分系数矩阵 H，见表 10：

Table 10. Component score coefficient matrix

表 10. 成分得分系数矩阵

成分得分系数矩阵			
	成分		
	1	2	3
x1	-0.240	0.130	0.549
x2	0.054	-0.109	0.176
x3	0.228	0.492	0.086
x4	0.068	-0.107	0.168
x5	0.410	0.172	-0.015
x6	0.109	-0.039	0.203
x7	0.193	0.455	0.044
x8	0.464	0.230	-0.069
x9	0.038	-0.058	-0.516

提取方法：主成分分析法；旋转方法：凯撒正态化等量最大法；组件得分。

由 $F = H'X - 1B'$ 计算其因子得分，其中 $F = (F1, F2, F3)$ ， F_i 为第 i 个公因子的得分，以各因子的加权方差贡献率作为权重，最后得到武汉市城区经济发展研究的综合评价公式为：

$$Y = (F1 * 0.59004 + F2 * 0.15209 + F3 * 0.12455) / 0.86668$$

根据综合得分对经济发展进行排名，得到表 11 如下：

Table 11. Factor score and comprehensive score
表 11. 主因子的得分以及综合得分表

	F1	第一主因子 得分排名	F2	第二主因子 得分排名	F3	第三主因子 得分排名	Y	综合排名
青山区	0.67806	5	2.82153	1	-1.31211	12	1.13	1
江汉区	1.62743	1	-0.84021	12	-0.23189	9	0.67	2
江岸区	1.0542	2	-0.61234	10	0.61533	5	0.53	3
汉阳区	0.22333	7	0.41234	4	0.92345	3	0.41	4
洪山区	0.82345	4	-0.4263	9	0.82316	4	0.39	5
武昌区	0.98742	3	-0.0003	5	-0.32189	10	0.28	6
东西湖区	0.0327	8	0.57476	2	0.51207	6	0.26	7
江夏区	-0.98234	9	0.4715	3	1.15876	2	-0.16	8
硚口区	0.35182	6	-0.77476	11	-0.92627	11	-0.24	9
黄陂区	-1.07169	10	-0.0782	7	1.21641	1	-0.37	10
新洲区	-1.2371	13	-0.09692	8	0.32643	7	-0.58	11
蔡甸区	-1.23101	11	-0.0094	6	-0.00087	8	-0.62	12
汉南区	-1.20121	12	-1.03121	13	-2.21252	13	-1.19	13

表 11 中不仅有各城区综合排名，还可以通过主因子得分看出各城区综合经济实力、工业发展水平、投资规模上的排名情况，有利于更好的对十三个城区的经济发展水平进行评价。由表可知，综合因子得分在平均水平之上的城区有 7 个，依次为青山区、江汉区、江岸区、汉阳区、洪山区、武昌区和东西湖区。

3.2.2. 利用 R 软件进行因子分析

利用 R 软件进行因子分析，结果如图 5 所示：

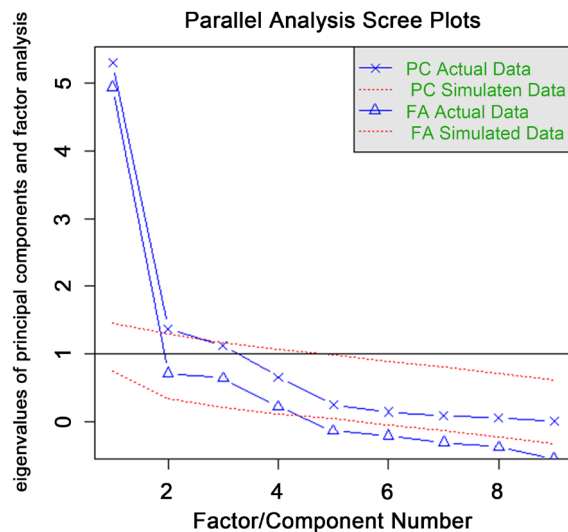


Figure 5. Factor analysis
图 5. 因子分析图

从图 5 中可以看出，保留三个主成分即可。将所得结果进行统计整理，所得数据如下表 12、表 13、表 14：

Table 12. Output results
表 12. 输出结果

	ML1	ML2	ML3
SS loadings	5.21	1.05	0.78
Proportion Var	0.58	0.12	0.09
Cumulative Var	0.58	0.70	0.78
Proportion Explained	0.74	0.15	0.11
Cumulative Proportion	0.74	0.99	1.00

Table 13. Output results
表 13. 输出结果

	ML1	ML2	ML3	u2	com
V1	-0.53	0	0.54	0.4339	2
V2	0.8	-0.03	0.09	0.3453	1
V3	-0.75	0.52	0.03	0.1639	1.8
V4	0.87	0.02	0.48	0.0232	1.6
V5	0.88	0.46	-0.03	0.0048	1.5
V6	0.83	0.08	0.43	0.1194	1.5
V7	-0.83	0.56	0.05	0.0050	1.8
V8	0.81	0.50	-0.24	0.0346	1.9
V9	-0.40	-0.01	-0.09	0.8301	1.1

Table 14. Output results
表 14. 输出结果

	ML1	ML2	ML3
Correlation of scores with factors	1.00	1.00	0.97
Multiple R square of scores with factors	1.00	0.99	0.94
Minimum correlation of possible factor scores	0.99	0.98	0.87

fa 函数中, 第一个参数为数据, 第二个参数说明要保留 3 个主成分, 第三个参数为旋转方法, 为 none, 即不进行主成分旋转, 第四个参数表示提取公因子的方法为最大似然法。上图表明输出结果与主成分输出结果基本一致, 并可看出该数据选取三个主成了解释了原始变量 80% 的方差。

因子分析的目的不仅是要找出公因子, 还要能解释每一个公因子所代表的实际含义, 未旋转的因子载荷矩阵可能某一行有多个载荷系数比较大, 也有可能某一列有多个载荷系数比较大, 这表明公因子可能解释了多个变量的信息, 导致该公因子代表性不强, 为此, 我们可以将初始公因子进行线性组合, 也就是进行因子旋转, 经过适当变换, 解决上述问题[5], 利用 R 软件所得结果如表 15、表 16、表 17 所示:

Table 15. Output results
表 15. 输出结果

	ML1	ML2	ML3	h2	u2	com
V1	0.02	0.38	-0.65	0.57	0.4339	1.6
V2	0.59	-0.44	0.35	0.65	0.3453	2.5
V3	-0.33	0.83	-0.17	0.84	0.1639	1.4
V4	0.91	-0.36	0.13	0.98	0.0232	1.3

Continued

V5	0.69	-0.09	0.71	1	0.0048	2
V6	0.87	-0.30	0.17	0.88	0.1194	1.3
V7	-0.36	0.91	-0.20	1.00	0.0050	1.4
V8	0.51	-0.06	0.84	0.97	0.0346	1.7
V9	-0.33	0.16	-0.16	0.17	0.8301	2

Table 16. Output results
表 16. 输出结果

	ML1	ML2	ML3
SS loading	3.04	2.11	1.88
Proportion Var	0.34	0.23	0.21
Cumulative Var	0.34	0.57	0.78
Proportion Explained	0.43	0.3	0.27
Cumulative Proportion	0.43	0.73	1.00

Table 17. Output results
表 17. 输出结果

	ML1	ML2	ML3
Correlation of (regression) scores with factors	0.98	1.00	0.98
Multiple R square of possible factor scores	0.97	0.99	0.97
Minimum correlation of possible factor scores	0.93	0.98	0.93

由图 5 可以看出，两个因子的方差比例不变，但在各观测值的载荷上发生了改变。
下图 6 是利用 R 软件进行数据可视化的结果：

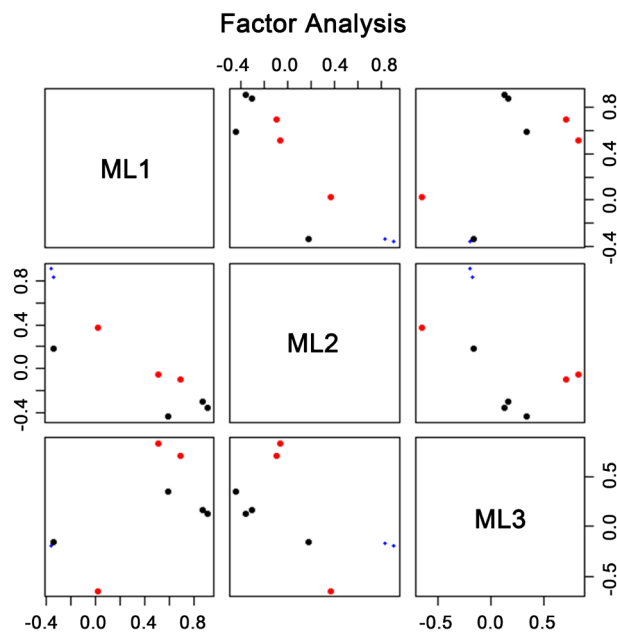


Figure 6. Data visualization
图 6. 数据可视化

3.3. 结果检验

运用 SPSS 25 对主成分分析和因子分析所得排名进行 Spearman 相关性检验，结果见下表 18:

Table 18. Correlation test
表 18. 相关性检验

		相关性		
			主成分分析排名	因子分析排名
主成分分析排名	相关系数		1.000	0.896**
	Sig.(双尾)		0.0	0.000
	N		13	13
斯皮尔曼 Rho	相关系数		0.896**	1.000
	Sig.(双尾)		0.000	0.0
	N		13	13

**在 0.01 级别(双尾), 相关性显著。

由表可知, 相关系数为 0.896, 说明两者之间存在较强的正相关, 同时 sig 值为 0, 通过显著性检验, 因此我们可以认为基于主成分分析和因子分析对武汉市城区经济发展研究的排名结果具有一致性, 其综合评价模型公式等价。

3.4. 聚类分析

将所得标准化数据利用 R 软件按照最短距离法、重心法和 Ward 法进行聚类分析, 得到分类结果(见图 7, 图 8, 图 9), 运行结果表明三种方法聚类效果高度吻合。

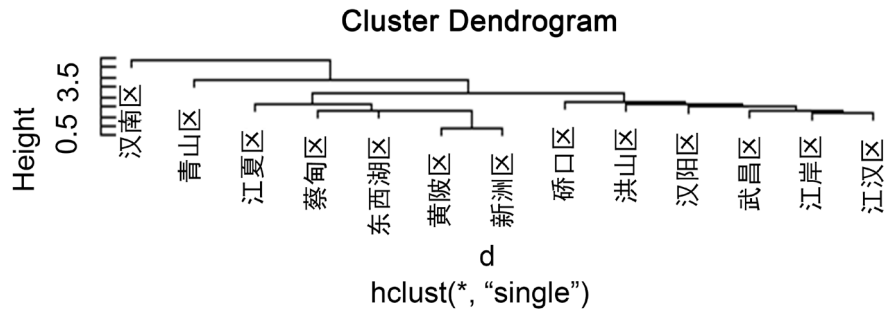


Figure 7. Shortest distance method
图 7. 最短距离法

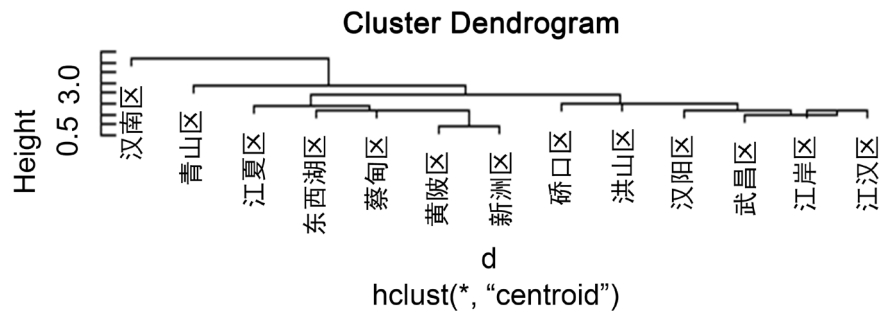


Figure 8. Centroid method
图 8. 重心法

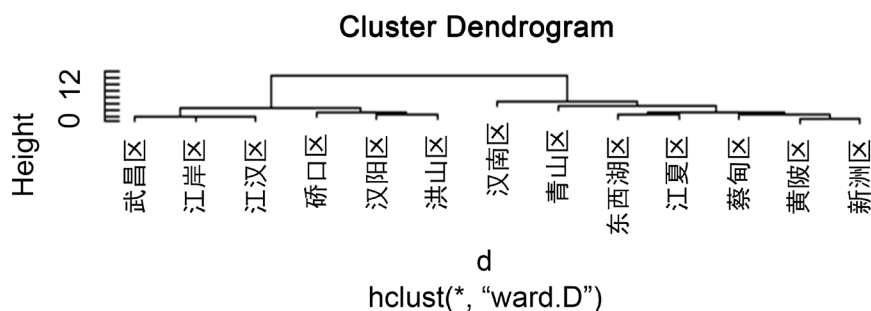


Figure 9. Ward method
图 9. Ward 法

从树状图中可以将城区进行划分类别，以聚类结果来看，武汉市 13 个城区其城区经济发展结构存在较明显的差异，结合原始数据，具体表现为青山区在主营业务收入、工业总产值、从业人员平均人数上远远高于其他城区，说明青山区工业发展势头很好，汉南区在社会消费品零售额、生产总值、财政总收入上明显低于其他城区，由此可见，汉南区总体经济实力发展较弱，剩下的其他城区发展结构较一致，再细分观察，发现中心城区与中心城区之间有更相似，远郊城区与远郊城区之间也更为相似[4]。

3.5. 判别分析

将聚类结果分成 4 类，第一类：青山区，汉南区；第二类，武昌区、江岸区、江汉区；第三类：硚口区、汉阳区、洪山区；第四类：蔡甸区、黄陂区、新洲区、东西湖区、江夏区。下面利用 R 软件做贝叶斯判别和距离判别(见表 19、表 20)，检验聚类分析结果的合理性，其中数字 1~4 表示类别，相同的数字代表同一个类别，结果表明判别分析误判率为 0.07，即约 93%的对象能被分到正确的类中，判别分析的结果与聚类分析几乎吻合，因此聚类分析是合理的。

Table 19. Distance discrimination results
表 19. 距离判别结果

城区	汉南区	青山区	武昌区	江汉区	江岸区	蔡甸区	东西湖区	黄陂区	新洲区	江夏区	硚口区	汉阳区	洪山区
类别	1	1	2	2	2	3	3	3	3	3	4	2	4

Table 20. Bayes discrimination results
表 20. 贝叶斯判别结果

城区	汉南区	青山区	武昌区	江汉区	江岸区	蔡甸区	东西湖区	黄陂区	新洲区	江夏区	硚口区	汉阳区	洪山区
类别	1	1	2	2	2	3	3	3	3	3	4	2	4

4. 讨论

4.1. 总结

本文在内容上运用了主成分分析、因子分析、聚类分析和判别分析等多种统计方法综合评价武汉市城区经济发展。用主成分分析和因子分析给出综合评价模型，并采用 Spearman 法对主成分分析和因子分析的结果进行一致性检验。再利用聚类分析研究出武汉市经济结构，并用判别分析检验其分类的合理性，进而分析出了各城区的经济发展状况。

4.2. 政策建议

经过以上研究,从宏观视角上对武汉市城区经济发展有了一个比较全面的了解,可以依据这个结果对武汉市城区经济发展进行更科学的宏观调控,建议可以从下述几个方面加以改善[6]。

1) 对城区经济较发达地区,合理优化其产业结构,推进城区产业均衡,提高江汉区、江岸区、洪山区、武昌区的工业发展水平,同时应当注重科技创新,在武汉高新技术开发区、武汉光谷等国家级高新技术产业园区的快速发展中,充分利用其科技优势资源,重点培养和发展生物制药、互联网、光缆通讯、精密仪器制造等产业。

2) 对城区经济一般发达地区,逐步缩小城区发展不平衡,全面协调可持续发展。这类城区多为远郊城区,应当充分发挥中心城区经济发展的辐射作用,带动远郊城区的经济发展,引导城区经济逐步走向一体化,因此政府应加快推进工业化进程,把远城区的发展看作是武汉新型工业化和新城镇化的战略支点,同时要促进消费,提高经济活跃度。

3) 打造城区特色,借助城区优势,发展城区重点经济。市政府应当提高对各城区自身发展的条件和可以利用的资源认识,建立不同城区在不同资源领域、不同技术方向上的竞争优势。如青山区继续推动武钢、武石化升级改造,建设绿色低碳循环发展的一流生态化工园区。武昌区进一步发展东湖生态旅游风景区,力求打造国际知名、国内一流的文化旅游生态风景名胜区等。另外要加大对经济实力较弱的汉南区的资金投入,扶植其优势产业,带动其经济发展。总之各城区应当相互借鉴、取长补短,将自己的优势产业做大做强,更好地促进全市经济稳步快速地发展[1]。

参考文献

- [1] 文乐. 基于多元统计分析的武汉市城区经济发展研究[D]: [硕士学位论文]. 武汉: 华中师范大学, 2016.
- [2] 樊长科, 吴雨. 广西各地市区域经济发展水平分析——基于系统聚类法和主成分分析法[J]. 安徽农业科学, 2010, 38(14): 7620-7622.
- [3] 杨曼琳. 基于主成分分析法的区域经济活力影响因素[J]. 中国科技信息, 2020(12): 106+108.
- [4] 姜天龙, 刘子宜, 单魁贤. 吉林省县域经济社会发展的综合评价及分析——基于主成分分析法[J]. 农村经济与科技, 2014, 25(10): 120-122.
- [5] 张小爱. 河北省区域经济的统计分析[J]. 统计与管理, 2018(12): 39-42.
- [6] 刘孟轩. 宏观经济背景下统计分析相关问题及其对策研究[J]. 现代交际, 2020(10): 42-43.