

基于我国各地区财政支出数据的因子分析与对比

1. 研究背景

在中国，财政支出是国家为实现其各种职能，由财政部门按照预算计划，将国家集中的财政资金向有关部门和方面进行支付的活动，这是其服务职能的保证与安排。作为政府分配的重要组成部分，是政府进行资源配置、实施宏观调控、促进产业合理有效配置的主要手段。随着改革开放进行，支出结构的每一次调整为宏观经济发展提供了方向指导，研究并对比分析我国各地区支出结构与差异具有重要意义，并为经济发展的不同阶段进行更合理提出调整与优化建议。

我国各地区的财政支出的指标体系此处为：一般公共服务、国防、公共安全、科学、教育等 14 项，为提高研究意义，选择中国发展的关键时期，于统计局官网上下载《2008 年统计年鉴》，选取当年我国 31 个省市自治区的各项财政支出总数据，并根据研究需要，预整理异常数据后结果见附件。目的在于分析不同地区下各类支出项目的结构，研究各地区财政支出的结构差异。

为软件处理与研究方便，构建基本的指标体系，变量表示如表 1。

表 1 指标——字母符号变量表示

X1	X2	X3	X4	X5	X6	X7
一般公共 服务	国防	公共 安全	教育	科学 技术	文体传 媒	保障 就业
X8	X9	X10	X11	X12	X13	X14
医疗卫生	环境 保护	社区 事务	农林水 事务	交通 运输	工商金 融等	其它 支出

对数据进行基本的统计分析，通过均值(mean)、最值(max、min)、方差(std)等，直接对比财政支出指标在各地区间的差异。选择性分析个别突出指标，采用 R 型因子分析方法，对指标体系进行降维，综合为少数几个公共因子，借此对各地区财政支出结构进行深入比较并分类。再求出各地区的因子得分并排名，分析各地区财政支出分配差异，为更合理安排提供依据。

2. 我国各地区财政支出的基本统计分析

本文采用 SPSS 软件处理并分析数据，思路及过程资料：附件 1 显示 SPSS

运行过程，附件 2 表示输出结果，附件 3 则是原始数据及后续处理整合。

在 SPSS 中对预处理后的经济数据进行简单的描述统计，分析了各个指标的最大值、最小值、平均值以及标准差等，初步了解了这些数据的大小关系等，反映了各地区不同指标下的支出结构差异，具体结果见表 2。

分析可得到，在一般公共服务、教育、保障就业、社区服务等 4 个项目均值较高，说明各地区在这些项目上支出明显，如北京、上海、广东等地区多项指标支出最高。但同时，这些指标的方差较大，最高达 1303120，反映了地区间支出结构差异明显，如宁夏、西藏等省市明显较低。从原始数据中分析得到，同一个地区各项支出同样存在明显差异，如安徽省在教育支出最大 2129665 元，而在国防上仅有 23101 元支出。

此外，对各地区不同指标的财政支出数据进行标准化后分组，进一步分析各地区财政支出分布情况，根据 3σ 准则，得到异常组的比例为 3.18%，大于理论值 3.0%，表明了我国各地区财政支出存在不均衡现象，具有研究必要性。

表 2 各地区财政支出描述性统计分布表

3. 我国各地区财政支出结构的因子分析

3.1 因子分析

对问题中指标的降维分析一般可采取主成分分析、因子分析。在主成分分析中，只要变量之间存在一定的相关性，前几个主成分往往能够较好地达到降维的目的。然而，多数情况下，对变量仅仅进行降维还是不够的，还必须对主成分给出符合实际背景和意义的解释。进行这种解释往往是主成分分析的困难之处。因子分析方法是主成分分析方法的推广和发展，它也是一种重要的降维方法。与主成分分析相比，因子分析较为灵活，体现在因子旋转上，这种灵活性使得变量在降维之后更易得到解释，这是因子分析比主成分分析更广泛应用的原因之一。

因子模型为： $x = \mu + Af + \varepsilon$ ，其中矩阵 A 被称为因子载荷矩阵。

$$\begin{cases} x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 \\ x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2 \\ \cdots \\ x_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + \varepsilon_p \end{cases}$$

因子分析的可行性分析

在基本的描述统计基础上，认识到了地区财政支出差异，为更深入地理解我国财政支出结构复杂、指标多且杂、相互依赖的关系，对整体进行因子分析。在基本的分析假设下，要求原来的各个变量间具有较高的相关性，即存在信息重叠。在较充足的样本容量下，通过分析指标间相关关系以及大小来判断因子分析的必要性，SPSS 运行预处理数据，得到 14 个指标间相关系数矩阵如下表 3 所示。

相关系数矩阵表明了各个指标之间存在较大的正相关关系，且多数大于 0.6，一定程度上说明了这些指标间关系密切，内部结构明显，则在分析经济问题时，可以利用少数代表性变量去代替描述其他多个变量，并称之为因子，具有可分析条件。

表3 Correlation Matrix

Corr elation	X 1	X 2	X 3	X 4	X 5	X 6	X 7	X 8	X 9	X 10	X 11	X 12	X 13	X 14
X1	1
	.000	774	907	969	626	792	714	877	512	650	912	615	644	618
X2	.	1
	774	.000	785	740	539	576	607	691	216	469	703	564	439	507
X3	.	.	1
	907	785	.000	948	847	852	641	907	273	738	761	497	649	841
X4	.	.	.	1
	969	740	948	.000	760	873	685	936	417	757	858	545	730	714
X5	1
	626	539	847	760	.000	877	491	808	044	852	440	200	709	885
X6	1
	792	576	852	873	877	.000	579	922	360	813	708	414	743	696
X7	1
	714	607	641	685	491	579	.000	674	561	603	694	363	758	489
X8	1
	877	691	907	936	808	922	674	.000	432	721	835	565	713	709
X9	1
	512	216	273	417	044	360	561	432	.000	195	650	631	343	007
X10	1
	650	469	738	757	852	813	603	721	195	.000	482	206	861	709
X11	1	.	.	.
	912	703	761	858	440	708	694	835	650	482	.000	775	562	393
X12	1	.	.
	615	564	497	545	200	414	363	565	631	206	775	.000	248	199
X13	1	.
	644	439	649	730	709	743	758	713	343	861	562	248	.000	576
X14	1
	618	507	841	714	885	696	489	709	007	709	393	199	576	.000

KMO 统计量与 Bartlett's 球性检验

为了更精确的表达因子分析的可行性，KMO 统计量可用于探查变量间的偏相关性，如果 KMO 统计量接近 1，主成分分析的效果好。一般认为，KMO 大于 0.9 效果最佳，0.7 以上时效果尚可，0.6 时效果很差，0.5 以下时不适宜进行因子分析。Bartlett's 球性检验统计量则用于检验各变量是否独立，如果结论为不拒绝

原假设，则说明变量可能各自独立提供信息，不适宜进行因子分析。

继续运行 SPSS，得到 KMO 和 Bartlett 的检验如表 4 所示。

表 4 KMO 检验表

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.788
Bartlett's Test of Sphericity	Approx. Chi-Square	678.52
		3
	Df	91
	Sig.	.000

其中，KMO值为0.788，根据判断标准，原有变量基本满足因子分析条件。Bartlett's 球度统计量的观测值为736.059，相应的概率为 $p = 0$ 小于显著水平 $\alpha = 0.05$ ，所以应拒绝原假设，认为相关系数矩阵与单位矩阵间有显著差异，进一步说明14个财政支出指标相关性强，适合作因子分析。

因子提取

运行SPSS时，参数估计采用主成分方法提取因子，根据每个因子解释原有变量总方差的数值大小与累计方差贡献率判断可取的公共因子，具体运行结果见表5。

表5 Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.424	67.311	67.311	9.424	67.311	67.311
2	1.959	13.996	81.307	1.959	13.996	81.307
3	.949	6.776	88.083			
4	.556	3.970	92.053			
5	.323	2.306	94.359			
6	.292	2.085	96.444			
7	.191	1.366	97.810			
8	.149	1.068	98.877			
9	.066	.474	99.351			
10	.039	.278	99.629			
11	.024	.169	99.797			
12	.015	.109	99.907			
13	.009	.062	99.968			
14	.004	.032	100.000			

从表中可看出，第一部分表示了各因子解释的方差百分比，第一列表示每个因子的特征值，第二列反映的是因子的方差贡献率，第三列则是累计方差贡献率。且大于1的特征根共有2个，第一个因子的特征根值为9.424，能解释原有变量总方差的67.311%，第2个因子的特征根值为1.959，对原有变量的方差解释能力为12.996%，前两个因子对所有变量的解释达到81.307%，这说明这两个因子基本包含了全部财政支出指标所具有的绝大部分信息，因子分析效果较想。

SPSS中运行同时，为了给提取的两个因子命名，因子载荷矩阵实施正交旋转，选择最大方差法实施正交旋转得到旋转载荷因子矩阵，使各个特殊因子能够得到更好的解释，具体结果见表6。

根据提取的旋转因子载荷矩阵，可得到关于因子与14个变量间的关系载荷函数如下：

$$\begin{cases} x'_1 = 0.055 f_1 + 0.704 f_2 + \varepsilon_1 \\ x'_2 = 0.417 f_1 + 0.764 f_2 + \varepsilon_2 \\ \dots \\ x'_{14} = 0.851 f_1 + 0.297 f_2 + \varepsilon_{14} \end{cases}$$

系数的大小及正负反映了各个指标在两个因子上的表现。其中，第一个因子在公共安全、科学技术、社区事务、医疗卫生、教育等指标上有较大载荷，表示为地区发展需要性支出，可解释为“地区发展贡献”因子；而第二个因子在一般公共服务、国防、交通运输等指标上载荷较大，说明其主要解释这几个变量，而这些指标基本表示为各地区公共基础服务支出，可解释为“地区服务贡献”因子。

因子得分

进一步在SPSS运行中，选择回归法计算因子得分系数矩阵，结果如表7所示。据此判断每个地区在提取的两个因子上的得分并进行排序，分别比较各地区在地区发展贡献、地区服务贡献上的排名情况。各地区在提取的两个因子上的得分函数如下：

$$\begin{cases} f_1 = 0.099x'_1 + 0.081x'_2 + 0.101x'_3 + 0.103x'_4 + 0.087x'_5 + 0.096x'_6 + 0.082x'_7 + 0.101x'_8 \\ \quad + 0.049x'_9 + 0.086x'_{10} + 0.090x'_{11} + 0.061x'_{12} + 0.084x'_{13} + 0.080x'_{14} \\ f_2 = 0.099x'_1 + 0.069x'_2 - 0.063x'_3 + 0.010x'_4 - 0.262x'_5 - 0.083x'_6 + 0.080x'_7 + 0.004x'_8 \\ \quad + 0.364x'_9 - 0.199x'_{10} + 0.237x'_{11} + 0.330x'_{12} - 0.101x'_{13} - 0.249x'_{14} \end{cases}$$

这2个因子变量间是线性无关的，利用它们可以对各地区的财政支出结构作出更深入的分析。在SPSS运行原数据后，得到各个地区在得分函数下的总得分及排名情况，结果如表8所示。

表6 Rotated Component Matrix

	Component	
	1	2
X1	.055	.704
X2	.417	.764
X3	.752	-.633
X4	.699	.688
X5	.937	-.241
X6	.791	-.408
X7	.533	.279
X8	.699	.589
X9	-.055	.359
X10	.900	.091
X11	.323	.762
X12	-.032	.821
X13	.797	.048
X14	.851	.297

表7 Component Score Coefficient Matrix

	Component	
	1	2
X1	.099	.099
X2	.081	.069
X3	.101	-.063
X4	.103	.010
X5	.087	-.262
X6	.096	-.083
X7	.082	.080
X8	.101	.004
X9	.049	.364
X10	.086	-.199
X11	.090	.237
X12	.061	.330
X13	.084	-.101
X14	.080	-.249

Extraction Method: Principal Component Analysis.

Component Scores.

表8 因子得分表

地区	因子 1 得分	因子 1 特征比	因子 2 得分	因子 2 特征比	综合得分	排名
北京	0.813066032	0.8279	-1.26751	0.172099	0.4550016	8
天津	-0.926078394	0.8279	-1.532428	0.172099	-1.03043	27
河北	0.275659521	0.8279	0.3863871	0.172099	0.2947156	10
山西	-0.266152632	0.8279	0.4299586	0.172099	-0.146353	18
内蒙古	-0.165272375	0.8279	1.1567855	0.172099	0.0622521	15
辽宁	0.73408717	0.8279	-0.009603	0.172099	0.606099	7
吉林	-0.514132411	0.8279	-0.03728	0.172099	-0.432067	22
黑龙江	-0.8279	0.8279	0.718538	0.172099	0.1052	14

江	0.022190953		9		877	
上	1.098064		-		0.3734	
海	242	0.8279	3.112379	0.172099	523	9
江	1.944396		0.638363		1.7196	
苏	607	0.8279	3	0.172099	299	2
浙	1.072778		-		0.8527	
江	883	0.8279	0.205744	0.172099	467	5
安	0.049146		0.609536		0.1455	
徽	862	0.8279	3	0.172099	892	13
福	-		-		-	
建	0.458697466	0.8279	0.682284	0.172099	0.497176	24
江	-		-		-	
西	0.496846889	0.8279	0.083371	0.172099	0.425688	21
山	1.331479		-		1.0382	
东	735	0.8279	0.372198	0.172099	79	4
河	0.783962				0.8251	
南	16	0.8279	1.023351	0.172099	607	6
湖	-				-	
北	0.030563716	0.8279	0.121912	0.172099	0.004323	16
湖	0.173871		0.636989		0.2535	
南	028	0.8279	1	0.172099	731	11
广	2.678773		-		1.9912	
东	676	0.8279	1.316377	0.172099	132	1
广	-		0.023295		-	
西	0.289966414	0.8279	1	0.172099	0.236054	19
海	-		-		-	
南	1.522144163	0.8279	0.947765	0.172099	1.423294	31
重	-		0.303684		-	
庆	0.700344797	0.8279	1	0.172099	0.527553	25
四	0.926859		2.271373		1.1582	
川	071	0.8279	7	0.172099	483	3
贵	-		0.503654		-	
州	0.576712419	0.8279	9	0.172099	0.390783	20
云	-		1.035282		0.1749	
南	0.003899118	0.8279	4	0.172099	427	12
西	-		-		-	
藏	1.368533843	0.8279	0.400791	0.172099	1.201987	28
陕	-				-	
西	0.225861937	0.8279	0.962156	0.172099	0.021406	17

甘	-	0.8279	0.29317	0.172099	-	26
肃	0.809178058				0.619465	
青	-	0.8279	-	0.172099	-	29
海	1.430543861		0.445948		1.261096	
宁	-	0.8279	-	0.172099	-	30
夏	1.512337101		0.820826		1.393329	
新	-	0.8279	0.120065	0.172099	-	23
疆	0.562688437		7		0.445187	

利用因子分析法提取的 2 个因子后，各地区综合得分大小与正负、排名前后差异明显。财政支出指标值都表现在不同地区的，因此排名需要按地区讨论，这样更体现地区财政支出结构的合理性。第一因子排名靠前的 5 个地区分别是广东、上海、江苏、浙江、北京，说明这些地区的地区发展支出多，重视地区发展，是在公共服务基础上，进一步扩大服务支出范围与质量。说明该因子得分与省市地区的经济发展水平有关，经济整体越发展则得分越高；反之，经济越落后得分越低，有些地区甚至达到负值，说明在这项因子上支出匮乏，需要调整并改善。第二因子排名靠前的 5 个地区则是四川、云南、内蒙古、河南、陕西，总体上说，这些地区经济较落后，它们的财政支出基本用于满足公共服务。

综合排名中，位于前 3 的地区则分别为广东、江苏、四川，两个因子单项得分靠前的地区均有包括，说明各个因子在总得分中权重存在差异，表明我国财政支出存在地区不平衡，需要根据各地区财政基础做出调整。

总结

财政支出结构一方面反映了该地区经济发展水平，另一方面也是影响地区发展的一个重要因素，本文调取了我国重要发展时期即改革开放 30 周年的 2008 年分地区财政支出数据，并对其进行预处理，算得各地区标准化财政支出数据，利用 SPSS 软件对此进行基本的描述统计分析和因子分析。研究表明，总体上我国财政支出存在地区不均衡现象，而这种不均衡主要体现在于经济发展落后如宁夏省的财政支出与其它地区差异过大，通过第二因子得分，说明这些地区的公共财政支出主要用于满足地区公共服支出了。

广东、上海、江苏等沿海地区的第一因子得分高，在满足了公共服务支出下，提供了更多的为促进地区进一步发展的支出，如科技技术。教育等项目。而云南、内蒙古、河南等中西部省区的经济发展情况对其财政支出结构存在很大的影响，

在公共服务支出上往往会更多，目的是促进私人经济的发展，而经济发达地区的各项经济基础设施建设已基本完成，会加大人力资本的投入，使经济能健康、平稳发展。