

我国高等教育资源地域性差异研究

暨南大学 王瑞峰、陈昭豪、冯慧阳

摘要

高等教育资源地域性差异常常成为社会关注的焦点，本文利用教育部和统计局公布的 2009 年我国高等教育统计数据，利用主成分分析方法、熵值法和“教育基尼系数”分析法从定量的角度对各地区的高等教育资源差异进行综合评价分析，得出各地区差异的情况并分析其深层次的原因，最后结合经济社会的知识给出一些合理的建议。

关键字: 高等教育资源 地域性差异 主成分分析 熵值法 教育基尼系数

目录

1 前言	3
1.1 问题的提出.....	3
1.2 研究现状及存在的问题.....	4
2 指标的选取与数据说明.....	4
2.1 指标的选取与说明.....	4
2.2 数据来源说明.....	6
2.3 数据的预处理.....	7
3 综合评价分析方法.....	7
3.1 方法一：基于主成分分析的综合评价.....	7
3.1.1 方法介绍.....	8
3.1.2 求解过程及结果.....	8
3.1.3 结果分析.....	11
3.2 方法二：基于熵值法的综合评价.....	12
3.2.1 方法介绍.....	12
3.2.2 分析过程及结果.....	13
3.2.3 结果分析.....	14
3.3 方法三：“教育基尼系数”分析法.....	15
3.3.1 方法介绍.....	15
3.3.2 求解过程.....	16
4 结论	16
5 建议	17
6 分析方法的优缺点及改进方向	18
6.1 优点	18
6.2 缺点	18
6.3 改进方向	19
参考资料	20
附录	21
附录 1. 原始数据:	21
附录 2 转换后的数据.....	22
附录 3 正文中用到的 R 语言程序代码:	23

1 前言

1.1 问题的提出

高等教育的发展历史可以追溯到中世纪的大学,后历经发展,主要是英国、德国、美国的大学的不断转型,形成了高等教育的三项职能即:培养专门人才;科学研究;服务社会。改革开放以来,我国高等教育事业获得长足发展,改革取得令人瞩目的成绩,初步形成了适应国民经济建设和社会发展需要的多种层次、多种形式、学科门类基本齐全的社会主义高等教育体系,为社会主义现代化建设培养了大批高级专门人才,在国家经济建设、科技进步和社会发展中发挥了重要作用。

当前我国的高等教育在办学规模、办学质量、办学理念、办学方式,以及师资水平、学生人数、高教投资、管理监控上都得到较快的发展,但是仍低于发展中国家的平均水平,未能赶上社会发展和经济发展。此外,我国的高等教育体制还面临着诸多问题,如高等教育的生存与竞争问题,大众化和质量监控问题,自身道德和科学精神问题等。而高等教育的公平直接关系到社会的公平,影响社会主义和谐社会的建设。

高等教育的不公平表现主要表现在教育资源的分配。高等教育资源分布是直接影响高校发展的重要基础性资源。由于高等教育资源具有有限性,这就决定了高教资源分布是否合理对于高等教育事业发展甚至区域经济发展具有重要意义。此外还有教育机会选择、地区性教育歧视、结构性供需失衡等问题。就教育资源的分配问题而言,在东部沿海地区和西部内陆地区,中心城市和边远地区,对高等教育的师资投入相差甚远,例如北京和海南比较,北京拥有许多经费庞大的部委高校,即使有些直属北京市地方的高校经费和教师队伍也要优于海南最好的高校。造成这种地域差异的原因之一是各地区经济发展不平衡,一些较落后的地区没有很多的财力可以投入到高等教育;原因之二是国家财政投入不足,很大程度上要求地方上的教育经费投入;原因之三是目前资源配置机制不合理,导致优质资源更加集中,加大了地区间差距。此外,我国教育资源相对于需求的不足也是出现分配不公平的重要原因。

国家要解决经济发展的地域差异问题,很重要一点是要解决好高等教育资源

配置的地域差异。要解决好高等教育资源配置的地域差异，首先要对高等教育资源配置的地域差异进行深入的了解，找出造成差异的根本原因，这样才能制定合理和有效的政策，解决好高等教育资源配置地域差异的问题，推动我国经济社会的可持续发展。

1.2 研究现状及存在的问题

康宁(2004)对改革开放以来我国高等教育资源配置制度转换过程进行考察，认为高等教育资源配置转换过程是学术、市场、政府三种配置力量不断调整与选择的过程。常晓宁(2004)分析了目前我国高等教育资源配置存在的弊端，提出了实现高等教育资源的优化配置的有效途径。岳武和李彬彬(2011)从高等教育系统内部各要素之间相对稳定的联系方式和比例关系入手，对我国转型期的高等教育资源配置中的结构布局特点进行了分析，并对当前高等教育因规模扩张而带来的问题进行了有效评论。Liefner I(2003)通过分析高等教育资金和资源的分配的各种形式对微观层面宏观层面的影响，认为基于绩效的资源配置往往会带来积极的效果，但也会产生一定副作用，同时资源配置的形式影响高校的学者和管理者。Ma Dan(2007)运用多元回归分析的方法对2002年中国大陆1085所高等教育机构经费开支及其他资源的分布数据进行分析，表明高等教育资源地域分布与地域的经济发展水平呈正相关。

从国内外的研究成果来看，对高等教育资源配置的研究主要集中在宏观层面上，并且以定性研究为主，强调市场的力量对高等教育资源配置的作用。在我国的高等教育资源配置的研究中，主要集中讨论政府在高等教育资源配置中的作用和我国高等教育资源配置存在的问题和改革方向等，对高等教育资源配置的计量分析不足，提出的建议很多未关注地域的差异。

本文利用教育部公布的2009年我国高等教育统计数据，利用主成分分析方法、熵值法和“教育基尼系数”分析法从定量的角度对各地区的高等教育资源差异进行综合评价分析，得出各地区差异的情况并分析其深层次的原因，最后结合经济社会的知识给出一些合理的建议。

2 指标的选取与数据说明

2.1 指标的选取与说明

为了能够反映各地区的高等教育资源地域性差异,本文引入了 8 个比较容易得到的指标构建指标体系,可能存在更好反映高等教育资源地域性差异的指标,但是这些指标的获得往往需要很高的成本和很长的时间,有时甚至是无法得到的,所以本指标体系放弃了这些指标。指标体系中的 8 个指标分别为:

- A1: 高等院校数
- A2: 211 高等院校数
- A3: 高级职称教师人数
- A4: 中级及以下职称教师人数
- A5: 学校占地面积
- A6: 办学经费
- A7: 计算机台数
- A8: 教学、科研仪器类固定资产值
- A9: 高等院校在校生人数

高等院校数:高等院校指按照国家规定的设置标准和审批程序批准举办,通过全国普通高等学校统一招生考试,招收普通高中毕业生及职业高中毕业生为主要培养对象,实施高等教育的全日制大学、独立学院和高等专科学校、高等职业学校和其他教育机构。高等院校数包括中央部委、本科院校和高职(专科)院校等普通高校数,同时还包括成人高等学校数和民办的其他高等教育机构。作为基础性的指标,高等院校数可以说是反映高等教育资源最基本的因素。

211 高等院校数:211 高等院校的来源,1993 年 2 月 13 日中共中央、国务院印发的《中国教育改革和发展纲要》及国务院《关于〈中国教育改革和发展纲要〉的实施意见》中,关于“211 工程”的主要精神是:为了迎接世界新技术革命的挑战,面向 21 世纪,要集中中央和地方各方面的力量,分期分批地重点建设 100 所左右的高等学校和一批重点学科、专业,使其到 2000 年左右在教育质量、科学研究、管理水平及办学效益等方面有较大提高,在教育改革方面有明显进展,力争在 21 世纪初有一批高等学校和学科、专业接近或达到国际一流大学的水平。并可概括表述为:“211 工程”就是面向 21 世纪,重点建设 100 所左右的高等学校和一批重点学科点。为此,211 高等院校数反映的是优质的高等教育资源,同时优质的高等教育资源能反映一个地区的文化竞争力。

高级职称教师人数：职称是专业技术资格，高级职称是职称中最高级别，分正高级和副高级，高校中高级职称教师是指正教授和副教授。教授的工作是在大学里针对他们所擅长的领域开课、或授与学生专业训练，如科学和文学等领域。另外教授也必须深耕自己专精的学科，以发表论文的方式来获得商业上的合作机会（包括了政府在科学上的顾问、或是商业发明等），同时训练自己的学生将来有足够的能力与他们交棒。为此，高级职称教师人数反映高校的优质师资力量，而师资力量可以说是各高校竞争的核心，不单代表高校的教学水平，还代表了高校的科研水平，为此该指标具有重要意义。

中级及以下职称教师人数：作为高校师资队伍的重要部分，同高级职称教师人数共同构成了高校的师资力量。

学校占地面积：进行高等教育必须要有一定的软硬设施，而教学楼、实验室和图书馆等硬件设施都是要占用一定土地，同时还需要向学生提供各种活动的场所，这些都需要占用一定的土地。而高等院校的学校占地面积作为一个空间限制指标很大程度上也决定了高校的办学规模。

办学经费：高等教育作为一项事业，无论是支付教师和行政人员的薪金，还是购买各种科研仪器，都需要经费的支持，高等院校办学经费包括中央下拨的经费和地方的各种经费。拥有充足的经费，高校才能运作起来。

计算机台数：现代的社会是信息化的社会，而计算机可以说是信息化的核心。为此高校拥有计算机台数，一定程度上反映了高校的信息化水平。

教学、科研仪器类固定资产：现代化教学需要使用各种教学和科研仪器，教学、科研仪器类固定资产的多少可以说一定程度上代表了高校的教学方式的水平。

高等院校在校生人数：是指各类高等院校在生人数，包括研究生、本科生还有大专生等。

2.2 数据来源说明

以上指标数据除开办学经费都来自《中国教育统计年鉴（2009）》，《中国教育统计年鉴(2009)》是一本全面反映中华人民共和国教育事业发展情况的资料性年鉴，是由教育部发展规划司根据全国各省、自治区、直辖市教育委员会、教育厅填报的学校基层报表数字整理汇编而成的。教育部教育管理信息中心承担了数

据的计算机处理汇总工作。年鉴包括以下部分：综合部分、高等教育、中等教育、初等教育、幼儿教育、特殊教育、全国各级各类学校的分布情况、办学条件、科学研究等。年鉴所列的资料，暂缺台湾省、香港特别行政区和澳门特别行政区的数字；凡未注明年份的均为 2009 年的数字。

办学经费指标的数据来自《中国教育经费统计年鉴（2010）》。

原始数据见附录 1。

2.3 数据的预处理

我们对高等院校数，211 高等院校数，学校占地面积，计算机台数和教学、科研仪器类固定资产值取其绝对值，因为这些都是硬指标，带有公共品的性质：增加一个学生，不会造成这类资源的效用的明显下降。这些指标可以说是代表高等教育办学规模；而高级职称教师人数，中级及以下职称教师人数和办学经费做了人均化处理，也就是用相关指标除以高等院校在校生人数，因为三个指标是软指标，流动和变化都比较容易，同时也比其他五个指标具有更多的排他性，故对其求人均比较合适，这些指标可以说是代表高等教育办学水平。这样得出以下转换的指标：

B1：高等院校数；

B2：211 高等院校数；

B3：人均高级职称教师数拥有数

B4：人均中级及以下职称教师拥有数

B5：学校占地面积

B6：人均办学经费

B7：计算机台数

B8：教学、科研仪器类固定资产值

转换后的数据见附录 2。

3 综合评价分析方法

3.1 方法一：基于主成分分析的综合评价

3.1.1 方法介绍

主成分分析 (principal component analysis, 简称 PCA) 是将多指标化为少数几个综合指标的一种统计分析方法, 是由 Pearson 提出, 后来被 Hotelling 发展起来的。

这种分析方法地基本思想是通过降维技术, 设法将原来众多具有一定相关性 (比如 P 个指标), 重新组合成一组新的互相无关的综合指标来代替原来的指标。通常数学上的处理就是将原来 P 个指标作线性组合, 作为新的综合指标。最经典的做法就是用 F1 (选取的第一个线性组合, 即第一个综合指标) 的方差来表达, 即 $\text{Var}(F1)$ 越大, 表示 F1 包含的信息越多。因此在所有的线性组合中选取的 F1 应该是方差最大的, 故称 F1 为第一主成分。如果第一主成分不足以代表原来 P 个指标的信息, 再考虑选取 F2 即选第二个线性组合, 为了有效地反映原来信息, F1 已有的信息就不需要再出现在 F2 中, 用数学语言表达就是要求 $\text{Cov}(F1, F2)=0$, 则称 F2 为第二主成分, 依此类推可以构造出第三、第四, …… , 第 P 个主成分。

主成分分析的步骤:

- (1) 将原始数据标准化, 消除变量间不同量纲的影响;
- (2) 计算相关系数矩阵;
- (3) 根据相关系数矩阵的特征值和特征向量确定主成分个数;
- (4) 根据原有线性组合计算 F_i 表达式;
- (5) 计算综合得分并排序。

3.1.2 求解过程及结果

利用 R 语言进行主成分分析, 得到标准化后的数据的主成分和主成分载荷如下:

表 1 主成分分析表

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
标准差	2.0493	1.6341	0.7937	0.6081	2.0493	1.6341	0.7937	0.6081
方差贡献率	0.5250	0.3338	0.0787	0.0462	0.5250	0.3338	0.0787	0.0462
累计方差贡献率	0.5250	0.8588	0.9375	0.9837	0.5250	0.8588	0.9375	1.0000

表2 主成分载荷表

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7	Comp.8
B1	-0.363	0.245		0.876	0.156	0.117		
B2	-0.433	-0.179	0.308	-0.286	0.674	0.310	-0.230	
B3	-0.260	-0.468	-0.455					0.707
B4	-0.258	-0.470	-0.454					-0.707
B5	-0.264	0.473	-0.287	-0.285	-0.316	0.663		
B6	-0.327	-0.321	0.623		-0.617	0.112		
B7	-0.396	0.334	-0.115	-0.192	-0.151	-0.541	-0.604	
B8	-0.461	0.172		-0.163		-0.381	0.757	

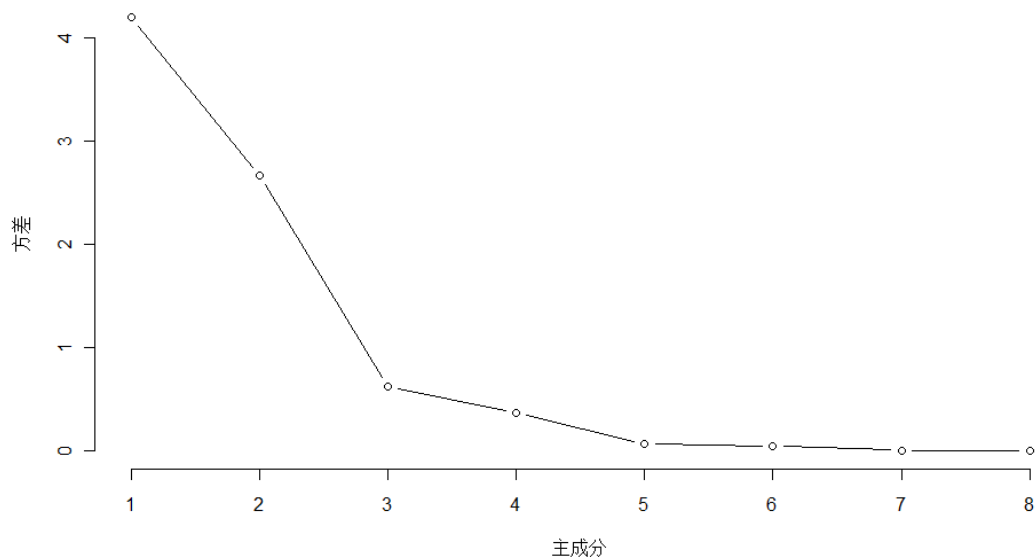


图1 碎石图

由于前两个主成分的累计方差贡献率，即解释能力达到了85%以上，所以可以选取这两个主成分来对数据进行评价。由图1的碎石图也可以看出取两个主成分是合适的。

由主成分载荷矩阵可以看出，主成分 C1 在 B2（211 高等院校数）、B7（计算机台数）、B8（教学、科研仪器类固定资产值）上的载荷都很大，可以视为反映高等教育资源规模的主成分；而主成分 C2 在 B3（人均高级职称教师数拥有数）、B4（人均中级及以下职称教师拥有数）和 B5（学校占地面积）上的载荷很

大，其中，表示人均教师拥有数的两个指标和表示占地面积的指标符号相反，存在一定的抵消作用，在结果中占主导地位的还是人均教师拥有数，这个主成分可以被视为反映高等教育资源水平的主成分。结合各个地区在这两个主成分上的得分和综合得分，就可以对各个地区的高等教育资源水平进行评价了。

表3 主成分综合评分表

	Comp.1	Comp.2	综合评分	排名
北京	-7.3133	-4.4326	-6.1936	1
天津	-0.2374	-2.1656	-0.9869	4
河北	0.1331	1.0277	0.4808	17
山西	1.2040	0.0523	0.7564	25
内蒙古	1.4975	-0.8851	0.5714	21
辽宁	-1.1864	0.1121	-0.6817	6
吉林	0.4681	-0.6913	0.0174	12
黑龙江	-0.2195	-0.1173	-0.1798	9
上海	-3.2788	-0.7862	-2.3100	2
江苏	-3.7582	2.4595	-1.3415	3
浙江	-0.3000	0.8588	0.1504	13
安徽	0.7244	1.1991	0.9089	27
福建	0.6094	0.1059	0.4137	16
江西	0.3466	0.6961	0.4825	19
山东	-2.1081	3.1643	-0.0588	11
河南	0.0424	2.2605	0.9045	26
湖北	-1.4837	1.1002	-0.4794	7
湖南	-0.2519	1.0356	0.2486	15
广东	-1.5257	2.1301	-0.1048	10
广西	1.6852	0.4999	1.2245	29
海南	2.8900	-0.0657	1.7412	31
重庆	1.0513	-0.2277	0.5542	20
四川	-0.5700	1.4325	0.2084	14
贵州	1.7162	-0.8269	0.7278	23
云南	1.4187	-0.2937	0.7531	24
西藏	2.7251	-0.9510	1.2963	30
陕西	-0.6792	0.0774	-0.3851	8
甘肃	1.8768	-0.0732	1.1189	28
青海	1.0107	-4.1092	-0.9793	5
宁夏	1.9404	-1.8129	0.4816	18
新疆	1.5721	-0.7736	0.6604	22

3.1.3 结果分析

在教育规模主成分 C1 上排名最高的地区是北京、上海和江苏，这些地区的经济水平都很高，而且拥有很多名牌大学，这些学校由于有中央政府和地方在多方面的支持，可以拥有一个比较大的高等教育规模。而排名最低的地区为海南、西藏和宁夏，这主要是因为这三个地区的学校数较少，而且缺乏名牌大学，只有一些规模较小的普通学校，这样的话很难形成很大的教育规模。在教育水平主成分 C2 上排名最高的是北京、青海和天津，青海在 C2 上排名高比较出乎意料，分析其原因可以看出：青海的高校在校学生人数较少，相对其教师资源也比较丰富，其平均每百人拥有 3.2 个高级职称教师，6.5 个专职教师，全国仅有北京地区拥有如此高的师生比；再者就是青海的人均办学经费也较高，所以其排名可以仅次于北京。而北京和天津排名这么高的原因则是显而易见的，尤其是北京，作为全国的政治与文化中心，其吸引了许多师资，再加上两院的院士主要集中在北京，所以其高等教育水平排名第一；而排名最低的则为山东、江苏和河南，这主要是由于反映教育水平的一些资源都具有一定程度的排他性，即多一人使用这项资源每个人得到的效用会明显变低，而这三个地区的在校学生数都非常多，虽然其教师数量和经费数额也很大，但是其相对于在校学生人数而言就要小于其他地区了。就总得分而言，北京、上海、江苏、天津和青海的排名最高，海南、西藏、广西和甘肃的排名最低，这说明综合得分是综合反映高等教育资源规模和水平的一个指标，其中教育规模的比重要更大一些。为了更加直观的比较各地区的差异，在此可以绘制中国地图，并根据各地区的综合得分的不同对其进行不同的着色，颜色越深的表示占有的高等教育资源越多。由图 2 可以看出，北京的色彩非常浓，与其他地区的区别非常明显，这说明在我国高等教育资源最明显的特点就是北京的资源占有量非常大，这是因为北京是我国的政治和文化中心，这就决定了政府对北京有很大的政策偏斜，如今北京很多的知名高校都是在建国后的几次大的高校调整中将非常多的资源调入北京后形成的，政府对其的大力支持更是其他地区没法可比的。



图2 各地区教育资源分布直观图(1)

3.2 方法二：基于熵值法的综合评价

3.2.1 方法介绍

在信息论中，熵是对不确定性的一种度量。信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也越大。根据熵的特性，我们可以通过计算熵值来判断一个事件的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，指标的离散程度越大，该指标对综合评价的影响越大。

计算熵值的步骤：

(1) 原始矩阵归一化。在评价体系内有 m 个评价指标， n 个评价对象，设原始矩阵 $A = (a_{ij})_{m \times n}$ ，归一化得到 $R = (r_{ij})_{m \times n}$ ，归一化公式为：

$$r_{ij} = \frac{a_{ij} - \min_j \{a_{ij}\}}{\max_j \{a_{ij}\} - \min_j \{a_{ij}\}}$$

(2) 定义熵。在评价体系中，令第 i 个指标的熵为 $h_i = -k \sum_{j=1}^n f_{ij} \ln f_{ij}$ ，其中，

$f_{ij} = r_{ij} / \sum_{j=1}^n r_{ij}$ ， $k = 1 / \ln n$ ，当 $f_{ij} = 0$ 时， $\ln f_{ij}$ 没有意义，故给整个矩阵加上非常

小的一个数后进行计算熵。

(3) 定义熵权。根据第 i 个指标的熵可以得到这个指标的熵权：

$$w_i = \frac{1 - h_i}{m - \sum_{i=1}^m h_i} \quad (0 \leq w_i \leq 1, \sum_{i=1}^m w_i = 1)。$$

3.2.2 分析过程及结果

利用 R 语言根据以上计算步骤计算得到四个因素的熵权（表 4），利用熵权计算得到综合评分并排序，得到表 5。

表 4 各指标熵权

指标	B1	B2	B3	B4	B5	B6	B7	B8
权重	0.0740	0.3249	0.0740	0.0743	0.0743	0.2074	0.0805	0.0958

表 5 基于熵权法的综合评分表

地区	综合评分	排名	地区	综合评分	排名
北京	2.7580	1	湖北	0.5496	4
天津	-0.0678	15	湖南	0.2360	10
河北	-0.2558	19	广东	0.3317	7
山西	-0.4351	24	广西	-0.6857	27
内蒙古	-0.3693	21	海南	-1.1192	30
辽宁	0.5415	5	重庆	-0.3599	20
吉林	-0.1785	17	四川	0.0712	13
黑龙江	0.1186	11	贵州	-0.6102	26
上海	1.0497	3	云南	-0.5713	25
江苏	1.1948	2	西藏	-1.1637	31
浙江	-0.1230	16	陕西	0.2541	9
安徽	-0.4178	23	甘肃	-0.7089	28
福建	-0.2199	18	青海	0.4317	6
江西	0.0851	12	宁夏	-0.9026	29
山东	0.2776	8	新疆	-0.4032	22
河南	0.0079	14			



图3 各地区教育资源分布直观图(2)

3.2.3 结果分析

在这八个权重中，B2（211 高等院校数）和 B6（人均办学经费）的权重最大，这是由于这两个指标的离散程度最大，例如北京拥有 26 所 211 工程大学，而且教学经费非常多，而其它地区，如河南，河北，不仅 211 工程大学数目和经费很少，而且在校学生人数也非常多，这些指标的人均值必然同北京这类地区有很大的差异，比如 2009 年北京的人均办学经费接近 20000 元，而河北只有 1000 多元，二者相差了将近 20 倍。同时 211 工程大学一般拥有更优秀的老师，这些老师通常可以更容易地申请到更多的经费。而其他指标的权重则基本相同，这说明其他指标所表示的资源对各地区教育资源的影响是没有特别大的区别的。故通过以上熵权法确定的权重计算得到的综合评分在一定程度上反映出高等教育的水平。熵权法得到的评分结果与主成分分析得到的排名相差不大，排名最高的 5 个地区仍然是北京、江苏、上海、湖北和辽宁，这与主成分分析的结果相差很小

而湖北能够排名靠前主要是因为其拥有一些非常出色的大学，如武汉大学、华中科技大学等。排名最低的地区分别是西藏、海南、宁夏和甘肃。同样可以利用地图来直观地表示地区间的差异^①。通过图 3 可以看出，北京依旧占有的高等教育资源非常多，与其它地区的差异非常明显；同时沿海的一些地区的得分相对于内陆都比较高，这说明经济水平对教育资源的影响还是很大的。

3.3 方法三：“教育基尼系数”分析法

3.3.1 方法介绍

基尼系数是国际上用来综合考察居民内部收入分配差异状况的一个重要分析指标，是意大利经济学家基尼于 1922 年提出的。其经济含义是：在全部居民收入中，用于进行不平均分配的那部分收入占总收入的百分比。如图 4 所示，基尼系数即阴影面积与直线 $y=x$ 以下面积的比值。基尼系数最大为“1”，即罗伦兹曲线与直线 $y=x$ 重合，最小等于“0”，即罗伦兹曲线自原点出发，沿 x 轴到 $x=100$ 后再沿垂直线到达终点 1 表示居民之间的收入分配绝对不平均，即全部的收入被一个单位的人全部占有了；而 0 则表示居民之间的收入分配绝对平均，即人与人之间收入完全平等，没有任何差异。但这两种情况只是在理论上的绝对化形式，在实际生活中一般不会出现。因此，基尼系数的实际数值只能介于 0 ~ 1 之间。

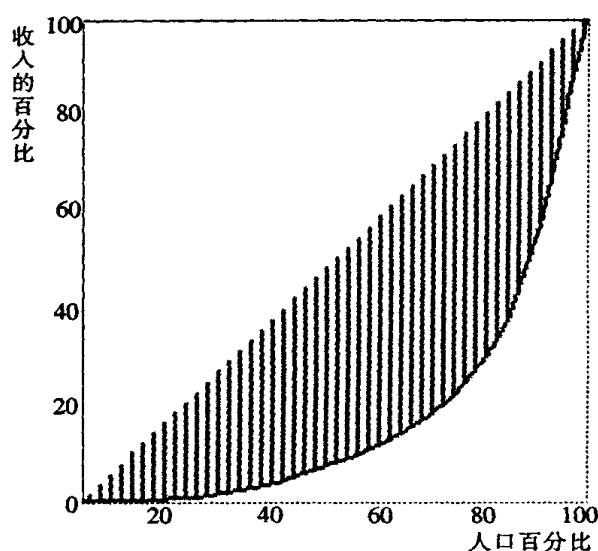


图 4 基尼系数图示

^① 利用 R 语言绘制地图有很多方法，本文中是的方法需要先下载中国地图的 GIS 数据，并将代码附在附录中。具体方法可以参考《用 R 软件绘制中国分省市地图》（邱怡轩）。

基尼系数是一种广义的分析工具，虽然目前主要用来解释收入分配，即贫富差距的问题，但是还可以用于其他关于分配均衡程度的问题分析。世界银行的学者早在上个世纪就提出了教育基尼系数的概念，但是这个系数是测量人口中教育成就和分布的指标，主要用来反映各国国家或地区的教育公平性。在这里，我们引入的基尼系数虽然与之类似，但是反映的却是高等教育资源地域性差异问题。

3.3.2 求解过程

国内外有许多读者进行关于基尼系数的研究，所以存在许多计算基尼系数的方法，本文在这里采用 R 的 ineq 程序包中的算法，该算法的计算公式如下：

$$G = \frac{2 \sum_{i=1}^n ix_i}{n \sum_{i=1}^n x_i} - 1 - \frac{1}{n} \quad (x \text{ 为原始按递增顺序排序后获得的新数据})$$

3.3.3 求解结果与分析

分别利用 R 语言对以上两种方法的结果来计算基尼系数。首先需要对这些数据进行非负化处理，得到一组新的非负数据，同时将每个数据重复其所在地区的在学生数次数，表示每个学生占有其所在地区的平均水平的资源。考虑个地区的人数较多，我们利用 10000 除人数得到新的数据，根据这两组新的数据得到所需的数据，然后计算得到粗略的基尼系数。

通过主成分的得分计算得到的基尼系数为 0.3107。

通过熵值法的得分计算得到的基尼系数为 0.2517。

二者都大于 0.25。参照联合国对基尼系数规定，这两个系数说明高等教育资源的分配是相对平均的，从总体上可以认为这种分配是相对合理的。

4 结论

根据以上综合分析结果，可以得到以下结论：

就总体情况而言，各个地区占有的高等教育资源是有差异的。其中东部沿海地区的高等教育资源较为丰富，而一些经济落后的内地地区的高等教育资源则比较匮乏，例外则是青海和陕西，其中青海主要是师资力量相对于学生数比较丰富，

而陕西则是拥有一些很出色的重点大学，依托这些学校可以申请到更多的科研经费，整个地区的高等教育水平也自然会随之提升。但是这种差距并不是很大，在可以达到高效率的前提下，这种差距是可以接受的。

本文中讨论的高等教育资源主要包括教育规模和教育水平，有些地区如北京、上海和江苏在两个方面都是佼佼者，有些地区则是在两个方面都很落后，如西藏和海南。同时有一些发展不平衡的地区，比如山东等地区的高等教育规模很大，但是由于在校学生基数过大，一些资源被稀释，造成某些资源水平上的相对落后；而青海则相反，由于学生数很少，其一些资源的相对量是很大的，但是整个教育规模却很小。

虽然总体上高等教育资源分配相对平均，但是北京占有的资源却是非常多的，其无论教育规模还是教育水平都远远超出其他地区，这对于其他地区存在不公平，同时这也是人们认为目前中国高等资源地域性分配不公平的主要原因。

5 建议

针对以上结论，可以对我国高等教育资源的配置提出一些建议：

(1) 继续加大对西部地区的高等教育经费的投入，同时利用国家政策号召吸引人才到西部就业，以提高西部地区高等教育的师资水平。

(2) 促进地区高等教育资源的平衡发展，比如青海省，拥有的专职教师数相对较多，但是由于学生数少，高等教育规模并没有实现很大的改进，在这种情况下如不能扩大其他类型资源的投入，会造成教师资源的浪费。

(3) 国家应对一些人口大省进行政策性的照顾，由于这些地区往往拥有数目庞大的在校学生人数，一些具有竞争性和排他性的资源对于每个学生的效用就非常小；同时地方政府也应该清楚地认识到这一问题，并针对这一问题投入更多相关的资源，以免产生方向上的错误。

(4) 应加强落后地区与发达地区的学术交流活动，这种交流通常可以在一定程度上对落后地区的高等教育水平提到正面作用。

(5) 中央政府应该注意到北京的教育资源占有量非常大的问题，并将其与其他地区的差异控制在警戒线以内，以免这一问题成为我国目前高等教育资源地域性不公平问题中的焦点，影响社会的公平。

6 分析方法的优缺点及改进方向

6.1 优点

在本文中使用的两种分析方法中，主成分分析方法是多元统计分析中一种非常经典的分析方法，这种方法虽然在一定条件下减少了变量个数，但是它利用指标间的相关关系通过变量变换得到综合的新指标，可以说基本没有损失原值标的信息。同时主成分分析还可以通过主成分来揭示变量之间的内在关系，有助于深化我们的研究。但是主成分分析利用方差贡献率作为权重，这一点包含了主观的成分。而另一种方法熵权法则是近年来比较流行的一种综合评价方法，其思想来自信息论中的熵。这种方法利用信息效用值来确定指标的权重，相对于主成分分析来说是一个更加可观的赋权方法，但是这种方法反映的问题比较少。可以说两种方法各有优缺点，而本文则将两种方法结合起来进行分析研究，使其起到相互补充的效果，可以得到更加客观、科学的评价结果，同时较为全面地反映了客观数据所提供的信息。

同时，本文利用了一些直观分析方法，引入经济中的基尼系数来比较各地区间高等教育资源占有量的差异程度，可以使读者对本文所论述的内容有更形象的认识。

6.2 缺点

本文基本在利用定量分析对高等教育资源的分配问题进行研究，并没有进行很有效的定性分析，这样并不能很好地反映社会大众和专家学者对高等教育资源分配问题的态度。

我们的研究都是基于截面数据进行分析的，未能从时间上反映高等教育资源分配地域性差别的变化状况。同时主成分分析和熵值法得到的权重往往是根据数据特点所得到的，有时并不能够很好地反映各种资源在高等教育资源中的地位。

另外，由于本文建立的指标体系中的指标只能反映一些可以量化的资源，而许多无形的资源，如名牌大学形成的品牌效应就无法很好地进行量化处理。同时有许多指标也由于政策性的影响可能存在偏差，比如 211 工程学校的数目，很多学校是因为政策性原因才进入 211 工程大学的，这些学校的资源很多时候还不如一些发达地区的非 211 工程的大学，所以在这一指标体系下得到的结果与实际可

能有些出入。

6.3 改进方向

在分析横截面数据的同时，还可以分析历史数据在时间上的变化，可以更好地反映高等教育资源地域性差别的动态变化，这样才能提出更加合理的建议。

由于指标体系的建立是在目前可得数据的基础上建立的，无法反映一些难以量化的资源，经过一些更深入的研究，对指标体系进行改进可以得到更合理的评价效果。

本文将综合评价的得分简单处理后综合各省在校学生数计算得到基尼系数，如果对评分进行更合理的预处理，或是找到可以更好地反映各地区高等教育资源的综合指标，应该可以获得更科学的结果。

参考资料

- [1] 王斌会.《多元统计分析及 R 语言建模》[M]. 广州:暨南大学出版社,2010. 01
- [2] 陆添超,康凯. 熵值法和层次分析法在权重确定中的应用[J]. 软件开发与设计, 2009. 09
- [3] 陶小龙. 教育基尼系数的计算及其分解[J]. 科技与社会, 2009. 02
- [4] 姚海娟. 论高等教育资源分布与区域经济发展的关系——以湖南省为例. 《黑龙江高教研究》, 2011 年第 4 期。
- [5] 康宁. 高等教育资源配置: 规律与变迁趋势——学术、市场、政府在优化高等教育资源配置中制衡的约束条件. 《教育研究》, 2004 年第 2 期。
- [6] 常晓宁. 高等教育资源配置的现状 & 实现优化配置的途径. 《中山大学学报》, 2004 年第 43 卷。
- [7] 岳武, 李彬彬. 转型期我国高等教育资源配置中的结构布局分析. 《东北师大学报》, 2011 年第 3 期。
- [8] 邱怡轩. 2009: 用 R 软件绘制中国分省市地图. 统计之都
<http://cos.name/2009/07/drawing-china-map-using-r/>
- [9] Liefner I, “Funding , resource allocation , and performance in higher education systems”, Higher Education 46: 469–489, 2003.
- [10] Ma Dan, “Analytical study of the effectiveness of resource distribution in higher education for each province in China”, PROCEEDINGS OF THE 4TH INTERNATIONAL CONFERENCE ON INNOVATION & MANAGEMENT 444–449, 2007.

附录

附录 1. 原始数据:

地区	高等院 校数 (所)	211 高 等院校 数(所)	高级职称 教师人数 (人)	中级及 以下职 称教师 人数 (人)	学校占地 面积(平方 米)	办学经费 (元)	计算机 台数 (台)	教学、科研 仪器类固定 资产值(万 元)	高等院 校在校 生人数 (人)
北 京	178	26	30537	30741	40631100	17353369058	327596	2077303.11	875166
天 津	70	4	12697	12771	32898495	2330866456	115786	524318.87	483501
河 北	149	1	22888	23038	62766805	1389274294	219390	687064.15	1278999
山 西	125	1	11733	11859	27905473	1132039760	112651	387862.01	677649
内 蒙 古	44	1	8284	8329	31375554	1102447083	73088	247145.29	420555
辽 宁	212	4	24461	24677	59070099	2048242301	242852	762234.78	1071178
吉 林	88	3	14567	14658	39034165	1610262450	128157	516241.74	690146
黑 龙 江	122	4	18549	18675	54304543	1340927842	177192	718201.27	873654
上 海	340	9	16893	17242	34517771	7754220091	236774	1164983.09	726107
江 苏	162	11	37568	37741	121941310	8451398679	521785	1802276.69	2025102
浙 江	132	1	20080	20213	48431379	3531962651	268332	1028542.83	1135756
安 徽	119	3	15218	15340	57990280	1371629057	165349	607126.71	1033951
福 建	93	2	12659	12754	43208861	2127597253	163190	569779.09	705755
江 西	117	1	16918	17036	64763904	1411445959	204953	545251.09	913853
山 东	251	3	33054	33308	125361302	2223128374	374982	1327254.68	1970317
河 南	137	1	23573	23711	92982732	2677283772	253687	807641.43	1646627
湖 北	139	7	29905	30051	80894540	2247523004	308396	1168853.63	1525161
湖 南	146	3	22509	22658	68093301	1829360485	231732	777748.83	1249973
广 东	180	4	26956	27140	86241299	9410709507	338108	1290810.27	1797484
广 西	75	1	9807	9883	35124877	1331349695	112623	363559.78	672759
海 南	18	1	2149	2168	12091371	386301271.5	27127	83121.43	168345
重 庆	64	2	11312	11378	42591884	1724313193	119535	399342.55	618843
四 川	147	5	21721	21873	77893544	3194265244	223618	974965.68	1338198
贵 州	54	1	7194	7249	23502676	1005467826	60371	197859.84	385281
云 南	63	1	9271	9335	33433037	2922135177	91103	284837.71	567265
西 藏	6	1	557	564	2846914	159281871.3	5842	30430.21	37222
陕 西	119	7	21240	21366	52201791	2422966131	208686	924230.55	1065649
甘 肃	79	1	7138	7218	22942627	856288001.7	67363	276611.01	451091
青 海	11	1	1880	1892	4012325	98655127.56	10180	33763.57	58014
宁 夏	16	1	2078	2095	9047785	486024158.4	21262	64138.32	103427
新 疆	45	2	5440	5487	32288432	871355084.3	57198	209311.05	293055

数据来源:《中国教育统计年鉴(2009)》.人民教育出版社出版,2010.10

《中国教育经费统计年鉴(2010)》.中国统计出版社,2011.4

附录 2 转换后的数据

地区	高等院 校数 (所)	211 高 等院校 数(所)	人均高级 职称教师 数拥有数	人均中级 及以下职 称教师拥 有数	学校占地 面积(平方 米)	人均办学 经费(元/ 人)	计算机台 数(台)	教学、科研仪 器类固定资 产值(万元)
北京	178	26	0.0349	0.0350	40631100	19828.66	327596	2077303.11
天津	70	4	0.0263	0.0264	32898495	4820.81	115786	524318.87
河北	149	1	0.0179	0.0180	62766805	1086.22	219390	687064.15
山西	125	1	0.0173	0.0174	27905473	1670.54	112651	387862.01
内蒙古	44	1	0.0197	0.0198	31375554	2621.41	73088	247145.29
辽宁	212	4	0.0228	0.0229	59070099	1912.14	242852	762234.78
吉林	88	3	0.0211	0.0212	39034165	2333.22	128157	516241.74
黑龙江	122	4	0.0212	0.0213	54304543	1534.85	177192	718201.27
上海	340	9	0.0233	0.0234	34517771	10679.17	236774	1164983.09
江苏	162	11	0.0186	0.0186	1.22E+08	4173.32	521785	1802276.69
浙江	132	1	0.0177	0.0178	48431379	3109.79	268332	1028542.83
安徽	119	3	0.0147	0.0148	57990280	1326.59	165349	607126.71
福建	93	2	0.0179	0.0181	43208861	3014.64	163190	569779.09
江西	117	1	0.0185	0.0186	64763904	1544.5	204953	545251.09
山东	251	3	0.0168	0.0168	1.25E+08	1128.31	374982	1327254.68
河南	137	1	0.0143	0.0144	92982732	1625.92	253687	807641.43
湖北	139	7	0.0196	0.0197	80894540	1473.63	308396	1168853.63
湖南	146	3	0.0180	0.0181	68093301	1463.52	231732	777748.83
广东	180	4	0.0150	0.0151	86241299	5235.49	338108	1290810.27
广西	75	1	0.0146	0.0147	35124877	1978.94	112623	363559.78
海南	18	1	0.0128	0.0129	12091371	2294.7	27127	83121.43
重庆	64	2	0.0183	0.0184	42591884	2786.35	119535	399342.55
四川	147	5	0.0162	0.0163	77893544	2386.99	223618	974965.68
贵州	54	1	0.0187	0.0188	23502676	2609.7	60371	197859.84
云南	63	1	0.0163	0.0165	33433037	5151.27	91103	284837.71
西藏	6	1	0.0150	0.0151	2846914	4279.24	5842	30430.21
陕西	119	7	0.0199	0.0200	52201791	2273.7	208686	924230.55
甘肃	79	1	0.0158	0.0159	22942627	1898.26	67363	276611.01
青海	11	1	0.0324	0.0326	4012325	1700.54	10180	33763.57
宁夏	16	1	0.0201	0.0202	9047785	4699.2	21262	64138.32
新疆	45	2	0.0186	0.0187	32288432	2973.35	57198	209311.05

附录 3 正文中用到的 R 语言程序代码:

```
#编制绘制地图并着色的函数
mapcol <- function(x,y){#x 为图像参数, y 为长度为 31 的向量
  getColor=function(mapdata,provname,provcol,othercol)
  {
    f=function(x,y) ifelse(x %in% y,which(y==x),0)
    colIndex=sapply(mapdata$NAME,f,provname)
    fg=c(othercol,provcol)[colIndex+1]
    return(fg)
  }
  provname=c("北京市","天津市","河北省","山西省","内蒙古自治区","辽宁省","吉林省",
"黑龙江省","上海市","江苏省","浙江省","安徽省","福建省","江西省","山东省","河南省",
"湖北省","湖南省","广东省","广西壮族自治区","海南省","重庆市","四川省","贵州省",
"云南省","西藏自治区","陕西省","甘肃省","青海省","宁夏回族自治区","新疆维吾尔自治区")
  provcol=rgb(green=1-y/max(y)/2,red=1-y/max(y)/2,blue=0)
  plot(x,col=getColor(x,provname,provcol,"white"),xlab="",ylab="")
  }
#读入数据
dat=read.table("clipboard",header=T)#读取主体分析数据
dat=scale(dat) #对数据进行标准化
library(maptools)
chinamap=readShapePoly('bou2_4p.shp') #读入图像文件
n=read.table('clipboard')$V1 #读入在校学生人数数据
#主成分分析
PCA = princomp(dat,cor=T)
summary(PCA)
PCA$loadings
screplot(PCA,type="l")
library(mvstats)
princomp.rank(PCA,2)
W0 = as.matrix(PCA[[1]]^2/sum(PCA[[1]]^2))
PCs = as.matrix(PCA$scores[, 1:2])
PC = PCs %*% W0[1:2]/sum(W0[1:2]) #得到综合得分
mapcol(chinamap,(max(PC)-PC)) #绘制地图并着色
#熵值法确定权重后进行综合评分
dat1=sweep(dat,2,apply(dat,2,min),"-")
R=sweep(dat1,2,apply(dat,2,max),"/")
```

```
R0=sweep(R,2,apply(R,2,sum),"/")
H=-apply(R0*log(R0+0.00000001),2,sum)/log(31)
W=(1-H)/(8-sum(H)) #计算得到权重
Si=apply(dat*W,1,sum) #计算综合得分
cbind(Si,rank(-Si))
mapcol(chinamap,(Si-min(Si))) #绘制地图并着色
#计算基尼系数
library(ineq)
pc=rep(max(PC)-PC,n) #处理数据
si=rep(Si-min(Si),n) #处理数据
Gini(pc)
Gini(si)
```