五一数学建模竞赛

承 诺 书

我们仔细阅读了五一数学建模竞赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与本队以外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其它公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们愿意承担由此引起的一切后果。

我们授权五一数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

参赛题号（从 A/B/C 中选择一项填写）： C

参赛队号： B18007528

参赛组别（研究生、本科、专科、高中）： 本科

所属学校（学校全称）： 兰州大学

参赛队员： 队员 1 姓名： 单闯

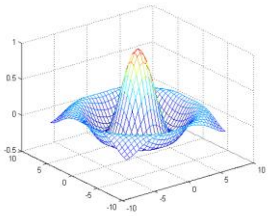
队员 2 姓名： 陈冠斌

队员 3 姓名： 李彦峰

联系方式： Email：2337818164@qq.com 联系电话：15294203791

日期： 2018 年 5 月 1 日

五 一 数 学 建 模 竞 赛



题 目：C题 江苏省本科教育质量综合评价

关键词： **量化、标准化、主成分分析、指标因子**

摘 要：

在人才资源越来越重要的当下，大学的本科教学质量成为关系到地方发展的重要因素。如何评价一个地区的本科教育质量，并采取相应措施缩小各地差异，更是关系到地区发展的重要问题。本文针对题中的要求，提出了对江苏各地级市本科教学质量评价的模型，并对提升江苏省本科教育质量提出了政策建议。

针对问题一，本文对题目中给出的 9 个指标划分了量化方式，并对这 9 个指标给出了具体的评价方式，并且通过自主搜集数据，完成了对江苏省 13 个地级市各指标的量化处理和相应的数据分析。

针对问题二，本文基于问题一所完成的量化处理结果，采用主成分分析的方式，计算出各指标对于综合评价结果的影响权重，构建了基于问题描述中的 9 个指标的数学模型，并采用该数学模型计算出各地级市的得分。

针对问题三，本文具体分析了可删减指标的出现情况，将可删减指标分为“低影响因子”和“高相关性指标”两类。通过比较各指标之间的相关系数确定了高相关性指标的去除，并通过原模型系数确定了低影响因子指标的去除。本文认为，去除了“师资队伍与结构”和“生师比”两个指标后，模型的评价结果仍然能保持一致。而后对删减指标后的模型进行了验证和解释，且验证结果符合预期。

针对问题四，本文首先通过原数学模型中各项指标系数初步判定关键指标为“专业建设与教学改革”，而后在后续评价中重新对 9 项指标进行了去除测试，并分别进行主成分分析重构模型，获得各地级市得分后计算各城市之间的方差。发现缩小 13 市的差距的关键指标为“双一流学科建设”，而缩小除南京外的 12 市的差距的关键指标为“专业建设与教学改革”。

针对问题五，我们结合前面建模的结论并分析数据规律，对江苏省提高本科教学质量提出了政策建议。

1. 问题重述

1.1 问题背景

随着中国改革开放的不断深化，人才的作用愈发凸显。高校作为高层次人才的主要输送地，各地也争相提高高校的本科教学质量。江苏省作为中国的教育大省，更是需要有一套合理的评价体系来评估各市高校的本科教学质量，从而帮助地方政府做出相应的政治决策，以培养更优质的人才，促进地方发展。

1.2 数据收集

本文所用数据均通过互联网采集自各高校官网、江苏省教育厅、江苏省统计局及百科等公开渠道，为保证各校数据的比较在同一年限下进行，均采集各校在2017年的统计数据。

1.3 问题要求

根据上述题目背景，并结合自主收集的各项数据，题目要求建立数学模型以讨论下列问题：

1. 查找资料，对如下9个指标（本科院校数量、招生人数、师资队伍与结构、生师比、教学条件与利用、专业建设与教学改革、学生就业、科研投入与产出、双一流学科建设）进行量化处理并完成表格，进行相关的数据分析。

2. 利用问题一中的指标数据，建立数学模型，对江苏省13个地级市的本科教育质量进行综合评价。

3.建立数学模型，找出问题一中可减少的指标并将其删去，并使得评价结果基本保持不变。

4. 建立数学模型，找出某个指标，满足当该指标普遍改善时能够缩小各市本

科教育的差异，并验证该模型的有效性。

5. 结合前四个问题的讨论，对提升江苏省本科教育质量提出政策建议。

2. 模型假设

1. 假设科研的投入和产出是成正比关系。
2. 假设高校本科教育质量由问题一中的 9 个指标体现。

3. 符号说明

符号	符号说明
Y	总体得分
y	主成分
x_j	第 j 个指标
x_{ij}	第 i 个城市的第 j 个指标
μ	均值
u	特征向量
λ	特征值
S	方差
R	相关系数矩阵
r_{ij}	第 i 个指标和第 j 个指标的相关系数
α	累计贡献率

4. 问题分析

4.1 问题一分析

问题一要求我们对 9 个指标（本科院校数量、招生人数、师资队伍与结构、生师比、教学条件与利用、专业建设与教学改革、学生就业、科研投入与产出、双一流学科建设）进行量化处理以满足后续建模需求。这其中要解决两个问题：

1. 因题目未给出具体数据，故需要自行查找相应数据。在查找时，需要保证各校数据是在同一年内的统计数据，以保证后续评价时是在同一时间年限下进行。

2. 部分指标例如“专业建设与教学改革”并无确切的量化方式，而部分指标如“本科院校数量”是有确定的值可以使用。故在做量化处理时，需要为无确切量化方式的指标设定评价体系，计算出具体的量化值。

4.2 问题二分析

问题二要求我们利用问题一中的 9 个指标，对江苏 13 个地级市的本科教育质量进行综合评价。在此，我们采用主成分分析的方法建立评价模型。首先搜集到了关于各个城市的详细数据，之后再根据数据将模型具体化。开始，需要对所有的数据进行标准化处理，以消除量纲影响。但是，我们注意到南京的各个指标过高，对于其它各市形成压倒性优势，直接参与模型建立，在标准化的过程中会极大地干扰模型的合理性，所以南京市不参与模型的建立，而是在模型建立好之后直接将数据带入。

4.3 问题三分析

问题三要找出 9 个指标中的冗余指标，即将这些指标去除后对于评价结果影响不大，这说明这些冗余指标可能有以下两种情况：

1. 所代表的评价意义能够被其他指标所替代。则该问题被简化成：找出那些能够互相取代的指标并且去除其中的冗余指标，更改其相对应指标的评价系数从

而保证评价结果保持不变。而能够互相取代的指标其增长趋势较为趋同，即相关性比较高，可以使用相关性分析进行寻找。

2. 所代表的评价意义对于最终评价结果贡献较小，即该指标的变化对于最终结果影响不大，在本题中意味着该指标对本科生教学质量的优劣影响极小，去除这类指标对于最终结果不大。

4.4 问题四分析

问题四要求确定一个关键指标，这个指标的普遍改善可以尽可能的缩小江苏省 13 个市的本科教育发展差异。意思是这个指标对于 13 个市来说均在相同条件下时，各市之间的差异能够被尽可能缩小。则该指标应当满足以下条件：

1. 该指标必须对于本科教学质量的提高足够重要，即占综合评价得分的权重足够大。

2. 该指标与其他指标有较高的正向相关性，能够较为全面地影响并提高本科教学质量。

3. 对于最终得分差距较大的两市，该指标的差距也较大，反之则该指标差距不明显。

5. 模型建立与求解

5.1 问题一：指标量化与分析

5.1.1 量化处理分类

首先，题目给出了 9 个指标，部分指标意思较为明确，其含义即为对应的数据，则认为该指标是可以直接通过数据进行量化的，定义该量化处理方式为“直接量化”。

但是部分指标无法直接用相应数据表示，需要手动给出评价系统，并通过评价系统内的子指标对其进行评判，从而转化为可量化的数据形式，对于这种指标

的处理方式，将其定义为“间接量化”。

5.1.2 计算方式

对于直接量化的方式，根据题目意思找到相应的直接数据，例如“本科院校数量”，“招生人数”等。对于间接量化方式，考虑哪些可量化因素会对该指标产生较大影响，从而对该指标设立具体的评价标准并计算该指标下得分。填写表格如下：

序	指标名称	量化处理方式	计算依据	备注
1	本科院校数量	直接量化	本科院校数量	
2	招生人数	直接量化	本科生招生人数	
3	师资队伍与结构	间接量化	教授人数*3+副教授人数*1.5+讲师*0.5	不同的老师等级，教学水平也不同，所以设置不同的权重
4	生师比	直接量化	老师数量/学生数	
5	教学条件与利	间接量化	教学财政投入(亿)	
6	专业建设与教学改革	间接量化	一级学科*3+ 二级学科*2+ 三级学科*1+ 新设学科*2	一级学科是学科大类，二级、三级学科是学科小类，要根据权值来体现它们之间的差异。新设学科可以体现教学改革。
7	学生就业	间接量化	毕业 5 年后学生就业率*平均工资	就业率与平均工资共同体现学生就业情况

8	科研投入与产出	间接量化	发表 SCI 论文数量	科研投入与产出可以通过论文发表数来体现，高水平层级论文可以体现出科研投入与产出的差距
9	双一流学科建	直接量化	双一流学科数量	

图 1-1 量化处理表

5.1.3 量化结果

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
c_1	35	329483	51392	20.4	276	93	9000	26772	38
c_2	6	53992	10089	17.1	54	78	7800	3078	2
c_3	7	67413	9872	18.7	73	70	8000	2317	2
c_4	2	34462	7802	18.9	39	67	7600	1136	1
c_5	5	19226	6752	18.5	32	64	7400	14	0
c_6	3	22405	6892	15.8	29	67	7600	1390	0
c_7	3	31641	6753	19.2	21	63	6800	815	0
c_8	3	39152	6309	16.6	19	59	6500	511	0
c_9	2	25760	4579	15.4	26	53	7000	2134	0
c_{10}	2	53992	4573	18.7	21	51	6400	10	0
c_{11}	2	18334	4678	15.2	17	48	6600	4	0
c_{12}	2	11545	4326	18.1	12	49	7000	7	0
c_{13}	1	4796	2105	21.8	7	49	6200	1	0

图 1-2 量化结果

c_1~c_{13}分别为南京、徐州、苏州、无锡、泰州、扬州、南通、常州、镇江、淮安、盐城、连云港、宿迁。

x_1~x_9分别为本科院校数量、招生人数、师资队伍与结构、生师比、教学条件与利用、专业建设与教学改革、学生就业、科研投入与产出、双一流学科建设。

可以大致看出，南京市各项数据均优于其他 12 个地级市，但其他 12 个地级市教育指标较为平均。

5.2 问题二：建立模型并对各市进行评价

1. 首先搜集到了关于各个城市的详细数据，如图 1-2。其次，我们对所有的数据进行标准化处理，以消除量纲影响。

做法为将第 i 个城市的第 j 个指标$\tilde{x}_{ij}$标准化处理，变成$\tilde{x}_{ij}$。

$$\tilde{x}_{ij} = \frac{(x_{ij}-\mu_j)}{s_j} \quad i = 1,2,\ldots,n, j = 1,2 \ldots m \qquad \text{figure2- 1}$$

其中：

μ_j为每个城市的相同指标的均值：

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

s_j为每个城市的相同指标的标准差：

$$s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\mu_j\right)^2} \quad j = 1,2,\ldots,m$$

n 为城市数目，m 为指标数目。

标准化的结果如表图 2- 2。我们发现南京的各个指标过高，对于其它各市形成压倒性优势，直接参与模型建立，在标准化的过程中会极大地干扰模型的合理性。验证结果如表 2-3，我们发现有很多指标之间的相关系数大于 0.95，甚至为 0.99，这是极不符合数学常识的，所以我们决定南京市不参与模型的建立，而是在模型建立好之后直接将数据带入。即 n=12，m=9。

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
c_1	3.26	3.25	3.28	1.21	3.22	2.33	2.27	3.29	3.32
c_2	0.04	-0.01	0.03	-0.48	0.08	1.19	0.74	0.02	-0.13
c_3	0.15	0.15	0.01	0.34	0.35	0.58	0.99	-0.09	-0.13
c_4	-0.4	-0.24	-0.15	0.44	-0.13	0.35	0.48	-0.25	-0.22
c_5	-0.07	-0.42	-0.23	0.24	-0.23	0.12	0.23	-0.4	-0.32
c_6	-0.29	-0.38	-0.22	-1.14	-0.27	0.35	0.48	-0.21	-0.32

c_7	−0.29	−0.27	−0.23	0.6	−0.38	0.05	−0.54	−0.29	−0.32
c_8	−0.29	−0.19	−0.27	−0.73	−0.41	−0.26	−0.93	−0.34	−0.32
c_9	−0.4	−0.34	−0.4	−1.35	−0.31	−0.71	−0.29	−0.11	−0.32
c_{10}	−0.4	−0.01	−0.4	0.34	−0.38	−0.86	−1.05	−0.4	−0.32
c_{11}	−0.4	−0.43	−0.39	−1.45	−0.44	−1.09	−0.8	−0.41	−0.32
c_{12}	−0.4	−0.51	−0.42	0.04	−0.51	−1.02	−0.29	−0.41	−0.32
c_{13}	−0.51	−0.59	−0.6	1.93	−0.58	−1.02	−1.31	−0.41	−0.32

图 2- 1 影响本科教育质量的各城市数据(规范化)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	0.99	0.99	0.35	0.99	0.8	0.78	0.99	0.99
x_2	0.99	1	0.99	0.34	0.99	0.77	0.74	0.98	0.99
x_3	0.99	0.99	1	0.33	0.99	0.81	0.78	0.99	0.99
x_4	0.35	0.34	0.33	1	0.34	0.23	0.12	0.32	0.37
x_5	0.99	0.99	0.99	0.34	1	0.82	0.82	0.99	0.98
x_6	0.8	0.77	0.81	0.23	0.82	1	0.91	0.77	0.74
x_7	0.78	0.74	0.78	0.12	0.82	0.91	1	0.75	0.72
x_8	0.99	0.98	0.99	0.32	0.99	0.77	0.75	1	1
x_9	0.99	0.99	0.99	0.37	0.98	0.74	0.72	1	1

图 2- 2 相关系数矩阵(使用南京的数据)

2. 之后对标准化的数据计算相关系数矩阵R：

$$R = \left(r_{ij}\right)_{m\times m}$$

$$r_{ij} = \sum_{k=1}^{17} \frac{\tilde{x}_{ki} * \tilde{x}_{kj}}{12-1} \quad i,j = 1,2,\dots,m$$

计算结果如表 2-3：

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
x_1	1	0.66	0.85	−0.06	0.87	0.78	0.75	0.61	0.75
x_2	0.66	1	0.72	−0.06	0.77	0.58	0.44	0.58	0.72
x_3	0.85	0.72	1	−0.18	0.89	0.94	0.85	0.72	0.81
x_4	−0.06	−0.06	−0.18	1	−0.07	−0.04	−0.19	−0.29	0.08
x_5	0.87	0.77	0.89	−0.07	1	0.81	0.86	0.77	0.9
x_6	0.78	0.58	0.94	−0.04	0.81	1	0.83	0.73	0.74
x_7	0.75	0.44	0.85	−0.19	0.86	0.83	1	0.71	0.74
x_8	0.61	0.58	0.72	−0.29	0.77	0.73	0.71	1	0.77
x_9	0.75	0.72	0.81	0.08	0.9	0.74	0.74	0.77	1

图 2- 3 相关系数矩阵(不使用南京的数据)

比较图 2-3 和图 2-4，我们发现相关系数的数值有了很大的改观，体现了我们新模型的合理性。

相关系数矩阵反应了各个指标之间的相关性，以此为依据，可以进行消除重复影响，这也是主成分分析的精髓所在。

3. 计算相关系数矩阵R的特征值$\lambda_1 \sim \lambda_m$与特征向量$u_1 \sim u_m$。

$$u_j = \left[u_{1j}, u_{2j}, \dots, u_{mj}\right]^T \quad j = 1,2,\cdots,m$$

计算结果如表 2-4：

	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9
$\tilde{x}_1$	0.35	0.08	0.03	−0.5	−0.34	0.68	0.02	0.19	−0.07
$\tilde{x}_2$	0.3	0.11	−0.78	−0.14	0.21	−0.23	0.28	0.24	−0.2
$\tilde{x}_3$	0.38	−0.04	0.07	−0.27	0.3	−0.19	−0.27	0.12	0.75
$\tilde{x}_4$	−0.05	0.93	0.16	0.15	0.09	0.05	0.21	0.09	0.15
$\tilde{x}_5$	0.38	0.07	−0.07	0	−0.35	−0.16	0.26	−0.78	0.14
$\tilde{x}_6$	0.36	0.04	0.32	−0.09	0.66	0.09	−0.09	−0.28	−0.48
$\tilde{x}_7$	0.35	−0.11	0.49	0	−0.26	−0.46	0.36	0.42	−0.17
$\tilde{x}_8$	0.33	−0.23	−0.05	0.71	0.14	0.43	0.27	0.12	0.21
$\tilde{x}_9$	0.36	0.21	−0.11	0.36	−0.3	−0.13	−0.72	0.09	−0.22

图 2- 4 特征向量

$\tilde{x}_1 \sim \tilde{x}_9$分别为$x_1 \sim x_9$的标准化指标变量。

$y_1 \sim y_9$分别为按照权重从大到小排序的主成分。

4. 接下来，可以得到主成分y_j与原指标$\tilde{x}_j$的表达$(j = 1,2,\cdots,m)$：

$$y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \cdots + u_{m1}\tilde{x}_m$$

$$y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \cdots + u_{m2}\tilde{x}_m$$

$$\cdots$$

$$y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \cdots + u_{mm}\tilde{x}_m$$

Figure2-2

但是所有主成分不应该全部使用，应当保留贡献率大的，剔除贡献率小的。

特征值反应的是每个主成分y_j的贡献率b_j和累计贡献率α_j（$j = 1,2,\cdots,m$）：

$$\mathrm{b}_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_K}$$

$$\alpha_j = \frac{\sum_{k=1}^{j} \lambda_k}{\sum_{k=1}^{m} \lambda_K}$$

计算结果如下：

6.34	1.10	0.61	0.41	0.25	0.17	0.10	0.02	0.01

图 2- 5 特征值$\lambda_1 \sim \lambda_m$

70.40	12.22	6.75	4.52	2.78	1.87	1.11	0.21	0.14

图 2- 6 贡献率

70.40	82.61	89.37	93.89	96.67	98.53	99.64	99.86	100.00

图 2- 7 累积贡献率

计算综合评价值：

$$\mathrm{Y} = \sum_{j=1}^{g} \mathrm{b}_j * y_j \qquad \text{Figure2- 3}$$

g 是选取的主成分的个数，选用第 1 个主成分至第 g 个主成分(按照贡献率递减)，使累积贡献值大于某个阈值。

当 g 等于 3 的时候，如图 2-8,累积贡献率达到89.37%，所以我们保留前三个主成分，剔除后面 6 个主成分。

结合 Figure2- 1、Figure2- 2、Figure2- 3

计算综合评价值：

$$\mathrm{Y} = \sum_{i=1}^{m} \left(\sum_{j=1}^{g} u_{ij} * \tilde{x}_i \right) = \sum_{i=1}^{m} c_i * \tilde{x}_i \qquad \text{figure2-4}$$

$$\mathrm{Y} = \sum_{i=1}^{m} \frac{c_i}{s_i} * x_i - \sum_{i=1}^{m} \frac{c_i * \mu_i}{s_i} \qquad \text{figure2-5}$$

最后的结果为(matlab 处理)：

$$Y = 70.40y_1 + 12.22y_2 + 6.75y_3 \qquad \text{Figure2- 6}$$

$$Y = 0.2594\tilde{x}_1 + 0.1747\tilde{x}_2 + 0.2686\tilde{x}_3 + 0.0087\tilde{x}_4 + 0.2751\tilde{x}_5 + 0.2791\tilde{x}_6 + 0.2643\tilde{x}_7 + 0.2022\tilde{x}_8 + 0.2704\tilde{x}_9 \qquad \text{Figure2- 7}$$

$$Y = 0.14018x_1 + 0.00001x_2 + 0.00011x_3 + 0.04590\,x_4 + 0.01476x_5 + 0.02838x_6 + 0.00044x_7 + 0.00019x_8 + 0.34101x_9 - 7.86283 \qquad \text{Figure2- 8}$$

5. 将 12 个城市的标准化指标代入 Figure2-7，将南京的原始指标代入 Figure2-8，最终得到综合评价值。为了更好地体现，对数据进行标准化处理，如图 2-9。

排名	城市	得分	排名	城市	得分
1	南京	35.639	1	南京	100
2	苏州	3.4675	2	苏州	14.9555
3	徐州	3.1558	3	徐州	14.1315
4	无锡	0.9053	4	无锡	8.1822
5	扬州	0.2462	5	扬州	6.4401
6	泰州	0.2163	6	泰州	6.3611
7	南通	-0.223	7	南通	5.1996
8	常州	-0.6573	8	常州	4.0516
9	镇江	-0.7148	9	镇江	3.8996
10	淮安	-1.0994	10	淮安	2.8829
11	连云港	-1.4721	11	连云港	1.8978
12	盐城	-1.6346	12	盐城	1.4682
13	宿迁	-2.19	13	宿迁	0

图 2- 4 各城市综合评价值分数及排名

观察结果发现，南京与其它城市的综合评价值看似差距很大，但是符合实际情况。

5.3 问题三：冗余指标的去除

5.3.1 高相关性指标的去除

对于两个指标 a 和 b 来说，可能其内在意义在本质上是相同的，或者其对于最终结果产生的影响是通过相同途径发生的，在这种情况下，可以使用 a 指标来代替 b 指标，将 b 指标删去，并通过修改 a 指标在最终模型中的系数来保证 a 指标发挥和原 a、b 指标共同作用的效果。本模型定义这样的 a、b 指标为高相关性指标。

皮尔逊积矩相关系数用于度量两个变量 X 和 Y 之间的相关性，其值介于-1~1，其中 1 表示完全正相关，0 表示没有相关关系，-1 表示完全负相关。

总体相关系数：

$$\rho_{X,Y}=\frac{cov(X,Y)}{\sigma_X\sigma_Y}=\frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}$$

样本相关系数：

$$\gamma=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\Big(\sum_{i=1}^{n}(X_i-\bar{X}\Big)^2}\sqrt{(\sum_{i=1}^{n}\Big(X_i-\bar{X}\Big)^2}}$$

观察图 2- 8 相关系数矩阵，我们发现第三项“师资队伍与结构”和第六项“专业建设与教学改革”相关性非常高，其相关系数高达 0.94，而其他任意两不同项的相关系数均不超过 0.9。在去除冗余项时，为了尽量保证去除该项后对最终结果影响不大，因此在本模型中，仅认为相关系数超过 0.9 的两项认为是高度相关项，可以选择其中一项作为冗余项进行删减。

故在此，选择第三项“师资队伍与结构”删减，因为它与第六项“专业建设与教学改革”相关性过高，满足本模型对于高相关性指标去除的条件。

5.3.2 低影响因子的去除

对于以下评价模型：

$$Y = \sum_{i=1}^{n} c_i \tilde{x}_i$$

其中Y为综合评价值，n为评价指标的个数，$\tilde{x}_i$为标准化后的指标值，c_i为该指标值的系数。在这些指标中可能存在某个指标$\tilde{x}_k$，其系数c_k远远小于其他指标的系数，此时我们认为该模型中，该指标对于最终结果影响极小，将其定义为低影响因子，为了简便模型可以将其作为删减指标去除。

观察 figure2- 7，可以看到第四项指标“生师比”的系数仅为 0.0087，远远小于其他指标的系数，故在这里将其认为是本模型的低影响因子，将其去除。

5.3.3 新模型的评价

新模型去除了指标三和指标四，重新按照问题二的建模方法进行主成分分析，得到最后的评分标准，其评分计算模型为：

$$Y = 0.2614x_1 + 0.3226x_2 + 0.3246x_5 + 0.2641x_6 + 0.2373x_7 + 0.3144x_8 + 0.3270x_9$$

将各市数据带入新模型，并进行标准化得到如下结果：

排名	城市	得分
1	南京	100.0000
2	苏州	14.3254
3	徐州	13.4843
4	无锡	7.3241
5	扬州	5.9112
6	泰州	5.2192
7	南通	4.6054
8	常州	4.5013
9	镇江	3.9568
10	淮安	3.3347
11	盐城	1.6654

12	连云港	1.6262
13	宿迁	0.0000

图 3-2 删减后的新模型下得分情况

与问题二中的得分情况相比较，仅仅是盐城和连云港的排名发生了调换，其他排名均保持不变。而盐城和连云港的得分原本差距过小，删减指标后发生了排名变化属于可接受的误差范围，故可认为新模型满足题目关于评价结果基本一致的要求

5.3.4 新模型的解释

对于“师资队伍与结构”的去除，因为其与“专业建设与改革”有极强的相关性，从第一问的量化处理过程来看，“师资队伍与结构”由教授、副教授、讲师的人数计算而来，“专业建设与改革”由等次学科数计算而来。而一个学校其专业建设程度本身即与教师队伍有着极强的相关性，故而从数学上显示出二者相关系数极高，因此在删减指标时，可以互相代替对方对于最终结果的影响。

对于“生师比”指标的去除，根据主成分分析的原理解释，其在最终评价得分模型中的系数极小，说明了该项指标对于各省市分布比较平均且没有较为明显的规律变化，与误差的分布类似，可以认为对于最终的结果影响极小，可以去除。

5.4 问题四：关键指标查找

5.4.1 查找方式

对于本问题，确定一个关键的指标，使得该指标值的普遍改善能够尽可能缩小江苏省 13 个地级市本科教育发展的差异，即选择影响综合评价值最大的指标，这个指标要满足占综合评价值的权重和与其它指标的相关系数尽量大。如果这个指标普遍改善，即假设各个城市的该指标是相同的，则在实际处理中，可以把该指标从评价体系中去掉，并采用问题 2 的模型进行计算分析，分析处理结果。

5.4.2 处理过程

根据问题二求得的加权计算式：

$$Y = 0.2594\mathrm{x}_1 + 0.1747x_2 + 0.2686x_3 + 0.0087x_4 + 0.2751x_5 + 0.2791x_6 + 0.2643x_7 + 0.2022x_8 + 0.2704x_9$$

第六项“专业建设与教学改革”在综合评价值的权重是最大的，说明其本身对于最终评价系统的影响最大。另外，根据问题三所求的相关性矩阵，其与其他几项指标的相关性也是最强的，甚至与第三项指标有着高达 0.94 的相关性导致其被删减（问题 3 中）。故可初步认为“专业建设与教学改革”较为关键，下面进行更进一步讨论。

5.4.3 关键指标的评价与修正

我们采用求方差的方式体现综合评价值的差异。

$$\sigma^2 = \frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}$$

（σ^2为总体方差，Y_i为第 i 个变量的综合评价值，μ为总体均值，N为变量数目）

在最终评价式分别去除 9 各指标，重新进行问题二方法下的主成分分析，并带入数据获得新的结果，并将数据分为包括南京和不包括南京。并对结果进行方差分析，从而对各市得分差距进行量化：

	x_0	x_1	x_2	x_3	x_4
r_1	648.9031	650.6276	655.5002	656.6750	645.3656
r_2	20.2337	20.0601	18.5750	20.3434	18.0930
	x_5	x_6	x_7	x_8	x_9
r_1	651.2096	654.3694	657.2231	647.8921	628.9659
r_2	18.4360	15.8018	16.4577	22.2146	35.3005

图 4-1 得分方差

r_1：包括南京在内的所有城市之间的综合评价值的差异

r_2：除南京外的其它城市之间的综合评价值的差异

x_0：不删除任何指标

$x_1 \cdots x_9$：被删除的关键指标

$$r_1 = \frac{\sum_{i=1}^{N}(Y_i - \mu)^2}{N}, i = 1 \cdots N$$

$$r_2 = \frac{\sum_{i=2}^{N}(Y_i - \mu)^2}{N}, i = 1 \cdots N$$

N 为城市数目，Y_i为第 i 个城市的综合评价值

可以看到，在计算组内方差时，只要考虑了南京的得分，无论将哪个指标视为关键指标进行处理，其最后的方差绝对值总是较大，但相对而言，将第九项“双一流学科建设”视为关键指标时，方差最小，但与此同时却导致其他 12 市数据结果方差最大。而若在求方差时不考虑南京市的数据，将第六项“专业建设与教学改革”视为关键指标时方差最小。下面列出不同情况下的得分数据。

保留所有指标			去掉专业建设与教学改革			去掉双一流学科建设		
排名	城市	得分	排名	城市	得分	排名	城市	得分
1	南京	100.0000	1	南京	100.0000	1	南京	100.0000
2	苏州	14.9555	2	苏州	13.7464	2	苏州	20.1337
3	徐州	14.1315	3	徐州	12.2547	3	徐州	18.9444
4	无锡	8.1822	4	无锡	6.9841	4	无锡	11.2057
5	扬州	6.4401	5	泰州	5.3192	5	扬州	10.0081
6	泰州	6.3611	6	扬州	5.1531	6	泰州	9.8381
7	南通	5.1996	7	南通	4.2146	7	南通	8.0055
8	常州	4.0516	8	镇江	3.6567	8	常州	6.2766
9	镇江	3.8996	9	常州	3.3484	9	镇江	6.0687
10	淮安	2.8829	10	淮安	2.8063	10	淮安	4.3998
11	连云港	1.8978	11	连云港	1.9054	11	连云港	2.9956

12	盐城	1.4682	12	盐城	1. 5413	12	盐城	2.3691
13	宿迁	0.0000	13	宿迁	0. 0000	13	宿迁	0.0000

图 4-2 关键指标去除后得分

具体分析来看，我们认为，南京市本科教学质量在多方面远远优于其他各市，无法简单地通过某个关键指标来大范围地缩小差距，但相对而言，“双一流学科建设”是对于缩小南京市与其他各市差距最为关键的。而在不考虑南京市时，其他地级市的学校并无特别多的顶尖学校进行双一流建设，双一流建设成果差距不大，故这个指标对于这些地级市来说并不关键，而“专业建设与教学改革”却关系到教学水平的优劣，是造成地级市本科生教学水平差异的主要影响因素。

因此，我们的结论是：

1. “双一流学科建设”的普遍改善能够尽可能缩小江苏省 13 个地级市本科教育发展的差异，其主要作用于南京市和其他 12 市的差距缩小。

2. “专业建设与教学改革”的普遍改善尽管不能缩小江苏省 13 个地级市本科教育发展的差异，却对缩小除南京市外的江苏省 12 个地级市本科教育发展的差异起到明显的作用。

5.5 问题五：政策建议

问题 2 中，我们观察发现南京市与其他市的本科教育质量相差很大。这样的优点是可以“先富带动后富”，南京市走在教学前沿，其它市可以借鉴教学经验，避免前车之鉴，使教学处于发展快车道。但是另外一方面，由于南京市与其它市本科教育质量的巨大差距，有可能导致其它市缺乏人才吸引能力，甚至导致从其他市到南京市的人才流失，使南京市与其它市的本科教育质量的差距更大。一个比较好的解决方式是加强南京市与其他市本科教育质量的合作，南京市需要更加广阔的教学空间，实现教育范围扩大化，其他市也需要更加优质的教学指引，实现教育质量提升化，从而达到双赢的局面。

问题 3 中，生师比是 9 个指标中的最小影响因子，这也给江苏省一些为了想快速提高本科教育质量而大幅扩张招生数量的学校提一个醒，并不是说扩张招生就能提高本科教育质量，如果一昧扩张招生，没有考虑学校的教学能力，那么反而得不偿失。

同时问题 3 中，通过观察相关系数矩阵，我们发现某些指标之间的相对系数很大，如师资队伍与结构和专业建设与教学改革，师资队伍与结构的促进能带动专业建设与教学改革的促进，同样的，专业建设与教学改革的促进能带动师资队伍与结构的促进，所以江苏省的一些学校如果想要提高本科教育质量，不能孤立地只看一两个指标，而是应该把这些指标作为一个整体，相互促进。实际上，除了生师比，其它指标之间都是正相关，其中一个指标的促进就能带动其它指标的促进。如果没有条件同时提高所有指标，不妨专注于某几个指标，把这几个指标搞好，自然地其它指标也能提高。如提高专业建设与教学改革，就能同时很大程度上地带动师资队伍与结构、学生就业、科研投入与产出、双一流学科建设的发展。

问题 4 中，我们得出两个结论，一个是双一流学科建设的普遍改善能够尽可能缩小江苏省 13 个地级市本科教育发展的差异，另一个是专业建设与教学改革的普遍改善对缩小除南京市外的江苏省 12 个地级市本科教育发展的差异起到明显的作用。为了平衡江苏省本科教育质量，使江苏省本科教育质量健康化发展，避免教育发展失衡的情况，两点一定要双管齐下，否则会出现其他市与南京市本科教育质量的差距缩小，但是出现一两个市的本科教育质量远远落下于其他市，使教育失衡更加明显的情况；或者是出现除南京市的其他市在齐头前进，但是与南京市的本科教育质量的差距不仅没有缩小，反而常态化增大的情况。

综合问题 2,3,4，我们认为，江苏省本科教育质量的提高仍然任重而道远，解决南京市与其他市的本科教育质量差距巨大的问题是刻不容缓的；而在发展过程中，应该避免一些为了提高教学质量而急功近利的事情，如一昧扩张招生而没有考虑学校的教学能力；同时要用科学的方式提高本科教育质量，如果不能同时兼顾所有教学指标的提高，不妨专注于某一两个教学指标的提高，而这一两个指标又能自然地带动其他指标的发展。我们坚信并祝福，在江苏省各高校的努力下，

江苏省本科教育质量可以更上一层楼，作为教育强省，为中国培养更多的高端人才。

6. 模型评价

6.1 优点：

1. 采用主成分分析法，可消除评价指标之间的相关影响。因为主成分分析在对原指标变量进行变换后形成了彼此相互独立的主成分，使得评价结果更为合理。

2. 对原始数据进行了标准化处理，消除了量纲影响。

3. 从相关系数入手，消除了重复项。对于最后的评价模型中的系数最小的一项，结合实际意义予以去除。

4. 考虑到南京的特殊情况，即本科教育对于江苏省其他市形成压倒性优势，在模型建立时会产生干扰，所以南京不参与原始评价模型的建立，而是在之后直接用模型进行评价。

6.2 缺点：

1. 因部分学校未即时在官方平台公布相关信息，导致建立模型时使用了部分往年数据，仅仅只能一定程度上代表该学校的发展情况。

2. 因为南京的情况较为特殊，容易对模型产生较大干扰，但未找到较好的去除干扰的方法，故在主成分分析时未引入其相关数据，显得较为草率，可能导致模型不够精准。

3. 在量化处理过程中，可能有部分指标评价不够全面，导致有些地方欠缺考虑，无法很好地量化该指标的具体情况。

7. 参考文献

[1] 姜启源，谢金星，叶俊.数学模型（第三版）[M]. 北京：高等教育出版社，2003.

[2] 司守奎，孙兆亮.数学建模算法与应用（第二版）[M]. 北京：国防工业出版社，2016.

8. 附录

Matlab 代码：

```
clear
clc

[data,city]=xlsread('data.xls');%data:data of city
                                  %city:city name(col) & component name(row)
    %problem-2
%数据相差太大，必须剔除南京这一项，否则结果不准确
%而且南京这一项不能参与标准化，即该语句放在标准化的前面
nj=data(1,:); data(1,:)=[];     %delete NanJing city(2,:)=[];

    % %delete component(s)
    % %problem-3
% data(:,4)=[]; nj(4)=[];              %delete 4
% data(:,[4,3])=[]; nj([4,3])=[]; %delete 4 and 3
    % %problem-4
%data(:,6)=[]; nj(6)=[];              %the most important component 6
%data(:,9)=[]; nj(9)=[];              %the most important component 9
%data(:,9)=[]; nj(9)=[];              %1~9

data1=data;                          %for further using
data=zscore(data);                   %standardize
r=corrcoef(data);                    %correlation matrix
[vec1,lamda,rate]=pcacov(r);         %principal component analysis
```

```
contr=cumsum(rate);
f=repmat(sign(sum(vec1)),size(vec1,1),1);
vec2=vec1.*f;
num=3;                                  %the amount of principal component
df=data*vec2(:,1:num);
tf=df*rate(1:num)/100;                  %score except NanJing

count=size(data,1);                     %the amount of city(except NanJing)
n=size(data,2);                         %the amount of principal component
a=zeros(n+1,1);
a(1:n)=vec2(:,1:num)*rate(1:num) ./100;    %y=a1*x1'+a2*x2'+…+an*xn'
                                           %x1',x2',…,xn'(standardization)
%print
fprintf('weight:\n');
for i=1:n
    fprintf('%.4f\t',a(i));
end
fprintf('\n');

for i=1:n
    a(i)=a(i)/std(data1(:,i));
    a(n+1)=a(n+1)-a(i)*mean(data1(:,i));
end
y=zeros(count,1);
for i=1:count
    y(i)=data1(i,:)*a(1:n)+a(n+1);         %y=tf(judge if the result is right)
end
% if y~=tf
```

```
%       fprintf('wrong');
%       return;
% end

    %problem-2
    %1.add NanJing
nj_score=nj*a(1:n)+a(n+1);
tf=[nj_score;tf];
count=count+1;
    %2.delete NanJing
% city(2,:)=[];

[stf,ind]=sort(tf,'descend');          %score & corresponding rank
delete('result.xls');                   %delete previous data
xlswrite('result.xls',[{'排名'},{'城市'},{'得分'}],1,'A1:C1');
xlswrite('result.xls',(1:count)',1,strcat('A2:A',num2str(count+1)));
xlswrite('result.xls',city(ind(:)+1,1),1,strcat('B2:B',num2str(count+1)));
xlswrite('result.xls',stf,1,strcat('C2:C',num2str(count+1)));

%print
[result1,result2]=xlsread('result.xls');
fprintf('\n%s\t%s\t\t%s\n',result2{1,1},result2{1,2},result2{1,3});
for i=1:size(result1,1)
    fprintf('%d\t\t%s\t\t%.4f\n',result1(i,1),result2{i+1,2},result1(i,3));
end
fprintf('\n');

%normalize(0~100)
```

```
score_max=max(result1(:,3));
score_min=min(result1(:,3));
diff=score_max-score_min;
[result1,result2]=xlsread('result.xls');
fprintf('\n%s\t%s\t\t%s\n',result2{1,1},result2{1,2},result2{1,3});
for i=1:size(result1,1)
    result1(i,3)=(result1(i,3)-score_min)/diff*100;
    fprintf('%d\t\t%s\t\t%.4f\n',result1(i,1),result2{i+1,2},result1(i,3));
end
fprintf('\n');

fprintf('correlation matrix:\n');
for i=1:n
    for j=1:n
        fprintf('%.2f',r(i,j));
        if (j~=n)
            fprintf('\t');
        else
            fprintf('\n');
        end
    end
end
fprintf('\n');

fprintf('principal component analysis:\n');
for i=1:n
    for j=1:n
        fprintf('%.2f',vec2(i,j));
```

```
        if (j~=n)
            fprintf('\t');
        else
            fprintf('\n');
        end
    end
end

var(result1(:,3),1)
var(result1(2:end,3),1)

%{
Z=a1*y1+a2*y2+…+an*yn;                          %a=rate
yi=b1i*x1'+b2i*x2'+…+bni*xn';                          %b=vec2(principal   component
analysis:created by data)

xi'= ( xi-mean(xji) ) / std(xji);               %j=1~count
Z=C1*x1+C2*x2+…+Cn*xn                          %basic
%}
```