

# 河南省高等教育质量综合评价

## 摘 要

随着中国改革开放的实施,我国的综合实力不断增强,中国高等教育发展整体已进入世界中上水平。作为一个教育大省,河南省的高等教育发展在全国名列中等,而河南省 18 个地级市的高等教育质量发展也不平衡。

针对问题一,采用用上网查询数据的方式对本、专科院校数量、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、科研投入与产出等八个指标的数据情况进行查询和预处理,得出结论:通过图 2 的变化趋势可知,发达的市区具有更好的本科教育,总体来说,本科高校的科研、师资队伍与结构在教育质量综合评价中占主要作用。

针对问题二,由于从网上获取了相应的数据,所以在对河南省本科教育质量进行评价时可以采用 Topsis 法以及熵权法,由于一些数据是定性数据可能会我们的评价结果不够准确,所以又采用模糊综合评价法,对 Topsis 进行检测。得出结论:郑州市的高等教育质量最高、洛阳市次之、然后是焦作市,而三门峡、济源市、驻马店市的高等教育质量最差。总体上经济发达的地区,高等教育水平也越高。

针对问题三,采用多元回归分析,将变量添加到一个模型中,建立一个回归方程,对其进行描述性统计,判断模型是否合理,再使用向后逐步回归分析筛选关键变量。得出结论:本、专科院校数量、招生人数、师资队伍与结构、生师比、教学条件与利用、学生就业、省级重点学科专业这些指标重要程度低于其他指标,可以剔除。

针对问题四,通过建立数学模型,采用因子分析法,从 8 个指标得出 2 个主因子,通过权重比例,计算得到综合因子,从而得出结论:提高三主因子本、专科院校数量、师资队伍与结构、专业建设与教学改革与结构数量因子,学生因子普遍改善能够尽可能缩小河南省 18 个地级市本科教育发展的差异。

**关键词:** Topsis 法; 模糊综合评价; 多元回归分析; 因子分析法; 熵权法

## 一、问题重述

### 1.1 背景资料与条件

随着中国改革开放的实施，国家的综合实力不断增强，中国高等教育发展整体已进入世界中上水平。作为一个教育大省，河南省的高等教育发展在全国名列中等，而河南省 18 个地级市的高等教育质量发展不平衡。

### 1.2 需要解决的问题

1. 影响高等教育质量的指标有很多。通过查找资料，对如下 8 个指标（本、专科院校数量、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）进行量化处理(完成表 1)，并进行相关的数据分析。
2. 根据问题 1 中的指标数据，建立数学模型，对河南省 18 个地级市的高等教育质量进行综合评价。
3. 通过建立数学模型，分析在所给的 8 个指标中，可以减少哪些指标使得评价结果与问题 2 中的评价结果基本一致。
4. 通过建立数学模型，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河南省 18 个地级市高等教育发展的差异，并利用查找的数据资料验证所建立模型的有效性。
5. 结合前面的讨论给出有效提升河南省高等教育质量的政策建议。

## 二、问题分析

### 2.1 问题一的分析

题目中需要查找的数据很多，可以从统计年鉴，教育厅网站，或者其他论文中的查找相关数据，也可从相似研究论文中查看他们的参考文献。查找完相应的数据，再对本、专科院校数量、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、省级重点学科专业这几个指标进行分析处理，建立了 18 个地级市总体折线图，和描述性统计的直方图。

### 2.2 问题二的分析

问题二要求对河南省 18 个地级市（郑州市、开封市、平顶山市、洛阳市、商丘市、安阳市、新乡市、许昌市、鹤壁市、焦作市、濮阳市、漯河市、三门峡市、周口市、驻马店市、南阳市、信阳市、济源市。）本科教育质量进行综合评价。结合问题一，采用优劣解距离法（Topsis）分别进行对 8 个指标的处理，通过熵权法计算权重和得分建立了河南省本科教育质量模型，并进行了综合评价；并又使用模糊综合评价检验结果是否满足 Topsis 结果。

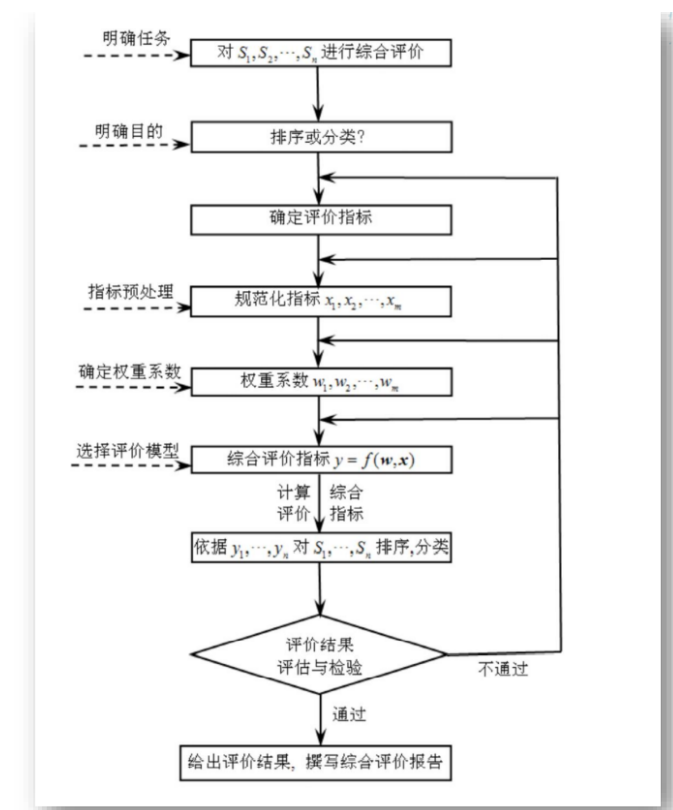


图 1 Topsis 法流程图

### 2.3 问题三的分析

问题三要求通过建立数学模型多元回归方程分析九个指标中。结合问题一的数据，通过向后逐步回归分析可以减少哪些指标使得评价结果与问题二中的评价结果基本一致。

### 2.4 问题四的分析

通过建立数学模型因子分析法，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河南省 18 个地级市本科教育发展的差异，并验证所建立模型的有效性。

## 三、模型假设

1. 除选取的指标外，其他因素均不影响河南省本科高等教育质量；
2. 本文查阅相关论文所以量化指标能代表过去趋势
3. 所有数据真实可靠
4. 假设对于每个指标数据每年的变化趋势趋于一个平稳阶段，没有多大起伏。

## 四、符号说明

符号	说明
$\beta_i$	估计参数
$H_0$	原假设
$VIF$	方差膨胀因子
$W_j$	权重向量
$X_{best}$	最优样本值
$X_i$	指标正向化元素
$Z_j^+$	最优向量
$Z_j^-$	最劣向量
$D_j^+$	评价对象与最大值的距离
$V$	评语集
$A$	权重集
$R$	模糊判断矩阵
$B$	综合判断结果

## 五、模型的建立与求解

### 5.1 问题一

#### 5.1.1 问题分析

本文依据河南省及 18 个地级市的统计年鉴数据和教学质量报告，在查阅相关文献资料的基础上，选取本、专科院校数量、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、等 8 个评价指标，构建了高等教育评价指标体系。并对相应的指标进行量化处理。

#### 5.1.2 数据分析

通过对网络收集的数据进行了一些列筛选和处理，并且引用了相关论文的研究结果，得到如下表格。

表格 1 量化数据表格

序号	指标名称	量化处理方式	计算依据
1	本、专科院校数量	相加	河南省教育厅官网
2	招生人数	相加	河南省统计年鉴
3	师资队伍与结构	七级评价法	论文引用[6]
4	生师比	平均	河南省教育厅报告
5	教学条件与利用	七级评价法	论文引用[5]
6	专业建设与教学改革	七级评价法	论文引用[7]
7	学生就业	比例	中国大学生就业报告
8	科研投入与产出	平均	论文收集

通过对上面 8 个指标的量化处理得到的数据，得到了 18 个市的数据图表。

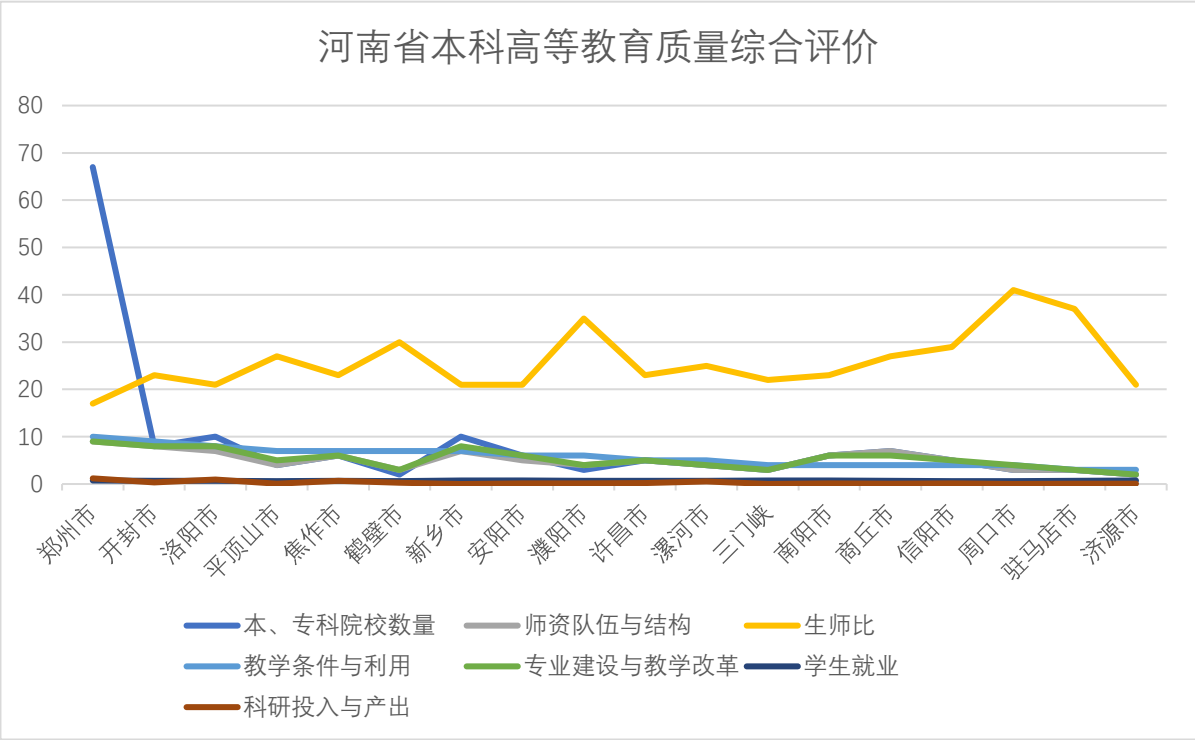


图 2 河南省 18 个地级市的高等教育质量发展情况

接下来通过 excel 计算出了每一种指标的描述性统计（具体数据见附录 2）

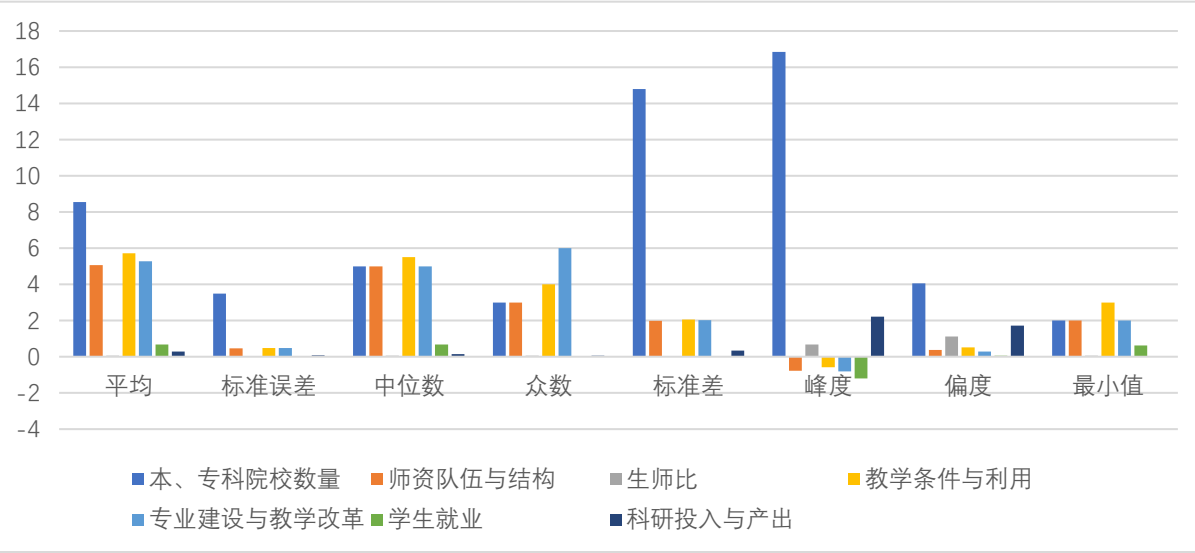


图 3 描述性统计

通过上面的表 1 和图 2、图 3，可以看到郑州市的教育资源最为丰富，其他 17 个地级市所拥有的本、专科院校数量，招生人数等等远远落后郑州市，可知河南省的教育资源分布非常不合理；在本、专科院校数量方面的平均数、标准误差、标准差、峰度波动都非常大，说明河南省的本、专科院校数量分布差别较大。

## 5.2 问题二模型建立与求解

### 5.2.1 模型的建立

#### 1. 建模思路:

由于是对河南省 18 个地级市高等教育质量进行综合评价, 并且我们可以从网上搜索到相关数据, 因此我们采用以 Topsis 法为主, 辅助模糊综合评价法进行检验的方法, 对河南省高等教育质量进行综合评价。

#### 2. 求解方法:

TOPSIS, 可译为逼近理想解排序法, 国内常简称为优劣解距离法。TOPSIS 法是一种常用的综合评价方法, 能充分利用原始数据的信息, 其结果能精确地反映各评价方案之间的差距。基本过程为先将原始数据矩阵统一指标类型 (一般正向化处理) 得到正向化的矩阵, 再对正向化的矩阵进行标准化处理以消除各指标量纲的影响, 并找到有限方案中的最优方案和最劣方案, 然后分别计算各评价对象与最优方案和最劣方案间的距离, 获得各评价对象与最优方案的相对接近程度, 以此作为评价优劣的依据。该方法对数据分布及样本含量没有严格限制, 数据计算简单易行。

同时使用熵权法, 这一种客观赋权方法, 对各项指标来确定权重。

### 5.2.2 Topsis 法

第一步: 统一指标类型: 将所有指标转化为极大型称为指标正向化。

(1) 极小型指标转换为极大型指标的公式:

$$\tilde{x}_i = \max - x_i \quad (1)$$

(2) 中间型指标转换变为极大型指标公式:

$$M = \max \{|x_i - x_{best}|\}, \tilde{x}_i = 1 - \frac{|x_i - x_{best}|}{M} \quad (2)$$

(3) 区间型指标转化为极大型指标:

$$M = \max \{a - \min \{x_i\}, \max \{x_i\} - b\}, \tilde{x}_i = \begin{cases} 1 - \frac{a - x_i}{M}, & x_i < a \\ 1, & a \leq x_i \leq b \\ 1 - \frac{x_i - b}{M}, & x_i > b \end{cases} \quad (3)$$

#### 第二步: 标准化处理

正向化矩阵标准化, 标准化的目的是消除不同指标量纲的影响。

有 18 个要评价的对象, 8 个评价指标 (已经正向化) 构成的正向化矩阵如下

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,8} \\ x_{2,1} & x_{2,2} & x_{2,3} & \dots & x_{2,8} \\ \dots & \dots & \dots & \dots & \dots \\ x_{18,1} & x_{18,2} & \dots & \dots & x_{18,8} \end{bmatrix}$$

第三步：熵权法确定权重矩阵

计算每个指标的信息熵，并计算信息效用值，并归一化得到每个指标的熵权。对第  $j$  个指标而言，其信息熵的计算公式为：

$$e_j = -\frac{1}{\ln n} p_{ij} \ln(p_{ij}) \quad (j=1, 2, 3 \dots m) \quad (4)$$

信息效用值：  $d_j = 1 - e_j$ ，将信息效用值进行归一化，就能得到每个指标的熵权

$$W_j = d_j / \sum_{j=1}^m d_j \quad (j=1, 2, \dots, m) \quad (5)$$

3 第三步：计算得分并归一化

那么对其标准化，矩阵记为  $Z$ ，有 18 个要评价的对象，8 个评价指标的标准化矩阵如下：

$$Z = \begin{bmatrix} Z_{1,1} & Z_{1,2} & Z_{1,3} & \dots & Z_{18,8} \\ Z_{2,1} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ Z_{18,1} & Z_{18,2} & Z_{18,3} & \dots & Z_{18,8} \end{bmatrix}$$

定义第  $i$  ( $i=1, 2, \dots, n$ ) 个评价对象与最大值的距离

$$D_i^+ = \sqrt{\sum_{j=1}^m w_j (Z_j^+ - Z_{ij})^2} \quad (6)$$

定义第  $i$  ( $i=1, 2, \dots, n$ ) 个评价对象与最小值的距离

$$D_i^- = \sqrt{\sum_{j=1}^m w_j (Z_j^- - z_{ij})^2} \quad (7)$$

那么，我们可以计算得出第  $i$  ( $i=1, 2 \dots, n$ ) 个评价对象为归一化的得分：

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (8)$$

#### 4. 模型的结论与分析

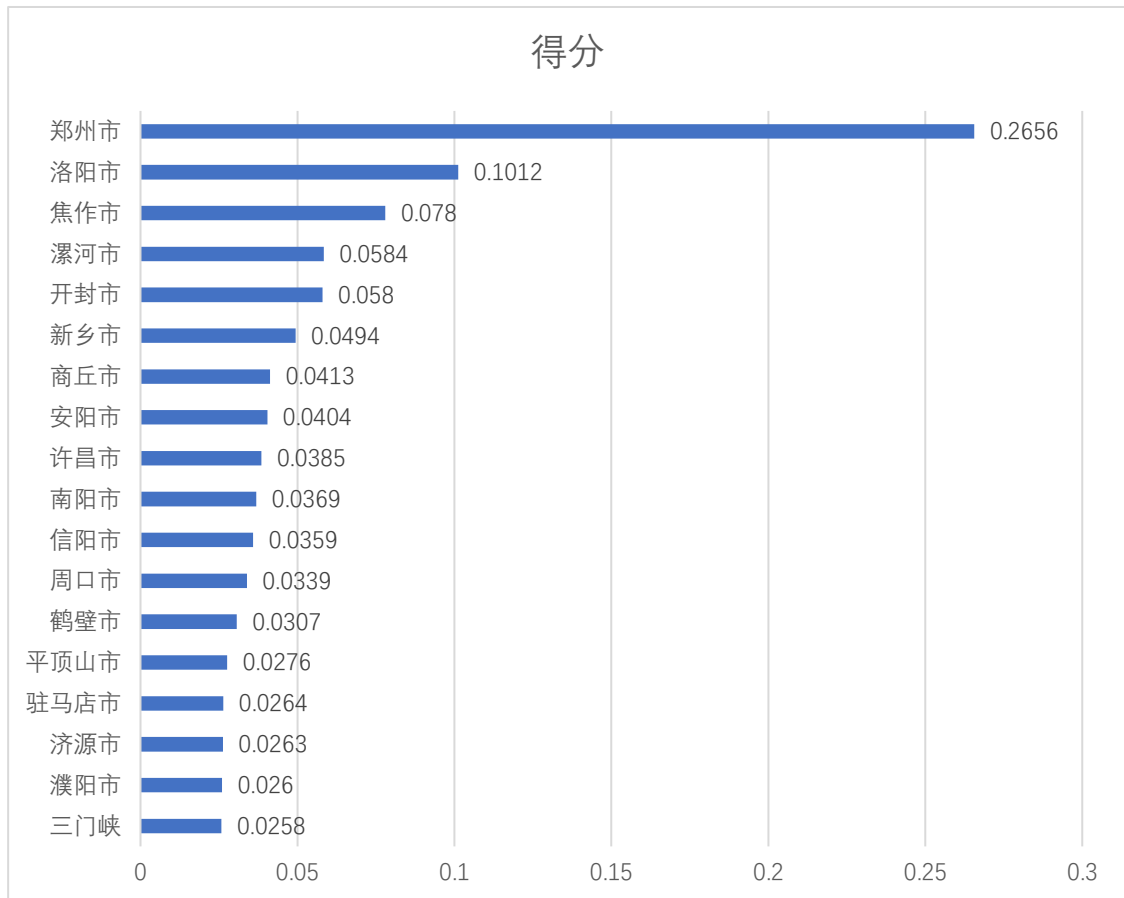


图 4 综合评价图

从上图可知郑州市的高等教育质量水平最高，洛阳市位居第二，而三门峡市的高等教育质量水平最低

#### 5. 2. 3 模糊综合评价法

##### 1. 指标量化求解

根据质表达指标和量表达指标的定义，在选取的 8 个指标中，量表达指标包括科研投入与产出、省级重点学科专业、生师比、本、专科院校数量、招生人数、学生就业。质表达指标包括师资队伍与结构、教学条件与利用、专业建设与教学改革。以河南省为例，通过对三个质表达指标采取七级评价法量化取值，得到了所有指标的量化数据。考虑到考虑各评价指标的原始数据对应的量纲不同，因此需数据进行无量纲化处理。此处采用直线法处理，即

$$y_i = \frac{x_i}{\max(x_i)} \times 100 \quad (9)$$

该方法的特点是评价价值随着指标实际值的增加而增加，且评价价值被控制在 0-100 之间。



## 2.确定评价指标和评语集

模糊综合评价模糊综合评价法是对受到多种因素影响的事物进行综合评价的一种常用方法。其特点是能够化定性为定量，对不确定的对象得出一个清晰的评价结果。

首先定义因素集  $U=\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9\}$ ,  $u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9$  分别代表影响本科教育质量的九个指标。定义评语集  $V=\{v_1, v_2, v_3, v_4, v_5\}$ , 其中  $v_1, v_2, v_3, v_4, v_5$  分别代表优秀、良好、中等、较差、差。根据本科教育质量模糊综合评价模型的因素集的数据特点, 此处选择半梯形组合的分布函数形式, 给出五组等级的模糊分界区间, 各等级状态的隶属函数如下表。

表 2 隶属度函数表格

评价等级	隶属函数表达式	区域
优秀	1	$X \geq 100$
	$(X-75)/25$	$75 \leq X < 100$
	0	$X < 75$
良好	$(X-50)/25$	$50 \leq X < 75$
	$(100-X)/25$	$75 \leq X < 100$
	0	$X \geq 100$ 或 $X < 50$
中等	$(X-25)/25$	$25 \leq X < 50$
	$(75-X)/25$	$50 \leq X < 75$
	0	$X \geq 75$ 或 $X < 25$
较差	$X/25$	$0 < X < 25$
	$(50-X)/25$	$25 \leq X < 50$
	0	$X \geq 50$ 或 $X < 0$
差	0	$X \geq 25$
	$(25-X)/25$	$0 \leq X < 25$
	1	$X \leq 0$

由上表可知, 当  $X$  处于不同区间时对应的隶属度函数不同, 得出的评价等级也不同。

## 3.权重矩阵

在确定权重的时候, 同样利用熵权法来计算权重。通过此方法得到了 13 个地级市的 9 个指标的权重集  $A=\{w_1, w_2, w_3, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$ 。通过 matlab 计算得出权重集为

$$A=\{0.3737, 0.1260, 0.0410, 0.0661, 0.0335, 0.0395, 0.0011, 0.3191\}$$

## 4.模糊评价矩阵

将数据分别带入隶属度函数表格, 即可求出模糊评价矩阵  $R$ 。以洛阳市为例, 将标准化 (去除量纲) 后处理的 8 个指标数据带入隶属函数, 结果如下表所示

表 3 洛阳市高等教育质量评价指标隶属度结果

因素集	评语集				
	优秀	良好	中等	较差	差
本、专科院校数量	0	0	0	0.5972	0.4082
招生人数	0	0	0.77	0.23	0
师资队伍与结构	0.1112	0.8888	0	0	0
生师比	0.332	0.6668	0	0	0
教学条件与利用	0.2	0.8	0	0	0
专业建设与教学改革	0.5556	0.4444	0	0	0
学生就业	0.7852	0.2148	0	0	0
科技投入与产出	0.1	0.9	0	0	0

因此模糊评价矩阵

$$R = \begin{pmatrix} 0.00 & 0.00 & 0.00 & 0.60 & 0.40 \\ 0.00 & 0.00 & 0.77 & 0.23 & 0.00 \\ 0.11 & 0.89 & 0.00 & 0.00 & 0.00 \\ 0.33 & 0.67 & 0.00 & 0.00 & 0.00 \\ 0.20 & 0.80 & 0.00 & 0.00 & 0.00 \\ 0.56 & 0.44 & 0.00 & 0.00 & 0.00 \\ 0.79 & 0.21 & 0.00 & 0.00 & 0.00 \\ 0.10 & 0.90 & 0.00 & 0.00 & 0.00 \end{pmatrix}$$

由此计算得出  $B=A*R=\{0.0879, 0.4123, 0.0970, 0.2522, 0.1525\}$ ,由最大隶属度原则,洛阳市本科教育质量的评价等级为“良好”。

按照相同的计算方法,可以得到河南省其余 17 个地级市的本科高等教育质量评价结果,如下表所示。

表 4 模糊综合评价表

郑州市	开封市	洛阳市	平顶山市	焦作市	鹤壁市	新乡市	安阳市	濮阳市
优秀	中等	良好	差	中等	差	中等	中等	差
许昌市	漯河市	三门峡	南阳市	商丘市	信阳市	周口市	驻马店市	济源市
较差	中等	差	较差	中等	较差	差	差	差

从表中可以查看出,除郑州市评价结果为“优秀”以外,其余 17 个地级市中高等教育质量最好的为洛阳市,评价结果为“良好”;而三门峡和濮阳市高等教育质量最差,评价结果为“差”。在 18 个地级市中,高等教育质量参差不齐,个别地级市之间差异

大，体现了高等教育质量水平不平衡的问题。

同时发现与采用 TOPSIS 进行综合评价的结果基本一致，相互验证了模型的有效性，展现了综合评价结果的合理性。

### 5.3 问题二模型的建立与求解

#### 5.3.1 模型的建立

##### 1. 问题分析

由于题目要求分析在给定的 8 个指标中，可以减少那些指标使得评价结果与综合评价结果基本一致。所以要得出 8 个指标的相关系数，通过相关系数分析其影响程度。

##### 2. 建模思路

先将所有变量均放入模型，之后尝试将其中的一个自变量从模型中剔除，看整个模型解释因变量的变异是否显著变化，最后将最没有解释力的那个自变量剔除；此过程不断迭代，直到没有自变量符合剔除条件。因此采用向后逐步回归分析。

#### 5.3.2 线性模型

根据原始数据我们可以对投入的元素进行简单的变量划分

表 5 变量分类

变量种类	变量类型	变量名称
自变量	定量变量	本、专科院校数量
		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
因变量		18 个地级市高等质量发展水平

根据题目中的条件我们可以以排名为因变量，其余 8 个条件为自变量建立如下的线性模型：

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \mu_i$$

其中  $\beta_i$  是待估计参数， $\mu_i$  是无法观测且满足一定条件的扰动项。

#### 5.3.3 回归分析

在进行描述性统计完成后，我们利用 Stata 软件对其进行回归性分析。

首先我们得到了一个方差分析：

表 6 方差分析

Source	Number of obs	18
	F(5, 12)	8.00
Model	Prob > F	0.0026
Residual	R-squared	0.8768
	Adj R-squared	0.7672
Total	Root MSE	2.5764

其中联合显著性检验  $H_0: \beta_1 = \beta_2 = \dots = \beta_8 = 0$ ，根据其 P 值为  $0.0026 < 0.01$  我们

可以拒绝原假设，即  $\beta_1, \beta_2, \dots, \beta_8$  不全等于 0，显著的异于 0。

$R^2 = 0.8768$ ， $R_{adj}^2 = 0.7672$  拟合优度比较大，验证模型的合理性。

回归分析如下：

表 7 回归分析表

排名	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
本专科院校数量	0.195092	0.254757	0.77	0.463	-0.38121	0.771391
招生人数	-7.9E-05	0.000165	-0.48	0.643	-0.00045	0.000293
师资队伍与结构	-0.23235	1.699347	-0.14	0.894	-4.07654	3.611838
生师比	0.344341	0.368772	0.93	0.375	-0.48988	1.178561
教学条件与利用	0.414138	1.205746	0.34	0.739	-2.31345	3.141724
专业建设与教学改革	-1.15568	2.77995	-0.42	0.687	-7.44436	5.133007
学生就业	25.73103	25.51905	1.01	0.34	-31.9971	83.45914
科研投入与产出	-9.79253	3.233893	-3.03	0.014	-17.1081	-2.47695
_cons	-7.77033	20.21326	-0.38	0.71	-53.4959	37.95524

由上表我们可以得知，联合显著性检验的 P 值  $< 0.1$ ，但每一个自变量的 P 值都  $> 0.1$ ，即在 90% 的置信水平下，接受原假设，这显然是不合理，与之前结论不符合。所以我们要探究其有没有异方差或完全多重共线性问题。

### 5.3.4 异方差分析

在之前的回归分析中，我们都默认了扰动项是球型干扰项，满足“同方差”和“无自相关”两个条件，但横截面数据容易出现异方差问题。

如果扰动项存在异方差

- (1) OLS 估计出来的回归系数是无偏、一致的
- (2) 假设检验无法使用（构造的统计量失效了）
- (3) OLS 估计量不再是最优线性无偏估计量（BLUE）

所以我们在接下来要检验其回归方程是否有异方差，利用怀特检验结合 stata 软件分析如下：

原假设  $H_0$ ：扰动项没有异方差

表 8 异方差检验表

Source	chi2	df	p
Heteroskedasticity	18.00	17	0.388
Skewness	6.02	8	0.6451
Kurtosis	0.01	1	0.9040
Total	24.03	26	0.5740

由于其 p 值>0.1，所以我们接受原假设，扰动项不存在异方差。由于没有异方差，所以我们接下来检验多重共线性。

### 5.3.5 完全多重共线性

根据以上我们进行的方差检验和回归性分析，可以看到，回归方程的 $R^2$ 较大，F 检验也很明显，但单个系数的 t 检验却不显著，或者说系数估计值不合理。所以现在对其进行完全多重共线性检验。

假设现在有 k 个自变量，那么第 m 个自变量的 $VIF_m = \frac{1}{1 - R_{1-k/m}^2}$ ， $R_{1-k/m}^2$ 是将第 m 个自变量作为因变量，对剩下的 k-1 个自变量进行回归的得到的拟合优度。 $VIF_m$ 越大，说明第 m 个变量和其他变量的相关性就是越大。

定义回归模型的 $VIF = \max\{VIF_1, VIF_2, VIF_3, \dots, VIF_K\}$ ，一个经验规则是，如果  $VIF > 10$ ，则认为该回归方程存在严重的多重共线性。

接下来利用 Stata 进行多重共线性检验：

表 9 方差膨胀因子表

Variable	VIF	1/VIF
专业建设革	81.1	0.012331
招生人数	42.63	0.023457
本专科院量	36.38	0.027487
师资队伍构	29.14	0.034314
教学条件用	15.69	0.063717
生师比	14.1	0.070922
学生就业	3.16	0.316895
科研投入出	3.13	0.31941
MAX VIF	42.63	

由上图可知 MAX VIF=42.63 远远大于 10，说明该回归方差具有严重的多重共线性。

### 5.3.6 逐步回归分析

经过前面的回归分析、怀特检验、完全多重共线性检验，发现该回归方程有严重的

多重共线性，所以采用向后逐步回归分析，分别计算各解释变量对被解释变量的贡献大小，遵循从大到小的顺序挑选相关程度最高的一个带入方程进行计算；判断以已代入方程的解释变量是否因为带入了新的解释变量而失去了统计意义。若是则剔除该解释变量，重新计算个解释变量与被解释变量间的相关程度。循环剔除，直到方程内不再有需要剔除的解释变量为止。

通过 stata 计算得出如下表格：

**表 10 方差表**

Source	Number of obs	18
	F(2, 15)	31.27
Model	Prob > F	0
Residual	R-squared	0.7807
Total	Root MSE	2.6613

**表 11 向后逐步回归结果表**

排名	Coef.	Std. Err.	t	P>t	Beta
科研投入与产出	-6.62352	2.609736	-2.54	0.023	0.42405
专业建设与教学改革	-1.4919	0.36506	-4.09	0.001	-0.5655
_cons	19.23955	1.931462	9.96	0	.

采用逐步回归法对影响指标进行初步筛选和剔除后。将科研投入和专业建设与教学改革这两个指标引入回归方程。由上表可以得知联合显著性检验  $F(2, 15) = 31.27$ ， $p$  值=0，说明拒绝原假设，方差为 0.7807，科研投入与产出的  $p$  值=0.023<0.1,专业建设与教学改革的  $p$  值=0.001<0.1,都拒绝原假设，显著的异于 0。接下来我们在进行多重共线性检测

**表 12 VIF 表**

Variable	VIF	1/VIF
专业建设革	1.52	0.656423
科研投入出	1.52	0.656423
Mean VIF	1.52	

由上图可知  $\text{MAX VIF}=1.52$  远远小于 10，说明该回归方差没有严重的多重共线性。最后我们通过 matlab 使用这两个指标进行综合评价得到如下图

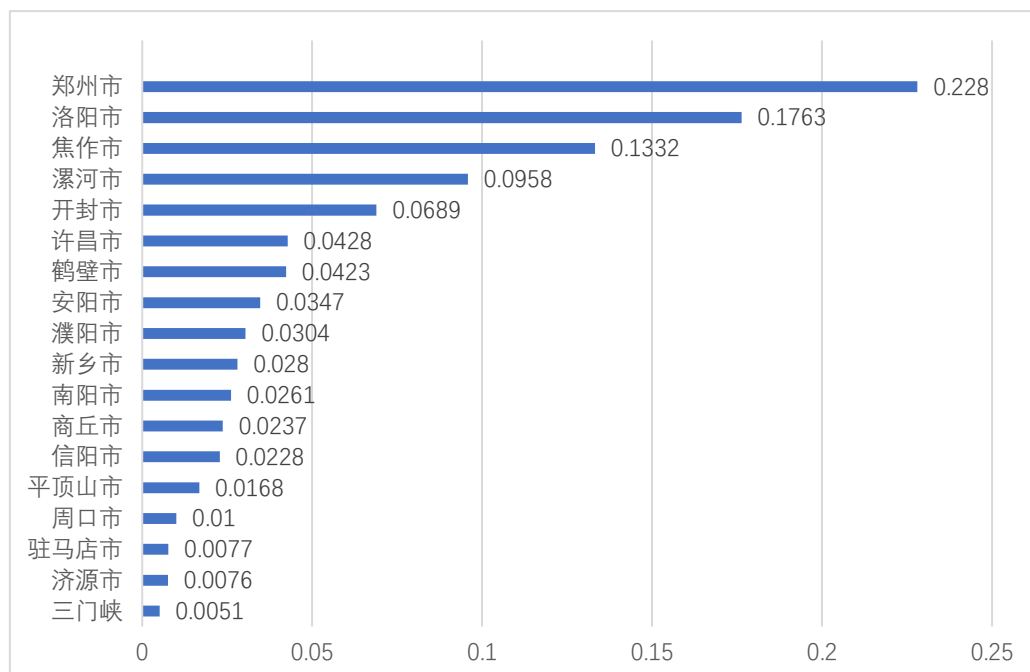


图 4 两指标的综合评价得分图

从上图可知郑州市的高等教育质量水平最高，而三门峡市的高等教育质量水平最低。与问题二中的评价结果基本一致。

## 5.4 问题三模型的建立与求解

### 5.4.1 模型的建立

先对数据做标准化处理，然后基于标准化数据进行以下操作

- 1、给出原始变量的相关系数矩阵。
- 2、KMO 和巴特利特检验。
- 3、用主成分法求公因子，公因子的提取按照默认提取（即特征值大于 1），给出公因子的方差贡献度表。
- 4、给出共同度表，并进行解释。
- 5、给出因子载荷矩阵，据之分析提取的公因子的实际意义。如果不好解释，请用因子旋转（采用正交旋转中最大方差法）给出旋转后的因子载荷矩阵，然后分析旋转之后的公因子，要求给各个公因子赋予实际含义；
- 6、先利用提取的每个公因子分别对各市进行排名并作简单分析。最后构造一个综合因子，计算各省市的综合因子的分值，并进行排序并作简单分析。

### 5.4.2 模型的求解

采用因子分析法，使用 spss 软件

#### 1、导入数据

选入标准化的 8 个变量，选择描述性统计中的 KOM 和巴特利特球形度检验。就可以得出 8 个变量的相关系数矩阵如下图。

表 13 相关性矩阵

相关性矩阵								
	本、专科 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
本、专科 院校数量	1.000	0.772	0.622	0.424	0.591	0.594	0.353	0.714
招生人数	0.772	1.000	0.583	-0.053	0.352	0.590	0.015	0.477
师资队伍 与结构	0.622	0.583	1.000	0.535	0.683	0.963	0.278	0.598
生师比	0.424	-0.053	0.535	1.000	0.435	0.523	0.690	0.465
教学条件 与利用	0.591	0.352	0.683	0.435	1.000	0.742	0.161	0.703
专业建设 与教学改革	0.594	0.590	0.963	0.523	0.742	1.000	0.317	0.586
学生就业	0.353	0.015	0.278	0.690	0.161	0.317	1.000	0.226
科研投入 与产出	0.714	0.477	0.598	0.465	0.703	0.586	0.226	1.000

由此表中数据可以看出大部分数据的绝对值都在 0.5 以上，说明变量间有较强的相关性。

## 2. KMO 和巴特利特检验

表 14 KOM 和巴特利特检验表

KMO 和巴特利特检验		
KMO 取样适切性量数。		0.508
巴特利特球形度检验	近似卡方	161.258
	自由度	36
	显著性	0.000

因子分析的 KMO 检验为  $0.508 > 0.5$ ，说明变量之间的相关性比较弱，但 Bartlett 球度检验结果 p 值等于 0.000 小于 0.05，认为在 95%置信水平下拒绝原假设，认为数据适合进行因子分析，所以可以进行因子分析。

3、依次点选在因子分析窗口点选抽取 a 方法：主成分分析,相关性矩阵；输出：未旋转的因子解，碎石图；抽取：基于特征值（特征值大于 1）；输出结果如下图。



表 15 总方差解释表

总方差解释						
成分	初始特征值			提取载荷平方和		
	总计	方差百分比	累积 %	总计	方差百分比	累积 %
1	4.637	57.957	57.957	4.637	57.957	57.957
2	1.477	18.467	76.423	1.477	18.467	76.423
3	0.770	9.623	86.047			
4	0.611	7.641	93.688			
5	0.261	3.262	96.950			
6	0.193	2.413	99.363			
7	0.046	0.571	99.934			
8	0.005	0.066	100.000			
提取方法：主成分分析法。						

上表第一列为特征值（主成分的方差），第二列为个各主成分的贡献率，第三列为累积贡献率，由上表看出前三个主成分的累计贡献率达到了 76.423%>75%，所以选取的主成分个数为 2 个。选  $y_1$  为第一主成分， $y_2$  为第二主成分。且这两个主成分的方差和占全部方差的 76.423%，即基本上保留了原来指标的信息。这样由原来的 10 个指标变为了 2 个指标。

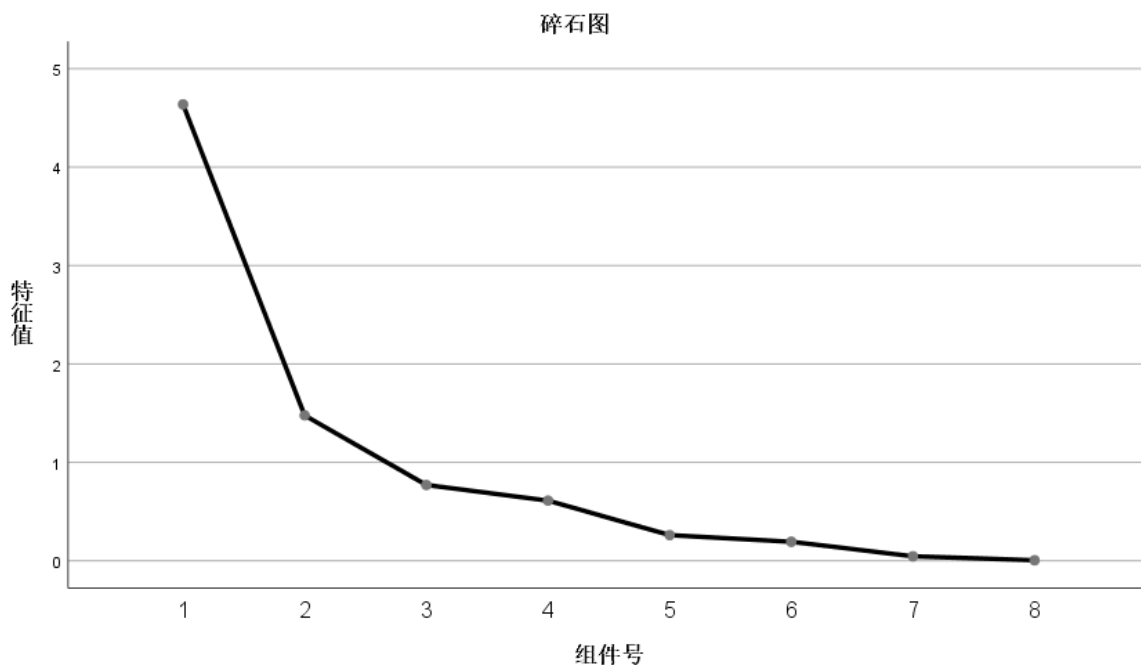


图 5 碎石图

由上图看出，成分数为 2 时，特征值的变化曲线趋于平缓，所以由碎石图也可大致确定出主成分个数为 2。与按累计贡献率确定的主成分个数是一致的。

4.共同度结果如下：

表 16 公因子方差

公因子方差		
	初始	提取
本、专科院校数量	1.000	0.742
招生人数	1.000	0.795
师资队伍与结构	1.000	0.809
生师比	1.000	0.903
教学条件与利用	1.000	0.650
专业建设与教学改革	1.000	0.818
学生就业	1.000	0.739
科研投入与产出	1.000	0.658
提取方法：主成分分析法。		

上表给出了该次分析从每个原始变量中提取的信息。由上表数据可以看出，主成分包含了各个原始变量的 80% 以上的信息。

5、在因子分析窗口，旋转 a 输出：载荷阵。输出结果如下：

表 17 成分矩阵

成分矩阵 <sup>a</sup>		
	成分	
	1	2
本、专科院校数量	0.843	-0.176
招生人数	0.646	-0.615
师资队伍与结构	0.897	-0.063
生师比	0.633	0.709
教学条件与利用	0.803	-0.070
专业建设与教学改革	0.903	-0.051
学生就业	0.438	0.740
科研投入与产出	0.807	-0.084
提取方法：主成分分析法。		
a. 提取了 2 个成分。		

由上表数据第一列表明：第一主成分与各个变量之间的相关性；第二列表明：第二主成分与各个变量之间的相关性；可以得出：x<sub>1</sub> x<sub>2</sub> x<sub>3</sub> x<sub>5</sub> x<sub>6</sub> x<sub>8</sub> 主要由第一主成分解释，x<sub>4</sub> x<sub>7</sub> 主要由第二主成分解释。

下面作因子旋转后的因子载荷阵。

在因子分析窗口，抽取 **a** 输出：旋转的因子解，继续；旋转 **a** 方法：最大方差法，输出结果如下图。

表 18 旋转后的成分矩阵

旋转后的成分矩阵 <sup>a</sup>		
	成分	
	1	2
本、专科院校数量	0.841	0.186
招生人数	0.841	-0.296
师资队伍与结构	0.844	0.311
生师比	0.286	0.906
教学条件与利用	0.761	0.266
专业建设与教学改革	0.844	0.324
学生就业	0.095	0.854
科研投入与产出	0.770	0.255
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 3 次迭代后已收敛。		

由上表数据可以得出， $x_1$   $x_2$   $x_3$   $x_5$   $x_6$   $x_8$  主要由第一主成分解释， $x_4$   $x_7$  主要由第二主成分解释。与第一因子关系密切的变量主要是招生人数、本科院校数量和师资队伍与教学条件与利用的变量，可以命名为数量因子；与第二因子关系密切的都是反映生师比方面的变量与学生就业的变量，可命名为学生因子。

表 19 总方差解释

总方差解释									
成分	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	4.637	57.957	57.957	4.637	57.957	57.957	4.104	51.299	51.299
2	1.477	18.467	76.423	1.477	18.467	76.423	2.010	25.125	76.423
3	0.770	9.623	86.047						
4	0.611	7.641	93.688						
5	0.261	3.262	96.950						
6	0.193	2.413	99.363						
7	0.046	0.571	99.934						
8	0.005	0.066	100.000						
提取方法：主成分分析法。									

由上表可以看出：第二列数据表明，各个主成分的贡献率与旋转前的有变化，但是 2 个主成分的累积贡献率相同都是 76.423%。

6、在因子分析窗口，得分 a 因子得分保存为变量 f\_1 f\_2；方法：回归。  
再按三个主成分降序排列：数据 a 排序个案：将 f\_1 选入排序依据，排列顺序：降序。同理得出按 f\_2 f\_3 排序的结果。  
最后，以各因子的方差贡献率占三个因子总方差贡献率的比重作为权重进行加权汇总，得出各城市的综合得分 f。即  
$$F = (\text{权重} \times f_1 + \text{权重} \times f_2) / 0.76423$$
  
f 得分在转换 a 计算变量中的出。最后再按 f 得分排序。排序结果如下：

表 20 得分顺序表

f_1	f_2
郑州市	安阳市
洛阳市	三门峡
开封市	济源市
新乡市	洛阳市
焦作市	新乡市
商丘市	南阳市
信阳市	许昌市
周口市	郑州市
平顶山市	开封市
南阳市	漯河市
安阳市	焦作市
濮阳市	平顶山市
漯河市	商丘市
许昌市	濮阳市
驻马店市	鹤壁市
鹤壁市	信阳市
三门峡	驻马店市
济源市	周口市

有了对各个公因子的合理的解释，结合各个城市在三个公因子的得分和综合得分，就可对各城市的经济发展水平进行评价了。在数量因子 f\_1 上得分最高的是郑州、洛阳、开封、新乡、焦作。其中郑州、洛阳数量能力最高，本科质量相对较高，从而推动城市本科教育质量发展；而鹤壁市、三门峡、济源市数量能力较差，可能由于地理位置的缘故本科教育质量发展相对落后。安阳市、三门峡市、济源市在学生因子上的得分较高，说明他们的生师比和学生就业方面工作比较好。

由两个因子可综合评价城市的本科教育质量发展水平，提高三主因子，本、专院校数量、师资队伍与结构、专业建设与教学改革与结构数量因子，学生因子普遍改善能够尽可能缩小河南省 18 个地级市本科教育发展的差异。

5.5 问题五

由题目所给的 8 个指标，上面综合评价了河南省 18 个市的本科教育质量水平，本

文通过了 Topsis 法与模糊综合评价法分别进行综合测评。并且通过因子分析法得到关键的因子,通过结果分析就如何有效提升河南省本科教育质量给出以下建议:

#### (一)加强师资队伍建设

高校应该加强学校专业教学团队建设,有计划地开展教师培训,对外增强教学交流,对内改革教师管理制度;可采取薪酬激励政策,实施绩效工资和基本工资相分离,以此提升教师的教学水平和能力。

#### (二)控制良好的生师比

生师比作为影响高等教育质量的关键因素之一,一方面难以严格把控,另一方面没有明确的数据可供参考。高校应当根据自身办学情况和当地普遍发展水平,控制良好的师生比,才能改善与提高教学质量。

#### (三)提高教学条件与利用

教学条件的优劣对教学质量也有较大的影响。为适应教学手段现代化和提高教学质量,高校可以适当增加办学投入,完善校园基础建设,注重校园信息化管理,提高校园教学条件的利用率。

#### (四)优化科研投入产出效率

高校可以通过严格把关科研活动的经费申请,发掘高质量科研项目,鼓励师生进行科研创新,使得在科研投入一定的条件下,产出得到提高。

#### (五)加强重点学科建设

学科建设是高效教学改革的重要举措之一。加强重点学科建设符合中国高等教育发展需要,对于增强我国高校学生 在国际人才市场的竞争力具有重要意义。因此高校需创新学科发展理念,建设重点学科以提升核心竞争力,争创“双一流”。

## 六、模型的评价、改进与推广

### 6.1 模型的优点

1.运用 Topsis 法来进行综合评价,能充分利用原始数据的信息,其结果能精确地反映各评价方案之间的差距,对数据分布、样本含量指标多少均无严格限制,既适用于小样本资料,也适用于多评价单元、多指标的大系统资料,既可用于横向(多单位之间)对比,也可用于纵向(不同年度)分析,应用灵活,数学计算比较简单,结果量化客观,因此该法在经济领域工作质量、经济效益等的综合评价中有一定的实用价值。

2.运用熵权法确定权重,避免了专家打分确定权重时出现了主观错误,导致结果产生很大影响。

3.采用模糊综合评价通过精确的数字手段处理模糊的评价对象,数学模型简洁,容易掌握;可以将不完全信息、不确定信息转化为模糊概念,使定性问题定量化,提高评估的准确性;

4.采用向后逐步回归法计算各解释变量对被解释变量的贡献大小,遵循从大到小的顺序挑选相关程度最高的一个首先代入方程进行运算;计算各解释变量和被解释变量间的相关程度,同时判断已代入方程的解释变量是否因为代入了新的解释变量而失去统计

意义。若是,则剔除该解释变量,重新计算各解释变量与被解释变量间的相关程度。

5. 因子分析简化系统结构,探讨系统内核在众多因素中找出各个变量最佳的子集合,从子集合所包含的信息描述多变量的系统结果及各个因子对系统的影响。“从树木看森林”,抓住主要矛盾,把握主要矛盾的主要方面,舍弃次要因素,以简化系统的结构,认识系统的内核。

## 6.2 模型的缺点

1. 因子分析的 KMO 检验为  $0.508 > 0.5$ , 说明变量之间的相关性比较弱。

## 6.3 模型的改进

在对河南省 18 个地级市的高等教育质量进行综合评价时,可以再使用 (APH) 层次分析法,完善结果的准确性。

# 七、参考文献

- [1] 姜启源 谢金星 数学模型第三版 2018-05-24
- [2] 林孝量 数学建模竞赛赛题简析与论文点评
- [3] 司守奎 孙玺菁 数学建模与应用 国防工业出版社 2011-08
- [4]胡颖[1], 骆润婷[2]. 本科高等教育质量综合评价——以江苏省 13 个地级市为例[J]. 今日财富, 2019, 000(004):168-169.
- [5]王鸣涛. 河南省地级城市科技实力评价实证研究[J]. 安阳师范学院学报, 2011(2):17-20.
- [6]曹思敏. 河南省农村幼儿园师资队伍现状、原因分析及发展对策[J]. 新课程研究(下旬刊), 2014(6 期):14-15.
- [7]余新创. 高等教育入学机会的地区差异研究——基于河南省 17 个地级市的实证研究[J]. 行政事业资产与财务, 2015, 000(007):36-39.
- [8]河南省教育统计年鉴:  
[https://xueshu.baidu.com/usercenter/paper/show?paperid=145s02s08r0p0rw0vs1t0460cb173775&site=xueshu\\_se](https://xueshu.baidu.com/usercenter/paper/show?paperid=145s02s08r0p0rw0vs1t0460cb173775&site=xueshu_se)
- [9]河南省教育厅官网: [河南省教育厅 \(henan.gov.cn\)](http://www.henan.gov.cn)
- [10]全国大学生就业信息网: [全国高校就业促进 --大学生就业信息网 \(bysjy.com.cn\)](http://bysjy.com.cn)

## 附录

附录 1

介绍：原始数据

地级市	学校数量	招生人数	师资队伍与结构	生师比	教学条件	专业建设	学生就业	科研投入
郑州市	67	113661.6	9	17	10	9	0.736	1.2
开封市	8	37047.2	8	23	9	8	0.6796	0.35
洛阳市	10	50299.2	7	21	8	7	0.7126	0.93
平顶山市	4	28042.4	4	27	7	4	0.6353	0.06
焦作市	6	31349.6	6	23	7	6	0.6527	0.71
鹤壁市	2	16138.4	3	30	7	3	0.6156	0.24
新乡市	10	50337.6	7	21	7	7	0.7165	0.03
安阳市	6	33364	5	21	6	5	0.753	0.17
濮阳市	3	31232.8	4	35	6	4	0.6823	0.17
许昌市	5	23235.2	5	23	5	5	0.6877	0.23
漯河市	4	24888.8	4	25	5	4	0.6733	0.52
三门峡	3	13925.6	3	22	4	3	0.7457	0.02
南阳市	6	34652.8	6	23	4	6	0.7139	0.11
商丘市	7	55158.4	7	27	4	7	0.6562	0.09
信阳市	5	55162.4	5	29	4	5	0.6196	0.11
周口市	3	70875.2	3	41	4	3	0.6336	0.02
驻马店市	3	53775.2	3	37	3	3	0.6472	0.05
济源市	2	7120	2	21	3	2	0.7182	0.06

附录 2

介绍：描述性统计完整数据

本、专 科院校 数量	招生人 数	师资队 伍与结 构	生师比
平均	8.555556	平均	5.055556
中位数	5	中位数	5
众数	3	众数	3
方差	218.732	方差	3.937908
最小值	2	最小值	2
最大值	67	最大值	9
求和	154	求和	91
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
平均	5.722222	平均	5.277778
		平均	0.682167
		平均	0.281667

中位数	5.5	中位数	5	中位数	0.68095	中位数	0.14
众数	4	众数	6	众数	0	众数	0.06
方差	4.212418	方差	4.094771	方差	0.001891	方差	0.116815
最小值	3	最小值	2	最小值	0.6156	最小值	0.02
最大值	10	最大值	9	最大值	0.753	最大值	1.2
求和	103	求和	95	求和	12.279	求和	5.07

### 附录 3

#### 介绍：matlab 编写的 Topsis 代码

```

clear;clc
load first.mat
[n,m] = size(X);
disp(['共有' num2str(n) '个评价对象,' num2str(m) '个评价指标'])
Judge = input(['这' num2str(m) '个指标是否需要经过正向化处理，需要请输入 1，不需要输入 0: ']);

if Judge == 1
    Position = input('请输入需要正向化处理的指标所在的列，例如第 2、3、6 三列需要处理，那么你需要输入[2,3,6]: '); % [2,3,4]
    disp('请输入需要处理的这些列的指标类型（1：极小型，2：中间型，3：区间型）')
    Type = input('例如：第 2 列是极小型，第 3 列是区间型，第 6 列是中间型，就输入 [1,3,2]: '); % [2,1,3]
    % 注意，Position 和 Type 是两个同维度的行向量
    for i = 1 : size(Position,2) % 这里需要对这些列分别处理，因此我们需要知道一共要处理的次数，即循环的次数
        X(:,Position(i)) = Positivization(X(:,Position(i)),Type(i),Position(i));
    end
    disp('正向化后的矩阵 X = ')
    disp(X)
end
disp('请输入是否需要增加权重向量，需要输入 1，不需要输入 0')
Judge = input('请输入是否需要增加权重: ');
if Judge == 1
    Judge = input('使用熵权法确定权重请输入 1，否则输入 0: ');
    if Judge == 1
        if sum(sum(Z<0))>0 % 如果之前标准化后的 Z 矩阵中存在负数，则重新对 X 进行标准化
            disp('原来标准化得到的 Z 矩阵中存在负数，所以需要对 X 重新标准化')
            for i = 1:n
                for j = 1:m
                    Z(i,j) = [X(i,j) - min(X(:,j))] / [max(X(:,j)) - min(X(:,j))];
                end
            end
        end
    end
end

```



```

        disp('X 重新进行标准化得到的标准化矩阵 Z 为: ')
        disp(Z)
    end
    weight = Entropy_Method(Z);
    disp('熵权法确定的权重为: ')
    disp(weight)
else
    disp(['如果你有 3 个指标，你就需要输入 3 个权重，例如它们分别为
0.25,0.25,0.5，则你需要输入[0.25,0.25,0.5]']);
    weight = input(['你需要输入' num2str(m) '个权数。' '请以行向量的形式输入这'
num2str(m) '个权重: ']);
    OK = 0; % 用来判断用户的输入格式是否正确
    while OK == 0
        if abs(sum(weight) -1)<0.000001 && size(weight,1) == 1 && size(weight,2) ==
m % 注意，Matlab 中浮点数的比较要小心
            OK =1;
        else
            weight = input('你输入的有误，请重新输入权重行向量: ');
        end
    end
end
end
else
    weight = ones(1,m) ./ m ; % 如果不需要加权重就默认权重都相同，即都为 1/m
end
D_P = sum([(Z - repmat(max(Z),n,1)) .^ 2] .* repmat(weight,n,1),2) .^ 0.5; % D+ 与最大值的
距离向量
D_N = sum([(Z - repmat(min(Z),n,1)) .^ 2] .* repmat(weight,n,1),2) .^ 0.5; % D- 与最小值的
距离向量
S = D_N ./ (D_P+D_N); % 未归一化的得分
disp('最后的得分为: ')
stand_S = S / sum(S)
[sorted_S,index] = sort(stand_S,'descend')

```

#### 附录 4

介绍：该代码是 matlab 语言编写的，作用计算熵权

```

[n,m] = size(Z);
D = zeros(1,m); % 初始化保存信息效用值的行向量
for i = 1:m
    x = Z(:,i); % 取出第i列的指标
    p = x / sum(x);
    e = -sum(p .* mylog(p)) / log(n); % 计算信息熵
    D(i) = 1 - e; % 计算信息效用值
end

```

```
end
W = D ./ sum(D); % 将信息效用值归一化，得到权重
end
```

## 附录 5

介绍：该代码是 stata 语言编写的，计算逐步回归分析

```
clear
import excel "G:\OneDrive - for personal\Desktop\regardless.xlsx", sheet("Sheet1") firstrow
// 1.回归分析
reg 排名 本专科院校数量 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
// 2.怀特检验，检测异方差
estat imtest,white
//3.检测完全多重共线性
estat vif
//4.向后逐步回归
stepwise reg 排名 本专科院校数量 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, r pr(0.05)
//增加了标准化的向后逐步回归
stepwise reg 排名 本专科院校数量 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, r pr(0.05)
```