

# 建模国家中心城市格局

## 一 摘要

本文主要分析了我国中心城市的评选指标，研究了具体应该设立多少个中心城市以及设立哪些城市为中心城市最有利于国家发展，并且预测了在 2035 年我国中心城市的格局。建立了层次分析模型、层次聚类模型以及回归模型。

问题一，对于中心城市的评选标准的评价指标本文主要考虑了包括地区生产总值、地方财政收入、人口总数等五个指标。建立层次分析模型，通过一致性检验后得出各个评价指标的相应权重以及目前各个城市在这五个指标上的加权得分值。

问题二，中心城市设立过多或者过少均不利于国家发展，考虑到中心城市的在经济、人口、就业等方面的相似程度，建立层次聚类模型将所有城市分成两类即中心城市与非中心城市，在中心城市这一类中城市的数量取为当前我国最适合设立中心城市的个数。根据聚类图的结果得出我国目前应该设立 11 个中心城市最有利于国家发展。这十一个城市分别是南京、成都、天津、广州、上海、北京、西安、重庆、武汉、沈阳、郑州，排名不分先后。

问题三，中心城市必然会随着经济等形势的变化而变化，对于问题三建立回归模型预测各个城市在 2035 年的评价指标值，问题一中已经利用层次分析法给出这些指标的相应权重值，在假设 2035 年这些指标权重值仍然不变的情况下通过加权求出分数值便可以判断出在 2035 年我国中心城市的格局。根据本文预测的结果，南京、沈阳、厦门三个城市按照本文预测的结果可以在 2035 年跻身在国家中心城市之内，而郑州在 2035 年可能在这五个准则标准中达不到中心城市的水准。

问题四，我们认为应该着重发展南京、沈阳、厦门这三个城市。因为无论是根据聚类的结果还是预测的结果，这三个城市在这五个准则指标下均可以达到国家中心的水准。

**关键词：**中心城市; 层次分析模型; 聚类; 回归分析;

## 二 问题分析

### 问题背景:

国家发展改革委在 2016 年《关于支持武汉建设国家中心城市的指导意见》中明确指出，国家中心城市是指居于国家战略要津、肩负国家使命、引领区域发展、参与国际竞争、代表国家形象的现代化大都市。在资源环境承载条件和经济发展基础较好的地区规划建设国家中心城市，既是引领全国新型城镇化建设的重要抓手，也是完善对外开放区域布局的重要举措。2010 年，住建部发布的《全国城镇体系规划（2010-2020 年）》中首次提出建设北京、天津、上海、广州、重庆五大国家中心城市。2016 年 5 月发布的《成渝城市群发展规划》，明确成都要“以建设国家中心城市为目标”。2016 年 12 月份，国家发改委发布《促进中部地区崛起“十三五”规划》，明确支持武汉、郑州建设国家中心城市。2018 年 2 月份发布的《关中平原城市群发展规划》明确提出“建设西安国家中心城市”。至此，我国已有 9 座城市进入国家中心城市建设之列。作为我国城镇体系层级的“塔尖”，目前我国已有北京、天津、上海、广州、重庆、成都、武汉、郑州、西安 9 座城市被明确定位为国家中心城市。

### 问题一分析:

针对问题一，可以考虑地区生产总值、人口总数、财政收入等几个常用评价指标建立层次分析模型，因为它特别适用于那些难于完全定量分析的问题。设定好评价准则以及判别矩阵，如果通过一致性检验便可以得到各个评选标准对应的权重。

**问题二分析:** 针对问题二，由于评定中心城市是由相同的指标因素来评定的，所以在这几个指标上中心城市会比非中心城市相似度更大，鉴于这一点可以建立层次聚类模型将中心城市聚为一类以此分析哪些城市应该算是中心城市。

### 问题三分析:

根据上述查阅的数据，将时间作为自变量，也成为回归变量，五个评价准则分别作为因变量建立多个回归模型分别预测出在 2035 年各个城市的这五个评价准则的预测值然后代入上述建立的层次分析模型，假设在每一个准则所对应的权值仍然不变，那么便可以计算出在 2035 年哪些城市符合国家中心城市的评选标准。

### 三 模型假设与符号说明

#### 模型假设

1. 假设在 2035 年各个城市内国家重点高校数量不变
2. 假设在 2035 年各个准则指标权值不变

#### 符号说明

$X_1$	人口总数
$X_2$	地区生产总值
$X_3$	国家重点高校数量
$X_4$	地方财政收入
$X_5$	人均可支配收入
$\hat{\sigma}^{*2}$	$\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
$D_{ij}$	$\min\{d_{ij}   x_i \in G_p, x_j \in G_q\}$ 类间最短距离

## 四 模型的建立与求解

### 问题一的求解：

针对问题一，本文考虑的评价指标包括地区生产总值、地方财政收入、人口总数等五个指标，建立层次分析模型。层次分析法是一种定性和定量相结合的、系统的、层次化的分析方法。建模步骤如下：

#### 1. 将因素分层

1. 目标层：中心城市的评选标准
2. 准则层：上述所提到的地区生产总值、人口总数等五个评价因素
3. 方案层：备选中心城市

#### 2. 构造判别矩阵

层次结构反映了因素之间的关系，但准则层中的各准则在目标衡量中所占的比重并不一定相同，在决策者的心目中，它们各占有一定的比例。构造判别矩阵的方法为：反复咨询专家然后根据各个准则因素，对元素两两比较哪个重要，重要多少，对重要程度按 1—9 赋值如下表：

重要性标度含义表

重要性标度	含义
1	表示两个元素相比较，具有同等重要性
3	表示两个元素相比较，前者比后者稍重要
5	表示两个元素相比较，前者比后者明显重要
7	表示两个元素相比较，前者比后者强烈重要
9	表示两个元素相比较，前者比后者极端重要
2,4,6,8	表示上述判断的中间值

显然判别矩阵具有反对称性。在使用判断矩阵求权重之前，必须对其进行一致性检验。

#### 3. 计算权向量与判别矩阵的一致性检验

本文计算权向量采用和法，具体步骤如下：

1. 将判别矩阵按列归一化：
$$b_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}$$

2. 按行求和:  $v_i = \sum_{j=1}^n b_{ij}$
3. 归一化:  $w_i = \frac{v_i}{\sum_{i=1}^n v_i}$ , 即为近似权向量
4. 为了进行一致性检验, 求最大化特征根  $\lambda_{max} = \frac{1}{n} \sum_{i=1}^n \frac{(Aw)_i}{w_i}$

一致性检验:

1. 计算一致性指标  $CI(consistencyindex)$ ,  $CI = \frac{\lambda_{max} - n}{n - 1}$
2. 查表确定相应的平均随机一致性指标  $RI(randomindex)$
3. 计算一致性比例  $CR(consistencyratio)$ ,  $CR = \frac{CI}{RI}$

当  $CR = 0$  时, 认为判别矩阵具有完全一致性, 当  $CR < 0.1$  时, 认为判别矩阵的一致性可以接受, 当  $CR > 0.1$  时, 认为判别矩阵不符合一致性要求需要对判别矩阵重新构造, 加以修正。

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
$X_1$	1	$\frac{1}{6}$	3	2	4
$X_2$	6	1	9	9	9
$X_3$	$\frac{1}{3}$	$\frac{1}{9}$	1	$\frac{2}{3}$	$\frac{4}{3}$
$X_4$	$\frac{1}{2}$	$\frac{1}{9}$	$\frac{3}{2}$	1	2
$X_5$	$\frac{1}{4}$	$\frac{1}{9}$	$\frac{3}{4}$	$\frac{1}{2}$	1

表 1 目标层到准则层的判别矩阵

上表就是目标层到准则层的判别矩阵, 利用上述所讲的模型计算步骤, 利用 Python 编程得到结果

$$CR = 0.0265, \quad w = [0.184 \quad 0.614 \quad 0.062 \quad 0.092 \quad 0.047]$$

$CR = 0.0265 < 0.1$ , 模型通过一致性检验且由计算得到的特征向量可见各个准则因素所占权重的比例由大到小依次为  $X_2, X_1, X_4, X_3, X_5$ 。

进一步, 通过查阅资料得到近几年来各个备选中心城市的以上五个评价准则因素的数据, 为了计算简便起见, 本文将这几年的数据取平均值然后构造了准则层到决策层的一共 5 个  $9 \times 9$  的判别矩阵。利用上述模型分别对这 5 个判别矩阵进行层次分析, 最终得到每一个备选中心城市对应的权重以及相应的得分见下表 2。

	指标权重	北京	天津	上海	广州	重庆	成都	武汉	郑州	西安
$X_1$	0.18379	0.14181	0.1027	0.15958	0.09545	0.20422	0.09717	0.07169	0.06669	0.0607
$X_2$	0.61445	0.17374	0.1078	0.1875	0.13131	0.11697	0.08773	0.08486	0.0625	0.04759
$X_3$	0.06225	0.35789	0.05263	0.13684	0.16842	0.02105	0.07368	0.09474	0.01053	0.08421
$X_4$	0.09237	0.20994	0.07244	0.25792	0.09854	0.08222	0.05168	0.10523	0.06905	0.05299
$X_5$	0.04718	0.15979	0.10123	0.16446	0.14118	0.06761	0.09261	0.10796	0.08469	0.08048

表 2 目前九个中心城市在这五个评价指标上的权重

	加权得分值
北京	0.182
上海	0.184
天津	0.099
广州	0.124
重庆	0.122
成都	0.085
武汉	0.086
郑州	0.062
西安	0.054

表 3 九个中心城市在这五个评价指标下的综合加权得分

最终得到每一个城市在这五个指标权重上的得分如表 3

### 问题二的求解:

考虑到各个中心城市在各个评选指标上的相似程度，建立层次聚类模型。聚类的迭代过程可以总结如下：

1. 计算  $n$  个样本两两之间的欧氏距离，记作距离矩阵  $D = [d_{ij}]$
2. 构造  $n$  个类，每一类仅仅包含一个样本
3. 合并类间距离最小的两个类，其中最短距离为类间距离，构造一个新类
4. 计算新类与当前各个类之间的距离，若类的个数为 1 则终止计算，否则回到步骤 3

为了简便，本文共考虑了 15 个城市，分别是北京、上海、重庆、天津、广州、成都、武汉、郑州、西安、南京、厦门、青岛、哈尔滨、合肥、沈阳。可以看作 15 个样本，以人口总数、地区生产总值等上述五个评价指标作为这些样本的特征。具体数值详见附录。通过这些数据得到样本与样本之间的距离矩阵如下：

利用 Python 编程并且画出聚类图如图所示

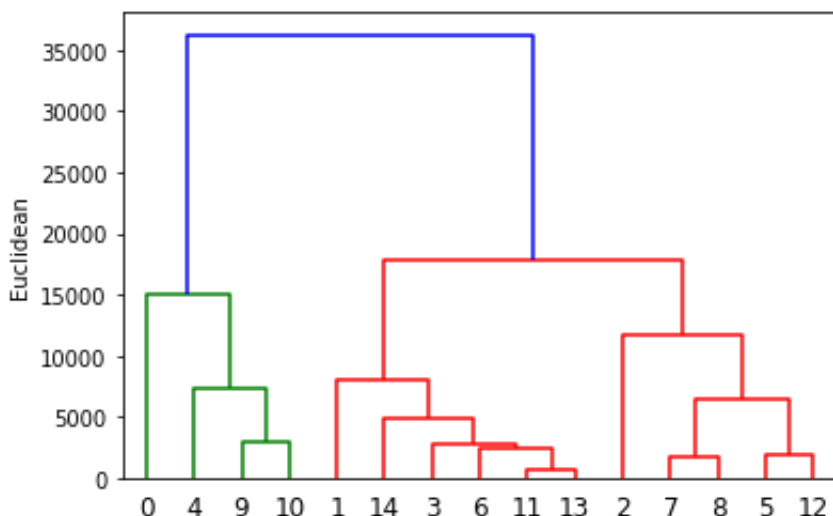


图 1 聚类图

其中看出数据被大致分成了两类，其中红色线所代表的类别为中心城市类别，下方横坐标数字分别对应南京、成都、天津、广州、上海、北京、西安、重庆、武汉、沈阳、郑州。绿色线所代表的类别为非中心城市类别包括青岛、合肥、哈尔滨、厦门。因此根据聚类图可以看出目前国内的城市在这五个指标上可以达到近似的一共有十一个城市。

### 问题三的求解:

以时间作为自变量，也称为回归变量，四个评价准则分别作为因变量建立多个回归模型分别预测出在 2035 年各个城市的这四个评价准则的预测值。其中重点高校数量与时间变量没有线性相关性，本文假设在 2035 年各个城市的重点高校数量不变。

一般的称下式所确定的模型为一元线性回归模型

$$\begin{cases} y = \beta_0 + \beta_1 x + \epsilon \\ E\epsilon = 0, D\epsilon = \sigma^2 \end{cases}$$

由上式对  $y = \beta_0 + \beta_1 x + \epsilon$  取期望即得到  $Y = \beta_0 + \beta_1 x$ ，称为  $y$  对  $x$  的回归方程。 $\beta_0$  和  $\beta_1$  是回归系数。常用最小二乘估计来估计回归系数。即对于  $\beta_0, \beta_1$  有：

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

当求出回归方程后需要对回归系数做显著性检验。

$$H_0 : \hat{\beta}_1 = 0, \quad H_1 : \hat{\beta}_1 \neq 0$$

构造统计量

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}^*} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}$$

且  $T \sim t(n-2)$ 。本文预测 15 个城市的未来的评价指标因此  $n = 15$ 。对于给定显著性水平  $\alpha$ ，查找分位数  $t_{\alpha/2}(n-2)$ 。本文取  $\alpha = 0.5$ ，因此即查找  $t_{0.025}(13) = 2.1604$  在给定的一组回归值代入统计量中若  $|t| \geq 2.1604$  则拒绝  $H_0$ ，否则接受  $H_0$ 。预测的结果见表 4。根据所预测的数值与问题一中已经求解的指标的权值对应相乘得到结果如表 5。

按照分值由高到低的顺序为北京、上海、重庆、南京、成都、广州、沈阳、厦门、西安、天津、武汉、郑州、哈尔滨、青岛、合肥。由此，如果仍以 11 个中心城市为标准的话，那么前十一个城市分别是北京、上海、重庆、南京、成都、广州、沈阳、厦门、西安、天津、武汉。

可以看出南京、沈阳、厦门三个城市按照本文预测的结果可以在 2035 年跻身在国家中心城市之内，而郑州在 2035 年可能在这五个准则标准中达不到中心城市的水准。

#### 问题四的求解：

近年来把“万亿元 GDP”作为衡量国家中心城市主要标准的说法在社会上广为流传，不仅直接影响到个别入选城市的发展思路和规划编制，也在一些有竞争力的大城市之间掀起了新一轮的“GDP”竞赛。但这个流行说法，在住房城乡建设部和国家发展改革委两方面均找不到任何依据。受其影响，一些大城市把心思和精力主要用于提升自身的首位度上，背弃了中心城市加强了对周边中



	$X_1$ (万人)	$X_2$ (万亿元)	$X_3$ (个)	$X_4$ (亿元)	$X_5$ (元)
北京	3224.423	5.339	34	11872.339	108815.781
天津	2280.827	3.76	5	3599.236	71202.029
成都	1525.19	2.361	7	2497.887	74478.789
重庆	2957.995	3.678	2	4880.363	51967.955
西安	917.135	1.111	8	2568.756	55064.852
沈阳	928.997	1.23	3	1339.304	80367.683
郑州	1881.6	1.775	1	3536.461	55040.712
武汉	1141.491	1.823	9	3061.692	79424.396
南京	1075.479	2.435	10	2289.361	70157.386
厦门	611.506	1.148	1	2372.302	83356.544
上海	3549.8	5.657	13	14654.057	60844.882
合肥	598.547	1.025	1	1247.245	63214.547
广州	3124.887	4.258	6	10245.52	72548.667
青岛	658.546	0.878	2	1178.658	59877.634
哈尔滨	879.258	0.997	5	1254.984	72546.335

表 4 2035 年部分城市的中心城市评价指标预测

北京	6828.590
上海	4113.587
重庆	4026.840
天津	3448.681
广州	4004.976
成都	4087.141
武汉	3270.455
郑州	3241.5265
西安	3721.27
南京	4065.047
厦门	3880.95
青岛	3208.369
哈尔滨	3246.537
合肥	3055.597
沈阳	3901.181

**表 5 2035 年 15 个城市在指标权重不变的情况下所得的分值**

小城市的截留、虹吸和压制。应尽快明确国家中心城市的核心职能，承担国家战略使命和战略意图，促进国家中心城市的健康培育和成长。因此本文考虑了五个评选准则因素并且进行加权求和得出最后的分值，我们认为应该着重发展南京、沈阳、厦门这三个城市。因为无论是根据聚类的结果还是预测的结果，这三个城市在这五个准则指标下均可以达到国家中心的水准。

## 参考文献

- [1] 中国统计年鉴 2018. <http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm>.
- [2] 司守奎, 孙兆亮. 数学建模算法与应用. 国防工业出版社. 2015(4): 150-156.
- [3] 塞巴斯蒂安·拉施卡 (Sebastian Raschka). Python 机器学习. 机器工业出版社. 2017(3): 56-65.

## 附录

### 判别矩阵的的权值计算以及一致性检验

```
1      import numpy as np
2      A=np.array([[1,1/6,3,2,4],[6,1,9,9,9],[1/3,1/9,1,2/3,4/3],[1/2,1/9,3/2,1,2],[1/4,1/9,3/4,1/
3      B=A/np.sum([np.transpose(A)[i] for i in range(5)])#因为是对列归一化，所以取先取转置
4      v=np.zeros(5)
5      for i in range(5):
6          v[i]=np.sum([B[i][j] for j in range(5)])
7      w=v/np.sum(v)
8      lambda_max=np.sum(A.dot(w)/w)/5
9      print(lambda_max)#5.11836095 这就是最大特征根
10     print(w)#[0.18373494 0.61445783 0.062249 0.09236948 0.04718876] 特征向量
11     CI=(lambda_max-5)/(5-1)
12     RI=1.12#查表得到的，如果n==9 则RI=1.46
13     CR=CI/RI
14     if CR<0.1:
15         print("通过一致性检验")
16     else:
17         print("未通过一致性检验，需要重新修改判别矩阵")
```

### 准则层到目标层的权值向量的计算

```
1      human_amount=[2154,1560,2424,1449.84,3102,1476,1089,1013,922]#九个中心城市的人口总数
2      gdp=[3.03,1.88,3.27,2.29,2.04,1.53,1.48,1.09,0.83]#九个中心城市的GDP
3      university=[34,5,13,16,2,7,9,1,8]#九个中心城市的重点高校数量
4      finance_revenue=[5785.9,1996.3,7108.1,2715.6,2266,1424.2,2900,1903,1460.39]#九个中心城市的财
5      per_personl=[62361,39506,64183,55099,26386,36142,42133,33051,31407]#九个中心城市的人均可支配
6      import numpy as np
7      def get_judge_matrix(judge_baseline):
8          length=len(judge_baseline)
9          judge_matrix=np.identity(length)
10         for i in range(length):
11             value=judge_baseline[i]
12             for j in range(length):
13                 judge_matrix[i][j]=value/judge_baseline[j]
14                 judge_matrix[j][i]=judge_baseline[j]/value
```

```

15         return judge_matrix#返回判别矩阵
16     human_amount_matrix=get_judge_matrix(human_amount)
17     gdp_matrix=get_judge_matrix(gdp)
18     university_matrix=get_judge_matrix(university)
19     finance_revenue_matrix=get_judge_matrix(finance_revenue)
20     per_personl_matrix=get_judge_matrix(per_personl)
21     #现在已经得到了各个评价准则指标的判别矩阵，下面计算权值向量，
22     def get_weights_vector(judge_matrix):
23         length=judge_matrix.shape[0]
24         B=judge_matrix/np.sum([np.transpose(judge_matrix)[i] for i in range(length)])
25         V=np.sum([judge_matrix[i] for i in range(length)],axis=1)#按行求和
26         W=V/np.sum(V)#归一化
27         lambda_max=np.sum(judge_matrix.dot(W)/W)/5
28         return W

```

### 利用层次聚类分析中心城市个数

```

1     import numpy as np
2     import pandas as pd
3     from scipy.cluster.hierarchy import dendrogram,linkage
4     from scipy.spatial.distance import pdist
5     indexs=[str(i) for i in np.arange(len(15))]
6     columns=['X_1','X_2','X_3','X_4','X_5']
7     X=np.ndarray((20,5))
8     df=pd.DataFrame(X,columns=columns,index=indexs)
9     clusters=linkage(df.values,method='complete',metric='euclidean')
10    #print(pd.DataFrame(clusters,columns=['city_ ','']))
11    dend_gram=dendrogram(clusters,labels=indexs)
12    import matplotlib.pyplot as plt
13    plt.tight_layout()
14    plt.show()

```

### 回归分析预测 2035 年各个城市的评价指标值

```

1     import pandas as pd
2     from sklearn.linear_model import LinearRegression
3     slr=LinearRegression()
4     def predict_(criterion_data,slr):

```

```

5     year_list_=np.reshape(criterion_data[0],(-1,1))
6     predict_result=[]
7     for data in criterion_data[1:]:
8         slr.fit(year_list_,data)
9         predict_result.append(slr.coef_[0]*2035+slr.intercept_)
10    return predict_result
11    predict_result=predict_(criterion_data,slr)
12    data=pd.read_excel("./model_data/data.xlsx")#将要读取的数据写成 data.xlsx 放到 data 文件夹下

```