



**大数据管理与应用专业**

**数学模型与决策报告**

**长江中游城市群生态效率分析**

**姓名：李宗霖**

**时间：2022 年 6 月 4 日**

**指导老师：韦庆明**

# 长江中游城市群生态效率分析

## 前言

在刚开始确定选题的时候，原本选定的是对游戏抽卡数据进行的分析。大概思路如下：

在热门《原神》游戏中，可以通过抽卡池的方式获取角色，每次可以单抽或者十连抽，当抽到第 180 抽时，必中五星角色。通过抽卡数据可以推算出每次抽卡可获得五星角色的概率，并将每十抽数据放做一组，计算每组可以得到五星角色的概率，列出一步概率状态矩阵，然后用初始状态矩阵乘于一步状态矩阵，依次可得第  $n$  组抽卡时得到的概率分布矩阵。最后使用离散型数学期望计算期望，使用马尔科夫概率转移矩阵进行计算，最后可以得出抽到角色的平均次数。在此基础上可以再进行测算真实情况下第几次抽卡有最高的抽中概率，起到实际研究价值。

其中重点注意第 180 次有“保底”机制，即概率矩阵需要调整，难点在于如何使用 Excel 计算矩阵的幂。最早准备使用 MATLAB 进行编程求解，但由于后来课程要求使用 Excel，所以决定使用 VBA 编程。可惜最后因为编程困难，数据难以收集，缺少实际意义，放弃了这个选题。

随后很长一段时间没有着手做课程作业，时间只剩下半个月的时候，才重新开始选题，原本因为时间紧张，想使用其他课程作业，重新用 Excel 计算后提交。但是在运算的过程中，又有了新的思路，还发现了更多 Excel 的用法，因此将课程报告重新设计，并加以完善，有了以下的一篇报告。

## 摘要

伴随着长江中游城市群工业化和城镇化的快速推进，环境保护和资源短缺问题日益凸显。本文选取 2008-2019 年长江中游城市群 28 个地级市的数据，利用熵值法测算生态效率，使用变异系数法与回归分析法对城市效率进行收敛性检验，运用核密度估计分析其动态演进的趋势，最后使用聚类分析将城市的生态效率划分为四个等级，方便之后的生态效率分组与归类。研究发现虽然长江中游城市群生态效率存在较大地区差距，但随时间推进，地区间的生态效率差距减小，有趋于同意的趋势。而使用核密度估计得出的结果则显示长江中游城市群的生态效率有下降的趋势，且不同地区下降趋势有差异。最后的聚类分析，将所有城市生态效率分为四个水平，就可以看出来哪几个城市是属于一个类别，然后可知这个地区效率的分布特征，并且提出针对性的建议。通过以上研究，能够为推动长江中游城市群生态环境治理和绿色低碳转型、促进生态一体化建设提供政策参考。

**关键词：**长江中游城市群 生态效率 熵值法 回归分析 核密度估计 聚类分析

## 一、 选题背景

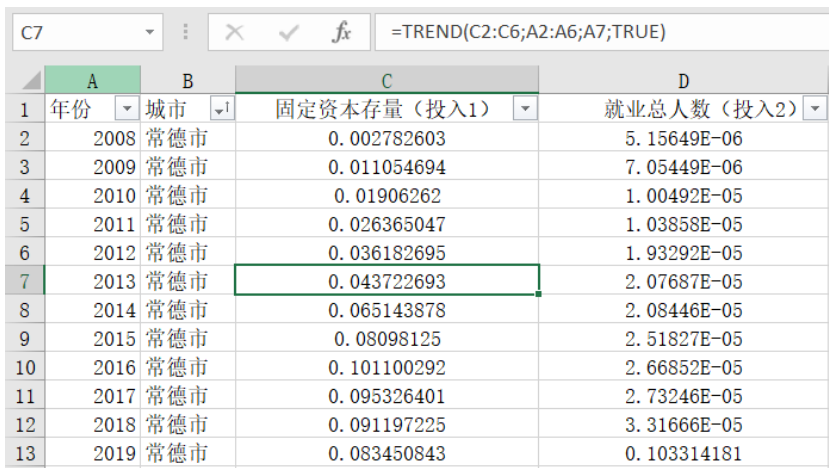
长江中游城市群地跨湖北、湖南、江西三省，承东启西、连南接北，是推动长江经济带发展、促进中部地区崛起、巩固“两横三纵”城镇化战略格局的重点区域，逐渐发展成为我国经济的“第五极”，在我国经济社会发展格局中具有重要地位。

然而，长江中游城市群工业化与城镇化的快速推进，同时也导致了許多生态环境问题。2022 年《长江中游城市群发展“十四五”实施方案》发展规划提出，要深化生态环境系统治理、协同治理，推动绿色低碳转型，生态一体化建设将逐渐成为长江中游城市群协同发展的重要方面。长江中游城市群作为全国首个“两型”社会建设引领区，应充分利用自身优势，加快从“两型”引领到“双碳”示范，推动城市绿色高质量发展和协同发展。在推进长江中游城市群发展过程中，综合协调经济发展、资源节约和环境保护三方面的关系，改善整体生态效率，同时探析长江中游城市群内部生态效率的地区差异和空间收敛，对于缩小城市间生态效率的非均衡性，促进长江中游城市群一体化发展具有重要的理论价值和现实意义。

## 二、 数据来源

由于县级市仙桃、天门、潜江相关数据的缺失，本文选取 2008-2019 年的长江中游城市群共计 28 个地级市的数据测算生态效率并进行实证评价。原始数据主要来源于《中国城市统计年鉴》、《中国城市建设统计年鉴》《中国环境统计年鉴》以及湖北、湖南、江西各省统计年鉴和各城市的国民经济和社会发展统计公报。部分缺失数据通过插值法补齐。

如图所示，使用 Excel 的内置 trend 函数可以快速进行数据补全。



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
1	年份	城市	固定资本存量（投入1）	就业总人数（投入2）
2	2008	常德市	0.002782603	5.15649E-06
3	2009	常德市	0.011054694	7.05449E-06
4	2010	常德市	0.01906262	1.00492E-05
5	2011	常德市	0.026365047	1.03858E-05
6	2012	常德市	0.036182695	1.93292E-05
7	2013	常德市	0.043722693	2.07687E-05
8	2014	常德市	0.065143878	2.08446E-05
9	2015	常德市	0.08098125	2.51827E-05
10	2016	常德市	0.101100292	2.66852E-05
11	2017	常德市	0.095326401	2.73246E-05
12	2018	常德市	0.091197225	3.31666E-05
13	2019	常德市	0.083450843	0.103314181

The formula bar at the top shows the formula: `=TREND(C2:C6;A2:A6;A7;TRUE)`. The cell C7 is highlighted, showing the result of the formula: 0.043722693.

图一

三、 具体研究

本部分将分为生态效率测算、变异系数及其回归分析、核密度测算与聚类分析四个部分进行具体说明。

3.1 生态效率测算

3.1.1 生态效率指标体系构建

本文根据现有研究对区域生态效率内涵的阐述，构建长江中游城市群生态效率水平指标评价体系。并且将二氧化碳排放量作为非期望产出纳入长江中游城市群生态效率水平指标评价体系，以期促进长江中游城市群能源绿色低碳发展，加快形成节约资源和保护环境产业结构、生产方式、生活方式、空间格局，实现经济绿色增长与经济社会协调发展。具体指标如表一：

类型	一级指标	二级指标
投入	资本投入	固定资本存量（亿元）
	劳动力投入	地区就业人数（万人）
	能源投入	能源消耗量（万 t 标准煤）
	水资源投入	供水量（亿 m <sup>3</sup> ）
	土地投入	建成区面积（km <sup>2</sup> ）
期望产出	经济增长	地区 GDP（亿元）
	城市绿化	建成区绿化面积（km <sup>2</sup> ）
非期望产出	废气排放	二氧化硫排放量（t）
		二氧化碳排放量（t）
	废水排放	工业废水排放量（万 t）

表一

3.1.2 运用熵值法进行生态效率测算

在最初的计划中，准备按照最热门的非期望产出的超效率 SBM 模型进行生态效率的测算，但在实际操作中，涉及到大量编程内容，而找到的（图二）DEA-Solver 也仅支持测算非期望产出或超效率模型，不支持同时满足两种条件，所以只好使用熵值法进行效率测度。



图二

(1) 数据标准化

使用如下公式进行数据标准化

$$x'_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \text{ (正向指标)} \quad \text{-----} \quad (1)$$

$$x'_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \text{ (负向指标)} \quad \text{-----} \quad (2)$$

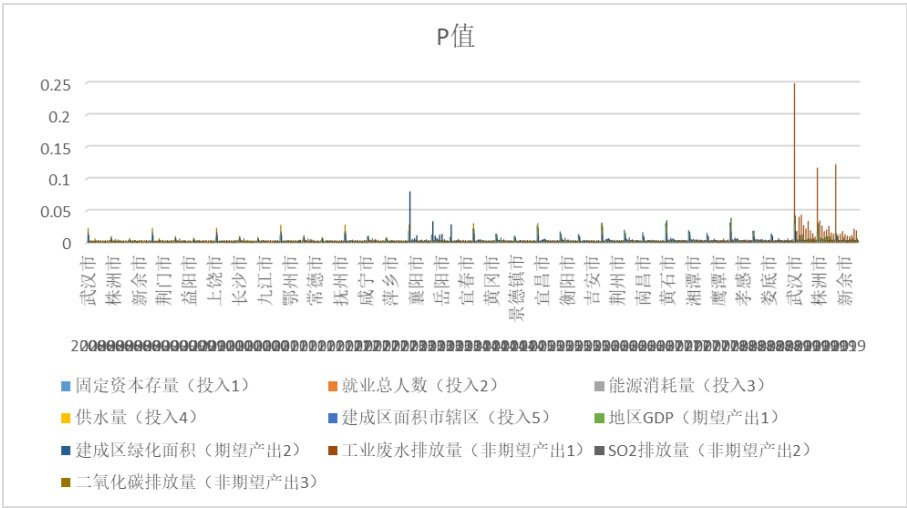
其中，将投入与期望产出记作正向指标，非期望产出记作负向指标（具体数据见标准化（处理后））。

此外，为便于之后计算，特将 0 换为无穷小 0.0000000001

(2) 比重 P 计算

使用如下公式对比重 P 进行计算，计算结果如下：

$$P_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} \quad \text{-----} \quad (3)$$



图三

使用以下公式进行计算

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), \text{ 其中 } k = \frac{1}{\ln(n)} \text{ 且 } k > 0, \text{ 满足 } e_j \geq 0$$

计算信息熵冗余度： $a_j = 1 - e_j$

计算各项指标的权值:  $w_j = \frac{d_j}{\sum_{j=1}^m d_j}$

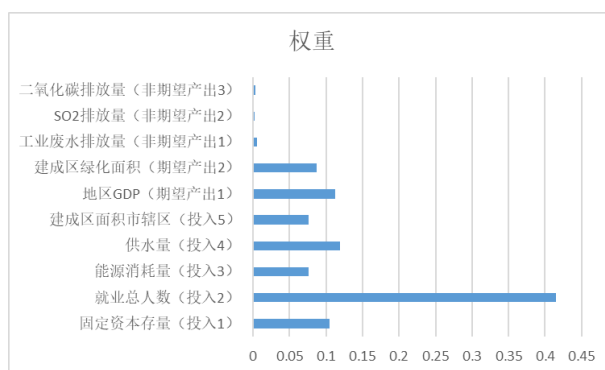
计算结果如表所示:

熵值e	固定资本	就业总人	能源消耗	供水量	建成区面积	地区GDP	建成区绿化	工业废水	SO2排放量	二氧化碳排放量	(非期望产出3)
	0.87	0.48293	0.905725	0.852085	0.905764	0.859498	0.891485	0.993269	0.998095	0.996624	
冗余度	固定资本	就业总人	能源消耗	供水量	建成区面积	地区GDP	建成区绿化	工业废水	SO2排放量	二氧化碳排放量	(非期望产出3)
	0.13	0.51707	0.094275	0.147915	0.094236	0.140502	0.108515	0.006731	0.001905	0.003376	
权重	固定资本	就业总人	能源消耗	供水量	建成区面积	地区GDP	建成区绿化	工业废水	SO2排放量	二氧化碳排放量	(非期望产出3)
	0.104457	0.415476	0.075752	0.118852	0.075721	0.112896	0.087194	0.005408	0.001531	0.002713	1

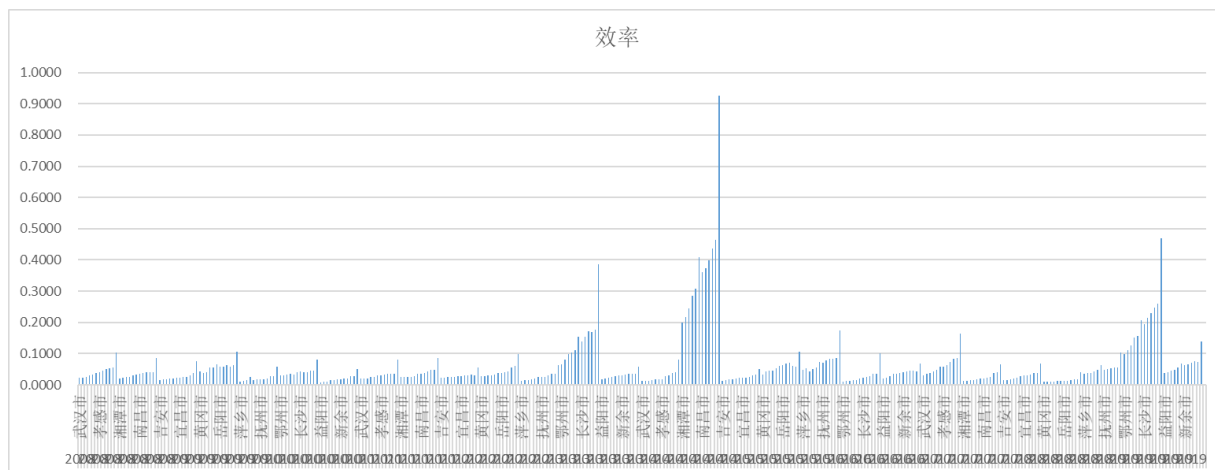
表二

#### (4) 生态效率测度

按照如图所示的权重，对标准化后的数据进行加权求和，得出效率。



图四



图五

3.2 变异系数及回归分析

3.2.1 变异系数

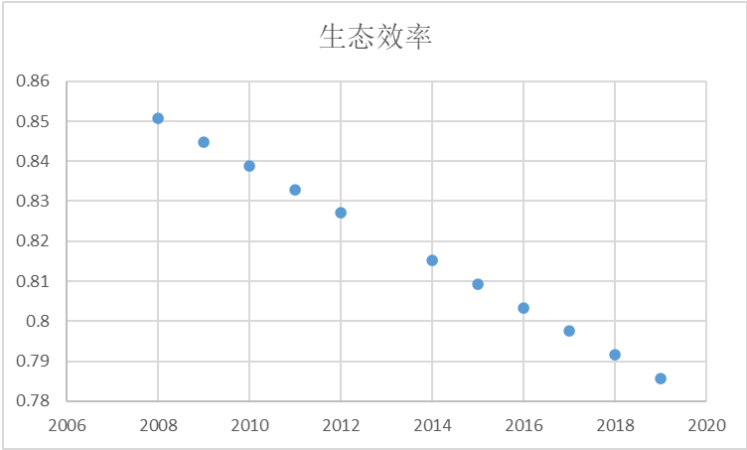
使用变异系数公式求得每年的长江中游城市群的变异系数。具体如表所示：

2008	0.850669
2009	0.844759
2010	0.838848
2011	0.832938
2012	0.827028
2014	0.815208
2015	0.809298
2016	0.803388
2017	0.797478
2018	0.791567
2019	0.785657

表二

3.2.2 回归分析

对表二数据进行散点图绘制有：



图六

可以看出图中数据有较为明显的降低趋势，使用回归分析后，有如下结果：

回归统计	
Multiple	0.772552
R Square	0.596836
Adjusted	0.55204
标准误差	0.019344
观测值	11

表三

	df	SS	MS	F	Significance F
回归分析	1	0.004985	0.004985	13.32344	0.005316467
残差	9	0.003368	0.000374		
总计	10	0.008353			

表四

可以直观的看出，R 方接近 0.6，方差分析显著性良好，说明变异系数有随时间变化而减小的趋势，即可说明长江中游城市群生态效率差异将逐渐减小。

### 3.3 核密度估计

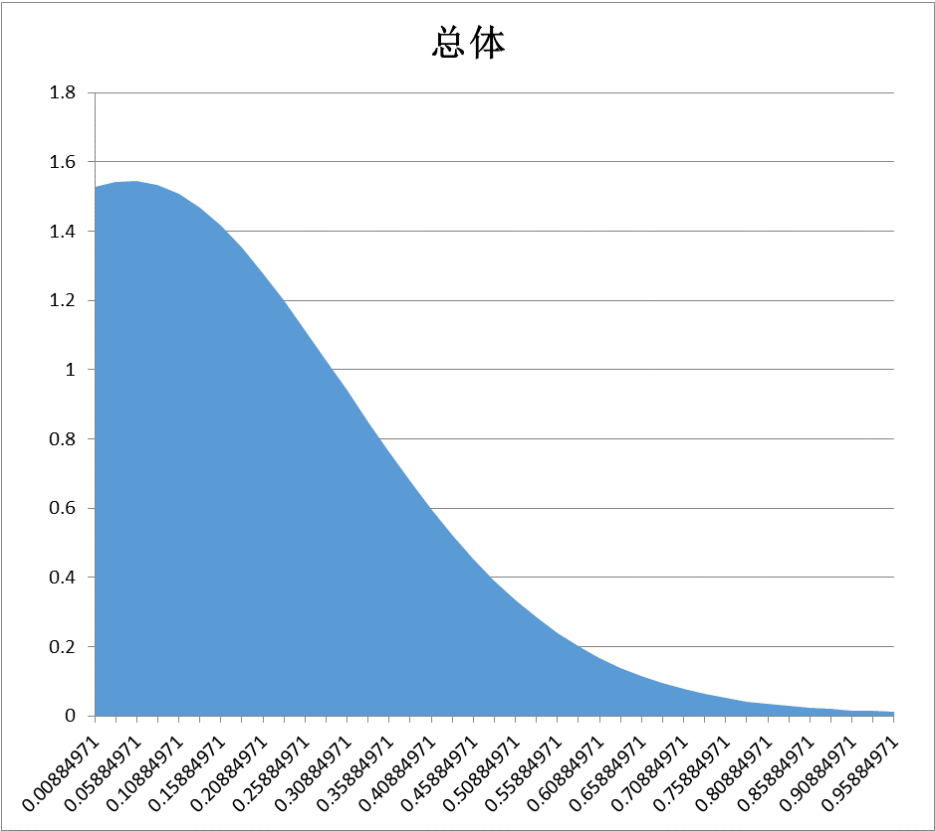
作为非参数估计方法，核密度估计用连续密度曲线刻画随机变量分布形态，反映出变量的分布位置、形态和延展性特征，被广泛应用于空间非均衡分析中。设定随机变量  $x$  的密度函数为  $f(x)$ ，点  $x$  密度函数由式(11)估计，其中  $N$ 、 $X_i$ 、 $h$  和  $K(x)$  分别为观测值个数、独立同分布观测值、带宽和核函数。核函数  $K(x)$  是一种加权函数或平滑转换函数，本研究采用常见的高斯核函数，计算方式见式(5)。此外，核密度估计对带宽  $h$  较为敏感，一般多选择较小带宽以提高估计精度。

$$f(x)=(1/Nh)\sum_{i=1}^N K[(X_i-x)/h] \tag{4}$$

$$K(x)=(1/\sqrt{2\pi})\exp(-x^2/2) \tag{5}$$

#### 3.3.1 总体核密度

通过使用 Easychart 插件，绘制了总体的核密度图像：

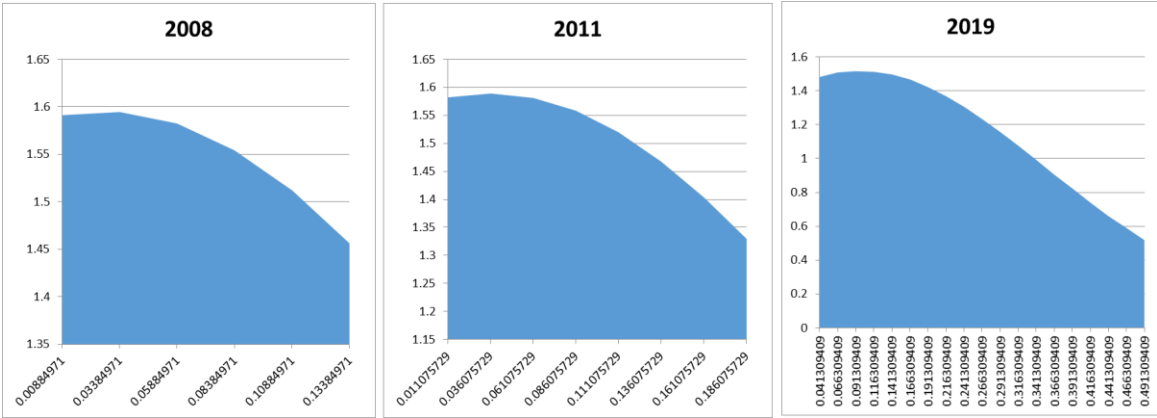


图七



### 3.3.2 核密度变化趋势

随后，绘制了每年的核密度图像，通过比对每年核密度图像的变化，可以得出相应结论，其中以 2008、2011、2019 三年的图像为例：



图八

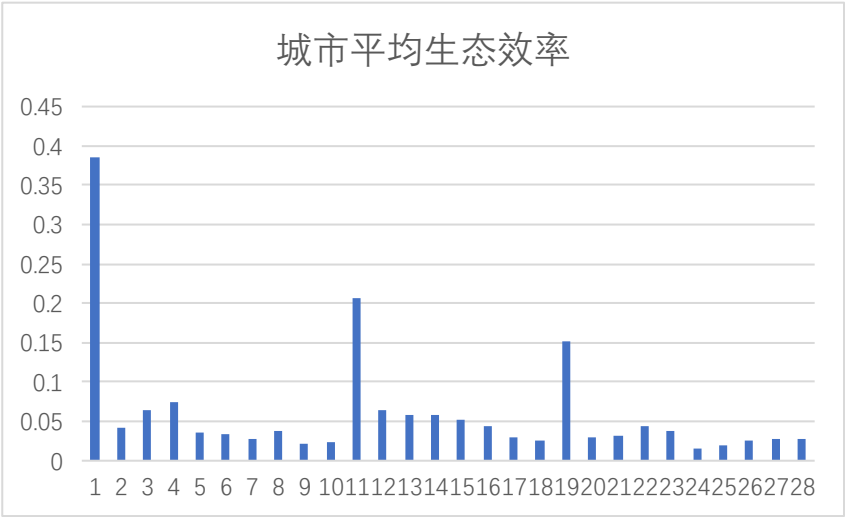
2008-2019 期间，长江中游城市群总体分布曲线随时间推移呈现小幅度左移趋势，表明城市群生态效率存在下行压力。因此长江中游城市群应加大生态保护重视力度，促进资源集约利用和减少环境污染，改善各地区的生态效率。

长江中游城市群总体分布曲线存在右拖尾现象，说明部分城市的生态效率明显较高。如果将长江中游城市群进行细分生态区，即可计算哪个区域内处于低生态效率发展的城市的数量，从而区域性的定制发展战略。

### 3.4 聚类分析

利用 K-means 方法对生态效率进行聚类处理。

首先将选取的 28 个城市进行 2008~2019 年平均的生态效率计算（其中 1 至 28 为城市编号，纤细信息见 2008-2019 长江中游城市群 5.28.xlsx），计算结果如图：

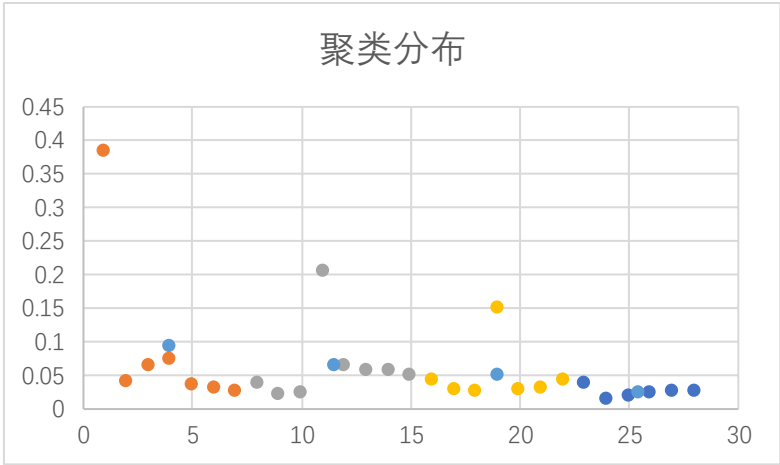


图九

随后，选取参数，k 为 4 类，最大迭代次数为 9，质心稳定精度为 1，进行聚类运算，结果如图所示：

类别	数目
1	6
2	7
3	8
4	7

表五



图十

根据聚类结果，我们可以将长江中游城市群的整体生态效率水平划分为 4 大类，每类对应一个区段的生态效率，将城市进行如上的划分，可以避免单纯从地理环境划分导致的不足（如：武汉生态效率很高，而周边城市却偏低），从而针对相似发展水平的城市采取相同的发展政策，这不光可以使各类城市充分发展，还可以减小差距，实现长江中游城市群一体化的提高。

四、 结论

从时序演进来看，长江中游城市群总体生态效率呈现出“下降-上升-下降-上升”的“w”形变化趋势，不同城市间的生态效率差异显著。但根据变异系数的变化趋势可知，各城市间的生态效率差距有明显的缩小趋势。

从分布动态演变来看，长江中游城市群总体和大部分生态区生态效率存在下降的趋势，生态效率差距有所增大。长江中游城市群总体不存在两极分化现象。长江中游城市群生态效率具有一定程度上保持原有水平的稳定性，向高水平提升的趋势不明显，有向低水平方向转移的趋势，存在“下行”风险，同时实现跨越式转移的概率较小。

## 五、模型缺点与不足

### 5.1 评价模型不足

只使用了熵值法进行评价，没有考虑到其他因素的影响，可以使用 Topsis 等评分方法，进行多种权重计算，最后使用均值处理。

### 5.2 生态效率分析不足

可以考虑时间序列进行分析，评价每个城市的生态效率增长率，并预测趋势。

### 5.3 地区发展水平预测不足

可以使用基尼系数法测度生态效率的地区相对差异，分析区域及其子地区的生态效率的空间特征。

## 六、课程心得

虽然在这学期的数学模型与决策的课上，和老师的交流不多，但在最后的实践过程中，还是很好地掌握了 Excel 的相关应用。

在平时的课程中，通过网课，学习了概率论和统计学的相关知识，复习回顾了上学期所学的概率论知识，和哲学亲所学的统计学对接，起到复习预习作用，收货颇多。

在最后的 Excel 实践中，计算各种模型时，充分地感受到了 Excel 的强大之处。同时，可以使用市面上的热门插件，进行快速准确的数据分析（如：easychart 画核密度估计图）。此外，还可以使用开源的宏代码进行模型的处理，做到了很多之前以为 Excel 无法做成的事。

在选题过程中，我学会了如何快速阅读文献，如何收集数据汇总，如何选取热门主流方法，可以说是收货很大。而在方法使用时，更是有了很多体会。

首先，选取的方法可能是全新的，之前未掌握的，这就要求我们在很短的时间内进行学习，并可以进行实际应用，这很有数模竞赛的感觉，不同的是，课程设计中我们有了更多的时间去学习一个方法，可以更好地去深入了解。但是，一个算法模型的学习，不能不深入仔细，所以在课程结束之后，我决定再深入地学习相关模型，在 B 站上观看相关网课，争取掌握它的工作原理。

在最后的实践中，我发现很多之前学过的方法都是可以投入到实际使用中的，比如最早的马尔卡夫，本以为从来不会用到，但却可以解决实际问题。还有相关系数法，在书上只有简单的定义，但在实际应用，却可以和回归分析结合计算反映差距减小的趋势，这让我有了一种融会贯通的感觉。

同时，也希望在课程结束后，更加深入地学习 Excel，掌握一些 VBA 语言，深入应用。