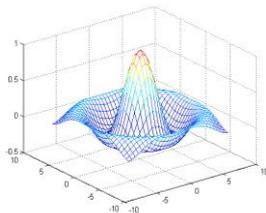


# 五一数学建模竞赛



## 题目：徐州潘安湖风景区游览路线设计问题

**关键词：**图论、规划模型、Lingo 软件、TSP 模型、穷举搜索

**摘要：**本文针对徐州潘安湖风景区的部分景点展开游览路线规划的研究，通过将问题图论化，TSP 问题的转化等，建立了最短路线模型，混合整数 0-1 规划模型，双目标混合整数规划模型等，运用了 lingo 算法，穷举法等得到了受时间约束的游览路线方案，并对结果进行了误差分析和相关验算。

问题一中，针对游览全部景点的最短路线规划问题，首先将问题图论化，将问题与 TSP 问题有机结合起来，通过引入 0-1 决策变量，建立了以线路距离最短为目标的 0-1 整数规划模型，利用 Lingo 软件求得总的步行距离为 1820 米，路线则为：景石→③→⑤→①→②→④→⑥→⑦。此外，我们还对不含有子巡回的约束条件表达式进行了严格的数学证明。

问题二中，由于游客在时间上存在约束，针对游览时间最长的路线规划问题，基于问题一，又因为总时间是由游览的总时间、景点之间的步行时间、在景区外的等待时间三部分组成。通过引入决策变量，建立了以总游览时间最长为目标的 0-1 混合整数规划模型，利用 Lingo 软件求得总游览时间为 270 分钟，总步行时间为 60 分钟，路线为：景石→②→④→③→⑤→①→⑥→⑦。

问题三中，针对 3 个旅游团的路线规划问题，通过顺延在景点外的等待时间来进行约束，由于每个景点只能容纳一个旅游团，基于问题二，建立了以总的游览时间最长为目标的混合 0-1 整数规划模型，采取穷举搜索结合 Lingo 软件的方法，求得最优结果：三个旅游团的最长的游览总时间分别为 270、261.9、268.8 分钟，而总步行时间分别是 60、68、61 分钟，无等待时间。此外，我们还利用 C 语言程序对各个旅游团在各个景点的时间节点是否冲突进行了验证，确保了结果的准确性。

问题四中，根据对各个旅行团的速度和平均速度的变化，基于问题三，建立了以游览总时间最长，总的等待时间最短为双目标的混合整数规划模型，仍然采取穷举搜索结合 Lingo 软件的方法，求得最优结果：3 个旅游团总游览时间分别为 270、268.8、268.8 分钟，总的步行时间分别为 60、61.2、61.2 分钟，且平均速度满足 2km/h，并用 C 语言冲突验证程序对结果的正确性进行了验证，特别注意的是 3 个旅行团的路线规划中有重复的路线选择，但每个旅游团在各个景点的游览时间并没有冲突。

问题五中，由于考虑了两个不确定因素：①不同旅游团从景点 1（景石）出发的时间具有不确定性；②每个景点的等待时间也存在不确定性。假设存在  $n$  个旅游团，基于问题四，建立了双目标的混合整数规划模型，并设定了数据进行了试算，得到可行的参考方案。

# 一、问题提出

## 1.1 背景

在党的十九大的积极推动下，习近平总书记来到徐州调研并在徐州调研贾汪区潘安采煤塌陷区整治工程时强调，塌陷区要坚持走符合国情的转型发展之路，打造绿水青山，并把绿水青山变成金山银山。对于新建设的潘安湖景区，游客在游览时如果被安排的游览路线在规定的时间内，更大限度地参观各个景点，则会令游客在旅程中感到景区服务的周到，从而使景区的人流量越来越多。但是人流量多也会带来一些不必要的麻烦，如多个旅游团在同一个景点进行参观，则会使参观体验变得很糟糕，倘若能很好地处理多个旅行团同一时间尽量不在相同的景点游览，则会很轻松地解决这一麻烦，这一类的问题都是徐州潘安湖景区游览路线设计中面临的一个重要课题。

## 1.2 需要解决的问题

试就徐州潘安湖风景区的游览路线设计建立数学模型，研究下列问题：

(1) 若游客从景石出发，步行游览的景点共有 8 处，分别为景石，游客服务中心，阳光草坪，森林小剧场，儿童科普体验区，儿童戏水场，湿地博物馆，湿地商业街。求出从景点 1（景石）出发到达景点 8（湿地商业街），并且经过景点 2 至 7 至少一次的距离最短的路线，计算该路线的长度，并将相关结果填入表 3。（在每个景点不用停留）

(2) 如果某游客 12:00 从景石出发，要求他 17:00 前到达湿地商业街，17:30 离开湿地商业街（根据表 2 的要求在湿地商业街游览时间至少为 30 分钟）。如何建立相关的数学模型，才能为该游客设计一条能游览完全部景点（景点 2 至 8）且游览总时间最长的游览路线（假设在各个景点没有等待时间），并完成表 4 的填写。

(3) 随着人流量的增加，相应的旅游团也会增多，在第二问的基础上，又额外增加了 3 个旅游团，除了湿地商业街以外每个景点只能容纳 1 个旅游团游览，按照时间顺序后到达的旅游团，需求等待先到达的旅游团游览结束后才能开始游览，如何建立数学模型，才能为三个旅游团分别设计一条能游览完全部 7 个景点且游览总时间最长的游览路线，并完成表 5 的填写。

(4) 假设 3 个旅游团的步行速度可以在 1km/h 到 3km/h 之间调节，但是总的平均速度不能超过 2km/h，3 个旅游团 12:00 同时从景点 1（景石）出发，要求三个旅游团 17:00 前到达湿地商业街，17:30 离开湿地商业街且在湿地商业街游览时间至少为 30 分钟，并且湿地商业街除外的每个景点，同时只能容纳 1 个旅游团游览，按照时间顺序后到达的旅游团，需要等待先到达的旅游团游览结束之后才能开始游览。建立数学模型，为三个旅游团分别设计一条能游览完全部景点且游览总时间长，总的等待时间短的游览路线，并完成表 6 的填写。

(5) 在现实中，如果考虑两个不确定性因素：①不同旅游团从景石出发的时间具有不确定性，例如，多个旅游团在不同的时间从景石出发开始游览，在此情况下到达湿地商业街的时间可以顺延；②每个景点的等待时间也存在不确定性因素，例如，旅游设施短时间的维护和清理，或者受到散客客流的影响，其他条件在满足问题 4 的基础上，建立数学模型，为多个旅游团分别设计一条能游览完全部景点且游览总时间长，总的等待时间短的游览路线。

## 二、基本假设

- (1) 假设正常情况下人的步行速度为  $V=2\text{km/h}$ 。
- (2) 假设游客在景区停留的时间由“景点之间的步行时间”、“景点游览时间（即在景点内游玩的时间）”和“在景区外的等待时间”三部分组成，其他时间忽略不计。
- (3) 假设在游玩过程中，游客们不会遇到突发事件，所有的游览均是正常进行。
- (4) 假设每个景点的旅游设备的维护与清理都不会影响游客们的正常游览时间。

## 三、符号说明

符号	意义	单位	备注
$t_b$	景点之间的步行时间	分钟	
$t_i$	在每个景点的游览时间	分钟	$i=1, 2\cdots 8$
$w_i$	在景区外的等待时间	分钟	$i=1, 2\cdots 8$
$d_{ij}$	景点 $i$ 到景点 $j$ 的最短距离	米	$i, j=1, 2\cdots 8$
$x_{ij}$	0-1 变量，表示是否选择从景点 $i$ 到景点 $j$ 的路线		$i, j=1, 2\cdots 8$
$u_i$	辅助变量，用于破圈		$i=1, 2\cdots 8$
$S$	路线的总长度	米	
$t_i$	游览景点 $i$ 的时间	分钟	$i=1, 2\cdots 8$
$y_i$	到达景点 $i$ 的时间	分钟	$i=1, 2\cdots 8$
$z_i$	离开景点 $i$ 的时间	分钟	$i=1, 2\cdots 8$
$w$	在景点 4（森林小剧场）前等待的时间	分钟	
$M$	一个充分大的正数		
$V$	人步行的速度	千米/小时	$V=2$ 千米/小时
$T$	游览的总时长	分钟	

P	辅助变量		自然数
K	旅行团的编号		$k=1,2,3$
$T_k$	旅行团游览的总时间	分钟	$k=1,2,3$
$W_k$	旅行团等待的总时间	分钟	$k=1,2,3$
$B_k$	旅行团步行的总时间	分钟	$k=1,2,3$

## 四、问题分析

### 4.1 背景

在前期的准备工作中，我们发现本问题与 TSP 问题有着相应的联系，并且可以用混合整数规划来替代，此外问题的图论化也是一个需要了解的方面，图论化会将许多现实中的实际问题简单化，对于此类问题的研究有着积极的作用。

### 4.2 所需问题的分析

(1) 根据问题一的说明，从景点 1（景石）出发，必须经过景点 2 至 7 至少一次，最后到达景点 8（湿地商业街），则可联想到问题的图论化，有利于我们对于路线筛选的准确性。此外图的连接也是一个需要解决的问题，需要连接 8 处景点，且必须从景点 1 进，景点 8 出。如果能将本问题的 TSP 模型转换成混合整数规划，则可通过相应的约束条件结合目标函数利用 Lingo 软件得出最短的路线及长度。

(2) 根据问题二的说明，必须中午 12:00 从景石出发，17:30 离开湿地商业街，则总的时间为 330 分钟，有根据表 2 给出的相应游览时间以及景点开放时间，我们需要对这些时间进行约束，结合问题一建立的模型，找出游览全部景点的路线，而总时间又由景点之间的步行时间，景点游览时间，和在景点外的等待时间三部分组成，则我们可以得到游览的总时间=总时间 - 景点之间的步行时间 - 在景区外的等待时间。而通过路程时间计算公式，在已知总路线长度的前提下，可以求出景点之间的步行时间；景点外等待的时间只有在景点 4（森林小剧场）才会出现，再对比路线的规划，求出到达景点 4（森林小剧场）时所耗的总时间，即可推算出到达景点 4（森林小剧场）的时间，从而对比相应时间段的开放时间，则可求出在景点外的等待时间，即运用公式，可得出游览的总时间，即为目标函数。

(3) 根据问题三的说明，若现有 3 个旅行团，同时从景点 1 出发，且时间同为 12:00，并于 17:00 前均到达景点 8，然后于 17:30 离开。而问题三就是在问题二的基础上增加一个变量，关于旅行团的编号。但是每个景点只能容纳 1 个旅游团，且按照时间顺序后到达的旅游团需等到先到达的团队游览结束后才能开始游览，则不同于问题二的相关约束就只是多了一个变量即旅行团的编号。同时对于每个景点只能容纳一个旅游团，那就只好通过增加在景点外的等待时间来实现这一问题的约束，从而建立相关模型与目标函数，总结出总的约束条件，得出 3 个旅行团对应的游览路线与时间规划。

(4) 根据问题四的说明，表明在问题三的基础上设置了对游客步行速度的限制，即旅行团的步行速度在 1km/h 到 3km/h 之间调节，但是总的平均速度不能超过 2km/h，

则只需再引入一个变量速度，在满足平均速度和各自旅行团的步行速度的约束下，就可以将总的约束条件分析完毕，但是问题四中的目标函数为双目标函数，既要分配三个旅游团的游览总时间长的路线，也要总的等待时间短的同一游览路线，所以就可以建立混合整数规划模型来得出双目标函数下的结果。

(5) 根据问题五的说明，只是在问题四条件的基础上考虑了两个不确定因素：①不同旅游团从景点 1（景石）出发的时间具有不确定性；②每个景点的等待时间也存在不确定性。为此可以设立一个不确定因素  $\varepsilon_{ik}$  来表明第  $k$  个旅游团在每个景点的不确定等待时间；再给出一个常量  $Y_k$  来表明第  $k$  个旅游团从景点 1（景石）出发的时间。又因为题设中只说明了多个旅行团，即表明可以用  $k$  个旅行团来参与模型的建立（ $k$  没有限制），再基于问题四，建立数学模型来完成考虑不确定因素的多个旅游团的路线选择的猜想。

## 五、模型的建立与求解

### 5.1 问题一模型的建立与求解

#### 5.1.1 问题一的分析

将本题中的 8 个景点看成一个图的顶点  $V$ ，连接景点的道路看成一个图的边  $E$ ，这样就得到一个图  $G=(V,E)$ 。在图的结构中，任意两个结点之间都可能相关，即结点之间的邻接关系是任意的，除去景点 1 与景点 8，我们将剩下的景点可看成无向图，倘若包括景点 1 和 8，那么则转换为有向图，而本题的求解即是混合图。而图与矩阵关系十分密切，又根据表 1 给出的路径，可以得到相应的距离矩阵，再与 0-1 变量和路径矩阵相结合，则可使问题简单化。查阅资料<sup>[1][2][3][4]</sup>知道，本问题与 **TSP 问题** 非常接近，而 TSP 问题又称为**旅行商问题**，迄今为止，这类问题中没有找到一个有效的算法。倾向于接受 NP 完全问题（NP-Complete 或 NPC）和 NP 难题（NP-Hard 或 NPH）不存在有效算法这一猜想，认为这类问题的大型实例不能用精确算法求解，必须寻求这类问题的有效的近似算法，而对于较少顶点的情况下可以使用整数规划的方法<sup>[1][2][3][4]</sup>，本题解法的一大特色就是将 TSP 模型转换成混合整数规划。

景点 1 至 8 分别为景石, 游客服务中心, 阳光草坪, 森林小剧场, 儿童科普体验区, 儿童戏水场, 湿地博物馆, 湿地商业街, 由题目条件可以得到图的距离矩阵如下:

$$(d_{ij})_{8 \times 8} = \begin{pmatrix} 0 & 300 & 360 & 210 & 590 & 475 & 500 & 690 \\ 300 & 0 & 380 & 270 & 230 & 285 & 200 & 390 \\ 360 & 380 & 0 & 510 & 230 & 765 & 580 & 770 \\ 210 & 270 & 510 & 0 & 470 & 265 & 450 & 640 \\ 590 & 230 & 230 & 470 & 0 & 515 & 260 & 450 \\ 475 & 285 & 765 & 265 & 515 & 0 & 460 & 650 \\ 500 & 200 & 580 & 450 & 260 & 460 & 0 & 190 \\ 690 & 390 & 760 & 640 & 450 & 650 & 190 & 0 \end{pmatrix}$$

很明显，这是一个**完全图**，任意两个顶点之间都有边相连接。相当于下面的图：

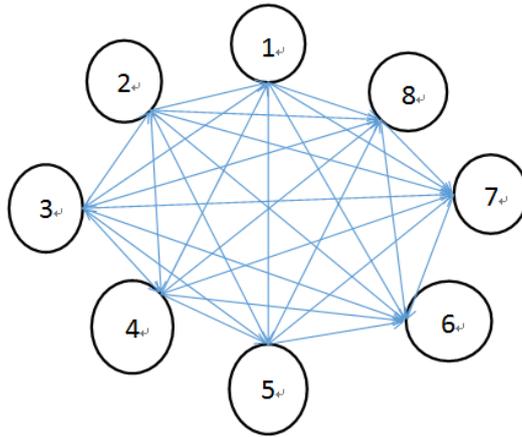


图 1: 完全图

### 5.1.2 问题一模型的建立

#### (1) 符号说明

$i, j$  —— 景点的编号,  $i, j=1, 2 \dots 8$

$x_{ij}$  —— 决策变量, 取 1 表示选择游览路线从景点  $i$  到景点  $j$ , 反之则不选

$d_{ij}$  —— 景点  $i$  到景点  $j$  的距离

#### (2) 目标函数分析

所求路线的总长度就是在相对应路线下各个景点之间最短步行距离的总和, 即得出目标函数:

$$\min S = \sum_{i=1}^8 \sum_{j=1}^8 d_{ij} \times x_{ij}$$

#### (3) 约束条件分析

① 只能从景点 1 开始出发到其他景点, 其他景点不能返回景点 1 的约束, 即

$$\sum_{j=2}^8 x_{1j} = 1, \sum_{i=2}^8 x_{i1} = 0$$

② 只能从其他景点结束于景点 8 的约束, 即

$$\sum_{i=1}^7 x_{i8} = 1, \sum_{j=1}^7 x_{8j} = 0$$

③ 进入各个景点且每个景点只游览一次的约束, 即

$$\sum_{i=1, i \neq j}^8 x_{ij} = 1, j = 2, 3, \dots, 7$$

④ 离开各个景点的约束, 即

$$\sum_{j=1, j \neq i}^8 x_{ij} = 1, i = 2, 3, \dots, 7$$

⑤ 路线选择的 0-1 变量的约束, 即

$$x_{ij} \in \{0,1\}, i=1,2,\dots,8; j=1,2,\dots,8$$

⑥在 TSP 问题中会出现所谓的圈，也称为子巡回，这是我们要避免的，增加约束如下，其证明见本问题的验证。

$$u_i - u_j + 8x_{ij} \leq 7; u_i > 0, u_j \geq 0; i=1,2,\dots,8; j=2,3,\dots,8$$

综上所述可得本问的 0-1 整数规划模型 I 如下：

$$\begin{aligned} \text{Min } S &= \sum_{i=1}^8 \sum_{j=1}^8 d_{ij} \times x_{ij} \\ \text{s.t. } &\begin{cases} \sum_{j=2}^8 x_{1j} = 1, \sum_{i=2}^8 x_{i1} = 0 \\ \sum_{i=1}^7 x_{i8} = 1, \sum_{j=1}^7 x_{8j} = 0 \\ \sum_{i=1, i \neq j}^8 x_{ij} = 1, j=2,3,\dots,7 \\ \sum_{j=1, j \neq i}^8 x_{ij} = 1, i=2,3,\dots,7 \\ x_{ij} \in \{0,1\}, i=1,2,\dots,8; j=1,2,\dots,8 \\ u_i - u_j + 8x_{ij} \leq 7; u_i > 0, u_j \geq 0; i=1,2,\dots,8; j=2,3,\dots,8 \end{cases} \end{aligned}$$

### 5.1.3 问题一模型的求解

#### (1) 求解步骤

【Step 1】 首先定义一个集合 cities，cities 中 8 个元素代表 8 个景点，定义一个矩阵 d，其为任意两个景点之间的最短距离矩阵，且矩阵大小为 8×8，且横纵坐标都是景点。

【Step 2】 定义 x 为 (0,1) 变量。

【Step 3】 给出 d 的数据集合。

【Step 4】 设 n=景点的个数，设计好目标函数为整个程序的求解目标 min=各个景点之间的最短距离×判断走不走这条路的 (0,1) 变量 x。

【Step 5】 定义约束条件：

1. 循环经历 x 的每一列， $x(1,j)=1$ ，且  $j \neq 1$ 。
2. 循环经历 x 的每一行， $x(i,8) = 1$ ，且  $i \neq 8$ 。
3. 循环经历 x 的每一列， $x(j,i)=1$ ，且  $j \neq i$ 。
4. 循环经历 x 的每一行， $x(i,j)=1$ ，且  $j \neq i$ 。
5. 循环经历 x 的每一行每一列，

$$level(j) \geq level(i) + x(i, j) - (n-2) \\ \times 1 - x(i, j) + (n-3) \times x(j, i), \text{且 } j > 1, j \neq i.$$

6. 限制0-1变量  $x(i, j)$

7. 限制

$$level(i) \leq n-1 - (n-1) \times x(i, j), \quad level(i) \geq 1 + (n-2) \times x(i, 1) \text{ 且 } i \geq 1$$

【Step 6】结合 Lingo 软件编写程序求解，详细程序见附录 1：求解问题一的最短距离的 LINGO 程序。

## (2) 计算结果

表1：问题一结果

出发景点	到达景点	步行距离(米)
景石	森林小剧场	210
森林小剧场	儿童戏水场	265
儿童戏水场	游客服务中心	285
游客服务中心	阳光草坪	380
阳光草坪	儿童科普体验区	230
儿童科普体验区	湿地博物馆	260
湿地博物馆	湿地商业街	190
总步行距离(最短路线距离)		1820
最短路线(请用①~⑧序号标出)		景石→③→⑤→①→②→④→⑥→⑦

### 5.1.4 问题一结果的分析及验证

(1) 对于上述结果的准确性我们通过验算求得从出发点景石到最终点湿地商业街的步行距离为 1820 米，且随机抽取了几组不同路线进行对比得出相应的结果均大于 1820 米，从而结果无误，此外我们又在问题一的分析中解题思路里运用反证法等方法来证明所附加的约束条件。为了使游览路线更好地从结果中呈现，于是做出简易的路线图：

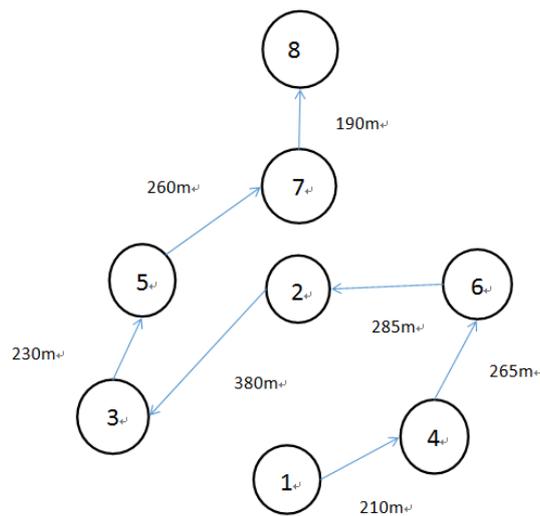


图 2：问题 1 结果路线图

1. 景石, 2. 游客服务中心, 3. 阳光草坪, 4. 森林小剧场,  
5. 儿童科普体验区, 6. 儿童戏水场, 7. 湿地博物馆, 8. 湿地商业街。

## (2) 限制无子巡回的证明

在约束条件中我们增加了下面的约束

$$u_i - u_j + 8x_{ij} \leq 7; u_i > 0, u_j \geq 0; i = 1, 2, \dots, 8; j = 2, 3, \dots, 8$$

其目的是避免出现圈，也就是子巡回，下面我们分两个方面来证明。

①任何含子巡回的路线一定不满足题目本身的约束条件。由反证法得，假设游览路线中存在子巡回，则至少存在两个子巡回，那么一定有一个子巡回中不含景点 1，如子巡回（景点 3-2-6-4-2），则一定有  $u_3 - u_2 + n \leq n-1$ ， $u_2 - u_6 + n \leq n-1$ ， $u_6 - u_4 + n \leq n-1$ ， $u_4 - u_2 + n \leq n-1$ ，将这些公式累加得  $n \leq n-1$ ，显然假设矛盾，则游览路线中不存在子巡回；

②全部不含子巡回的路线都可以满足题目本身的约束条件。对于全部的游览路线即总巡回，若  $u_i$  取成合适的值，均可以满足该题的约束条件。当对于总巡回上的边时， $x_{ij} = 1$ ，而令每游览一个景点， $i$  加 1，即  $u_i - u_j = -1$ ，则附加的约束条件显然成立（ $-1 + n \leq n-1$ ）；当对于不是总巡回上的边时， $x_{ij} = 0$ ，则变为  $-1 < n-1$ ，显然成立。综上所述，附加的约束条件成立。

## 5.2 问题二模型的建立与求解

### 5.2.1 问题二的分析

通过表 2 给出的各景点的游览时间及相应的开放时间，为了计算的简便，将小时换算成分钟，即某游客中午 12:00 从景石出发一直到下午 17:30 离开湿地商业街，共计时长 330 分钟。而总时长=景点之间的步行时间 + 景点游览时间（即在景点内游玩的时间 + 在景区外的等待时间。而通过表 1 给出的任意两景点之间的最短步行距离的路径矩阵，且游客步行速度恒为  $V=2\text{km/h}$ ，则可以求出相应的步行时间矩阵，利用问题一建立的模型可得出游览全部景点的相应路线。而对于在景区外的等待时间只有在景点 4（森林小剧场）才会出现，则只要考虑到达景点 4 时当前所耗的时间，由于森林小剧场只在整点或者半点开放，则可以通过简单的运算法则就可得出在景区外的等待时间。从而可作出相应的约束条件，再由游览的总时间=总时间 - 景点之间的步行时间 - 在景区外的等待时间则可确立相应的目标函数，再得出一条能游览全部景点且游览时间最长的游览路线，从而完成求解。

### 5.2.2 问题二模型的建立：

#### (1) 符号说明

$x_{ij}$ ——决策变量，取 1 表示选择游览路线从景点  $i$  到景点  $j$ ，反之则不选

$d_{ij}$ ——景点  $i$  到景点  $j$  的距离

$t_i$ ——游览景点  $i$  的时间， $i = 1, 2, \dots, 8$ ，单位：分钟

$y_i$ ——到达景点  $i$  的时间， $i = 1, 2, \dots, 8$ ，单位：分钟

$z_i$  ——离开景点  $i$  的时间,  $i = 1, 2, \dots, 8$ , 单位: 分钟

$w$  ——在景点 4 (森林小剧场) 前的等待的时间, 单位: 分钟

$M$  ——表示一个充分大的正数, 比如取 99999

$V$  ——人步行的速度

## (2) 目标函数分析

由于问题二中所求游览总时间是在每个景区所游览时间的总和, 即得出目标函数:

$$\max T = \sum_{i=1}^8 t_i$$

## (3) 约束条件分析

又因为问题二相对于问题一的基础上又增加了时间上的相关约束, 所以只要将时间加以限制, 而其他约束条件不变, 关于时间方面的相关约束, 可结合表 2 给出的游览时间及开放时间, 每一次进入一个景点都必须要在时间范围内才能成功游览, 而景点 4 (森林小剧场) 总是半点或整点开放, 所以等待时间再加上到达景点 4 的时间也必须时半点或整点, 再结合进出景点以及路线选择的约束, 就可以得到问题二中总的约束条件。

①只能从景点 1 开始出发到其他景点, 其他景点不能返回景点 1 的约束, 即

$$\sum_{j=2}^8 x_{1j} = 1, \sum_{i=2}^8 x_{i1} = 0$$

②只能从其他景点结束于景点 8 的约束, 即

$$\sum_{i=1}^7 x_{i8} = 1, \sum_{j=1}^7 x_{8j} = 0$$

③进入各个景点且每个景点只游览一次的约束, 即

$$\sum_{i=1, i \neq j}^8 x_{ij} = 1, j = 2, 3, \dots, 7$$

④离开各个景点的约束, 即

$$\sum_{j=1, j \neq i}^8 x_{ij} = 1, i = 2, 3, \dots, 7$$

⑤路线选择的 0-1 变量的约束, 即

$$x_{ij} \in \{0, 1\}, i = 1, 2, \dots, 8; j = 1, 2, \dots, 8$$

⑥在 TSP 问题中会出现所谓的圈, 也称为子巡回, 这是我们要避免的, 增加约束:

$$u_i - u_j + 8x_{ij} \leq 7; u_i > 0, u_j \geq 0; i = 1, 2, \dots, 8; j = 2, 3, \dots, 8$$

⑦每个景点的游览时间=离开时间-到达时间, 即

$$z_i - y_i = t_i, i = 1, 2, 3, 5, 6, 7, 8$$

⑧每个景点的开放时间、包括到达时间和离开时间、特别是小剧场的半点或整点开放的特殊要求, 得到约束条件详见下面的大模型。

⑨如果  $x_{ij} = 1$ ，则离开景点  $i$  的时间  $z_i$ 、景点  $i, j$  之间的步行时间、到达景点  $j$  的时间  $y_j$  满足下面的约束：

$$(x_{ij} - 1) \cdot M \leq y_j - z_i - \frac{d_{ij}}{V} \leq (1 - x_{ij}) \cdot M, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j$$

综上所述可得问题 2 的混合整数规划模型 II 如下：

$$\begin{aligned} \max \quad & T = \sum_{i=1}^8 t_i \\ \text{s.t.} \quad & \begin{cases} \sum_{j=2}^8 x_{1j} = 1, \sum_{i=2}^8 x_{i1} = 0 \\ \sum_{i=1}^7 x_{i8} = 1, \sum_{j=1}^7 x_{8j} = 0 \\ \sum_{i=1, i \neq j}^8 x_{ij} = 1, j = 2, 3, \dots, 7 \\ \sum_{j=1, j \neq i}^8 x_{ij} = 1, i = 2, 3, \dots, 7 \\ x_{ij} \in \{0, 1\}, i = 1, 2, \dots, 8; j = 1, 2, \dots, 8, i \neq j \\ u_i - u_j + 8 * x_{ij} \leq 7, u_i, u_j \geq 0, i = 2, 3, \dots, 8, j = 2, 3, \dots, 8, i \neq j \\ z_i - y_i = t_i, i = 1, 2, 3, 5, 6, 7, 8 \\ y_1 = z_1 = 12 \times 60 \\ y_2 \geq 9 \times 60, z_2 \leq 16 \times 60 \\ y_i \geq 9 \times 60, z_i \leq 17 \times 60, i = 3, 5, 6, 7 \\ y_4 + w = 30 \times p, p \in Z \\ y_8 \leq 17 \times 60, z_8 = 17.5 \times 60 \\ 10 \leq t_2 \leq 30, 20 \leq t_3 \leq 60, t_4 = 30, 30 \leq t_5 \leq 60 \\ 20 \leq t_6 \leq 60, 30 \leq t_7 \leq 60, t_8 \geq 30 \\ (x_{ij} - 1) \cdot M \leq y_j - z_i - \frac{d_{ij}}{V} \leq (1 - x_{ij}) \cdot M \\ (i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j) \end{cases} \end{aligned}$$

### 5.2.3 问题二模型的求解:

#### (1) 求解步骤

【Step 1】定义三个一维长度为 8 的集合 cities，属性有 u, y, z;

【Step 2】定义三个二维尺寸为 8x8 的集合，属性有 d, t, x;

【Step 3】给变量 d 赋值为最短距离矩阵，给 t 赋值为最短路径用时矩阵;

【Step 4】定义变量 n 为集合 cities 的长度;

【Step 5】定义目标函数为  $\min = \sum_{i=1}^8 (z(i)-y(i))$ ;

【Step 6】给出  $z(i)-y(i)$  ( $i \in 1, 2, \dots, 8$ ) 的大小范围以及相等关系的约束条件;

【Step 7】给出限制不出现子巡回的约束条件;

【Step 8】结合 Lingo 软件编写程序求解，详细程序见附录 2: 求解问题二的最长游览时间的 LINGO 程序。

#### (2) 计算结果

表 2: 问题二结果

序号	景点名称	达到时间点	游览时间 (停留时间, 单位分钟)	离开时间点
1	景石	12:00	0	12:00
2	阳光草坪	12:10:48	60	13:10:48
3	儿童科普体验区	13:17:42	58.2	14:15:54
4	森林小剧场	14:30	30	15:00:00
5	儿童戏水场	15:07:57	30	15:37:57
6	游客服务中心	15:46:30	10	15:56:30
7	湿地博物馆	16:25	30	16:32:30
8	湿地商业街	16:38:12	51.8	17:30
总的游览时间:			270 分钟	
总的步行时间:			60 分钟	

### 5.2.4 问题二结果的分析及验证:

对于上述结果的准确性，我们通过画路线图以及多方位的验算得知在游览全部景点的前提下，总的游览时间最长为 270 分钟，又由于总时间为 330 分钟，则相应的步行时间也会求出，为 60 分钟。此外我们有随机抽取了几条不同的全部游览景点的路线，对比之下，总游览时间均低于 270 分钟，则结果无误。问题 2 结果的路线图如下:

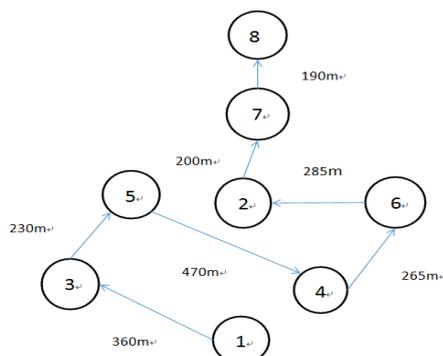


图 3: 问题 2 结果路线图

### 5.3 问题三模型的建立与求解

#### 5.3.1 问题三的分析

由于问题三相当于在问题二的基础上只是增加了旅行团的数量，但对于游览的景点也存在限制，即每个景点同时只能容纳 1 个旅游团，且按照时间顺序后到达的旅游团需等到先到达的团队游览结束后才能开始游览。所以只有约束在景点外等待的时间和后一个旅游团等待前一个旅游团的时间才能满足这一限制，从而建立目标函数来得出结果。

#### 5.3.2 问题三模型的建立

##### (1) 符号说明

$k$  ——旅行团的编号， $k=1,2,3$

$x_{ijk}$  ——取 1 表示第  $k$  旅行团选择游览路线从景点  $i$  到景点  $j$ ，反之则不选

$t_{ik}$  —— $k$  号旅游团游览景点  $i$  的时间， $i=1,2,\dots,8$ ，单位：分钟

$y_{ik}$  —— $k$  号旅游团到达景点  $i$  的时间， $i=1,2,\dots,8$ ，单位：分钟

$z_{ik}$  —— $k$  号旅游团离开景点  $i$  的时间， $i=1,2,\dots,8$ ，单位：分钟

$w_{ik}$  —— $k$  号旅游团在在景点  $i$  前的等待的时间， $i=1,2,\dots,8$ ，单位：分钟

##### (2) 目标函数分析

对于总的游览时间可以由 3 个旅游团的游览时间的和组成，则得出目标函数：

$$\max T = T_1 + T_2 + T_3$$

##### (3) 约束条件分析

问题三中的很多约束与问题二大致相同，只是需要增加一个下标  $k$ ；每个旅游团的总时间都是由相应  $k$  下的总游览时间所决定。不同的就是每个景点只能容纳一个旅游团，则只有增加在景点外等待的时间才能约束每个景点只能容纳一个旅游团。此外每一个旅游团到达一个景点之前与所离开的景点之间的时间也会受路线前进与否，即 0-1 变量所限制，则总结所有就可得出相关的约束条件。

①每个旅游团的游览时间就是相应  $k$  值下的总游览时间，即

$$T_k = \sum_{i=1}^8 t_{ik}, k=1,2,3$$

②第  $k$  个旅游团只能从景点 1 出发且不能返回景点 1 的约束，即

$$\sum_{j=2}^8 x_{1jk} = 1, \sum_{i=2}^8 x_{i1k} = 0, k=1,2,3$$

③第  $k$  个旅游团只能结束于景点 8 且从景点 8 开始不能在游览其他景点的约束，即

$$\sum_{i=1}^7 x_{i8k} = 1, \sum_{j=1}^7 x_{8jk} = 0, k=1,2,3$$

④第  $k$  个旅游团进入各个景点的约束，即

$$\sum_{i=1, i \neq j}^8 x_{ijk} = 1, j = 2, 3, \dots, 7, k = 1, 2, 3$$

⑤第 k 个旅游团离开各个景点的约束，即

$$\sum_{j=1, j \neq i}^8 x_{ijk} = 1, i = 2, 3, \dots, 7, k = 1, 2, 3$$

⑥第 k 个旅游团的路线中不含有子巡回的约束，即

$$u_{ik} - u_{jk} + 8x_{ijk} \leq 7, u_{ik}, u_{jk} \geq 0, i = 1, 2, \dots, 8, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3$$

⑦第 k 个旅游团到达时间、等待时间、游览时间、离开时间的关系约束，即

$$y_{ik} + w_{ik} + t_{ik} = z_{ik}, w_{ik} \geq 0, i = 1, 2, 3, 4, 5, 6, 7, 8, k = 1, 2, 3$$

⑧第 k 个旅游团对于每个景点的开放时间、包括到达时间和离开时间、特别是小剧场的半点或整点开放的特殊要求，满足约束条件

$$\begin{cases} y_{1k} = z_{1k} = 12 \times 60, y_{2k} \geq 9 \times 60, z_{2k} \leq 16 \times 60, k = 1, 2, 3 \\ y_{ik} \geq 9 \times 60, z_{ik} \leq 17 \times 60, i = 3, 5, 6, 7, k = 1, 2, 3 \\ y_{4k} + w_{4k} = 30 \times G_k, G_k \in Z, k = 1, 2, 3 \\ y_{8k} \leq 17 \times 60, z_{8k} = 17.5 \times 60, k = 1, 2, 3 \\ 10 \leq t_{2k} \leq 30, 20 \leq t_{3k} \leq 60, t_{4k} = 30, 30 \leq t_{5k} \leq 60, k = 1, 2, 3 \\ 20 \leq t_{6k} \leq 60, 30 \leq t_{7k} \leq 60, t_{8k} \geq 30, k = 1, 2, 3 \end{cases}$$

⑨第 k 个旅行团在每个到达离开景点的时间与紧接着到达的景点的时间上的约束

$$(x_{ijk} - 1) \cdot M \leq y_{jk} - z_{ik} - \frac{d_{ij}}{V} \leq (1 - x_{ijk}) \cdot M$$

$$i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3$$

⑩任意两个旅行团在各个景点中，后一个旅行团需要等待前一个旅行团游览完毕，不能有时间冲突。如果有冲突，则需要等待，等待时间计算如下：

$$w_{ik} = z_{is} - y_{ik}, \text{if } (y_{is} < y_{ik} < z_{is}), i = 2, 3, \dots, 7, k, s = 1, 2, 3, k \neq s$$

综上分析可得问题 3 的混合整数规划模型Ⅲ如下：

$$\max T = T_1 + T_2 + T_3$$

$$\begin{cases}
 T_k = \sum_{i=1}^8 t_{ik}, k = 1, 2, 3 \\
 \sum_{j=2}^8 x_{1jk} = 1, \sum_{i=2}^8 x_{i1k} = 0, k = 1, 2, 3 \\
 \sum_{i=1}^7 x_{i8k} = 1, \sum_{j=1}^7 x_{8jk} = 0, k = 1, 2, 3 \\
 \sum_{i=1, i \neq j}^8 x_{ijk} = 1, j = 2, 3, \dots, 7, k = 1, 2, 3 \\
 \sum_{j=1, j \neq i}^8 x_{ijk} = 1, i = 2, 3, \dots, 7, k = 1, 2, 3 \\
 x_{ijk} \in \{0, 1\}, i = 1, 2, \dots, 8; j = 1, 2, \dots, 8, i \neq j, k = 1, 2, 3 \\
 y_{ik} + w_{ik} + t_{ik} = z_{ik}, w_{ik} \geq 0, i = 1, 2, 3, 4, 5, 6, 7, 8, k = 1, 2, 3 \\
 s.t. \quad u_{ik} - u_{jk} + 8x_{ijk} \leq 7, u_{ik}, u_{jk} \geq 0 \\
 (i = 1, 2, \dots, 8, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3) \\
 y_{1k} = z_{1k} = 12 \times 60, y_{2k} \geq 9 \times 60, z_{2k} \leq 16 \times 60, k = 1, 2, 3 \\
 y_{ik} \geq 9 \times 60, z_{ik} \leq 17 \times 60, i = 3, 5, 6, 7, k = 1, 2, 3 \\
 y_{4k} + w_{4k} = 30 \times G_k, G_k \in Z, k = 1, 2, 3 \\
 y_{8k} \leq 17 \times 60, z_{8k} = 17.5 \times 60, k = 1, 2, 3 \\
 10 \leq t_{2k} \leq 30, 20 \leq t_{3k} \leq 60, t_{4k} = 30, 30 \leq t_{5k} \leq 60, k = 1, 2, 3 \\
 20 \leq t_{6k} \leq 60, 30 \leq t_{7k} \leq 60, t_{8k} \geq 30, k = 1, 2, 3 \\
 (x_{ijk} - 1) \cdot M \leq y_{jk} - z_{ik} - \frac{d_{ij}}{V} \leq (1 - x_{ijk}) \cdot M \\
 (i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3, \\
 w_{ik} = z_{is} - y_{ik}, \text{if } (y_{is} < y_{ik} < z_{is}) \\
 i = 2, 3, \dots, 7, k, s = 1, 2, 3, k \neq s
 \end{cases}$$

### 5.3.3 问题三模型的求解

问题 3 变得复杂起来，复杂程度相比问题 2 提高 3 倍不止，因为还需要对旅行团之间游览景点的时间进行冲突检测。由于 TSP 问题的复杂性，所以本问题我们采用穷举搜索的方法进行求解。

(1) 求解算法步骤

【Step 1】利用问题 2 中求游览时间最长的 lingo 程序，通过增加限制语句（程序详见附录 3）求出出发点是景点 1 到其他剩余景点的所有 6 条可能的备选最优路线。具体结果为：

表 3:六种备选最优路线信息表

起点	游览时间 (分钟)	等待时间 (分钟)	路程 (米)	完整路线
1-2	268.8	0	2040	1-2-6-4-3-5-7-8
1-3	270	0	2000	1-3-5-4-6-2-7-8
1-4	251.7	23.7	1820	1-4-6-2-3-5-7-8
1-5	261.9	0	2270	1-5-3-4-6-2-7-8
1-6	267.9	0	2070	1-6-4-2-3-5-7-8
1-7	256.8	0	2440	1-7-5-3-4-6-2-8

【Step 2】对可能出现的走同一线路的情况进行分析，由于这种情况下必须等待，计算等待时间，看是否超过 6 分钟以上。计算表明均超过了 6 分钟，故排除这种选择。

【Step 3】选取游览时间最大的 3 条路线。

【Step 4】计算 3 个旅游团分别在这 3 条路线上各个景点的到达和离开时间。

【Step 5】利用编写的 c 语言验证程序（程序详见附录 5）检测各个旅游团在各个景点的到达和离开时间有无冲突。

【Step 6】若有冲突，则计算到达和离开时间冲突时的差值作为等待时间。用 3 个旅游团的总游览时间-等待时间=等待时间，得到实际游览时间。

【Step 7】若无冲突，则等待时间=0，3 个旅游团的总游览时间+等待时间=实际游览时间。

【Step 8】返回【Step 2】，选择次优的路线继续计算。直到选择方法全部选完。

【Step 9】比较得出的所有实际游览时间，得出游览时间最大的 3 条线路。

(2) 计算结果：详细的计算比较结果可以见附录四。

表 4: 问题三结果表

	第一旅游团			第二旅游团			第三旅游团		
	达到 时间点	游览时间(停留 时间,单位分钟)	离开 时间点	达到 时间点	游览时间(停留 时间,单位分 钟)	离开 时间点	达到 时间点	游览时间(停留 时间,单位分 钟)	离开 时间点
景石	12:00	0	12:00	12:00	0	12:00	12:00	0	12:00
游客服务中心	12:10:48	58.2	13:09	12:18	30	12:48	12:09	14.5	12:23:24
阳光草坪	13:16:12	60	14:16:12	12:54:36	50.1	13:45	12:31:48	20	12:52:12
森林小剧场	14:30	30	15:00	14:00	30	14:30	13:00	30	13:30
儿童科普体验区	15:07:48	20	15:28:12	14:37:48	60	15:37:48	13:45:36	20	14:05:24
儿童戏水场	15:36:36	10	15:46:48	15:46:48	13.5	16:00	14:12	30	14:42
湿地博物馆	15:52:48	60	16:52:48	16:06	30	16:36	14:49:48	30	15:19:48
湿地商业街	16:58:12	61.8	17:30	16:42	48.3	17:30	15:25:48	124.3	17:30
总步行时间	60 分钟			68 分钟			61 分钟		
总游览时间	270 分钟			261.9 分钟			268.8 分钟		
总等待时间	0 分钟			0 分钟			0 分钟		

5.3.4 问题三结果的分析与验证

对于上述结果的准确性我们通过验算求得 3 个旅游团的各自的总步行时长分别为 60, 68, 61 分钟，总等待时长分别为 0, 0, 0 分钟，总游览时长分别为 270, 261.9, 268.8 分钟。而对于在各个景点的时间 3 个旅行团是否相交的问题上，我们采用了 C 言程序来

验证是否相交，我们发现没有相交的时间节点，即结果准确无误。

## 5.4 问题四模型的建立与求解

### 5.4.1 问题四的分析

在问题三建立的数学模型的基础上，只要增加对旅游团速度的约束，就可以完成整个的约束条件。而问题四中的结果牵扯了游览总时间长的和总的等待时间短的路线，则在双目标函数的条件下，就要对 3 个旅游团进行分别求和运算，从而得出总的和，即双目标函数的确立，从而建立双目标混合整数规划模型来解答相关问题。

### 5.4.2 问题四模型的建立

#### (1) 符号说明

$V_{ijk}$  ——第  $k$  个旅行团从景点  $i$  到景点  $j$  的步行速度 ( $i, j = 0, 1, \dots, 7, k = 1, 2, 3$ )

$T_k, W_k, B_k$  ——分别为 3 个旅行团的游览总时间，等待总时间，步行总时间， $k = 1, 2, 3$

$t_{ik}$  —— $k$  号旅游团游览景点  $i$  的时间， $i = 1, 2, \dots, 8$ ，单位：分钟

#### (2) 目标函数

要求 3 个旅游团游览的总时间长且总等待时间短，故为双目标函数：

$$\begin{aligned} \max T &= T_1 + T_2 + T_3 \\ \min W &= W_1 + W_2 + W_3 \end{aligned}$$

#### (3) 约束条件的构造

本问题与问题 3 的区别不大，所以与问题 3 相同的约束这里不再重复叙述。

① 3 个旅游团游览的时间等于其在各个景点游览的时间，3 个旅游团等待的时间等于其在各个景点等待的时间，3 个旅游团行走的总时间=其整个旅游所花的总时间-其游览的总时间-其等时间，即

$$T_k = \sum_{i=1}^8 t_{ik}, W_k = \sum_{i=1}^8 w_{ik}, B_k = 330 - \sum_{i=1}^8 (t_{ik} + w_{ik}), k = 1, 2, 3$$

②第  $k$  个旅行团在每个到达景点的时间与紧接着到达的景点的时间上的约束，即

$$(x_{ijk} - 1) \cdot M \leq y_{jk} - z_{ik} - \frac{d_{ij}}{V_{ijk}} \leq (1 - x_{ijk}) \cdot M$$

$$i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, \dots, n$$

③每个旅游团在去各个景点时的速度限制，即

$$\frac{1000}{60} \leq V_{ijk} \leq \frac{3000}{60}, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3$$

④每个旅游团所行走的总路程=平均速度×每个旅游团各自行走的总时间的约束，即

$$\sum_{i=1}^8 \sum_{j=1}^8 d_{ij} \cdot x_{ijk} = \frac{2000}{60} B_k, k = 1, 2, 3$$

综上所述可得问题 4 的双目标混合整数规划模型 IV 如下：

$$\max T = T_1 + T_2 + T_3$$

$$\min W = W_1 + W_2 + W_3$$

$$\begin{cases}
 T_k = \sum_{i=1}^8 t_{ik}, W_k = \sum_{i=1}^8 w_{ik}, B_k = 330 - \sum_{i=1}^8 (t_{ik} + w_{ik}), k = 1, 2, 3 \\
 \sum_{j=2}^8 x_{1jk} = 1, \sum_{i=2}^8 x_{i1k} = 0, k = 1, 2, 3 \\
 \sum_{i=1}^7 x_{i8k} = 1, \sum_{j=1}^7 x_{8jk} = 0, k = 1, 2, 3 \\
 \sum_{i=1, i \neq j}^8 x_{ijk} = 1, j = 2, 3, \dots, 7, k = 1, 2, 3 \\
 \sum_{j=1, j \neq i}^8 x_{ijk} = 1, i = 2, 3, \dots, 7, k = 1, 2, 3 \\
 x_{ijk} \in \{0, 1\}, i = 1, 2, \dots, 8; j = 1, 2, \dots, 8, i \neq j, k = 1, 2, 3 \\
 u_{ik} - u_{jk} + 8 * x_{ijk} \leq 7, u_{ik}, u_{jk} \geq 0, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3 \\
 \text{s.t. } y_{ik} + w_{ik} + t_{ik} = z_{ik}, w_{ik} \geq 0, i = 1, 2, 3, 4, 5, 6, 7, 8, k = 1, 2, 3 \\
 y_{1k} = z_{1k} = 12 \times 60, y_{1k} \geq 9 \times 60, z_{1k} \leq 16 \times 60, k = 1, 2, 3 \\
 y_{ik} \geq 9 \times 60, z_{ik} \leq 17 \times 60, i = 3, 5, 6, 7, k = 1, 2, 3 \\
 y_{4k} + w_{4k} = 30 \times G_k, G_k \in Z, k = 1, 2, 3 \\
 y_{8k} \leq 17 \times 60, z_{8k} = 17.5 \times 60, k = 1, 2, 3 \\
 10 \leq t_{2k} \leq 30, 20 \leq t_{3k} \leq 60, t_{4k} = 30, 30 \leq t_{5k} \leq 60, k = 1, 2, 3 \\
 20 \leq t_{6k} \leq 60, 30 \leq t_{7k} \leq 60, t_{8k} \geq 30, k = 1, 2, 3 \\
 (x_{ijk} - 1) \cdot M \leq y_{jk} - z_{ik} - \frac{d_{ij}}{V_{ijk}} \leq (1 - x_{ijk}) \cdot M, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3 \\
 \frac{1000}{60} \leq V_{ijk} \leq \frac{3000}{60}, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, 3 \\
 \sum_{i=1}^8 \sum_{j=1}^8 d_{ij} \cdot x_{ijk} = \frac{2000}{60} B_k, k = 1, 2, 3 \\
 w_{ik} = z_{is} - y_{ik}, \text{if } (y_{is} < y_{ik} < z_{is}), i = 2, 3, \dots, 7, k, s = 1, 2, 3, k \neq s
 \end{cases}$$

### 5.4.3 问题四模型的求解

问题 4 在问题 3 的基础上增加了步行速度的调节，因此理论上的结果必然比问题 3 的结果要优。我们仍然采用穷举搜索的方法求解。

#### (1) 求解算法步骤

【Step 1】首先将双目标转换为单目标，这里我们采用

$$\max Q = T - W$$

【Step 2】利用问题 2 中求游览时间最长的 lingo 程序求出出发点是景点 1 到其他剩余景点的路线。详细线路见表 3：六种备选最优路线信息表。

【Step 3】对可能出现的不同旅行团走同一线路的情况进行分析，由于步行速度可

以调整，要在问题 3 的结果上继续优化，这种情况是极有可能的选择。

【Step 4】选取游览时间最大的 3 条路线，注意这里要包括不同的旅行团走同一线路的选择。

【Step 5】计算 3 个旅游团分别在这 3 条路线上各个景点的到达和离开时间，包括可能的等待时间。

【Step 6】利用编写的 c 语言验证程序检测各个旅游团在各个景点的到达和离开时间有无冲突。

【Step 7】若有冲突，则计算到达和离开时间冲突时的差值。用 3 个旅游团的总游览时间-差值得到实际游览时间。筛选出影响最大的那条路线，并删除，且返回 step4，选择次优的路线继续计算。

【Step 8】若无冲突，则 3 个旅游团的总游览时间=实际游览时间，转【Step 6】。

【Step 9】比较所有得出的实际游览时间，选出游览时间最大的那 3 条线路。不同组合下的具体计算数据主要结果为：

情况种类	组合情况一			组合情况二			组合情况三			组合情况四		
线路最长游览时间(分钟)	270	270	268.8	270	268.8	268.8	270	268.8	267.9	270	267.9	267.9
等待时间(分钟)	5.6			0			3			1		
相比问题 3 提升时间(分钟)	8.1			6.9			6			5.1		
时间差	2.5			6.9			3			4.1		

(2) 计算结果：通过比较，得到最优结果如下表（详细计算数据比较结果见附录七）。

表 5: 第四问结果表格

	第一旅游团				第二旅游团				第三旅游团			
	达到时间点	游览时间(单位分钟)	离开时间点	达到下一个景点的步行速度 $km/h$	达到时间点	游览时间(停留时间, 单位分钟)	离开时间点	达到下一个景点的步行速度 $km/h$	达到时间点	游览时间(停留时间, 单位分钟)	离开时间点	达到下一个景点的步行速度 $km/h$
景石	12:00	0	12:00	2	12:00	0	12:00	3	12:00	0	12:00	1
游客服务中心 2	15:46	11	15:57	2	12:06	10	12:16	2	12:18	30	12:48	2
阳光草坪 3	12:10	60	13:10	2	13:45	25	14:10	2	14:13	40	14:52	3
森林小剧场 4	14:30	30	15:00	2	13:00	30	13:30	2	13:30	30	14:00	2.336
儿童科普区 5	13:18	58.2	14:16	2	14:17	30	14:47	2	14:57	30	15:27	3
儿童戏水场 6	15:07	30	15:37	2	12:24	27.5	12:52	2	12:56	25.5	13:22	2
湿地博物馆 7	16:03	30	16:33	2	14:55	30	15:25	1.31	15:33	30	16:03	3
湿地商业街 8	16:39	50.8	17:30	0	15:33	116.3	17:30	0	16:06	83.3007	17:30	0
总步行时间	60 分钟				61.20 分钟				61.20 分钟			
总游览时间	270 分钟				268.8 分钟				268.8 分钟			
总等待时间	0 分钟				0 分钟				0 分钟			
平均速度	2 $km/h$				2 $km/h$				2 $km/h$			

#### 5.4.4 问题四结果的分析与验证

对于上述结果的准确性，我们通过验算求得 3 个旅游团的各自的总步行时长分别为 60, 61.2, 61.2 分钟，总等待时长分别为 0, 0, 0 分钟，总游览时长分别为 270, 268.8, 268.8 分钟，平均速度为 2km/h。除此之外我们又算出了 3 个旅行团到达并离开各个景点的准确时间和打到下一个景点时步行的速度。而对于在各个景点的时间 3 个旅行团是否相交的问题上，我们采用了 C 言程序来验证是否相交（详情见附录五），我们发现没有相交的时间节点，即结果准确无误。

### 5.5 问题五模型的建立与分析

#### 5.5.1 问题五的分析

在考虑了两个不确定因素后，基于问题四的解题思路，我们只需要在每个景点的到达时间与离开时间的约束中加入了不确定因素  $\varepsilon_{ik}$ ，再在问题四中相关时间的约束上添加另一不确定因素  $Y_k$ ，即旅游团的出发时间。之后再通过问题四所建立的模型作参考，构造双目标函数，建立本题中的混合整数规划的相关数学模型，从而得出结果。

#### 5.5.2 问题五模型的建立：

##### (1) 符号说明

$n$  —— 常量，旅行团总个数

$Y_k$  —— 常量，第  $k$  个旅行团从景石出发的时间， $Y_k \geq 12 \times 60, k = 1, 2, 3, \dots, n$

$\varepsilon_{ik}$  —— 不确定因素引起的第  $k$  个旅行团在景点  $i$  的随机等待时间，

$i = 1, 2, \dots, 6, k = 1, 2, \dots, n$ ，单位：分钟

##### (2) 目标函数的建立

同问题 4，要求  $k$  个旅游团游览的总时间长且总等待时间短，所以其目标函数为：

$$\begin{aligned} \max T &= T_1 + T_2 + \dots + T_n \\ \min W &= W_1 + W_2 + \dots + W_n \end{aligned}$$

##### (3) 约束条件的构造

在第 4 问的基础上修改或加入以下约束条件：在每个景点的到达时间与离开时间的约束中加入了不确定因素  $\varepsilon_{ik}$ ；每个旅行团出发的时间变动；在不确定因素的影响下到达湿地商业的时间延后。即

$$\begin{cases} y_{ik} + w_{ik} + t_{ik} + \varepsilon_{ik} = z_{ik}, w_{ik} \geq 0, i = 1, 2, 3, 4, 5, 6, 7, 8, k = 1, 2, \dots, n \\ y_{1k} = z_{1k} = Y_k, y_{2k} \geq 9 \times 60, z_{2k} \leq 16 \times 60, k = 1, 2, \dots, n \\ y_{ik} \geq 9 \times 60, z_{ik} \leq 17 \times 60, i = 3, 5, 6, 7, k = 1, 2, \dots, n \\ y_{4k} + w_{4k} = 30 \times G_k, G_k \in Z, k = 1, 2, \dots, n \\ y_{8k} \leq 17 \times 60 + Y_k - 12 \times 60, z_{8k} = 17.5 \times 60, k = 1, 2, \dots, n \\ 10 \leq t_{2k} \leq 30, 20 \leq t_{3k} \leq 60, t_{4k} = 30, 30 \leq t_{5k} \leq 60, k = 1, 2, \dots, n \\ 20 \leq t_{6k} \leq 60, 30 \leq t_{7k} \leq 60, t_{8k} \geq 30, k = 1, 2, \dots, n \end{cases}$$

综上所述可得问题 5 的双目标混合整数规划模型 V 如下：

$$\begin{aligned}
 \max \quad & T = \sum_{i=1}^n T_i \\
 \min \quad & W = \sum_{i=1}^n W_i \\
 \text{s.t.} \quad & \left\{ \begin{aligned}
 & T_k = \sum_{i=1}^8 t_{ik}, W_k = \sum_{i=1}^8 w_{ik}, B_k = 330 - \sum_{i=1}^8 (t_{ik} + w_{ik}), k = 1, 2, \dots, n \\
 & \sum_{j=2}^8 x_{1jk} = 1, \sum_{i=2}^8 x_{i1k} = 0, k = 1, 2, \dots, n \\
 & \sum_{i=1}^7 x_{i8k} = 1, \sum_{j=1}^7 x_{8jk} = 0, k = 1, 2, \dots, n \\
 & \sum_{i=1, i \neq j}^8 x_{ijk} = 1, j = 2, 3, \dots, 7, k = 1, 2, \dots, n \\
 & \sum_{j=1, j \neq i}^8 x_{ijk} = 1, i = 2, 3, \dots, 7, k = 1, 2, \dots, n \\
 & x_{ijk} \in \{0, 1\}, i = 1, 2, \dots, 8; j = 1, 2, \dots, 8, i \neq j, k = 1, 2, \dots, n \\
 & u_{ik} - u_{jk} + 8 \cdot x_{ijk} \leq 7, u_{ik}, u_{jk} \geq 0, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, \dots, n \\
 & y_{ik} + w_{ik} + t_{ik} + \varepsilon_{ik} = z_{ik}, w_{ik} \geq 0, i = 1, 2, 3, 4, 5, 6, 7, 8, k = 1, 2, \dots, n \\
 & y_{1k} = z_{1k} = Y_k, y_{1k} \geq 9 \times 60, z_{1k} \leq 16 \times 60, k = 1, 2, \dots, n \\
 & y_{ik} \geq 9 \times 60, z_{ik} \leq 17 \times 60, i = 3, 5, 6, 7, k = 1, 2, \dots, n \\
 & y_{4k} + w_{4k} = 30 \times G_k, G_k \in Z, k = 1, 2, \dots, n \\
 & y_{8k} \leq 17 \times 60 + Y_k - 12 \times 60, z_{8k} = 17.5 \times 60, k = 1, 2, \dots, n \\
 & 10 \leq t_{2k} \leq 30, 20 \leq t_{3k} \leq 60, t_{4k} = 30, 30 \leq t_{5k} \leq 60, k = 1, 2, \dots, n \\
 & 20 \leq t_{6k} \leq 60, 30 \leq t_{7k} \leq 60, t_{8k} \geq 30, k = 1, 2, \dots, n \\
 & (x_{ijk} - 1) \cdot M \leq y_{jk} - z_{ik} - \frac{d_{ij}}{V_{ijk}} \leq (1 - x_{ijk}) \cdot M, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, \dots, n \\
 & \frac{1000}{60} \leq V_{ijk} \leq \frac{3000}{60}, i = 1, 2, \dots, 7, j = 2, 3, \dots, 8, i \neq j, k = 1, 2, \dots, n \\
 & \sum_{i=1}^8 \sum_{j=1}^8 d_{ij} \cdot x_{ijk} = \frac{2000}{60} B_k, k = 1, 2, \dots, n \\
 & w_{ik} = z_{is} - y_{ik}, \text{if } (y_{is} < y_{ik} < z_{is}), i = 2, 3, \dots, 7, k, s = k = 1, 2, \dots, n, k \neq s
 \end{aligned} \right.
 \end{aligned}$$

### 5.5.3 问题五模型的求解

#### (1) 求解算法步骤

【Step 1】利用问题 2 中求游览时间最长的 lingo 程序（改时间约束）求出出发点 是景点 1 到其他剩余景点的路线。

【Step 2】选取游览时间最大的 n 条路线。

【Step 3】计算个 k 旅游团分别在这 n 条路线上各个景点的到达和离开时间。

【Step 4】从  $n \times n$  条路线中任意选  $k-n$  条路线。计算  $k-n$  个旅游团在这  $k-n$  条路线中到达和离开的时间且  $X_{k-n}$  作为延续后的旅行团。

【Step 5】利用编写的 c 语言验证程序检测各个旅游团在各个景点的到达和离开时间有无冲突。

【Step 6】若有冲突，则计算到达和离开时间冲突时的差值作为等待时间。用  $k$  个旅游团的总游览时间-等待时间=等待时间得到实际游览时间。若无冲突，则等待时间=0， $k$  个旅游团的总游览时间+等待时间=实际游览时间。接着返回【Step 2】，计算下一线路。

【Step 7】比较各个实际游览总时间-等待时间的差值，取差值较大的那  $n$  条路线。

(2) 计算结果：选定 4 个旅行团，设定了一些数据，进行了试算，得到下表结果。

表 6: 问题五的试算结果参考

	第一旅游团				第二旅游团			
	达到时间	游览时间	离开时间	步行速度	达到时间	游览时间	离开时间	步行速度
景石 1	8:45	0	525.75	2	12:00	0	12:00	2
游客服务中心 2	10:08	30	10:38	2	15:17	10	15:26	2
阳光草坪 3	10:49	20	11:09	2	13:25	20.1	13:45	2
森林小剧场 4	9:30	30	10:00	2	14:00	30	14:30	2
儿童科普区 5	11:16	30	11:46	2	12:30	60	13:00	2
儿童戏水场 6	9:00	22.1	9:22	2	14:37	30	15:07	2
湿地博物馆 7	11:54	60	12:54	2	15:32	30	16:02	2
湿地商业街 8	12:59	270	17:30	0	16:08	81.8	17:30	0
总步行时间	97.35 分钟				68 分钟			
总游览时间	462.15 分钟				261.9 分钟			
总等待时间	0 分钟				0 分钟			
平均速度	2km/h				2km/h			
	第三旅游团				第四旅游团			
景石 1	12:00	0	12:00	3	13:02	0	13:02	1
游客服务中心 2	12:06	10	12:16	2	16:38	10	16:48	2
阳光草坪 3	13:45	25	14:10	2	15:24	20	15:44	3
森林小剧场 4	13:00	30	13:30	2	16:00	30	16:30	2.34
儿童科普区 5	14:17	30	14:47	2	13:20	30	13:50	3
儿童戏水场 6	12:24	27.5	12:52	2	14:41	20	15:01	2
湿地博物馆 7	14:55	30	15:25	1.3	13:57	30	14:27	3
湿地商业街 8	15:33	116	17:30	0	16:59	30.2	17:30	0
总步行时间	61.20 分钟				62.1 分钟			
总游览时间	268.8 分钟				170.2 分钟			
总等待时间	0 分钟				0 分钟			
平均速度	2km/h				2km/h			

#### 5.5.4 问题五结果的分析与验证

对于上述结果的准确性，我们通过验算求得  $k$  个旅游团各自的总步行时间为 97.35, ..., 62.1 分钟；总的游览时间为 462.15, ..., 170.2 分钟；总的等待时间为 0, ..., 0 分钟；平均速度为 2, ..., 2km/h。除此之外，我们又算出了  $k$  个旅行团到达并离开各个景点的准确时间和打到下一个景点时步行的速度。而对于在各个景点的时间  $k$  个旅行团是否相交的问题上，我们采用了 C 言程序来验证是否相交，我们发现没有相交的时间节点，即结果准确无误。

## 六、评价与推广

### 6.1 模型的评价：

#### 1. 模型的优点：

(1) 在数据处理方面，我们对结果进行了规范，使结果在具有可用性的同时，简洁明了。

(2) 模型在图像的处理和显示上，我们采用 visio 作图，使结果路线更加清晰，条理和直观。

(3) 我们利用穷举法从数据中筛选出旅游团游览时间长且不等待或等待时间短的合理数据，从而分析给出科学直观的结果。

(4) 在模型的建立上，我们引入了决策变量  $(0, 1)$  用双目标混合整数规划模型，同时简化了无用的公式接近了数学建模中“用简单的方法解决最难问题”的思想。

#### 2. 模型的缺点：

(1) 在问题 3 和问题 4 中我们仅改变时间约束，多次应用程序运算结果，这样虽然结果准确但过程繁琐。

(2) 在结果数据的处理中，我们采用的是人工辅助计算机处理，虽然提高了准确性，但仍然可能出现人工错误。

### 6.2 模型的推广：

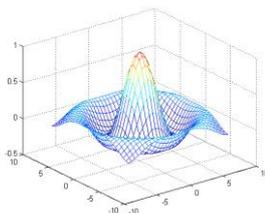
(1) 根据题目中提供的各景点开放时间数据，和各景点之间的最短距离数据，建立了长游览时间和短等待时间模型，采用穷举法确定了多个旅游团的最佳路线，对广大游客对路线的安排方面有一定的参考价值。

(2) 模型中分析问题，解决问题的一些特别的方法，对其他数学问题及一般模型仍然适用。

## 七、参考文献

- [1] 谢金星. 优化建模与 Lingo/Lingo 软件. 北京：清华大学出版社，2005：P301
- [2] 姜启源等. 数学模型（第四版）. 北京：高等教育出版社，2003：P21
- [3] 司守奎，孙玺箴. 数学建模算法与应用. 北京：国防工业出版社，2012：P57
- [4] 袁新生等. Lingo/Excel 在数学建模中的应用. 北京：科学出版社，2007：P75

# 五一数学建模竞赛



## 题目：徐州潘安湖风景区游览路线设计

**关键词：**旅行商问题；0-1 规划；多目标规划；模拟退火算法；新起点策略

**摘要：**本文研究的风景区游览路线设计问题属于起终点确定的旅行商问题（TSP）。全文以步行距离最短、游览总时间最长、总等待时间最短等为目标函数，分别建立 0-1 规划模型、双目标规划模型、多目标规划模型，逐步完善改进模型，采用蚁群算法和模拟退火算法，利用 MATLAB 进行求解。结合实际情况，考虑现实生活中的不确定因素，为旅游团设计出一条最优游览路线。

问题一中，要求为游客设计距离最短的路线，遍历每个景点，是起终点确定的旅行商问题（TSP）。建立了路线距离最短的改进 0-1 规划模型，结合蚁群算法，运用 MATLAB 编程求解，得到游览最短距离为 1820m，游览线路为景石→③→⑤→①→②→④→⑥→⑦。

问题二中，在不考虑等待时间的情况下，兼顾各景点开放时间，为游客设计一条能游览全部景点且游览时间最长的线路，建立了单游客路径优化模型，利用模拟退火算法，MATLAB 编程求解，得到游览线路为景石→②→④→③→⑤→①→⑥→⑦（2000m, 270min），并给出了 3 种不同的在各景点游览时长方案。

问题三中，针对 3 个旅游团，考虑各旅游团之间的相互影响，以 3 个旅游团总游览时间最长、游览时间方差最小的目标，建立双目标规划模型，利用模拟退火算法，MATLAB 编程求解，3 个旅游团分别的游览路线：景石→②→④→③→⑤→①→⑥→⑦（2000m, 270min），景石→⑤→③→①→②→④→⑥→⑦（2070m, 267.9min），景石→①→④→②→③→⑤→⑥→⑦（2185m, 264.45min），3 个旅游团总游览时间为：802.35min，并给出了 3 个旅游团各自的游览时长方案。

问题四中，针对 3 个旅游团，在步行速度可调节基础上，要求游览总时间最长、游览时间方差最小和总的等待时间最短，在第三问基础上对线路进一步优化设计。建立了多目标规划模型，改进的模拟退火算法，MATLAB 编程求解，3 个旅游团其中 1 个选择：景石→②→④→③→⑤→①→⑥→⑦（2000m, 270min），另外 2 个选择：景石→①→⑤→③→②→④→⑥→⑦（2040m, 268.8min），3 个旅游团总游览时间为：807.6min，并给出了 3 个旅游团各自的游览时长方案。

问题五中，考虑了现实中出发时间、等待时间等不确定因素的影响，提出针对随机情况的新起点策略，建立了不确定多目标约束模型，可以应用到较复杂问题的处理上，游览线路为景石→③→⑤→①→②→④→⑥→⑦。

本文的特色：1) 假设条件考虑周密，提出了利用统计数据的新起点策略；2) 对合理制定游览路线问题进行了从局部到整体、从单一到具体全面深刻的探讨；3) 全面、科学地建立了路线选择模型，严谨地研究了不确定因素对游览路线选择的影响，为合理化选择游览路线提供了科学的依据，有一定的参考价值。

# 目 录

一、问题提出.....	1
1.1 问题背景.....	1
1.2 问题重述.....	1
二、基本假设.....	2
三、符号说明.....	2
四、问题分析.....	3
五、模型的建立与求解.....	4
5.1 问题一模型建立与求解.....	4
5.1.1 问题一的分析.....	4
5.1.2 问题一模型的建立.....	5
5.1.3 问题一模型的求解.....	6
5.2 问题二模型建立与求解.....	6
5.2.1 问题二的分析.....	6
5.2.2 问题二模型的建立.....	7
5.2.3 问题二模型的求解.....	8
5.3 问题三模型建立与求解.....	11
5.3.1 问题三的分析.....	11
5.3.2 问题三模型的建立.....	11
5.3.3 问题三模型的求解.....	13
5.3.4 问题三结果的分析.....	15
5.4 问题四模型建立与求解.....	15
5.4.1 问题四的分析.....	15
5.4.2 问题四模型的建立.....	15
5.4.3 问题四模型的求解.....	17
5.4.4 问题四结果的分析.....	19
5.5 问题五模型建立与求解.....	19
5.5.1 问题五的分析.....	19
5.5.2 问题五模型的建立.....	19
5.5.3 问题五模型的求解.....	21
六、模型的评价与推广.....	22

6.1 模型的评价 .....	22
6.2 模型的推广 .....	22
七、参考文献 .....	22
八、附录 .....	1
8.1 附录清单 .....	1
8.2 附录正文 .....	1

# 一、问题提出

## 1.1 问题背景

习近平总书记在党的十九大报告中明确提出“支持资源型地区经济转型发展”。2017年12月12日至13日，习近平总书记来到徐州调研，这是十九大后习总书记的首次调研。习近平总书记在徐州调研贾汪区潘安采煤塌陷区整治工程时强调，塌陷区要坚持走符合国情的转型发展之路，打造绿水青山，并把绿水青山变成金山银山。

## 1.2 问题重述

本题选取潘安湖景区的部分景点(如图1所示)，完成徐州潘安湖风景区游览路线设计问题。从景石出发，步行游览以下景点：①游客服务中心，②阳光草坪，③森林小剧场，④儿童科普体验区，⑤儿童戏水场，⑥湿地博物馆，⑦湿地商业街。

**问题1：**建立数学模型，找出从景石出发，到达⑦湿地商业街，并且经过①—⑥所有景点至少1次的距离最短的路线，计算该路线的长度，并将相关结果填入表格（注：在每个景点不用停留）。

**问题2：**如果某游客12:00从景石出发，要求他17:00前到达湿地商业街，17:30离开湿地商业街(注：根据表2的要求在湿地商业街游览时间至少为30分钟)。建立数学模型，为该游客设计一条能游览完全部景点(景点①—⑦)且游览总时间最长的游览路线（假设在各个景点没有等待时间）。

**问题3：**如果有3个旅游团，12:00同时从景石出发，要求3个旅游团17:00前到达湿地商业街，17:30离开湿地商业街，并且每个景点（湿地商业街除外）同时只能容纳1个旅游团游览，按照时间顺序后到达的旅游团，需要等待先到达的旅游团游览结束之后才能开始游览。建立数学模型，为3个旅游团分别设计一条能游览完全部7个景点且游览总时间最长的游览路线。

**问题4：**假设3个旅游团的步行速度可以在 $1\text{km}/\text{h}$ 到 $3\text{km}/\text{h}$ 之间调节，但是总的平均步行速度不能超过 $2\text{km}/\text{h}$ ，3个旅游团12:00同时从景石出发，要求3个旅游团17:00前到达湿地商业街，17:30离开湿地商业街，并且每个景点（湿地商业街除外）同时只能容纳1个旅游团游览，按照时间顺序后到达的旅游团，需要等待先到达的旅游团游览结束之后才能开始游览。建立数学模型，为3个旅游团分别设计一条能游览完全部7个景点且游览总时间长，总的等待时间短的游览路线。

**问题5：**在现实中，考虑如下两个不确定性因素：

(1) 不同旅游团从景石出发的时间具有不确定性，例如，多个旅游团在不同的时间从景石出发开始游览，在此情况下到达湿地商业街的时间可以顺延。

(2) 每个景点的等待时间也存在不确定性因素，例如，旅游设施短时间的维护和清理，或者受到散客客流的影响。

考虑上述两个不确定性因素，其它条件与问题4相同，建立数学模型，为多个旅游团分别设计一条能游览完全部7个景点且游览总时间长，总的等待时间短的游览路线。



图 1 各景点位置示意图

## 二、基本假设

假设 1: 任意两个景点之间的最短步行距离如题目给出。

假设 2: 第二问、第三问假设步行速度  $V = 2\text{km}/\text{h}$ 。

假设 3: 游客在景区停留的时间由“景点之间的步行时间”、“景点游览时间（即在景区内游玩的时间）”和“在景区外的等待时间”三部分组成，其他时间忽略不计，游览时间必须符合所给的要求。

假设 4: 问题一中，游客在每个景点不用停留。

假设 5: 问题二中，游客在各个景点开放时间内没有等待时间，即游客到达每个景点均无需排队，在开放时间内可直接进入。

假设 6: 游客在游览过程中，每个景点仅经过一次。

假设 7: 为保证游览路径尽可能优化，如果等待时间超过两个景点间最长步行时间，则不等待，直接选择其他景点游览。

假设 8: 游客游览时间、步行时间等不受天气等自然条件影响。

假设 9: 在问题五中，仅当旅游团到达某景点时，才能获得后续点的堵塞信息。

假设 10: 非正常等待（因景点容量限制产生的等待和因旅游设施维修清洁产生的等待）时间  $\mu_i$  服从  $N(\bar{\mu}_i, \sigma_i^2)$ ， $\sigma_i$  为旅游团在景点  $i$  上产生非正常等待时间的标准差。

## 三、符号说明

符号	意义	单位
$O$	景点编号集合, $O = \{0, 1, 2, 3, 4, 5, 6, 7\}$	-
$i$	景点编号, $i \in O$	-
$j$	景点编号, $j \in O$ ;	-
$k$	除起终点外的游览路径景点编号, $k = 1, 2, \dots, 6$ ;	-
$q$	除景点 3 的游览路径景点编号, $q = 0, 1, 2, 4, 5, 6, 7$ ;	-
$n$	旅游团编号, $n = 1, 2, 3$	-
$r_{ij}$	判断游客是否走从景点 $i$ 到景点 $j$ 的路径,	-

	$r_{ij} = \begin{cases} 1, & \text{游客从第}i\text{个景点到第}j\text{个景点} \\ 0, & \text{游客不从第}i\text{个景点到第}j\text{个景点} \end{cases}$ , 上标为旅游团	
$s_{ij}$	旅客从景点 <i>i</i> 到景点 <i>j</i> 的距离, 上标为旅游团	$m$
$t_i$	游客在景点 <i>i</i> 的游览时间, 上标为旅游团	min
$t_{ij}$	游客从景点 <i>i</i> 到景点 <i>j</i> 的最短步行时间, 上标为旅游团	min
$t_{ai}$	游客到达景点 <i>i</i> 的时间, 上标为旅游团	min
$t_{li}$	游客离开景点 <i>i</i> 的时间, 上标为旅游团	min
$u_i$	判断游客在景点 <i>i</i> 是否由于未开放需要等待 $u_i = \begin{cases} 1, & \text{第}i\text{个景点由于未开放需要等待} \\ 0, & \text{第}i\text{个景点由于开放无需等待} \end{cases}$ , 上标为旅游团	-
$t_{w3}$	游客在第三个景点的等待时间, 上标为旅游团	min
$T^n$	第 <i>n</i> 个旅游团在景区总游览时间	min
$t_i^n$	第 <i>n</i> 个旅游团在景点 <i>i</i> 的游览时间	min
$t_{ai}^n$	第 <i>n</i> 个旅游团到达景点 <i>i</i> 的时间	min
$t_{li}^n$	第 <i>n</i> 个旅游团离开景点 <i>i</i> 的时间	min
$t_{wi}^n$	第 <i>n</i> 个旅游团在景点 <i>i</i> 处的等待时间	
$s_i^n$	判断第 <i>n</i> 个旅游团在景点 <i>i</i> 处是否由于景点内有旅游团需要排队等待, $s_i^n = \begin{cases} 1, & \text{第}n\text{个旅游团在景点}j\text{处需要排队等待} \\ 0, & \text{第}n\text{个旅游团在景点}j\text{处不需要排队等待} \end{cases}$	-
$u_3^n$	判断第三个景点是否由于未开放需要等待 $u_3^n = \begin{cases} 1, & \text{第三个景点未开放需要等待} \\ 0, & \text{第三个景点开放无需等待} \end{cases}$	-
$r_{ij}^n$	判断第 <i>n</i> 个旅游团是否走从景点 <i>i</i> 到景点 <i>j</i> 的路径 $r_{ij}^n = \begin{cases} 1, & \text{旅游团}n\text{从第}i\text{个景点到第}j\text{个景点} \\ 0, & \text{旅游团}n\text{从第}i\text{个景点到第}j\text{个景点} \end{cases}$	-
$s_{ij}^n$	景点 <i>i</i> 到景点 <i>j</i> 间的距离	$m$
$s_{ij\max}$	两景点间的最大步行距离	$m$
$t_{ij\max}$	两景点间的最大步行时间	$m / \text{min}$
$v_{ij}^n$	第 <i>n</i> 个旅游团从景点 <i>i</i> 到景点 <i>j</i> 的速度	$m / \text{min}$
$R$	路线解矩阵	-
$\mu_i$	旅游团在 <i>i</i> 景点非正常等待时间	min
$\sigma_i$	旅游团在 <i>i</i> 景点非正常等待时间的标准差	-
$p_i$	<i>i</i> 景点产生非正常等待的概率	-

#### 四、问题分析

为模型建立方便给各个旅游景点分别进行编号,①游客服务中心,②阳光草坪,③森林小剧场,④儿童科普体验区,⑤儿童戏水场,⑥湿地博物馆,⑦湿地商业街,则这7个景点和景石 $v_0$ 构成了一个网络图 $G$ 。

在本文中,问题一首先需要考虑的是安排这样一个游览路线计划:在每个景点

不用停留的情况下，既满足经过所有景点一次的要求，又满足所选路径距离最短要求，与 TSP 问题有类似之处，再针对题目中的所给条件进行分析。这些景点可看作该网络图的节点，这些节点由相应的步行线路相连，节点之间的边就是游览线路。将题目所给景点之间的最短步行距离数据进行整理，只留下各点之间直接相连的线路距离，如图 2 所示。

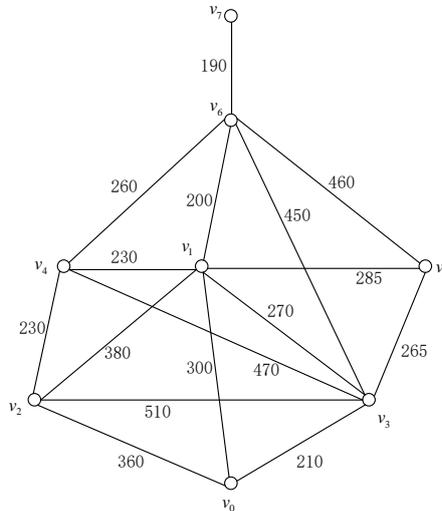


图 2 景区网络图  $G$

问题二则是为游客设计一条能游览完全部景点且游览时间最长线路，由于不用考虑等待时间，还是可以将问题转化为尽可能选择一条步行距离最短的线路，但必须考虑到③森林小剧场  $v_3$  的开放时间问题，以及各个景点的开放时间和游览时间必须满足。问题三中，需要为 3 个旅游团分别设计一条游览总时间最长的路线，由于每个景点同时只能容纳 1 个旅游团，所以在问题的基础上，考虑各个旅游团之间相互影响，尽量避免由于旅游团之间游览时间重叠而发生的等待。问题四中，又将考虑旅游团步行速度的可调节性，更灵活地调节各旅游团到达景点的时间点，因此在问题三所选线路的基础上，线路设计的可调性更大。针对问题五，考虑不管未来的堵塞点序列如何产生，先根据景区各景点路径网络图找到由起点到目的地、满足约束条件和和目标函数的路径，然后旅游团沿此路行进。当旅游团达到景点需要等待时，则需要对进行决策：1) 继续等待；2) 改走其它路径。以动态规划的最优性定理为基础，当旅游团遇到等待时，考虑堵塞时间的期望，重新寻找达到重点的最优子路径，这种策略成为新起点策略。本问中提出针对 TSP 问题中不确定性问题提出策略，参数的标定需要统计数据，由于缺乏实际数据，故需要假定某一时段参数给定用所建立的模型、算法进行求解。

## 五、模型的建立与求解

### 5.1 问题一模型建立与求解

#### 5.1.1 问题一的分析

问题一中，要求游客在步行游览时，需从景石出发，最后到达湿地商业街，途中经过游客服务中心，阳光草坪，森林小剧场，儿童科普体验区，儿童戏水场，湿地博物馆，并要求经过所有景点至少一次且距离最短。

传统图与网络模型中，旅行推销员问题<sup>[1]</sup> (Travelling salesman problem, TSP) 是给定一系列景点和每对景点之间的距离，求解访问每一座景点一次并回到起始景点的最短回路。此问题与本题类似，本题中要求游客在步行游览时，经过每一个景点

且距离最短，但步行路径不构成回路。因此，本题选用改进 0-1 规划模型进行模型建立。

为经过所有景点并保证距离最短，考虑以游览步行距离最短以目标函数，以各项约束及相关合理约束为约束条件建立模型，进行求解。

### 5.1.2 问题一模型的建立

问题一中，利用改进 TSP 问题进行模型建立，考虑非回路的 TSP 问题。模型建立中不考虑游客在景点的停留。

**Step1:** 确立目标函数。

考虑步行距离最短，建立目标函数

$$\min Z_1 = \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} s_{ij} \quad (5.1.1)$$

**Step2:** 综合考虑各个影响因素，确定约束条件

(1) 考虑在经过景点①—⑥时，每个景点仅经过 1 次，建立约束条件：

$$\sum_{i=0}^7 r_{ik} = \sum_{j=0}^7 r_{kj} = 1 \quad (5.1.2)$$

(2) 考虑旅客在游览过程中，不会走重复路径，即一条路径仅经过一次，建立约束条件：

$$r_{ij} \times r_{ji} = 0, i \in O, j \in O \quad (5.1.3)$$

(3) 考虑每个点仅经过一遍，游览过程中无回路产生，建立约束条件：

$$\sum_{i=0}^7 \sum_{j=0}^7 r_{ij} = 7 \quad (5.1.4)$$

(4) 考虑起点为景石，建立约束条件：

$$\sum_{i=0}^7 r_{i0} = 0 \quad (5.1.5)$$

(5) 考虑终点为湿地商业街，建立约束条件如下：

$$\sum_{j=0}^7 r_{7j} = 0 \quad (5.1.6)$$

**Step3:** 建立模型

综合考虑目标函数及各项约束条件，可得模型如下：

$$\min Z_1 = \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} s_{ij}$$

$$s.t. \left\{ \begin{array}{l} \sum_{i=0}^7 r_{ik} = \sum_{j=0}^7 r_{kj} = 1 \\ r_{ij} \times r_{ji} = 0, i \in O, j \in O \\ \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} = 7 \\ \sum_{i=0}^7 r_{i0} = 0 \\ \sum_{j=0}^7 r_{7j} = 0 \end{array} \right.$$

### 5.1.3 问题一模型的求解

问题一是求解最短路中已知起终点且需遍历所有中间点的问题,若采用枚举法,一共有 720 种情况,计算时长较长。而蚁群算法<sup>[2]</sup>由于其较好的鲁棒性、计算的并行性、易于其他仿生优化方法,故在本问题中采用该算法。

求解步骤:

**Step1:** 对相关参数进行初始化,并将数据读入程序;

**Step2:** 随机分别将旅游团放在在不同的结点,每个旅游团的第一个访问节点为该结点,至所有旅游团访问完所有结点;

**Step3:** 计算每个旅游团的总游览时长,记录当前最优解,同时更新连接路径访问状态  $r_{ij}^n=1$ ;

**Step4:** 判断是否达到最大迭代次数,是则终止,否则返回步骤 2;

**Step5:** 输出结果。

在本题中通过 MATLAB 程序实现过程,代码见附录 1。求解结果如下表。

表 1 第一问结果表格

出发景点	到达景点	步行距离(米)
景石	森林小剧场	210
森林小剧场	儿童戏水场	265
儿童戏水场	游客服务中心	285
游客服务中心	阳光草坪	380
阳光草坪	儿童科普体验区	230
儿童科普体验区	湿地博物馆	260
湿地博物馆	湿地商业街	190
总步行距离(最短路线距离)		1820
最短路线(请用①~⑥序号标出)		景石→③→⑤→①→②→④→⑥→⑦

由上表可知,经过各景点至少 1 次的最短路线长度为 1820m,线路方案为景石→③森林小剧场→⑤儿童戏水场→①游客服务中心→②阳光草坪→④儿童科普体验区→⑥湿地博物馆→⑦湿地商业街,该游览方案只需经过各景点 1 次即可。

## 5.2 问题二模型建立与求解

### 5.2.1 问题二的分析

问题二中,要求游客 12:00 从景石出发,17:00 到达湿地商业街,游览途中经过所有景点且游览时间最长。

由题目所给数据可知,游客在到达景点 3 时,仅在整点和半整点可以进入景点,其他时间到达均需等待到景点开放时方可进入,景点 1 的开放时间为 9:00-16:00,游览时间为 10-30min,即若要游览游客服务中心,则需在 15:50 之前到达景点 1,否则不能完成整体游览。假设 5 中,在开放时间内游客无需等待,即在不考虑到达景点 3 时的等待时间时,游客在景区内时间由步行时间和游览时间及游客在景点 3 的等待时间组成。为模型建立方便,考虑将时间化为 min,游客从景石出发时间为 0min,其余时间顺延,时间节点与分钟数转化见表。

表 2 各景点游览时间

时间 景点	游览时间 (min)	转化开放时间 (min)
游客服务中心	10-30	0-240

阳光草坪	20-60	0-300
森林小剧场	30	0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300
儿童科普体验区	30-60	0-300
儿童戏水场	20-60	0-300
湿地博物馆	30-60	0-300
湿地商业街	大于 30	0-300

考虑游览时间最长为目标，结合到达时间约束及问题一各项约束条件，建立模型。

### 5.2.2 问题二模型的建立

由于问题二中，要求游览时间最长，游客在景区内停留时间为 330min，景区停留时间由“景点之间的步行时间”、“景点游览时间（即在景点内游玩的时间）”和“在景区外的等待时间”三部分组成，其他时间忽略不计，故游览时间最长<sup>[3]</sup>，可转化景点之间的步行时间与景区外的等待时间之和最短。

**Step1:** 以景点之间的步行时间与景区外的等待时间之和最短为目标，建立目标函数

$$\min Z_2 = \sum_{i=1}^7 \sum_{j=1}^7 t_{ij} r_{ij} + t_{w3} \quad (5.2.1)$$

**Step2:** 综合考虑各个影响因素，确定约束条件

(1) 考虑游客服务中心开放时间为9:00-16:00，需在16:00前离开游客服务中心，而游览时间为10-30min，即到达游客服务中心时间需在15:50，转化为分钟即为230，建立约束条件如下：

$$t_{a1} \leq 230 \quad (5.2.2)$$

(2) 考虑游客在 12:00 点由景石出发，建立约束条件：

$$t_{l0} = 0 \quad (5.2.3)$$

(3) 考虑游客在 17:00 前要到达湿地商业街，建立约束条件：

$$t_{a7} \leq 300 \quad (5.2.4)$$

(4) 考虑游客到达景点*i*与离开景点*j*之间存在关系，建立约束条件：

$$t_{ai} = t_{lj} + r_{ij} \times t_{ij} \quad (5.2.5)$$

(5) 考虑游客到达景点与*i*离开景点*i*之间存在关系，建立约束条件：

$$t_{li} = t_{ai} + t_{wi} + t_i \quad (5.2.6)$$

(6) 考虑游客在每个景点的游览时间范围，建立约束条件：

$$\begin{aligned} 10 \leq t_1 \leq 30, \quad 20 \leq t_2 \leq 60, \quad t_3 = 30, \quad 30 \leq t_4 \leq 60, \\ 20 \leq t_5 \leq 60, \quad 30 \leq t_6 \leq 60, \quad t_7 \geq 30 \end{aligned} \quad (5.2.7)$$

(7) 考虑除景点3外，到达其他景点均不存在等待时间，建立约束条件：

$$t_{wq} = 0, q = 0, 1, 2, 4, 5, 6, 7 \quad (5.2.8)$$

(8) 考虑景点3（森林小剧场）仅在整点及半整点开放，若在其他时间到达则需要等待，建立约束条件：

$$t_{w3} = \left( \left\lceil \frac{t_{a3}}{30} \right\rceil \times 30 - t_{a3}^n \right) u_3$$

其中， $u_3$  为判断旅游团  $n$  到达森林小剧场时是否需要等待，为0-1变量，建立表达式：

$$u_3 = \left\lceil \frac{t_{a3}}{30} \right\rceil - \left\lfloor \frac{t_{a3}}{30} \right\rfloor$$

则：

$$t_{w3} = \left( \left\lceil \frac{t_{a3}}{30} \right\rceil \times 30 - t_{a3}^n \right) u_3 = \left( \left\lceil \frac{t_{a3}}{30} \right\rceil \times 30 - t_{a3} \right) \times \left( \left\lceil \frac{t_{a3}}{30} \right\rceil - \left\lfloor \frac{t_{a3}}{30} \right\rfloor \right) \quad (5.2.9)$$

**Step3:** 综合考虑目标函数及约束条件以及问题一约束条件，建立模型

$$\begin{aligned} \min Z_2 = & \sum_{i=1}^7 \sum_{j=1}^7 t_{ij} r_{ij} + t_{w3} \\ \text{s.t.} & \left\{ \begin{array}{l} \sum_{i=0}^7 r_{ik} = \sum_{j=0}^7 r_{kj} = 1 \\ r_{ij} \times r_{ji} = 0, i \in O, j \in O \\ \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} = 7, \sum_{i=0}^7 r_{i0} = 0, \sum_{j=0}^7 r_{7j} = 0 \\ t_{a1} \leq 230, t_{10} = 0, t_{a7} \leq 300 \\ t_{ai} = t_{ij} + r_{ij} \times t_{ij}, t_{li} = t_{ai} + t_{wi} + t_i \\ 10 \leq t_1 \leq 30, 20 \leq t_2 \leq 60, t_3 = 30 \\ 30 \leq t_4 \leq 60, 20 \leq t_5 \leq 60 \\ 30 \leq t_6 \leq 60, t_7 \geq 30 \\ t_{wq} = 0 \\ t_{w3} = \left( \left\lceil \frac{t_{a3}}{30} \right\rceil \times 30 - t_{a3}^n \right) u_3 \\ u_3 = \left\lceil \frac{t_{a3}}{30} \right\rceil - \left\lfloor \frac{t_{a3}}{30} \right\rfloor \end{array} \right. \end{aligned}$$

### 5.2.3 问题二模型的求解

本题中，要求游客 12:00 出发，17:00 前到达湿地商业街，求其最长游览时间。若要利用排列组合的方法将所有的可行解计算出来，并逐一比较，选出最大游览时长，由于可行解数量巨大，求解时间难以接受。无论是用分枝与界限法还是用动态规划，想得到最优解几乎是不可能的，选择模拟退火算法<sup>[4]</sup>，是因为它与一般纯粹的下降算法不同，想得到最优解几乎是不可能的，只能用近似算法。选择模型退火算法，是因为它与一般纯粹的下降算法不同，函数值可在一定条件下随机上升，从一个局部极小值局域跃过波峰到达另一极小值局域，可使近似解更接近全局最优解。在解决此类问题中有着其他算法所没有的优势，本题针对该问题进行模拟退火算法的编程和求解。

**Step1:** 问题的解矩阵和初始值

旅行问题的解矩阵<sup>[5]</sup>R 是游客是否从  $i$  点出发达到  $j$  点的组合，本问题的解矩阵 R 如下：

$$R = \begin{bmatrix} r_{00} & r_{01} & r_{02} & r_{03} & r_{04} & r_{05} & r_{06} & r_{07} \\ r_{10} & r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} & r_{17} \\ r_{20} & r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{26} & r_{27} \\ r_{30} & r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{36} & r_{37} \\ r_{40} & r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{46} & r_{47} \\ r_{50} & r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{56} & r_{57} \\ r_{60} & r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & r_{66} & r_{67} \\ r_{70} & r_{71} & r_{72} & r_{73} & r_{74} & r_{75} & r_{76} & r_{77} \end{bmatrix}$$

其中，每一矩阵  $R_k$  表示游客是否从  $i$  点出发达到  $j$  点的一个路径 ( $k \leq 2^{64}$ )。由于模拟退火算法的最优解与初始解没有很大的依赖关系，故初始解为最简单的。

$$R_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Step2: 目标函数设定**

目标函数即为遍历 8 个景点的路径总长度，即：

$$Z = R_k \times D = \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} d_{ij} \quad (5.2.11)$$

**Step3: 约束条件设定**

根据模型建立中的约束条件进行设定。

**Step4: 新解的产生**

按照约束条件进行变换产生新解。

**Step5: 目标函数差**

计算变换前和变换后的目标函数的差值：

$$\Delta Z = Z' - Z \quad (5.2.12)$$

**Step6: Metreoplis 接受准则**

以新解与当前的目标函数差定义接受概率，即：

$$P = \begin{cases} 1 & \Delta Z < 0 \\ \exp(-\Delta Z / T), & \Delta Z > 0 \end{cases} \quad (5.2.13)$$

**Step7: 程序编写，运用 MATLAB 编程（代码见附录 2）得到结果。**

方案路线长度为 2000m，总的游览时间：270 分钟，总的步行时间：60 分钟，在各个景点均无等待时间，该游览方案只需经过各景点 1 次即可。由于森林小剧场开放时间为半点或整点，经求解可得，为保证无等待时间，确定出此方案森林小剧场的观看时间可为 13:30、14:00、14:30，其余景点时间分配方案有无穷多组，在此本文对这 3 个时间点各示例出一种游览时间分配方案，其中具体游览时间方案见下表。

表 3 (a) 第二问结果表格(森林小剧场 13:30 方案)

序号	景点名称	达到时间点	游览时间(停留时间,单位分钟)	离开时间点
1	景石	12:00	0	12:00
2	阳光草坪	12:10.80	20	12:30.80
3	儿童科普体验区	12:37.70	38.2	13:15.90
4	森林小剧场	13:30.00	30	14:00.00
5	儿童戏水场	14:07.95	20	14:27.95
6	游客服务中心	14:36.50	30	15:06.50
7	湿地博物馆	15:12.50	60	16:12.50
8	湿地商业街	16:18.20	71.8	17:30
总的游览时间: 270 分钟				
总的步行时间: 60 分钟				

由上表可知,该游览线路方案为景石→②阳光草坪(20min)→④儿童科普体验区(38.2min)→③森林小剧场(30min)→⑤儿童戏水场(20min)→①游客服务中心(30min)→⑥湿地博物馆(60min)→⑦湿地商业街(71.8min)。

表 3 (b) 第二问结果表格(森林小剧场 14:00 方案)

序号	景点名称	达到时间点	游览时间(停留时间,单位分钟)	离开时间点
1	景石	12:00	0	12:00
2	阳光草坪	12:10.80	40.00	12:50.80
3	儿童科普体验区	12:57.70	48.20	13:45.90
4	森林小剧场	14:00.00	30.00	14:30.00
5	儿童戏水场	14:37.95	60.00	15:37.95
6	游客服务中心	15:46.50	30.00	16:16.50
7	湿地博物馆	16:22.50	31.80	16:54.30
8	湿地商业街	17:00.00	30.00	17:30
总的游览时间: 270 分钟				
总的步行时间: 60 分钟				

由上表可知,该游览线路方案为景石→②阳光草坪(40min)→④儿童科普体验区(48.2min)→③森林小剧场(30min)→⑤儿童戏水场(60min)→①游客服务中心(30min)→⑥湿地博物馆(31.8min)→⑦湿地商业街(30min)。

表 3 (c) 第二问结果表格(森林小剧场 14:30 方案)

序号	景点名称	达到时间点	游览时间(停留时间,单位分钟)	离开时间点
1	景石	12:00	0	12:00
2	阳光草坪	12:10.80	60.00	13:10.80
3	儿童科普体验区	13:17.70	58.20	14:15.90
4	森林小剧场	14:30.00	30.00	15:00.00
5	儿童戏水场	15:07.95	20.00	15:27.95
6	游客服务中心	15:36.50	23.50	16:00.00
7	湿地博物馆	16:06.00	40.00	16:46.00

8	湿地商业街	16:51.70	38.30	17:30
		总的游览时间:		270 分钟
		总的步行时间:		60 分钟

由上表可知,该游览线路方案为景石→②阳光草坪(60min)→④儿童科普体验区(58.2min)→③森林小剧场(30min)→⑤儿童戏水场(20min)→①游客服务中心(23.5min)→⑥湿地博物馆(40.0min)→⑦湿地商业街(38.3min)。

### 5.3 问题三模型建立与求解

#### 5.3.1 问题三的分析

问题三中,有3个旅游团在12:00同时从景石出发,在17:00前到达湿地商业街,且每个景点仅能容纳一组旅游团游览,要求建立模型设计游览路线,实现3个旅游团均能游览完3个旅游团且游览总时长最长。

由于旅游团均在12:00同时从景石出发,在17:00前到达湿地商业街,17:30离开湿地商业街,即在景区停留时间为5个半小时(330分钟)。由于游客服务中心开放时间在9:00-16:00,森林小剧场仅在整点及半整点开放,因此,需要建立约束条件,进行这两个景点的游览时间约束。由于每个景点仅能容纳一组旅游团游览,需要设计0-1变量,对是否需要等待进行判断,并建立约束条件。由于若等待时间较长,则选择不等待先行游览其他景点的步行时间可能小于等待时间,为保证游览时间最大,认为若等待时间大于最大步行时长(到达除湿地商业街外的其他景点的最大步行时间)则不等待。最大步行时长为:

$$t_{ij\max} = \frac{s_{ij\max}}{v_{ij}} = \frac{765m}{2km/h} = 22.95 \text{ min}$$

由于要求实现3个旅游团均能游览完3个旅游团且游览总时长最长,因此目标函数应考虑3个旅游团的总游览时间,为使得各个旅游团均获得条件允许下的最长游览时长,目标函数同时应考虑3个旅游团游览时长方差最小。因此,以总游览时长最大和3个旅游团游览时长方差最小为目标,考虑各项约束条件,进行多目标优化问题的模型建立。同时,为方便模型建立,将所有时间用分钟表示,认为12:00为0分钟。

考虑目标函数及各项约束条件,进行模型建立。

#### 5.3.2 问题三模型的建立

由问题三分析可知,问题三目标函数应为3个旅游团总游览时间最长及3个旅游团游览时长方差最小。

##### Step1: 确立目标函数

以总游览时间最长为目标,建立目标函数如下:

$$\max Z_3^1 = \sum_{n=1}^3 T^n \quad (5.3.1)$$

由于游客在景区停留的时间<sup>[6]</sup>由“景点之间的步行时间”、“景点游览时间(即在景点内游玩的时间)”和“在景区外的等待时间”三部分组成,其他时间忽略不计,游览时间可由步行时间及等待时间表达。游览时间与其他时间的表达式如下:

$$T^n = 330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n \quad (5.3.2)$$

则以3个旅游团总游览时间最长为目标的目标函数可化简为:

$$\max Z_3^1 = 330 \times 3 - \sum_{n=1}^3 \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{n=1}^3 \sum_{k=1}^6 t_{wk}^n \quad (5.3.3)$$

考虑3个旅游团均在条件允许情况下尽可能获得较大的游览时间，以总游览时长的方差最小为目标，建立目标函数如下：

$$\begin{aligned} \min Z_3^2 &= \frac{1}{3} \times \sum_{n=1}^3 (T^n - \frac{1}{3} \sum_{n=1}^3 T^n)^2 \\ &= \frac{1}{3} \times \sum_{n=1}^3 [ (330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n) - \frac{1}{3} \sum_{n=1}^3 (330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n) ]^2 \end{aligned} \quad (5.3.4)$$

**Step2:** 综合考虑各个影响因素，确定约束条件

(1) 考虑每个旅游团均需在12:00从景石出发，17:00前到达湿地商业街，即每个旅游团的景点间步行时间、景点1-6的游览时间、景点1-6的景区外等待时间之和应小于300，建立约束条件如下：

$$\sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} + \sum_{i=1}^6 t_i^n + \sum_{i=1}^6 t_{wi}^n \leq 300 \quad (5.3.5)$$

(2) 考虑每个旅游团均需在17:30离开湿地商业街，建立约束条件如下：

$$t_{17}^n = 330 \quad (5.3.6)$$

(3) 考虑每个旅游团均需在12:00从景石出发，建立约束条件如下：

$$t_{10}^n = 0 \quad (5.3.7)$$

(4) 考虑游客服务中心①开放时间为9:00~16:00，旅游团需在16:00前离开游客服务中心，建立约束条件如下：

$$t_{11}^n \leq 240 \quad (5.3.8)$$

(5) 考虑如果旅游团  $n$  在景点  $i$  的等待时间超过了最大步行时间，则不等待，建立约束条件如下：

$$t_{wi}^n \leq 22.95 \quad (5.3.9)$$

(6) 考虑若旅游团  $n$  到达景点  $i$  时，景点内若有旅游团正在游览，则需排队等待，判断旅游团  $n$  是否需要等待，即旅游团  $n$  是否在前一组旅游团  $m$  离开之前到达，建立表达式如下：

$$s_k^n = 1 - \left\lfloor \frac{t_{lk}^m - t_{ak}^n}{500} \right\rfloor \quad (5.3.10)$$

(7) 考虑森林小剧场仅在整点及半整点开放，若在其他时间到达则需要等待，判断旅游团  $n$  到达森林小剧场时是否需要等待，建立表达式：

$$u_3^n = \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor - \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor \quad (5.3.11)$$

(8) 考虑旅游团  $n$  到达景点  $j$  的时间与离开景点  $i$  时间关系，建立表达式：

$$t_{aj}^n = (t_{li}^n + t_{ij}^n) r_{ij}^n \quad (5.3.12)$$

(9) 考虑除景点3外，其余景点的等待时间均由排队产生，建立表达式：

$$t_{wq}^n = (t_{lq}^m - t_{aq}^n) s_q^n \quad (5.3.13)$$

(10) 考虑景点3的等待时间由于景点3仅在整点及半整点开放，建立表达式如下：

$$t_{w3}^n = \left( \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor \times 30 - t_{a3}^n \right) u_3^n \quad (5.3.14)$$

(11) 考虑除景点3外，其他景点的等待时间均由排队等待产生，建立游览时间与到达离开景点时间之间表达式如下：

$$t_q^n = (t_{lq}^n - t_{lq}^m) s_i^n + (t_{lq}^n - t_{aq}^n) (1 - s_q^n) \quad (5.3.15)$$

(12) 考虑景点3的等待时间，建立景点3游览时间表达式如下：

$$t_3^n = (1 - u_3^n) (t_{l3}^n - t_{a3}^n) + u_3^n (t_{l3}^n - t_{a3}^n - t_{w3}^n) \quad (5.3.16)$$

(13) 考虑在各个景点的游览时间，结合上述分析推导的表达式，建立约束条件如下：

$$\begin{aligned} 10 \leq t_1^n \leq 30, \quad 20 \leq t_2^n \leq 60, \quad t_3^n = 30, \\ 30 \leq t_4^n \leq 60, \quad 20 \leq t_5^n \leq 60, \quad 30 \leq t_6^n \leq 60, \quad t_7^n \geq 30 \end{aligned} \quad (5.3.17)$$

**Step3:** 综合上述目标函数及约束条件以及问题一相关约束条件，建立模型

$$\begin{aligned} \max Z_3^1 &= 330 \times 3 - \sum_{n=1}^3 \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{n=1}^3 \sum_{k=1}^6 t_{wk}^n \\ \min Z_3^2 &= \frac{1}{3} \times \sum_{n=1}^3 \left[ \left( 330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n \right) - \frac{1}{3} \sum_{n=1}^3 \left( 330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n \right) \right]^2 \\ s.t. & \left\{ \begin{aligned} & \sum_{i=0}^7 r_{ik} = \sum_{j=0}^7 r_{kj} = 1 \\ & r_{ij} \times r_{ji} = 0, i \in O, j \in O \\ & \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} = 7, \sum_{i=0}^7 r_{i0} = 0, \sum_{j=0}^7 r_{7j} = 0 \\ & \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} + \sum_{i=1}^6 t_i^n + \sum_{i=1}^6 t_{wi}^n \leq 300 \\ & t_{l7}^n = 330, t_{l0}^n = 0, t_{l1}^n \leq 240, t_{wi}^n \leq 22.95 \\ & t_{aj}^n = (t_{li}^n + t_{ij}^n) r_{ij}^n, t_{wq}^n = (t_{lq}^n - t_{aq}^n) s_q^n \\ & t_{w3}^n = \left( \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor \times 30 - t_{a3}^n \right) u_3^n \\ & t_i^n = (t_{li}^n - t_{li}^m) s_i^n + (t_{li}^n - t_{ai}^n) (1 - s_i^n) \\ & t_3^n = (1 - u_3^n) (t_{l3}^n - t_{a3}^n) + u_3^n (t_{l3}^n - t_{a3}^n - t_{w3}^n) \\ & s_k^n = 1 - \left\lfloor \frac{t_{lk}^m - t_{ak}^n}{500} \right\rfloor, u_3^n = \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor - \left\lfloor \frac{t_{a3}^n}{30} \right\rfloor \\ & 10 \leq t_1^n \leq 30, 20 \leq t_2^n \leq 60, t_3^n = 30 \\ & 30 \leq t_4^n \leq 60, 20 \leq t_5^n \leq 60 \\ & 30 \leq t_6^n \leq 60, t_7^n \geq 30 \end{aligned} \right. \end{aligned}$$

### 5.3.3 问题三模型的求解

在本题中，要求分别设计一条能游览全部景点且游览总时间最长，总等待时间

最短的游览路线。这是一个典型的双目标函数<sup>[7]</sup>，我们的决策目标函数有 2 个： $Z_3^1$  和  $Z_3^2$ 。对于“ $Z_3^1$  尽量大， $Z_3^2$  尽量小”的双目标要求，我们很难给出一个完全贴合实际的界定函数  $F$ ，使得  $F(Z_3^1, Z_3^2)$  取最大（或最小）时的  $(Z_3^1, Z_3^2)$  就是最优解。通过构造新目标函数，将问题转化为“单目标决策问题”，使问题的求解变的简单可行。

其基本思路是在将双目标转化为单目标的基础上，构造恰当的优化条件（用不等式形式表达），不断从可行解的集合中除去不满足优化条件的点，直到可行解的集合只有有限的几个或更少为止；最后从剩下的为数不多的可行解中，通

过人脑进行主观判断，得到最贴合实际的最优可行解，因为具体什么才算最优是一个模糊的概念，并且还是一对矛盾。所以最后一步需要人来判断。

**Step1:** 将双目标函数转化为单目标函数。根据约束法将  $Z_3^2$  转化为约束条件。在本题中，我们认定，当 3 个旅游团的游览时长方差小于 10 时，认定其为本文中的可行解，否则认定为不可行解。即  $Z_3^2 \leq 10$ 。

**Step2:** 求解所有可行解。若采用穷举法，将所有可行解一一列出，再验证每条路是否满足约束条件，从而可以得到所有可行解。但穷举路径法的时间复杂度为较大，即使用计算机也是永远穷举不完的。故此，我们利用问题二中模型算法进行计算。

可行解：满足约束条件的解均称为可行解<sup>[8]</sup>。

**Step3:** 从所有可行解中得到非劣解，并组成非劣解集合。

劣解： $L_0$  称为劣解，如果存在可行解  $L_1$ ，且不等于  $L_0$ ，满足  $Z_4^1(0) < Z_4^1(1)$ ， $Z_4^2(0) > Z_4^2(1)$ ，即  $L_0$  得到的总游览时长比  $L_1$  小，总等待时间却比  $L_1$  大。

非劣解（有效解）：不是劣解的可行解称为非劣解或有效解。

**Step4:** 由非劣解集合生成目标空间，优化目标空间，并得到优化后的目标空间。

目标空间：所有非劣解对应的目标函数集  $(Z_4^1, Z_4^2)$  组成的集合称为目标空间。

**Step5:** 找到优化后的目标空间中元素所对应的非劣解，并人为主观的从中判断出最优解。

用 MATLAB 编程实现该模型，程序见附录 3，结果见下表。

表 4 第三问结果表格

	第一旅游团			第二旅游团			第三旅游团		
	达到 时间点	游览时间 (停留时间,单位 分钟)	离开 时间点	达到 时间点	游览时间 (停留时间, 单位分钟)	离开 时间点	达到 时间点	游览时间 (停留时间,单位 分钟)	离开 时间点
景石	12:00	0	12:00	12:00	0	12:00	12:00	0	12:00
游客服务 中心	15:36.50	10.00	15:46.50	13:38.10	20.10	13:58.20	12:09.00	10	12:19.00
阳光草坪	12:10.80	58.20	13:09.00	14:09.60	60.00	15:09.60	13:09.00	35.70	13:44.70
森林小剧 场	14:30.00	30.00	15:00.00	13:00.00	30.00	13:30.00	14:00.00	30	14:30.00
儿童科普 体验区	13:15.90	60.00	14:15.90	15:16.50	60.00	16:16.50	12:25.90	36.2	13:02.10
儿童戏水 场	15:07.95	20.00	15:27.95	12:14.25	37.80	12:52.05	14:37.95	30	15: 07.95
湿地博物 馆	15:52.50	30.00	16:22.50	16:24.30	30	16:54.30	15:21.75	30	15:51.75

湿地商业街	16:28.20	61.80	17:30	17:00.00	30	17:30	15:57.45	92.55	17:30
总步行时间	60 分钟		62.1 分钟		65.55 分钟				
总游览时间	270 分钟		267.9 分钟		264.45 分钟				
总等待时间	0 分钟		0 分钟		0 分钟				

由上表可知，第一旅游团的游览线路方案为景石→②阳光草坪（58.2min）→④儿童科普体验区（60min）→③森林小剧场（30min）→⑤儿童戏水场（20min）→①游客服务中心（10min）→⑥湿地博物馆（30min）→⑦湿地商业街（61.8min）。方案路线长度为2000m，总的游览时间：270分钟，总的步行时间：60分钟，在各个景点均无等待时间，该游览方案只需经过各景点1次即可。

第二旅游团的游览线路方案为景石→②阳光草坪（60min）→④儿童科普体验区（60min）→③森林小剧场（30min）→⑤儿童戏水场（37.8min）→①游客服务中心（20.1min）→⑥湿地博物馆（30min）→⑦湿地商业街（30min）。方案路线长度为2070m，总的游览时间：267.9分钟，总的步行时间：62.1分钟，在各个景点均无等待时间，该游览方案只需经过各景点1次即可。

第三旅游团的游览线路方案为景石→②阳光草坪（35.7min）→④儿童科普体验区（36.2min）→③森林小剧场（30min）→⑤儿童戏水场（30min）→①游客服务中心（10min）→⑥湿地博物馆（30min）→⑦湿地商业街（92.55min）。方案路线长度为2185m，总的游览时间：264.45分钟，总的步行时间：65.55分钟，在各个景点均无等待时间，该游览方案只需经过各景点1次即可。

3个旅游团总游览时间为802.35min，总等待时间为0，总步行时间为187.65min。

### 5.3.4 问题三结果的分析

由于森林小剧场开放时间为半点或整点，游客服务中心开放时间到16:00，经求解可得，为尽量减少等待时间且游览时间尽可能长，首先确定出方案中森林小剧场和游客服务中心游的观看时间，其余景点时间分配方案有无穷多组，在此本文对这3个旅游团的路线各示例出一种游览时间分配方案。

## 5.4 问题四模型建立与求解

### 5.4.1 问题四的分析

问题四中，在问题三的基础上，步行速度可以变化，在 $1\text{km}/\text{h}$ 到 $3\text{km}/\text{h}$ 之间调节，但总平均速度不能超过 $2\text{km}/\text{h}$ 。因此，问题四在问题三的基础上，步行时间存在变化，利用步行速度的约束建立约束条件，其他约束条件与问题三相同。为模型建立方便，将单位统一化简为 $\text{m}/\text{min}$ 。由于问题四要求在游览完全部7个景点的前提下，保证游览时间尽可能长且等待时间尽可能短，3个旅游团游览时长方差最小。因此模型应为多目标约束模型，目标函数应为游览时间最长与等待时间最短以及3个旅游团的游览时长方差最小。

考虑目标函数及各项约束条件，进行模型建立。

### 5.4.2 问题四模型的建立

由问题四分析可知，问题四目标函数应为多目标，即3个旅游团总游览时间最长且总等待时间最短，3个旅游团总游览时间方差最小。

**Step1:** 确立目标函数

以总游览时间最长为目标，建立目标函数如下：

$$\max Z_4^1 = 330 \times 3 - \sum_{n=1}^3 \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{n=1}^3 \sum_{i=1}^7 t_{wi}^n \quad (5.4.1)$$

以总等待时间最短，建立约束条件如下：

$$\min Z_4^2 = \sum_{n=1}^3 \sum_{i=1}^7 t_{wi}^n \quad (5.4.2)$$

考虑3个旅游团均在条件允许情况下尽可能获得较大的游览时间，以总游览时长的方差最小为目标，建立目标函数如下：

$$\begin{aligned} \min Z_4^3 &= \frac{1}{3} \times \sum_{n=1}^3 (T^n - \frac{1}{3} \sum_{n=1}^3 T^n)^2 \\ &= \frac{1}{3} \times \sum_{n=1}^3 [(330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n) - \frac{1}{3} \sum_{n=1}^3 (330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n)]^2 \end{aligned} \quad (5.4.3)$$

**Step2:** 综合考虑各个影响因素，确定约束条件

(1) 考虑步行速度在在  $1km/h$  到  $3km/h$  之间调节，且总平均速度不超过  $2km/h$ ，单位化为  $m/min$ ，建立约束条件：

$$\frac{50}{3} \leq v_{ij}^n \leq 50 \quad (5.4.4)$$

$$\frac{s_{ij}}{v_{ij}^n} \leq \frac{100}{3} \quad (5.4.5)$$

(2) 考虑每个旅游团均需在12:00从景石出发，17:00前到达湿地商业街，即每个旅游团的景点间步行时间、景点1~6的游览时间、景点1~6的景区外等待时间之和应小于300，建立约束条件：

$$\sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times \frac{s_{ij}}{v_{ij}^n} + \sum_{i=1}^6 t_i^n + \sum_{i=1}^6 t_{wi}^n \leq 300 \quad (5.4.6)$$

**Step3:** 综合考虑目标函数及上述约束条件，结合问题三及问题一中的相关约束条件，建立模型

$$\max Z_4^1 = 330 \times 3 - \sum_{n=1}^3 \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{n=1}^3 \sum_{i=1}^7 t_{wi}^n$$

$$\min Z_4^2 = \sum_{n=1}^3 \sum_{i=1}^7 t_{wi}^n$$

$$\min Z_4^3 = \frac{1}{3} \times \sum_{n=1}^3 \left[ (330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n) - \frac{1}{3} \sum_{n=1}^3 (330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij} - \sum_{k=1}^6 t_{wk}^n) \right]^2$$

$$s.t. \left\{ \begin{array}{l} \sum_{i=0}^7 r_{ik} = \sum_{j=0}^7 r_{kj} = 1, r_{ij} \times r_{ji} = 0, i \in O, j \in O \\ \sum_{i=0}^7 \sum_{j=0}^7 r_{ij} = 7, \sum_{i=0}^7 r_{i0} = 0, \sum_{j=0}^7 r_{7j} = 0 \\ \frac{50}{3} \leq v_{ij}^n \leq 50, \frac{s_{ij}}{v_{ij}^n} \leq \frac{100}{3} \\ t_{ij}^n = \frac{s_{ij}}{v_{ij}^n}, \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}^n \times t_{ij}^n + \sum_{i=1}^6 t_i^n + \sum_{i=1}^6 t_{wi}^n \leq 300 \\ t_{l7}^n = 330, t_{l0}^n = 0, t_{l1}^n \leq 240, t_{wi}^n \leq 22.95 \\ t_{aj}^n = (t_{li}^n + t_{ij}^n) r_{ij}^n, t_{wq}^n = (t_{lq}^m - t_{aq}^n) s_q^n \\ t_{w3}^n = \left\lceil \frac{t_{w3}^n}{30} \right\rceil \times 30 - t_{w3}^n u_3^n \\ t_i^n = (t_{li}^n - t_{li}^m) s_i^n + (t_{li}^n - t_{ai}^n) (1 - s_i^n) \\ t_3^n = (1 - u_3^n) (t_{l3}^n - t_{a3}^n) + u_3^n (t_{l3}^n - t_{a3}^n - t_{w3}^n) \\ s_k^n = 1 - \left\lceil \frac{t_{lk}^m - t_{ak}^n}{500} \right\rceil, u_3^n = \left\lceil \frac{t_{a3}^n}{30} \right\rceil - \left\lceil \frac{t_{a3}^n}{30} \right\rceil \\ 10 \leq t_1^n \leq 30, 20 \leq t_2^n \leq 60, t_3^n = 30 \\ 30 \leq t_4^n \leq 60, 20 \leq t_5^n \leq 60 \\ 30 \leq t_6^n \leq 60, t_7^n \geq 30 \end{array} \right.$$

### 5.4.3 问题四模型的求解

在本题中，在速度可变的条件在，要求分别设计一条能游览全部景点且游览总时间最长，总等待时间最短的。这是一个典型的多目标函数，我们的决策目标函数有 3 个： $Z_4^1$ 、 $Z_4^2$ 和 $Z_4^3$ 。对于“ $Z_4^1$ 尽量大， $Z_4^2$ 、 $Z_4^3$ 尽量小”的多目标要求，我们很难给出一个完全贴合实际的界定函数 F，使得 F ( $Z_4^1$ ,  $Z_4^2$ ,  $Z_4^3$ ) 取最大（或最小）时的 ( $Z_4^1$ ,  $Z_4^2$ ,  $Z_4^3$ ) 就是最优解。通过构造新目标函数，将问题转化为“单目标决策问题”，使问题的求解变的简单可行。

其基本思路是通过构造恰当的优化条件<sup>[9]</sup>（用不等式形式表达），不断从可行解的集合中除去不满足优化条件的点，直到可行解的集合只有有限的几个或更少为止；最后从剩下的为数不多的可行解中，通过人脑进行主观判断，得到最贴合实际的最优可行解，因为具体什么才算最优是一个模糊的概念，并且还是一对矛盾。所以最

后一步需要人来判断。

**Step1:** 将双目标函数转化为单目标函数。根据约束法将  $Z_4^2$ 、 $Z_4^3$  转化为约束条件。在本题中，我们认定，当 3 个旅游团的游览时长方差小于 10 时，认定其为本文中的可行解，否则认定为不可行解。即  $Z_4^3 \leq 10$ ，且由于速度可变，可以降低等待时间，从而增加总游览时间，故总游览时间会优于第三问答案，即  $Z_4^2 \geq 802$ 。

**Step2:** 求解所有可行解。若采用穷举法，将所有可行解一一列出，再验证每条路是否满足约束条件，从而可以得到所有可行解。但穷举路径法的时间复杂度为较大，即使用计算机也是永远穷举不完的。故此，我们利用问题二中模型算法进行计算。

**Step3:** 从所有可行解中得到非劣解，并组成非劣解集合。

**Step4:** 由非劣解集合生成目标空间，优化目标空间，并得到优化后的目标空间。

**Step5:** 找到优化后的目标空间中元素所对应的非劣解，并人为主观的从中判断出最优解。

用 MATLAB 编程实现该模型，程序见附录 4，结果如下：

表 5 第四问结果表格

	第一旅游团				第二旅游团				第三旅游团			
	达到 时间点	游览 时间 (停留 时间, 单位 分钟)	离开 时间点	达到下 一个景 点的步 行速度 $km/h$	达到 时间点	游览时 间(停 留时 间,单 位分 钟)	离开 时间点	达到下 一个景 点的步 行速度 $km/h$	达到 时间点	游览 时间 (停留 时间, 单位 分钟)	离开 时间点	达到下 一个景 点的步 行速度 $km/h$
景石	12:00	0	12:00	2	12:00	0	12:00	3	12:00	0	12:00	1.13
游客服 务中心	15:36.50	10.00	15:46.50	2	12:06.00	10.00	12:16.00	1.48	12:16.00	30.00	12:46.00	2.83
阳光草 坪	12:10.80	58.20	13:09.00	2	13:45.30	22.50	14:07.80	2	14:43.30	24.50	15:07.80	2
森林小 剧场	14:30.00	30.00	15:00.00	2	13:00.00	30.00	13:30.00	2	14:00.00	30.00	14:30.00	2.30
儿童科 普体验 区	13:15.90	60.00	14:15.90	2	14:14.70	60.00	15:14.70	2	15:14.70	60.00	16:14.70	2.69
儿童戏 水场	15:07.95	20.00	15:27.95	2	12:27.55	24.50	12:52.05	2	12:52.05	60.00	13:52.05	2
湿地博 物馆	15:52.50	30.00	16:22.50	2	15:22.50	30.00	15:52.50	2	16:20.50	32.00	16:52.50	2.19
湿地商 业街	16:28.20	61.80	17:30		15:58.20	91.80	17:30		16:58.20	32.30	17:30	
总步行 时间	60 分钟				61.2 分钟				61.2 分钟			
总游览 时间	270 分钟				268.8 分钟				268.8 分钟			
总等待 时间	0 分钟				0 分钟				0 分钟			
平均速 度	2 $km/h$				2 $km/h$				2 $km/h$			

由上表可知，第一旅游团的游览线路方案为：景石→②阳光草坪（58.2min）→④儿童科普体验区（60min）→③森林小剧场（30min）→⑤儿童戏水场（20min）→①游客服务中心（10min）→⑥湿地博物馆（30min）→⑦湿地商业街（61.8min）。方案路线长度为 2000m，总的游览时间：270 分钟，总的步行时间：60 分钟，在各

个景点均无等待时间，该游览方案只需经过各景点 1 次即可。

第二旅游团的游览线路方案为：景石→①游客服务中心（10min）→⑤儿童戏水场（24.5min）→③森林小剧场（30min）→②阳光草坪（22.5min）→④儿童科普体验区（60min）→⑥湿地博物馆（30min）→⑦湿地商业街（91.8min）。方案路线长度为 2040m，总的游览时间：268.8 分钟，总的步行时间：61.2 分钟，在各个景点均无等待时间，该游览方案只需经过各景点 1 次即可。

第三旅游团的游览线路方案为：景石→①游客服务中心（30min）→⑤儿童戏水场（60min）→③森林小剧场（30min）→②阳光草坪（24.5min）→④儿童科普体验区（60min）→⑥湿地博物馆（32min）→⑦湿地商业街（32.3min）。方案路线长度为 2040m，总的游览时间：268.8 分钟，总的步行时间：61.2 分钟，在各个景点均无等待时间，该游览方案只需经过各景点 1 次即可。

3 个旅游团总游览时间为 807.6min，总等待时间为 0，总步行时间为 182.4min。

#### 5.4.4 问题四结果的分析

由于旅游团速度可调节，所以出现了二、三旅游团游览线路一致的情况。根据速度的变化，让第二旅游团以最快的速度抵达游客服务中心进行游览，第三旅游团则以较缓速度抵达，可将二个旅游团的游览时间完全错开，其余景点时间分配方案也有无穷多组，在此本文对这 3 个旅游团的路线只示例出一种游览时间分配方案即可。

### 5.5 问题五模型建立与求解

#### 5.5.1 问题五的分析

在第四问的基础上，需要考虑不同旅游团出发时间的不确定性和每个景点等待时间的不确定性，其它条件同第四问相同。对于出发时间的不确定性而言，由于受各景点开放时间、景点之间距离和旅游团走行速度的影响，能确定最迟出发时间为 12:55，故旅游团的出发时间是一个区间数。此外，本问中要求考虑旅游设施短时间的维护和清理及散客的影响，对某一旅游团而言，其它旅游团和散客对该团的影响是相同的，故该问题可以简化成为一个旅游团设计一条能游览全部 7 个景点且游览总时间长、总等待时间短的游览路线。因此在第四问的基础上，建立不确定多目标约束模型，目标函数为单旅游团的游览时间最长和等待时间最短。

当旅游团面临非正常等待时，由两种方案：1) 沿原路径完成非正常等待后游览完毕继续前进<sup>[10]</sup>；2) 改走其它路径。这两种选择都是有风险的，因为方案一中非正常等待时间  $\mu_i$  是随机的，而方案二中也无法保证改变路径后到达的下一个景点不会产生非正常等待。

根据动态规划<sup>[11]</sup>的最优性定理推论可知：若允许策略是最优策略，则对任意阶段而言，它的子策略对于以该阶段状态为起点的后续过程而言，必是最优策略。基于此，我们就以该点为新的起点，考虑堵塞时间的期望，重新寻找达到终点、满足约束条件且是目标函数最优的路径，把这种策略成为新起点策略。

#### 5.5.2 问题五模型的建立

##### Step1: 确立目标函数

由问题五分析可知，该问题为单旅游团的多目标约束模型，建立游览时间最长目标函数如下：

$$\max Z_5^1 = 330 - \sum_{i=0}^7 \sum_{j=0}^7 r_{ij}' \times \frac{s_{ij}'}{v_{ij}'} - \sum_{i=1}^7 t_{wi}' \quad (5.5.1)$$

以总等待时间最短，建立约束条件如下：

$$\min Z_5^2 = \sum_{i=1}^7 t_{wi}' \quad (5.5.2)$$

**Step2:** 综合考虑各个影响因素，确定约束条件

(1)该问中旅游团到达湿地商业街的时间可以根据出发时间进行顺延，同样地，每个旅游团的景点间步行时间、景点1~6的游览时间、景点1~6的景区外等待时间之和也应小于300，建立约束条件：

$$\sum_{i=0}^7 \sum_{j=0}^7 r_{ij}' \times \frac{s_{ij}'}{v_{ij}'} + \sum_{i=1}^6 t_i' + \sum_{i=1}^7 t_{wi}' \leq 300 \quad (5.5.3)$$

(2)考虑受各景点开放时间的限制，在旅游团采用最短路时可得到从景石最晚出发时间为13: 51.1，建立约束条件：

$$0 \leq t_{i0}' \leq 111.1 \quad (5.5.4)$$

(3)考虑各景点的开放时间，建立约束条件：

$$t_{i1}' \leq 240, t_{ii}' \leq 300 (i = 2, 3, 4, 5, 6), t_{i7}' \leq 570 \quad (5.5.5)$$

(4)考虑如果旅游团在景点 $i$ 的等待时间超过了最大步行时间，则不等待，建立约束条件：

$$t_{wi}' \leq 22.95 \quad (5.5.6)$$

(5)考虑若旅游团到达景点 $i$ 时，可能产生非正常等待（由于景点容量限制和旅游设施维护清理导致的等待），假设 $X$ 中的非正常等待时间均值 $\bar{\mu}_i$ 和非正常等待概率 $p_i$ 均可以用过资料统计分析获得，得下式：

$$\sum_{i=1}^7 p_i \bar{\mu}_i = \sum_{i=1}^7 \theta_i \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{\mu}_i)^2}{2\sigma^2}} dx \quad (5.5.7)$$

(6)考虑森林小剧场仅在整点及半整点开放和非正常等待的可能性，在没有非正常等待情况下，旅游团到达时间不满足该景点开放要求则会产生正常等待；若该景点因占用或维修清理而产生非正常等待时间，在非正常等待结束后的时间点不满足进场要求，同样也会产生正常等待，判断旅游团在该景点是否需要等待，建立表达式：

$$u_3' = \left\lfloor \frac{t_{a3}' + \theta_i \mu_i}{30} \right\rfloor - \left\lfloor \frac{t_{a3}' + \theta_i \mu_i}{30} \right\rfloor \quad (5.5.8)$$

(7)考虑旅游团到达景点 $j$ 的时间与离开景点 $i$ 时间关系，建立表达式：

$$t_{aj}' = (t_{ii}' + t_{ij}') r_{ij}' \quad (5.5.9)$$

(8)考虑景点3仅在整点及半整点开放，建立在景点3的总等待时间表达式：

$$t_{w3}' = \left( \left\lfloor \frac{t_{a3}' + \theta_i \mu_i}{30} \right\rfloor \times 30 - t_{a3}' - \theta_i \mu_i \right) u_3' + \theta_i \mu_i \quad (5.5.10)$$

其它景点的等待时间：

$$t_{wi}' = \theta_i \mu_i, i = 1, 2, 4, 5, 6, 7 \quad (5.5.11)$$

即：

$$t_{wi}' = \begin{cases} \left( \left\lfloor \frac{t_{a3}' + \theta_i \mu_i}{30} \right\rfloor \times 30 - t_{a3}' - \theta_i \mu_i \right) u_3' + \theta_i \mu_i, i = 3 \\ \theta_i \mu_i, i = 1, 2, 4, 5, 6, 7 \end{cases} \quad (5.5.12)$$

(9) 建立旅游团在某景点游览时间与到达、离开景点时间和总等待时间之间表达式:

$$t'_i = t'_{li} - (t'_{ai} + t'_{wi}) \quad (5.5.13)$$

(10) 考虑在各个景点的游览时间, 结合上述分析推导的表达式, 建立约束条件:

$$10 \leq t'_1 \leq 30, \quad 20 \leq t'_2 \leq 60, \quad t'_3 = 30, \quad 30 \leq t'_4 \leq 60, \quad 20 \leq t'_5 \leq 60, \quad 30 \leq t'_6 \leq 60, \quad t'_7 \geq 30$$

(11) 考虑步行速度在在  $1\text{km}/\text{h}$  到  $3\text{km}/\text{h}$  之间调节, 且总平均速度不超过  $2\text{km}/\text{h}$ , 单位化为  $\text{m}/\text{min}$ , 建立约束条件:

$$\frac{50}{3} \leq v'_{ij} \leq 50, \quad \frac{s_{ij}}{v'_{ij}} \leq \frac{100}{3} \quad (5.5.14)$$

**Step3:** 综合考虑目标函数及上述约束条件, 结合问题四及问题一中的相关约束条件, 建立模型

$$\max Z_5^1 = 330 - \sum_{i=0}^7 \sum_{j=0}^7 r'_{ij} \times \frac{s_{ij}}{v'_{ij}} - \sum_{i=1}^7 t'_{wi}$$

$$\min Z_5^2 = \sum_{i=1}^7 t'_{wi}$$

$$\left\{ \begin{array}{l} \sum_{i=0}^7 r'_{ik} = \sum_{j=0}^7 r'_{kj} = 1, r'_{ij} \times r'_{ji} = 0, i \in O, j \in O \\ \sum_{i=0}^7 \sum_{j=0}^7 r'_{ij} = 7, \sum_{i=0}^7 r'_{i0} = 0, \sum_{j=0}^7 r'_{7j} = 0 \\ \frac{50}{3} \leq v'_{ij} \leq 50, \frac{s_{ij}}{v'_{ij}} \leq \frac{100}{3}, t'_{ij} = \frac{s_{ij}}{v'_{ij}} \\ \sum_{i=0}^7 \sum_{j=0}^7 r'_{ij} \times \frac{s_{ij}}{v'_{ij}} + \sum_{i=1}^6 t'_i + \sum_{i=1}^7 t'_{wi} \leq 300 \\ t'_{l1} \leq 240, t'_{li} \leq 300 (i = 2, 3, 4, 5, 6), t'_{l7} \leq 570, t'_{wi} \leq 22.95 \\ s.t. \left\{ \begin{array}{l} t'_{aj} = (t'_{li} + t'_{ij})r'_{ij}, 0 \leq t'_{l0} \leq 111.1, t'_i = t'_{li} - (t'_{ai} + t'_{wi}) \\ \sum_{i=1}^7 p_i \bar{\mu}_i = \sum_{i=1}^7 \theta_i \int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\bar{\mu}_i)^2}{2\sigma_i^2}} dx \\ u'_3 = \left[ \frac{t'_{a3} + \theta_i \mu_i}{30} \right] - \left[ \frac{t'_{a3} + \theta_i \mu_i}{30} \right] \\ t'_{wi} = \begin{cases} \left( \left[ \frac{t'_{a3} + \theta_i \mu_i}{30} \right] \times 30 - t'_{a3} - \theta_i \mu_i \right) u'_3 + \theta_i \mu_i, i = 3 \\ \theta_i \mu_i, i = 1, 2, 4, 5, 6, 7 \end{cases} \\ 10 \leq t'_1 \leq 30, 20 \leq t'_2 \leq 60, t'_3 = 30, 30 \leq t'_4 \leq 60, 20 \leq t'_5 \leq 60, 30 \leq t'_6 \leq 60, t'_7 \geq 30 \end{array} \right. \end{array} \right.$$

### 5.5.3 问题五模型的求解

由于任何一个景点都可能因为非正常等待而成为新的起点, 在本问中主要是提出在随机情况下的决策模型<sup>[12]</sup>, 缺乏统计数据进行参数标定, 故假设某一时段的参

数状态值给定,见附表,在前问的基础上反复用模拟退火算法进行 MATLAB 编程,程序代码见附录 5,在随机情况下得到最优路线方案为景石→③森林小剧场→⑤儿童戏水场→①游客服务中心→②阳光草坪→④儿童科普体验区→⑥湿地博物馆→⑦湿地商业街,该游览方案只需经过各景点 1 次即可。参数的取值都会影响决策。

## 六、模型的评价与推广

### 6.1 模型的评价

模型的优点:模型直观易懂,灵活运用 MATLAB 软件对数据进行处理,直观、快捷,对实际问题的描述是合理的。在本文中,相比于其他模型,此模型变量较多,求解过程比较困难,但它却能求得最优解。在计算中,我们采用了合理的算法及相关软件,并通过合理假设,使得求解难度大大降低。对结果我们也做了相应的优化,使结果更符合实际情况。

模型的缺点:在模型假设中忽略了天气等因素对旅游路线选择的影响,但是在实际情况下必须考虑;由于影响旅游路线选择的因素众多,本文在这里做了适当简化处理,与实际情况有些许偏差,这些都需要在以后的模型建立中进行适当改进。

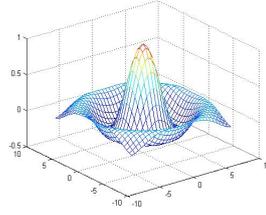
### 6.2 模型的推广

本文第一问建立全部景点游览距离最短的遍历模型,改进了经典的旅游商模型;然后,第二问在不考虑等待时间的情况下,建立了游客游览线路选择的 0-1 模型;在第三、四问中逐步改进,考虑多目标优化,结合问题一模型制定出 3 个旅游团分别的游览路线;本文第五问分析了现实中的不确定因素影响,提出全新的包含不确定因素的旅游路径优化问题,建立了新的线路选择模型,可以应用到较复杂问题的处理上,对旅游中游览线路的制定有一定的参考价值。

## 七、参考文献

- [1]杜鹏楨,唐振民,孙研.一种面向对象的多角色蚁群算法及其 TSP 问题求解[J].控制与决策,2014,29(10):1729-1736.
- [2]霍星,史海滨,杨松益,李泽鸣,龚雪文,杨晓.基于层次分析-蚁群算法的内蒙古大型灌区节水改造综合评价[J].农业工程学报,2014,30(17):132-140.
- [3]王沛栋.改进蚁群算法及在路径规划问题的应用研究[D].中国海洋大学,2012.
- [4]孙宇兴,谭芬.利用动态规划解决 01 背包问题[J].现代交际,2010(05):121-122+120.
- [5]姚明海,王娜,赵连朋.改进的模拟退火和遗传算法求解 TSP 问题[J].计算机工程与应用,2013,49(14):60-65.
- [6]张涛,陈忠,吕一兵.利用一种改进的模拟退火算法求解多目标规划问题[J].武汉工业学院学报,2013,32(02):74-76.
- [7]裴小兵,贾定芳.基于模拟退火算法的景点物流多目标配送车辆路径优化研究[J].数学的实践与认识,2016,46(02):105-113.
- [8]蒋正威,曹一家,孙维真.基于 01 整数规划的多目标最优 PMU 配置算法[J].电力系统保护与控制,2008(21):12-17.
- [9]张水舰,刘学军,杨洋.动态随机最短路径算法研究[J].物理学报,2012,61(16):1-10.
- [10]冯树民,吴海月,王弟鑫.基于理想点法的多目标最短路径求解算法研究[J].公路交通科技,2016,33(03):97-101.
- [11]胡静,李金龙,曹先彬.模拟退火算法与遗传算法结合及多目标优化求解研究[J].计算机应用与软件,2000(11):19-23+54.
- [12]林澜,闫春钢,蒋昌俊,周向东.动态网络最短路径问题的复杂性与近似算法[J].计算机学报,2007(04):608-614.

# 五一数学建模竞赛



## 题目：徐州潘安湖风景区游览路线设计

**关键词：**整数规划 k 短路算法 0-1 变量 Matlab 软件 多目标规划

### 摘要：

本文主要研究了在给定条件下如何规划徐州潘安湖风景区游览路线问题，重点研究分析了不同情况下的游览路线设计，通过建立整数规划模型，并利用 k 短路算法和枚举算法对提出的模型进行求解，具体结果如下：

针对问题一：考虑到每个景点可能重复经过，故将经过两个景点之间路段的次数作为变量，建立整数规划模型，利用枚举法和约束优化算法并通过 Matlab 编程，得出了最佳旅游路径，此时游览总路程为 1820 米，游览路径为：景石→③→⑤→①→②→④→⑥→⑦。

针对问题二：在问题一的模型的基础上建立新的整数规划模型，由于在模型中加入了新的约束条件，故引入 k 短路算法并利用 Matlab 编程对模型进行求解，得到预览时间最长的路线，此时游览时间为 270 分钟，具体路线为：景石→②→④→③→⑤→①→⑥→⑦。

对于问题三：根据问题一二中提出的整数规划模型，将问题二中的模型变量扩展到三个旅游团，并建立新的整数规划模型，利用 K 短路算法进行求解，得到最优方案如下：旅行团 1：景石→②→④→③→⑤→①→⑥→⑦；旅行团 2：景石→①→⑤→③→②→④→⑥→⑦；旅行团 3：景石→⑤→③→②→④→①→⑥→⑦。

对于问题四：问题中要求总的游览时间尽可能长，总的等待时间尽可能短。据此，通过建立新的目标函数，将模型建立为一个多目标规划问题。考虑到多目标规划问题求解上的特点，本题将通过为等待时间函数与游览时间的函数添加不同的权重，将目标函数转化为单目标模型，运用 Matlab 编程进行求解，得到最优游览路线方案：旅行团 1：景石→②→④→③→⑤→①→⑥→⑦，旅行团 2：景石→①→⑤→③→②→④→⑥→⑦，旅行团 3：景石→⑤→③→①→②→④→⑥→⑦

对于问题五：题目中要求考虑出发时间以及等待时间两个不确定因素。据此在问题四模型的基础上，增加一个旅行团，考虑等待时间和游览时间两个函数的权重，并对等待时间添加新的约束条件，接着使用 Matlab 编程进行求解，得到四个旅行团的最优旅行路径如下：A: 景石→③→⑤→①→②→④→⑥→⑦；B: 景石→②→④→③→⑤→①→⑥→⑦；C: 景石→⑤→③→①→②→④→⑥→⑦；D: 景石→①→⑤→③→②→④→⑥→⑦。

最后，深入研究了论文中提出模型的优缺点，并对模型的优化提供了一些改进思路。

# 徐州潘安湖风景区游览路线设计

## 一、问题重述

习近平总书记在党的十九大报告中明确提出“支持资源型地区经济转型发展”。2017年12月12日至13日，习近平总书记来到徐州调研，这是十九大后习总书记的首次调研。习近平总书记在徐州调研贾汪区潘安采煤塌陷区整治工程时强调，塌陷区要坚持走符合国情的转型发展之路，打造绿水青山，并把绿水青山变成金山银山。

徐州是一个老工业基地和资源型城市，煤炭开采历史长达130年。长期煤炭开采在徐州累计形成采煤塌陷区达数十万亩。位于徐州市贾汪区西南部、紧邻马庄的潘安湖湿地公园原来就是徐州最大的、塌陷最严重的采煤塌陷区。据统计，1949年以来，贾汪煤田累计为国家提供原煤3.5亿余吨，也造成了11.3万亩土地塌陷。2010年，江苏省单体投资最大的一宗土地整理项目在潘安湖启动，即利用采煤塌陷形成的开阔水面，同步展开基本农田整理、采煤塌陷地复垦、生态环境修复等，在一片废墟上建成一个6500亩湖面的国家级水利风景区。2016年，贾汪被列为“国家全域旅游示范区”首批创建单位，全区共接待游客560万人次，旅游综合收入达到16.8亿元。本题选取潘安湖景区的部分景点(如图1所示)，请在如下的假设下，完成徐州潘安湖风景区游览路线设计问题。

假设：

(1) 任意两个景点之间的最短步行距离如表1给出。

(2) 第二问、第三问假设步行速度 $V = 2\text{km}/\text{h}$ 。

(3) 游客在景区停留的时间由“景点之间的步行时间”、“景点游览时间(即在景点内游玩的时间)”和“在景区外的等待时间”三部分组成，其他时间忽略不计，游览时间必须符合表2的要求。

**问题1：**从景石出发，步行游览以下景点：①游客服务中心，②阳光草坪，③森林小剧场，④儿童科普体验区，⑤儿童戏水场，⑥湿地博物馆，⑦湿地商业街。建立数学模型，找出从景石出发，到达⑦湿地商业街，并且经过①—⑥所有景点至少1次的距离最短的路线，计算该路线的长度，并将相关结果填入表格3。注：在每个景点不用停留。

**问题2：**如果某游客12:00从景石出发，要求他17:00前到达湿地商业街，17:30离开湿地商业街(注：根据表2的要求在湿地商业街游览时间至少为30分钟)。建立数学模型，为该游客设计一条能游览全部景点(景点①—⑦)且游览总时间最长的游览路线(假设在各个景点没有等待时间)，并完成表4的填写。

**问题3：**如果有3个旅游团，12:00同时从景石出发，要求三个旅游团17:00前到达湿地商业街，17:30离开湿地商业街(注：根据表2的要求在湿地商业街游览时间至少为30分钟)，并且每个景点(湿地商业街除外)同时只能容纳1个旅游团游览，按照时间顺序后到达的旅游团，需要等待先到达的旅游团游览结束之后才能开始游览。建立数学模型，为三个旅游团分别设计一条能游览全部7个景点且游览总时间最长的游览路线，并完成表5的填写。

**问题4：**假设3个旅游团的步行速度可以在 $1\text{km}/\text{h}$ 到 $3\text{km}/\text{h}$ 之间调节，但是总的平均步行速度不能超过 $2\text{km}/\text{h}$ ，3个旅游团12:00同时从景石出发，要求三个旅游团17:00前到达湿地商业街，17:30离开湿地商业街(注：根据表2的要求在湿地

商业街游览时间至少为 30 分钟), 并且每个景点 (湿地商业街除外) 同时只能容纳 1 个旅游团游览, 按照时间顺序后到达的旅游团, 需要等待先到达的旅游团游览结束之后才能开始游览。建立数学模型, 为三个旅游团分别设计一条能游览完全部 7 个景点且游览总时间长, 总的等待时间短的游览路线, 并完成表 6 的填写。

**问题 5:** 在现实中, 考虑如下两个不确定性因素

- (1) 不同旅游团从景石出发的时间具有不确定性, 例如, 多个旅游团在不同的时间从景石出发开始游览, 在此情况下到达湿地商业街的时间可以顺延。
- (2) 每个景点的等待时间也存在不确定性因素, 例如, 旅游设施短时间的维护和清理, 或者受到散客客流的影响。

考虑上述两个不确定性因素, 其它条件与问题 4 相同, 建立数学模型, 为多个旅游团分别设计一条能游览完全部 7 个景点且游览总时间长, 总的等待时间短的游览路线。

	景石	游客服务中心	阳光草坪	森林小剧场	儿童科普体验区	儿童戏水场	湿地博物馆	湿地商业街
景石	0	300	360	210	590	475	500	690
游客服务中心	300	0	380	270	230	285	200	390
阳光草坪	360	380	0	510	230	765	580	770
森林小剧场	210	270	510	0	470	265	450	640
儿童科普体验区	590	230	230	470	0	515	260	450
儿童戏水场	475	285	765	265	515	0	460	650
湿地博物馆	500	200	580	450	260	460	0	190
湿地商业街	690	390	760	640	450	650	190	0

表 1 景点之间的最短步行距离(单位: 米)

景点 \ 时间	游览时间	开放时间
游客服务中心	10-30 分钟	9:00-16:00
阳光草坪	20-60 分钟	9:00-17:00
森林小剧场	30 分钟	9:00, 9:30, 10:00, 10:30, 11:00, 11:30, 12:00, 12:30, 13:00, 13:30, 14:00, 14:30, 15:00, 15:30, 16:00, 16:30, 17:00 (半点和整点开放)
儿童科普体验区	30-60 分钟	9:00-17:00
儿童戏水场	20-60 分钟	9:00-17:00
湿地博物馆	30-60 分钟	9:00-17:00
湿地商业街	30 分钟以上	9:00-21:30

表 2 各景点游览时间

## 二、问题分析

对于问题 1, 是求解以景石为初始点, 湿地商业街为终止点, 经过①—⑥所有景点至少 1 次距离最短的路线, 问题可以转化为: 在给定的网络图中从起始点景石出发, 所有景点至少经过一次且回到最终回到湿地商业街, 使得总距离最小。考虑用建立整数规划模型, 结合 k 短路算法求解。

对于问题 2, 假设在各个景点没有等待时间的情况下, 求解某个游客游玩总时间最长的问题, 可以将此问题转化为: 在景区待的总时间一定和景区游览时间限制的情况下, 行人在路上和等待时间花费最短的问题。可以结合问题 1 的最短路线进

行求解，建立整数规划模型求解。

对于问题 3，是在考虑等待时间和限制条件的情况下，三个旅行团游览总时间最长的问题。可以将此问题转化为：三个旅行团在景区待的总时间一定和景区游览时间限制的情况下，三个旅行团在路上和等待时间花费最短的问题。可以结合问题 2 的最短路线进行求解。

对于问题 4，是在问题 3 的基础上增加约束条件：3 个旅游团的步行速度可以在  $1km/h$  到  $3km/h$  之间调节，但是总的平均步行速度不能超过  $2km/h$ 。问题 4 属于多目标规划问题，需在问题 3 的模型中多入目标函数和约束条件，转化为单目标函数问题。

对于问题 5，假设 4 个旅行团在以上问题限制条件的情况下，考虑从景石出发的时间和在每个景点等待的时间两个不确定性因素，求解这 4 个旅游团游览完全部景点，游览总时间最长，总等待时间短的游览路线。

### 三、问题假设

- 1.任意两个景点之间的最短步行距离如表 1 给出。
- 2.第二问、第三问假设步行速度  $V = 2km/h$ 。
- 3.游客在景区停留的时间由“景点之间的步行时间”、“景点游览时间（即在景区内游玩的时间）”和“在景区外的等待时间”三部分组成，其他时间忽略不计，游览时间必须符合表 2 的要求。

### 四、符号说明

$x_{ij}$	从节点 $i$ 到节点 $j$ 走过的次数
$w_{ij}$	节点 $i$ 到节点 $j$ 的距离
$W_{ij}$	$w_{ij}$ 元素构成的矩阵
$L$	问题 1 路线最短距离
$T$	除游览景点外的总时间
$t_1$	除游览景点外的总时间
$t_2$	等待时间
$r_{ij}$	$r_{ij} = \begin{cases} 1 & \text{游客直接从第 } i \text{ 个景点到达第 } j \text{ 个景点} \\ 0 & \text{其他} \end{cases}$
$\omega_{ij}$	游客从第 $i$ 个景点到第 $j$ 个景点需要等待的时间
$\omega_i$	表示游客在第 $i$ 个景点的游玩时间

$t_{ijk}$	旅行团 $k$ 从景点 $i$ 到景点 $k$ 的步行时间
$\omega_{ijk}$	旅行团 $k$ 从景点 $i$ 到景点 $k$ 需要等待的时间
$\varpi_{ijk}$	旅行团 $k$ 从景点 $i$ 到景点 $k$ 游览时间
$T_1$	3 个旅行团步行时间和等待时间之和
$r_{ijk}$	旅行团 $k$ 是否从第 $i$ 个景点直接到第 $j$ 个景点
$p$	旅游设施的维护和清理时间
$q$	散客客流影响的时间
$Y$	旅行团 $k$ 总的等待时间减去旅行团 $k$ 总的游览时间

## 五、模型建立与问题求解

### 5.1 问题 1 的模型建立与求解

这八个景点和其交通线路构成了一个网络图，这些景点可以看做该网络图的节点，这些节点则由相应的交通线路连接，问题一求解以节点 8 为初始点，节点 7 为终止点，本题所要实现的目标是：经过①—⑥所有景点至少 1 次距离最短的路线，考虑到路径中可能出现环路，现用整数规划建立以下模型。

#### 5.1.1 目标函数的建立

设  $x_{ij}$  为最终行驶路径中，从节点  $i$  到节点  $j$  走过的次数，其中  $1 \leq i, j \leq 8$ ，且为整数， $w_{ij}$  为节点  $i$  到节点  $j$  的距离，建立目标函数：

$$\text{Min } L = \sum_{i=1}^8 \sum_{j=1}^8 x_{ij} w_{ij}$$

其中矩阵  $W_{ij}$  如下：

$$W_{ij} = \begin{bmatrix} 0 & 380 & 270 & 230 & 285 & 200 & 390 & 300 \\ 380 & 0 & 510 & 230 & 765 & 580 & 760 & 360 \\ 270 & 510 & 0 & 470 & 265 & 450 & 640 & 210 \\ 230 & 230 & 470 & 0 & 515 & 260 & 450 & 590 \\ 285 & 765 & 265 & 515 & 0 & 460 & 650 & 475 \\ 200 & 580 & 450 & 260 & 460 & 0 & 190 & 500 \\ 390 & 770 & 640 & 450 & 650 & 190 & 0 & 690 \\ 300 & 360 & 210 & 590 & 475 & 500 & 690 & 0 \end{bmatrix}$$

#### 5.1.2 约束条件

根据所有景点都需要至少游览一次，以及最终线路要形成路径，可得到以下约

束条件;

$$\left\{ \begin{array}{l} \sum_{j=1}^8 x_{ij} \geq 1 \quad i = 1,2,3,4,5,6,8 \quad \text{①} \\ \sum_{j=1}^8 x_{ij} = \sum_{j=1}^8 x_{ji} \quad i = 1,2,3,4,5,6 \quad \text{②} \\ \sum_{j=1}^8 x_{8j} - 1 = \sum_{i=1}^8 x_{i8} \quad \text{③} \\ \sum_{j=1}^8 x_{7j} + 1 = \sum_{i=1}^8 x_{i7} \quad \text{④} \\ x_{ij} \geq 0 \quad x_{ij} \in \mathbf{N}, \quad 1 \leq i, j \leq 8 \quad \text{⑤} \end{array} \right.$$

③式表示除起点和终点外，进入景点的次数和从景点出来的次数相等；④式表示从节点 8 出去的次数比进入起点的次数多 1，这保证节点 8 是起点；⑤式表示从节点 7 出去的次数比进入节点 7 的次数少 1，这保证节点 7 是终点。

根据问题一中的约束条件，利用 MATLAB 模拟，得到如下结果：

景石→③→⑤→①→⑥→⑦，②→④，④→②

此方案总距离为 1610 米，存在一条闭回路，而路径如果要连接此闭回路，至少再增加  $(230 + 260 - 200)$  米的路程，保留此结果，考虑不存在②→④，④→②这样一条闭回路的方案，添加约束条件

$$x_{24} \neq x_{42}$$

利用 matlab 进行一次计算，得到如下方案：

景石→②→④→①→⑥→⑦，③→⑤，⑤→③

此方案总距离为 1740 米，存在一条闭回路，而路径如果要连接此闭回路，至少要再增加  $(270 + 285 - 265)$  米的路程，保留此结果，考虑不存在③→⑤，⑤→③这样一条闭回路的方案，添加约束条件

$$x_{35} \neq x_{53}$$

利用 matlab 再进行一次计算，得到如下方案：

景石→③→⑤→①→②→④→⑥→⑦

此方案总路程 1820 米，为不包含③→⑤，⑤→③或者②→④，④→②闭回路的最优解，较之于存在闭回路的方案，此方案更优，故 1820 为此问题的全局最优解。最终的 matlab 程序由附录 1 给出，方案如下图 1 所示：



图 1 问题 1 最优路线图

具体方案见下表 3

出发点	到达景点	步行距离(米)
景石	森林小剧场	210
森林小剧场	儿童戏水场	265
儿童戏水场	游客服务中心	285
游客服务中心	阳光草坪	380
阳光草坪	儿童科普体验区	230
儿童科普体验区	湿地博物馆	260
湿地博物馆	湿地商业街	190
总步行距离(最短路线距离)		1820
最短路线(请用①~⑥序号标出)		景石→③→⑤→①→②→④→⑥→⑦

表 3 问题 1 最优路线表

## 5.2 问题 2 的模型建立与求解

由问题 1 得出的最短路线：景石→③→⑤→①→②→④→⑥→⑦，是在不考虑等待时间和限制条件的情况下，在路上行走所花费的时间最短。经过对问题 2 的分析，可知本题所要实现的目标是：游客在 17:00 前到达湿地商业街，17:30 离开湿地商业街(注：根据表 2 的要求在湿地商业街游览时间至少为 30 分钟)下，游客游览全部景点(景点①—⑦)且游览总时间最长的游览路线。可将问题转化为游客步行时间和等待时间之和最短的问题。

### 5.2.1 目标函数的建立

在不考虑其它因素的情况下，除游览景点外的时间由两部分组成，分别为游客步行时间和等待时间，只有游客步行时间和等待时间之和最短，才能保证游览总时间最长，下面定义：

T——除游览景点外的总时间

$t_1$ ——游客步行时间

$t_2$ ——等待时间

从而得到目标函数： $\text{Min } T = t_1 + t_2$

### (1) 游客步行时间

因为 $t_{ij}$ 表示从第*i*个景点到第*j*个景点所需的时间， $w_{ij}$ 表示从第*i*个景点到第*j*个景点之间的距离，所以 $t_{ij} = w_{ij} / v$ 而 $r_{ij}$ 是判断游客是否从第*i*个景点直接到第*j*个景点的 0-1 变量，因此可以得出游客步行时间为：

$$t_1 = \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij}}{v}$$

### (2) 等待时间

$r_{ij}$ 也可以判断游客是否从第*i*个景点直接到第*j*个景点， $\omega_{ij}$ 表示游客从第*i*个景点到第*j*个景点需要等待的时间，由问题一可知，游客不会重复游览某个景点，因此 $\sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij} r_{ij}$ 实际上将游客需要等待的时间计算了两遍，因此可以得出游客等待时间为：

$$t_2 = \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij} r_{ij}$$

从而可以得到目标函数：

$$\text{Min } T = t_1 + t_2$$

$$= \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij}}{v} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij} r_{ij}$$

## 5.2.2 约束条件

### (1) 时间约束

如果游客 12:00 从景石出发，17:30 离开湿地商业街，旅游时间一共为 330 分钟，而这些时间包括游客步行时间、景点等待时间和在景点游览时间。因为 $t_{ij}$ 表示从第*i*个景点到第*j*个景点步行的时间， $\omega_{ij}$ 表示游客从第*i*个景点到第*j*个景点需要等待的时间， $\omega_i$ 表示游客在第*i*个景点的游玩时间，故游客在景点的总游玩时间为：

$\frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 r_{ij} \times (\omega_i + \omega_j)$ ，要求游客 17:00 前到达湿地商业街，时间约束为：

$$\frac{1}{2} \sum_{i=1}^7 \sum_{j=1}^7 r_{ij} \times (\omega_i + \omega_j) + \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij}}{v} + \frac{1}{2} \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij} \leq 300$$

(2) 0—1 变量约束

游客不可能在两个景点之间往返旅游，因为这样显然不满足游览时间尽量多的原则，因此可得约束  $r_{ij} \times r_{ji} = 0$  ( $i, j = 1, 2, 3, 4, 5, 6$ )，除起点和终点外，进入景点的次数和从景点出来的次数相等；可得约束： $\sum_i r_{ij} = \sum_j r_{ij} \leq 1$  ( $i, j = 1, 2, \dots, 8$ )，

当  $i=8$  时，即从起始点景石出发，所以  $\sum_{i=8} r_{ij} = 1$ ；

当  $j=7$  时，即回到终点湿地商业街，所以  $\sum_{j=7} r_{ij} = 1$ ；

综上所述，可以得到总的模型为：

$$\text{Min } T = t_1 + t_2$$

$$= \sum_{i=1}^8 \sum_{j=1}^8 t_{ij} r_{ij} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij} r_{ij}$$

$$\text{s.t.} \left\{ \begin{array}{l} \frac{1}{2} \sum_{i=1}^7 \sum_{j=1}^7 r_{ij} \times (\omega_i + \omega_j) + \sum_{i=1}^8 \sum_{j=1}^8 t_{ij} r_{ij} + \frac{1}{2} \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij} \leq 300 \\ \sum_i r_{ij} = \sum_j r_{ij} \leq 1 (i, j = 1, 2, \dots, 8) \\ r_{ij} \times r_{ji} = 0 (i, j = 1, 2, 3, 4, 5, 6) \\ \sum_{i=8} r_{ij} = 1 \\ \sum_{j=7} r_{ij} = 1 \end{array} \right.$$

根据问题 2 的模型,利用 MATLAB 编程 (程序见附录 2)，得到方案：景石 → ② → ④ → ③ → ⑤ → ① → ⑥ → ⑦，如图 2 所示：

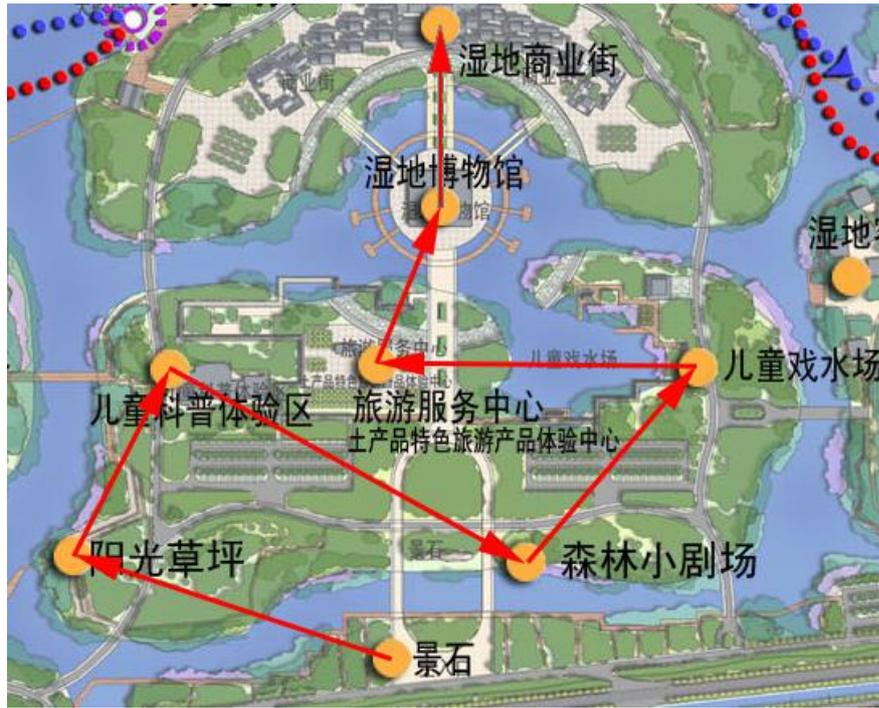


图 2 问题 2 最优路线图

具体方案见下表 4:

序号	景点名称	达到时间点	游览时间 (停留时间,单位分钟)	离开时间点
1	景石	12:00	0	12:00
2	阳光草坪	12:10:48	30	12:40:48
3	儿童科普体验区	12:47:42	58:12	13:45:54
4	森林小剧场	14:00:00	30	14:30:00
5	儿童戏水场	14: 37:57	44:30	15:21:27
6	游客服务中心	15:30	30	16:00
7	湿地博物馆	16:06	48:18	16:54:18
8	湿地商业街	17:00:00	30	17:30
总的游览时间:		270 分钟		
总的步行时间:		60 分钟		

表 4 问题 2 最优路线表

### 5.3 问题 3 的模型建立与求解

由问题 2 得出的最短路线: 景石→②→④→③→⑤→①→⑥→⑦, 是在考虑等待时间和限制条件的情况下, 一个游客游览总时间最长的问题。经过对问题 3 的分析, 可知本题所要实现的目标是: 3 个旅游团, 12:00 同时从景石出发, 在条件 17:00 前到达湿地商业街, 17:30 离开湿地商业街(注: 根据表 2 的要求在湿地商业街游览时间至少为 30 分钟)下, 游览完全部景点(景点①—⑦)且游览总时间最长的游览路线。可将问题转化为 3 个旅行团步行时间和等待时间之和最短的问题。5.3.1 目标函数的建立

下面定义:

$t_{ijk}$  ——旅行团 k 从景点 i 到景点 k 的步行时间

$\omega_{ij_k}$ ——旅行团 k 从景点 i 到景点 k 需要等待的时间

$\varpi_{ij_k}$ ——旅行团 k 从景点 i 到景点 k 游览时间

$T_1$ ——3 个旅行团步行时间和等待时间之和

### (1) 游客步行时间

因为  $t_{ij_k}$  表示旅行团 k 从第 i 个景点到第 j 个景点所需的时间， $w_{ij}$  表示从第 i 个景点到第 j 个景点之间的距离，所以  $t_{ij_k} = \frac{w_{ij}}{v}$ ，而  $r_{ij_k}$  是判断旅行团 k 是否从第 i 个景点直接到第 j 个景点的 0-1 变量，因此可以得出旅行团 k 步行时间为：

$$t_k = \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij_k}}{v} \quad (k=1,2,3)$$

### (2) 等待时间

$\omega_{ij_k}$  表示旅行团 k 从第 i 个景点到第 j 个景点需要等待的时间， $r_{ij_k}$  也可以判断旅行团 k 是否从第 i 个景点直接到第 j 个景点，结合问题 2 可以得出旅行团 k 等待

时间为  $\frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k}$

从而可以得到目标函数：

$$\begin{aligned} \text{Min } T_1 &= \sum_{k=1}^3 t_k + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k} \\ &= \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij_k}}{v} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k} \end{aligned}$$

## 5.3.2 约束条件

### (1) 时间约束

3 个旅行团 12:00 从景石出发，17:30 离开湿地商业街，旅游时间一共为 330 分钟，同样这些时间包括 3 个旅行团步行时间、景点等待时间和在景点游览时间。因为  $t_{ij_k}$  表示旅行团 k 从第 i 个景点到第 j 个景点的步行时间， $\omega_{ij_k}$  表示旅行团 k 从第 i

个景点到第 j 个景点需要等待的时间， $\varpi_{i_k}$  表示旅行团 k 在第 i 个景点的游玩时间，

故旅行团 k 在景点的总游玩时间为： $\frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 r_{ij_k} \times (\varpi_{i_k} + \varpi_{j_k})$ ，要求 3 个旅行团 17:00

前到达湿地商业街，时间约束为：

$$\frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 r_{ij_k} \times (\varpi_{i_k} + \varpi_{j_k}) + \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij_k} r_{ij_k}}{v} + \frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij_k} \leq 300$$

(2) 0—1 变量约束

正如问题 2 中所述，任何一个旅行团也不可能在两个景点之间往返旅游，因此可得约束  $r_{ij_k} \times r_{ji_k} = 0$  ( $k=1,2,3, i, j=1,2,3,4,5,6$ )，除起点和终点外，每个旅行团进入景点的次数和从景点出来的次数相等；可得约束：

$$\sum_i r_{ij_k} = \sum_j r_{ij_k} \leq 1 \quad (i, j=1,2,\dots,8),$$

当  $i=8$  时，即从起始点景石出发，所以  $\sum_{i=8} r_{ij_k} = 1$ ；

当  $j=7$  时，即回到终点湿地商业街，所以  $\sum_{j=7} r_{ij_k} = 1$ ；

综上所述，可以得到总的模型为：

$$\begin{aligned} \text{Min } T_1 &= \sum_{k=1}^3 t_k + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k} \\ &= \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij_k} r_{ij_k}}{v} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k} \end{aligned}$$

S.t.

$$\frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 r_{ij_k} \times (\varpi_{i_k} + \varpi_{j_k}) + \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij_k} r_{ij_k}}{v} + \frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij_k} \leq 300$$

$$r_{ij_k} \times r_{ji_k} = 0 \quad (k=1,2,3, i, j=1,2,3,4,5,6)$$

$$\sum_i r_{ij_k} = \sum_j r_{ij_k} \leq 1 \quad (i, j=1,2,\dots,8)$$

$$\sum_{i=8} r_{ij_k} = 1$$

$$\sum_{j=7} r_{ij_k} = 1$$

根据问题 3 的模型,修改附录 2 中的程序得到程序 3 (附录 2)，运行得到以下方案：

旅行团 1：景石→②→④→③→⑤→①→⑥→⑦

旅行团 2：景石→①→⑤→③→②→④→⑥→⑦

旅行团 3：景石→⑤→③→②→④→①→⑥→⑦

如图 3 所示：

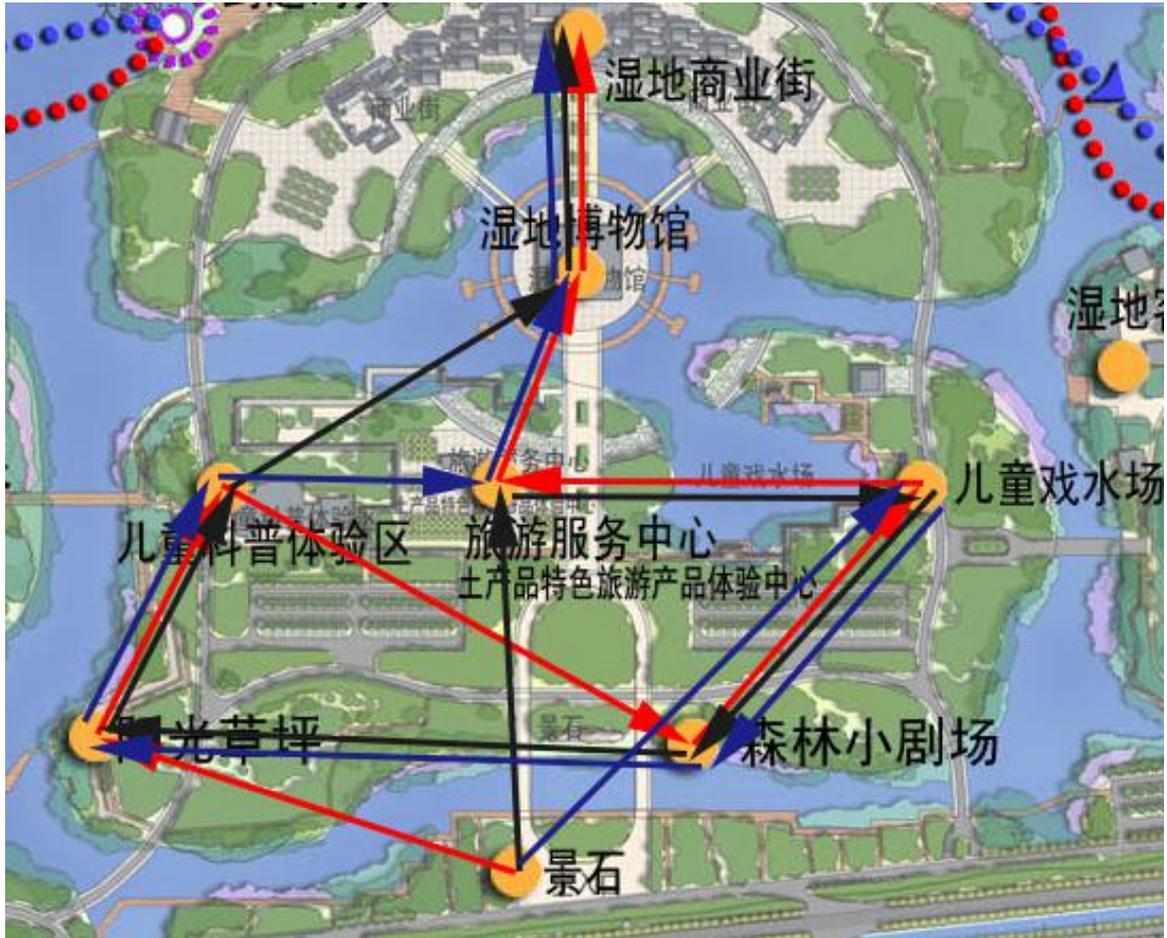


图 3 问题 3 最优路线图

具体方案见下表 5

	第一旅游团				第二旅游团				第三旅游团		
	到达时间 点	游览时间 点(停留时 间,单位分 钟)	离开时间 点		到达时间 点	游览时间 点(停留时 间,单位分 钟)	离开时间 点		到达时间 点	游览时间 点(停留时 间,单位分 钟)	离开时间 点
景石	12:00	0	12: 00	景石	12:00	0	12:00	景石	12:00	0	12:00
阳光草坪	12:10:48	30	12:40:48	游客服务 中心	12:09	30	12:39	儿童 戏水场	12:14:15	37.8	12:52:03
儿童科普 体验区	12:47:42	58.2	13:45:54	儿童 戏水场	12:47:33	34.5	13:22:03	森林 小剧场	13:00	30	13:30
森林 小剧场	14:00	30	14: 30	森林 小剧场	13:30	30	14:00	阳光草坪	13:45:18	20	13:05:18
儿童 戏水场	14:37:57	20	14:57:57	阳光草坪	14:15:18	35.2	14:50:30	儿童科普 体验区	14:12:12	30	14:42:12
游客服务 中心	15:06:30	23.5	15:30	儿童科普 体验区	14:57:24	60	15:57:24	游客服务 中心	14:49:06	10	14:59:06
湿地 博物馆	15:36	30	16:05	湿地 博物馆	16:17	30	16:47	湿地 博物馆	15:05:06	30	15:35:06

湿地 商业街	16:11:42	78.3	17:30	湿地 商业街	16:40:54	49.1	17:30	湿地 商业街	15:40:48	109.2	17:30
总步行时间	60			总步行时间	61.2			总步行时间	63		
总游览时间	270			总游览时间	268.8			总游览时间	267		
总等待时间	0			总等待时间	0			总等待时间	0		

表 5 问题 3 最优路线表

## 5.4 问题 4 的模型建立与求解

问题 4 是在问题 3 的基础上,增加条件: 3 个旅游团的步行速度可以在  $1km/h$  到  $3km/h$  之间调节, 但是总的平均步行速度不能超过  $2km/h$ 。

本题目属于多目标规划问题, 只需在问题 3 的模型中多加目标函数和速度约束条件, 为等待时间函数与游览时间的函数添加不同的权重 1 和-1, 将问题转化为单目标函数。

### 5.4.1 目标函数的建立

符号  $Y$  表示旅行团  $k$  总的等待时间减去旅行团  $k$  总的游览时间, 在  $k$  最小的情况下, 才能保证三个旅行团总的游览时间最长。

$$\text{Min } Y = \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k} - \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij_k} r_{ij_k}$$

### 5.4.2 约束条件

只需在问题 3 的约束条件中加入约束条件:

$$1 \leq v \leq 3 \text{ 且 } v_{\text{平均}} \leq 2$$

$$\text{s.t.} \left\{ \begin{array}{l} \frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 r_{ij_k} \times (\omega_{i_k} + \omega_{j_k}) + \sum_{k=1}^3 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij_k} r_{ij_k}}{v} + \frac{1}{2} \sum_{k=1}^3 \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij_k} \\ \sum_i r_{ij_k} = \sum_j r_{ij_k} \leq 1 \quad (i, j = 1, 2, \dots, 8) \\ \sum_{i=8} r_{ij_k} = 1 \\ \sum_{j=7} r_{ij_k} = 1 \\ 1 \leq v \leq 3 \text{ 且 } v_{\text{平均}} \leq 2 \end{array} \right.$$

根据问题 4 的模型,修改附录 3 中的程序得到程序 4 (附录 4), 运行得到以下方案:

旅行团 1: 景石 → ② → ④ → ③ → ⑤ → ① → ⑥ → ⑦

旅行团 2: 景石 → ① → ⑤ → ③ → ② → ④ → ⑥ → ⑦

旅行团 3: 景石 → ⑤ → ③ → ① → ② → ④ → ⑥ → ⑦

具体方案见下表 6:

	第一旅游团					第二旅游团					第三旅游团			
	达到时间点	游览时间	离开时间	达到下一个景点的步行速度		达到时间点	游览时间	离开时间	达到下一个景点的步行速度		达到时间点	游览时间	离开时间	达到下一个景点的步行速度
景石	12: :00	0	12:00	2	景石	12: :00	0	12:00	2	景石	12: :00	0	12:00	2
阳光草坪	12:10:48	30	12:40:48	2	游客服务中心	12:09	30	12:39	2	儿童戏水场	12:14:15	37.8	12:52:03	2
儿童科普体验区	12:47:42	58.2	13:45:54	2	儿童戏水场	12:47:33	34.5	13:22:03	2	森林小剧场	13:00	30	13:30	3
森林小剧场	14:00	30	14:30	2	森林小剧场	13:30	30	14:00	1.645	游客服务中心	13:35:24	10	13:45:24	3
儿童戏水场	14:37:57	20	14:57:57	2	阳光草坪	14:18:36	31.9	14:50:30	2	阳光草坪	13:52:36	20	14:18:36	1.45
游客服务中心	15:06:30	23.5	15:30	2	儿童科普体验区	14:57:24	60	15:57:24	3.25	儿童科普体验区	14:26:42	31.3	14:58	1.25
湿地博物馆	15:36	30	16:05	2	湿地博物馆	16:13:42	33.3	16:47	2	湿地博物馆	15:35	30	16:05	24
湿地商业街	16:11:42	78.3	17:30	0	湿地商业街	16:40:54	49.1	17:30	0	湿地商业街	16:08:45	81.25	17:30	0
总步行时间	60 分钟				总步行时间	61.2 分钟				总步行时间	62.1 分钟			
总游览时间	270 分钟				总游览时间	268.8 分钟				总游览时间	267.9 分钟			
总等待时间	0 分钟				总等待时间	0 分钟				总等待时间	0 分钟			
平均速度	2Km/h				平均速度	2Km/h				平均速度	2Km/h			

表 6 问题 4 最优路线表

## 5.5 问题 5 的模型建立与求解

在这里假设有 4 个旅行团 ABCD, 有两个不确定性因素: 从景石出发的时间; 每个景点等待的时间 (比如: 旅游设施短时间的维护和清理, 或者受到散客客流的影响)。在问题 4 中, 我们可知旅游团的步行速度可以在  $1\text{km}/\text{h}$  到  $3\text{km}/\text{h}$  之间调节, 但是总的平均步行速度不能超过  $2\text{km}/\text{h}$ , 4 个旅游团从景石出发, 要求 4 个旅游团 17:00 前到达湿地商业街, 17:30 离开湿地商业街 (注: 根据表 2 的要求在湿地商业街游览时间至少为 30 分钟), 并且每个景点 (湿地商业街除外) 同时只能容纳 1 个旅游团游览, 按照时间顺序后到达的旅游团, 需要等待先到达的旅游团游览结束之后才能开始游览。要求为 4 个旅游团分别设计一条能游览完全部 7 个景点且游览总时间长, 总的等待时间短的游览路线。

### 5.5.1 条件假设

- (1) 假设 4 个旅行团两两出发的时间间隔少于 20 分钟。
- (2) 假设旅游设施的维护和清理时间为  $p$  分钟, 散客客流影响的时间为  $q$  分钟。
- (3) 假设旅行团的步行速度为  $v \text{ km/h}$ , 其中  $1 \leq v \leq 3$ 。
- (4) 旅行团从景石出发的时间在 9:00 以后。
- (5) 除以上条件外, 其它影响因素忽略不计。

### 5.5.2 目标函数的建立

#### (1) 游客步行时间

因为  $t_{ij\theta}$  表示旅行团  $\theta$  从第  $i$  个景点到第  $j$  个景点所需的时间,  $w_{ij}$  表示从第  $i$  个景点到第  $j$  个景点之间的距离, 所以  $t_{ij\theta} = w_{ij} / v$  而  $r_{ij\theta}$  是判断旅行团  $k$  是否从第  $i$  个景点直接到第  $j$  个景点的 0-1 变量, 因此可以得出旅行团  $k$  步行时间为:

$$t_{\theta} = \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij\theta}}{v} \quad (\theta = A, B, C, D)$$

#### (2) 等待时间

考虑到 4 个旅行团在旅游线路上可能存在重合而造成在同一个景点处产生等待时间, 目标函数中还应该包括每个旅行团在各个景点处可能产生的等待时间。这个等待时间可由每个景点处的等待时间累计而得到。假设

$\omega_{ij\theta}$  表示旅行团  $k$  从第  $i$  个景点到第  $j$  个景点需要等待的时间,  $r_{ij\theta}$  也可以判断旅行团  $k$  是否从第  $i$  个景点直接到第  $j$  个景点, 结合问题 2 可以得出旅行团  $k$  等待时间

$$\text{为 } \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij\theta} r_{ij\theta}$$

从而可以得到目标函数:

$$\begin{aligned} \text{Min } T_1 &= \sum_{\theta=1}^4 t_{\theta} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij\theta} r_{ij\theta} + p + q \\ &= \sum_{\theta=1}^4 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij\theta}}{v} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij\theta} r_{ij\theta} + p + q \end{aligned}$$

### 5.5.2 约束条件

#### (1) 时间约束

由条件可知，4 个旅行团以不同的时间从景石出发，要求 4 个旅游团 17:00 前到达湿地商业街，17:30 离开湿地商业街。因为  $t_{ij\theta}$  表示旅行团  $k$  从第  $i$  个景点到第  $j$  个景点的步行时间， $\omega_{ij\theta}$  表示旅行团  $k$  从第  $i$  个景点到第  $j$  个景点需要等待的时间， $\omega_{i\theta}$  表示旅行团  $\theta$  在第  $i$  个景点的游玩时间，故旅行团  $\theta$  在景点的总游玩时间为：

$\frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 r_{ij\theta} \times (\omega_{i\theta} + \omega_{j\theta})$ ，要求 4 个旅行团 17:00 前到达湿地商业街，时间约束为：

$$\frac{1}{2} \sum_{\theta=1}^4 \sum_{i=1}^8 \sum_{j=1}^8 r_{ij\theta} \times (\omega_{i\theta} + \omega_{j\theta}) + \sum_{\theta=1}^4 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij\theta}}{v} + \frac{1}{2} \sum_{\theta=1}^4 \sum_{i=1}^7 \sum_{j=1}^7 \omega_{ij} r_{ij} + p + q \leq 480$$

#### (2) 0—1 变量约束

正如问题 2 中所述，任何一个旅行团也不可能在两个景点之间往返旅游，因此可得约束  $r_{ij\theta} \times r_{ji\theta} = 0$  ( $\theta = A, B, C, D$  ;  $i, j = 1, 2, 3, 4, 5, 6$ )，除起点和终点外，每个旅行团进入景点的次数和从景点出来的次数相等；可得约束：

$$\sum_i r_{ij\theta} = \sum_j r_{ij\theta} \leq 1 \quad (i, j = 1, 2, \dots, 8),$$

当  $i=8$  时，即从起始点景石出发，所以  $\sum_{i=8} r_{ij\theta} = 1$ ；

当  $j=7$  时，即回到终点湿地商业街，所以  $\sum_{j=7} r_{ij\theta} = 1$ ；

综上所述，可以得到总的模型为：

$$\begin{aligned} \text{Min } T_1 &= \sum_{k=1}^3 t_k + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij} r_{ij\theta} + p + q \\ &= \sum_{\theta=1}^4 \sum_{i=1}^8 \sum_{j=1}^8 \frac{w_{ij} r_{ij\theta}}{v} + \frac{1}{2} \sum_{i=1}^8 \sum_{j=1}^8 \omega_{ij\theta} r_{ij\theta} + p + q \end{aligned}$$

### 5.5.4 模型得求解与结果分析

根据模型，运用 matlab 编程，在假设条件允许的情况下，得出 4 个旅行团的最优游览路线为：

- A: 景石 → ③ → ⑤ → ① → ② → ④ → ⑥ → ⑦
- B: 景石 → ② → ④ → ③ → ⑤ → ① → ⑥ → ⑦
- C: 景石 → ⑤ → ③ → ① → ② → ④ → ⑥ → ⑦
- D: 景石 → ① → ⑤ → ③ → ② → ④ → ⑥ → ⑦

## 5.6 模型的评价、改进及推广

### 5.6.1. 模型的评价

1. 本文思路清晰，模型恰当，得出的方案合理；

- 2.本文成功的使用了 0—1 变量，使模型的建立和编程得以顺利进行；
- 3.在第 2 题中采用了 k 短路算法，简化了模型的求解难度；

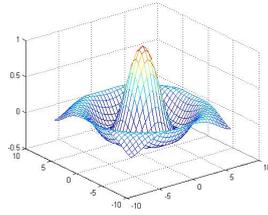
#### 5.6.2. 模型的改进与推广

- 1.对于问题 5 在实际情况中，不确定因素太多，尽可能增加这些考虑，结果会更加合理。
- 2.因数据资料搜集的不完整，准确性也有待商榷，而且没有对最终方案进行更为细致的讨论研究，这些方面有待改进。

#### 参考文献：

- [1]姜启源 谢金星 叶俊，《数学模型（第三版）》，北京：高等教育出版社，2003。
- [2]谢金星 薛毅，《优化建模与 LINDO/LINGO 软件》，北京：清华大学出版社，2005。
- [3]周仁郁，《SPSS13.0 统计软件》，成都，西南交通大学出版社，2005。
- [4] 周康，强小莉，同小军，许进，《求解 TSP 算法》，计算机工程与应用，2009，43（29），43-48.
- [5] 吴斌，史忠植，《一种基于蚁群算法的 TSP 问题分段求解算法》，计算机学报，2001,24（12）：1—6.

# 五一数学建模竞赛



## 题目 商业银行人民币贷款规模分配及盈利问题

**关键词：**主成份回归分析 多目标规划模型 存贷比 流动性风险

**摘要：**本文通过 AR 预测模型、主成份回归分析、目标规划法，建立了一个考虑收益水平、流动性风险、均衡地区间发展差异的多目标非线性贷款规模分配规划模型。对商业银行的贷款规模分配方案结果见附录一。

针对问题一，首先运用 AR 自回归预测模型对数据进行了宏观经济指标缺失值的处理，接着进行量纲一致化的预处理过程，为了预测 2018 年的存贷款增量，我们先通过分析宏观经济指标和存贷款增量的历史数据，由于宏观经济指标数量较多，所以运用主成份回归分析，建立存贷款增量和宏观经济指标之间的关系式，最后结合 2018 年的宏观经济指标的预测值，得到存贷款增量的预测值。以北京市为例，2018 年存款增量为 353 亿元，贷款增量为 99 亿元。为了对模型的结果有进一步的了解，我们根据存贷比的高低和变化趋势对 30 个地区进行了分类。

针对问题二，首先对 2018 年的存贷款利率相对应的宏观经济指标进行筛选，接下来运用多元线性回归最小二乘估计方法进行了 2018 年利率的预测。针对贷款规模的分配问题，除了使收益最大化，还要考虑贷款总额度约束条件和平衡地区差异等条件，所以运用目标规划法，建立一个多目标线性规划模型，并引入偏差变量  $d^+$ ,  $d^-$  描述实际值和目标值之间的偏差。以北京市为例，求解结果为 2018 年贷款分配额为 78.9 亿元。

针对问题三，首先根据 30 个区域的存贷比进行聚类，从而确定不同区域的债券利率。我们将债券分配和贷款分配看成是两个独立的过程，分别进行最佳分配方案的求解。对于这两个问题，均采用多目标规划的方法建立模型，运用 lingo 编程求解。

针对问题四，首先对附件 4 的数据进行预处理，运用 Box-Cox 变换对数据进行处理，得到每日存取款交易数据的分布。在分析取款金额和备付金额之间的关系，得到备付资金满足正态分布的结果，根据计算机模拟求解结果。

针对问题五，综合考虑收益水平、流动性风险和均衡地区之间发展差异三个方面的因素，基于对前面问题所建立的模型进行改进，量化了流动性风险和均衡地区差异两个目标，最后建立起一个综合地确定最优贷款规模分配方案的多目标非线性规划模型。

## 一、问题重述

商业银行贷款投放的简单模型是：从客户端吸收存款，缴存法定准备金（法定准备金率：大型金融机构 15.5%，中小金融机构 12%；法定准备金利率 1.62%），预留一定比例备付水平（商业银行承担支付结算金融中介作用，满足流动性安全），剩余资金用于贷款投放或其他资产配置。

现有某中型商业银行 A，其各家省级分行近三年存贷款业务发展情况见附件 2。请查阅相关文献和资料，在对商业银行经营模式有所掌握，对未来经济、利率、汇率等政策及趋势有充分了解和预判的基础上回答以下问题。

1. 假设该银行除客户存款外无其他资金来源，且暂不考虑备付水平。请根据附件 2 商业银行 A 各项存贷款历史数据及附件 1 宏观经济指标历史数据，建立数学模型，预测该银行 2018 年存、贷款增量情况。
2. 假设该银行除客户存款外无其他资金来源，且暂不考虑备付水平。请根据问题 1 预测的结果并结合附件 3 相关数据，建立数学模型，给出 2018 年商业银行 A 各分行贷款规模的分配方案，使得全行增量存贷款利息净收入最大，并将该分配方案填入表 1。
3. 若商业银行 A 将于 2018 年 5 月 1 日发行 500 亿规模的 15 年期商业银行普通债（利率约为 5.1%），请结合该条件对问题 2 进一步优化，暂不考虑备付水平情况下，重新设计商业银行 A 各分行贷款规模的分配方案，使得全行增量存贷款利息净收入最大，并将该分配方案填入表 1。
4. 为保证每日交易正常进行，各家分行每日需预留一定的备付资金（备付资金不足易引起客户不满，严重的会引起社会恐慌，引发挤兑；预留资金过多，会降低银行盈利水平），以确保最低的备付水平（备付水平=备付资金÷存款余额）。假设每个客户存取款的行为是随机的，请根据附件 4 各分行 2017 年每日存取款交易数据，建立数学模型，计算在置信水平 99%的情况下，2018 年商业银行 A 各分行日常经营所需最低备付金额，并将结果填入表 1。
5. 在贷款规模分配问题上，为了帮助商业银行 A 处理好收益与风险、企业经营与国际政策导向，区域化差异与分行公平考核等之间的关系，以期达到双赢或多赢，请对以上模型进行改进，并给出相关建议。

## 二、基本假设

- 假设一：贷款、备付金额分配好后年中不再进行各分行间的调配；
- 假设二：贷款全部能收回；
- 假设三：债券可以全部按照面值发放出去；
- 假设四：存贷款利率年初确定后不再进行调整；

## 三、问题分析

### 3.1 对于问题一的分析

问题一要求预测该银行 2018 年存、贷款增量情况，由附件历史数据得，宏观经济指标和存、贷款量，两组变量之间存在多重相关关系，并且附件只提供了三年（一年一组）的样本数据。对于这种研究多个变量和多个变量之间的相关联系，且样本数量较少，

宜采用偏最小二乘回归方法。用该方法来求解该银行存、贷款增量与宏观经济指标之间的回归方程式，最后结合 2018 年的宏观经济指标的预测量求得该银行 2018 年存、贷款增量情况。

### 3.2 对于问题二的分析

问题二要求建立模型，给出贷款的分配方案。存贷款利率和多项指标有关，首先剔除关联度较小的指标，再通过最小二乘法求得利率的预测值。贷款的分配方案是一个多目标优化模型，我们引入线性规划模型以及偏量  $d^+, d^-$  对问题进行优化，约束条件中存在刚性约束和柔性约束，接下来确定优先级以及权重系数，建立算法，最后通过 lingo 程序得到最优解。

### 3.3 对于问题三的分析

问题三在问题二的基础上增加了发行 500 亿债券的条件，需要对问题进行进一步优化，得到最优的贷款分配方案。对于发行的 500 亿债券，其利率不是定制，各分行间债券利率受到多种因素影响，通过聚类分析的方法，以存贷比为分类指标，将各省分行间的利率进行分类。再基于问题二的线性规划模型，增加多个约束条件，调整目标函数，确定权重系数，建立算法，运用 lingo、matlab 程序求得最优解。

### 3.4 对于问题四的分析

问题四要求根据附件 4 中各分行 2017 年每日存取款交易数据，建立模型计算 2018 年商业银行 A 各分行的备付金额。首先，对附件 4 中数据进行处理，得到每日存取款交易数据的分布，接下来分析备付资金与存取款金额的关系，建立备付资金模型。设计模拟算法，求解备付资金模型，得到各分行备付资金。最后对备付资金求解结果进行检验，评价所建立的模型和算法的优劣性。

## 四、符号说明

1.  $M$ ：总金额
2.  $m$ ：各省分行金额
3.  $\alpha$ ：各省分行利率
4.  $\hat{M}$ ：预测总金额
5.  $\hat{m}$ ：各省分行金额预测值
6.  $\hat{\alpha}$ ：各省分行利率预测值

注：上标  $(c)$  表示存款， $(d)$  表示贷款， $(q)$  表示取款， $(b)$  表示备付金额， $(z)$  表示债券分配量。

## 五、问题一的模型的建立和求解

### 5.1 数据预处理

对于附件给出的 2010-2017 年的宏观经济指标和 2018 年的指标预测值，存在一部分数据的缺失，主要是 2017 年 12 月部分数据的缺失和 2018 年预测值的缺失，这部分对缺失值进行了填补。同时，还对数据进行了标准化处理。

#### 5.1.1 缺失值的处理

附件一的数据缺失包含是以下三种情况：

- (1) 缺少 2017 年 12 月数据，有 2018 年的预测数据

(2) 有 2017 年 12 月数据，缺少 2018 年的预测数据

(3) 同时缺少 2017 年 12 月数据和 2018 年的预测数据

对于情况 (1)，用缺失数据两边数据的平均值代替缺失值；

对于情况 (2) 和 (3)，运用 matlab 软件（代码见附件），采用 AR 自回归模型填补缺失值。AR 自回归模型是经济预测过程中一种广泛使用的方法，它既考虑了经济现象在时间上的依存性，又考虑了随机波动的干扰性，对经济运行的短期趋势的预测准确率较高。对于情况 (2)，将计算一次预测值，对于情况 (3)，将计算两次预测值。

AR 预测模型的基本步骤：

Step1: 对原始时间序列  $a_t(t=1,2,\dots,n)$  进行平稳性 daniel 检验，取显著水平为  $\alpha=0.05$ ；

Step2: 若序列为非平稳序列，对序列  $a_t(t=1,2,\dots,n)$  做一阶差分运算  $b_t = a_{t+1} - a_t$  得到序列  $b_t(t=1,2,\dots,n)$ ，即为平稳序列；

Step3: 对序列  $b_t$  建立如下的自回归预测模型，采用最小二乘法计算模型的参数；

$$y_t = c_1 y_{t-1} + c_2 y_{t-2} + \varepsilon_t (c_1, c_2 \text{ 为待定参数, } \varepsilon_t \text{ 为随机扰动项})$$

为了验证该预测模型的正确度，我们将原始预测值和 AR 模型预测值进行了对比，结果两者的相对误差在合理范围之内。

表 5.1: 已知 2018 年数据预测值和 AR 模型的预测值之间比较及相对误差

指标编号	2018 年预测值	AR 模型预测值	相对误差%	指标编号	2018 年预测值	AR 模型预测值	相对误差%
I1	6.70	6.95	3.68	I11	101.53*	101.53	0.00
I2	1.43	1.88	31.38	I12	3.61	4.01	11.21
I3	6.75	6.45	-4.49	I13	0.41*	0.41	0.00
I4	4.00	6.63	65.75	I14	99.65	103.07	3.43
I5	124.13	128.95	3.88	I15	16.16	14.93	-7.60
I6	13.00	10.64	-18.17	I16	1653117*	1653117	0.00
I7	9.00	8.05	-10.51	I17	4.75	4.35	-8.42
I8	6.50	5.90	-9.30	I18	1.50	1.50	0.00
I9	5.89	7.34	24.59	I19	2.61	2.71	3.72
I10	4.42	5.82	31.72	I20	4.72	4.64	-1.77

注：\*表示原始预测值缺失

### 5.1.2 数据一致化处

对于宏观经济指标，大部分数据为同比数据，即上一年对应时间该指标的相对改变量。对于其他不是同比数据的指标值通过以下公式化为同比数据：

$$x'_t = \frac{x_t - x_{t-1}}{x_{t-1}} \times 100\%$$

对存贷款量做相同的处理，得到存贷款量的增量同比数据：

$$y'_t = \frac{y_t - y_{t-1}}{y_{t-1}} \times 100\%$$

由于存贷款量只有三年的三个数据，所以将对应年份的四个季度的宏观经济指标求平均值作为本年度的宏观经济指标值。

## 5.2 基于主成份回归分析的存贷款预测模型

为了预测 2018 年的存贷款增量，我们先根据附件的数据建立存贷款量和宏观经济指标之间的关系，再代入 2018 年的宏观经济指标得到 2018 年存贷款增量的预测值。

我们对附件所给的数据进行分析，得到已知数据为各省 2016 年和 2017 年存贷款增量和对应年份的宏观经济指标，考虑到宏观经济指标数量较多，且数据存在多重共线性，直接进行最小二乘估计结果不稳定，故采用主成分分析法进行降维处理，用主成分代替指标，以达到简化指标和优化模型的目的。

主成份分析的原理如下：

设  $X_1, X_2, \dots, X_p$  表示以  $x_1, x_2, \dots, x_p$  为样本观测值的随机变量，如果有  $c_1, c_2, \dots, c_p$ ，使得：

$$Var(c_1X_1 + c_2X_2 + \dots + c_pX_p)$$

的值取到最大，其中  $c_1^2 + c_2^2 + \dots + c_p^2 = 1$ ，由于方差反映了数据的变异程度，上式的最优解就是一个主成份方向，主成份可以有多个，几何上两个主成份之间的方向正交。

设  $Z_i$  表示第  $i$  个主成分， $i = 1, 2, \dots, p$ ，则  $p$  个变量的主成份表示为：

$$\begin{cases} Z_1 = c_{11}X_1 + c_{12}X_2 + \dots + c_{1p}X_p \\ Z_2 = c_{21}X_1 + c_{22}X_2 + \dots + c_{2p}X_p \\ \vdots \\ Z_p = c_{p1}X_1 + c_{p2}X_2 + \dots + c_{pp}X_p \end{cases}$$

本题中，我们先建立预处理后的宏观经济指标观测矩阵：

$$A = \begin{bmatrix} a_{2015,1} & a_{2015,2} & \dots & a_{2015,14} \\ a_{2016,1} & a_{2016,2} & \dots & a_{2016,14} \\ a_{2017,1} & a_{2017,2} & \dots & a_{2017,14} \end{bmatrix}$$

为了消除变量的量纲效应，使每个变量具有同等的表现力，需对指标值  $a_{ij}$  进行标准化处理：

$$\tilde{a}_{ij} = \frac{a_{ij} - \bar{a}_j}{s_j}, i = 1, 2; j = 1, 2, \dots, 20$$

其中  $\tilde{a}_{ij}$  表示标准化后的数据， $\bar{a}_j$  和  $s_j$  分别表示第  $j$  个指标的样本数据均值和标准差。

接下来计算经济指标的相关系数矩阵  $R = \tilde{A}^T \tilde{A} / (n-1)$ ，并计算矩阵的特征值和特征向量，得到特征值的累计贡献率。

保留累计贡献率为 85% 以上的特征值和对应的特征向量。

$$\eta_j^T = [m_{j1}, m_{j2}, \dots, m_{j,14}]$$

则主成分可表示为

$$z_j = m_{j1} \tilde{x}_1 + m_{j2} \tilde{x}_2 + \dots + m_{j,14} \tilde{x}_{14}$$

作主成分做最小二乘回归分析，得到回归方程

$$\hat{y} = n_1 z_1 + n_2 z_2 + \dots + n_r z_r$$

化成标准化变量的回归方程为

$$\hat{y} = k_1 \tilde{x}_1 + k_2 \tilde{x}_2 + \dots + k_{14} \tilde{x}_{14}$$

恢复到原始自变量，即得到主成分回归方程：

$$\hat{y} = c_0 + c_1 x_1 + c_2 x_2 + \dots + c_{14} x_{14}$$

最后，代入 2018 年的宏观经济指标数据，得到预测 2018 年的存贷款增量模型：

$$\begin{cases} \hat{y}_{2018}^{(c)} = c_0^{(c)} + c_1^{(c)} x_1 + c_2^{(c)} x_2 + \dots + c_{20}^{(c)} x_{20} \\ \hat{y}_{2018}^{(d)} = c_0^{(d)} + c_1^{(d)} x_1 + c_2^{(d)} x_2 + \dots + c_{20}^{(d)} x_{20} \end{cases}$$

其中为  $c_i^{(c) \text{ or } (d)}$  为主成份回归求得的自变量系数。

### 5.3 模型的求解

以北京市存款增量为例，详细表述求解过程：

对宏观经济指标矩阵和北京市存款增量历史数据进行一致化和标准化的处理，得到标准化的数据矩阵：

$$x^{(c)} = \begin{bmatrix} 6.735 & 2.008 & \dots & 5.391 & -1.695 \\ 6.875 & 1.554 & \dots & 4.725 & 10.093 \end{bmatrix}_{2 \times 20}$$

$$y^{(c)} = [11.27, 18.65]^T$$

对矩阵进行主成份分析，求的主成份的累计贡献率为在第一个主成份以及达到 1，所以只保留第一个主成份：

$$\begin{aligned} z_1 = & -0.234 \tilde{x}_1 + 0.234 \tilde{x}_2 - 0.234 \tilde{x}_3 - 0.234 \tilde{x}_4 \\ & + 0.234 \tilde{x}_5 + 0.234 \tilde{x}_6 + 0.234 \tilde{x}_7 + 0.234 \tilde{x}_8 - 0.234 \tilde{x}_9 \\ & + 0.234 \tilde{x}_{10} + 0.234 \tilde{x}_{11} - 0.234 \tilde{x}_{12} + 0.234 \tilde{x}_{13} \\ & + 0.234 \tilde{x}_{14} - 0.234 \tilde{x}_{15} + 0.234 \tilde{x}_{16} + 0.234 \tilde{x}_{19} - 0.234 \tilde{x}_{20} \end{aligned}$$

用主成份做回归分析，得到回归方程：

$$\hat{y}^{(c)} = 0.2357z_1$$

转化成标准化变量回归方程：

$$\begin{aligned} \hat{y}^{(c)} = & -0.056 \tilde{x}_1 + 0.056 \tilde{x}_2 - 0.056 \tilde{x}_3 - 0.056 \tilde{x}_4 \\ & + 0.056 \tilde{x}_5 + 0.056 \tilde{x}_6 + 0.056 \tilde{x}_7 + 0.056 \tilde{x}_8 - 0.056 \tilde{x}_9 \\ & + 0.056 \tilde{x}_{10} + 0.056 \tilde{x}_{11} - 0.056 \tilde{x}_{12} + 0.056 \tilde{x}_{13} \\ & + 0.056 \tilde{x}_{14} - 0.056 \tilde{x}_{15} + 0.056 \tilde{x}_{16} + 0.056 \tilde{x}_{19} - 0.056 \tilde{x}_{20} \end{aligned}$$

恢复到原始的宏观经济指标为自变量，得到主成份回归方程：

$$\begin{aligned} \hat{y}^{(c)} = & 10.454 - 0.337 x_1 + 0.111 x_2 - 0.084 x_3 - 0.006 x_4 \\ & + 0.008 x_5 + 0.007 x_6 + 0.020 x_7 + 0.194 x_8 - 0.053 x_9 \\ & + 0.014 x_{10} + 0.027 x_{11} - 0.023 x_{12} + 0.007 x_{13} \\ & + 0.077 x_{14} - 0.003 x_{15} + 0.108 x_{16} + 0.076 x_{19} - 0.004 x_{20} \end{aligned}$$

代入预处理后的 2018 年的宏观经济指标值，得到：

$$y_{2018}^{(c)} = 6.845$$

2018 年预测存款增量和存款总额为：

$$\begin{aligned} \Delta_{2018}^{(c)} &= y_{2018}^{(c)} \times m_{2017}^{(c)} = 353(\text{亿元}) \\ m_{2018}^{(c)} &= (1 + y_{2018}^{(c)}) \times m_{2017}^{(c)} = 5503(\text{亿元}) \end{aligned}$$

因此，对 30 个省份的存款和贷款增量的预测结果进行了求解，结果见表 5.2：

表 5.2 各省份存款贷款预测结果

省份	存款增量	贷款增量	省份	存款增量	贷款增量
北京	353	99	陕西	81	70
江苏	361	324	天津	42	96
广东	337	377	广西	52	135
浙江	279	317	吉林	42	18
上海	229	152	云南	42	56
山东	146	105	福建	51	96
河南	134	76	江西	42	60
湖北	143	278	黑龙江	51	36
安徽	93	32	新疆	32	73
四川	83	64	贵州	21	56
辽宁	74	3	甘肃	21	23
河北	91	53	海南	30	175
重庆	82	31	内蒙古	20	12
湖南	82	141	宁夏	10	20
山西	81	51	青海	10	10
			合计	3116	3040

## 5.4 模型的结果的经一步解释

这里先引入一个概念：

$$\text{存贷比} = \frac{\text{银行贷款总额}}{\text{存款总款}} \times 100\%$$

由文献[2]，得知，各个省份的存贷比跟地区之间的差异密切相关，所以我们对各个省份的 2016-2018 年的商业银行的增量存贷比进行了统计分析，根据前两年的存贷比数据的均值将 30 个省份化为两类：高存贷比（ $\geq 75\%$ ）、低存贷比（ $< 75\%$ ）。高存贷比的省份包括：{海南、福建、内蒙古、宁夏、湖南、江西、云南、贵州、湖北、广东、浙江、新疆、山西、青海、四川、安徽、江苏、辽宁、河南、甘肃、广西}；低存贷比的省份为：{吉林、山东、陕西、重庆、上海、河北、北京、天津、黑龙江}

接下来我们分别对高存贷比的省份和低存贷比的省份再根据三年的增量存贷比变化情况划分为以下四类：

{上升趋势、下降趋势、平稳趋势、波动趋势}

其结果如下（其余图见附录 6）：

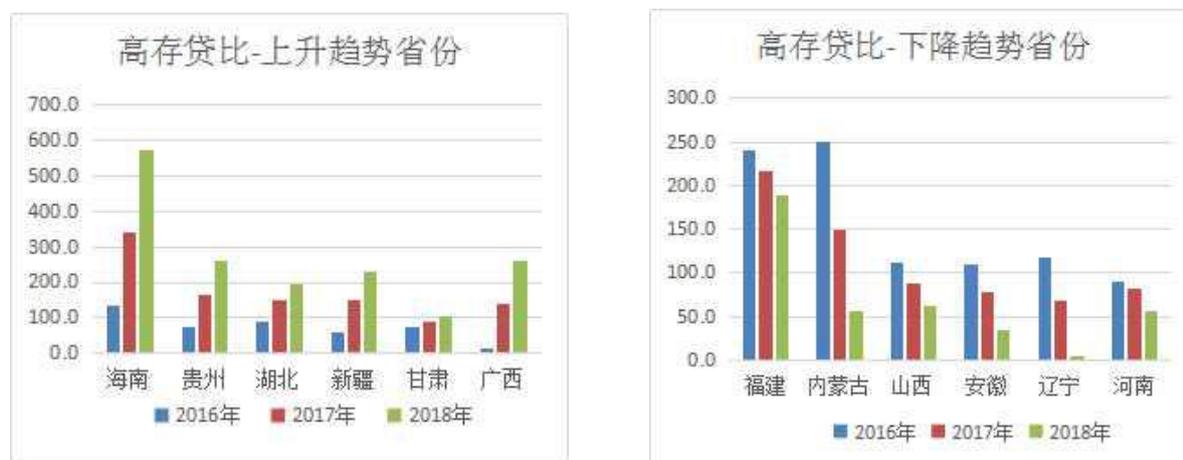


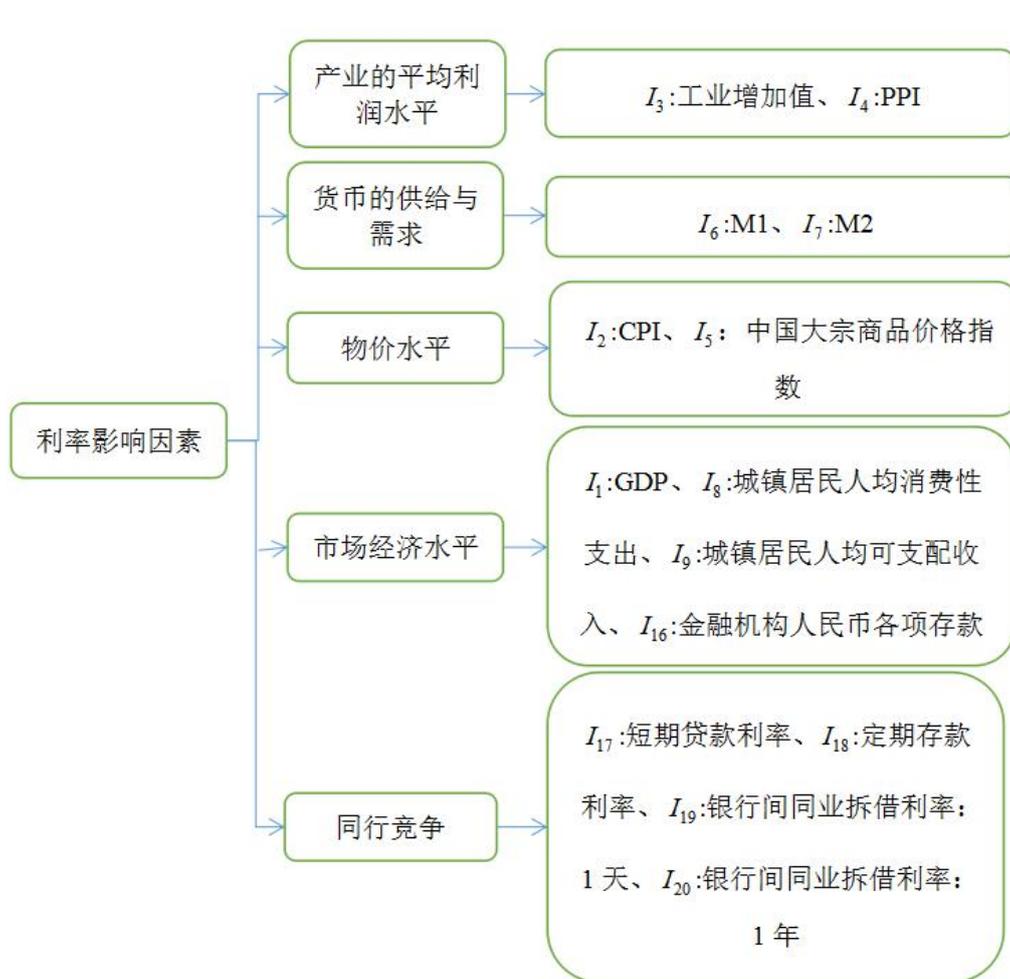
图 5.1 存贷比趋势图

## 六、问题二的模型的建立和求解

### 6.1 2018 年利率的预测

#### 6.1.1 指标的筛选

通过上网查询相关信息以及论文，我们依据附件 1 中的宏观经济指标对存贷利率的影响程度，对指标进行了遴选和剔除。根据相关文献资料所述，利率受利润率水平、资金供求状况、物价变动幅度、国际经济环境以及政策性因素的影响。基于此，我们保留了以下宏观经济指标。



### 6.1.2 基于线性回归的最小二乘估计方法的利率预测

附件只给出了 2016 年和 2017 年的各省存贷款利率，为了得到 2018 年的存贷款收益情况，必须先要预测得到 2108 年各省的存贷款利率。

将 14 个宏观经济指标建立一个新的指标向量  $\mathbf{X} = x_{ij}, i = 1, 2, 3; j = 1, 2, \dots, 14$ ，取对应年份的贷款利率或者存款利率为因变量向量  $Y = y_{ik}, i = 1, 2, 3; k = 1, 2$ （其中  $k$  表示存款利率或者贷款利率）。

这是一个多元线性回归问题，其基本模型为：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \dots + \beta_{14} x_{14} + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases}$$

其中， $\beta_0, \beta_1, \dots, \beta_{14}$  和  $\sigma^2$  都为未知的参数，前者称为回归系数。

参数  $\beta_0, \beta_1, \dots, \beta_{14}$  用最小二乘法估计，即选取适当的估计值  $\hat{\beta}_j$ ，使得误差的平方和

$$Q = \sum_{i=1}^n \varepsilon_i^2$$

最小，为此求解方程：

$$\frac{\partial Q}{\partial \beta_j} = 0, j = 0, 1, 2, \dots, 14$$

其解为:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

将其解  $\hat{\beta}$  为带回原模型, 就得到因变量  $y$  的估计值:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_{14} x_{14};$$

通过此回归方程可以估计因变量

### 6.1.3 预测利率的求解

以北京 2018 年存款利率的预测为例, 由于每年数据是按季度划分, 故使用四个季度的均值代替年度指标数据, 则北京各年存款利率与 14 个指标之间的关系数据共 3 组, 见表 6.1。

表 6.1 14 个指标数据表

指标	2015	2016	2017	指标	2015	2016	2017
$x_1$	6.925	6.725	6.875	$x_9$	5.6050	5.7175	6.6736
$x_2$	1.438	2.0081	1.5541	$x_{10}$	4.9125	4.3500	4.3500
$x_3$	6.092	6.039	6.6383	$x_{11}$	1.9375	1.5000	1.5000
$x_4$	-5.217	-1.308	6.4917	$x_{12}$	2.0175	2.0761	2.6250
$x_5$	98.99	102.51	125.79	$x_{13}$	3.8990	3.1027	4.2908
$x_6$	8.667	22.725	15.325	$x_{14}$	1315379	1466165	1604231
$x_7$	12.317	12.041	9.533	$y_1$	3.06	2.79	2.85
$x_8$	5.605	5.717	6.673				

运用公式  $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$  得到的原始自变量的回归方程各项系数见表 6.2。

表 6.2 回归方程各项系数表

$i$	0	1	2	...	12	13	14
$\hat{\beta}_i$	0.9175	0.2289	-0.0775	...	-0.0117	0.0253	0.0000

最终计算出北京 2018 年存款利率预测值

$$y_1 = 2.93$$

同理可得到北京 2018 年贷款利率及各省份存贷款利率见表 6.3

表 6.3 2018 年各省预测存、贷款利率表

省份	存款利率 (%)	贷款利率 (%)	省份	存款利率 (%)	贷款利率 (%)
北京	2.93	4.86	陕西	2.06	5.36
江苏	2.01	5.26	天津	2.66	4.75
广东	2.44	5.12	广西	2.11	5.23
浙江	2.57	5.11	吉林	2.20	5.56
上海	2.30	4.92	云南	2.01	5.25
山东	2.53	4.89	福建	2.74	5.04
河南	2.22	5.45	江西	2.32	5.51
湖北	2.17	5.58	黑龙江	2.40	4.86
安徽	2.28	5.28	新疆	1.44	5.26
四川	2.25	5.39	贵州	1.88	5.91
辽宁	2.51	5.13	甘肃	2.08	5.58
河北	2.41	5.28	海南	2.22	5.59
重庆	2.03	5.43	内蒙古	2.02	5.22
湖南	1.88	5.76	宁夏	2.22	5.66
山西	2.13	5.33	青海	2.06	5.30

## 6.2 贷款额分配的多目标优化模型

基于第一问的分析与预测结果，得到了 2018 年存贷款的预测值。存款的去向包含两个部分：法定准备金部分、库存现金部分，由于不用预留备付资金，故可将库存现金全部用于贷款投放，由于不同分行存贷款利率不同，目标求得存贷款利息净收入的最大值，同时要考虑贷款总额度约束条件和平衡地区差异等情况，故可采用目标规划法建立多目标优化模型求解该问题。

### 6.2.1 确定问题目标

对于贷款投放规模的问题，首先要考虑以下目标：

- (1) 各个分行贷款投放量总额度不能超过 2018 年预测的库存现金总额；
- (2) 全行增量存贷款净收入尽可能最大；
- (3) 为平衡各地区发展，加强民生建设，各分行贷款投放量尽可能在 2018 年各城市贷款预估值上下波动，且波动幅度尽可能小；

### 6.2.2 处理目标与约束

用偏差变量来表示实际值和目标值之间的差异，设：

$$\begin{cases} d^+ : \text{正偏差变量} \\ d^- : \text{负偏差变量} \end{cases}$$

其中： $d^+, d^- \geq 0$  且  $d^+ \cdot d^- = 0$

- (1) 库存现金总量约束

$$\min \{d_0^+\}$$

$$s.t. \begin{cases} \sum_{i=1}^{30} m_i^{(d)} + d_0^- - d_0^+ = \hat{M}^{(c)}(1-p) \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ m_i^{(d)}, \hat{m}_i^{(c)} \geq 0; i=1,2,\dots,30 \end{cases}$$

(2) 全行增量存贷款净收入最大约束

$$\min\{-M\}$$

$$s.t. \begin{cases} M = \sum_{i=1}^{30} m_i^{(d)} \hat{\alpha}_i^{(d)} - \sum_{i=1}^{30} \hat{m}_i^{(c)} \hat{\alpha}_i^{(c)} + \hat{M}^{(c)} p\alpha \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ \hat{m}_i^{(c)or(d)}, m_i^{(d)} \geq 0; i=1,2,\dots,30 \end{cases}$$

(3) 各城市贷款预计值约束

$$\min\{\sum_{i=1}^{30} d_i^+ + \sum_{i=1}^{30} d_i^-\}$$

$$s.t. \begin{cases} m_i^{(d)} - d_i^+ + d_i^- = \hat{m}_i^{(d)} \\ d_i^+, d_i^- \geq 0; i=1,2,\dots,30 \end{cases}$$

### 6.2.3 确定目标的优先级与权系数

对于贷款投放规模的问题各个目标的等级和优先权系数如下：

$P_1$  : 各个分行贷款投放量总额度不能超过 2018 年库存现金总额

$P_2$   $\begin{cases} w_1 : \text{全行} \blacklozenge \text{量存贷行增量存尽可能最大} \\ w_2 : \text{为平衡各地区发展，加强民生建设，各分行贷款投放量尽可能} \\ \text{在2018年各城市贷款预估值上下波动，且波动幅度尽可能小} \end{cases}$

取权系数向量为： $\mathbf{w} = [6,4]$

综上所述，建立如下所示的求解贷款规模分配方案的多目标规划模型：

$$\min = P_1 d_0^+ + P_2 \times [w_1 \times (-M) + w_2 \times (\sum_{i=1}^{30} d_i^+ + \sum_{i=1}^{30} d_i^-)]$$

$$\begin{cases}
 \sum_{i=1}^{30} m_i^{(d)} + d_0^- - d_0^+ = \hat{M}^{(c)}(1-p) \\
 M = \sum_{i=1}^{30} m_i^{(d)} \hat{\alpha}_i^{(d)} - \sum_{i=1}^{30} \hat{m}_i^{(c)} \hat{\alpha}_i^{(c)} + \hat{M}^{(c)} p \alpha \\
 s.t. \begin{cases} m_i^{(d)} - d_i^+ + d_i^- = \hat{m}_i^{(d)}; i = 0, 1, \dots, 30 \\
 \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\
 \hat{m}_i^{(c) \text{ or } (d)}, m_i^{(d) \text{ or } (c)}, d_i^+, d_i^- \geq 0 \text{ 且 } d_i^+ \times d_i^- \geq 0
 \end{cases}
 \end{cases}$$

### 6.2.4 模型求解

问题二中存在多个约束，其中刚性约束以及柔性约束均含有，故对于不同的约束条件，其优先级以及权系数存在差异，各约束条件的优先级会影响到最终目标的最优解。该算法设计分为以下三个步骤：

Step1: 输入  $P_1, P_2$  的权系数；

Step2: 通过误差分析，调整  $P_1, P_2$  的权系数值；

Step3: 将调整后的值输入程序，重复运行，获得最优解。

求解结果见表 6.4

表 6.4 问题二的贷款分配方案表

省份	2018 年贷款预测增量	贷款分配额度 (亿元)	相对差值 %	省份	2018 年贷款预测增量	贷款分配额度 (亿元)	相对差值 %
北京	99	78.92	-20.28	陕西	70	73.74	5.34
江苏	324	280.04	-13.57	天津	96	76.62	-20.19
广东	377	301.76	-19.96	广西	135	108.01	-19.99
浙江	317	253.26	-20.11	吉林	18	19.38	7.67
上海	152	121.6	-20.00	云南	56	44.8	-20.00
山东	105	84.29	-19.72	福建	96	77.13	-19.66
河南	76	79.9	5.13	江西	60	63.36	5.60
湖北	278	291.9	5.00	黑龙江	36	28.42	-21.06
安徽	32	33.47	4.59	新疆	73	58.4	-20.00
四川	64	67.2	5.00	贵州	56	58.8	5.00
辽宁	3	2.4	-20.00	甘肃	23	23.82	3.57
河北	53	55.72	5.13	海南	175	183.75	5.00
重庆	31	32.94	6.26	内蒙古	12	9.23	-23.08
湖南	141	147.84	4.85	宁夏	20	21.12	5.60
山西	51	53.91	5.71	青海	10	10.66	6.60
				合计	3040	2742.4	-9.79

### 6.3 模型的检验和解释

对于求解结果，我们将其与预测结果进行比较，发现除个别省市以外，偏差值都不超过 20%。对于像北京这样偏差较大的省份，通过数据校核发现是由于该省利率较低引起的。对于整体模型，贷款规模分配方案与预测数据符合程度高，对于个别数据，我们可通过改进优化模型，提高数据的准确度。

## 七、问题三的模型的建立和求解

### 7.1 债券分配模型

对于债券利率这类受多种因素影响的数值，我们通过聚类分析的方法，得到了各分行发行的债券的利率

#### 7.1.1 数据处理

由于各省存贷款利率与存贷比有差异，故而债券利率  $\alpha^{(2)}$  无法之际统一使用 5.1%，需要根据实际情况适当调整债券利率  $\alpha^{(2)}$ 。经分析，存贷比越高，可用于购买债券的流动资金就越少，债券的吸引力也就越差。采用基于存贷比的聚类分析将各省债券利率分成三个部分，存贷比偏高的  $\alpha^{(2)}$  调整为 5.0%，偏低的则调为 5.2%，其余仍为 5.1%。

聚类分析结果见附表，各省调整后的利率表见表 7.1

表 7.1 调整后各省债券利率表

城市	债券利率%	城市	债券利率%
北京	4.9	陕西	4.9
江苏	5.0	天津	4.9
广东	5.0	广西	5.0
浙江	5.0	吉林	4.9
上海	4.9	云南	5.0
山东	4.9	福建	5.0
河南	5.0	江西	5.0
湖北	5.0	黑龙江	4.9
安徽	4.9	新疆	5.0
四川	4.9	贵州	5.0
辽宁	5.0	甘肃	5.0
河北	4.9	海南	5.0
重庆	4.9	内蒙古	5.0
湖南	5.0	宁夏	5.0
山西	5.0	青海	4.9

#### 7.1.2 确定问题目标

由于公民对银行债券接受度普遍较大，故假设债券能全部发行完。

(1) 各个分行发行债券总和为 500 亿元。

(2) 为平衡各地区发展，缩小地域差异，各分行债券发行量在发放区间内。

### 7.1.3 处理目标与约束

(1) 各个分行发行债券总和为 500 亿元。

$$\begin{cases} \sum_{i=1}^{30} m_i^{(z)} = 500 \\ m_i^{(z)} \geq 0; i = 1, 2, \dots, 30 \end{cases}$$

(2) 为平衡各地区发展，缩小地域差异，各分行债券发行量在发放区间内。

$$\begin{cases} 0.1\hat{m}_i^{(c)} \leq m_i^{(z)} \leq 0.2\hat{m}_i^{(c)} \\ m_i^{(z)} \geq 0; i = 1, 2, \dots, 30 \end{cases}$$

综上所述，建立如下线性规划模型  $\min = \sum_{i=1}^{30} \alpha_i^{(z)} m_i^{(z)}$

$$\begin{cases} \sum_{i=1}^{30} m_i^{(z)} = 500 \\ 0.1m_i^{(c)} \leq m_i^{(z)} \leq 0.2m_i^{(c)} \\ m_i^{(z)} \geq 0; i = 1, 2, \dots, 30 \end{cases}$$

通过 matlab 程序算法求解结果见表 7.2

表 7.2 各省债券发行量表

城市	债券发行量 (亿元)	城市	债券发行量 (亿元)
北京	16.96	陕西	16.96
江苏	16.46	天津	16.96
广东	16.46	广西	16.46
浙江	16.46	吉林	16.96
上海	16.96	云南	16.46
山东	16.96	福建	16.46
河南	16.46	江西	16.46
湖北	16.46	黑龙江	16.96
安徽	16.96	新疆	16.46
四川	16.96	贵州	16.46
辽宁	16.46	甘肃	16.46
河北	16.96	海南	16.46
重庆	16.96	内蒙古	16.46
湖南	16.46	宁夏	16.46
山西	16.46	青海	16.96

## 7.2 贷款分配模型建立

银行通过发行债券，筹得了五百亿的流动资金，为提高银行收益，可将该筹得资金用于放贷或长期存入中央银行。目标函数是使得银行增量存贷款利息取得最大值，故可

用线性规划来解决该问题，建立相关约束。

### 7.2.1 确定问题目标

对于贷款规模投放的问题，首先要考虑以下目标：

- (1) 各个分行贷款投放量总额度不能超过 2018 年预测的库存现金与债券筹资总额；
- (2) 全行增量存贷款净收入尽可能最大；
- (3) 为平衡各地区发展，加强民生建设，各分行贷款投放量尽可能在 2018 年各城市贷款预估值上下波动，且波动幅度尽可能小；

### 7.2.2 处理目标与约束

- (1) 库存现金与债券筹资总额约束

$$\begin{cases} \min \{d_0^+\} \\ \sum_{i=1}^{30} m_i^{(d)} + d_0^- - d_0^+ = \hat{M}^{(c)}(1-p) + \hat{M}^{(c)} \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ m_i^{(d)}, \hat{m}_i^{(c)} \geq 0; i = 1, 2, \dots, 30 \end{cases}$$

- (2) 全行增量存贷款净收入最大约束

$$\begin{cases} \min \{-M\} \\ M = \sum_{i=1}^{30} m_i^{(d)} \hat{\alpha}_i^{(d)} - \sum_{i=1}^{30} \hat{m}_i^{(c)} \hat{\alpha}_i^{(c)} + \hat{M}^{(c)} p\alpha - \sum_{i=1}^{30} m_i^{(z)} \alpha_i^{(z)} \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ \hat{m}_i^{(c)}, m_i^{(d)}, m_i^{(z)} \geq 0; i = 1, 2, \dots, 30 \end{cases}$$

- (3) 各城市贷款预计值约束

$$\begin{cases} \min \left\{ \sum_{i=1}^{30} d_i^+ + \sum_{i=1}^{30} d_i^- \right\} \\ m_i^{(d)} - d_i^+ + d_i^- = \hat{m}_i^{(d)}; i = 1, 2, \dots, 30 \\ d_i^+, d_i^- \geq 0 \end{cases}$$

- (4) 确定目标的优先级与权系数

对于贷款投放规模的问题各个目标的等级和优先权系数如下：

$P_1$ : 各个分行贷款投放量总额度不能超过 2018 年库存现金与债券筹资总额

$P_2$   $\begin{cases} w_1: \text{全行增量存贷款净收入尽可能最大} \\ w_2: \text{为平衡各地区发展, 加强民生建设, 各分行贷款投放量尽可能} \\ \text{在2018年各城市贷款预估值上下波动, 且波动幅度尽可能小} \end{cases}$

权系数向量为:  $\mathbf{w} = [6, 4]$

综上所述，建立如下线性规划模型：

$$\begin{cases} \min = P_1 d_0^+ + P_2 \times [w_1 \times (-M) + w_2 \times (\sum_{i=1}^{30} d_i^+ + \sum_{i=1}^{30} d_i^-)] \\ \sum_{i=1}^{30} m_i^{(d)} + d_0^- - d_0^+ = \hat{M}^{(c)}(1-p) + \hat{M}^{(c)} \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ M = \sum_{i=1}^{30} m_i^{(d)} \hat{\alpha}_i^{(d)} - \sum_{i=1}^{30} \hat{m}_i^{(c)} \hat{\alpha}_i^{(c)} + \hat{M}^{(c)} p \alpha - \sum_{i=1}^{30} m_i^{(z)} \alpha_i^{(z)} \\ m_i^{(d)} - d_i^+ + d_i^- = \hat{m}_i^{(d)}; i = 0, 1, \dots, 30 \\ \hat{m}_i^{(c)}, m_i^{(d) \text{ or } (c) \text{ or } (z)}, d_i^+, d_i^- \geq 0 \text{ 且 } d_i^+ \times d_i^- \geq 0 \end{cases}$$

### 7.3 模型求解

(一) 问题三沿用了问题二的线性规划模型，在添加了相关约束之后，寻求该条件下最优解，也即最佳贷款分配方案。

(二) 算法设计

算法设计也采用问题二的 lingo 程序，故不多赘述

(三) 求解结果见表 7.3

表 7.3 各省贷款分配表

省份	贷款分配额度 (亿元)	省份	贷款分配额度 (亿元)
北京	102.71	陕西	73.20
江苏	331.28	天津	93.35
广东	371.80	广西	140.21
浙江	315.24	吉林	17.87
上海	153.33	云南	55.35
山东	104.50	福建	94.07
河南	72.53	江西	57.99
湖北	268.10	黑龙江	37.09
安徽	30.50	新疆	74.50
四川	62.77	贵州	54.34
辽宁	3.09	甘肃	21.73
河北	58.99	海南	167.74
重庆	36.53	内蒙古	11.88
湖南	147.17	宁夏	19.21
山西	53.85	青海	9.97

### 7.4 模型的检验和解释

该模型在第二问的基础上进一步优化，增加了新的约束条件，从程序运行结果来看，数据分布较为合理，在预测值上下波动不大，符合程度良好，置信度高。

## 八、问题四的模型的建立和求解

### 7.1 数据的分析和处理

对每日存取款金额，由于只需要求解备付水平，而备付金额满足

$$m^{(b)} = m^{(q)} - 0.88 \times m^{(c)}$$

故只需要分析存取款金额的差值  $\Delta$  即可，其中

$$\Delta = -m^{(b)} = 0.88 \times m^{(c)} - m^{(q)}$$

通过绘制各省份存取款金额  $\Delta$  的直方图，发现差值都类似于正态分布，在此基础上，

计算了偏度系数和峰度系数，并对数据进行了显著性检，从结果分析， $\Delta$  并不符合正态分布，需要对其进行修正变换才能服从正态分布。经分析，采用 Box-Cox 变换对数据进行处理，Box-Cox 变换可以在满足线性、独立性、方差齐性以及正态性的同时，又不丢失信息，其变换公式为

$$\Delta^{(\lambda)} = \begin{cases} \frac{\Delta^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log(\Delta), \lambda = 0 \end{cases}$$

其中  $\lambda$  为待定变换参数，可由极大似然法确定。在进行变换后由逆变换公式

$$\Delta = \begin{cases} (1 + \lambda \Delta^{(\lambda)})^{\frac{1}{\lambda}}, \lambda \neq 0 \\ \exp(\Delta^{(\lambda)}), \lambda = 0 \end{cases}$$

可以得到对应的返回值。由于  $\Delta$  中含有负值，所以在实际处理过程中先对其进行预处理，使其均为正数

$$\Delta' = \Delta + c$$

其中  $c$  为常数，可由  $\Delta$  中的最小值所确定。

### 7.2 备付资金模型的建立

本问题要求根据附件 4 中商业银行 A 各分行每日存取款交易数据，建立备付金分配模型并设计算法，求出 2018 年各分行日常经营所需最低备付金额。

根据备付金额与存取款金额关系可知，当备付金额  $m^{(b)}$  不小于  $\Delta$  时，即可满足备付需求，为了在不引起客户不满的基础上降低备付水平，提高盈利，需要对 2018 年备付金额进行模拟预测，在置信水平为 99% 的条件下得到最低的备付金额。

各分行一年的总备付金额可根据每日实际备付金额需求来确定，表示为

$$m^{(b)} = -\sum_{i=1}^{365} \Delta_i$$

由于  $\Delta_i$  服从正态分布，则  $m^{(b)}$  也服从正态分布，故可以对备付金额进行假设检验，得到置信区间的下限即为最低备付金额

$$m_{\min}^{(b)} = \bar{m}^{(b)} - t_{\alpha/2} \frac{\sigma_{m^{(b)}}}{n}$$

其中  $\bar{m}^{(b)}$  为备付资金的样本均值， $\sigma_{m^{(b)}}$  为备付资金样本的标准差， $n$  为样本容量， $t_{\alpha/2}$  为标准正态分布的  $\alpha/2$  分位点。

### 7.3 备付资金模型的求解

本问题需要预测 2018 年的备付金额，在已经得到  $\Delta$  的分布情况下，可以进行模拟求解，以下为算法步骤：

Step 1: 根据分布生成各分行每日备付金额  $\Delta_i$  的随机数

Step 2: 计算各分行一年的备付金额总量  $m^{(b)}$

Step 3: 重复 Step 1 和 Step 2，得到足够大的备付资金样本

Step 4: 根据样本求得置信水平为 99% 的条件下得到最低的备付金额  
通过算法得到的结果见表 7.1

表 7.1 各省预备金额表

省份	预备金额	省份	预备金额
北京	190.489	陕西	-26.398
江苏	105.623	天津	20.334
广东	44.375	广西	28.611
浙江	39.453	吉林	20.219
上海	11.393	云南	76.851
山东	34.659	福建	33.228
河南	149.179	江西	120.546
湖北	97.274	黑龙江	17.461
安徽	0.605	新疆	118.459
四川	-7.739	贵州	22.555
辽宁	67.396	甘肃	8.333
河北	31.853	海南	3.869
重庆	42.684	内蒙古	5.468
湖南	80.024	宁夏	14.614

其中，预备资金负值表示本省存款金额足以保证客户顺利取款，且有富余。

## 九、问题五的模型的建立和求解

### 9.1 模型的建立

综合考虑三个方面的因素：

- (1) 收益水平;
- (2) 流动性需求和风险水平;
- (3) 地区之间的协调发展;

关于第一个因素已在前面进行了讨论,下面分别对(2)、(3)因素进行分析,进而改进贷款规模的分配模型。

### 9.1.1 流动性需求和风险水平

我们参考文献[《商业银行流动性风险管理指引》]对流动性风险的定义:

流动性风险指商业银行虽然有清偿能力,但无法及时获得充足资金或无法以合理成本及时获得充足资金以应对资产增长或支付到期债务的风险。

对于银行和企业来说,流动性的极度不足会导致破产,因此资金的流动性能是银行和企业经营能力的指标之一。

根据文献[1],结合附件所给的数据,从可行性的角度,选取了能够反映流动性风险的两个主要指标:

指标	计算公式
存贷比率 $r_1$	贷款余额/存款余额
备付金比例 $r_2$	(存放中央银行+库存现金)/存款余额

下面一次对每个指标进行分析:

(1) 存贷比:存贷比过高意味着银行库存现金、存款准备金不足,应付客户日常支取和结算的能力严重下降,代表商业银行资产的流动性越差,因此,流动性风险也会越大,假设流动风险性和存贷比之间成正比,即:

$$R \propto \frac{m^{(d)}}{\hat{m}^{(c)}}, (m^{(d)} \text{ 表示实际的分行贷款量, } \hat{m}^{(c)} \text{ 表示预测的存款量})$$

(2) 备付金比例:可以用来衡量商业银行应付存款支取不确定性的能力。该指标值越高,表示商业银行资产流动性越强,风险越低。假设:

$$R \propto -(r_2 - \frac{m^{(b)}}{\hat{m}^{(c)}}), (m^{(b)} \text{ 表示第四问所求得各分行最低备付金额})$$

所以,流动性风险可以表示为:

$$R = k_1 \frac{m^{(d)}}{\hat{m}^{(c)}} - k_2 (r_2 - \frac{m^{(b)}}{\hat{m}^{(c)}}), (k_1, k_2 \geq 0)$$

### 9.1.2 考虑地区差异的改进模型

存贷比的增长率,即一阶倒数的大小:  $r_1'$ , 反映了该地区贷款相比存款的增长速度,对于该地区应该在分配贷款额时施加更多的关照。将问题二的模型的多目标之一,贷款的投放量尽量靠近预测值:

$$\min \left\{ \sum_{i=1}^{30} d_i^+ + \sum_{i=1}^{30} d_i^- \right\}$$

改为:

$$\min\left\{\sum_{i=1}^{30} r'_{li} \times d_i^+ + \sum_{i=1}^{30} r'_{li} \times d_i^-\right\}$$

用  $r'_{li}$  的大小来表达各地区存贷比增长率对贷款分配的影响。

综合以上的分析，对于贷款投放规模的问题，要考虑以下目标：

- (1) 各个分行贷款投放量总额度不能超过 2018 年预测的库存现金总额；
- (2) 全行增量存贷款净收入尽可能最大；
- (3) 为平衡各地区发展，加强民生建设，各分行贷款投放量尽可能在 2018 年各城市贷款预估值上下波动，且波动幅度尽可能小；
- (4) 各个分行要保留最低的备付金比例；
- (5) 流动风险系数  $R$  要尽量小；
- (6) 模型要考虑地区之间的差异，积极扶持。

我们建立以下多目标非线性规划模型来分配贷款规模：

(其中  $m_i^{(d)}$  为各个分行实际的贷款量)

$$\min = P_1 d_0^+ + P_2 \times [w_1 \times (-M) + w_2 \times (\sum_{i=1}^{30} r'_{li} \times d_i^+ + \sum_{i=1}^{30} r'_{li} \times d_i^-) + w_3 \times R + w_4 \times d_{31}^-]$$

$$s.t. \begin{cases} \sum_{i=1}^{30} m_i^{(d)} + d_0^- - d_0^+ = \hat{M}^{(c)}(1-p) \\ M = \sum_{i=1}^{30} m_i^{(d)} \hat{\alpha}_i^{(d)} - \sum_{i=1}^{30} \hat{m}_i^{(c)} \hat{\alpha}_i^{(c)} + \hat{M}^{(c)} p \alpha \\ 0.88 \times \hat{m}_i^{(c)} - m_i^{(d)} + d_{31}^- - d_{31}^+ = m_{\min,i}^{(b)} \\ R = k_1 \frac{m^{(d)}}{\hat{m}^{(c)}} - k_2 (r_2 - \frac{m^{(b)}}{\hat{m}^{(c)}}), (k_1, k_2 \geq 0) \\ m_i^{(d)} - d_i^+ + d_i^- = \hat{m}_i^{(d)} \\ \hat{M}^{(c)} = \sum_{i=1}^{30} \hat{m}_i^{(c)} \\ \hat{m}_i^{(c)or(d)}, m_i^{(d)or(c)}, d_i^+, d_i^- \geq 0 \text{ 且 } d_i^+ \times d_i^- \geq 0; i = 0, 1, \dots, 30 \\ m_i^{(d)} < \hat{m}_i^{(d)}; m_i^{(d)} < m_i^{(d)} \end{cases}$$

## 9.2 企业经营与政策导向建议

首先一个企业的经营模式直接与企业的利润相挂钩，若组织构架、绩效考核等相对合理且对员工有激励作用，那么该企业的效益将会提高。其次，企业领导者需要有把握全局的目光，对政策有较高的敏感性，比如在哪个地区贷款政策放宽或者对特定人群的贷款方式发生了变化，都将对银行今后一段时间的贷款对象、贷款利率、贷款金额等起到一定导向作用。再次，企业管理人员需要关注贷款的行业集中度、地区集中度以及人群集中度，若集中度过高，将加大贷款金额收回的风险，以及影响贷款的发放，故而需要注意调控集中度，尽量使产业多样化、地区差异特色化、人群类别化，这样才能让贷款模式多层次且更稳定。

## 十、模型的评价和改进

问题一：针对多指标采用主成分分析法，在保证各指标贡献率的基础上实现了降维简化的目的，使得回归分析更加准确。但由于样本数较少，AR 自回归预测模型预测效果不是很好，可以采用多种预测方式相结合，通过比较检验来确定最佳预测值。

问题二：贷款分配模型考虑了实际值与预测值的偏差变量，使得优化结果更符合实际。预测存贷款利率时通过指标含义筛选指标，没有结合相关数据，可以在了解内涵的基础上再对相应指标数据进行相关性分析检验再剔除。

问题三：采用聚类分析法确定债券在各省的实际利率，考虑了地方性差异，对债券分配起到了一定约束作用。发放债券后，没有充分考虑各省贷款的需求也会发生改变。

问题四：采用 Box-Cox 对每日存取款金额分布进行修正，在满足线性、独立性、方差齐性以及正态性的同时，又不丢失信息。在模拟过程可以适当加大样本量让数据更有说服力。

问题五：建立了一个多目标非线性规划模型，考虑了多个指标对贷款分配的影响，但是仍然有一些指标难以量化，无法用于优化贷款分配模型。

## 十一、参考文献

[1] 付强, 刘星, 计方. 商业银行流动性风险评价[J]. 金融论坛, 2013, 18(04):9-16.

[2] 宋保庆. 地区存贷比差异的再认识[J]. 华北金融, 2014(12):9-11+53.

[3] 司守奎, 孙兆亮. 数学建模算法与应用. 国防工业出版社, 2017 年.

[4] 韩中庚. 数学建模方法及其应用. 解放军信息工程大学, 2009, 6.

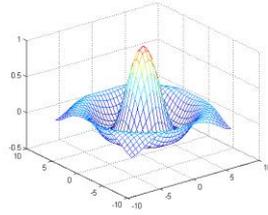
附录：

附录目录：

附录 1：

单位：亿元		问题 2	问题 3	问题 4
序号	分行	贷款规模分配方案	贷款规模分配方案	2018 年备付资金
1	北京	78.92	102.71	190.48
2	江苏	280.04	331.28	105.62
3	广东	301.76	371.80	44.37
4	浙江	253.26	315.24	39.45
5	上海	121.6	153.33	11.39
6	山东	84.29	104.50	34.65
7	河南	79.9	72.53	149.17
8	湖北	291.9	268.10	97.27
9	安徽	33.47	30.50	0.60
10	四川	67.2	62.77	-7.73

# 五一数学建模竞赛



## 题目：商业银行人民币贷款规模分配及盈利问题

**关键词：**支持向量机 分位数回归 主成分分析 层级分析法 正态分布 线性规划

**摘要：**本文利用支持向量机（svm）回归模型、主成分分析、分位数回归、层次分析法等方法，使用 python 语言对方法进行实现，建立了银行贷款额度分配模型，并引入了收益与风险、企业经营与国际政策导向，区域化差异与分行公平考核对于模型进行优化，最后针对商业银行 A 给出了合理的贷款额度分配方案。

针对问题一，将季度数据合并为年度数据，对数据特征采用主成分分析进行处理并修复缺失值，利用 svm 模型进行回归预测得到 A 行 2018 年贷款增量。

针对问题二，对利率和存款进行 svm 回归预测得到 2018 年 A 行各地预测值，然后对贷款值进行分位数回归求得贷款规模的上下限，将利率和存款作为固定参数，贷款规模分配作为自变量，贷款规模上下限和问题 1 求得贷款总额作为约束条件，以利息收入为目标，建立数学规划模型，得到使得全行增量存贷款利息净收入最大的贷款规模分配方案。

针对问题三，加入普通债所获资金 500 亿，调整贷款总额度和贷款规模上限后使用 python 语言的 scipy 包中的 linprog 线性规划模块求解最优贷款规模分配方案。

针对问题四，统计各地分行每日的存取款之差发现其规律符合正态分布，求得其均值估计和方差估计。利用 2017 年的各分行存款数额、分布均值和方差，得到 2018 年的各分行存款预测数额、分布均值和方差。利用其对各分行贷款额度分配在 99%置信区间对上下界区间进行预测，下界即每日存款小于取款的差额的最大值，可以将此值看作各分行日常经营所需最低备付金额。

针对问题五，为了实现收益与风险、企业经营与国际政策导向、区域化差异与分行公平考核的多赢。利用层次分析法赋予这三对因素权重，引入风险系数、地区国际化指数、不公平差异，对三对因素在不同贷款规模分配方案上进行量化，最终得到三对因素收益最大的贷款规模分配方案。

## 一、 问题重述

### 1.1 问题背景

商业银行贷款投放的简单模型是：从客户端吸收存款，缴存法定准备金（法定准备金率：大型金融机构 15.5%，中小金融机构 12%；法定准备金利率 1.62%），预留一定比例备付水平（商业银行承担支付结算金融中介作用，满足流动性安全），剩余资金用于贷款投放或其他资产配置。

对商业银行而言，贷款规模增长受限于其存款规模增长，只有在存款有效增长的情况下银行才有充足资金用于贷款投放。具体来说，某家商业银行在全国存款总额中所占比例相对稳定，每年可增长的存款量受限于社会资金总量的增长情况。从经济指标上来看，社会存款增长与 GDP、CPI、工业增加值等宏观经济指标密切相关（附件 1 为 2010-2017 宏观经济指标及 2018 年预测情况）。

在贷款分配模式上，商业银行过去多采用年初预分配、全年限额管理模式，即年初一次性预分配全年规模至各家分行，年度内不再调增。该模式存在较多弊端，如未考虑全年实际可支配贷款规模、环境变化、需求变化等，效率低、弹性差。目前，商业银行多采用以存定贷、存贷结合、表内外协同、资产负债动态平衡模式，既要努力获取全行最大收益，也要平衡各区域发展差异，调动各单位展业积极性，同时也要对国家重大项目、重点政策、民生工程等倾斜扶持，支持实体经济有效发展。

现有某中型商业银行 A，其各家省级分行近三年存贷款业务发展情况见附件 2。请查阅相关文献和资料，在对商业银行经营模式有所掌握，对未来经济、利率、汇率等政策及趋势有充分了解和预判的基础上回答以下问题。

### 1.2 问题概述

问题 1：假设该银行除客户存款外无其他资金来源，且暂不考虑备付水平。请根据附件 2 商业银行 A 各项存贷款历史数据及附件 1 宏观经济指标历史数据，建立数学模型，预测该银行 2018 年存、贷款增量情况。

问题 2：假设该银行除客户存款外无其他资金来源，且暂不考虑备付水平。请根据问题 1 预测的结果并结合附件 3 相关数据，建立数学模型，给出 2018 年商业

银行 A 各分行贷款规模的分配方案，使得全行增量存贷款利息净收入最大，并将该分配方案填入表 1。

问题 3：若商业银行 A 将于 2018 年 5 月 1 日发行 500 亿规模的 15 年期商业银行普通债（利率约为 5.1%），请结合该条件对问题 2 进一步优化，暂不考虑备付水平情况下，重新设计商业银行 A 各分行贷款规模的分配方案，使得全行增量存贷款利息净收入最大，并将该分配方案填入表 1。

问题 4：为保证每日交易正常进行，各家分行每日需预留一定的备付资金（备付资金不足易引起客户不满，严重的会引起社会恐慌，引发挤兑；预留资金过多，会降低银行盈利水平），以确保最低的备付水平（备付水平=备付资金÷存款余额）。假设每个客户存取款的行为是随机的，请根据附件 4 各分行 2017 年每日存取款交易数据，建立数学模型，计算在置信水平 99%的情况下，2018 年商业银行 A 各分行日常经营所需最低备付金额，并将结果填入表 1。

问题 5：在贷款规模分配问题上，为了帮助商业银行 A 处理好收益与风险、企业经营与国际政策导向，区域化差异与分行公平考核等之间的关系，以期达到双赢或多赢，请对以上模型进行改进，并给出相关建议。

## 二、 建模假设

1. 假设 2018 年经济情况平稳发展。
2. 假设城市国际化指数和城市风险系数真实有效。
3. 假设每个客户存取款的行为是随机的；
4. 假设银行普通债能够全部卖出，并在普通债到期日付清本息。
5. 假设按照某地存款量来分配贷款额度是公平的。

## 三、 符号说明

符号	说明
$w$	模型参数

$C$	惩罚项
$\xi$	松弛变量
$b$	截距项
$x$	输入特征
$y$	实际值
$y'$	预测值
<i>Desposit Prediction</i>	各地存款
<i>Upper</i>	贷款规模上限
<i>Lower</i>	贷款规模下限
<i>TotalLoan</i>	贷款总额度
<i>TotalLoanl</i>	调整后的贷款总额度
<i>Upper'</i>	调整后的贷款规模上限
<i>Difference</i>	每日存取款差额
$Lr$	贷款利率向量
$Dr$	存款利率向量
$\alpha$	贷款规模分配比例向量
$\alpha_k$	k 地贷款规模分配比例
$\mu$	存款规模比例向量
$\mu_k$	k 地存款规模比例
$A$	收益与风险因素重要性权重
$B$	企业经营与国际政策导向因素重要性权重

$C$	区域化差异与分行公平考核因素重要性权重
$\rho$	风险系数向量
$SBII$	分行国际化指数向量

## 四、 问题分析

### 4.1 问题一的分析

第一问中要求预测2018年存、贷款增量情况。筛选出附件1主要指标建立与金融机构人民币各项存款余额差值之间的数学模型，对于2018年金融机构人民币各项存款增量进行预测。题目中给出某家商业银行在全国存款总额中所占比例相对稳定，每年可增长的存款量受限于社会资金总量的增长情况，可以由此建立回归预测模型得出该银行存、贷款增量情况。

### 4.2 问题二的分析

第二问要求建立数学模型，给出2018年商业银行A使得全行增量存贷款利息净收入最大时的各分行贷款规模的分配方案。可以考虑根据附件1中2015-2018年短期贷款利率以及附件3的各地区存贷款利率水平数据，使用线性回归方法建立宏观指标与地区平均贷款利率的关系式，得到2015-2018年的综合平均贷款利率。再使用线性回归方法分别建立2015-2017年每个地区的贷款利率与其相对应的综合平均贷款利率的关系，进而得到2018年商业银行A各地区分行贷款规模分配方案。

### 4.3 问题三的分析

第三问要求在第二问的基础上追加500亿15年期利率为5.1%的商业银行普通债，重新设计使得全行增量存贷款利息净收入最大时的贷款规模分配方案。考虑到商业银行普通债在期末时才会给客户支付利息，因此只代表商业银行A贷款总规模在第二个问的基础上增加了500亿。所以将总贷款规模加500亿带入第二问的数学模型中，得到使得全行增量存贷款利息净收入最大时新的贷款规模分配方案。

### 4.4 问题四的分析

第四问要求考虑备付水平，根据各分行2017年每日存取款交易数据建立数学模型，计算在置信水平99%的情况下，2018年商业银行A各分行日常经营所需最低备付金额。在这里，备付资金主要用于客户取款，不考虑其他用途。因此我们统计各

地分行每日的存取款之差，作频率图可发现大多数具有类似正态分布的形状，考虑到用户存取款的行为存在随机性，所以假设每日存取款的额度服从正态分布。根据作图观察分析，我们发现分行存款总额和分行存取款差额的均值、分行存款总额和分行存取款差额的方差都存在相关性。因此本文假设 2018 年的各分行的存取款数额也符合正态分布，以 2017 年的数据为基础，建立回归预测模型对 2018 年各分行存取款数额分布的参数进行预测。根据预测获得的参数构建分布，最终获得置信水平 99%的情况下，2018 年商业银行 A 各分行日常经营所需最低备付金额。

#### 4.5 问题五的分析

第五问要求在贷款规模分配问题上，为了达到双赢或多赢的目标，考虑收益与风险、企业经营与国际政策导向、区域化差异与分行公平考核等之间的关系，对贷款规模分配模型进行优化，并给出相关建议。考虑到实际问题中，贷款规模分配不只是依据利率水平的高低，而是由许多内外部因素综合影响决定的。因此，我们引入量化指标：风险系数向量、各分行所在地区的国际化指数SBII、总体区域化差异以及各地区公平差异向量，分别建立收益与风险、企业经营与国际政策导向、区域化差异与分行公平考核与贷款规模分配比例向量的函数关系，进而得到商业银行A总体效益与贷款规模分配比例向量的关系式。最终即可计算得到使得商业银行A总体效益最大时的贷款规模分配方案。

### 五、模型建立与求解

经过以上的分析和准备，建立了数学模型，求得问题2-4的分配方案如下：

表5 问题2-4贷款分配方案结果

单位：亿元		问题 2	问题 3	问题 4
序号	分行	贷款规模分配方案	贷款规模分配方案	2018 年备付资金
1	北京	3206.966953	3206.966974	167.16812
2	江苏	4756.121517	4880.007678	161.208657
3	广东	3873.057608	3873.057704	131.807727
4	浙江	3907.338599	3907.338777	124.094153
5	上海	2709.536293	2709.536324	110.516718
6	山东	1828.689509	1828.689527	76.031349
7	河南	1501.007807	1540.105685	59.179048
8	湖北	1698.547996	1742.791358	54.976219
9	安徽	1184.299950	1215.148283	43.059693
10	四川	1028.578481	1055.370627	42.098096
11	辽宁	1157.738793	1157.738863	40.263087

12	河北	859.111589	881.4894815	33.914674
13	重庆	734.061208	753.1818351	34.008316
14	湖南	1214.490766	1246.125526	33.133276
15	山西	813.907864	835.1083126	30.320751
16	陕西	705.376612	723.7500598	29.148729
17	天津	783.039898	783.0399131	24.6178
18	广西	798.675545	820.193015	25.570082
19	吉林	552.446634	566.8366139	22.564667
20	云南	690.998179	694.3826687	22.271826
21	福建	1016.437342	1016.43738	21.889466
22	江西	683.943893	701.7590792	21.393474
23	黑龙江	427.422939	427.4229827	21.00738
24	新疆	534.269119	548.1855876	18.387902
25	贵州	454.345207	466.1798712	13.625914
26	甘肃	335.837476	344.5852639	13.333374
27	海南	460.219403	472.2070702	10.886036
28	内蒙古	427.709500	438.8503241	8.758563
29	宁夏	164.132965	168.4082409	4.577643
30	青海	131.096031	134.5106516	3.407456
全行合计		38639.405677	39139.405677	1403.220197

下面我们将进一步阐述模型的实际建立过程与求解。

## 5.1 问题一的模型建立与求解

### 5.1.1 模型的建立

#### 5.1.1.1 建立回归预测模型

基于附件 1 宏观经济指标历史数据和附件 2 商业银行 A 各项存贷款历史数据，建立商业银行 A 存贷款金额关于这些宏观经济指标的回归预测模型。本文使用支持向量机 (svm) 回归模型[1]，该回归模型可以很灵活的解决各种非线性的分类回归问题，而且在数据量非海量的时候，分类准确率高、泛化能力强。

建立svm回归模型如下：

$$\begin{aligned} \text{最小化目标} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\lambda} (\xi_i + \xi_i^*) \\ \text{约束条件} \quad & \begin{cases} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned}$$

其中， $w$ ：模型参数； $\xi$ ：松弛变量； $b$ ：截距项； $x$ ：输入特征； $y$ ：存贷款

实际值；  $y'$ ：存贷款预测值；惩罚项  $C = 100, \varepsilon = 0.001$ 。

在满足约束条件的情况下最小化目标便能求得一组参数使得  $y_{\text{预测值}} = wx$  贴近  $y_{\text{实际值}}$ 。

### 5.1.1.2 输入特征处理

步骤一：由于最终是对2018年全年预测的存贷款数量进行预测，且考虑到附件2提供的数据也是全年的数据，因此本文对附件一的季度数据进行合并，得到按年度划分的数据，处理方法如下：

首先，对于绝对值数据，将每一季度的数据除以上一季度的数据求出当前季度的百分比提升，化为同比数据。

$$X_{\text{季度}i\text{同比}} = \frac{X_{\text{季度}i\text{绝对值}}}{X_{\text{季度}i-1\text{绝对值}}}$$

然后，使用下列公式对同比数据进行处理，即可得到年度数据。

$$X_{\text{年}} = \left(\frac{X_{\text{季度}1}}{100} + 1\right) * \left(\frac{X_{\text{季度}2}}{100} + 1\right) * \left(\frac{X_{\text{季度}3}}{100} + 1\right) * \left(\frac{X_{\text{季度}4}}{100} + 1\right)$$

步骤二：由于数据特征较多，本文采用主成分分析[2]对宏观数据进行处理。主成分分析如下，求得一个转换矩阵  $w \in R^{n*k}$ ，使原数据特征使用转换矩阵进行投影求得的  $t_i = w \cdot x_i$  方差最大，即尽可能在投影后保留的原数据方差。

### 5.1.1.3 实际值

由于附件2的存贷款数据是各地区分行的存贷款，因此需要进行加总，求得商业银行A年度总存贷款  $y_{\text{总}}$ 。

$$y_{\text{总}} = \sum y_{\text{分行}}$$

## 5.1.2 模型的求解

根据上述本文所建立的回归预测模型，结合实际数据详细描述求解过程如下。

### 5.1.2.1 处理绝对值数据

将金融机构人民币各项存款余额（亿元）特征列处理为等比数据，如表5-1-1所示。

**表5-1-1 金融机构人民币各项存款余额（亿元）特征列等比数据**

季度	2010-03	2010-06	2010-09	2010-12	.....
存款余额（亿元 同比增长）	1.056431	1.056431	1.039944	1.024556	.....
季度	2016-12	2017-03	2017-06	2017-09	2017-12
存款余额（亿元 同比增长）	1.013904	1.033617	1.025795	1.016361	1.011268

### 5.1.2.2 缺失值处理

由于宏观数据存在缺失值影响回归模型的输入，所以需要修复缺失值。将不含缺失值的属性列  $X$  作为输入特征，含有缺失值的属性列  $y$  作为目标，利用svm模型回归预测。预测补全的缺失值如表5-1-2所示，红色即为预测值。

**表5-1-2 预测补全缺失值的宏观数据**

年份 \ 指标	国房景气指数:季(%)	70个大中城市新建住宅价格指数:同比:季(%)	固定资产投资完成额:实际当季同比(%)	固定资产投资价格指数:季	铁路货运量:同比:季(%)	金融机构人民币各项存款余额(亿元)同比
2017-12	98.804028	4.26978	14.457951	101.626112	3.70047	1.011268
2018(预测)	99.167213	3.61	12.587315	99.65	16.16	1.029508

### 5.1.2.3 计算年度宏观经济指标数据

将补全的宏观经济指标数据进行输入特征处理，得到年度宏观经济指标数据，部分数据如表5-1-3所示，全部数据见附录。

**表5-1-3 年度宏观经济指标数据**

年份 \ 指标	70个大中城市新建住宅价格指数:同比:季(%)	CPI:同比:季(%)	GDP::当季同比(%)	M1:同比:季(%)	M2:同比:季(%)	...
2010	1.584296	1.139604	1.501509	2.569785	2.125769	...
2011	1.176602	1.234518	1.442848	1.542367	1.729287	...
2012	0.972919	1.110281	1.351675	1.206064	1.661366	...
2013	1.255222	1.109089	1.349183	1.48651	1.739794	...

2014	1.105788	1.08199	1.325553	1.21663	1.631564	...
2015	0.855909	1.058744	1.307123	1.390661	1.590999	...
2016	1.267919	1.082743	1.297372	2.267097	1.575535	...
2017	1.357604	1.063661	1.304681	1.767111	1.439117	...
2018	1.152409	1.058439	1.296157	1.630474	1.411582	...

#### 5.1.2.4 主成分分析

设定主成分分析中保留90%以上的原数据方差，求得成分数为1，其数值如表5-1-4所示。

表5-1-4 主成分

年份	2010	2011	2012	2013	2014	2015	2016	2017	2018
主成分	0.313038	0.440366	0.409925	0.37265	0.324459	0.219306	0.235775	0.316886	0.302989

#### 5.1.2.5 使用svm模型进行回归预测

将主成分作为输入特征  $X$ ，总存贷款作为  $y_{\text{存款}}, y_{\text{贷款}}$ ，分别建立svm模型进行回归预测得到2018年的预测值如表5-1-5所示。

表5-1-5 2018年存贷款预测值

单位（亿元）	2018年预测值
存款	46944.55794439
贷款	38639.40567697

### 5.2 问题二的模型建立与求解

#### 5.2.1 模型的建立

对利率和存款进行 svm 回归预测得到 2018 年的预测值，然后对贷款值进行分位数回归[3]求得贷款规模的上下限，将利率和存款作为固定参数，贷款规模分配作为自变量，贷款规模上下限和问题 1 求得贷款总额作为约束条件，以利息收入为目标，建立数学规划模型。

##### 5.2.1.1 利率预测模型

采用和问题 1 一样的 svm 模型，输入特征  $X$  为对宏观数据进行主成分分析后的特征，实际值  $y$  为历年各地的利率，进行预测可得到贷款利率向量和存款利率向量  $L_r, D_r$ 。

### 5.2.1.2 存款预测

采用和问题 1 一样的 svm 模型，输入特征  $X$  为对宏观数据进行主成分分析的特征，实际值  $y$  为转化为等比数据的存款增长比例，进行预测可得各地存款 *Deposit Prediction*。

### 5.2.1.3 贷款规模分位数回归

分位数回归模型如下：

$$\text{最小化目标 } \min \left[ \sum q |y_i - x_i b| + \sum (1-q) |y_i - x_i b| \right]$$

其中， $q$  为分位量， $b$  为参数， $y$  为贷款规模等比增长实际值， $x$  为宏观数据进行主成分分析后的特征，进行预测可得贷款规模上下限 *Upper, Lower*。

### 5.2.1.4 贷款规模线性规划模型

该模型目标和约束如下

$$\text{最大化目标 } Lr \cdot X - Dr \cdot \text{Deposit Prediction}$$

$$\text{约束条件 } \begin{cases} \sum X_i = \text{Total Loan Prediction} \\ X_i \leq \text{Upper}_i, X \geq \text{Lower}_i \end{cases}$$

### 5.2.2.1 利率预测

使用 python 语言的 sklearn 包中的 svm 回归预测得利率如表 5-2-1 所示。

**表5-2-1 2018年各地区存贷款利率预测表**

地区	存款利率 (%)	贷款利率 (%)	地区	存款利率 (%)	贷款利率 (%)
北京	2.874199	4.597808	陕西	2.041477	5.066269
江苏	1.948459	5.010707	天津	2.647217	4.453429
广东	2.34408	4.850648	广西	2.068518	4.947748
浙江	2.527098	4.847808	吉林	2.211666	5.280589
上海	2.297217	4.690766	云南	1.979998	4.92047
山东	2.44698	4.57621	福建	2.636921	4.739109
河南	2.177098	5.154849	江西	2.265619	5.19337

湖北	2.147158	5.287748	黑龙江	2.337039	4.675204
安徽	2.244258	5.019228	新疆	1.435797	5.022186
四川	2.198518	5.096269	贵州	1.857158	5.624908
辽宁	2.507217	4.826269	甘肃	2.082956	5.216091
河北	2.400057	4.96621	海南	2.161359	5.303429
重庆	2.027217	5.139168	内蒙古	1.982838	5.022246
湖南	1.913016	5.464849	宁夏	2.230116	5.387748
山西	2.091418	5.082127	青海	1.97124	4.997689

### 5.2.2.2 存款预测

使用 python 语言的 sklearn 包中的 svm 回归预测得利率如表 5-2-2 所示。

**表5-2-2 2018年各地区存款预测表**

地区	存款	地区	存款
北京	5627.787346	陕西	970.25135
江苏	5426.682187	天津	817.352808
广东	4434.532531	广西	849.488059
浙江	4174.233957	吉林	748.068758
上海	3716.056361	云南	738.186695
山东	2552.329693	福建	725.283733
河南	1983.640058	江西	708.546227
湖北	1841.813418	黑龙江	695.51731
安徽	1439.684055	新疆	607.121658
四川	1407.234471	贵州	446.425905
辽宁	1345.311123	甘肃	436.553977
河北	1131.080633	海南	353.967288
重庆	1134.240628	内蒙古	282.174621

湖南	1104.711964	宁夏	141.08731
山西	1009.801836	青海	101.598723

### 5.2.2.3 贷款分位数预测

使用 python 语言的 statsmodels 包中的 QuantReg 分位数回归，选取 0.05 分位数和 0.95 分位数，分别预测得贷款上下界如表 5-2-3 所示。

**表5-2-3 2018年各地区贷款上下限预测表**

地区	贷款下限	贷款上限	地区	贷款下限	贷款上限
北京	3206.966953	3209.030001	陕西	705.207934	705.376612
江苏	4753.0146	4756.121517	天津	783.039898	785.551234
广东	3873.057608	3875.90424	广西	796.914787	799.371187
浙江	3907.338599	3910.216808	吉林	551.741436	552.446634
上海	2709.536293	2718.422491	云南	690.998178	694.108877
山东	1828.689508	1829.639835	福建	1016.437342	1017.550886
河南	1499.364126	1501.007808	江西	683.150797	683.943893
湖北	1696.661529	1698.547996	黑龙江	427.422939	428.031047
安徽	1181.765327	1184.299951	新疆	533.353024	534.26912
四川	1027.10141	1028.578481	贵州	453.77709	454.345207
辽宁	1157.738793	1160.396284	甘肃	335.597236	335.837476
河北	858.908499	859.11159	海南	457.971298	460.219403
重庆	732.962628	734.061208	内蒙古	426.830202	427.709501
湖南	1213.022475	1214.490767	宁夏	163.997988	164.132965
山西	812.813481	813.907864	青海	131.064466	131.096032

### 5.2.2.4 规划求解

使用 python 语言的 scipy 包中的 linprog 线性规划模块求解得最优贷款规模如表 5-2-4 所示。

表5-2-4 2018年各地区贷贷款规模分配方案

单位：亿元		问题 2
序号	分行	贷款规模分配方案
1	北京	3206.966953
2	江苏	4756.121517
3	广东	3873.057608
4	浙江	3907.338599
5	上海	2709.536293
6	山东	1828.689509
7	河南	1501.007807
8	湖北	1698.547996
9	安徽	1184.299950
10	四川	1028.578481
11	辽宁	1157.738793
12	河北	859.111589
13	重庆	734.061208
14	湖南	1214.490766
15	山西	813.907864
16	陕西	705.376612
17	天津	783.039898
18	广西	798.675545
19	吉林	552.446634
20	云南	690.998179
21	福建	1016.437342
22	江西	683.943893
23	黑龙江	427.422939
24	新疆	534.269119
25	贵州	454.345207
26	甘肃	335.837476
27	海南	460.219403
28	内蒙古	427.709500
29	宁夏	164.132965
30	青海	131.096031
全行合计		38639.405677

### 5.3 问题三的模型建立与求解

#### 5.3.1 模型的建立

本题在题目 2 的模型基础上将贷款总额度增加 500 亿，贷款规模上限按照贷款

总额度的上涨比例进行调整，其余部分和题目 2 相同。

$$\text{贷款总额调整 } TotalLoan' = TotalLoan + 500$$

$$\text{贷款规模上限调整 } Upper' = Upper * \frac{TotalLoan'}{TotalLoan}$$

### 5.3.2 模型的求解

使用 python 语言的 scipy 包中的 linprog 线性规划模块求解得最优贷款规模如表 5-3-1 所示。

**表5-3-1 2018年各地区贷贷款规模分配方案**

单位：亿元		问题 2
序号	分行	贷款规模分配方案
1	北京	3206.966974
2	江苏	4880.007678
3	广东	3873.057704
4	浙江	3907.338777
5	上海	2709.536324
6	山东	1828.689527
7	河南	1540.105685
8	湖北	1742.791358
9	安徽	1215.148283
10	四川	1055.370627
11	辽宁	1157.738863
12	河北	881.4894815
13	重庆	753.1818351
14	湖南	1246.125526
15	山西	835.1083126
16	陕西	723.7500598
17	天津	783.0399131
18	广西	820.193015
19	吉林	566.8366139
20	云南	694.3826687
21	福建	1016.43738
22	江西	701.7590792
23	黑龙江	427.4229827
24	新疆	548.1855876
25	贵州	466.1798712
26	甘肃	344.5852639
27	海南	472.2070702
28	内蒙古	438.8503241
29	宁夏	168.4082409

30	青海	134.5106516
全行合计		39139.405677

## 5.4 问题四的模型建立与求解

### 5.4.1 模型的建立

统计各地分行每日的存取款之差，作频率图可发现大多数具有类似正态分布的形状，考虑到用户存取款的行为存在随机性，所以假设每日存取款的额度服从正态分布，部分地区分行存取款频率图如图5-1所示。

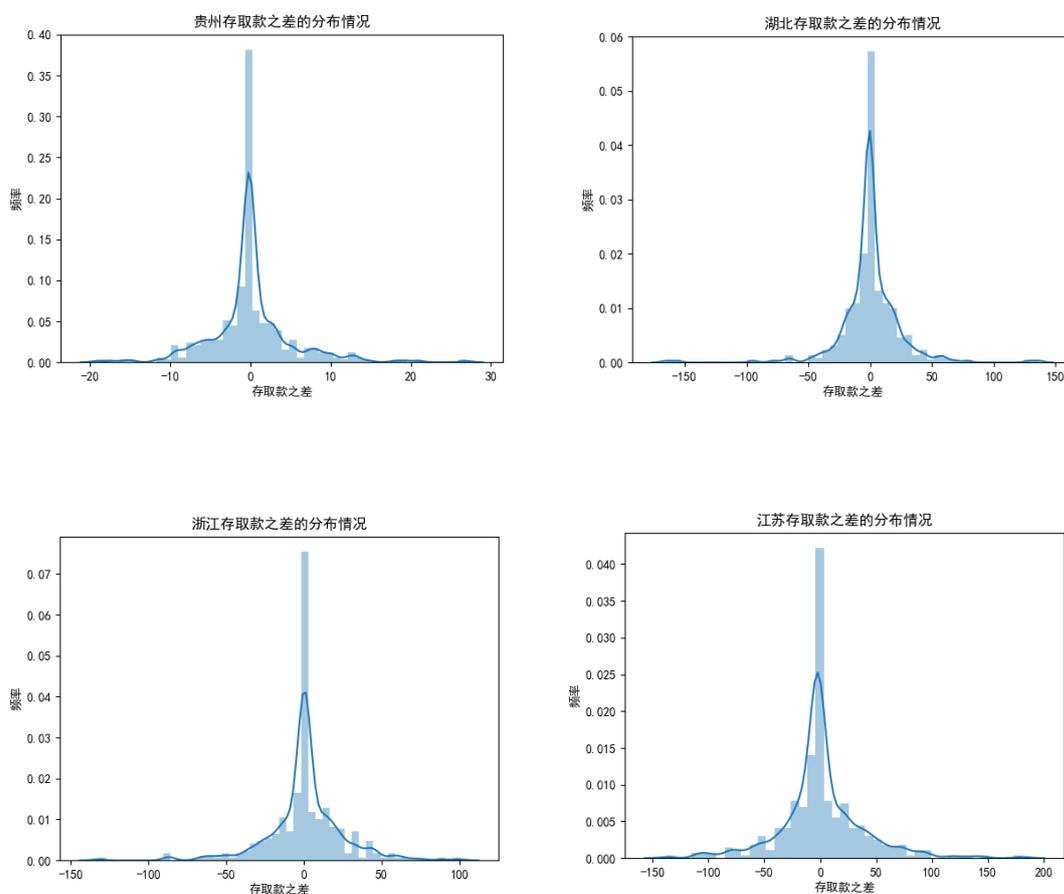


图 5-1 部分地区分行存取款频率图

以各分行的年度存款总额为和存取款差额的均值作散点图如下，可以推测分行存款总额和分行存取款差额的均值存在相关性。

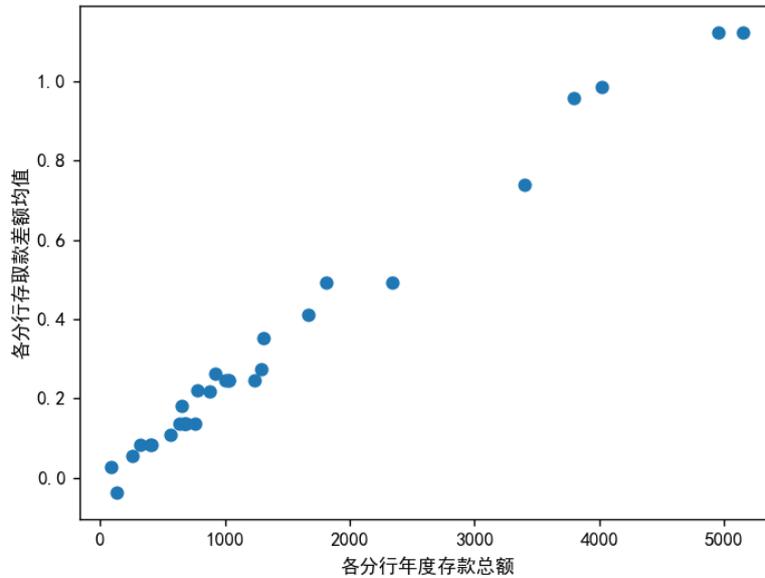


图 5-2 各分行存取款差额均值与年度存款总额散点图

以各分行的年度存款总额和存取款差额的方差作散点图如下，可以推测分行存款总额和分行存取款差额的方差存在相关性。

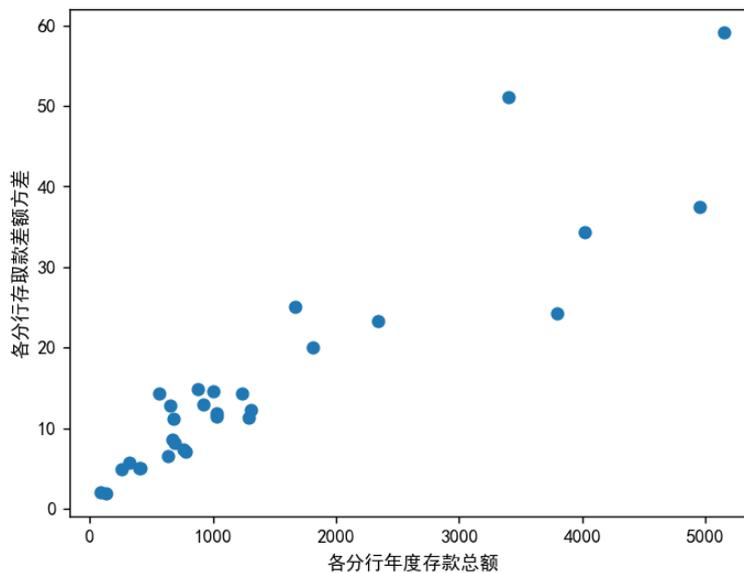


图 5-3 分行存取款差额方差与年度存款总额散点图

基于上述分析，可假设2018年的各分行的存取款数额也符合正态分布，可以2017年的数据为基础，建立回归预测模型对2018年各分行存取款数额分布的参数进行预测。

#### 5.4.1 模型的建立

##### 5.4.1.1 正态分布模型

根据上述分析，本文假设商业银行 A 各个地区分行的存取款差值符合正态分布，则其均值估计和方差估计可分别由下列公式求得：

$$\mu' = \frac{\sum \text{Difference}}{365}$$

$$\sigma' = \sqrt{\frac{365 \cdot \sum \text{Difference}^2 - (\sum \text{Difference})^2}{365 \cdot (365 - 1)}}$$

其中，*Difference* 为每日存取款的差额，365为样本数量。

#### 5.4.1.2 svm回归预测模型

以2017年的各分行存款数额为输入特征 *X*，各个分行的分布均值  $\mu'$  和方差  $\sigma'$  分别为实际值 *y*，建立与题目1相同的svm回归模型。

将2018年的各分行存款预测数额作为模型的输入特征，即可对2018年各个分行的分布均值  $\mu'$  和方差  $\sigma'$  进行预测。

#### 5.4.2 模型的求解

##### 5.4.2.1 回归预测均值和方差

使用python语言的sklearn包中的svm模块进行计算，以题目2中获得的2018年各分行存款预测数据作为svm模型的输入特征，使用预测2018年各行存取款差值分布的均值和方差如表5-4-1所示。

**表 5-4-1 2018 年各行存取款差值分布均值和方差**

地区	均值	方差	地区	均值	方差
北京	1.174091	65.354568	陕西	0.200325	11.394021
江苏	1.132045	63.024635	天津	0.168358	9.622594
广东	0.924614	51.529944	广西	0.175077	9.994901
浙江	0.870192	48.514218	吉林	0.153873	8.819893
上海	0.774399	43.205936	云南	0.151807	8.705403
山东	0.531095	29.723415	福建	0.149109	8.555914
河南	0.412198	23.134781	江西	0.14561	8.361999
湖北	0.382546	21.491628	黑龙江	0.142886	8.211051
安徽	0.298471	16.832701	新疆	0.124405	7.186931
四川	0.291687	16.456752	贵州	0.090807	5.325167
辽宁	0.27874	15.73933	甘肃	0.088744	5.210795
河北	0.23395	13.257332	海南	0.071477	4.253975
重庆	0.234611	13.293943	内蒙古	0.056467	3.422211
湖南	0.228437	12.951834	宁夏	0.026969	1.787623
山西	0.208594	11.852239	青海	0.018713	1.330123

#### 5.4.2.2 置信区间预测

根据表5-4-1的均值和方差，对各分行贷款额度分配上下界区间进行预测，预测结果如表5-4-2所示。

**表 5-4-2 各分行贷款额度分配上下界**

地区	下界	上界	地区	下界	上界
北京	-167.16812	169.516302	陕西	-29.148729	29.54938
江苏	-161.208657	163.472748	天津	-24.6178	24.954517
广东	-131.807727	133.656954	广西	-25.570082	25.920236
浙江	-124.094153	125.834537	吉林	-22.564667	22.872413
上海	-110.516718	112.065516	云南	-22.271826	22.57544
山东	-76.031349	77.09354	福建	-21.889466	22.187684
河南	-59.179048	60.003444	江西	-21.393474	21.684693
湖北	-54.976219	55.74131	黑龙江	-21.00738	21.293152
安徽	-43.059693	43.656635	新疆	-18.387902	18.636711
四川	-42.098096	42.68147	贵州	-13.625914	13.807529
辽宁	-40.263087	40.820567	甘肃	-13.333374	13.510861
河北	-33.914674	34.382575	海南	-10.886036	11.02899
重庆	-34.008316	34.477538	内蒙古	-8.758563	8.871497
湖南	-33.133276	33.590151	宁夏	-4.577643	4.631582
山西	-30.320751	30.737939	青海	-3.407456	3.444883

上表是由存取款差额分布的 99%置信区间计算得来的，该表的下界即每日存款小于取款的差额的最大值，可以将此值看作各分行日常经营所需最低备付金额，因此可得最低备付金如表 5-4-3 所示。

**表 5-4-3 2018 年各分行备付资金**

单位：亿元		问题 4
序号	分行	2018 年备付资金
1	北京	167.16812
2	江苏	161.208657
3	广东	131.807727
4	浙江	124.094153
5	上海	110.516718
6	山东	76.031349
7	河南	59.179048
8	湖北	54.976219
9	安徽	43.059693
10	四川	42.098096
11	辽宁	40.263087

12	河北	33.914674
13	重庆	34.008316
14	湖南	33.133276
15	山西	30.320751
16	陕西	29.148729
17	天津	24.6178
18	广西	25.570082
19	吉林	22.564667
20	云南	22.271826
21	福建	21.889466
22	江西	21.393474
23	黑龙江	21.00738
24	新疆	18.387902
25	贵州	13.625914
26	甘肃	13.333374
27	海南	10.886036
28	内蒙古	8.758563
29	宁夏	4.577643
30	青海	3.407456
合计		1403.220197

## 5.5 问题五的模型建立与求解

### 5.5.1 模型的建立

在实际问题中，贷款规模分配不仅仅是依据各个地区分行存贷款的利率，而是需要综合考虑许多内外部影响因素的[8]。本文为了帮助商业银行A处理好收益与风险、企业经营与国际政策导向、区域化差异与分行公平考核等之间的关系，进而达到双赢或多赢的目标，在上述提出的贷款规模分配模型的基础上引入风险系数 $\rho$ 、各分行所在地区的国际化指数 $SBII_k$ 、不公平差异 $d_k$ ，建立贷款规模分配综合模型。

$$W = Af(x)' + Bg(x)' + Ch(x)'$$

其中，A、B、C分别代表收益与风险( $A_1$ )、企业经营与国际政策导向( $A_2$ )、区域化差异与分行公平考核( $A_3$ )这三对因素对商业银行A所得效益的影响程度的权数。 $f(x)'$ 、 $g(x)'$ 、 $h(x)'$ 分别是这三对影响因素关于贷款规模分配比例向量 $\alpha$ 的函数，且 $f(x)'$ ， $g(x)'$ ， $h(x)'$   $\in (-1,1)$ 。

$x$ 代表商业银行A分配给各地区分行的贷款额度，贷款规模分配比例向量 $\alpha$ 与 $x$ 的关系以及 $x_k$ 的约束条件如下式所示。

$$\alpha_k = \frac{x_k}{\sum_{k=1}^{30} x_k}$$

$$\text{约束条件} \begin{cases} \sum_{k=1}^{30} x_k = TotalLoan Prediction \\ x_k \leq Upper_k, x_k \geq Lower_k \end{cases}$$

#### 5.5.1.1 AHP方法确定因素权重[4]

步骤一：明确系统目标，即追求商业银行A的整体效益 $W$ 最大。

步骤二：把系统中各要素规划不同层次，建立层次结构模型，如图5-4所示。

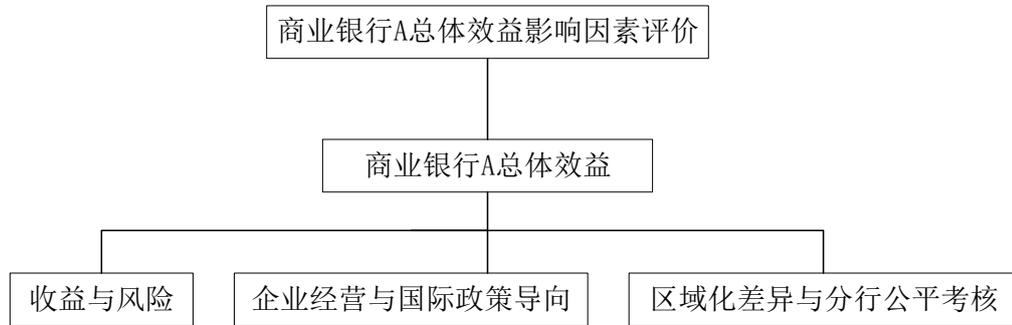


图 5-4 AHP 层次结构图

步骤三：建立判断矩阵。结合已有文献综合分析，按照语言评价集{“1”同等重要，“3”稍微重要，“5”明显重要，“7”强烈重要，“9”极端重要}建立判断矩阵，如表5-1所示。矩阵中的元素 $a_{ij}$ 代表以商业银行A的效益 $W$ 为准则，因素 $A_i$ 对 $A_j$ 的相对重要程度，且  $i, j = 1, 2, 3$ 。

表 5-5-1 AHP 判断矩阵

$W$	$A_1$	$A_2$	$A_3$
$A_1$	1	7	9
$A_2$	1/7	1	1/3
$A_3$	1/9	3	1

步骤四：层次单排序。基于上述判断矩阵，首先将判断矩阵每一列做归一化处理：

$$\bar{a}_{ij} = \frac{a_{ij}}{\sum_{t=1}^n a_{tj}} = \begin{bmatrix} 0.797 & 0.636 & 0.871 \\ 0.113 & 0.091 & 0.032 \\ 0.089 & 0.273 & 0.097 \end{bmatrix}$$

完成归一化后，求出每一行各元素之和：

$$\bar{\theta}_i = \sum_{j=1}^n \bar{a}_{ij} = \begin{bmatrix} 2.305 \\ 0.237 \\ 0.458 \end{bmatrix}$$

对进行归一化处理：

$$\theta_i = \frac{\bar{\theta}_i}{\sum_{j=1}^n \bar{\theta}_j} = \begin{bmatrix} 0.768 \\ 0.079 \\ 0.153 \end{bmatrix}$$

即得到收益与风险 ( $A_1$ )、企业经营与国际政策导向 ( $A_2$ )、区域化差异与分行公平考核 ( $A_3$ ) 这三对因素对商业银行A所得效益的影响程度的权数分别为：

$$A = 0.768, B = 0.079, C = 0.153$$

### 5.5.1.2 收益与风险

本文为了计算收益与风险关于贷款规模分配向量 $\alpha$ 的函数引入城市风险系数向量 $\rho$ [5-7]。所谓风险系数向量 $\rho$ ，指的是各个地区分行客户不按期缴付贷款的违约事件发生率(%)，主要考虑信用风险，可通过各地区不良信用评分得到。则收益与风险因素关于贷款规模向量 $\alpha$ 的函数 $f(x)$ 可由下列公式得到。

$$f(x) = \alpha \cdot Lr - \alpha \cdot \rho \cdot Lr = \sum_{k=1}^{30} \alpha_k Lr_k (1 - \rho_k)$$

$$\text{约束条件} \begin{cases} \sum_{k=1}^{30} x_k = TotalLoan Prediction \\ x_k \leq Upper_k, x_k \geq Lower_k \end{cases}$$

表 5-5-2 城市风险系数

地区	风险系数	地区	风险系数
北京	5.1%	陕西	14.9%
江苏	4.9%	天津	10.1%
广东	15.4%	广西	7.4%
浙江	5.2%	吉林	2.2%
上海	6.0%	云南	2.8%

山东	7.0%	福建	5.4%
河南	9.4%	江西	12.3%
湖北	22.1%	黑龙江	2.5%
安徽	4.6%	新疆	10.9%
四川	2.6%	贵州	3.7%
辽宁	3.6%	甘肃	2.6%
河北	4.7%	海南	16.2%
重庆	3.0%	内蒙古	3.3%
湖南	12.3%	宁夏	4.8%
山西	9.2%	青海	9.3%

数据来源：信用中国（<https://creditcity.creditchina.gov.cn/>）

### 5.5.1.3 企业经营与国际政策导向

商业银行企业经营与国际政策导向因素主要可以量化为各分行所在地区的国际化指数 $SBII_k$ 等指标。分行地区的开放程度越高且分行利润越高，说明商业银行A在企业经营与国际政策导向方面做的越好。

$$g(x) = SBII \cdot \alpha = \sum_{k=1}^{30} SBII_k \alpha_k$$

$$\text{约束条件} \begin{cases} \sum_{k=1}^{30} x_k = TotalLoan Prediction \\ x_k \leq Upper_k, x_k \geq Lower_k \end{cases}$$

表5-5-3 地区国际化指数表

地区	国际化指数	地区	国际化指数
北京	44.87436667	陕西	27.14116667
江苏	31.6382375	天津	34.70025
广东	35.34161667	广西	20.267866
浙江	34.34117167	吉林	22.33826
上海	40.3371875	云南	24.2167325

山东	27.80824833	福建	37.19485833
河南	21.56343	江西	21.79247
湖北	26.10281571	黑龙江	22.8049375
安徽	24.203404	新疆	16.7407442
四川	28.34092	贵州	19.1211
辽宁	22.3948095	甘肃	18.1334685
河北	23.1448	海南	32.7335436
重庆	32.71888889	内蒙古	16.878976
湖南	21.430387	宁夏	17.965654
山西	23.0169083	青海	17.43111

数据来源：“城区国际化与决胜全面小康”分论坛

对 $g(x)$ 进行归一化处理，将数据映射到0~1范围之内，得到企业经营与国际政策导向因素关于贷款规模分配比例向量 $\alpha$ 的函数  $g(x)'$ 。

$$g(x)' = \frac{g(\alpha)}{\sum_{k=1}^{30} (\max\{SBII\}\alpha_k)}$$

其中， $k = 1, 2, \dots, 30$ 。

#### 5.5.1.4 区域化差异与分行公平考核

为了平衡区域化差异和分行公平考核这一对相互矛盾的因素,本文引入区域不公平差异 $d_k$ ，我们认为按照各地存款分配各地贷款额度是公平合理的，所以用某地贷款额度分配比例与存款比例之差的平方代表不公平的地区差异，即：

$$d_k = (\alpha_k - \beta_k)^2$$

其中， $k = 1, 2, \dots, 30$ ， $n=30$ 。

对各地不公平差异进行求和得到总的该银行不公平差异，对进行归一化处理，由于在目标函数 $W = Af(x)' + Bg(x)' + Ch(x)'$ 中不公平差异是一个负向影响，将数据映射到-1~0范围之内，可以得到区域化差异与分行公平考核因素关于贷款规模分

配比例向量 $\alpha$ 的函数 $h(x)$ 。

$$h(x) = -\frac{\sum_{k=1}^{30} d_k}{30}$$

$$\text{约束条件} \begin{cases} \sum_{k=1}^{30} x_k = \text{TotalLoan Prediction} \\ x_k \leq \text{Upper}_k, x_k \geq \text{Lower}_k \end{cases}$$

其中,  $k = 1, 2, \dots, 30$ 。

### 5.5.2 模型的求解

使用python语言的scipy包中的optimize模块对指标 $W$ 进行优化求得如下贷款规模方案, 且该贷款方案下指标 $W = 0.7716$ , 求得2018年贷款规模预测值如表5-5-4所示。

**表5-5-4 2018年贷款规模预测值**

单位: 亿元		问题 5
序号	分行	贷款规模分配方案
1	北京	3206.966955
2	江苏	4753.014601
3	广东	3873.057609
4	浙江	3907.338600
5	上海	2709.536294
6	山东	1828.689508
7	河南	1499.364126
8	湖北	1696.661529
9	安徽	1181.765327
10	四川	1027.101410
11	辽宁	1157.738793
12	河北	858.908499
13	重庆	732.962628
14	湖南	1213.022475
15	山西	812.813481
16	陕西	705.207934
17	天津	783.039898
18	广西	796.914787
19	吉林	551.741436
20	云南	690.998178
21	福建	1016.437343

22	江西	683.150797
23	黑龙江	427.422939
24	新疆	533.353024
25	贵州	453.777090
26	甘肃	335.597236
27	海南	457.971298
28	内蒙古	426.830202
29	宁夏	163.997988
30	青海	131.064466
全行合计		38616.446451

## 六、模型评价与推广

### 6.1 模型评价

问题一将季度数据合并为年度数据，采用主成分分析对宏观经济指标数据进行处理修复其缺失值，建立了svm回归预测模型。该模型能够比较好的反映各宏观经济指标与商业银行A各地区分行存贷款的关系，分类准确率高、泛化能力强，使得2018年各分行存贷款的预测结果较为准确。

问题二以利息收入为目标，以贷款总额及各地区分行贷款额度上下限为约束条件，建立了基于多次svm回归预测的数学规划模型。对利率和存款进行svm回归预测求得2018年A行各地预测值，对贷款值进行分位数回归求得贷款规模的上下限。使得模型预测结果更加合理准确，具有比较高的可靠性。

问题三在问题二模型的基础上，将贷款总额度增加500亿，并调整各地区分行的贷款规模上限，得出的预测结果更加贴近实际。

问题四通过描绘各地区分行存取款差额数据图示发现其正态分布规律，进而建立了回归预测模型。该模型能够较好的拟合数据，具有比较高的预测精度。

问题五在原贷款分配模型的基础上，综合考虑收益与风险、企业经营与国际政策导向、区域化差异与分行公平考核这三对影响因素，并结合层次分析法建立了数学规划模型。该模型综合考虑了现实情况中商业银行A内外部因素对贷款分配方案的影响，能够较为客观的反映实际状况，有利于帮助商业银行A平衡因素之间的矛盾，实现多赢。

### 6.2 模型推广

本文所建立的预测模型准确度较高，实用且贴近实际，可通过海量的数据达

到较好的预测效果，预测结果准确可靠，能够为商业银行的经营提供一定的参考依据，同时该模型也可以应用于其他预测问题。

## 七、参考文献

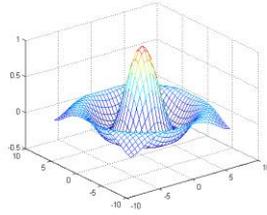
- [1]周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 133-137
- [2]周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 229-232
- [3]Koenker, Roger. Quantile Regression[M]. Cambridge: Cambridge University Press, 2005
- [4]楚存坤, 孙思琴, 韩丰谈. 基于层次分析法的高校图书馆学科服务评价模式[J]. 大学图书馆学报, 2014, 32(06):86-90.
- [5]原利斌, 潘焕学, 俞惠倩. 基于风险调整收益的银行贷款定价模式研究[J]. 山西财经大学学报, 2011, 33(S1):50-51.
- [6]许亚菲. 应用多元数据建模对我国商业银行经营管理能力评价[J]. 中国商界(下半月), 2010(08):44-47.
- [7]张开春. A 银行 B 分行风险管理研究[D]. 江南大学, 2008.
- [8]张玉庆, 马树江, 宋利鹏. 现代商业银行运营管理模式探讨[J]. 华北金融, 2011(05):30-32.

## 八、附录

附录 1 问题一的宏观经济指标年度历史数据

	2010	2011	2012	2013	2014	2015	2016	2017	2018
70 个大中城市新建住宅价格指数:同比:季(%)	1.584296	1.176602	0.972919	1.255222	1.105788	0.855909	1.267919	1.357604	1.152409
CPI:同比:季(%)	1.139604	1.234518	1.110281	1.109089	1.081990	1.058744	1.082743	1.063661	1.058439
GDP:同比:季(%)	1.501509	1.442848	1.351675	1.349183	1.325553	1.307123	1.297372	1.304681	1.296157

# 五一数学建模竞赛



## 题目：商业银行存贷款预测及最优化贷款分配盈利方案

**关键词：**线性回归预测；熵值法多变量综合评价；线性规划最优解；正态分布

**摘要：**本文基于商业银行人民币贷款规模分配及盈利问题，通过线性回归预测、熵值法多变量综合评价、线性规划最优解、正态分布拟合等方法建立了预测商业银行下一年存贷款总量和贷款增量分配盈利方案的数学模型，具有较强的现实意义。

针对问题一，通过拟合得到缺失的宏观经济学指标，再由熵值法确定各宏观经济学指标在总体评价中所占权重，进而得到反映各季度客户存取款能力的量化指标。对每年的存款能力量化指标与往年存款总量拟合确定两者相关程度，从而由 2018 年的存款能力指标预测出当年的存款总量为 46370 亿元。通过线性回归研究各年存贷款之间的关系，可以得到 2018 年的贷款总量为 38038 亿元。和 2017 年实际的数据作差，得到 2018 年存贷款的增量分别为 3490 亿元、3161 亿元。

针对问题二，通过对附件 2 和附件 3 的数据处理，得到各分行利率和存贷款需求的矩阵，以此建立线性规划模型，根据实际的存贷款需求和市场波动对存款增量和贷款增量进行限制。通过 LINGO 软件编程，得最优化的贷款增量分配方案，将其与 2017 年贷款金额相加作为 2018 年贷款规模分配方案，净收入最大为 82.55 亿元。

针对问题三，考虑到在 500 亿元普通债的作用下，一方面总行的存款增量会增大导致其市场的波动也会有所改变；另一方面债券并非在年初发行，一年的贷款增量的配额会与两种贷款分配模式相关。由此对问题二中的模型进一步优化，调整波动范围并将两种贷款分配模式取不同的权重，用问题二的模型结果的处理方法求得此时最优化的贷款分配方案，净收入最大为 73.81 亿元。最后将两个问题中的净收入进行比较，给出合理的解释。

针对问题四，引入备付金额成本作为判断银行每日交易正常进行的依据，由理论推导得出当备付金额与资金流量近似相等时为最低备付水平。因此首先计算出 2017 年各分行日资金流量，选取北京和河南两个分行为例，绘制频数分布直方图，拟合曲线较好符合正态分布规律。进而通过对标准正态分布概率密度函数积分得到置信水平为 99% 时的日资金流量，利用模型一中的存款能力量化指标预测得到该分行 2018 年日常经营所需最低备付金额，至此完成 2018 年贷款分配方案与备付资金，填入表 5。

针对问题五，本文从收益与风险，经营与政策，差异与公平三个角度对上述模型进行优化，相继引入风险系数、利率修正系数、历史经验系数、公平系数对分配方案进行修正与改进，在此基础上对商业银行达到多赢提出了五点建议。

# 目录

一、问题重述 .....	1
1.1 问题背景 .....	1
1.2 问题分析 .....	1
二、模型假设与符号定义 .....	3
2.1 模型假设 .....	3
2.2 符号定义 .....	3
三、模型建立与求解 .....	4
3.1 问题一模型的建立与求解 .....	4
3.1.1 模型假设 .....	4
3.1.2 模型建立 .....	4
3.1.3 模型求解 .....	7
3.2 问题二模型的建立与求解 .....	9
3.2.1 模型假设 .....	9
3.2.2 模型建立 .....	9
3.2.3 模型求解 .....	10
3.3 问题三模型建立与求解 .....	10
3.3.1 模型假设 .....	10
3.3.2 模型建立 .....	11
3.3.3 模型求解 .....	11
3.3.4 与问题二结果比较 .....	12
3.4 问题四模型的建立与求解 .....	12
3.4.1 模型假设 .....	12
3.4.2 模型建立 .....	12
3.4.3 模型求解 .....	14
四、模型优化与评价 .....	17
4.1 模型优化 .....	17
4.1.1 收益与风险 .....	17
4.1.2 经营与政策 .....	17
4.1.3 差异与公平 .....	17
4.2 模型评价 .....	18
4.2.1 模型优点 .....	18
4.2.2 模型缺点 .....	18
4.3 改进建议 .....	18
参考文献 .....	19
附件 .....	19

# 一、问题重述

## 1.1 问题背景

商业银行的运营模式一般是从客户端吸收资金，用于发放贷款和其他投资。从贷款和投资中获得较大的收益以支付客户的利息并获得收入。同时商业银行需要缴纳法定的准备金并且预留一些资金来保障流通性安全。

商业银行的贷款规模受制于其存款规模。某家商业银行在全国存款总额中占的比例一般是稳定的。全国存款总额受限于社会资金总量的增长情况，与 GDP、CPI 等宏观经济指标密切相关。因而一个商业银行的存款规模受限于社会资金总量的增长情况。

在此之前，商业银行需要根据往年的存、贷款数据，预测下一年商业银行的存款和贷款总量。影响商业银行存款和贷款总量的宏观经济学指标有很多，银行需要根据已有的宏观经济学指标，分析这些指标对存贷款总量有无影响，有多大的影响。通过量化这些指标建立宏观经济学指标和存贷款总量的函数，以预测下一年商业银行的存款和贷款总量。根据预测的存、贷款指标，结合已有的各分行的存、贷款利率，要研究各分行的贷款分配数量使得全行的收益最大。

以前银行多采取这种年初预确定分配全年的贷款规模至各家分行，年度内不再调增。不考虑全年的全年实际可支配贷款规模、环境变化、需求变化等问题，资金利用总效益低。

目前，商业银行多改进了贷款分配的模式，采用以存定贷、存贷结合等方式努力获取全行最大收益，协调各区域发展，支持实体经济发展。多种贷款分配模式有助于规避风险、提高收益，同时能够协调各方面关系，以达到共赢的目标。商业银行需要考虑多种多样灵活实用的贷款分配的模式以达到银行最优的结果。

## 1.2 问题分析

针对**问题一**，宏观经济指标历史数据实质反映了不同季度客户存款和贷款能力。客户存款和贷款能力受各种宏观经济指标的影响。这些指标对存贷款能力的影响并不完全相同。通过研究宏观经济指标历史数据，并且量化这些指标，应当会发现这些指标和存、贷款变化的关系。

通过这些量化的指标以及各指标和存贷款变化的关系可以量化另一个数，用来表示某季度客户存款和贷款的能力。表示能力的数在研究的范围内应当与实际存、贷款数据之间成线性的关系。

表示出附件 1 中三年各个季度客户的存贷款能力之后，再与附件 2 中这三年实际的存贷款数据建立函数关系。建立多种函数关系，分析选择最合适的一种函数作为表示能力的数与实际的存贷款数据之间函数关系。然后带入 2018 年的预测的客户存款和贷款能力数便可以得到 2018 年预测的存、贷款数据。与 2017 年相比较，便可得出该银行 2018 年存、贷款增量情况。

针对**问题二**，通过问题一，可以预测得到该银行 2018 年存、贷款增量情况。以此数据为依据，考虑给各个分行的贷款增量的分配额，使得全行的增量存贷款利息净收入最大。

假设各分行的存款指标全部达到，贷款全部能放出。设第  $i$  个城市的存款增量为  $x_{i1}$ ，分配的贷款数量为  $x_{i2}$ 。由附件 2 每个分行的存款利息、贷款利息与该地的实

际情况有关。一般而言，各分行的利率会在一个值上下波动，类似于价格围绕价值上下波动的规律。可以认为 2018 年各个银行的利率是前三年利率的平均值。

决策变量是  $x_{i1}$  和  $x_{i2}$ 。目标函数是每个分行的贷款增量收入减去存款增量的利息支出后总的收入（净收入）最大。限制条件有总的贷款增量不能大于总行存款增量。理论的情况仅是每个分行的存款和贷款数量大于等于 0 即可，但是考虑到客户对各个分行实际的存款和贷款需求并且避免总行把贷款都较多地分给贷款利率高的分行。本文认为 2018 年实际贷款增量的分配会延续其前三年贷款分配的趋势，但是实际存款增量和贷款增量的情况可能会偏离趋势一个较小的范围。本文可以通过附件 3 中三年的数据得到前三年的贷款分配趋势并且给出存款增量和贷款增量各自波动的上限比例和下限比例，把 2018 年各个分行的存、贷款数量限制在存款和贷款各自波动的上下限内。

贷款会带来相应的收入，且贷款规模很大程度上受限于存款规模。在达到预测的存款时本文得出：各个分行分配的贷款增量总和等于总行的贷款增量。将决策变量和存、贷款利率各分行的写成矩阵的形式，通过矩阵相乘，表示出目标函数。通过 LINGO 软件求解决策变量并且求出目标函数的最佳值，并得到各分行贷款增量的分配方案，将其分别与 2017 年该行贷款金额相加，得到 2018 年各分行贷款规模分配方案。

针对**问题三**，500 亿元的商业银行普通债使得全行存款增量增加，相应的其贷款增量也必定增加。本文假设 5 月 1 日后相当短的时间内 500 亿元的普通债完全售出，转化为存款当量，相当于初始的存款增量增加了 500 亿。关于银行利率，本文假设以固定的年利率 5.1%按单利计算利息。本文假设 5 月 1 日之后 500 亿元的债券才对存贷款产生影响。在 5 月 1 日之前按问题 2 中的存贷款增量和约束条件进行分配。5 月 1 号之后初始的存款增量增加了 500 亿，按此时的存贷款增量和对应的约束条件计算。以 5 月 1 日为时间节点，目标函数由两种分配方式所获得的最大收益分别乘以相应的权重后相加得到。限制条件为总的贷款增量不能大于总行存款增量；各行的贷款增量的分配额要符合市场的趋势，把各行的贷款增量的分配额限制在市场趋势的波动范围内。与问题 2 相比，此时贷款增量的上限和下限有所变化，其他处理方式相同。

针对**问题四**，对附件 4 各分行 2017 年每日存取款交易数据进行分析，计算出各分行各天取款和存款的差值。随机选取若干分行（以北京和河南为例），对各取款和存款的差值进行分组。利用差值范围和和差值范围内的天数列出直方图，直方图连接各长方形的中点，观察差值范围和和差值范围内的天数的变化趋势。把离散型随机变量连续化处理，提出准备金概率密度的概念。通过拟合，得到概率密度函数，对概率密度函数积分求概率分布函数  $F(x)$ 。令  $F(x) > 0.99$  得到  $x$  的值，即为在置信水平 99%的情况下 2017 年该分行日常经营所需最低备付金额。对每个分行进行上述步骤就可以得到 2017 年各个分行日常经营所需最低备付金额。

本文认为一年的最低备付金额由该年的客户存款能力和贷款能力决定，而贷款能力又限制于存款能力。本文简化处理为最低备付金额于客户存款能力成正比。通过问题一得到的存款能力值和 2017 年商业银行 A 各分行日常经营所需最低备付金额便得出 2018 年商业银行 A 各分行日常经营所需最低备付金额。

针对**问题五**，本文在收益与风险，经营与政策，差异与公平这三个方面对以上模型做出了优化，使贷款分配方案既能满足全行的利息收益最大化，又能实现双赢或多赢的目标。

## 二、模型假设与符号定义

### 2.1 模型假设

- (1) 在所研究的十几年的时间内，商业银行 A 在全国存款总额中所占比例不变。
- (2) 银行的存款增量受限于社会资金增长总量，银行的贷款增量受限于存款增量。
- (3) 在所研究的时间内，不考虑重大国际金融危机等引起经济发生巨大变化的低概率事件，商业银行 A 正常运行，不发生企业倒闭重组等极端情况。

### 2.2 符号定义

表 1 基本符号及其说明

符号	说明	单位
$B$	宏观经济指标拟合后各项系数	
$r_{ij}$	相对偏差矩阵中第 $i$ 行第 $j$ 列的元素	
$w_j$	各项经济指标权重向量	
$x_F$	年平均存款能力	
$y_S$	商业银行历史存款总量	亿元
$\max$	全行增量存贷款利息净收入	亿元
$r$	各分行存贷款利率的系数矩阵	
$L$	债券利息	亿元
$m_1$	商业银行的存款增量	亿元
$m_2$	商业银行的贷款增量	亿元
$a, b, e, f$	浮动上限	亿元
$c, d, g, h$	浮动下限	亿元
$M_f$	商业银行日资金流量	亿元
$C$	备付资金成本	亿元
$P$	最低备付金额	亿元
$\lambda$	备付资金成本损失系数	
$T$	备付金保留时间	日

### 三、模型建立与求解

#### 3.1 问题一模型的建立与求解

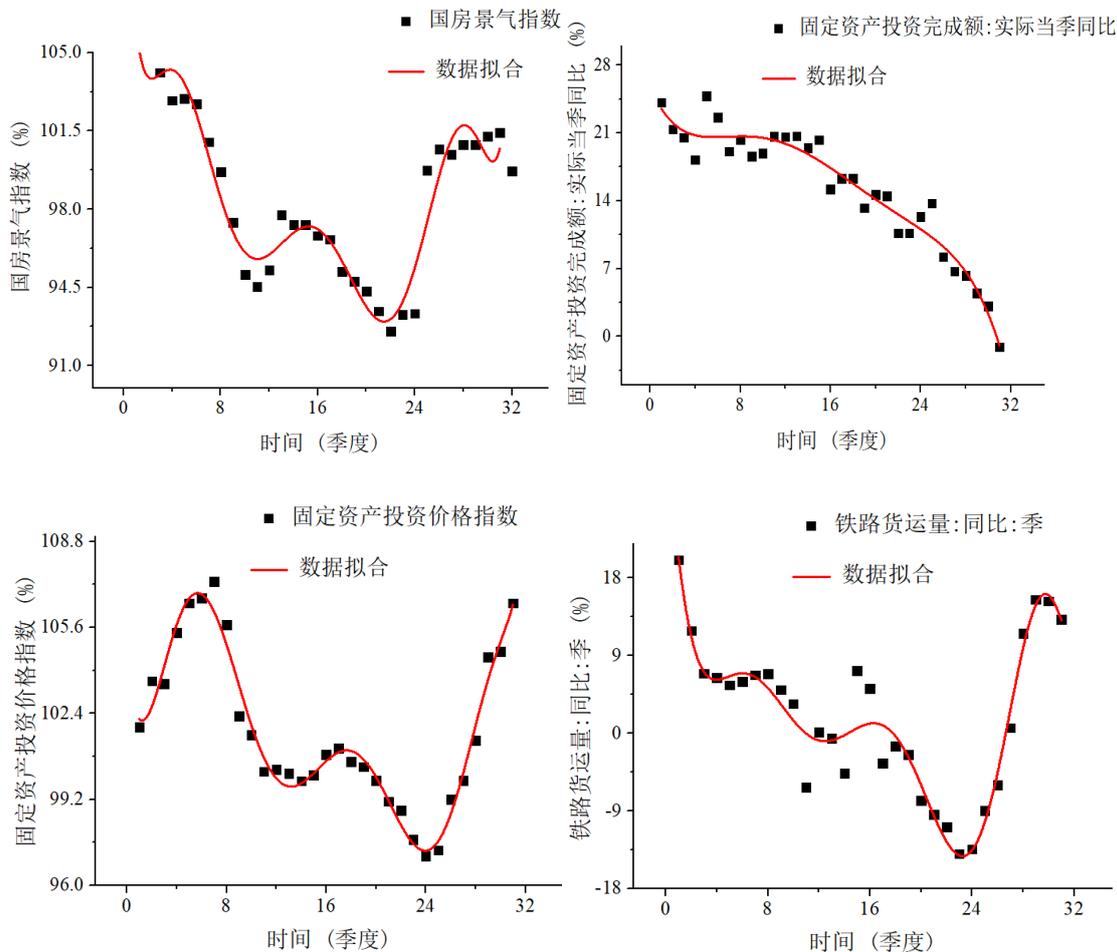
##### 3.1.1 模型假设

- (1) 该银行除客户存款外无其他资金来源。
- (2) 不考虑备付水平。

##### 3.1.2 模型建立

###### (1) 拟合得到缺失的宏观经济指标

银行的存贷款增量受到宏观经济指标的影响。由于附件 1 中 2017 与 2018 年的部分宏观经济指标缺失，本文利用已知的 2010 至 2017 年的数据为纵坐标，以时间（单位：年）为横坐标，使用专业的绘图软件 *Origin* 进行多项式拟合求得附件 1 中缺失的 2017 与 2018 年的部分宏观经济指标，如图 1 所示。



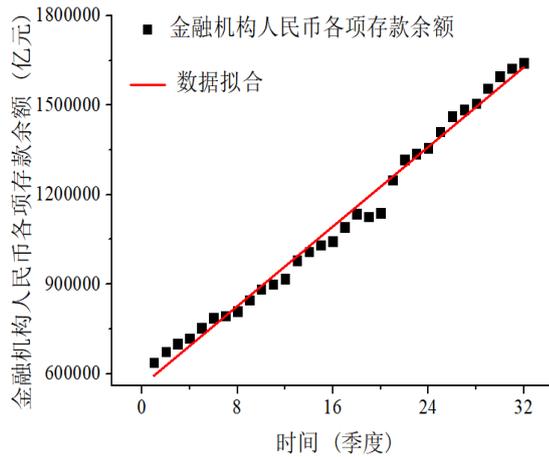


图 1 2010 至 2017 年部分宏观经济指标拟合

由图 1 并结合表 2 可知，通过数据拟合的方式得到的缺失数据可信。

表 2 多项式拟合部分宏观经济指标的各项系数与相关指数

宏观经济指标	$B_0$	$B_1$	$B_2$	$B_3$	$B_4$	$B_5$	$B_6$	$B_7$	$B_8$	$B_9$	$R^2$
国房景气指数	116	-15.14	7.292	-1.65	0.1948	-0.01	4.64	-8.86	6.90	0.00	0.913
		195	72	759	9	271	E-04	E-06	E-08	E+00	83
固定资产投资完成额	26	-2.681	0.505	-0.04	0.0014	-1.82	0	0	0	0	0.926
		88	08	147	4	E-05					44
固定资产投资价格指数	104	-3.633	1.968	-0.26	-0.003	0.003	-2.86	1.13	-2.19	1.67	0.956
		58	26	422	15	25	E-04	E-05	E-07	E-09	88
铁路货运量	36	-18.52	2.678	0.507	-0.197	0.024	-0.00	5.44	-9.82	7.20	0.899
		172	51	42	63	66	157	E-05	E-07	E-09	71
各项存款余额	559	33334.	0	0	0	0	0	0	0	0	0.988
	983	31771									05

根据表 2，得到多项式拟合的各项系数，将相应的时间值代入即可求得 2017 及 2018 年缺失的部分宏观经济指标。

### (2) 熵值法确定宏观经济指标的权重

熵值法是一种确定指标权重的客观赋权法。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，指标的离散程度越小，该指标对综合评价的影响越大，适用于量纲不统一的指标的权重计算<sup>[1]</sup>。对于本问题，由于各宏观经济指标的量纲不统一，且权重的确定需要尽量避免主观因素的影响，因此选用熵值法作为宏观经济指标权重的确定方法。

#### Step 1. 构造相对偏差矩阵，数据标准化

对于指标  $j$ ，相对偏差矩阵的元素由下式确定

$$r_{ij} = \frac{|a_{ij} - u_j|}{\max(a_{ij}) - \min(a_{ij})}, i = 1, 2, 3 \dots 13, j = 1, 2, 3 \dots 20 \quad (1)$$

其中,  $a_{ij}$  为 2015 年 3 月至 2018 年 3 月的每个指标的绝对值,  $u_j$  为每个指标的最优值。

$u_j$  的确定依据是该指标对存款增量是正向影响还是负向影响。若该指标的值越大, 存款增量越大, 则  $u_j$  取第  $j$  列的最大值, 反之取该列的最小值。

由此得到相对偏差矩阵

$$R = (r_{ij})_{i \times j} \quad (2)$$

相对偏差矩阵  $R$  的元素消除了量纲, 指标之间具有可比性。

**Step 2.** 计算权重系数

权重系数  $v_j$  由下式得到

$$\begin{cases} v_j = \frac{s_j}{|r_j|}, j = 1, 2, \dots, 20 \\ \bar{r}_j = \frac{\sum_{i=1}^{13} r_{ij}}{13} \\ s_j^2 = \frac{\sum_{i=1}^{13} (r_{ij} - \bar{r}_j)^2}{13-1} \end{cases} \quad (3)$$

其中  $s_j$  为 A 类不确定度。

**Step 3.** 归一化处理, 得到权重向量

将权重系数归一化处理

$$w_j = \frac{v_j}{\sum_{j=1}^{20} v_j} \quad (4)$$

权重向量是  $w_j$ 。

**Step 4.** 得到权重

将相对偏差矩阵  $(r_{ij})_{i \times j}$  与权重向量  $w_j$  相乘, 即可得到权重向量  $F_i$ 。

(3) 确定 2018 年存款增量

得到 2015 年 3 月至 2018 年 3 月的存款增量的权重向量后, 对每年四个季度的权重求均值作为该年的权重。根据每年的存款总量权重与历史存款总量的一一对应关系, 将 2018 年的存款总量权重代入拟合图像的表达式, 可以由 2018 年的存款总量权重预测其存款总量, 再由该存款总量与 2017 年的历史存款总量作差即可求得 2018 年的存款增量。

(4) 确定 2018 年贷款增量

由假设条件可知, 贷款增量受制于存款增量。建立贷款总量与存款总量之间的函数关系式可以进一步研究贷款增量与存款增量之间的关系。将 2015 至 2017 年的历史存贷款总量拟合, 可以得到存贷款总量之间的关系。

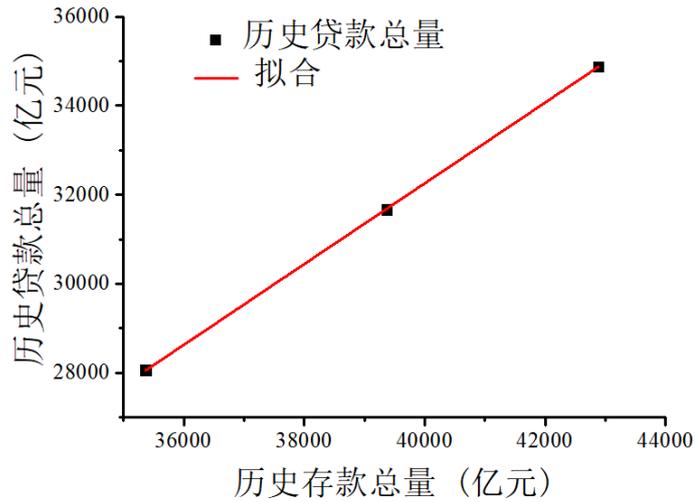


图 2 2015 至 2017 年历史存贷款总量关系

由图 2 可求得存贷款总量之间满足如下关系

$$y_{\text{贷}} = 0.90761 \times x_{\text{存}} - 4048.2837 \quad (5)$$

该直线拟合的相关指数  $R^2 = 0.99998$ ，可以通过式(5)预测 2018 年的贷款总量。将该贷款总量预测值与 2017 年历史贷款总量作差即可求得 2018 年的贷款增量。

### 3.1.3 模型求解

通过对 2010 至 2017 年宏观经济指标与时间的拟合，将相应的时间代入得到缺失的部分指标后，按照熵值法的四个步骤，计算得到权重向量如表 3 所示。

表 3 熵值法确定权重向量

指标序号	$\bar{r}_j$	标准差	方差	$s_j^2$	$s_j$	$v_j$	$w_j$
1	0.22	0.22	0.05	0.00	0.06	0.28	0.09
2	0.27	0.16	0.03	0.00	0.05	0.17	0.05
3	0.48	0.32	0.10	0.01	0.09	0.19	0.06
4	0.48	0.37	0.14	0.01	0.11	0.22	0.07
5	0.60	0.32	0.10	0.01	0.09	0.15	0.05
6	0.71	0.30	0.09	0.01	0.09	0.12	0.04
7	0.49	0.02	0.00	0.00	0.01	0.01	0.00
8	0.58	0.29	0.08	0.01	0.08	0.14	0.04
9	0.70	0.29	0.08	0.01	0.08	0.12	0.04
10	0.50	0.30	0.09	0.01	0.09	0.17	0.05
11	0.64	0.39	0.15	0.01	0.11	0.17	0.05
12	0.58	0.34	0.12	0.01	0.10	0.17	0.05
13	0.45	0.32	0.10	0.01	0.09	0.20	0.06
14	0.63	0.31	0.09	0.01	0.09	0.14	0.04
15	0.47	0.41	0.17	0.01	0.12	0.25	0.08
16	0.41	0.31	0.09	0.01	0.09	0.22	0.07
17	0.82	0.30	0.09	0.01	0.09	0.11	0.03
18	0.44	0.14	0.02	0.00	0.04	0.09	0.03
19	0.54	0.37	0.14	0.01	0.11	0.20	0.06
20	0.34	0.14	0.02	0.00	0.04	0.12	0.04

将权重向量  $w_j$  与相对偏差矩阵  $(r_{ij})_{i \times j}$  相乘得到每季的历史存款总量权重，将每年四个季度的历史存款总量权重加权平均作为该年的历史存款总量权重。

表 4 年平均存款能力

时间	存款能力 (季)	年平均存款能力	存款金额 (亿元)
2015-03	0.3663		
2015-06	0.2495		
2015-09	0.3590	0.3370	35370
2015-12	0.3731		
2016-03	0.3522		
2016-06	0.3999		
2016-09	0.3596	0.3774	39370
2016-12	0.3978		
2017-03	0.4084		
2017-06	0.3943		
2017-09	0.4172	0.3933	42880
2017-12	0.3534		
2018 (预测)	0.4258	0.4802	46370

每年的历史存款总量权重反映了银行年平均存款能力。这种能力直接影响该年

银行存款总量。为了得到两者的函数关系式，将年平均存款能力与历史存款总量拟合如图 3 所示。

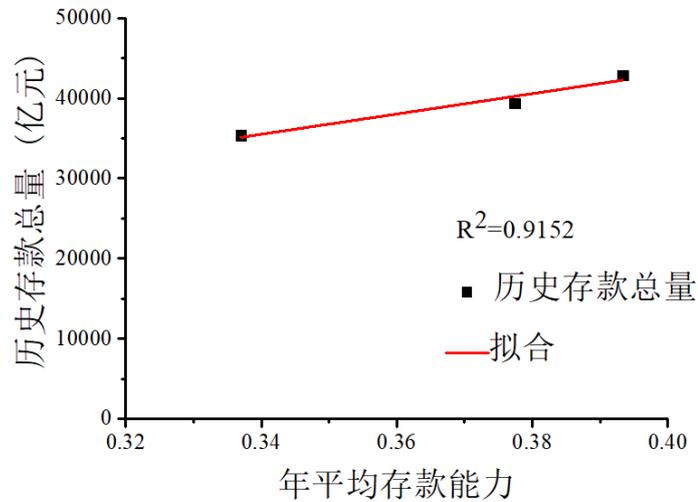


图 3 2015 至 2017 年银行的年平均存款能力与历史存款总量

图 3 拟合得到年平均存款能力与历史存款总量的函数关系式

$$y_S = 0.90761 \times x_F - 4048.2837 \quad (6)$$

将 2018 年的年平均存款能力  $x_F = 0.4258$  代入式(6)可得 2018 年的存款总量

$y_S = 46370.27$ ，与 2017 年历史存款总量作差得到 2018 年存款增量为 3490 亿元，

将 2018 年存款总量代入式(5)得到 2018 年贷款总量为 38038 亿元，与 2017 年历史贷款总量作差得到 2018 年贷款增量为 3161 亿元。

### 3.2 问题二模型的建立与求解

由问题一的结果得到 2018 年商业银行 A 的存贷款增量，并结合各个分行对存贷款的需求，以其为约束条件，建立线性规划模型确定贷款规模的分配方案使得全行增量存贷款利息净收入最大。

#### 3.2.1 模型假设

(1) 各分行只能从商业银行 A 获得贷款，彼此之间没有存贷款往来。

(2) 2015 至 2018 年金融市场稳定，银行的存贷款利率在均值上下波动，将各行存贷款利率的均值作为 2018 年的存贷款利率值。

#### 3.2.2 模型建立

记商业银行 A 的各分行为  $i$  ( $i=1,2,\dots,30$ )，2018 年各分行的存款增量为  $x_{i1}$ ，存款利率为  $r_{i1}$ ；贷款增量为  $x_{i2}$ ，贷款利率为  $r_{i2}$ 。全行增量存贷款利息净收入的目标函数为

$$\max = \sum_{i=1}^{30} (r_{i2} \times x_{i2} - r_{i1} \times x_{i1}), (i=1,2,\dots,30) \quad (7)$$

接下来讨论该目标函数的约束条件。

结合实际情况，各分行在 2018 年的存款增量之和与贷款增量之和受限于商业银行 A 在 2018 年的存贷款增量。又因为受到市场经济环境的变化，商业银行 A 的存贷款增量有一定的浮动范围，因此各分行的存款增量之和与贷款增量之和在一定的范围内波动。记存款增量和贷款增量的浮动上限分别为  $a, b$ ，浮动下限分别为  $c, d$ 。

存款增量约束

$$(1-c) \times m_1 \leq \sum_{i=1}^{30} x_{i1} \leq (1+a) \times m_1 \quad (8)$$

贷款增量约束

$$(1-d) \times m_2 \leq \sum_{i=1}^{30} x_{i2} \leq (1+b) \times m_2 \quad (9)$$

### 3.2.3 模型求解

为了找到目标函数式(7)的最优解，首先需要确定 2018 年各分行的存贷款利率。存贷款利率受市场变化如通货膨胀等影响会在均值上下波动。根据 2015 至 2017 年的银行 A 各个分行的存贷款利率，用加权平均的方法得到平均值，该平均值与存贷款利率均值十分接近，在一定误差范围内可以将该值作为存贷款利率的均值。根据假设条件 2，2018 年各分行的存贷款利率即为该平均值。由此得到 2018 年各分行存贷款利率的系数矩阵

$$U = \begin{bmatrix} 2.90 & 1.99 & 2.42 & 2.54 & 2.27 & 2.48 & 2.18 & 2.14 & 2.26 & 2.25 \\ 4.76 & 5.15 & 5.00 & 4.99 & 4.81 & 4.75 & 5.31 & 5.48 & 5.14 & 5.27 \\ 2.48 & 2.39 & 2.00 & 1.84 & 2.11 & 2.05 & 2.64 & 2.08 & 2.20 & 1.99 \\ 5.03 & 5.13 & 5.34 & 5.64 & 5.22 & 5.23 & 4.63 & 5.14 & 5.42 & 5.16 \\ 2.70 & 2.29 & 2.37 & 1.41 & 1.83 & 2.04 & 2.22 & 2.01 & 2.18 & 2.00 \\ 4.90 & 5.41 & 4.79 & 5.16 & 5.82 & 5.47 & 5.44 & 5.06 & 5.52 & 5.16 \end{bmatrix} \quad (10)$$

由假设条件可知，2018 年市场稳定，存贷款增量的波动范围可记为  $a=0.15$ ， $b=0.20$ ， $c=0.10$ ， $d=0.15$ 。利用 LINGO 求得该线性规划问题的解<sup>[2]</sup>，即为各分行的贷款增量分配方案，再加上 2017 年各分行的贷款金额就得到各分行的贷款规模分配方案，填入表 5，并得到此时全行增量存贷款利息净收入为 82.55 亿元。

## 3.3 问题三模型建立与求解

### 3.3.1 模型假设

在问题二建立的单目标线性规划问题的假设的基础上，再补充三条假设。

(1) 商业银行 A 于 2018 年 5 月 1 日发行的 500 亿规模的 15 年期商业银行普通债短时间内售完。

(2) 发行债券的 15 年内，利率为 5.1% 保持不变，不考虑通货膨胀等因素对利率的影响。

(3) 2018 年各行的存贷款利率保持不变，仍为式(10)。

(4) 商业银行 A 以本息合一的方式支付债券本息。即通过债券到期后的一次还本付息而支付利息，即在每一个年度都按同一的固定利率来支付利息<sup>[3]</sup>。

### 3.3.2 模型建立

商业银行A发行债券后，使得A银行的收入增加，这部分收入定义为存款当量，算入商业银行A的存款增量中。同时，由于贷款增量受限于存款增量，由问题一的模型中可以得到存款量与贷款增量的定量关系，可以得到银行A存款增量增加后相应的贷款增量。记此时商业银行A的存款增量和贷款增量分别为  $m_1'$ ， $m_2'$ 。

由于商业银行A于2018年5月1日开始发行500亿元债券，在5月1日之前的4个月的时间里，各个分行的存款增量和贷款增量仍遵循问题二的模型的结果，约束条件同式(8-9)。在5月1日之后，由于债券的发行使得商业银行A增加了存款当量，各分行的存款增量之和与贷款增量之和受到的约束条件发生变化。同时存贷款增量的波动范围也相应地发生变化，即

$$\begin{cases} (1-g) \times m_1' \leq \sum_{i=1}^{30} y_{i1} \leq (1+e) \times m_1' \\ (1-h) \times m_2' \leq \sum_{i=1}^{30} y_{i2} \leq (1+f) \times m_2' \end{cases} \quad (11)$$

其中存款增量和贷款增量的浮动上限分别为  $e, f$ ；浮动下限分别为  $g, h$ ； $y_{i1}, y_{i2}$  为2018年5月1日之后的存贷款增量。

接下来讨论本模型的目标函数。

为了使2018年商业银行A分配给各分行的贷款金额获得最大的存贷款利息净收入，需要以2018年5月1日为时间节点，分段讨论存贷款利息收入，并以各段所占的时间比重为权重，线性求和得到全年的存贷款利息收入。容易得到，5月1日之前所占的比重为1/3,5月1日之后所占的比重为2/3。另外，由于商业银行A发行债券后要在规定日期15年后连本带息缴纳发行债券的费用，因此要在存贷款利息收入的基础上减去发行债券应支付的利息。由此得到全行增量存贷款利息净收入的目标函数

$$\max' = \left( \frac{1}{3} \times \sum_{i=1}^{30} (r_{i2} \times x_{i2} - r_{i1} \times x_{i1}) + \frac{2}{3} \times \sum_{i=1}^{30} (r_{i2} \times y_{i2} - r_{i1} \times y_{i1}) - L \right) \quad (12)$$

其中  $L$  为债券利息均摊到2018年5月1日之后的8个月对应的金额。

### 3.3.3 模型求解

根据本息合一支付利息方式的要求，2018年5月1日之后的8个月对应的债券利息为

$$L = 500 \times 0.051 \times \frac{2}{3} \quad (13)$$

发行债券后，商业银行A的存款增量为

$$m_1' = 3490 + 500 = 3990 \quad (14)$$

将  $m_1'$  的值代入式(5)得到商业银行A的贷款增量  $m_2' = 3621$ 。

将  $m_1', m_2'$  的值与  $m_1, m_2$  的值比较后，可得商业银行A的贷款增量和存款增量均增大。存款增量增加后浮动范围不变，贷款增量的浮动上限增大，下限减小，有

$$\begin{cases} e = 0.15 \\ f = 0.30 \\ g = 0.10 \\ h = 0.10 \end{cases} \quad (15)$$

利用LINGO求得该线性规划问题的解，和问题二的模型的解作相同的处理，将2018年各分行的贷款规模分配的结果填入表5，并得到此时全行增量存贷款利息净收入为73.81亿元。

### 3.3.4 与问题二结果比较

比较问题二和问题三求得的 2018 年发行债券前后全行增量存贷款利息净收入的结果可知，发行债券后全行增量存贷款利息净收入有所减少，由 82.55 亿元减少为 73.81 亿元。然而，仅仅通过 2018 年的增量存贷款净收入不能说明债券的发行使得全行增量存贷款利息净收入减少，原因是发行债券后的利息是按照时间均摊到每一年，在目标函数上减去此部分利息得到增量存贷款净收入。然而，在发行债券的 15 年期限中，由于发行债券而获得了存款当量，全行的存款增量比没有发行债券时提高了，贷款增量也相应地提高，贷款能力提高，增量存贷款总收入也提高。在 15 年的时间内，由于资金的累积效应，总收入有逐年增加的趋势，说明银行 A 在发行债券的中后期比前期更有偿还债券利息的能力。因此，在问题三的模型中将债券利息按照时间均分到每一年得到的目标函数值是偏小的。采用问题三的贷款分配方式，商业银行 A 的实际所得增量存贷款利息净收入大于所得的目标函数值。当债券到期后，发行债券获得的增量存贷款利息净收入会超过不发行债券获得的增量存贷款利息净收入。

## 3.4 问题四模型的建立与求解

### 3.4.1 模型假设

(1) 对于商业银行正常经营过程中的资金流入量只考虑最主要的来源，即存款金额；商业银行正常经营过程中资金流出量则近似地认为全部来自取款金额，忽略其他情况。

(2) 每个客户存取款的行为是随机的，并且商业银行每日资金流量受到大量的随机因素的影响，各个因素所起的作用都很微小。

(3) 管理成本指商业银行在管理备付资金过程中所发生的各项支出，因其所包含的具体成本项目往往难以准确界定，因此不考虑银行管理成本对备付资金持有量的影响。

### 3.4.2 模型建立

商业银行的资金流动和备付金管理之间存在着非常密切的关系<sup>[4]</sup>。资金流入可以看成是对备付金的供给，而资金流出可以看成是对备付金的需求。备付资金持有量的大小，则是由这些供给和需求之间差额的大小决定的。某一时期内供给和需求之间的这部分差额称作这一时期内的资金流量，即资金流出量减去流入量。

$$M_o - M_i = M_f \quad (16)$$

其中  $M_o$  为银行每日取款金额， $M_i$  为银行每日存款金额， $M_f$  为商业银行每日基金

流量。

商业银行备付金确定的目的，就是在保证日常资金流转的前提下，尽量减少备付金的持有量，以实现流动性和盈利性的统一。因此，本文从商业银行的资金流量着手，研究商业银行 A 各分行日常经营所需最低备付金额。

### (1) 资金流量分析

由假设可知，该商业银行每个客户存取款的行为是随机的，并且商业银行每日资金流量受到大量的随机因素的影响，各个因素所起的作用都很微小。由中心极限定理可知，在自然界与生产中，一些现象受到许多相互独立的随机因素的影响，如果每个因素所产生的影响都很微小时，总的影响可以看作随机变量序列部分和分布渐近于正态分布。因此本文认为商业银行每日资金流量是连续随机变量概率分布的一种，即商业银行每日资金流量服从期望值为  $\mu$ 、方差为  $\sigma$  的正态分布<sup>[5]</sup>，其概率密度公式

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (17)$$

记作 
$$X \sim N(\mu, \sigma^2) \quad (18)$$

下面我们对商业银行 A 各分行 2017 年每日存取款交易数据进行处理可以得到各分行 2017 年每日资金流量的分布，随机选取两分行（以北京和河南为例）研究它们的每日资金流量概率分布，若与正态分布的拟合相关性较好，则证明了上述对于每日资金流量的分析。

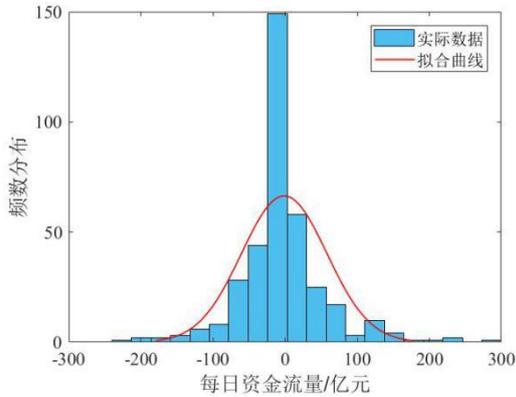


图 4.北京分行每日资金流量概率分布

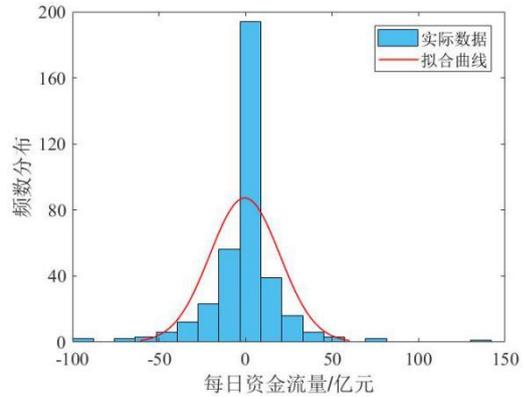


图 5.河南分行每日资金流量概率分布

由上面的拟合图像，本文得出北京和河南两分行的每日资金流量概率分布可近似认为符合正态分布，由此可以求得两分行每日资金流量的概率密度函数

$$f(x)_{\text{北京}} = \frac{1}{\sqrt{2\pi} \times 59.1213} \exp\left(-\frac{(x+1.1)^2}{2 \times (59.1213)^2}\right) \quad (19)$$

$$f(x)_{\text{河南}} = \frac{1}{\sqrt{2\pi} \times 20.1556} \exp\left(-\frac{(x+0.4)^2}{2 \times (20.1556)^2}\right) \quad (20)$$

同理也可以得到其余 28 个分行的每日资金流量也符合正态分布规律，将 30 个

地区分行拟合所得正态分布的均值与方差汇总在下表当中。

## (2) 备付成本分析

商业银行为进行正常的经营活动，必须要保证资金流入和流出的大体平衡，同时要保持相当数量的资金准备，以应付不时之需。备付金管理的主要目的，就是以最低的成本实现流动性、安全性与盈利性的统一。备付金与资金流量的关系主要体现在备付金成本上，成本也是影响备付金额的最重要因素。我们将备付金成本分为三部分，分别为过量成本、不足成本与管理成本，下面分别对三部分成本进行逐一分析。

过量成本是指商业银行因持有较多的备付金而丧失的潜在收益，假设  $P$  为备付金存有量， $\lambda_1$  为过量损失系数， $T$  为备付金保留时间，则可以得到过量成本计算公式：

$$C_{\text{过量成本}} = \lambda_1 \times P \times T \quad (21)$$

不足成本是指商业银行因持有备付金不足而造成的银行盈利水平的降低。当每日资金流量与备付金余额之和较小时，商业银行需通过其他方式取得资金以满足需要，从而付出的代价即为不足成本，假设  $\lambda_2$  为不足损失系数，则有不足成本计算公式

$$C_{\text{不足成本}} = \lambda_2 \times (M_f - P) \times T \quad (22)$$

结合备付金额的管理成本，本文得到到每日备付金额的总成本公式如下：

$$C_{\text{总成本}} = \lambda_1 \times P + \lambda_2 \times (M_f - P) + C_{\text{管理成本}} \quad (23)$$

从前述成本的构成可知，随着备付水平的变化，备付金的过量成本和不足成本互为消长。即商业银行持有的备付金越多，过量成本越高，不足成本越低；反之，如果降低备付水平，可降低过量成本，但不足成本将会随之上升。备付金额管理的目的，就是要寻找使总成本最低时的备付金持有量，即为符合要求的最低备付金额。

### 3.4.3 模型求解

我们分别对 30 个分行进行模型的求解，由于方法相同，这里仅以北京分行为例介绍。由上文可知各分行的每日资金流量分布符合正态分布，通过对北京分行一年间每日的资金流量数据进行处理，可以得到

$$\mu = -1.1, \sigma = 59.1213 \quad (24)$$

将此一般正态分布标准化，利用公式

$$U = \frac{X - \mu}{\sigma} \sim N(0,1) \quad (25)$$

同时也可以得到标准正态分布概率密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (26)$$

及其概率分布公式

$$\Phi(x) = P(U \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (27)$$

为使备付金额的总成本达到最低，在  $\lambda_1 \ll \lambda_2$  的前提下，我们以减小不足成本为主要途径，也就是保证资金流量与备付金额与之差趋近 0，即保证备付金额与资金流量在数值上相等，此时不足成本达到最小值。同时在置信水平达到 99% 的情况下，为满足过量成本时达到最小值，备付金额无需继续增大，故存在一满足银行日常经营的最低备付金额值。通过查阅标准正态分布表可知

$$\Phi(2.33) = 0.9901 > 0.99 \quad (28)$$

由此可以求出当  $P_{2017} = 136.63$ ，即在置信水平 99% 的情况下，2017 年商业银行 A 北京分行日常经营所需最低备付金额为 136.63 亿元，此时备付金额的成本最低。接下来我们利用问题一当中提出的银行年平均存款能力这一概念，将 2017 年备付金额乘以两年平均存款能力的比值，预测得到 2018 年北京分行的最低备付金额  $P_{2018} = 136.63$ 。

同理可以算出 2018 年银行 A 其他 29 处分行的最低备付金额，并填写于表 5 中。

**表 5 2018 年各分行贷款增量分配方案与备付资金**

单位：亿元		问题二	问题三	问题四
序号	分行	贷款规模分配方案	贷款规模分配方案	2018 年备付资金
1	北京	3194	3199	138.95
2	江苏	4668	4687	87.76
3	广东	3751	3765	80.42
4	浙江	3814	3827	57.02
5	上海	2798	2805	120.31
6	山东	1799	1804	54.79
7	河南	1478	1520	47.34
8	湖北	1652	1664	58.80
9	安徽	1182	1186	28.83
10	四川	1007	1041	26.64
11	辽宁	1170	1174	33.61
12	河北	847	849	27.05
13	重庆	729	749	27.91
14	湖南	1205	1219	35.07
15	山西	801	825	30.52
16	陕西	685	701	35.11
17	天津	742	742	41.58
18	广西	745	746	16.78
19	吉林	566	569	19.16
20	云南	713	739	26.46
21	福建	987	990	20.26
22	江西	686	695	30.43
23	黑龙江	416	416	15.43
24	新疆	504	516	33.83
25	贵州	444	447	12.03
26	甘肃	341	343	11.90
27	海南	395	399	13.39
28	内蒙古	427	428	11.50
29	宁夏	164	165	4.51
30	青海	129	132	5.10
全行合计		38035	38343	1152.57

至此，通过问题一至四中所建立的模型，本文从问题一建立的模型中得到了 2018 年商业银行 A 存贷款增量的预测值；从问题二、三建立的模型中得到了是否发行 15 年期的 500 亿元债券对应的贷款分配方案；从问题四建立的模型中得到了 2018 年各分行的备付资金，四个基本模型成功地解决了相应的问题。

## 四、模型优化与评价

### 4.1 模型优化

#### 4.1.1 收益与风险

通过以上四个问题的分析可知，商业银行 A 的增量存贷款利息净收入取决于贷款利息收入与存款利息支出的差值（不考虑发行债券等对增量存贷款利息净收入的影响）。存款增量一定的条件下，贷款增量越大，利息净收入越大。根据式(5)可以得到存贷款增量满足

$$\Delta y_{\text{贷}} = k \times \Delta x_{\text{存}} \quad (29)$$

其中  $k$  为式(5)直线的斜率 0.90761。在存款增量一定的情况下，为了得到更大的贷款增量需要承担一定的风险，引入风险系数  $\alpha$  对式(29)修正，有

$$\Delta y_{\text{贷}} = \frac{1}{1-\alpha} \times k \times \Delta x_{\text{存}} \quad (30)$$

其中风险系数  $\alpha$  的范围为  $[0,1)$ 。贷款增量越大， $\alpha$  越大，银行所承担的风险越大。通过风险系数  $\alpha$  可以衡量银行收益与风险之间的关系

$$z_{\text{收益}} \propto \Delta y_{\text{贷}} \propto \frac{1}{1-\alpha} \quad (31)$$

其中  $z_{\text{收益}}$  为存款增量一定时全行增量存贷款利息净收入。

#### 4.1.2 经营与政策

企业经营与国际政策的变化也会对全行增量存贷款利息净收入有影响，需要调整贷款分配方案以适应这种变化。这种变化的影响主要体现在存贷款利率的变化上。考虑到同行竞争，商业银行可能会通过适当提高存款利率以吸收客户存款，增大收入；考虑到国家重点事业、民生工程的投入，国家会通过行政等手段降低贷款利率以促进重点工程的建设。由以上两种情况的分析可得，各分行的存贷款利率矩阵  $U$  会根据企业经营状况和国际政策的变化而改变。引入利率修正矩阵  $U'$ ，其中每个元素都是存贷款利率矩阵  $U$  的元素  $u_{ij}$  乘以相应的利率修正系数  $\beta_{ij}$  得到。修正后得到的存贷款利率矩阵

$$V = U + U' \quad (32)$$

修正后的存贷款利率矩阵  $V$  中的元素为  $v_{ij} = u_{ij} + u'_{ij} = (1 + \beta_{ij})u_{ij}$ 。

#### 4.1.3 差异与公平

为了协调不同分行之间的发展，促进发展公平，在解决贷款分配问题时应充分考虑到各个分行的发展情况。对任一分行的贷款增量  $x_{i2}$ ，修改线性规划模型中的贷款增量的约束条件，有

$$\begin{cases} \zeta m_2 \leq x_{i2} \leq \zeta(1+\gamma)m_2 \\ \zeta = \frac{n_i}{n_0} \end{cases} \quad (33)$$

其中  $n_i$  表示上一年  $i$  分行贷款总量,  $n_0$  表示上一年商业银行 A 的贷款总量,  $\zeta$  为历史经验系数, 是衡量分行历史贷款水平的指标。 $\gamma$  为公平系数, 与  $\zeta$  有相互抑制的关系。当某分行历史发展水平较低时,  $\zeta$  值偏小, 此时  $\gamma$  增大, 在保证该分行得到基本水平的贷款增量保障后, 使得分配给该分行的贷款增量的波动上限增大, 体现了缩小区域化发展差异, 促进贷款增量分配公平的理念。

## 4.2 模型评价

### 4.2.1 模型优点

(1) 问题一模型中, 选用熵值法为各项宏观经济指标赋权, 客观地完成多变量综合评价。并创新地应用评估结果与银行年存款能力建立联系, 从而合理、准确地预测存贷款增量。

(2) 问题二模型中, 在多元线性规划的基础上充分地考虑到浮动上下限的变化, 结果经验证后在合理的变化范围当中, 且尽量做到缩小区域化差异, 保证分配公平。

(3) 问题三模型中, 将发行债券对于收益与风险等影响综合地表达, 通过合理调整波动范围将目标函数的最优解更加贴近现实地求解出来。

(4) 问题四模型中, 巧妙地将每日现金流量拟合为正态分布, 通过随机事件代替复杂的经济背景, 预测 2018 年最低备付金额结果适应性较强。

### 4.2.2 模型缺点

(1) 问题一模型中, 忽略了各宏观指标间的横向对比, 近似地认为各变量为线性组合关系。

(2) 问题二模型中, 假设复杂多变的经济环境对于存贷款等无影响在现实生活中很难实现。

(3) 问题三模型中, 将发行债券后的利息按照时间均摊到每一年的做法严谨性稍有不足。

(4) 问题四模型中, 未结合资产负债结构、利率风险等综合分析, 导致模型较精确性欠佳。

## 4.3 改进建议

通过上述模型的建立可以看出, 影响一年各分行存款和贷款量的因素是很复杂的, 仅仅通过在年初初预分配、全年限额管理的模式的缺点是很明显: 基层银行缺乏必要的经营管理自主权, 各分行之间银行没有直接联系, 资金的运行缺乏灵活性, 不能考虑到实际存取款需求的变化, 资金的实用效率低、获益也较少。对此, 本文结合建模结果给出以下五点建议。

(1) 加强制度建设。要有效地加强信贷资金的计划管理, 需要减少行政手段, 加强经济调节手段, 落实各级银行的经济责任制。从制度上解决信贷资金分配上的供给制, 建立起符合银行特点和信贷资金运动规律的、采取经济办法和经济利益进行调节的计划管理体制。

(2) 加强各分行之间的联系, 强调和实行相互融通; 各在相互融通的条件下, 各行之间就能够通过支付利息的有偿原则, 相互拆借, 调剂余缺, 全面发挥资金融通的横向调剂作用, 最大限度地运用资金。

(3) 支持各地区协调的发展; 采用以存定贷、存贷结合、等运营模式, 努力获取全行最大收益, 同时注意平衡各区域发展差异, 对国家重大项目、重点政策、民生工程等倾斜扶持, 支持实体经济有效发展。

(4) 提升安全运营意识和能力, 规避风险; 银行的制度改革和投资都不能忘记安全运营意识, 同时商业银行需要建立起有效的风险控制系统。“商业银行尤其应高度重视运营集约化改革过程中的风险控制, 任何运营集中项目的实施, 都不能以牺牲风险控制来换取效率的提升; 运营模式的任何变革都必须以操作风险的切实可控为前提。”<sup>[6]</sup>

(5) 不断与时俱进, 改革创新; 在银行运营体系中引入制造业的生产理念, 充分发挥社会化分工的优势, 最大限度地获取规模效益。建立健全运营人才队伍体系, 特别是培育富有经验的管理队伍和专业的研发队伍体系不断提升队伍的创新意识和引领变革的能力。

## 参考文献

- [1] 熵值法, <https://baike.sogou.com/v8324674.htm>, 2018年4月29日.
- [2] 姜启源, 谢金星, 叶俊. 数学模型(第四版)[M]. 北京:高等教育出版社, 2011, 95-100.
- [3] 债券利率, <https://baike.sogou.com/v7817848.htm>, 2018年4月29日.
- [4] 汤谷良, 杨瑾. 从现金流量分析我国商业银行最佳备付金持有量——一个基于盈利性目标的数学模型. 国际金融研究, 2005(03): 第12-17页.
- [5] 盛骤, 谢式千. 概率论与数理统计(第四版)[M]. 北京:高等教育出版社, 2008, 56-60.
- [6] 杨坚旭. 商业银行运营能力建设探讨[J]. 金融经济, 2012(20):40-41.

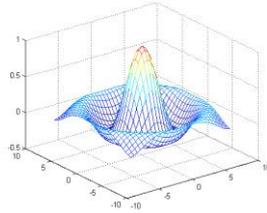
## 附件

问题2的LINGO程序:

```
a=0.15;
b=0.20;
c=0.10;
d=0.15;
sets:
jine/1 2/:Y;
fenhang/1..30/;
Link(fenhang,jine):r, x;
endsets

data:
Y=3490 3161;
```

# 五一数学建模竞赛



## 题目：江苏省本科教育评价问题

### 摘要：

本文利用主成分分析、熵权法、灰色关联分析等方法，通过建立指标量化模型、综合评价模型，从可减少指标的确定、关键指标的确定等方面对江苏省本科教学质量评价问题进行了研究。

对于问题一，首先将 9 项指标分为可直接量化的指标与不可直接量化指标两类，对于 5 项可直接量化的指标利用对地级市中各院校统计数据求和、加权平均等方法进行量化，对于 4 项不可直接量化指标依据相关统计报告和选取原则共遴选出了 22 项子指标见表 5.2；然后使用主成分分析法，求出最终的评价值作为该指标的量化值，所有指标的量化方法及计算依据汇总见表 5.4，13 个地级市的最终量化统计数据见表 5.5；最后对图表进行相关的数据分析，发现生师比这一指标的地区差异最小，而师资队伍与结构、专业建设与教学改革等指标的地区差异明显。

对于问题二，首先利用问题一中的量化数据，首先建立基于熵权法的本科教育质量综合评价模型，求解得出评价值及排名；然后建立基于灰色关联分析的本科教育质量综合评价模型求解相关评价结果；将两种评价结果相互验证，证明所建立的模型的准确性。

对于问题三，首先明确可减少的指标需满足的条件，即剔除指标后剩余的指标体系所得地级市排名与问题二中的排名一致，同时保留绝大部分的原数据信息；然后利用关键因子识别的方法进行求解，得出两个可减少的指标为“学生就业”与“生师比”的结论；最后使用剔除指标前后评价值作雷达图验证了结果的合理性。

对于问题四，本文以一组评价值的方差来衡量各市本科教育质量差异；然后利用模拟退火算法，寻找出 9 维空间的一个使得评价值方差最小的“方向”向量，最大的分量值对应的指标——“师资队伍与结构”即为关键指标；最后单方向变动进行验证

对于问题五，结合上述模型的求解结果，从完善评价体系、缩小地区差异、改进关键指标、提升整理水平等方面给出了相关的政策建议。

本文最大的特色在于综合利用了主成分分析、熵权法、灰色关联分析多种方法对江苏省本科教学质量评价问题进行了研究，利用退火算法确定了可减少指标与关键指标，为未来教学质量的评估提供了参考。

**关键词：**主成分分析、熵权法、灰色关联分析、模拟退火算法、综合评价模型

---

## 一. 问题重述

随着中国的改革开放，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，江苏省的本科教育发展在全国名列前茅，而江苏省 13 个地级市的本科教育质量发展并不平衡。

1. 影响本科教育质量的指标有很多。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(完成下表)，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，对江苏省 13 个地级市（常州、淮安、连云港、南京、南通、宿迁、苏州、泰州、无锡、徐州、盐城、扬州、镇江）的本科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 9 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江苏省 13 个地级市本科教育发展的差异，并利用查找的数据资料验证所建立模型的有效性。
5. 请结合前面的讨论给出有效提升江苏省本科教育质量的政策建议。

## 二. 模型假设

1. 假设教育部高等教育教学评估中心及各学校官方网站等权威机构得到的数据真实可靠，不存在明显误差。
2. 假设选取的指标能较为全面的反映本科教育质量水平。
3. 假设各院校的本科教学质量指标不会在短期内发生重大变化。

## 三. 符号说明

序号	符号	符号含义
1	$i$	城市序号
2	$j$	指标序号
3	$x_{ij}$	第 $i$ 个城市的第 $j$ 项指标的值
4	$Z_i$	第 $i$ 个城市的本科教育质量评价值
5	$D_i$	第 $i$ 个城市的本科教育质量排名

---

## 四. 问题分析

### 4.1 问题一的分析

问题一要求查找相关资料对 9 个指标进行量化处理并进行数据分析。即要对各地级市中的学校数据进行搜集处理，对不同的指标选取不同的量化处理方法，使其可以客观全面反映整个市该项指标的水平。对于 4 个不能直接量化的指标，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和科研投入与产出，应依据导向性、可操作性、全面性、普适性和相对独立性的原则进行指标的遴选，分别建立主成分分析综合评价模型得出指标值作为量化处理结果。并基于量化处理结果进行相关的数据分析。

### 4.2 问题二的分析

问题二要求建立对本科教育质量的综合评价模型，并对 13 个城市进行评价。对此，我们首先利用问题一得到的 9 个指标的量化数据，分别建立基于熵权法的综合评价模型和基于灰色关联分析的综合评价模型和基于灰色关联分析的综合评价模型，分别计算各城市的评价值及排名，并将两种方法下得到的排名进行对比分析。

### 4.3 问题三的分析

问题三要求分析在所给的 9 个指标中，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。评价结果一致不仅意味着排名的一致，还要使剩余的指标，也就是关键因子仍具有代表性和可靠性，即需要保留足够的信息量，因此我们需要找出减少后仍能使剩余的指标体系满足上述条件的指标。

### 4.4 问题四的分析

问题四要求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，并验证所建立模型的有效性。各地级市本科教育发展的差异即缩小最终评价值之间的方差。指标值的普遍改善即使 13 个样本点在 9 维空间中逼近由各指标最大值构成的理想点。在改善方法相同的情况下，不同的方向可以得到不同的方差。由方向向量的分量大小，确定哪个因素的普遍改善会减小市级本科教育的差异，值越大则该指标的改善带来的影响越大，沿此指标改善越能缩小地区之间的差距。

## 4.5 问题五的分析

问题五要求结合前面的讨论给出有效提升江苏省本科教育质量的政策建议。既要结合前面模型的求解结果从多角度进行分析，在完善评价体系、缩小地区差异、改进关键指标、提升整理水平等方面给出客观全面的建议。

# 五. 问题一模型的建立与求解

本文所有的数据均来自各学校教务网，信息公开网、教育部高等教育教学评估中心、江苏省教育厅、教育部科学技术司等权威机构发布的近两年的《本科教育质量报告》、《毕业生就业质量报告》、《“十三五”江苏省重点学科立项学科名单》、《高等学校科技统计资料汇编》、《关于公布世界一流大学和一流学科建设高校及建设学科名单的通知》、《中国大学评价》等统计报告及相关数据。

我们将所给的 9 个指标分为两类，可直接量化的指标与不可直接量化的指标。可直接量化的指标包括本科院校数量、招生人数、生师比、学生就业与双一流学科建设。对于需要间接量化的指标，包括师资队伍与解雇、教学条件与运用、专业建设与教学改革和科研投入与产出，我们首先选取对应的子指标，然后分别利用主成分分析的对这些指标进评价，将得到的评价价值作为最后的量化值。

## 5.1 直接量化指标的定义与说明

- 本科院校数量  $x_1$ : 各地级市的本科院校数量为所有改市所有本科院校数量之和。
- 招生人数  $x_2$ : 根据各学校的统计报告显示，各学校每年近几年的招生数量均没有发生较大幅度的变动，维持在一个相对稳定的水平，考虑数据搜集的完整性，我们选取了各本科院校 2016 年的实际录取人数，并以该市各学校的实际录取人数的加权平均代表该市的平均招生人数。

$$x_2 = \frac{1}{n} \sum_{i=1}^n p_i \quad (5-1)$$

其中  $n$  为各市院校数量， $p_i$  为第  $i$  个学校的实际录取人数

- 生师比  $x_4$ : 各学校的生师比为各学校的全日制在校生与专任教师的比例。

$$\frac{\sum_{i=1}^n s_i}{\sum_{i=1}^n t_i} \quad (5-2)$$

其中  $n$  为各市学校数量， $s_i$  为各校学生数量， $t_i$  为各校专任老师数量

- 学生就业  $x_7$ : 选取各本科院校 2017 届本科毕业生截止到 2016-2017 学年末的就业率，以各地级市所有本科院校就业率的加权平均代表该市的学生就业情况。

$$\frac{\sum_{i=1}^n g_i}{\sum_{i=1}^n w_i} \quad (5-3)$$

$g_i$  为第  $i$  个学校应届毕业生就业人数， $w_i$  为第  $i$  个学校应届毕业生人数

- 双一流学科建设  $x_9$ ：以各地级市中所有本科院校的一流学科数量的加权平均代表该是的一流学科建设水平。

$$\frac{1}{n} \sum_{i=1}^n o_i \quad (5-4)$$

$i$  为各市学校数量， $o_i$  为各校双一流学科数。

## 5.2 间接量化指标的主成分分析综合评价模型

### 5.2.1 间接量化指标评价子指标的遴选与说明

对于师资队伍与结构、教学条件与利用、专业建设与教学改革、科研投入与产出四个指标，每个指标均包含多方面的内容，根据各院校公布的《本科教学质量评估报告》和相关文献中<sup>[1]</sup>关于教学评估指标体系的研究，对 4 个指标分别选取子指标进行评价，用评价值作为该指标的量化值，各指标说明如下：

表 5.1 间接量化指标子指标的遴选与说明

指标	子指标	指标解释
师资队伍与结构	● 专任教师的正高级职称比例	● 各院校正高级职称教师数 / 各院校专任教师数
	● 专任教师的副高级职称比例	● 各院校副高级职称教师数 / 各院校专任教师数
	● 专任教师的博士学位比例	● 各院校博士学位教师数 / 各院校专任教师数
	● 教授授课情况	● 各院校教授授课程比例之和 / 院校数量
	● 教师学术水平	● 各院校教师学术水平得分 / 院校数量
教学条件与运用	● 校园占地面积（平方米）	● 各院校校园占地面积之和 / 院校数量
	● 图书馆藏书（万册）	● 各院校馆藏图书数量 / 院校数量
	● 国家重点实验室个数	● 各院校国家重点实验室数量之和

	<ul style="list-style-type: none"> <li>● 教育部重点实验室个数</li> <li>● 省重点实验室个数</li> <li>● 本科专项教育经费（万元）</li> </ul>	<ul style="list-style-type: none"> <li>● 各院校教育部国家重点实验室数量之和</li> <li>● 各院校省重点实验室数量之和</li> <li>● 各院校本科专项教研经费之和 / 院校数量</li> </ul>
专业建设与教学改革	<ul style="list-style-type: none"> <li>● 专业数量</li> <li>● 国家级特色专业数量</li> <li>● 一级国家重点学科数量</li> <li>● 二级国家重点学科数量</li> <li>● 省重点学科数量</li> <li>● 省优势学科数量</li> </ul>	<ul style="list-style-type: none"> <li>● 各院校开始专业之和 / 院校数量</li> <li>● 各院校之和</li> <li>● 各院校之和</li> <li>● 各院校之和</li> <li>● 各院校之和</li> <li>● 各院校之和</li> </ul>
科研投入与产出	<ul style="list-style-type: none"> <li>● 教学与科研人员</li> <li>● 研究与发展人员</li> <li>● 科技经费（千元）</li> <li>● 科技课题数</li> <li>● 成果授奖</li> </ul>	<ul style="list-style-type: none"> <li>● 各院校人数之和</li> <li>● 各院校人数之和</li> <li>● 各院校之和</li> <li>● 各院校之和</li> <li>● 各院校奖项之和</li> </ul>

注：表中所列科研投入与产出数据为 2016 年度数据，教师水平得分为 2017 年公布数据，其余均为 2016-2017 学年数据（部分还未更新数据的院校采用上一年的统计数据），统计的截止时间为学年末结束时。

### 5.2.2 间接量化指标主成分分析综合评价模型的建立与求解

对于 4 项不可直接量化的指标，在进行了子指标遴选的条件下分别建立主成分分析的综合评价模型，计算评价值作为其最终量化值。

以师资队伍与结构为例说明主成分分析的过程与结果：

**Step1:**数据的预处理：

本部分共涉及 5 个子指标且均为“极大型”指标，为消除不同变量的量纲的影响，对变量进行极差化处理。对于 13 个评价对象的 5 项指标的指标值

对于 13 个地级市的 5 个指标  $a_{nm}$  ( $n = 1, 2, \dots, 13; m = 1, 2, 3, 4, 5$ ),

令  $M_m = \max \{ a_{1m}, a_{2m}, \dots, a_{13m} \}$ ,  $m_m = \min \{ a_{1m}, a_{2m}, \dots, a_{13m} \}$

则新的指标为：

$$\tilde{a}_{nm} = \frac{a_{nm} - m_m}{M_m - m_m} \quad (5-5)$$

**Step2:**计算标准化相关系数矩阵：

$$r_{mm'} = \frac{\sum_{k=1}^{13} \tilde{a}_{km} \tilde{a}_{km'}}{13-1}, m, m' = 1, 2, 3, 4, 5 \quad (5-6)$$

则相关系数矩阵为  $R = (r_{mm'})_{5 \times 5}$

**Step3:** 计算特征值与特征向量:

计算相关系数矩阵  $R$  的特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_5 \geq 0$ , 及其对应的特征向量  $\xi_1, \xi_2 \dots \xi_5$ , 其中  $\xi_m = [\xi_{1m}, \xi_{2m}, \dots, \xi_{5m}]^T$ , 由特征向量组成 5 个新的指标变量

$$\begin{cases} y_1 = \xi_{11} \tilde{a}_1 + \xi_{21} \tilde{a}_2 + \xi_{31} \tilde{a}_3 + \xi_{41} \tilde{a}_4 + \xi_{51} \tilde{a}_5 \\ y_2 = \xi_{12} \tilde{a}_1 + \xi_{22} \tilde{a}_2 + \xi_{32} \tilde{a}_3 + \xi_{42} \tilde{a}_4 + \xi_{52} \tilde{a}_5 \\ \dots \\ y_5 = \xi_{15} \tilde{a}_1 + \xi_{25} \tilde{a}_2 + \xi_{35} \tilde{a}_3 + \xi_{45} \tilde{a}_4 + \xi_{55} \tilde{a}_5 \end{cases}$$

其中,  $y_i$  为第  $i$  主成分,  $i=1, 2, \dots, 5$

**Step4** 选择  $p$  个主成分, 计算综合评价

根据以上步骤, 本文利用 SPSS 统计软件, 首先求的各指标的相关系数表见附录 1, 从表中可以发现, 某些指标具有很强的相关性, 主成分分析可以把多个指标转化成少数几个不相关的综合指标。求得的主成分分析结果如下表:

表 5.2 主成分分析结果

累计贡献率						
主成分	提取平方和载入			旋转平方和载入		
	合计	方差的%	累计%	合计	方差的%	累计%
1	4.205	84.092	84.092	4.205	84.092	84.092
2	.511	10.219	94.311	.511	10.219	94.311
3	.181	3.629	97.940			
4	.065	1.298	99.238			
5	.038	.762	100.000			

可以看出, 前两个特征根的累计贡献率达到了 94% 以上, 主成分分析的效果很好。下面选取前两个主成分进行综合评价。前两个特征根对应的特征向量表如下:

### 5.3 标准化变量的前 2 个主成分对应的特征向量

	$\tilde{a}_1$	$\tilde{a}_2$	$\tilde{a}_3$	$\tilde{a}_4$	$\tilde{a}_5$
第一特征向量	0.459603	0.435292	0.440535	0.453161	0.447057
第二特征向量	-0.27372	-0.60175	0.545187	-0.15834	0.490587

由此得到 2 个主成分分别为:

$$y_1 = 0.459\tilde{a}_1 + 0.435\tilde{a}_2 + 0.440\tilde{a}_3 + 0.453\tilde{a}_4 + 0.447\tilde{a}_5$$

$$y_2 = -0.273\tilde{a}_1 - 0.601\tilde{a}_2 + 0.545\tilde{a}_3 - 0.158\tilde{a}_4 + 0.490\tilde{a}_5$$

分别以 2 个主成分的贡献率为权重，构建主成分评价模型：

$$x_3 = 0.8409y_1 + 0.1022y_2 \quad (5-7)$$

教学条件与运用、学科建设与教学改革、科技投入与产出所得到的主成分分析结果及特征向量见附录 1。其主成分综合评价模型分别为：

1. 教学条件与运用：

$$x_5 = 0.5202z_1 + 0.4172z_2 \quad (5-8)$$

2. 学科建设与教学改革：

$$x_6 = 0.829w_1 + 0.116w_2 \quad (5-9)$$

3. 科技投入与产出

$$x_8 = 0.988q_1 + 0.007q_2 \quad (5-10)$$

其中 $z_i, w_i, q_i$ 分别为三项指标对应的主成分， $i=1, 2$

### 5.3 指标量化方式与计算依据汇总

#### 5.4 量化处理方式及计算依据汇总

序号	指标名	量化处理方式	计算依据	备注
1	本科院校数量	地级市本科院校数量相加求和	$X_i$	i 为城市序号
2	招生人数	地级市各本科院校的平均实际录取人数的加权平均	$\frac{1}{n} \sum_{i=1}^n p_i$	n 为各市院校数量， $p_i$ 为第 i 个学校的实际录取人数
3	师资队伍与结构	利用主成分分析法进行评价	$\sum_{i=1}^n k_i y_i$	n 为主成分个数， $k_i$ 为第 i 个主成分贡献度， $y_i$ 为第 i 个主成分
4	生师比	地级市中各本科院校生师比的加权平均	$\frac{\sum_{i=1}^n s_i}{\sum_{i=1}^n t_i}$	n 为各市学校数量， $s_i$ 为各校学生数量， $t_i$ 为各校专任老师数量

5	教学条件与利用	利用主成分分析法进行评价	$\sum_{i=1}^n k_i y_i$	n 为主成分个数, $k_i$ 为第 i 个主成分贡献度, $y_i$ 为第 i 个主成分
6	专业建设与教学改革	利用主成分分析法进行评价	$\sum_{i=1}^n k_i y_i$	n 为主成分个数, $k_i$ 为第 i 个主成分贡献度, $y_i$ 为第 i 个主成分
7	学生就业	各地级市所有本科院校就业率的加权平均	$\frac{\sum_{i=1}^n g_i}{\sum_{i=1}^n w_i}$	n 为各市学校数量, $g_i$ 为第 i 个学校应届毕业生就业人数, $w_i$ 为第 i 个学校应届毕业生人数
8	科研投入与产出	利用主成分分析法进行评价	$\sum_{i=1}^n k_i y_i$	n 为主成分个数, $k_i$ 为第 i 个主成分贡献度, $y_i$ 为第 i 个主成分
9	双一流学科建设	各市内院校双一流学科数量加权平均	$\frac{1}{n} \sum_{i=1}^n o_i$	i 为各市学校数量, $o_i$ 为各校双一流学科数

#### 5.4 地级市量化指标数据汇总

对于不可量化的 4 项指标, 把各城市的主成分值分别带入公式 (5-7) (5-8) (5-9) (5-10), 求的 13 个城市的量化值。所有指标最后的量化情况如下:

表 5.5 所有指标量化统计数据汇总

城市 \ 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
南京	25	3728	1.517	16.923	0.837	1.894	95.60%	2.212	1.52
无锡	2	4601	1.285	17.913	0.282	0.074	95.97%	0.215	1
徐州	4	4746	1.105	17.100	0.470	0.305	98.56%	0.256	0.5
常州	3	4018	0.723	18.000	0.354	0.105	96.51%	0.079	0
苏州	4	4407	1.319	14.723	0.272	0.217	96.06%	0.278	0.25
南通	1	6340	1.546	14.270	0.983	0.089	98.52%	0.078	0
连云港	1	4421	0.588	16.810	0.245	0.067	98.28%	0.019	0
淮安	2	5419	0.750	16.670	0.369	0.084	98.11%	0.035	0
盐城	2	5634	0.744	16.210	0.344	0.080	98.32%	0.070	0
扬州	1	6414	2.022	17.210	0.921	0.157	98.42%	0.147	0
镇江	2	4845	1.567	17.730	0.595	0.121	98.36%	0.236	0
泰州	1	1977	0.156	16.100	0.000	0.000	94.46%	0.001	0
宿迁	1	3500	0.103	17.400	0.122	0.000	92.50%	0.001	0

通过表中数据可以分析得到如下结论：

- 在 9 项指标中，生师比这一指标的地区差异最小，除常州、苏州和南通之外均在 16~17 之间。
- 在 9 各指标中，师资队伍与结构、专业建设与教学改革、科研投入与产出以及双一流学科建设这几项指标在地区之间有很大不同。说明江苏省教学资源的分布和科研水平存在着显著差异。

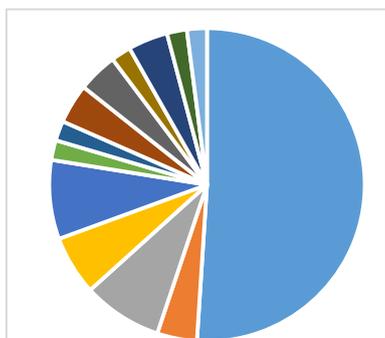


图 5.1 江苏省本科院校数量分布

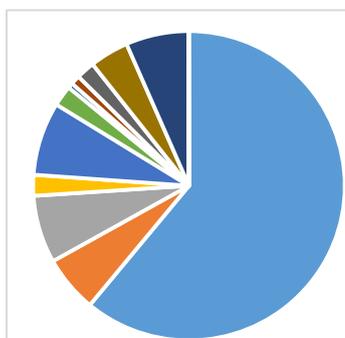


图 5.2 江苏省科研投入与产出分布

由图可以看出，在很多指标上，南京市都处于绝对的领先地位，比如在江苏省 49 所本科院校中南京就占了 25 所，另外其中还包含 8 所 958、211 院校，占到江苏省 985、211 院校的 60% 以上。在科研投入与产出、专业建设与教学改革及双一流学科建设等方面南京的量化值也超过了其他所有地级市之和。

## 六. 问题二模型的建立与求解

### 6.1 基于熵权法本科教学质量综合评价模型

#### 6.1.1 数据的预处理

##### 1. 数据的一致化

在一个综合评价问题中，可能同时含有多种类型的指标，需要先将这些指标做一致化处理。对于问题一中所得到的 9 项指标，有 8 项为“极大型”指标，但生师比为“极小型”数据，即生师比越小意味着教师可以更加有针对性的进行教学，从而提升教学质量，将其变为“极大型”数据，只需取他的倒数即生师比。

$$x'_{i4} = \frac{1}{x_{i4}} \quad (6-1)$$

##### 2. 数据的无量纲化

各评价指标之间存在着明显的度量单位及数量级的差别，为了消除量纲的影响，用极差化的方法对 9 项指标做无量纲化的处理。

对于 13 个地级市的 9 个指标  $x_{ij}$  ( $i = 1, 2, \dots, 13; j = 1, 2, 3, 4, 5$ )，令

$$M_j = \max \{x_{1j}, x_{2j}, \dots, x_{13j}\} m_j = \min \{x_{1j}, x_{2j}, \dots, x_{13j}\}$$

则新的指标为:

$$x_{ij}^* = \frac{x_{ij} - m_j}{M_j - m_j} \in [0, 1]$$

即 $x_{ij}^*$  ( $i=1, 2, \dots, 13; j=1, 2, 3, 4, 5$ )为无量纲的标准值。

### 6.1.2 基于熵权法的本科教学质量综合评价模型的建立

**Step1:** 计算标准值 $x_{ij}^*$ 的比重 $R_{ij}$ , 其中:

$$R_{ij} = \frac{x_{ij}^*}{\sum_{i=1}^{13} x_{ij}^*} \quad (6-2)$$

**Step2:** 计算第 $j$ 项指标的熵值 $e_{jk}$ , 其中:

$$e_{jk} = - \left( \frac{1}{\ln 13} \right) \sum_{i=1}^{13} R_{ij} \quad (6-3)$$

**Step3:** 计算第 $j$ 项指标的差异性系数 $g_{jk}$ , 其中

$$g_{jk} = 1 - e_{jk} \quad (6-4)$$

**Step4:** 计算指标 $x_j$ 的权重 $w_{jk}$ , 计算公式为

$$w_{jk} = \frac{g_{jk}}{\sum_{i=1}^{13} g_{jk}} \quad (6-5)$$

可以得到各地级市本科教学质量的综合得分 $Z_i$ 为:

$$Z_i = \sum_{j=1}^{13} w_{jk} x_{ij}^* \quad (6-6)$$

至此, 基于熵权法的综合评价模型建立完毕。

### 6.1.3 基于熵权法的本科教育质量综合评价模型的求解

根据上文的步骤，利用 MATLAB 对熵权编程计算，得到各项指标的权重如下表：

表 6.1 各指标熵权表

指标名称	权重
本科院校数量	22.57%
招生人数	2.57%
师资队伍与结构	3.62%
生师比	5.59%
教学条件与运用	4.36%
专业建设与教学改革	17.22%
学生就业	2.16%
科研投入与产出	18.62%
双一流学科建设	23.27%

根据得出的本科教学质量综合评价模型，利用 MATLAB 就算可得 13 个地级市的本科教育质量如下：

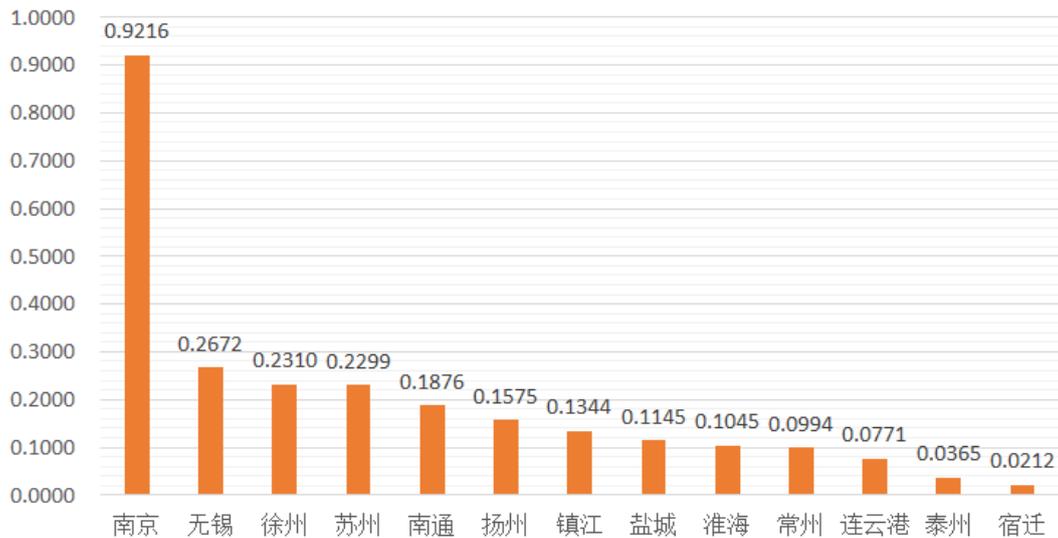


图 6.1 各地级市本科教学质量评价价值

由图表可以看出

- 江苏省各地级市本科教学质量水平差异较大，其中南京的教学质量水平明显优于其他地级市。
- 无锡、徐州的本科教学水平紧随其后，但与南京已有了较大的差别。
- 苏州、南通、扬州、镇江四个城市的等地级市的教学水平较为平均，处于中间水平。
- 盐城、淮安、常州、连云港及泰州 5 个城市的本科教学质量水平较为平均，在江苏省处于比较落后的水平。
- 宿迁市的本科教学质量水平在整个江苏省排名最低。

## 6.2 基于灰色关联分析的本科教学质量综合评价模型

### 6.2.1 基于灰色关联分析的本科教学质量综合评价模型的建立

**Step1:**确定比较对象和参考数列

评价对象为 13 个城市，评价指标为 9 个，参考数列取各指标的最大值。

参考数列为  $X_0 = \{X_0(k) | k=1, 2, \dots, 9\}$ ，比较数列为  $X_i = \{X_i(k) | k=1, 2, \dots, 9\}$ ， $i=1, 2, \dots, 13$

**Step 2:**确定各指标的权重

各指标对应权重  $w = [w_1, \dots, w_9]$ ，其中  $w_k (k = 1, 2, \dots, 9)$  为第  $k$  个评价指标对应权重

**Step3:**计算灰色关联系数

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \quad (6-7)$$

**Step4:**计算灰色加权关联度

灰色加权关联度的计算公式为:

$$r_i = \sum_{k=1}^n w_k \xi_i(k) \quad (6-8)$$

至此，基于灰色关联分析的综合评价模型建立完毕。

### 6.2.2 基于灰色关联分析的本科教学质量综合评价模型的求解

利用 MATLAB 求解得到各城市关联度值如下:

表 6.2 关联度值

城市	$r_i$	城市	$r_i$
南京	0.925865	连云港	0.38696
无锡	0.437175	淮安	0.368733
徐州	0.404301	盐城	0.372195
常州	0.407085	扬州	0.38304
苏州	0.41297	镇江	0.351809
南通	0.397382	泰州	0.369918
		宿迁	0.338736

## 6.3 本科教学质量水平排行

基于上述两种综合评价方法，对各城市的本科教学质量进行排名如下:

表 6.3 两种综合评价方法排名

城市	熵权法排名	灰色关联排名
南京	1	1
无锡	2	2
徐州	3	6
常州	10	11
苏州	4	4
南通	5	5
连云港	11	10
淮安	9	9
盐城	8	8
扬州	6	3
镇江	7	7
泰州	12	12
宿迁	13	13

由表示数据可以看出，在两种综合评价方法下各城市的排名基本一致，有部分城市的排名在较小范围内发生了变动，说明了两种评价模型的一致性和合理性。

## 七. 问题三模型的建立与求解

### 7.1 可减少指标的确定

对于影响各地级市本科教学质量的 9 个指标，如果要找出减少哪些指标可以使评价结果基本一致，不仅应保证排名顺序的一致，还应该使剩余因子（关键因子）仍具有代表性和可靠性，如果剩余的指标体系只能反应中极少部分原始信息，那么这样的评价结果也并不具有说服力，因此剔除后的指标体系需要满足以下两个条件：

1. 对本科教学质量的综合评价排名结果与问题二中的综合评价模型排名结果一致。
2. 剩余因子保留了原始指标体系绝大部分信息，本文选取 90% 为界。

设  $w$  为问题二中基于熵权法确定的各指标权重集合

$$w = \{w_1, w_2, \dots, w_9\}$$

$$\text{求 } Y \in P(\{1, 2, 3, 4, 5, 6, 7, 8, 9\})$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{13} \sum_{j \in Y} (D_{ij} - D_{ij}')^2 = 0 \\ \prod_{j \in Y} (1 - w_j) \geq 90\% \end{cases}$$

其中  $D_{ij}$ ,  $D_{ij}'$  分别为剔除第  $j$  个指标前后各城市本科教学质量的排名。

计算减少指标后的剩余指标权重、新的排名和可减少指标的算法如下图：

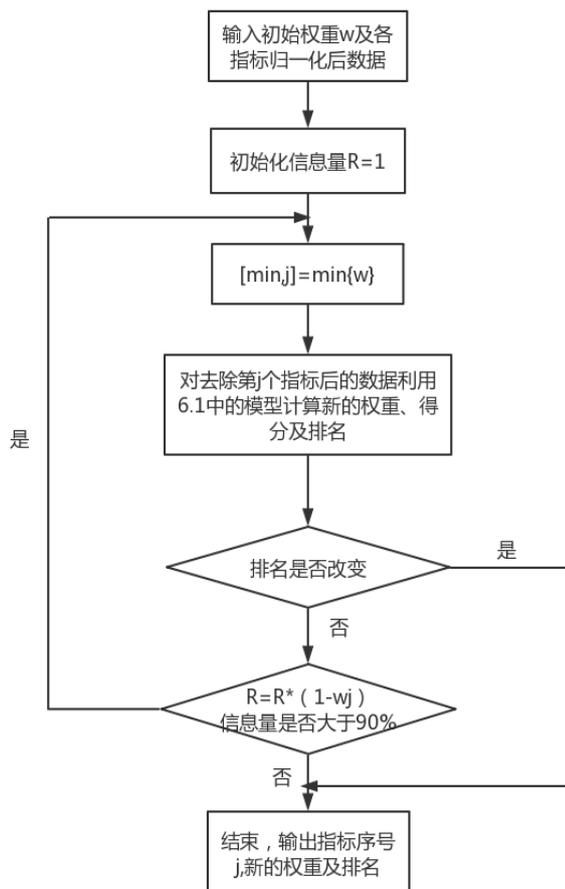


图 7.1 算法流程图

## 7.2 问题三模型的求解

利用 Matlab 对结果进行求解，求解结果见表 7.1 和表 7.2

表 7.1 剩余指标权重及保留信息结果

指标序号	指标名称	筛选次数			
		原始指标 权重	1	2	3
1	本科院校数量	22.57%	23.07%	23.69%	24.63%
2	招生人数	2.57%	2.63%		
3	师资队伍与结构	3.62%	3.70%	3.80%	
4	生师比	5.59%	5.72%	5.87%	6.10%
5	教学条件与利用	4.36%	4.46%	4.58%	4.76%
6	专业建设与教学改革	17.22%	17.60%	18.07%	18.79%
7	学生就业	2.16%			
8	科研投入与产出	18.62%	19.03%	19.55%	20.32%
9	双一流学科建设	23.27%	23.79%	24.43%	25.40%
合计权重		100.00%	100.00%	100.00%	100.00%
保留信息		100.00%	97.84%	95.27%	91.65%

表 7.2 评价值及排名结果

城市	剔除指标次数				原始排名	第三次剔除后排名
	0	1	2	3		
南京	0.9216	0.9307	0.9426	0.9480	1	1
无锡	0.2672	0.2605	0.2499	0.2341	2	2
徐州	0.2310	0.2220	0.2198	0.2069	3	3
常州	0.0994	0.0863	0.0747	0.1934	5	4
苏州	0.2299	0.2140	0.2114	0.1246	6	5
南通	0.1876	0.1698	0.1478	0.0812	10	6
连云港	0.0771	0.0578	0.0426	0.0727	11	7
淮安	0.1045	0.0870	0.0668	0.0648	9	8
盐城	0.1145	0.0958	0.0754	0.0636	8	9
扬州	0.1575	0.1394	0.1161	0.0557	4	10
镇江	0.1344	0.1161	0.1003	0.0340	7	11
泰州	0.0365	0.0302	0.0269	0.0255	12	12
宿迁	0.0212	0.0217	0.0103	0.0108	13	13

根据表中数据可知，一共进行了三次指标的剔除，前两次提出后各地级市的排名均没有发生变化且保留的信息在 95% 以上，在剔除第三个指标后虽然保留的信息仍在 90% 以上但排名已发生了较为明显的变化。原始评价值与每次剔除指标后的评价值雷达图如下：

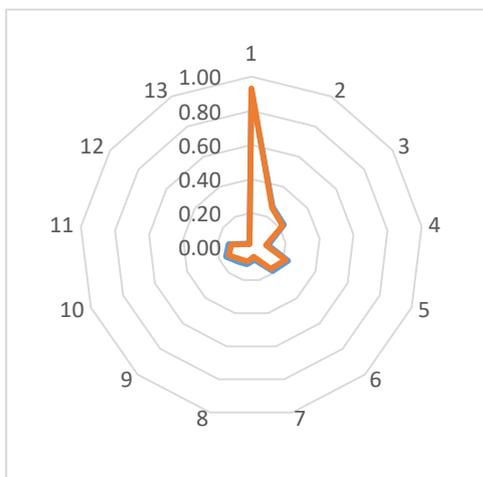


图 7.1 原始值与剔除一次指标后评价值

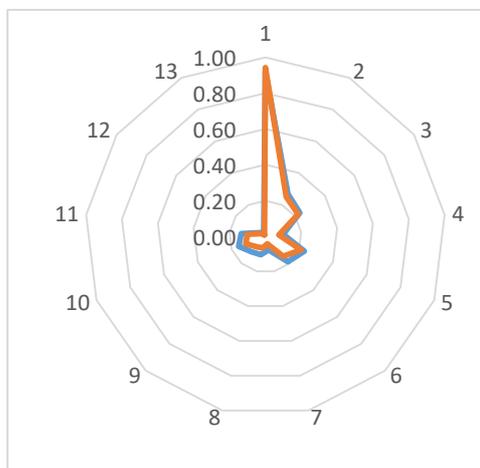


图 7.2 原始值与剔除两次次指标后评价值

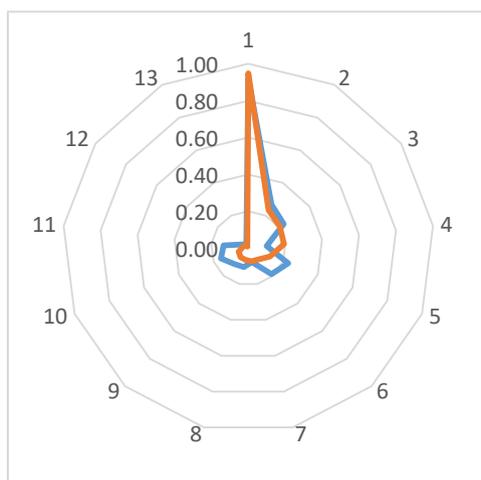


图 7.3 原始值与剔除三次指标后的评价值

由雷达图可以看出，在前两次剔除数据之后，两条曲线几乎重合，前后的评价结果变化，但剔除第三个指标后，两条曲线产生了明显的不同，即前后的评价结果存在明显的差异。

综合上述的分析可知，剔除掉学生就业、招生人数这两项指标可以使评价结果与问题二结果基本一致。

## 八. 问题四模型的建立与求解

### 8.1 关键指标的确定

所谓关键指标，即其普遍改善能尽可能缩小江苏省各地级市本科教育发展的差异，而差异可由一组评价结果的方差体现。这里以二维空间为例，说明关键指标的确定方法。

假设各市本科教学质量只与指标  $u$  和  $v$  有关，要使得指标值获得普遍改善，即越落后的地区越需要改善，我们以各市的指标值与最大指标值的差为依据，认为差距越大，对其改善值应越大。为指标  $u$  与  $v$  提供一方向向量，如  $(1, 0)$  表示只改善指标  $x$ ， $(0, 1)$  表示只改善指标  $y$ 。以各市各指标的最大值作为理想点的指标值，则任给一方向向量，我们就可以让样本点以相同的改善指标值的方式逼近理想点。使得方差最小的方向向量的分量值即表明其是否是关键指标。现将其扩展到 9 维空间。

$$\text{令 } x_{im} = \max(x_{1i}, x_{2i}, \dots, x_{9i}), i = 1, 2 \dots 9,$$

取  $(x_{1m}, x_{2m}, x_{3m}, x_{4m}, x_{5m}, x_{6m}, x_{7m}, x_{8m}, x_{9m})$  为理想点，用不同的方向向量转换矩阵，再利用熵权法对矩阵进行权重计算得到分数，对得到的一组评价求方差。求得方差取最小值时的方向向量，根据方向向量值得大小确定关键指标，分量值最大则为关键指标。

### 8.2 基于模拟退火算法的模型求解

模拟退火算法描述如下:

**Step1:解空间的确定**

解空间可表示为 $B\{b_1, b_2, \dots, b_9\}$ , 且满足 $\sum_{i=1}^9 b_i = 1$

样本空间的处理:

样本空间统一采用 $x_{ij} = x_{ij} + b_j \times (x_{\max j} - x_{ij})$ 进行矩阵变换得到矩阵 $X_{ij}$ , 熵权法求得得分为scores。

**Step2:目标函数**

目标函数为熵权法得分的方差。要求

$$\min f(\text{scores}_1, \text{scores}_2, \dots, \text{scores}_{13}) = \sum_{i=1}^{13} \frac{(\text{scores}_i - \overline{\text{scores}})^2}{n}$$

**Step3:新解的产生。**

设上一步迭代的解为 r (r1,r2,...,r9), 在此基础上增加扰动并保持其之和为 1.

得到新解为 $r' = \{r'_1, r'_2, \dots, r'_9\}$ ;

**Step4:代价函数差。**

$$\Delta f = \exp^{(E_{\text{new}} - E_{\text{best}})/t}$$

其中  $E_{\text{new}}$  为当前产生新解求得方差,  $E_{\text{best}}$  为当前保存最佳方差,  $t$  为当前温度。

**Step:接受准则**

$$P = \begin{cases} 1, & E_{\text{new}} - E_{\text{best}} < 0 \\ \exp(-\Delta f/t) & E_{\text{new}} - E_{\text{best}} > 0 \end{cases}$$

如果  $E_{\text{new}} - E_{\text{best}} < 0$  则接受新的路径, 否则以概率  $\exp(-\Delta f/t)$  接受新的路径。

**Step6:降温**

利用选定的降温系数 $\alpha$ 进行降温, 取新的温度 $t$ 为 $\alpha t$ , 这里选定 $\alpha=0.99$

**Step7:结束条件**

选定终止温度为 $t_f = 89.9^\circ\text{C}$ , 判断退火过程是否结束。若 $t \geq t_f$ ,则算法结束, 最输出当前状态

终求得解如下表:

表 8.1 各指标方向向量数据

指标序号	1	2	3	4	5	6	7	8	9
方向分量值	0.05	0.02	0.4	0.05	0	0.09	0.1	0.07	0.14

### 8.3 模型的检验

根据求解得到的方向向量，可以看出，第三个指标(师资队伍与结构)的明显改善可以减小市级间本科教学质量的差异，第二个指标(招生数量)的改善对减小本科教学质量的差异没有明显帮助。

分别针对这两个方向，进行单方向逼近最优解，检验他们的方差变化值，发现第三个指标的改善导致的方差变化最大，第二个指标的改善导致的方差变化最小。这两种指标以相同方式改善指标值后的方差及方差变化如下：

表 8.2 方差及方差变化

指标	方差	方差变化百分比
招生数量	0.053042	0.72%
师资队伍与结构	0.050512	5.45%

与通过模拟退火算法得到的结果吻合。

## 九. 问题五的求解

本科教学对于一个省的整体发展起着至关重要的作用，江苏作为全国的教育大省，始终以全面提高教育质量为核心，大力推进高等教育内涵建设，加强优势学科、品牌专业和“双一流”建设，持续深化高等教育综合改革，在本科教育人才培养模式改革、高职教育和研究生教育创新发展等方面深化推进。但是仍存在教育水平差异大，教学资源分布不均衡的问题，基于上文建立的综合评价模型，本文对江苏省的本科教育发展提出如下建议：

### 1. 更大程度的鼓励教师投入本科教学

提升本科教学质量根本还是要提高教师教学质量。师资队伍的结构普遍改善可以尽可能地缩小各地级市之间的差异。学校可以通过津贴方案、绩效奖励、建设经费支持等多种措施调动教师投入教学的积极性。但学校对各类教师相应的工作职责、工作内容、工作量还没有做出明确要求，还没有充分考虑专业课与基础课教师之间、理论课教师与实验室技术人员之间、学科专业之间、教学与科研之间的工作差异

### 2. 增强地区间学术交流

根据统计数据表明，江苏省各地级市之间在多个本科教育质量的指标方面均存在这较大的差异，教学资源、科研投入与产出等在地区之间的分布很不均衡，为了改善整体的教学水平，应充分发挥地区优势，积极推进各地区高校之间的学术文化交流等活动，使各院校的资源可以得到充分的利用。

### 3. 建立起以“学”为中心的人才培养模式

其核心就是在人才培养过程实现从重“教”向促“学”转变，改革知识本位和社会本位的人才培养目标，树立以能力培养、素质提升、个性彰显为重点的新型人才培养目标。把促进人的全面发展和适应社会需要作为衡量人才培养水平的根本标准，树立多样化人才和人人成才观念，努力造就信念执著、品德优良、知识丰富、本领过硬的高素质人才。重视文理交融，促进科研与教学互动，健全与科研院所、行业企业协同育人的机制，探索国际合作培养人才。加快建设高校教学联盟，打破

---

学校之间、专业模块之间的壁垒，推进跨专业、跨学科、跨学校的学分互认。

#### 4. 重视实践教学环节，提升高素质创新人才培养力度

结合专业特点和人才培养要求，增加实践教学比重，改革实践教学内容，改善实践教学条件，建立基础实验、综合性实验、创新性实验、研究性实验等多种实验构成的实验教学体系。配齐配强实验室人员，鼓励高水平教师承担实验教学，加大学科实验室和科研项目等资源向学生开放的力度。同时应注重对教师实践教学质量的检查和评估，而且将其制度化并落实到学校、院系级教学计划、教学单位和教师个人各个层面，明确基层教学单位对实践教学的质量监控职责。

#### 5. 以信息化建设为中心提升整体教学服务质量

在信息技术飞速发展的今天，移动技术、人工智能、大数据等高新技术对高等教育的影响愈加广泛。一方面，新兴的微博、微信等社交平台以及青年学生的网络社交偏好对稍显传统的高校教学管理信息化平台提出了更高的挑战；另一方面，慕课、“翻转课堂”等新型课程迫切需要更为灵活有效的线上教学方式。各本科院校也应该积极推进教学设备、教学方式的信息化更新，更好实现院校之间的信息共享，提升整体的教与质量水平。

## 十. 模型的评价、改进与推广

本文利用主成分分析法对无法直接量化的指标进行量化，结合熵权法建立综合评价模型，避免了模型建立的主观性。同时将结果与灰关联法进行结合，验证了模型的准确性。第三问利用关键因子识别法，既保证了评价结果排名不变，又保留了原指标中绝大部分的信息。

对于问题一，在对指标通过主成分分析法量化时，选取指标较少，且有部分数据缺失，可能会导致指标量化值的不准确。可以基于更全面的考虑，从更多的角度选取更加全面的指标进行评价，是评价的结果更具有合理性。对于问题二，在进行灰色关联分析时对于各指标权重的确定方法带有一定的主观性，可能会使结果存在偏差。对于问题四中的算法，可以进行学习，对当前最优解增加随机扰动而不是任意给方向向量赋值，增加了程序的运行效果。

目前本文建立的综合评价模型具有一定推广价值。在已知某些本科教育条件的情况下，可以对该市的本科教育质量作出一个综合的评价，同时第四问建立的模型对提高本科教育质量具有一定的启发性。

---

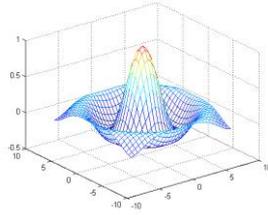
## 十一. 参考文献

- [1]王淼. 研究型大学专业评估指标体系的构建研究[D].兰州大学,2016.
- [2]史艳伟. 中共党史党建学科建设的改革与创新[J].中国高新区,2018(08):292.
- [3]孙绪华. 我国科技资源配置的实证分析与效率评价[D].华中农业大学,2011.
- [4]吴薇,刘璐璐. “双一流”建设背景下研究型大学师资队伍建设的现状与对策——基于 16 所“985”高校师资队伍基本状态数据的分析[J].中国高等教育评论,2018,9(01):209-226.
- [5]陈德胜,刘永凤. 加强教师队伍建设 努力办高品质教育——江苏省扬州市师资队伍建设[J].基础教育参考,2017(24):18-20.

## 附录一

次序	城市	分数
1	南京	1.893775 222
2	无锡	0.073985 482
3	徐州	0.304574 834
4	常州	0.105188 791
5	苏州	0.216510 069
6	南通	0.089005 964
7	连云港	0.066529 628
8	淮安	0.084185 366
9	盐城	0.079981 984
10	扬州	0.156501 392
11	镇江	0.120779 679
12	泰州	0
13	宿迁	0

# 五一数学建模竞赛



题目：\_\_\_\_\_江苏省本科教育质量综合评价\_\_\_\_\_

关键词：熵权法 *min-max* 标准化法 主成分分析法 *TOPSIS* 法 皮尔逊相关  
灰色理论优势分析 线性回归模型 *F* 检验和 *t* 检验

**摘要：** 本文根据江苏省各市的本科高校统计数据，对 13 个地级市的本科教育质量进行全方位的综合评价，根据指标得分对其进行了排名并找出了对本科教育质量影响较大和较小的指标，找出缩小各市发展差距的途经，提出了一系列政策建议。

针对问题一，对 9 个指标进行量化处理，针对一些泛指类指标建立了三级指标体系，根据三级指标复杂程度将 9 个二级指标分为 I 类、第 II 类、第 III 类，利用 *MATLAB* 运用熵权法、主观赋权法、*min-max* 标准化法分别进行量化，得到更准确的量化结果。

针对问题二，首先用主成分分析法，利用问题一的量化结果计算各市本科教学质量的综合得分，利用 *SPSS* 得到主成分得分  $F = 0.115x_1 + 0.114x_2 + 0.109x_3 + 0.114x_4 + 0.115x_5 + 0.115x_6 + 0.114x_7 + 0.108x_8 + 0.115x_9$ ，得到南京市第一、泰州市最后，同时引进了 *TOPSIS* 法(逼近理想解排序法)，运用 *MATLAB* 相关函数，得到对理想解的逼近程度，对排名靠后的几个城市进行排名的修正，得到南京市第一、宿迁市最后，即为更为准确的排名情况。

针对问题三，根据问题二中的主成分分析法的结论，进行因子分析，得到其成分矩阵和相关系数矩阵，再利用 *SPSS* 得出皮尔逊相关系数并进行二次筛选，通过皮尔逊相关表与成分得分矩阵综合分析，发现减少招生人数、师生比、专业建设与教学改革这三项指标对问题二中的评价结果影响不大。

针对问题四，利用灰色理论中的优势分析，利用 *MATLAB* 相关函数得到关联矩阵，由矩阵可以看出，师资队伍与结构  $r_{15} = 0.936$ ，大于任意  $r_{ij}$ ，即为优势子因素，然后运用线性回归模型中的 *F* 检验和 *t* 检验对优势因子进行检验，最终拟合曲线结果  $R^2$  为 0.992、*F* 值为 1445.783、*P* 值 (*sig*) 为 0.000、*sig* 均  $< 0.05$  即通过 *F* 检验和 *t* 检验，即可验证结论。查阅相关资料也可证实，最终得到关键指标为：师资队伍与结构。

针对问题五，根据本文建立的模型，向政府部门提出了有效提升江苏省本科教学质量的政策建议；并从持续推进校内师资队伍与结构的改善、杜绝盲目扩招，制定人才培养模式、着力推进科研工作深入全面发展等五个方面对江苏省本科教育的“保持整体繁荣，促进协同发展”提出了科学的参考意见。

## 一：问题的重述

### 1.1 问题背景

江苏省作为全国的教育大省，拥有本科高校 52 所，在校大学生 190 余万人，十多所高校入选中国大学综合实力排行榜百强，与北京市并列全国第一。经济的快速发展要求高教强省建设的持续推进，在此背景下，省委、省政府出台了《关于深入推进教育现代化建设努力办好人民满意教育的意见》，明确提出全省教育事业改革将在 2020 年实现“一达到、两提高”的总目标。

在全省高等教育总体繁荣的背景下，江苏省仍存在着高等教育资源地区发展极度不平衡的现状。南京市作为江苏省省会，独揽省内半数高校，其掌握的教育资源远超其他地级市，甚至占据全省资源的半壁江山。高等学校是人才输送的通道，也是经济发展的活力源泉。高校资源的分布严重不均就会导致地区经济发展的不平衡，其中南京、苏州、无锡作为苏中、苏南的教育大市控制了江苏省绝大多数的高校资源，赢得了经济发展的便利条件，而苏北地区只有徐州是唯一拥有四所本科高校的城市，省内南北之间的高等教育资源差距直接导致了其经济发展的差距。因此，推进全省本科教育的均衡发展是江苏省提高高等教育质量的最重要的途径之一。

### 1.2 提出问题

问题 1. 影响本科教育质量的指标有很多。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，并进行相关的数据分析。

问题 2. 根据问题 1 中的指标数据，建立数学模型，对江苏省 13 个地级市（常州、淮安、连云港、南京、南通、宿迁、苏州、泰州、无锡、徐州、盐城、扬州、镇江）的本科教育质量进行综合评价。

问题 3. 通过建立数学模型，分析在所给的 9 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。

问题 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江苏省 13 个地级市本科教育发展的差异，并利用查找的数据资料验证所建立模型的有效性。

问题 5. 请结合前面的讨论给出有效提升江苏省本科教育质量的政策建议。

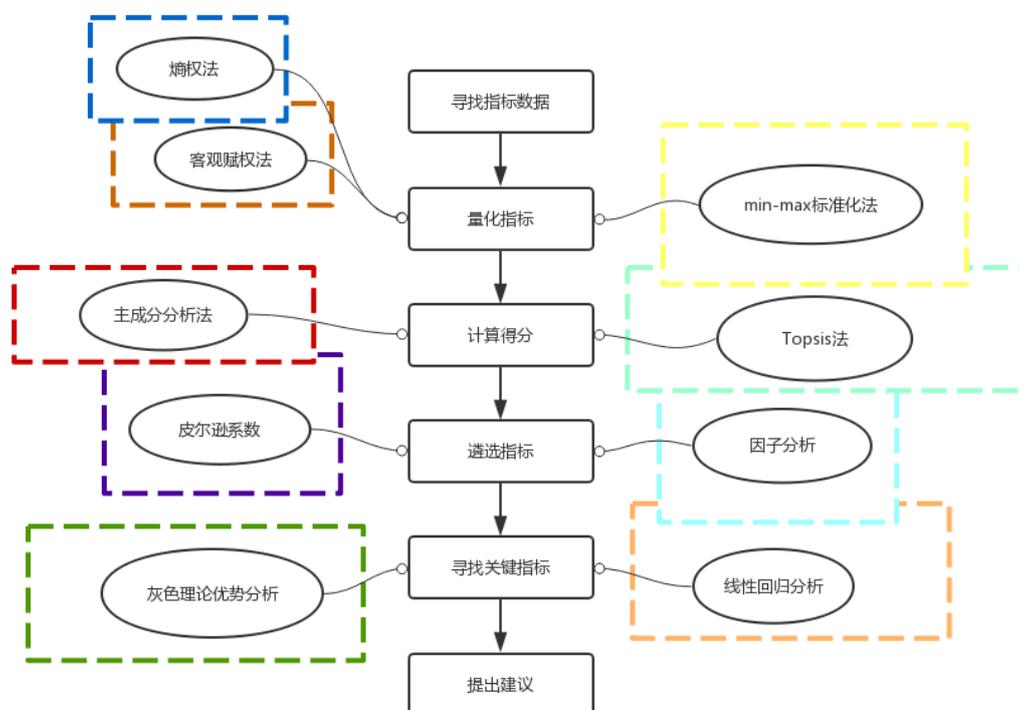
## 二：问题的假设

1. 找到的数据准确，可以反映事实；
2. 查找到的资料文件没有个别案例，均可反映事实；
3. 短期内指标数据不发生巨大变化；

## 三：符号说明

符号	符号意义
$W$	权重
$P$	指标标准化后的数据
$S$	量化结果
$C$	相对接近度
$F$	主成分得分
$Z$	综合得分

## 问题总分析流程图



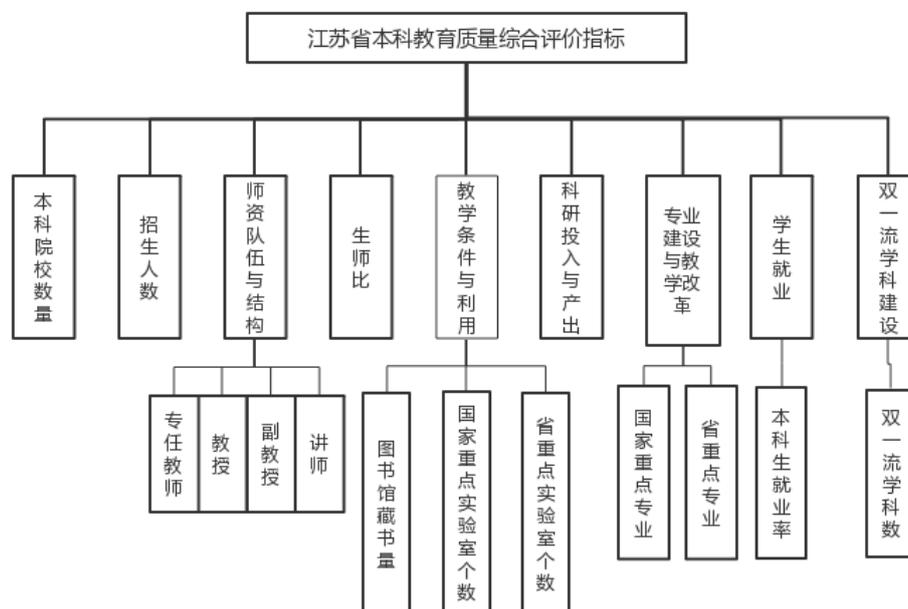
## 四：问题一：量化九个标准指标

### 4.1 问题的分析

面对问题一给的 9 个指标，想要进行量化处理，首先，我们在各学校官网及《2015-2016 学年江苏省及 49 所本科院校质量报告》找到相关准确数据（具体数据见附录三），其中，一些泛指类指标，比如教学条件与利用、专业建设与教学改革，我们查阅相关资料，将其用建立三级指标体系的方法表示出来，面对不同三级指标的数量，我们按其复杂度将二级指标分为**第 I 类**、**第 II 类**、**第 III 类**，然后我们采用不同的量化方法，有熵权法、主观赋权法、*min-max* 标准化法，这样就能更加准确的对指标进行量化。

## 4.2 江苏省本科教育质量综合评价体系

查阅相关资料，建立如图一所示的江苏省本科教育质量综合评价三级指标体系。



图一：江苏省本科教育质量综合评价指标

由图一，结合三级指标复杂程度，将本科院校数量、招生人数、生师比、双一流学科建设、科研投入与产出、学生就业归为第 I 类，将专业建设与教学改革、教学条件与利用归为第 II 类，将师资队伍与结构归为第 III 类。

## 4.3 基于熵权法的第 III 类二级指标量化

### 4.3.1 基于熵权法的分析

熵表示一种能量在空间中的均匀程度。在实际应用在系统论中，熵越大说明系统越混乱，携带的信息越少；熵越小说明系统越有序决定携带的信息越多。熵值法是一种客观的赋权方法，它通过计算指标的信息熵，根据指标的相对变化程度对系统的整体的影响来决定指标的权重，相对变化程度大的指标具有较大的权重，相对变化程度大的指标具有较大的权重。子指标数已超过 4 个，正适用于师资队伍与结构。

### 4.3.2 熵权法原理与步骤

**Step1:** 数据标准化处理。假定评价指标体系包括  $n$  个指标，由  $m$  个样本组成。由于各指标的量纲，数量级均有差异，为消除量纲不同对评价结果的影响，需要对各指标进行标准化处理，本文采用极值法对数据进行处理：

$$\text{当为正向指标时，或为阈值指标，} x_{\min} \leq x_i \text{ 且 } x'_{ij} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}}$$

$$\text{当为负向指标时，或为阈值指标，} x_{\max} \geq x_i \text{ 且 } x'_{ij} = \frac{x_{\max} - x_{ij}}{x_{\max} - x_{\min}}$$

式中， $x_{ij}$  为第  $j$  项指标值， $x_{\max}$  为第  $j$  项指标的最大值， $x_{\min}$  为第  $j$  项指标的

最小值  $x_j$  为第  $j$  项指标的阈值, 为标准化值; 当  $x_{ij} = 0$  时,  $x_{ij} = x_j + \Delta$  ( $\Delta$  为极小数)

**Step2:** 计算  $x_{ij}$  在第  $j$  项指标的比重:

$$S_{ij} = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}} (0 \leq S_{ij} \leq 1)$$

**Step3:** 计算第  $j$  项指标的信息熵值:

$$e_j = -K \sum_{i=1}^m s_{ij} \ln s_{ij}$$

式中,  $K$  为常数,  $K = \frac{1}{\ln n}$

**Step4:** 计算第  $j$  项指标的信息效用度

$$d_j = 1 - e_j$$

指标的信息效用的价值取决于该指标的信息熵值与 1 之间的差值, 它的值直接影响权重的大小, 信息效用值越大, 对评价的重要性就越大, 权重也就越大。

**Step5:** 计算第  $j$  项评价指标权重。

$$w_j = \frac{d_j}{\sum_{i=1}^m d_j}$$

利用熵值法估算各指标的权重, 其本质是利用信息熵值估算各指标的权重。计算指标信息的价值系数, 其价值系数越高; 其价值系数越高, 对评价的重要性就越大。

#### 4.3.3 熵权法结果

将数据导入 *MATLAB* 中, 建立 *shang.m* 文件 (代码见附录一), 用于求解熵权, 由 *MATLAB* 求解得到教授、副教授、讲师、专任教师的权值为下:

$$W_1 = [0.3213, 0.2522, 0.2466, 0.1800]$$

可以发现在重要性上各指标表现为:

教授 > 副教授 > 讲师 > 专任教师

与实际相符。

由公式

$$s_i = \sum_{j=1}^m w_j \cdot p_{ij}$$

得出各地级市师资队伍与结构量化结果为

表 1: 师资队伍与结构量化结果

师资队伍与结构	量化结果	淮安市	0.16599802
---------	------	-----	------------

南京市	3.105724588	连云港	0.115059749
徐州市	0.734311766	盐城市	0.147620032
苏州市	0.411043761	宿迁市	0.029099476
镇江市	0.247479082	泰州市	0.196385742
南通市	0.17304039	无锡市	0.260939564
常州市	0.237440717	扬州市	0.20272163

#### 4.4 基于客观赋权法的第Ⅱ类二级指标量化

客观赋权法是综合评价的一种手段，根据原始数据之间的关系通过一定的数学方法来确定权重，其判断结果不依赖于人的主观判断，有较强的数学理论依据。

##### 4.4.1 客观赋权法的实现

参考资料《中国医学院院科学研究排行榜》（详情见附录二）可以得出：国家重点学科与省级重点学科的权重分别是

$$W_2 = [0.7403, 0.2596]$$

国家重点实验室数与省重点实验室数的权重分别是

$$W_3 = [0.8333, 0.1667]$$

图书馆藏书量即与这两类同等重要，即近似为 0.3333。

由公式

$$s_i = \sum_{j=1}^m w_j \cdot p_{ij}$$

计算可得出专业建设与教学改革、教学条件与利用量化结果分别为：

表 2：专业建设与教学改革量化结果

专业建设与教学改革	量化结果	淮安市	0.143987974
南京市	2.616971962	连云港	0.093750427
徐州市	0.543791966	盐城市	0.089072778
苏州市	0.292020621	宿迁市	0.034157583
镇江市	0.164063247	泰州市	0.018759958
南通市	0.045559899	无锡市	0.231745438
常州市	0.295755237	扬州市	0.146667968

表 3：教学条件与利用量化结果

教学条件与利用	量化结果	淮安市	0.16599802
南京市	3.105724588	连云港	0.115059749
徐州市	0.734311766	盐城市	0.147620032
苏州市	0.411043761	宿迁市	0.029099476
镇江市	0.247479082	泰州市	0.196385742
南通市	0.17304039	无锡市	0.260939564
常州市	0.237440717	扬州市	0.20272163

#### 4.5 基于 *min-max* 标准化的第Ⅲ类二级指标量化

*min-max* 标准化 (*Min-max normalization*) 也叫离差标准化, 是对原始数据的线性变换, 使结果落到 [0, 1] 区间。其中 *max* 为样本数据的最大值, *min* 为样本数据的最小值。

转换函数如下:

$$x^* = \frac{x - \min}{\max - \min}$$

鉴于生师比为负向指标, 我们在计算量化结果时暂时将生师比转化为师生比, 由此变为正向指标, 使计算更加方便。

将数据导入 *MATLAB* 中, 运用 *min-max* 标准化函数 (代码见附录四), 由 *MATLAB* 求解得到本科院校数量、招生人数、生师比、双一流学科建设、科研投入与产出、学生就业量化结果分别为:

表 4: 本科院校数量量化结果

学校个数	量化结果	淮安市	0.076138699
南京市	0.951733735	连云港	0.038069349
徐州市	0.152277398	盐城市	0.076138699
苏州市	0.152277398	宿迁市	0.038069349
镇江市	0.076138699	泰州市	0.038069349
南通市	0.076138699	无锡市	0.076138699
常州市	0.114208048	扬州市	0.038069349

表 5: 招生人数量化结果

招生人数	量化结果	淮安市	0.317573952
南京市	2.914757247	连云港	0.158786976
徐州市	0.642293318	盐城市	0.325513301
苏州市	0.627295888	宿迁市	0.103211534
镇江市	0.435870249	泰州市	0.055575442
南通市	0.062625583	无锡市	0.200603526
常州市	0.365210045	扬州市	0.19848372

表 6: 师生比量化结果

师生比	量化结果	淮安市	0.100056829
南京市	2.206280262	连云港	0.043351777
徐州市	0.431008272	盐城市	0.110749605
苏州市	0.22242039	宿迁市	0.048790812
镇江市	0.12374729	泰州市	0.057493266
南通市	0.802517692	无锡市	0.282821971
常州市	0.172568665	扬州市	0.085312588

表 7: 双一流学科建设量化结果

双一流学科建设	量化结果	淮安市	0
南京市	1.9728617	连云港	0
徐州市	0.103834826	盐城市	0

苏州市	0.051917413	宿迁市	0
镇江市	0	泰州市	0
南通市	0	无锡市	0.103835
常州市	0	扬州市	0

表 8: 科研投入与产出量化结果

科研投入与产出	量化结果	淮安市	0.069616711
南京市	1.430371763	连云港	0.040205885
徐州市	0.429689306	盐城市	0.058079495
苏州市	0.666549559	宿迁市	0.030069948
镇江市	0.269862178	泰州市	0.037245505
南通市	0.05814764	无锡市	0.381454842
常州市	0.113734954	扬州市	0.349890273

表 9: 学生就业量化结果

学生就业	量化结果	淮安市	0.287192113
南京市	3.417077388	连云港	0.132808841
徐州市	0.57566719	盐城市	0.291424418
苏州市	0.574944602	宿迁市	0.142600654
镇江市	0.28837185	泰州市	0.136067863
南通市	0.286218831	无锡市	0.286734965
常州市	0.433420457	扬州市	0.145137088

#### 4.6 量化处理的结果

通过上述分类、讨论，得出量化处理的方式、计算依据，通过运用 *MATLAB* 计算、查找相关资料，最终得出江苏省十三个地级市的 9 个指标的量化结果。

表 10: 量化处理方式表

序	指标名称	量化处理方式	计算依据	备注
1	本科院校数量	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类
2	招生人数	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类
3	师资队伍与结构	熵权法	熵权法原理 ( <i>MATLAB</i> 实现)	第 III 类
4	生师比	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类
5	教学条件与利用	客观赋权法	参考相关资料 (附录二)	第 II 类
6	专业建设与教学改革	客观赋权法	参考相关资料 (附录二)	第 II 类
7	学生就业	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类
8	科研投入与产出	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类
9	双一流学科建设	<i>min-max</i> 标准化	<i>min-max</i> 标准化原理 ( <i>MATLAB</i> 实现)	第 I 类

表 11: 13 个地级市 9 个指标量化结果

指标 城市	教学条件 与利用	学校个 数	生师比	专业建设与 教学改革	科研投入 与产出	师资队伍 与结构	招生人 数	双一流学 科建设	学生就 业
南京市	1.843	0.952	2.206	2.617	1.430	3.106	2.915	1.973	3.417
徐州市	0.587	0.152	0.431	0.544	0.430	0.734	0.642	0.104	0.576
苏州市	0.441	0.152	0.222	0.292	0.667	0.411	0.627	0.052	0.575
镇江市	0.378	0.076	0.124	0.164	0.270	0.247	0.436	0.000	0.288
南通市	0.131	0.076	0.803	0.046	0.058	0.173	0.063	0.000	0.286
常州市	0.220	0.114	0.173	0.296	0.114	0.237	0.365	0.000	0.433
淮安市	0.138	0.076	0.100	0.144	0.070	0.166	0.318	0.000	0.287
连云港	0.034	0.038	0.043	0.094	0.040	0.115	0.159	0.000	0.133
盐城市	0.155	0.076	0.111	0.089	0.058	0.148	0.326	0.000	0.291
宿迁市	0.078	0.038	0.049	0.034	0.030	0.029	0.103	0.000	0.143
泰州市	0.020	0.038	0.057	0.019	0.037	0.196	0.056	0.000	0.136
无锡市	0.301	0.076	0.283	0.232	0.381	0.261	0.201	0.104	0.287
扬州市	0.163	0.038	0.085	0.147	0.350	0.203	0.198	0.000	0.145

## 五：问题二：十三个地级市综合评价

### 5.1 问题分析

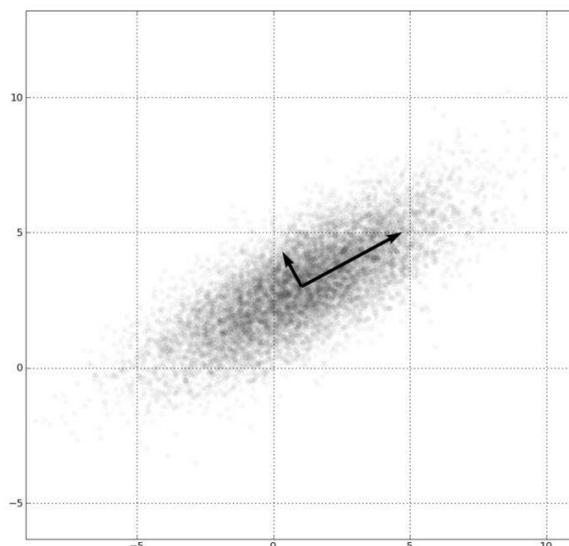
想要对 13 个地级市进行本科教育质量进行综合评价，必须求出 13 个地级市的本科教育质量的综合得分。为了求出本科教育质量的综合得分，我们首先用主成分分析法，得到一些主成分，用这些主成分来对于城市进行计算综合得分，但我们都知道，主成分分析法筛选出的一些主成分代表大部分的特征，而对于一些值都很小的城市，这样误差就会很大，并且这次主成分分析法只提取出一个主成分，误差较大，于是我们又引进了 *TOPSIS* 法（逼近理想解排序法），用来对排名靠后的几个城市进行排名的修正。

### 5.2 基于主成分分析的教育城市排名

计算本科教育质量的综合得分，由于指标过多，主成分分析可以较为准确快速的得出各个城市的本科教育质量的综合得分。

#### 5.2.1 主成分分析

PCA 认为一个随机信号最有用的信息体包含在方差里。为此我们需要找到一个方向  $w_1$ ，使得随机信号  $x$  在该方向上的投影  $w_1^T x$  的方差最大化。接下来，我们在与  $w_1$  正交的空间里到方向  $w_2$ ，使得  $w_2^T x$  的方差最大，以此类推直到找到所有的  $n$  个方向  $w_1, \dots, w_n$ 。用这种方法我们最终可以得到一系列不相关的随机变量： $w_1^T x, \dots, w_n^T x$ 。



【图片来自 wikipedia】

### 5.2.2 主成分分析法原理与步骤

主成分分析也称主分量分析，旨在利用降维的思想，把多指标转化为少数几个综合指标（即主成分），其中每个主成分都能够反映原始变量的大部分信息，且所含信息互不重复。这种方法在引进多方面变量的同时将复杂因素归结为几个主成分，使问题简单化，同时得到的结果更加科学有效的数据信息。在实际问题研究中，为了全面、系统地分析问题，我们必须考虑众多影响因素。这些涉及的因素一般称为指标，在多元统计分析中也称为变量。

**step1.** 计算相关系数矩阵  $R$

相关系数矩阵  $R = (r_{ij})_{m \times m}$

$$r_{ij} = \frac{\sum_{k=1}^n a'_{ki} \cdot a'_{kj}}{n-1}, (i, j = 1, 2, \dots, m)$$

式中  $r_{ij} = 1, r_{ij} = r_{ji}$ ,  $r_{ij}$  是第  $i$  个指标与第  $j$  个指标的相关系数。

**step2.** 计算特征值和特征向量

计算相关系数矩阵  $R$  的特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ ，及对应的特征向量，

$\mu_1, \mu_2, \dots, \mu_m$ 。

其中  $\mu_j = (\mu_{1j}, \mu_{2j}, \dots, \mu_{mj})^T$ ，是由特征向量组成的  $m$  个新的指标变量。

$$y_1 = \mu_{11}\bar{x}_1 + \mu_{21}\bar{x}_2 + \dots + \mu_{m1}\bar{x}_m$$

$$y_2 = \mu_{12}\bar{x}_1 + \mu_{22}\bar{x}_2 + \dots + \mu_{m2}\bar{x}_m$$

...

$$y_m = \mu_{1m}\bar{x}_1 + \mu_{2m}\bar{x}_2 + \dots + \mu_{mm}\bar{x}_m$$

式中  $y_1$  是第一主成分， $y_2$  是第二主成分， $y_m$  是第  $m$  主成分。

**step3.** 特征值  $\lambda_j (j=1, 2, \dots, m)$  的方差贡献率和累积贡献率，称

$$b_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} (j=1, 2, \dots, m)$$

为主成分  $y_j$  的方差贡献率。

$$\alpha = \sum_{j=1}^p b_j$$

为主成分  $y_1, y_2, \dots, y_p$  的累积贡献率。

当特征值小于 1 或者累计贡献率大于 85% 时，则选择前  $p$  个指针变量  $y_1, y_2, \dots, y_p$  作为  $p$  个主成分，代替原来  $m$  个指针变量，从而对  $p$  个主成分进行综合分析。

综合公式如下：

$$M = \sum_{j=1}^p b_j y_j$$

其中  $b_j$  为第  $j$  个主成分的方差贡献率，根据综合的分值就可以进行评价。

### 5.2.3 主成分分析法分析结果

用主成分分析法，把 13 个城市的 9 个指标看为 9 个变量，用 SPSS 分析 9 个变量之间的相关系数，并通过 SPSS 进行数据处理，得出相关系数矩阵，如图 3 所示。

	本科学校数量	招生人数	师生比	学生就业	教学条件与利用	专业建设与改革	双一流学科建设	科研投入与产出	教师结构
相关	1.000	.990	.946	1.000	.973	.994	.991	.905	.992
	.990	1.000	.908	.991	.985	.990	.973	.926	.987
	.946	.908	1.000	.945	.915	.932	.941	.832	.941
	1.000	.991	.945	1.000	.975	.994	.990	.907	.992
	.973	.985	.915	.975	1.000	.981	.954	.952	.982
	.994	.990	.932	.994	.981	1.000	.987	.915	.996
	.991	.973	.941	.990	.954	.987	1.000	.885	.986
	.905	.926	.832	.907	.952	.915	.885	1.000	.918
	.992	.987	.941	.992	.982	.996	.986	.918	1.000

图 3: SPSS 相关矩阵

图 3 为 9 个变量的相关系数矩阵，其中全部数据在 0.8 以上，结果验证分析中的假设，说明它们之间的相关性非常高，如果直接用于分析，可能会带来严重的共线性问题导致结果重叠。

因此利用主成分分析提取主要信息，然后使用提取出的主成分代替原变量进行分析，就可以避开元变量的共线性问题。

通过 SPSS 相关系数矩阵特征向量的方差贡献率，得到图 4

解释的总方差

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	8.665	96.276	96.276	8.665	96.276	96.276
2	.195	2.168	98.445			
3	.088	.973	99.418			
4	.028	.312	99.730			
5	.014	.157	99.887			
6	.006	.062	99.948			
7	.003	.037	99.985			
8	.001	.015	100.000			
9	5.756E-6	6.395E-5	100.000			

提取方法：主成份分析。

图 4：总特征值表

图 4 为整个分析的核心部分：它把主成分按照特征根从大到小的次序排列，可见第一个主成分的特征根为 8.665，它解释了数据总变异的 96.276%，这说明该成分的解释力足够大。这六个变量只需要提取第一个变量作为主成分即可。

另一方面，下图给出了特征根的碎石图：

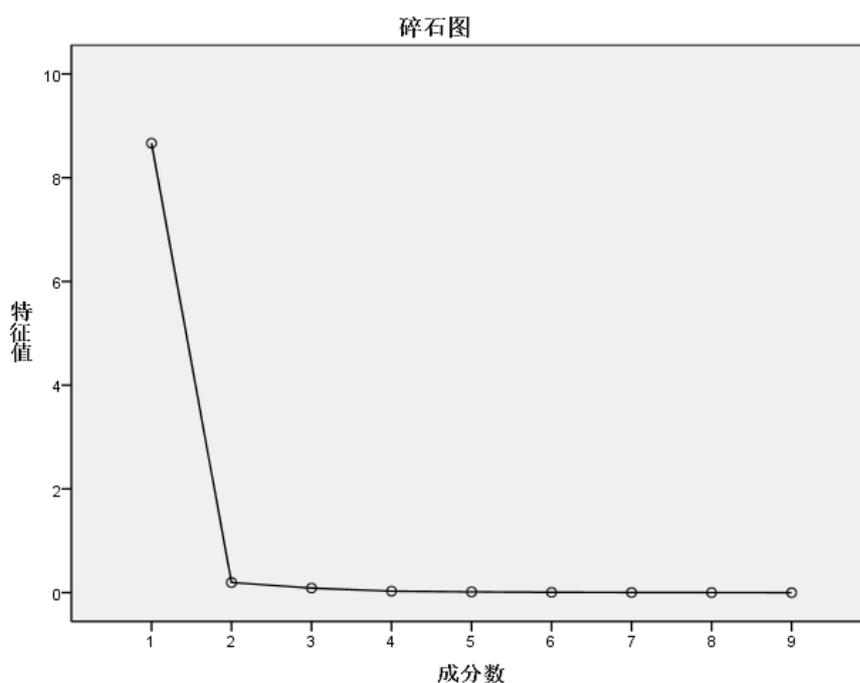


图 5：碎石图

该图中，具有较强解释能力的因子一般在图中表现为较大的斜率，从上图可以看出第一个因子在非常陡峭的斜率上，后面的因子斜率开始变得平缓，因此选择第一个因子为主成分。

为了确定每个变量的线性组合系数，还要计算主成分对应的变量特征值，结果如下图：

成份得分系数矩阵

	成份
	1
本科学校数量	.115
招生人数	.114
师生比	.109
学生就业	.115
教学条件与利用	.114
专业建设与改革	.115
双一流学科建设	.114
科研投入与产出	.108
教师结构	.115

提取方法:主成份。  
构成得分。

图 6: 特征向量图

根据图 5 可以得到主成分得表达式

$$F = 0.115x_1 + 0.114x_2 + 0.109x_3 + 0.114x_4 + 0.115x_5 + 0.115x_6 + 0.114x_7 + 0.108x_8 + 0.115x_9$$

其中 $x$ 为各量标准化数据。

以第一个主成分的贡献率为权重，构建主成分综合评价模型

$$Z = 0.9628F$$

可以得到各地级市本科教育质量综合评价得分以及排序结果如下：

表 12: 主成分分析法综合得分表

城市	得分	淮安市	-0.37
南京市	3.25	连云港	-0.49
徐州市	0.19	盐城市	-0.38
苏州市	0.08	宿迁市	-0.51
镇江市	-0.22	泰州市	-0.51
南通市	-0.3	无锡市	-0.18
常州市	-0.24	扬州市	-0.34

相应的，我们得到综合评价得分图，更直观的可以看出本科教育质量得分情况。



图 7: 综合评价得分图

### 5.3 基于 TOPSIS 本科教育综合质量评分改进

#### 5.3.1 TOPSIS 法基本原理

TOPSIS 法是 *Technique for Order Preference by Similarity to Ideal Solution* 的缩写, 即逼近于理想解的技术, 它是一种多目标决策方法。方法的基本思路是定义决策问题的理想解和负理想解, 然后在可行方案中找到一个方案, 使其距理想解的距离最近, 而距负理想解的距离最远。

#### 5.3.2 距离的测度

采用相对接近测度。设决策问题有  $m$  个目标  $f_j (j = 1, 2, \dots, m)$ ,  $n$  个可行解  $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{im}), i = (1, 2, \dots, n)$ ; 并设该问题的规范化加权目标的理想解是  $Z^*$ , 其中

$$Z^+ = (Z_1^+, Z_2^+, \dots, Z_m^+),$$

那么用欧几里得范数作为距离的测度, 则从任意可行解  $Z_i$  到  $Z^+$  的距离为:

$$S_i^+ = \sqrt{\sum_{j=1}^m (Z_{ij} - Z_j^+)^2} \quad i=1, \dots, n, \quad (8.1)$$

式中,  $Z_{ij}$  为第  $j$  个目标对第  $i$  个方案 (解) 的规范化加权值。

同理, 设  $Z^- = (Z_1^-, Z_2^-, \dots, Z_m^-)^T$  为问题的规范化加权目标的负理想解, 则任意可行解  $Z_i$  到负理想解  $Z^-$  之间的距离为:

$$S_i^- = \sqrt{\sum_{j=1}^m (Z_{ij} - Z_j^-)^2} \quad i=1, \dots, n, \quad (8.2)$$

那么, 某一可行解对于理想解的相对接近度定义为:

$$C_i = \frac{S_i^-}{S_i^- + S_i^+} \quad 0 \leq C_i \leq 1, i=1, \dots, n, \quad (8.3)$$

于是, 若  $Z_i$  是理想解, 则相应的  $C_i=1$ ; 若  $Z_i$  是负理想解, 则相应的  $C_i=0$ 。 $Z_i$  愈靠近理想解,  $C_i$  愈接近于 1; 反之, 愈接近负理想解,  $C_i$  愈接近于 0。那么, 可以对  $C_i$  进行排队, 以求出满意解。

#### 5.3.3 TOPSIS 法计算步骤

第一步: 设某一决策问题, 其决策矩阵为  $A$ . 由  $A$  可以构成规范化的决策矩阵  $Z'$ , 其元素为  $Z'_{ij}$ , 且有

$$Z'_{ij} = \frac{f_{ij}}{\sqrt{\sum_{i=1}^n f_{ij}^2}} \quad i = (1, 2, \dots, n), j = (1, 2, \dots, m) \quad (8.4)$$

式中,  $f_{ij}$  由决策矩阵给出。

$$A = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \vdots & \vdots & \dots & \vdots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{bmatrix} \quad (8.5)$$

第二步: 构造规范化的加权决策矩阵  $Z$ , 其元素  $Z_{ij}$

$$Z_{ij} = W_j Z'_{ij} \quad i=1, \dots, n; \quad j=1, \dots, m \quad (8.6)$$

$W_j$  为第  $j$  个目标的权。

第三步： 确定理想解和负理想解。如果决策矩阵  $Z$  中元素  $Z_{ij}$  值越大表示方案越好，则

$$Z^+ = (Z_1^+, Z_2^+, \dots, Z_m^+) = \{ \max_i Z_{ij} | j = 1, 2, \dots, m \} \quad (8.7)$$

$$Z^- = (Z_1^-, Z_2^-, \dots, Z_m^-) = \{ \min_i Z_{ij} | j = 1, 2, \dots, m \} \quad (8.8)$$

第四步： 计算每个方案到理想点的距离  $S_i$  和到负理想点的距离  $S^-_i$ 。

第五步： 按式(8.3)计算  $C_i$ ，并按每个方案的相对接近度  $C_i$  的大小排序，找出满意解。

多目标综合评价排序的方法较多，各有其应用价值。在诸多的评价方法中，*TOPSIS* 法对原始数据的信息利用最为充分，其结果能精确的反映各评价方案之间的差距，*TOPSIS* 对数据分布及样本含量，指标多少没有严格的限制，数据计算亦简单易行。不仅适合小样本资料，也适用于多评价对象、多指标的大样本资料。利用 *TOPSIS* 法进行综合评价，可得出良好的可比性评价排序结果。

### 5.3.4 TOPSIS 法计算结果

将数据带人 *MATLAB* 中，由 *TOPSIS* 函数（代码见附录五）计算出与最优值的相对接近程度与排序结果：

表 13： 不同城市指标值与最优值的相对接近程度及排序结果

城市	$D^+$	$D^-$	$C_i$	排序结果
南京市	0.045	0.078	0.96542549	1
徐州市	0.034	0.036	0.35486977	2
苏州市	0.065	0.056	0.223103061	3
镇江市	0.069	0.064	0.136466582	5
南通市	0.024	0.089	0.11336352	7
常州市	0.069	0.045	0.130690816	6
淮安市	0.064	0.048	0.093148342	9
连云港	0.025	0.041	0.05849375	10
盐城市	0.046	0.065	0.090260459	11
宿迁市	0.084	0.061	0.052717985	12
泰州市	0.048	0.054	0.052717985	9
无锡市	0.036	0.075	0.148018112	4
扬州市	0.047	0.045	0.10181199	8

由表可知连云港市、宿迁市、泰州市排名发生变化，也正好验证开头分析所说，一些数据较少的城市，它们可能会有较大的误差。

通过 *TOPSIS* 法改进之后的新排名为下表所示：

表 14： *TOPSIS* 法改进后的本科教育质量排名

城市	排名	淮安市	9
南京市	1	连云港	10
徐州市	2	盐城市	11
苏州市	3	宿迁市	12
镇江市	5	泰州市	9
南通市	7	无锡市	4
常州市	6	扬州市	8

## 六：问题三：筛选指标

### 6.1 问题分析

对于问题三的要求筛选数据，我们可以根据问题二中的主成分分析法的结论，也就是因子分析，得到其成分矩阵和相关系数矩阵，再利用 SPSS 得出皮尔逊相关系数进行二次筛选。

### 6.2 基于因子分析的指标筛选

问题二中的主成分分析法具有很好的降维处理技术，能将多个指标化为几个不相关的综合因子，并且综合因子变量能反映原始指标变量的绝大部分信息，能较好地解决多指标遴选的问题。通过主成分分析得出若干主成分，筛选掉主成分载荷绝对值小于 0.1 的指标。求出剩下指标的皮尔逊相关系数再筛选。而这就是因子分析的步骤。相关步骤与问题二相同。

### 6.3 基于因子分析求解步骤

用主成分分析法，把气象相关的 9 个指标变为六个变量，用 SPSS 分析 9 个变量之间的相关系数，并通过 SPSS 进行数据处理，得出其相关系数矩阵，如上图 3 所示。

下图为各指标成分矩阵

	成份
	1
本科学校数量	.115
招生人数	.114
师生比	.109
学生就业	.115
教学条件与利用	.114
专业建设与改革	.115
双一流学科建设	.114
科研投入与产出	.108
教师结构	.115

提取方法：主成份。  
构成得分。

图 6：特征向量图

该图在上文中出现过，从上图可以看出，每个指标的主成分载荷都高于 0.1，但都没有超过 0.2，也就是说这些指标都占了较大比重，但又不是很大，下面再进一步进行遴选。

任意两指标间的相关系数还不足以反映这两个指标变量之间实际的相关性，所以对它们进行相关性的统计检验。其置信水平为 0.05 的双尾检验的临界值 0.707，即若样本相关系数绝对值超过 0.707，便可认为这两个指标变量之间存在显著的线性相关，否则，就不相关。当两指标建相关系数大于 0.707 时我们视作高度相关，若关于指标意义相近，则根据重要性删减为一个，若意义相差较大，

则都保留。

下图为皮尔逊相关系数表

		相关性				
		本科学校数量	招生人数	师生比	教学条件与利用	学生就业
本科学校数量	Pearson 相关性	1	.990 <sup>**</sup>	.946 <sup>**</sup>	.973 <sup>**</sup>	1.000 <sup>**</sup>
	显著性 (双侧)		.000	.000	.000	.000
	N	13	13	13	13	13
招生人数	Pearson 相关性	.990 <sup>**</sup>	1	.908 <sup>**</sup>	.985 <sup>**</sup>	.991 <sup>**</sup>
	显著性 (双侧)	.000		.000	.000	.000
	N	13	13	13	13	13
师生比	Pearson 相关性	.946 <sup>**</sup>	.908 <sup>**</sup>	1	.915 <sup>**</sup>	.945 <sup>**</sup>
	显著性 (双侧)	.000	.000		.000	.000
	N	13	13	13	13	13
教学条件与利用	Pearson 相关性	.973 <sup>**</sup>	.985 <sup>**</sup>	.915 <sup>**</sup>	1	.975 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000		.000
	N	13	13	13	13	13
学生就业	Pearson 相关性	1.000 <sup>**</sup>	.991 <sup>**</sup>	.945 <sup>**</sup>	.975 <sup>**</sup>	1
	显著性 (双侧)	.000	.000	.000	.000	
	N	13	13	13	13	13
专业建设与改革	Pearson 相关性	.994 <sup>**</sup>	.990 <sup>**</sup>	.932 <sup>**</sup>	.981 <sup>**</sup>	.994 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000	.000
	N	13	13	13	13	13
双一流学科建设	Pearson 相关性	.991 <sup>**</sup>	.973 <sup>**</sup>	.941 <sup>**</sup>	.954 <sup>**</sup>	.990 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000	.000
	N	13	13	13	13	13
科研投入与产出	Pearson 相关性	.905 <sup>**</sup>	.926 <sup>**</sup>	.832 <sup>**</sup>	.952 <sup>**</sup>	.907 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000	.000
	N	13	13	13	13	13
教师结构	Pearson 相关性	.992 <sup>**</sup>	.987 <sup>**</sup>	.941 <sup>**</sup>	.982 <sup>**</sup>	.992 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000	.000
	N	13	13	13	13	13

图 7: 皮尔逊相关表 1

		相关性			
		专业建设与改革	双一流学科建设	科研投入与产出	教师结构
本科学校数量	Pearson 相关性	.994 <sup>**</sup>	.991 <sup>**</sup>	.905 <sup>**</sup>	.992 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000
	N	13	13	13	13
招生人数	Pearson 相关性	.990 <sup>**</sup>	.973 <sup>**</sup>	.926 <sup>**</sup>	.987 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000
	N	13	13	13	13
师生比	Pearson 相关性	.932 <sup>**</sup>	.941 <sup>**</sup>	.832 <sup>**</sup>	.941 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000
	N	13	13	13	13
教学条件与利用	Pearson 相关性	.981 <sup>**</sup>	.954 <sup>**</sup>	.952 <sup>**</sup>	.982 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000
	N	13	13	13	13
学生就业	Pearson 相关性	.994 <sup>**</sup>	.990 <sup>**</sup>	.907 <sup>**</sup>	.992 <sup>**</sup>
	显著性 (双侧)	.000	.000	.000	.000
	N	13	13	13	13
专业建设与改革	Pearson 相关性	1	.987 <sup>**</sup>	.915 <sup>**</sup>	.996 <sup>**</sup>
	显著性 (双侧)		.000	.000	.000
	N	13	13	13	13
双一流学科建设	Pearson 相关性	.987 <sup>**</sup>	1	.885 <sup>**</sup>	.986 <sup>**</sup>
	显著性 (双侧)	.000		.000	.000
	N	13	13	13	13
科研投入与产出	Pearson 相关性	.915 <sup>**</sup>	.885 <sup>**</sup>	1	.918 <sup>**</sup>
	显著性 (双侧)	.000	.000		.000
	N	13	13	13	13
教师结构	Pearson 相关性	.996 <sup>**</sup>	.986 <sup>**</sup>	.918 <sup>**</sup>	1
	显著性 (双侧)	.000	.000	.000	
	N	13	13	13	13

图 8: 皮尔逊相关表 2

通过上表与成分得分矩阵综合, 即可遴选出代表性的指标。

考察指标:

1. 本科院校数量和招生人数高度相关, 根据其在第一主成分中的载荷大小, 删去招生人数。
2. 双一流学科建设与学生就业高度相关但意义迥异, 不做删除。

3. 生师比与师资队伍与结构高度相关，根据其在第一主成分中的载荷，删去生师比。
4. 科技投入与产出和专业建设与教学改革高度相关，根据其在第一主成分中的载荷，删去专业建设与改革
5. 双一流学科建设与科研投入与产出高度相关但意义迥异，不做删除。

遴选后的指标体系如下图：

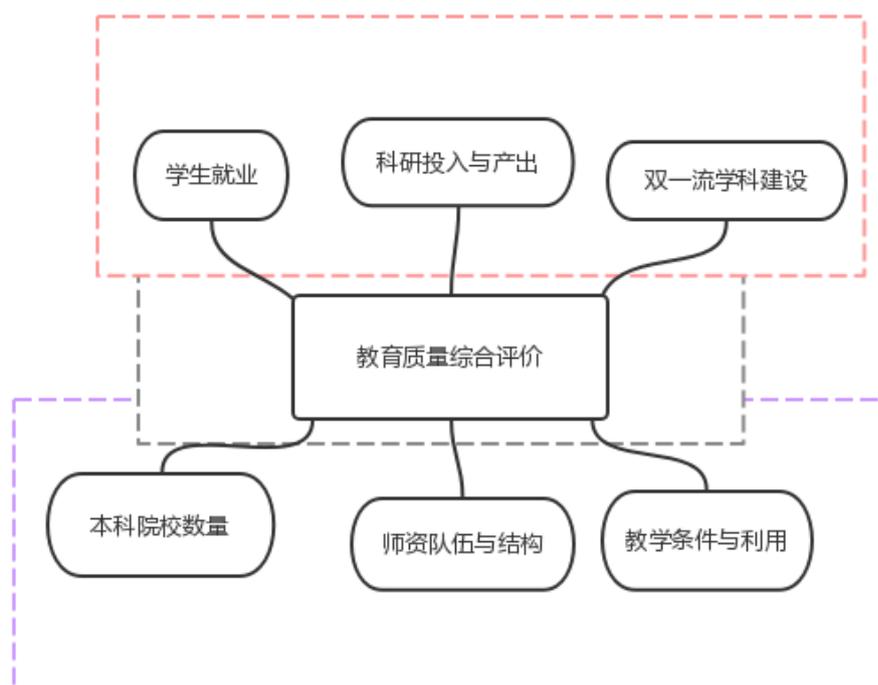


图 9:遴选后指标体系

即通过分析，得出结论：减少招生人数、师生比、专业建设与教学改革三项指标，使评价结果基本一致。

## 七：问题四：关键指标的查找

### 7.1 问题的分析

对于要确定一个关键指标的问题，一个很好的算法就是灰色理论中的优势分析，运用优势分析得到优势子因素之后，运用非线性回归模型中的  $F$  检验和  $t$  检验对优势因子进行检验，即可验证结论。

### 7.2 基于灰色理论优势分析

#### 7.2.1 优势分析步骤

当参考数列不止一个，被比较的因素也不止一个时，则需进行优势分析。假设有  $m$  个参考数列（宜称母因素），记为  $y_1, y_2, \dots, y_m$ ，再假设有  $l$  个比较数列（亦称子因素），记为  $x_1, x_2, \dots, x_l$ 。显然，每一个参考数列对  $l$  个比较数列有  $l$  个关联度，设  $r_{ij}$  表示比较数列  $x_j$  对参考数列  $y_i$  的关联度，可构造关联（度）矩阵  $R = (r_{ij})_{m \times l}$ 。根据矩阵  $R$  的各个元素的大小，可分析判断出哪些因素起主要

影响，哪些因素起次要影响。起主要影响的因素称之为优势因素。再进一步，当某一列元素大于其它列元素时，称此列所对应的子因素为优势子素；若某一行元素均大于其它行元素时，称此行所对应的母元素为优势母元素。

例如，矩阵  $R$  的第 3 列元素大于其它各列元素，

$$r_{i3} > r_{ij}, i = 1, 2, \dots, m; \quad j \neq 3$$

则称  $x_3$  为优势子因素。

### 7.2.2 优势分析结果

将表 11 当中的数据导入到 *MATLAB* 中，运用 *MATLAB* 程序（代码见附录六）程序，计算出各个子因素对母因素的关联度（这里取  $\rho = 0.5$ ），去掉问题四得出的结论之后，得到关联矩阵为

$$R = \begin{bmatrix} 0.802 & 0.761 & 0.557 & 0.810 & 0.936 \\ 0.689 & 0.666 & 0.529 & 0.885 & 0.800 \\ 0.891 & 0.858 & 0.579 & 0.577 & 0.675 \\ 0.678 & 0.663 & 0.568 & 0.780 & 0.731 \\ 0.811 & 0.774 & 0.565 & 0.804 & 0.921 \\ 0.743 & 0.766 & 0.562 & 0.607 & 0.632 \end{bmatrix}$$

从关联矩阵  $R$  可以看出：师资队伍与结构为  $r_{15}=0.936$  最大，表明师资队伍与结构对本科生教育质量影响最大。

由此即遴选出师资队伍与结构为优势因子，即为关键指标。

## 7.3 线性回归分析的检验

### 7.3.1 线性回归分析的 $F$ 检验

因变量  $y$  与自变量  $x_1, \dots, x_m$  之间是否存在如模型所示的线性关系是需要检验的，显然，如果所有的  $|\hat{\beta}_j| (j = 1, \dots, m)$  都很小， $y$  与  $x_1, \dots, x_m$  的线性关系就不明显，所以可令原假设为：

$$H_0: \beta_j = 0 (j = 1, \dots, m)$$

当  $H_0$  成立时由分解式 (14) 定义的  $U, Q$  满足

$$F = \frac{U/m}{Q/(n-m-1)} \sim F(m, n-m-1)$$

在显著性水平  $\alpha$  下由  $1 - \alpha$  分位数被  $F_{1-\alpha}(m, n-m-1)$ ，若  $F < F_{1-\alpha}(m, n-m-1)$ ，接受  $H_0$ ；反之，则拒绝。

### 7.3.2 线性回归分析的 $t$ 检验

当上面的  $H_0$  被拒绝时， $\beta_j$  不全为 0，但是不排除其中若干个为 0 的情况。

所以应进一步做如下  $m$  个检验 ( $j = 1, \dots, m$ ):  $H_0^{(j)}: \beta_j = 0$

由于服从正态分布， $\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$ ， $c_{jj}$  是  $(X^T X)^{-1}$  对角线上的元素，用  $s^2$  代替  $\sigma^2$ ，当  $H_0^{(j)}$  成立时：

$$t_j = \frac{\hat{\beta}_j / \sqrt{c_{jj}}}{\sqrt{Q/(n-m-1)}} \sim t(n-m-1)$$

对给定的 $\alpha$ ，若 $|t_j| < t_{1-\frac{\alpha}{2}}(n - m - 1)$ ，接受 $H_0^{(j)}$ ；反之，则拒绝。

### 7.3.3 线性回归分析检验结果

首先，对于上文遴选出的关键指标师资队伍与结构，若与第二问求出的本科生教育质量评分分别作为横纵坐标，这两项数据所组成的若干点，若与其线性回归分析拟合而成的一次函数通过  $F$  检验和  $t$  检验，即完全可以说明此指标为关键指标。

将数据导入到 SPSS 中（横坐标为师资队伍与结构，纵坐标为综合得分），得到结果如图 10、表 14、表 15、表 16 所示：

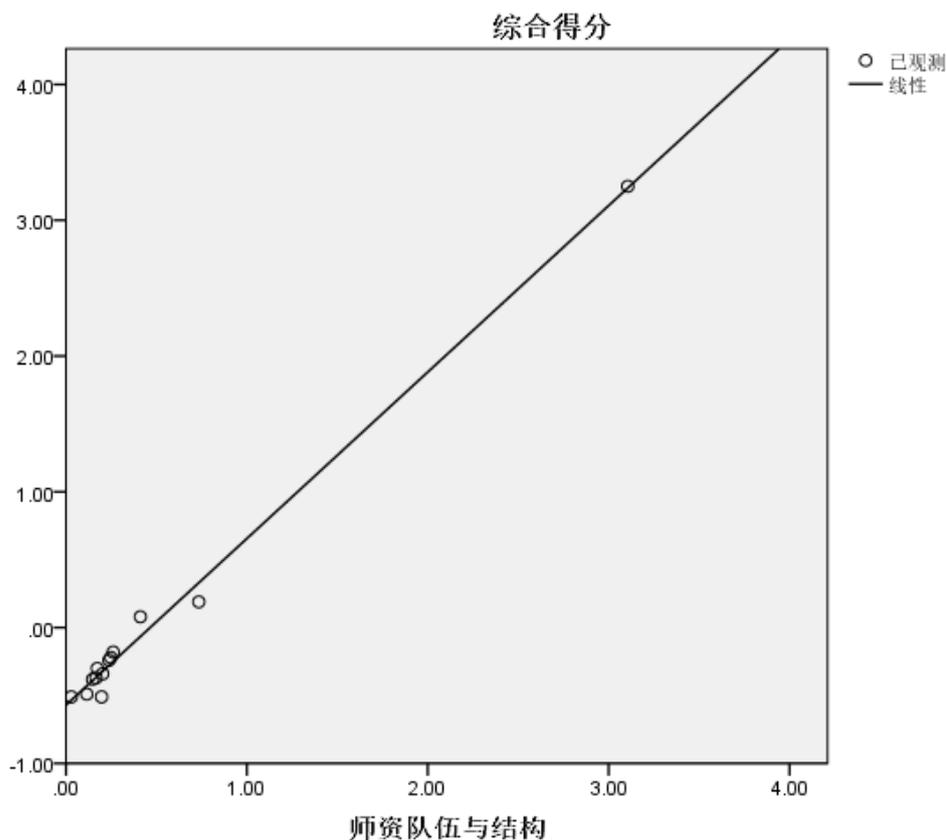


图 10：线性拟合结果图

表 14：模型汇总

模型汇总			
R	R 方	调整 R 方	估计值的标准误差
.996	.992	.992	.091

自变量为师资队伍与结构。

表 15：ANOVA

ANOVA

	平方和	df	均方	F	Sig.
回归	11.900	1	11.900	1445.783	.000
残差	.091	11	.008		
总计	11.991	12			

自变量为师资队伍与结构。

由表可知，模型相关系数  $R$  为 0.996，计算所得  $R^2$  为 0.992，表明模型的拟合度效果较好，通过拟合优度检验；同时统计量  $F$  值为 1445.783， $P$  值（ $sig$ ）为 0.000，小于 0.01，表明回归方程效果显著，通过  $F$  检验。

表 16：回归系数检验

系数

	未标准化系数		标准化系数	t	Sig.
	B	标准误	Beta		
师资队伍与结构	1.226	.032	.996	38.023	.000
(常数)	-.570	.029		-19.472	.000

表中给出了对各个系数的检验结果（ $t$  检验），由表可知  $sig$  均  $<0.05$ ，表明回归系数均有显著意义，通过  $t$  检验。

综合上述理论与方法，我们查询了相关资料（附录七），得出**师资队伍与结构**是本科生教育综合评价的关键数据。

## 八：问题五：政策建议

江苏省作为全国的教育大省，拥有本科高校 52 所，在校大学生 190 余万人，十多所高校入选中国大学综合实力排行榜百强，与北京市并列全国第一。高教强省建设持续推进，省委、省政府出台了《关于深入推进教育现代化建设努力办好人民满意教育的意见》，明确提出全省教育事业改革将在 2020 年实现“一达到、两提高”的总目标。

在总体繁荣的背景下，江苏省高等教育仍存在着地区发展极度不平衡的现状。南京市作为江苏省省会，独揽省内半数高校，其掌握的教育资源远超其他地级市。高等学校是人才输送的通道，也是经济发展的活力源泉。高校资源的分布严重不均就导致了地区经济发展的不平衡，其中南京、苏州、无锡作为苏中、苏南的教育大市控制了江苏省绝大多数的高校资源，徐州是苏北唯一拥有四所本科高校的城市，省内南北之间的高校资源差距直接导致了其经济发展的差距。因此，我们认为“保持整体繁荣，促进协同发展”，是江苏省提升本科教育质量的发展途径之一。

根据 2016 年的统计数据，我们在分析之后提出如下具体政策建议：

**持续推进校内师资队伍与结构的改善。**制定高校合理的教师队伍结构方案，制定教师人才战略，吸引更多高水平教师人才，促进教师队伍结构的改善。同时，基于“教师-课程”质量监控体系，对近 5 年内学生课堂测评分持续较低的课程进行预警。筛选 2010-2014 学年的学生测评分数据，对测评分相对较低的课程进行跟踪监控，安排督导员听课帮扶，并由教学名师、教学督导先期对其中 45 岁以下的青年教师进行个性化的教学帮扶与指导。以先进帮后进，着力提升高校整

体师资水平。

**探索高校之间教师人才流动方案。**地区之间经济、社会发展存在不平衡现象，不同地区的高校对人才的吸引强度大不相同。因此，可以制定不同地区高校一对一或多对多教师人才流动战略，利用高回报率吸引高水平教师在高校之间形成流动教学和经验分享。以此带动高等教育落后地区的快速发展。

**杜绝盲目扩招，制定人才培养模式。**近年来，高等学校招生数在逐年上升，这对高校的基础设施和教学资源产生了较大的压力，过分扩招容易稀释教学资源，持续缩小师生比，不利于人才的培养和高校的长远发展。因此，高校应根据自身情况，制定相应的人才培养模式，招收适合学校规模的学生数量，保证教育质量，提高人才素质。同时，给予落后地区更多的生源，保证其有足够的发展动力。

**着力推进科研工作深入全面发展。**科学研究成果是推动社会前进的强大动力。要针对社会需求进行科学的、审慎的分析和判断，妥善处理好四大关系：一是处理好国家需要和公众需要的关系；二是处理好经济需要和社会需要的关系；三是处理好全局需要和区域需要的关系；四是处理好近期需要和长远需要的关系。一方面，积极引导研究者关注社会发展需要。不能单纯为了评职称、发论文、完成规定的工作量，而要切实解决实际问题。力求通过科研，回答理论问题、实践问题、学术问题、专业问题、社会问题以及公众的生活问题等。另一方面，高校科研要积极引导研究者注重科研成果的转化。避免课题验收之后、论文发表之后、成果获奖之后，或束之高阁、或打入冷宫，而要真正使科研成果转化成推动社会进步的强大动力。

**大力提升教学条件，利用资源充分发展。**一般说来，高等教育的增长速度应不高于或略低于国民经济增长的速度。但实际上，高等教育的增长速度实际上已远远超过了国民经济的增长速度，中央财政、地方财政单一的投资机制已无法满足要求。应鼓励高校走多元化投资的道路，提倡和鼓励企业及经济实体与财团投资教育的发展思路。高校可以利用自身资源优势通过合作、服务或开展捐资、集资办学及运用金融、信贷等方式筹措资金，发展教学基础条件建设。重点提升图书馆、实验室等科研基础设施建设，为高校的持续发展打好坚实的基础。

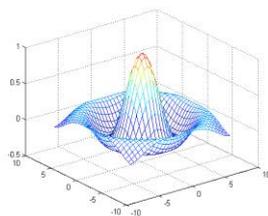
## 九：参考文献

- [1]程启月，评测指标权重确定的结构熵权法，北京，中国国防大学信息作战与指挥训练教研部.2010.07
- [2]李艳双.曾珍香.张闽.于树江，主成分分析法在多指标综合评价方法中的应用，石家庄，河北工业大学管理学院.1999.01
- [3]胡永宏.对 TOPSIS 法用于综合评价的改进,西安，西安统计学院统计学系.2002.04
- [4]刘军.史东林.刘蕾，国内外全能运动成绩灰色理论“优势分析”的比较研究，济南.石家庄，山东体育学院.河北体育学院山东体育学院，2006.03

## 十：附录

附录一：*shang.m* 求熵权的 *MATLAB* 文件  
*function [s,w]=shang(x,ind)*

# 五一数学建模竞赛



## 题目：——江苏省本科教育质量综合评价——

关键词：熵权法 主成分分析法 虚拟控制调节法 灰色关联分析 敏感度分析

### 摘要：

本文运用熵权法、主成分分析法、虚拟控制调节法、灰色关联分析等方法，建立了江苏省 13 个地级市本科教育质量的综合评价模型，并对该模型中的指标进一步优化与筛选，对江苏省的本科教育质量问题进行了研究。

问题一中，本文基于给定的九个指标，参考现存的多种评价体系与原则，综合考虑分析后完成一套完备的四级指标评价体系，并基于熵权法构建评估本科教育质量的综合评价模型，对给定的九个指标进行了合理的量化。

问题二中，本文综合考虑指标的客观与主观因素，利用熵权法和主成分分析法互为验证，得到更有效的综合评价模型。基于问题一各三级指标的量化值，进行无量纲化处理。根据处理后的无量纲值，先利用主成分分析，确定各成分的信息贡献率和累积贡献率，利用其信息贡献率得到江苏省 13 个城市本科质量的综合评价模型。本文再利用熵权法，对模型中各指标的权重进行求解赋值，并建立最终的江苏省 13 个地级市本科教育质量的综合评价模型。对于两种不同的模型，本文分别计算得到 13 个城市的本科质量综合得分，得到基于这两个模型的江苏省 13 个地级市的本科教育质量排名。对比分析两种模型得到的排名，最后得到的结果为：南京>苏州>无锡>南通>镇江>徐州>扬州>盐城>淮安>常州>连云港市>泰州>宿迁。

问题三中，引入虚拟控制城市，通过控制 13 个城市三级指标数据进行相同幅度的调节，计算调节前后 13 个城市的本科教育质量综合评分间的相对欧式距离。本文利用灰色关联分析，验证这一方法的可靠，得到各三级指标与本科教育质量的关联度。本科院校数量、招生人数这两项指标的相对欧式距离与关联度均较小，说明这两项指标对本科教育质量的影响较小。减少这两项指标后再带入建立的综合评价模型的排名变化浮动较小，因此可以减少本科院校数量、招生人数这两项三级指标。

问题四中，利用敏感度分析，选择其余 7 个指标为主要三级指标，每次仅改变一个指标，与其他未改变的组成一个新指标集，重新计算其综合得分，得到敏感度。将敏感度最大的专业建设与改革作为关键指标，比较仅改变该值与未改变该值的江苏省 13 个地级市本科教育质量的得分，发现各市评分差距均明显降低，且得分越低的城市的增量明显越高，该指标确实可以缩小江苏省 13 个地级市本科教育发展的差异。

问题五中，根据前四问的综合评价模型分析出江苏省各地级市本科教育质量不均衡的原因，并针对现实情况，结合当下政策，提出了合理有效的建议。

## 一、问题重述

随着中国的改革开放，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，江苏省的本科教育发展在全国名列前茅，而江苏省 13 个地级市的本科教育质量发展并不平衡。

- 问题 1：影响本科教育质量的指标有很多。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（完成下表），并进行相关数据分析。
- 问题 2：根据问题 1 中的指标数据，建立数学模型，对江苏省 13 个地级市（常州、淮安、连云港、南京、南通、宿迁、苏州、泰州、无锡、徐州、盐城、扬州、镇江）的本科教育质量进行综合评价。
- 问题 3：通过建立数学模型，分析在所给的 9 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。
- 问题 4：通过建立数学模型，确定一个关键指标，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，并利用查找的数据资料验证所建立模型的有效性。
- 问题 5：请结合前面的讨论给出有效提升江苏省本科教育质量的政策建议。

## 二、问题分析

江苏省位于长江三角洲经济发达地区，综合实力强劲，高等教育资源优渥，但仍存在着教育发展不平衡的现象，省内十三个地级市的高等教育资源分配差异较为显著。对江苏省本科教育质量进行综合评估，可以体现出江苏省高等教育所存在的缺陷与问题，利于教育改革的推进和各高校的健康发展。因此，进行江苏省本科教育质量综合评估具有十分重要的意义。

### 2.1 关于问题一的分析

问题一要求对给出的九个指标进行量化处理，并进行相关的数据分析。由于江苏省本科教育质量受多种复杂因素作用的影响，且在选取评价指标时需要考虑指标的可行性、指标选取的主观性等因素的影响，因此可以基于熵权法进行客观赋权，建立一个多层次的江苏省本科教育质量评价指标体系。

### 2.2 关于问题二的分析

问题二要求对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。考虑到指标针对这十三个地级市的可行性，将收集到的数据标准化后用熵权法计算各指标权重，最后代入第一问的评价体系即可求出综合评分进行排名。

### 2.2 关于问题三的分析

问题三要求在所给的九个指标中，可以减少那些指标使评价结果与问题二中的评价结果基本一致。可以考虑根据分析各项指标的灵敏性来判断指标对目标层影响的大小。为了能更直观地调控各指标的变量，且使各变量值对该指标的影响是等效

的，本文考虑插入一个虚拟控制变量。通过调节改虚拟控制变量的各指标数据，计算各指标的灵敏度，从而确定出各指标变化对目标层的影响效果。

## 2.2 关于问题四的分析

问题四要求确定一个关键的指标，使得该指标值得普遍该晒能够尽可能缩小江苏省 13 个地级市本科教育发展差异。可以考虑敏感度分析，任意改变主要三级指标的综合得分，对比改变前后 13 个地级市本科教学质量的得分，来判断该关键指标是否符合要求。

## 2.2 关于问题五的分析

问题五要求我们结合前面的讨论，给出有效提升江苏省本科教育质量的政策建议。考虑根据问题一至问题四中获得的所有可能结果进行多方面地分析，采取重点对策来提出了对有效提升江苏省本科教育质量的科学且有依据的政策建议。

## 三、问题假设

1. 假设网上搜集的数据真实可靠，不存在异常数据。
2. 假设各高校自然发展，没有特殊情况导致其综合实力剧变。
3. 假设不考虑民办与中外合办本科院校时，不影响教育综合质量的评价。
4. 假设对某些指标采取的调整措施不会影响其他指标

## 四、符号说明

序号	符号	符号说明
1	$L'_{\max}$	指标中最大值
2	$L'_{\min}$	指标中最小值
3	$w_i$	第 $i$ 项指标的熵权
4	$Z_t$	第 $t$ 个地级市的综合得分

## 五、数据的预处理

本文通过教育部高等教育司、江苏省教育厅、江苏省各大高等本科院校等权威机构的官方网站对需要使用到的数据进行了检索与收集。数据涵盖方面十分广泛，存在部分原始数据缺失的情况。此外，某些模型的建立与求解需要进行数据标准化处理。此部分即给出对原始数据的预处理过程。

### 5.1 缺失数据的处理

原始数据收集过程中，存在数据缺失现象，主要由两种情况导致。一是单个独

立的数据无从查得，二是统计方法的精确度不足，导致某年的相关数据具体值无法求得。这两种情况皆可使用均值替换法对缺失数据进行处理。

### 5.1.1 单个独立数据无法查得的情况

将项目除去缺失项以外的数据求平均值来替换缺失数据。缺失值计算公式为：

$$X'_{ij} = \frac{1}{n} \left[ \sum_{i=1, i \neq j}^n X_{ij} \right]$$

其中， $X'_{ij}$  为第  $i$  个三级指标下的第  $j$  个四级指标的缺失值。

以 2017 年徐州市各本科院校师资队伍中教授职称的人数数量为例进行修补。徐州医科大学的教授人数无法查得，修补后得到的数据如下表所示。

表 5-1 2017 年徐州市各本科院校师资队伍中教授职称的人数

高校名称	中国矿业大学	江苏师范大学	徐州医科大学	徐州工程学院
修补前数据	410	227	数据缺失	92
修补后数据	410	227	243	92

### 5.1.2 统计方法精确度不足的情况

用给出数据的总数除以年数求平均值来替换缺失数据。缺失值计算公式为：

$$X'_{ij} = N \cdot \frac{1}{n} \quad (5-1)$$

其中， $X'_{ij}$  为第  $i$  个三级指标下的第  $j$  个四级指标的缺失值， $N$  为数据总数， $n$  为总年数。

以 2017 年各高校 SCI/EI 论文发表数量为例进行修补。宿迁学院官网信息中相关描述为“自 2015 年以来，我校被 SCI/EI 收录的论文篇目数量为 120 篇”。因而可得宿迁学院 2017 年的 SCI/EI 收录篇目为 40 篇。

## 5.2 数据的无量纲化处理

各个评价指标之间度量单位一般不同，为了消除不同变量量纲的影响，需要首先进行无量纲化处理。无量纲化处理方法多种，如极值化、标准化、均值化等，本文中均采用极值化处理。公式如下：

$$L'_i = \frac{L_i - L_{i \min}}{L_{i \max} - L_{i \min}} \quad (5-2)$$

其中， $L'_i$  为各指标无量纲化后的数据， $L_{i \max}$ 、 $L_{i \min}$  分别为该指标中的极大、极小值。

即可对数据无量纲化至区间[0,1]之内。

### 5.3 数据的一致化处理

给出的所有指标中，存在“极大型”和“极小型”两种类型的数据。为了方便对这两种类型的数据同时进行评价和预测，需要对数据进行一致化处理。本

文将所有数据处理成“极大型”数据，即值越大本科教学质量越高。

本文中存在的极小型数据指标为：生师比  $C_4$ ，对于“极小型”数据的处理公式如下：

$$g_i' = \max(C_i) - C_i \quad (5-3)$$

其中， $\max(C_i)$  为第  $i$  个指标中的极大值。

对于“极大型”数据： $g_i = C_i$

## 六、问题一的模型建立与求解

### 6.1 本科教育质量的观念

随着社会的进步与发展，国家对高等素质人才的需求大量增加，本科教育受到越来越多的重视。近年来出现大量本科院校，而本科教育的质量是重中之重，其在服务国家战略、引领经济社会发展、分层分类培养多样化人才等方面发挥越来越积极的作用。

### 6.2 本科教育质量评估指标体系的建立

#### 6.2.1 确定指标的原则

##### 1) 硬实力评估

院校的硬件设施水平是教育工作开展的基础，能够体现教育工作的执行能力。保障了人力资源能够发挥的最大潜力。

##### 2) 软实力评估

院校的师资力量及文化底蕴，是教育工作的内核。优秀的师资队伍和深厚的校园文化才能由内而外地展现综合实力。

##### 3) 社会贡献

反映院校的教育成果与质量，其对社会带来的效益是人才培养的根本目的。

#### 6.2.2 现有的评价指标体系

现有的本科教育质量评价指标体系一般分为三至四层：目标层（一级指标）均为本科教育质量，其指数数值越高，则该城市本科教育质量越高。准则层（二级指标）、领域层（三级指标）和指标层（四层）根据实际情况及理论依据而各有不同。根据江苏省教育厅发布的《江苏普通高校本科教学质量报告》、《南京大学本科质量报告》总结出两个较为全面且可学的评价体系，如表 6-1 所示：

表 6-1 本科教学质量的准则层指标及选取原则

评价体系	准则层分层数	准则层主要指标	准则层选取原则
(1)	3	<ul style="list-style-type: none"><li>• 教学质量</li><li>• 教学建设与改革</li><li>• 教学研究与质量保障</li></ul>	<ul style="list-style-type: none"><li>• 以设施与师资为基础</li><li>• 以教学为根本</li><li>• 以改革为方向</li></ul>

(2)	4	<ul style="list-style-type: none"> <li>• 教学基本情况</li> <li>• 教学建设与改革</li> <li>• 质量保障与效益</li> <li>• 特色发展</li> </ul>	<ul style="list-style-type: none"> <li>• 保障教学基础建设</li> <li>• 注重教学改革与创新</li> <li>• 强调质量</li> <li>• 具有特色的发展</li> </ul>
-----	---	--	--

### 6.2.3 本科教育质量的评价指标体系

基于表 6-1 中本科教育质量指标设立的原则，并结合江苏省本科教育现况来综合考虑，我们对以上两种体系进行了优化并应用于准则层。又由于题目已给出具体的九个评价指标，我们根据其数量与指标涵盖范围宽度，将这个九个指标归为领域层，即三级指标。对于三级指标，我们根据指标性质及涵盖内容对其进行细化，最终确立出了四层评价指标体系：具体如表 6-2 所示：

表 6-2 本科教学质量的评价指标体系

目标层（一级指标）	准则层（二级指标）	领域层（三级指标）	指标层（四级指标）	
地级市本科教育质量	基本教学条件	本科院校数量	本科院校数量	
		招生人数	招生人数	
		师资队伍与结构	教师（不含外聘）	教师（不含外聘）
			教授	教授
			院士	院士
			千人计划	千人计划
		生师比	生师比	
		教学条件与利用	校区	校区
			占地面积/亩	占地面积/亩
			藏书量（纸质，万册）	藏书量（纸质，万册）
	教学科研仪器/件		教学科研仪器/件	
	本科经费投入/万元（日常）		本科经费投入/万元（日常）	
	日常		日常	
	学科实力	专业建设与教学改革	学院数	学院数
			专业数	专业数
			教学改革项目	教学改革项目
			国家级精品课程	国家级精品课程
			国家重点学科	国家重点学科
		科研投入与产出	国家实验室	国家实验室
			国家重点实验室	国家重点实验室
			国家级平台	国家级平台
省级平台			省级平台	
科研经费/亿			科研经费/亿	
质量评估	学生就业	就业率	就业率	
		升学率	升学率	
		出国率	出国率	
	双一流学科建设	双一流学科建设		
	双一流学科数	双一流学科数		

### 6.3 三级指标的综合量化评价模型

首先对数据进行预处理。由于数据收集、量化的过程限制，且本科教育质量评价体系需要相对客观性，本文采用了熵值赋权法，其特点是相对客观性较强，能充分利用数据特征。定义第  $i$  个三级指标量化值  $C_i$  ( $i=0,1,2\dots 9$ )，则  $C_i$  的计算公式为：

$$C_i = \sum_{j=1}^n D_{ij} \cdot w_{ij} \quad (6-1)$$

其中， $D_{ij}$  为经过无量纲化后的第  $i$  个三级指标下的第  $j$  项四级指标数值， $w_{ij}$  为  $D_{ij}$  的熵权。 $D_{ij}$  的熵权  $w_{ij}$  的计算步骤如下：

*Step1*：级指标数据  $D_j$  标准化后数据的比重  $G_{ij}$  的计算：

$$G_{ij} = \frac{D_{ij}}{\sum_{i=1}^n D_{ij}} \quad (6-2)$$

*Step2*： $j$  项四级指标  $D_i$  的熵值  $e_{ij}$  的计算：

$$e_{ij} = - \left( \frac{1}{\ln n} \right) \sum_{i=1}^n G_{ij} \ln G_{ij} \quad (6-3)$$

*Step3*：第  $j$  项四级指标  $D_i$  的差异性系数  $f_{ij}$  的计算：

$$f_{ij} = 1 - e_{ij} \quad (6-4)$$

$f_{ij}$  的值越大，则  $D_i$  的重要性越强。

*Step4*：第  $j$  项四级指标  $D_i$  熵权  $w_{ij}$  的计算

$$w_{ij} = \frac{g_{ij}}{\sum_{i=1}^n g_{ij}} \quad (6-5)$$

至此，基于熵权法的综合评价模型建立完毕。

### 6.3 三级指标的综合量化评价模型的求解

法，求三级指标  $C_i$  下，四级指标  $D_j$  的权重。这里仅列出三级指标  $C_3$ （师资队伍与结构）下，四级指标  $D_{31}$ 、 $D_{32}$ 、 $D_{33}$ 、 $D_{34}$  的权重，见表 6-3。其余指标权重参见附录 1。

表 6-3 三级指标  $C_3$  下各四级指标的熵权

熵权 ( $w_{ij}$ )	$w_{31}$	$w_{32}$	$w_{33}$	$w_{34}$
权值	0.112	0.158	0.29	0.43

由此可以得到三级指标  $C_3$  的计算式为:

$$C_3 = \sum_{j=1}^4 D_{3j} \cdot w_{3j}$$

同理可得到  $C_1$ 、 $C_2$ 、 $C_3$ ..... $C_9$  的数值,即完成了对三级指标的量化。以苏州市的三级指标为例,见表 6-4。其余地级市的三级指标量化值见附录 2。

表 6-4 三级指标量化表

$C_{ij}$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$
量化值	0.8333	0.1616	0.0579	0.5355	0.1644	0.0444	0.4575	0.0684	0.0263

由此填写题中给出的表格,见表 6-5。

表 6-5 三级指标量化处理表

序号	指标名称	量化处理方式	计算依据	备注
1	本科院校数量	熵权法	$C_1 = D_1$	无
2	招生人数	熵权法	$C_2 = D_2$	无
3	师资队伍与结构	熵权法	$C_3 = 0.11D_{31} + 0.16D_{32} + 0.3D_{33} + 0.4D_{34}$	无
4	生师比	熵权法	$C_4 = D_4$	无
5	教学条件与利用	熵权法	$C_5 = 0.15D_{51} + 0.3D_{52} + 0.3D_{53} + 0.06D_{54} + 0.22D_{55}$	无
6	专业建设与教学改革	熵权法	$C_6 = 0.16D_{61} + 0.11D_{62} + 0.18D_{63} + 0.85D_{64} + 0.26D_{65}$	无
7	学生就业	熵权法	$C_7 = 0.47D_{71} + 0.43D_{72} + 0.10D_{73}$	无
8	科研投入与产出	熵权法	$C_8 = 0.3D_{81} + 0.23D_{82} + 0.13D_{83} + 0.01D_{84} + 0.11D_{85} + 0.10D_{86} + 0.12D_{87}$	无

9	双一流学科建设	熵权法	$C_9 = D_9$	无
---	---------	-----	-------------	---

## 6.4 相关数据分析

### 6.4.1 数理统计

通过 SPSS 求解江苏省各个地级市的不同指标的极大值、极小值、平均数、偏度、方差、标准偏差、峰度，得到的数据处理结果如表 5-1-3 所示：其中指标 1 到指标 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。统计数据见表 6-6。

表 6-6 各指标描述性统计数值表

	N	全距	极小值	极大值	和	均值	标准差	方差	偏度	峰度			
	统计量	统计量	统计量	统计量	统计量	标准误	统计量	统计量	统计量	标准误	统计量	标准误	
专业建设与教学改革综合得分	13	1	0	1	1	0.1	0.075	0.27	0.073	3.562	0.616	12.768	1.191
招生人数综合得分	13	1	0	1	2	0.15	0.073	0.262	0.068	3.264	0.616	11.153	1.191
教学条件与利用综合得分	13	1	0	1	2	0.15	0.074	0.266	0.071	3.14	0.616	10.411	1.191
学生就业综合得分	13	1	0.00E+00	1	4	0.307692308	0.075250888	0.271320934	0.074	1.491	0.616	2.477	1.191
师资队伍与建设综合得分	13	1	0	1	1	0.1	0.075	0.271	0.073	3.562	0.616	12.769	1.191
生师比综合得分	13	1	0.00E+00	1	5.976777939	0.459752149	0.072932763	0.262962815	0.069	0.369	0.616	0.685	1.191

双一流 学科数 综合得 分	13	1	0	1	1	0.09	0.076	0.275	0.076	3.573	0.616	12.825	1.191
科研投 入综合 得分	13	1	0	1	1	0.09	0.076	0.273	0.074	3.564	0.616	12.783	1.191
本科院 校数量 得分	13	1	0	1	1	0.11	0.075	0.271	0.073	3.463	0.616	12.244	1.191

根据表中各地级市各项指标的平均、偏斜度、标准偏差和峰度等参数的分析，能看出江苏省 13 个地级市的本科教育质量存在很大差异，南京、徐州和苏州教育水平明显高于其他地级市。

#### 6.4.2 数据分析

通过数据的统计和分析，得到了以下结论：

- 1) 江苏省各个地级市的双一流学科数、科研投入、专业建设与教学改革等指标在地区内相对集中，呈现不均衡状态。
- 2) 学生就业指标在各个地级市差异较小。

## 七、问题二的模型建立与求解

问题二要求我们根据 9 个指标数据，建立数学模型，对江苏省的 13 个地级市的本科教育质量进行综合评价。根据问题一指标的量化处理，我们利用熵权法和主成分分析法，从而求解模型。两种方法互为验证，从而得到更有利的综合评价。

### 7.1 基于主成分分析法的综合评价指标体系

由于这 9 个三级指标实际上是有关联性的，而主成分分析可以把这多个相关联的指标转化为少数几个不相关的综合指标，以此来降低指标集密度。在问题一的基础上，已经得到了标准化后三级指标的值，利用这些值进行以下步骤的计算。

**Step1:** 建立模型，选择  $p$  ( $p \leq 9$ ) 计算本科教育质量的综合评分  $\delta_i$ ：

$$\delta_i = \sum_{e=1}^n \beta_e f_e \quad (7-1)$$

其中  $f_e$  表示第  $e$  个主成分，第  $i$  个主成分的权重用  $\beta_e$  ( $e=1,2,\dots,p$ ) 表示。

**Step2:** 计算各成分  $\chi_i$  ( $i \leq 9$ ) 特征值  $\lambda_i$  ( $i=1,2,\dots,9$ ) 的信息贡献率和累计贡献率：

$$b_i = \frac{\lambda_i}{\sum_{\theta=1}^i \lambda_{\theta}}, i=1,2,\dots,9 \quad (7-2)$$

$$\alpha_e = \frac{\sum_{\theta=1}^e \lambda_\theta}{\sum_{\theta=1}^p \lambda_\theta} \quad (7-3)$$

其中， $b_i$ 为各成分 $\chi_i$ 信息贡献率， $\alpha_e$ 为各成分 $\chi_1, \chi_2, \dots, \chi_p$ 的累积贡献率。当 $\alpha_e$ 接近1 ( $\alpha_e=0.85, 0.90, 0.95$ )时，前 $p$ 个指标变量 $\chi_1, \chi_2, \dots, \chi_p$ 作为主成分，代替原来的9个指标变量，从而可对 $p$ 个指标变量成分进行综合分析。

Step3: 计算综合得分:

$$\delta_i = \sum_{e=1}^i \beta_e b_e \quad (7-4)$$

其中： $b_e$ 为第 $e$ 个主成分的信息贡献率，根据综合得分值就可以进行评价。

根据以上步骤，本文利用SPSS统计软件，由主成分分析和特征贡献率表可知特征根 $\lambda_1 = 6.911$ ，特征根 $\lambda_2 = 1.329$ 前两个主成分的累计方差贡献率达91.489%，即涵盖了大部分信息。这表明前两个主成分能够代表最初的9个指标来分析江苏省13个地级市的本科教育质量，故提取前两个指标即可。前两个主成分分别记作 $f_1, f_2$ 。

部分结果见表7-1，全表见附录3。

表7-1 主成分分析结果表

	初始特征值			提取平方和载入		
	合计	方差的%	累积%	合计	方差的%	累积%
1	6.911	76.787	76.787	6.911	76.787	76.787
2	1.329	14.762	91.549	1.329	14.762	91.549
3	0.683	7.587	93.135			
4	0.058	0.643	96.117			

成份矩阵表明专业建设与教学改革、招生人数、教学条件与利用、师资队伍建设、科研投入与产出、双一流学科建设、本科院校数量各指标在第一主成分上有较高的载荷，相关性高。学生就业与生师比在第二主成分上有较高载荷，相关性高。成分得分系数矩阵（因子得分系数）列出了强两个特征根对应的特征向量，即各主要成份解析表达式中的标准化变量的系数向量，部分结果见表7-2与7-3，全表见附录4。

表7-2 成份矩阵表

项目	成份	
	1	2
专业建设与教学改革综合得分	0.998	-0.023

招生人数综合得分	0.99	0.06
教学条件与利用综合得分	0.975	0.03
学生就业综合得分	-0.043	-0.807

表 7-3 成分得分系数矩阵表

项目	成份	
	1	2
专业建设与教学改革综合得分	0.144	-0.018
招生人数综合得分	0.143	0.045
教学条件与利用综合得分	0.141	0.022
学生就业综合得分	-0.006	-0.608

将各指标的标准化数据带入个主成分表达式中，求得主成分值。分别以 2 个主成分的贡献率为权重，构建第  $t$  个地级市的本科教育综合评价  $Z_t (t=1,2,\dots,13)$ ，即

$$Z_t = 0.7098f_{t1} + 0.1136f_{t2} \quad (7-5)$$

其中  $f_{t1}, f_{t2}$  是针对于第  $t$  个地级市的主成分。可以得到各城市的本科教育质量水平的综合得分以及排序结果，部分结果见表 7-4，全表见附录 5。

表 7-4 基于主成分分析法得出的城市综合得分表

排名	城市	综合得分
1	南京	7.259731978
2	苏州	2.124733901
3	无锡	0.064676967
4	南通	0.013528803
5	徐州	0.019736643
6	镇江	-0.123371951

## 7.2 基于熵权法综合评价指标体系

定义第  $t$  市本科教育质量的综合评分  $Z_t (t=1,2,\dots,13)$ ，则  $Z_t$  的计算公式为：

$$Z_t = \sum_{\rho=1}^n C_{\rho} w_{\rho} \quad (7-6)$$

其中， $C_{\rho}$  为第  $t$  个城市第  $\rho$  个三级指标的量化值， $w_{\rho}$  为  $C_{\rho}$  的熵权。

引用 6.3 的模型计算  $C_i$  的熵权  $w_i$ ，由此得到了各项指标对地级市本科教育质量的影响所占的权重，则可计算第  $t$  个地级市本科教育质量的综合评分值，该评分值越大，该城市本科教育质量越高。

至此，利用熵权法对三级量化指标  $C_i (i=1,2,\dots,9)$  进行赋权从而计算出市级本科

教育质量的模型建立完毕。

根据以上步骤，将以上的标准化数据带入该模型中进行计算，得出三级量化指标的熵权部分，部分结果见表 7-5，全表见附件 6。

表 7-5 三级指标  $C_i$  的熵权表

三级指标 $C_i$	$C_1$	$C_2$	$C_3$	$C_4$
熵权	0.148991	0.041864	0.156894	-0.32670

注：生师比  $C_4$  应呈现负相关，即每一名教师对应的学生越少，则越好，所以该项指标的权重是非异常值。

将上表中各指标的权重代入式 (7-1) 中，计算出 13 个地级市的本科教育质量的综合评分  $Z_t (t=1,2,\dots,13)$ ，这里仅列出第三个城市（无锡）的综合评分  $Z_3$ ：

$$Z_3 = \sum_{i=1}^9 C_{3i} w_{3i} = 5.716009633 \quad (7-7)$$

由此可以得到江苏省各城市的综合得分。

并按照得分的高低对其进行排名。为了便于观察我们将综合评分扩大 100，部分结果见表 7-6，全表见附录 7。

表 7-6 基于熵权法得出的城市综合得分表

排名	城市	综合得分
1	南京	95.86566666
2	苏州	11.25913778
3	无锡	5.716009633
4	南通	3.153344101
5	镇江	0.766770225
6	徐州	0.484868691

### 7.3 综合评价结果分析

分别利用主成分分析法与熵权法对江苏省的 13 个地级市的本科教育质量进行综合排名，排名对比见表 7-7。

表 7-7 城市综合得分对比表

城市	熵权值排名	主成分分析排名	浮动排名（绝对值）
南京	1	1	0
苏州	2	2	0
无锡	3	3	0
南通	4	4	0
镇江	5	6	1
徐州	6	5	1
扬州	7	7	0
盐城	8	9	1
淮安	9	10	1

常州	10	8	2
连云港市	11	11	0
泰州	12	12	0
宿迁	13	13	0

二者的排名结果浮动较小，变动均在两名以内。我们认为模型建立较为成功，综合排名的结果可靠性较高。二者排名的前四名与后三名均未发生变动，这是因为 7 个城市的本科教育质量确实存在较大的差异，教育资源很不平衡，而排名处于中间的城市的教育资源较为接近，在用不同的方法进行综合分析时，存在客观与主观性的差异，导致其有微小的浮动。南京市的本科教育综合评分是远高于他其他地级市，这与南京作为江苏省的政治、经济与文化中心的地位是吻合的，宿迁仅有一所民办的本科学校，其教育质量自然远落后于其他地级市。

## 八、问题三的模型建立与求解

问题三要求建立数学模型，分析在所给的 9 个指标中，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。我们插入一个虚拟控制量（即控制一个城市）来调控各三级指标的变量，且使每项三级指标的变化量对于该指标的影响是等效的，从而得到各指标灵敏度的排名。并运用了灰色关联分析法，将定量与定性的方法有机结合，将各指标与本科教育质量的关联度进行排序，与虚拟控制量法进行验证，确定到底可以减少哪些三级指标。

### 8.1 虚拟控制调节法

*Step1*: 插入一个虚拟控制量，则该城市的各项三级指标数据为：

$$v'_{i0} = \frac{1}{n} \sum_{j=1}^n v'_{ij} \quad (8-1)$$

其中， $v_{i0}$  为虚拟控制城市的第  $i$  个三级指标数据。将标准数据按照 5.2，进行无量纲化的处理。将标准化处理后的数据代入模型 6.2 的公式中进行计算，得到这

13 个城市的评分： $\hat{O}_0 = \{\hat{O}_{01}, \hat{O}_{02}, \dots, \hat{O}_{09}\}$

为了观察各指标产生同一波动对目标层产生的影响，分别按  $v'_{10}, v'_{20}, v'_{30}, \dots, v'_{i0}$  增大了 20% 后重复上述无量纲化过程及评价过程，即可得到调节控制城市第  $i$  个指标后

所产生的 9 个城市的综合评分  $\hat{O}_i = \{\hat{O}_{i1}, \hat{O}_{i2}, \dots, \hat{O}_{i9}\}$

若直接通过排名变化来对指标的灵敏度进行评定，不仅评定精度低，且易受偶然因素影响，因此，本文将指标调节前后各城市的综合评分的相对距离大小作为该指标的灵敏性评价的标准。

*Step2*: 参照 5.2，将综合评分进行无量纲化处理，计算变化距离。

对  $i$  个指标数据进行调节前后，第  $t$  个城市的综合评分与调节之前相比较会产生

一段一维变化距离： $|O'_{it} - O'_{or}|$ ，那么  $t$  个城市的相对变化距离可用  $n$  维欧式距离来计算，公式如下：

$$dis_i = \sum_{t=1}^n |O'_{it} - O'_{or}|^2 \quad (8-2)$$

其中， $dis_i$  为第  $i$  个指标数据调节前后  $t$  个城市综合排名的相对变化距离，本文中  $t=9$ 。

*Step3*：判断灵敏度。

相对距离  $dis_i$  越大，该指标的波动对于城市评分的影响越大，说明该指标的灵敏性越强对城市排名的影响越显著。

在 *Matlab* 中进行上述步骤，插入虚拟城市后，其各项指标的综合评分部分结果见表 8-1。全表见附录 8。

表 8-1 虚拟控制城市的各项三级指标综合得分表

三级指标	虚拟控制城市综合得分
本科院校数量	0.108974359
招生人数	0.154872253
师资队伍与建设	0.101153456
生师比	0.459752149

已知虚拟控制城市的本科院校数量的综合评分为  $v'_{10} = 0.108974359$ ，将其调节增大 20%， $v''_{10} = (1+20\%)v'_{10} = 0.1307692308$ ，其他指标综合评分不变。这里仅计算调节第一个三级指标前后的 9 个城市评分的相对距离：

$$dis_i = \sum_{t=1}^n |O'_{it} - O'_{or}|^2 = 0.1878$$

将调节后的数据重复上述过程可以得到 9 个指标数据的 9 维相对距离，见表 8-2

表 8-2 相对距离及权重表

三级指标	相对距离	权重
生师比综合得分	0.8116	-0.32670361
双一流学科数综合得分	0.2081	0.244116517
科研投入综合得分	0.1878	0.169851252
专业建设与教学改革综合得分	0.1804	0.15232664
师资队伍与建设综合得分	0.1792	0.156894983
教学条件与利用综合得分	0.1698	0.05737263
本科院校数量综合得分	0.1357	0.148991183
学生就业综合得分	0.0873	0.075252506

招生人数综合得分	0.0863	0.041864651
----------	--------	-------------

对于各指标产生相对相同的波动，9个城市本科教育质量的综合评分相对距离越大则该指标的灵敏性越高。我们可以发现，灵敏度最高的指标为“生师比的综合评分”，其权重的绝对值也是最高的。此外，其他数据也基本符合“权重越大，灵敏性越大”的规律。

### 8.2 灰色关联分析法

*Step1*：利用问题二中的熵权法中得到的权重为各指标值对应的权重  $w = [w_1, \dots, w_n]$ ，其中  $w_i (i=1, 2, \dots, n)$  为第  $i$  个指标对应的权重。

*Step2*：计算灰色关联系数：

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \quad (8-3)$$

为比较数列  $x_i$  对参考数列  $x_0$  在第  $i$  个指标上关联系数，其中  $\rho \in [0, 1]$  为分辨系数。

*Step3*：计算灰色关联加权重：

$$r_i = \sum_{k=1}^n w_i \xi_i(k) \quad (8-4)$$

其中： $r_i$  为第  $i$  个指标的灰色关联加权重。

*Step4*：根据灰色关联度的大小，对各评价指标关联度进行排序。

根据以上步骤，在 *Matlab* 和 *Excel* 完成上述步骤。结果见表 8-3。

表 8-3 三级指标灰色关联度表

三级指标	灰色关联度
生师比综合得分	0.5445
教学条件与利用综合得分	0.2555
师资队伍与建设综合得分	0.2447
专业建设与教学改革综合得分	0.2438
科研投入综合得分	0.2438
学生就业综合得分	0.2397
双一流学科建设综合得分	0.2389
本科院校数量得分	0.2247
招生人数综合得分	0.2201

### 8.3 指标的筛选及评价分析

虚拟控制法中的相对距离越大，则该指标灵敏度越高，灰色关联分析法中指标的关联度越高，则越应该优先考虑其对本科教育质量的影响，根据两种方法对于不

同指标的影响程度，由高到低进行排序，如表 8-4 所示。

表 8-4 排序对比表

三级指标	灰色关联法关联排序	虚拟控制的相对距离排序
本科院校数量	8	7
招生人数	9	9
师资队伍与建设	3	5
生师比	1	1
教学条件与利用	2	6
专业建设与教学改革	4	4
学生就业	6	8
科研投入与产出	5	3
双一流学科建设	7	2

我们可以发现，两种方法下，本科院校数量、招生人数对本科教育质量的影响均较小，其排序变化的浮动均在一名以内，因此我们认为减少这两个指标对 13 个地级市的本科教育质量的评价结果影响较小。这是因为这两个指标受院校性质、地市级的地表面积的影响较大，综合性院校的招生人数可能会大于专业性的大学，该市的面积越大，其建立更多的院校，所以可认为这两个指标的波动对本科教育质量关系较小。

#### 8.4 指标筛选的验证

为了验证这一结论，对于剩余的数据，我们再次利用问题二中的熵值法与主成分分析再次建立数学模型进行求解，过程不再赘述，其结果见图 8-1 与图 8-2。对应数据表格见附录 9。

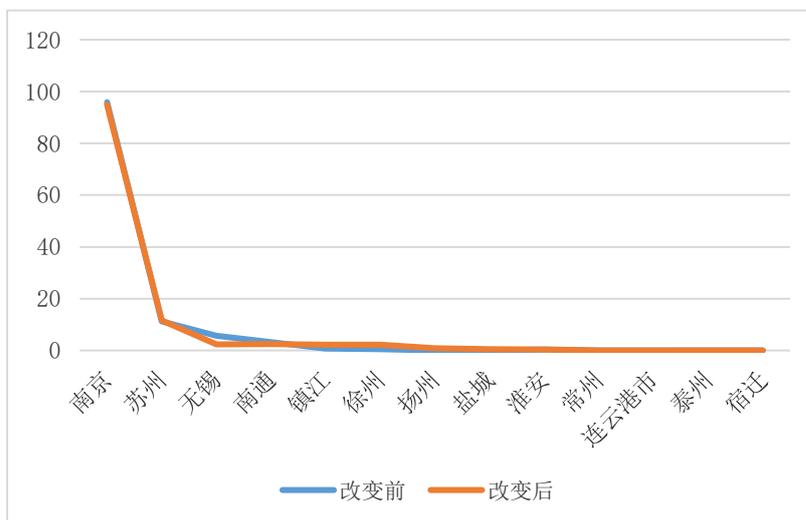


图 8-1 基于熵权法的评估对比

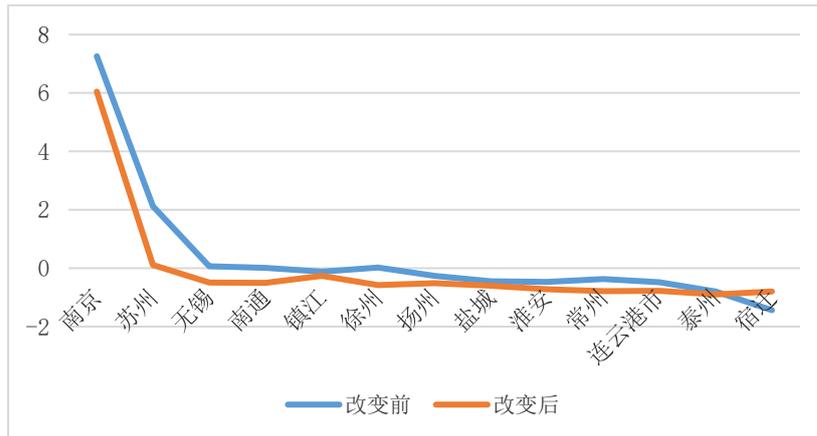


图 8-1 基于主成分分析法的评估对比

由上图可以看出，筛出这两个指标后对于该市的本科教育质量的综合评分及排名影响较小，因此可以减少这两个指标。

## 九、问题四的模型建立与求解

问题四要求确定一个关键的指标，使该指标的普遍改善可以尽可能缩小江苏省 13 个地级市本科教育发展的差异。我们选用敏感度模型，任意改变主要三级指标的综合得分，对比改变前后 13 个地级市本科教学质量的得分，来判断该关键指标是否符合要求。

### 9.1 确定主要指标

*Step1*：确定影响本科教育质量发展的主要三级指标。

问题一、二与三中，我们通过熵权法、主成分分析法、灰色关联分析等方法，已经从 28 个影响本科教育质量的指标中筛选出了 7 个主要指标：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。定义这 7 个主要指标为  $\psi_h (h=1,2,\dots,7)$ 。

*Step2*：确定指标的变化范围。

任意选择将这 7 个主要指标变化，变化幅度为  $\nu$ ，变化后的指标为  $\psi'_h$ ：

$$\psi'_h = (1+\nu)\psi_h \begin{cases} \nu = \pm 20\% \\ \nu = \pm 40\% \end{cases} \quad (9-1)$$

若改变  $\psi_h (h=1,2,\dots,7)$  的变化幅度大，则  $\nu = \pm 40\%$ ；若改变指标的变化幅度小，则  $\nu = \pm 20\%$ 。

### 9.2 敏感度模型

*Step1*：选择第  $\gamma$  项指标，将其代入 (9-1)，

得到第  $\gamma$  项指标集  $\{\psi_1, \psi_2, \dots, \psi'_\gamma, \dots, \psi_7\}$ ，

Step2: 将第 $\gamma$ 项的指标集 $\{\psi_1, \psi_2, \dots, \psi'_\gamma, \dots, \psi_7\}$ 代入 7.2 建立的熵权法模型中,

计算第 $\gamma$ 项的敏感度 $\mu_h$ :

$$\mu_h = \sum_{a=1}^n \left| \frac{Z'_{th} - Z_{th}}{Z_{th}} \right| \quad (9-2)$$

其中 $Z_{th}, Z'_{th}$ 是变化前后针对于第 $\gamma$ 项的第 $t$ 个城市的综合评分值, 在本文中 $n=9$ 。

在 *Matlab* 和 *Excel* 完成上述步骤, 完成表 9-1。

表 9-1 敏感度表

变动因素	变动比例 (%)	得分	敏感度
本科院校数量	+20	1.5426	1.2111
	+40	1.6824	
	-20	1.4988	
	-40	1.3824	
招生人数	+20	1.4302	1.8423
	+40	1.8753	
	-20	1.2214	
	-40	1.0012	
师资队伍与结构	+20	1.8785	5.9862
	+40	2.2231	
	-20	1.6546	
	-40	1.2212	
教学条件与利用	+20	1.4335	3.6126
	+40	1.5424	
	-20	1.3412	
	-40	1.2213	
专业建设与教学改革	+20	1.3142	7.4121
	+40	3.4564	
	-20	1.2011	
	-40	0.0021	
科研投入与产出	+20	1.9864	5.8311
	+40	2.7543	
	-20	3.6537	
	-40	0.8970	
双一流学科建设	+20	1.7767	3.7642
	+40	2.1444	
	-20	3.5420	
	-40	1.1686	

由表中的敏感度我们可以看出, 专业建设与教学改革的敏感度是其他指标敏感度最大, 且与其余敏感度差别较大, 则该指标变化将会使江苏省 13 个地级市本科教育质量综合评分变化较大, 从而使得排名发生显著变化。则我们认为专业建设与教学改革可以作为关键性的指标。

### 9.3 利用查找的数据资料验证所建立模型的有效性

本文将专业建设与教学改革指标数据升高 40%，代入 7.3 建立的模型，得到指标数据变化后本科教育综合质量综合得分，并与未改变该指标的综合得分对比。得分如下图所示：

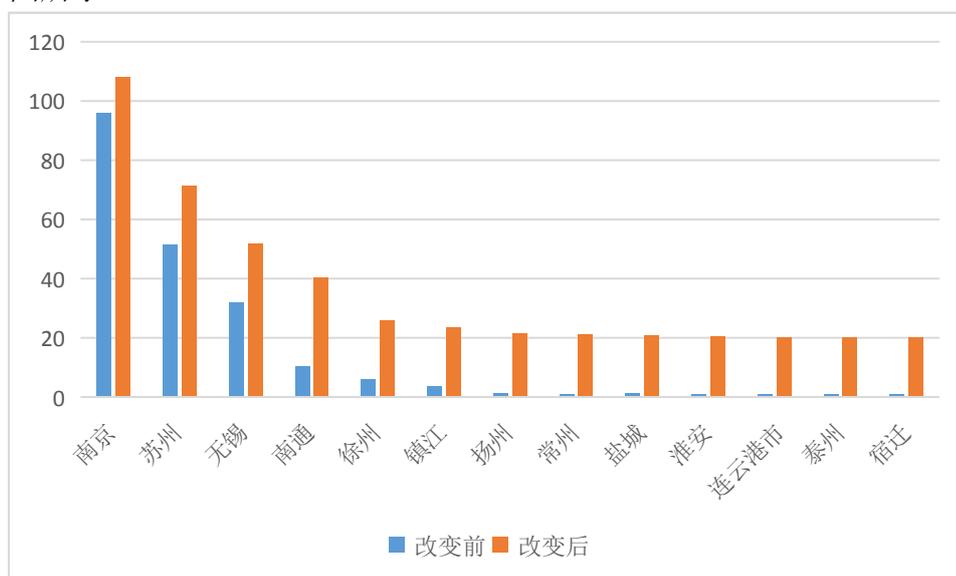


图 9-1 专业建设与教学改革指标改变前后评分图

改变后各地级市的评分与各市的评分差距均明显降低，且得分越低的城市的增量明显越高，说明该指标确实可以缩小江苏省 13 个地级市本科教育发展的差异，有力验证了该关键指标的有效性。

## 十、对有效提升江苏省本科教育质量的政策建议

本文收集能反映江苏省各个地级市本科教育质量的四级指标数据，基于熵权法与主成分分析法建立了综合评价模型，并对 13 个地级市的本科教育质量进行了排名。并在后续的指标筛选与优化中，对各指标有了定性与定量的理解。基于问题二三四，我们给出了如下的政策建议。

### 10.1 增强专业学科建设，加大教学改革

问题四中可知专业学科建设与教学改革的敏感度最高，是改善各地级市之间本科教育质量不均衡的关键指标，因此，应针对这一点进行大力的教育改革。本科大学之间要加大联合研究课题的力度，学科之间采取薄弱学科优秀教师“巡回走教”以解决校际之间的学科教学不平衡，校级领导之间互换挂职锻炼。

### 10.2 加大科研投入经费

问题三中的虚拟控制调节法可知该指标的相对距离较大，即灵敏度很高，则该指标对 13 个地级市本科教育质量的影响较大。加大部分教育水平较落后的地级市对教育的重视和投入，积极鼓励师生创新，提高科研水平。创立一些科研活动，积极调动学生们的兴趣，并让学生可以提前进入实验室学习。

### 10.3 平衡师资力量与人才引进

师资队伍与结构、生师比在灰色关联分析法与虚拟控制调节法中均表现出与 13 个地级市本科教育质量有较强的影响。可以采取区域内互相借调方式，每两年至三年为一周期，学校之间分批交换教师挂职教学。这样可达到相互借鉴，相互交流，相互提高并监督，从而达到师资力量，教育质量更加均衡。教育资源均衡化，能一定程度改善择校热，避免教育资源不平衡的进一步加剧。

#### 10.4 有方向性地平衡教育资源

同时，江苏省政府也应该加强政策体系建设，按照各个高校的背景不同，发展方向不同，定位不同合理的分配资源，优化教育资源配置，让资源的投放更加具有针对性。

### 十一、模型的评价、改进及推广

#### 11.1 模型的评价

本文模型考虑的十分详细全面，例如：

- 1) 本文在指标量化的过程中利用熵权法，更加客观。利用软件 SPSS 进行数据峰度，偏度等参数的计算，进行多方面分析，数据更加可靠。
- 2) 其次利用熵权法与主成分分析法对指标再次分析，将二者的结果二者相互认证，提高了结果精确度与可信度。
- 3) 利用虚拟控制调节法、灰色关联分析等方法分析各个指标对评分的影响性与关联度；

#### 11.2 模型的改进

- 1) 用于预测的数据样本稍少，导致存在一定误差。在此模型以后的应用中，为使结果更加准确，应对数据的搜集更加官方与准确。
- 2) 对于部分指标没有继续细化评价，没有做到更全面的分析。可用 AHP 进行进一步的筛选。
- 3) 问题四中的敏感度分析不能很好的体现各个指标的交互作用，如果能够对指标进行进一步的进行因子分析，结果将更具可靠性。

#### 11.3 模型的推广

本文所建立的评价模型简单实用，结果准确可靠，为这些城市的发展提供了一定的参考作用，该模型也可用于其他评价问题。

### 十一、参考文献

- [1] 高校本科教育质量评价理论探索与体系构建，山东科技大学，2017
- [2] 司守奎，孙玺菁，数学建模算法与应用，国防工业出版社，2013.8
- [3] 姜启源，谢金星，叶俊，数学建模（第四版），高等教育出版社，2011.1
- [4] 何璠，基于 BP 神经网络的环境质量评价模型研究，成都：四川大学 2006.
- [5] 杨志禄，增强学士学位意识，提高本科教育质量，中国高等教育，1995（6）

### 十二、附录

#### 附录 1

表 1 三级指标  $C_1$  下各四级指标的熵权

熵权 ( $w_{ij}$ )	$w_{11}$
权值	1

表 2 三级指标  $C_2$  下各四级指标的熵权