

参赛队号#1248

## 第六届“认证杯”数学中国

## 数学建模网络挑战赛

## 承 诺 书

我们仔细阅读了第六届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

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

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

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：1248

参赛队员 (签名)：

队员 1：陈胜

队员 2：李玮

队员 3：张思德

参赛队教练员 (签名)： 无

参赛队伍组别：大学本科组

参赛队号#1248

## 第六届“认证杯”数学中国

### 数学建模网络挑战赛

#### 编 号 专 用 页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：  
1248

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

参赛队号#1248

# 2013 年第六届“认证杯”数学中国 数学建模网络挑战赛

题 目 计及城市差异的公路运输业对 GDP 的影响分析

关 键 词 产业关联理论 地区差异 K 均值聚类

基于 AHP 的模糊综合指数评价 相关性分析 支持向量机

## 摘 要：

本文研究公路运输业对国民生产总值（GDP）的影响，对题目中大量数据（异常数据予以剔除）利用统计学原理建立相关模型求解。

对于问题 1，由题意可知，公路运输业对 GDP 增长的贡献分交通建设和客货运输两个阶段。对于第一阶段，运用投入产出方法和产业关联理论定量分析各个部门产值的增加对 GDP 增长的贡献得分，得出在所有 42 个部门中，交通建设的贡献得分排名第 3，故相对于其余部门，可认为该省交通建设对 GDP 增长有很大显著性影响。

对于问题 1 的第二阶段，考虑到该省各个城市生产力水平的差异，不能用统一的指标衡量客货运输对 GDP 的影响。首先筛选出由生产力水平的差异造成的调查项目，标准化后用 K 均值聚类将 12 个城市分为 4 类（依次为发达、较发达、中度发达、欠发达）；然后将客货运输的影响细分为 9 个方面，各类地区将每个方面的影响值归一化，采用基于层次分析法（AHP）的模糊综合指数评价，得出 4 类地区的影响指数为 2.85, 3.01, 2.64, 2.48（指数为 3 表示影响较显著，2 表示影响一般），说明随着生产力水平的提高，客货运输对 GDP 的影响越显著，但当生产力水平的发展趋于饱和时，客货运输对生产力的影响指数也趋于饱和。最后综合各类城市影响指数可得，该省客货运输对 GDP 的影响较为显著。

综合两个阶段对 GDP 的影响，可知该省公路运输业对 GDP 的影响较为显著。

在分析问题一的基础上，进一步求出各类城市 9 项指标与客货运输对 GDP 的影响之间的相关系数，发现有 5 项指标对 GDP 影响较大，说明这 5 项指标在公路运输业对城市 GDP 的贡献中占最主要地位，调查项目可以对这些指标进行细化，调查尽可能多的影响因素；同时有 2 项指标对于不同生产力水平的城市 GDP 影响有较大差异，应考虑城市差异制定不同的调查项目；此外，有 2 项指标对于 GDP 的影响较小，在调查项目有限的情况下，可剔除一些影响较小的项目；最后还考虑了增加“满意度”调查项、调整模型参数、调查消费情况等途径提高模型精确度。

最后，采用支持向量机（SVM）的方法验证了模型的可靠性。

参赛队号： 1248

参赛密码 \_\_\_\_\_  
(由组委会填写)

所选题目： C 题

## 参赛队号#1248

## ABSTRACT

This essay studies the highway transportation's contribution to gross national product growth. Based on the Principle of Statistics to build mathematical models for solving large amounts of data in the title.(deleting the abnormal date)

For problem one, according to the title, the highway transportation's contribution to gross national product growth is divided into two stages: transportation construction and passenger and freight transport. For the first stage, the authors use the input-output method and industry correlation theory to each department's contribution to GDP growth score quantitatively, then reach the conclusion that the score of the transportation construction ranks NO.3 among the 42 departments. So, compared with other departments, transportation construction has an evident influence on GDP growth in this province.

For the second stage, considering the differences of the level of productivity in every city of the province, the authors cannot measure the impact of passenger and freight transport to GDP growth with the same unified indicators. Firstly, the authors screen out the survey items caused by the difference of the level of productivity and divide the 12 cities into 4 types (developed, more developed, moderately developed, less developed area) with the k-means clustering after Standardization. Subsequently, passenger and freight transport is divided into nine areas and all kinds of areas' the impact of every aspect of the normalization, using fuzzy composite index based on the Analytic Hierarchy Process (AHP) evaluation to know the influence index of the 4 types is: 2.85, 3.01, 2.64, 2.48 (index for the 3 effect is more significant, 2 is general). At the end, taking various types of urban influence index available into account, the authors conclude that the impact of passenger and freight transport to GDP growth is obvious.

In conclusion, the highway transportation's contribution to gross national product growth is relatively notable.

Furthermore, the authors solve out the correlation coefficient between the nine indicators and passenger and freight transport impact on GDP in every city. Finding five indicators play an important role in it that illustrates these indicators are important when we analysis the issue. The survey is recommended to refine these indicators and to look into effecting factors as much as possible. At the same time, two indicators in it differ from each other for the differences of the level of productivity, so we can make the different surveys according to the city differences. Coupled with these, we can delete some items having little impact on the GDP growth in poor condition. At last, to increase the accurate of it, the survey is suggested to add the item 'satisfaction', adjust the model parameters, investigate the consuming condition and so on.

At the end of the essay, by using support vector machine (SVM) method to verify the reliability of the model.

## 参赛队号#1248

## 1. 问题重述

交通运输作为国民经济载体，沟通生产和消费，在经济发展中扮演者极其重要的角色。纵观几百年来交通运输与经济发展的相互关系，生产水平越高，就越要求基础结构超前发展。工业化时期的基础结构，已经不允许交通运输滞后。进入现代化社会，经济社会对交通运输的要求本质上就是超前的，交通运输是国民经济的先行官，发展经济，交通先行，是经济发展的内在规律。公路运输是在公路上运送旅客和货物的运输方式，是交通运输系统的组成部分之一，主要承担短途客货运输。发展公路运输对国内生产总值（GDP）增长的贡献产生于交通运输和客货运输两个阶段，表现为公路运输对国民经济的直接贡献、波及效果。对于相关行业的直接消费以及创造就业机会等几个方面。

某省的统计部门想通过调查研究的方法估计公路运输业对于 GDP 的影响，通过随机发放问卷，获得了附件 1 中所示的数据，该数据为真实调查所得到的原始数据。请参照该数据完成如下问题：

问题 1：请你建立合理的模型，估计该省公路运输业对于 GDP 的影响。

问题 2：考虑所获得数据的情况，如果由你来设计调查项目，为了提高问题 1 中模型的精度，需要对现有调查项目做哪些调整，并陈述理由。

## 2. 问题分析

本题着眼于公路运输业与国民生产总值之间的联系。公路运输业是国民经济中的基础产业，它与国民经济其他产业相互依存、紧密联系，公路运输业的发展依赖于其他产业的发展，也促进其他产业的发展。这种相辅相成的密切关系，既说明国民经济其他产业的发展会对公路运输业和国民经济的发展作出贡献，也表明公路运输业的发展会对其他部门和国民经济的发展作出贡献。公路运输系统是一个复杂的系统，可以从不同的层次、不同的阶段研究其对国民经济的贡献[1]。

问题一：

第一问是建立模型分析该省公路运输业对于 GDP 的影响。依据题干所给信息，将公路运输业对 GDP 增长的贡献分为交通建设和客货运输两个阶段。第一阶段，我们又考虑公路运输业对 GDP 的直接贡献、波及效果、对于相关行业的直接消费等三个方面，其中波及效果可分为后向波及效果和前向波及效果。我们先利用投入产出方法和产业关联理论分别建立分析各个方面对 GDP 的贡献模型，最后综合考虑各个方面，建立分析公路运输业对 GDP 总贡献的模型。

第二阶段，分析客货运输对 GDP 的影响，考虑到该省各个城市的生产力水平差异，其经济水平、产业结构、消费水平等方面均存在差异，故不同生产力水平的城市的客货运输对 GDP 的贡献不尽相同，也不宜用统一的指标去衡量，故考虑将 12 个城市分类，其次对城市分类后将客货运输对 GDP 的影响细分为很多方面，且其对 GDP 的影响难以直接计算，可考虑用模糊综合评价进行分析，其权重的确定可采用层次分析法。

问题二：

第二问是为了提高模型的精确度，要对那些调查项进行调整。可以看出第二问是在第一问的基础上对模型进行推广，改进。要提高模型精度，我们可以从细化或增加调查项、调整模型参数、分析对 GDP 具体表现的影响等方面入手。我们先用相关系数法计算评价指标与不同发达程度城市的相关系数，分析哪些指标对

## 参赛队号#1248

GDP 贡献显著，对于贡献显著的方面，调查细化，贡献小的方面，可弱化其调查项目，同时可计及城市的差异，对不同发展水平的城市，设定不同的调查表格。还可以通过调查数据估计模型参数，增加未考虑项等方法来达到提高模型精度的要求。

此外，虽然本调查样本是真实的，但不代表所有受访者给出的调查数据都是符合实际情况的，对于异常数据所在样本应予以剔除。

### 3. 模型假设

1. 假设该省的人口基本保持不变，没有大规模的迁移
2. 假设调查样本在该省各个城市是随机分布
3. 假设没有政策性的变化，没有对公路进行大规模的整改
4. 假设附件 2 中各生产部门是保持有计划按比例协调发展的

### 4. 符号说明

$Z_j$ : 某部门  $j$  单位产值所引起的 GDP 的增值

$a_{ij}$ : 某部门  $j$  单位产值所需要的其它部门  $i$  的中间投入产值

$b_{ij}$ : 某部门  $j$  单位产品产值需要完全消耗部门  $i$  的产品产值

$c_{ij}$ : 某部门增加一个单位的最终使用对各产业部门的完全需要量

$\{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9\}$ : 交通运输基础设施经营、交通运输自身的产值、加

工制造业、贸易、金融业、第三产业（服务业）、修理业、直接就业机会、间接就业机会

$B$ : 模糊综合评价向量

$FCI$ : 计算综合指数

参赛队号#1248

## 5. 模型的建立与求解

### 5.1 问题一

第一问是建立模型分析该省公路运输业对于 GDP 的影响。依据题干所给信息，将公路运输业对 GDP 增长的贡献分为交通建设和客货运输两个阶段，下分别研究这两个阶段对该省 GDP 的影响。

#### 5.1.1 模型一（产业关联理论模型）

##### <1>初步分析

公路交通建设与国民经济的其它产业相互依存，紧密联系，交通建设的发展依赖于其它产业的发展也促进其它产业的发展，从而促进国民生产总值的增长。主要表现为本身对国民生产总值的直接贡献、波及效果、对相关部门的直接消费，创造就业机会四个方面。其中波及效果[2]分为后向波及效果和前向波及效果。

我们以附件 2：2007 公路建筑业投入产出表(43 部门)为研究样本，以投入产出方法为基础，结合产业关联模型，利用 MATLAB 进行编程先得到四个方面各自对 GDP 做出的贡献，分析比较后得到交通建设对 GDP 做出的总贡献。

我们要研究的是交通建设业对 GDP 的影响，而附件表格中分别给出的公路建设业和交通运输及其仓储业两个部门的数据，我们可以将这两个部门合为一个部门，即交通建设业。

利用附件 2 数据抽象出投入产出表：

投入	中间产出				最终产出			净出口			总产出	社会总需求	
	部门 1	部门 2	...	部门 n	中间使用合计	资本形成总额	最终消费	合计	进口	出口			合计
部门 1	$x_{11}$	$x_{12}$		$x_{1n}$	$\sum x_{1j}$	$y_{11}$	$y_{12}$	$Y_1$	$p_{11}$	$p_{12}$	$P_1$	$X_1$	$Q_1$
部门 2	$x_{21}$	$x_{22}$		$x_{2n}$	$\sum x_{2j}$	$y_{21}$	$y_{22}$	$Y_2$	$p_{21}$	$p_{22}$	$P_2$	$X_2$	$Q_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮					
部门 n	$x_{n1}$	$x_{n2}$		$x_{nn}$	$\sum x_{nj}$	$y_{n1}$	$y_{n2}$	$Y_n$	$p_{n1}$	$p_{n2}$	$P_n$	$X_n$	$Q_n$
中间投入合计	$\sum x_{i1}$	$\sum x_{i2}$		$\sum x_{in}$	$\sum x_{ij}$	$\sum y_{i1}$	$\sum y_{i2}$	$\sum Y_i$	$\sum p_{i1}$	$\sum p_{i2}$	$\sum P_i$	$\sum X_i$	$\sum Q_i$
增加值	$g_1$	$g_2$		$g_n$	$\sum g_j$								
总投入	$X_1$	$X_2$		$X_n$	$\sum X_i$								

根据百度百科[3]的知识，可知该表利用了 GDP 计算的两种方法分别是：纵列是收入法，横列是支出法。

#### ◆ 收入法：

总投入=中间投入合计+劳动者报酬+生产税净额+固定资产折旧+营业盈余  
 营业盈余=增加值—劳动者报酬—生产税净额—固定资产折旧

$$\text{所以：} \sum X_j = \sum x_{ij} + \sum g_j$$

## 参赛队号#1248

## ◆ 支出法：

总产出=中间使用合计+资本形成总额+最终消费+出口—进口

$$\sum X_i = \sum x_{ij} + \sum Y_i + \sum P_i$$

所以： 
$$\sum P_i = \sum P_{i2} - \sum P_{i1}$$

另外： 
$$\sum Q_i = \sum x_{ij} + \sum Y_i + \sum P_{i1} + \sum P_{i2}$$

初步结论：通过表中的数据可以初步看出：两种计算方法所求得 GDP 是近似相等的，公路建设业对 GDP 是有贡献的；可以初步得到各个物理量如上述所示的关系；比较 X、Q 列数值发现各行业都存在一定的供给关系的不平衡，有升值空间，能够增加就业机会。

## &lt;2&gt;模型建立

## 1) 模型分析的基本系数

① GDP 增值系数  $Z_j$ 

该系数是指某部门 j 单位产值所引起的国民生产总值增值，即  $Z_j = \frac{g_j}{X_j}$ 。各

部门国民生产总值增值系数所组成的向量可表示为：  $Z = (Z_1, Z_2, \dots, Z_n)^T$ 。

② 直接消耗系数  $a_{ij}$ 

该系数是指某部门 j 单位产值所需要的其他部门 i 的中间投入产值。据上表： $a_{ij} = x_{ij} / X_j (i, j = 1, 2, \dots, n)$ 。由  $a_{ij}$  所组成的矩阵 A 即为直接消耗系数矩阵，它反映了部门各种产品之间的经济联系，是考核各部门经营情况是否改善的一个重要技术经济指标。式中： $a_{ij}$  为部门 j 所需要的部门 i 的中间投入产值； $X_j$  为部门 j 的生产总值。

③ 完全消耗系数  $b_{ij}$ 

部门 j 在生产中除了要直接消耗部门 i 的产品以外，还会通过消耗其他各部门的产品形成对部门 i 的间接消耗。所有直接消耗与间接消耗之和，就构成了部门 j 的产品对部门 i 产品的完全消耗。

完全消耗系数  $b_{ij}$  是指部门 j 单位产品产值需要完全消耗部门 i 的产品产值，即

$$b_{ij} = \sum_{k=1}^n a_{ik} b_{kj} + \sum_{s=1}^n \sum_{k=1}^n a_{is} a_{sk} a_{kj} + \dots$$



## 参赛队号#1248

完全消耗系数矩阵  $B = (I - A)^{-1} - I$ 。

④ 逆阵系数  $c_{ij}$ 

逆阵 C 即为列昂捷夫矩阵  $(I - A)$  的逆矩阵  $(I - A)^{-1}$ 。 $(I - A)^{-1}$  中的每一个元素即为逆阵系数  $c_{ij}$ ，它表示某部门增加一个单位的最终使用对各产业部门的完全需要量。

## 2) 公路运输业对 GDP 的直接效果分析

公路运输业对 GDP 的直接效果是指公路运输业本身对 GDP 所作的净贡献，其计算公式为

$$d_k = Z^T \Delta X$$

式中： $\Delta X = (0, 0, \dots, \Delta X, \dots, 0)^T$  为各部门产值增值向量，该处仅考虑公路运输部门 k 的直接效益，故其他各部门产值增值为零。

## 3) 运输业对 GDP 的波及效果分析

所谓公路运输业波及，是指国民经济产业体系中，公路运输产值的变化会沿着不同的产业关联方式，引起与其直接相关的产业部门产值的变化，并且这些相关产业部门产值的变化又会导致与其直接相关的其他产业部门产值的变化，依次传递，这一过程就是波及。这种波及对 GDP 的影响，就是公路运输业的波及效果。

## ① 后向波及效果

我们把公路运输业与那些提供公路运输生产所需的中间产品的部门之间的联系称为公路运输业的后向波及，把公路运输业这种因需要其他部门的产品作为自己中间投入而产生的波及效果的总和称为后向波及效果。

可利用投入产出原理中的完全消耗系数矩阵来计算后向波及效果。

根据完全消耗系数矩阵，我们可得出公路运输部门 k 单位产值需要其他中间投入部门的产品价值，分别为：

$$W = (B_{1j}, \dots, B_{nj}) \Delta x_k$$

其矩阵形式为  $W = B \bullet \Delta X$ 。

式中：B 为完全消耗系数矩阵； $\Delta X$  为各部门产值增值向量，根据 GDP 增值系数 Z，可以得出公路运输部门 k 产值增加  $\Delta x_k$  后，所引起的其他部门 GDP 增值，即公路运输部门 k 的后向波及效果为：

$$b_k = Z^T W = Z^T B \Delta X$$

由于  $B = (I - A)^{-1} - I$ ，A 为直接消耗系数矩阵，代入上式得

$$b_k = Z^T (I - A)^{-1} \Delta X - Z^T \Delta X$$

## 参赛队号#1248

## ② 前向波及效果

我们把公路运输与那些以公路运输生产为其中间投入的部门之间的关系，称为公路运输业的前向波及，把运输业因这种充当其他部门的中间投入而产生的波及效果的总和称为前向波及效果。

公路运输部门  $k$  增加产值为  $\Delta x_k$  时，部门  $j$  需要的公路运输部门  $k$  产值为：

$$u_j = \begin{cases} \frac{x_{kj} \Delta x_k}{x_{kj} - x_{kk}} & j=1, \dots, n, j \neq k \\ 0 & j=k \end{cases}$$

这些部门得到运输部门  $k$  产值的一部分增值后，即可扩大生产，这时部门  $j$  所能增加的产值为：

$$\Delta x_j' = \begin{cases} \frac{u_j}{a_{kj}} & a_{kj} \neq 0 \\ 0 & a_{kj} = 0 \end{cases}$$

式中： $a_{kj}$  为部门  $j$  单位产值所需要运输部门  $k$  的中间投入产值。有公路运输部门  $k$  前向波及导致的各部门产值增值相应为

$$\Delta x' = (\Delta x_1', \Delta x_2', \dots, \Delta x_n')^T$$

故各部门所能创造的 GDP 为  $Z^T \Delta x'$ 。

上述各部门扩大生产，除以运输部门为其中间投入以外，还需其他相关部门产品作为中间投入。这样，这些部门也存在着各自后向波及效果，即为  $Z^T B \Delta x'$ 。

因此，运输业的前向波及效果  $f_k$  为

$$f_k = Z^T \Delta x' + Z^T B \Delta x' = Z^T (I - A)^{-1} \Delta x'$$

## 4) 消费波及效果

消费波及效果是指上述三项效果由于消费  $c$  而引起的个生产部门所创造的 GDP 增值。

根据凯恩斯乘数原理，如果投资增加一个单位，GDP 增值将增加  $1-c$  个单位（其中  $c$  为消费乘数， $0 < c < 1$ ，这里我们取  $c$  为 0.5）。

上述三项效果所引起的消费量为  $(d_k + f_k + b_k)c$ 。由于这些消费作用而引起的 GDP 增值，即消费波及效果  $c_k$  为： $(d_k + f_k + b_k) \frac{c}{1-c}$

通过分析运输业对 GDP 的直接效果和波及效果，可以得出公路运输业对 GDP 的贡献，记为  $g_k$ ，则：

## 参赛队号#1248

$$g_k = d_k + f_k + b_k + c_k = X^T(I - A)^{-1}(\Delta X + \Delta x') \frac{c}{1 - c}$$

## &lt;3&gt;模型求解

根据以上模型，我们计算得到交通建设对 GDP 的贡献  $g_k=6050240$ ，而这并不能看出交通建设对 GDP 贡献的大小。进而，我们根据同样的方法计算其余部门的  $g_k$  值，如下表：

各部门对 GDP 的贡献表 ( $\times 10^5$ )

部门	1	2	3	4	5	6	7	8	9	10
$g_k$	62.09	48.18	4.79	21.78	7.11	45.78	7.26	1.48	3.29	2.75
部门	11	12	13	14	15	16	17	18	19	20
$g_k$	1.26	17.25	14.78	60.69	0.12	1.83	2.72	1.10	1.56	0.15
部门	21	22	23	24	25	26	交通建设	28	29	30
$g_k$	0.67	0.06	38.71	3.17	0.71	36.93	60.50	0.28	7.89	43.11
部门	31	32	33	34	35	36	37	38	39	40
$g_k$	7.88	28.40	14.17	8.07	1.89	2.88	3.08	12.42	12.56	14.27
部门	41	42								
$g_k$	4.73	29.42								

结论：通过比较可知，交通建设对 GDP 的贡献在所有部门中排第三，影响较突出，说明交通建设对 GDP 的贡献很大。

## 5.1.2 模型二（基于 K 均值分类的城市分类模型）

## &lt;1&gt;数据预处理

观察附件一中 2012 年公路运输调查数据，一些数据显示异常，出现异常数据的原因可能是数据录入时漏录、错录，被调查者提供了虚假的信息，在数据保存过程中丢失等。对于这些错误数据，我们要予以剔除。

异常数据的判别方法：对于每个城区，如果某个样本调查项目值远大于均值（一般取  $x_i > 15\mu$  或  $x_i > 10\mu$ ），则认为该数偏离过大，被判定为异常数据。

对于附录一中的大量数据，通过编程（详见附录）对数据首先进行筛选，按照数值大于平均值 10 倍或 15 倍的方式筛选出其中的坏数据并与表中的数据进行核查，发现购置金额异常的行数是：354、1189、3401；运载范围异常的行数是：

## 参赛队号#1248

295；全年运货量异常的行数是：342、587、706、727、1680、3255；按照每个地区的平均值进行数值大于每个地区的平均值的 10 倍进行筛选出路途堵塞停驶天数异常的行数是：1646、3196；将这些坏数据以及未填写的数据删除后进行数据的进一步分析。

## &lt;2&gt;模型建立

分析客货运输对 GDP 的影响，考虑到该省各个城市的生产力水平差异，其经济水平、产业结构、消费水平等方面均存在差异，故不同生产力水平的城市的客货运输对 GDP 的贡献不尽相同，也不宜用统一的指标去衡量，故考虑将 12 个城市分类。

## 1) 聚类分析变量的选择

根据反应城市发达程度的情况，我们要选择聚类分析的变量。购置金额反应了当地的物价水平，城市越发达的物价越高，购置金额可以作为变量；全年运货量反应了当地物流水平，物流水平与城市发达程度成正相关，所以全年运货量可以作为一变量；路途堵塞停放天数反应城市车辆数量，也与发达程度有关，可以作为变量之一；产权形式反应了城市的经济结构，一般认为国有所占比重越高城市越发达，产权形式可以作为一个变量；年运输收入、运输所得报酬、收入盈余三项都反应了人们收入状况，可以将它们的和归为一个变量收入；另外，途中住宿花销、途中餐饮花销、途中其他花销都属于途中花费，反应城市消费水平、消费观念等，将总的途中花销作为变量之一；还有燃油消耗、过路费、过桥费占用运输费用比例更换润滑油、滤清、防冻等费用、更换轮胎费用、机械故障、更换零部件等花销、正常保养费用、特殊原因费用、图中通讯费用、车辆折旧费、备用零部件支出、罚没款支出等可以看成一大类，为维护费用，反应了车辆使用频率等，将它们的和作为一变量。这样，我们选择了 7 个与城市发达程度最相关的因素来作为聚类分析变量，分别依次记为  $x_1$ 、 $x_2$ 、 $x_3$ 、 $x_4$ 、 $x_5$ 、 $x_6$ 、 $x_7$ 。

## 2) 数据处理

先统计每个城市的 7 个变量的平均值：

$$\bar{x}_{ij} = \frac{\sum_{j=1}^N x_{ij}}{N}$$

其中， $x_{ij}$  表示第  $i$  ( $i=1, 2, 3, 4, 5, 6, 7$ ) 个选择变量第  $j$  个值， $N$  表示剔除错误数据后剩下的样本容量。

因为各聚类变量量纲不同，数据差异较大，所以需对聚类变量均值做标准化处理。将各变量值转换成标准化变量  $\tilde{x}_{ij}$ ，有：

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{s_j}$$

其中， $x_{ij}$  意义与上述相同， $s_j = \sqrt{\frac{1}{N-1} \sum_{j=1}^N (x_{ij} - \bar{x}_{ij})^2}$  ( $N$  意义与上述相同)。

## 参赛队号#1248

标准化结果如下（原均值样本见附录）：

各城市聚类变量标准化数据

城市编号	购置金额(千元)	全年运货量(吨)	路途堵塞停驶天数(天)	产权形式(1、国有 2、集体 3、个体私营 4、承包租赁)	收入(十万元)	途中花费(万元)	维护费用(十万元)
城市1	0.3583	-0.4115	1.4109	0.2983	0.9408	0.2707	0.4798
城市2	0.5596	2.4618	-0.2236	-2.6434	1.0855	1.4621	1.2114
城市3	0.1689	-0.6648	2.1607	0.0385	0.3083	0.0972	0.6338
城市4	-0.9508	-0.5243	-0.5310	-0.9010	1.1303	0.4771	0.3337
城市5	0.9001	-0.6606	-1.0071	0.4552	-1.3420	-1.1208	-1.4708
城市6	0.8057	-0.0857	-0.6601	1.5180	-0.2809	-1.3959	-0.9325
城市7	-1.8985	-0.6620	-0.6798	0.6284	-1.5858	-1.3373	-1.4520
城市8	-0.1751	-0.2196	0.1796	0.0038	0.2878	1.4066	1.0291
城市9	0.5427	1.5772	0.9491	-0.0003	0.1596	0.0262	0.6547
城市22	-1.4489	-0.7277	-0.9585	0.2642	-1.5527	-0.6440	-1.2787
城市25	-0.2940	-0.2677	-0.1781	0.3650	0.0438	-0.2833	0.0703
城市29	1.4321	0.1849	-0.4620	-0.0267	0.8052	1.0414	0.7214

### <3>模型求解

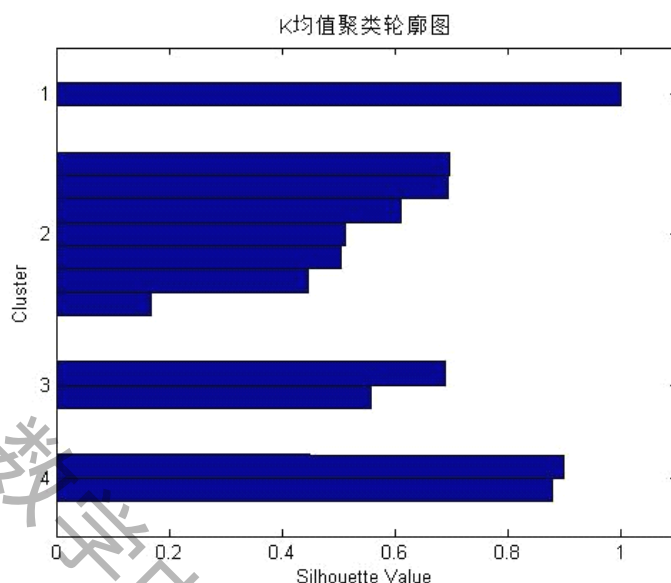
用 K 均值聚类法聚类[4]。我们选取高效、快速的 K 均值聚类法对城市进行聚类分析，K 均值聚类法步骤如下：

- ① 选择 k 个样品作为初始凝聚点（我们选择的是城市 1、3、5、10），或者将所有样品分成 k 个初始类，然后将 k 个类的重心（均值，这里取城市得分）作为初始凝聚点。
- ② 对除凝聚点之外的所有样品逐个归类，将每个样品归入离它最近的凝聚点所在的类，该类的凝聚点更新为这一类目前的均值，直至所有样品都归了类。
- ③ 重复上一步骤，直至所有样品都不能再分配为止。

#### 1) 聚类分析求解

聚类结果为：

## 参赛队号#1248



每个观测的轮廓值都是正的，仅有一个处于 0.2 附近，其余都在 0.4 以上，说明将 12 个城市分为 4 类是非常合适的。

得到城市分类为：

A 类：城市 2

B 类：城市 1、3、4、8、9、25、29

C 类：城市 5、6

D 类：城市 7、22

## 2) 基于层次分析法的比较各城市差异的模型

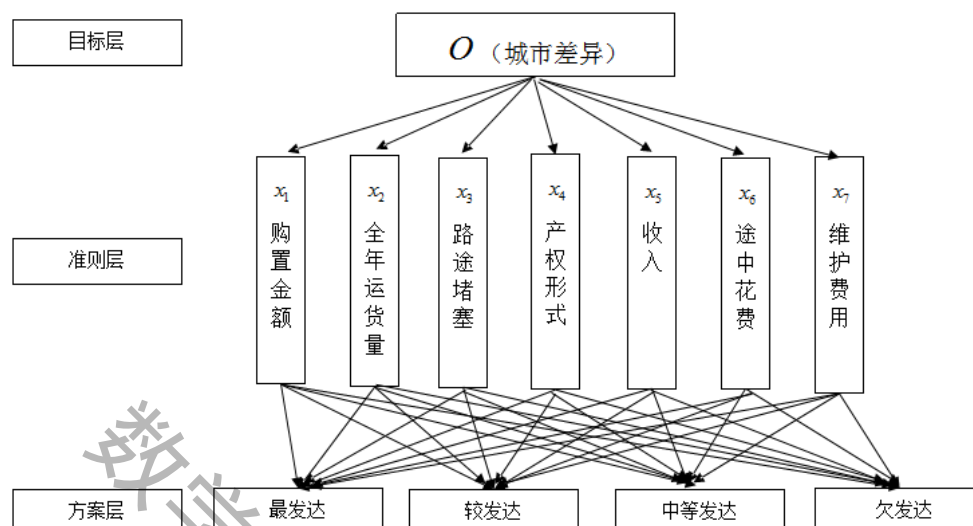
聚类仅把结果聚于 4 类，但不知道每类对于城市相对的生产力水平下，故构造每个城市的综合评价指标，得出每个城市得分（记为  $y$ ）。由于我们只需大概的区别各城市的发达程度，所以我们可以粗略的将各个聚类变量的线性加权组合作为城市得分。除产权形式（国有、集体企业越多，城市生产力水平越高）外，每个聚类变量与城市发达程度成正相关，至于各权重变量可用层次分析法获得。

层次分析法[5]的基本步骤：

- ① 建立层次分析结构模型，结合实际，将有关因素自上而下分为：目标层、准则层、方案层。
- ② 构造成对比较阵，用成对比较法和 1-9 尺度的方法。
- ③ 计算权向量并通过计算特征跟和特征向量做一致性检验。
- ④ 计算组合权向量作组合一致性检验，组合权向量可作为决策的定量依据。

### (1) 建立层次分析模型

## 参赛队号#1248



(2) 构造成对比较阵。两两元素进行对比，对比采用相对尺度。设要比较各准则： $x_1$ ,  $x_2$ ,  $x_3$ ,  $x_4$ ,  $x_5$ ,  $x_6$ ,  $x_7$  对目标  $O$  的重要性。

$$x_i : x_j \rightleftharpoons x_{ij}$$

$$A = (a_{ij})_{n \times n}, a_{ij} > 0, a_{ji} = \frac{1}{a_{ij}}$$

尺度采用 1-9 尺度—— $a_{ij}$  取值 1, 2, ..., 9 及其倒数 1, 1/2, ..., 1/9, 便于定性到定量的转化。详细见相对尺度参照表：

尺度 $a_{ij}$	1	2	3	4	5	6	7	8	9
$x_i : x_j$ 的重要性	相同		稍强		强		明显强		绝对强

得到准则层对于目标的对比阵矩阵：

$$A = \begin{bmatrix} 1 & 2 & 3 & 3 & 1 & 4 & 4 \\ \frac{1}{2} & 1 & 2 & 1 & \frac{1}{2} & 2 & 2 \\ \frac{1}{3} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & 1 & 1 \\ \frac{1}{3} & 1 & 2 & 1 & \frac{1}{2} & 2 & 2 \\ 1 & 2 & 3 & 2 & 1 & 3 & 3 \\ \frac{1}{4} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & 1 & 1 \\ \frac{1}{4} & \frac{1}{2} & 1 & \frac{1}{2} & \frac{1}{3} & 1 & 1 \end{bmatrix}$$

方案层对准则层的对比阵同样采取上述方法，不再详细说明。

(3) 计算权向量并做一致性检验。

参赛队号#1248

$$\lambda_{\max} = 7.0421, CI = \frac{\lambda_{\max} - n}{n - 1} = 0.007 < 1.32$$

$$w = [0.2792 \quad 0.1357 \quad 0.0740 \quad 0.1291 \quad 0.2408 \quad 0.0707 \quad 0.0707]$$

通过一致性检验。

由于  $x_4$  与  $y_1$  负相关，故：

$$y_1 = 0.2792 * x_1 + 0.1357 * x_2 + 0.0740 * x_3 - 0.1291 * x_4 + 0.2408 * x_5 + 0.0707 * x_6 + 0.0707 * x_7$$

根据此公式计算各城市得分如下：

城市得分	得分	对应等级
城市1	0.3897	B
城市2	1.2651	A
城市3	0.2377	B
城市4	0.0699	B
城市5	-0.4779	C
城市6	-0.2637	C
城市7	-1.3302	D
城市8	0.1756	B
城市9	0.5223	B
城市22	-1.1179	D
城市25	-0.1832	B
城市29	0.7126	B

由此可以看出 A 类属于生产力水平最高，属于发达城市，同理 B, C, D 分别对应于较发达、中度发达、欠发达（这里所说的发达程度仅指该省各城市相互间的比较，并非全国意义上的），与实际情况符合，即每个省很发达及欠发达城市较少，多数城市处于较发达或中度发达生产力水平。并且综合各城市聚类结果及生产力水平得分，可以快速看出该聚类及各类生产力水平的判定模型是正确可靠的。

### 5.1.3 模型三（基于 AHP 的模糊综合评价模型）

模型二中已将该省 12 个城市分为发达、较发达、中度发达、欠发达 4 类，由于城市生产力水平的差异，各城市的经济水平、产业结构、消费能力等方面均有差异，继而不能以统一的指标衡量客货运输对 GDP 的增长贡献，因而本模型分别考虑 4 类地区客货运输对 GDP 的增长贡献，评估其影响力。

#### <1>数据预处理

① 剔除模型二中找出的异常数据所在样本（有一项调查不符合常情，即可认为该样本不够真实可靠）。

② 对于存在数据缺失的样本，虽难以认定其所在样本是否绝对可靠，但考虑到数据缺失的令我们无法对该样本进行综合评价，且含数据缺失的样本相对于总样本的量较



## 参赛队号#1248

小，可不考虑存在数据缺失的样本，予以剔除。

③ 分别对 4 类地区的样本调查值进行标准化处理，处理原则为：

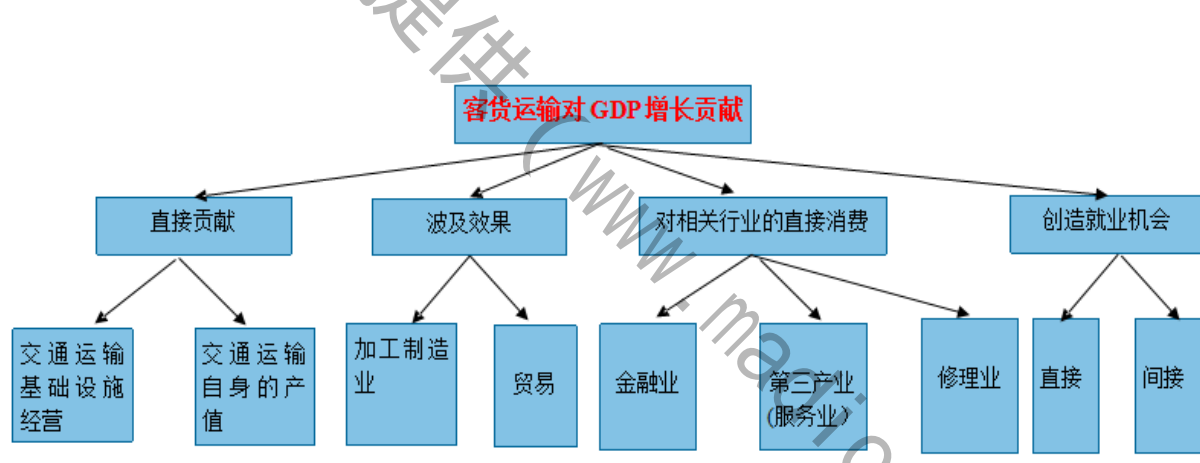
$$z_{bz} = \frac{z - z_{\min}}{z_{\max} - z_{\min}}$$

（其中  $z$  为某项调查值， $z_{bz}$  为标准化后的结果， $z_{\min}$  为  $z$  的最小值， $z_{\max}$  非绝对意义上的  $z$  最大值，考虑到某些样本值过大， $z_{\max}$  取大于 95% 调查值，当  $z > z_{\max}$  时，取  $z_{bz} = 1$ ）

## <2>建立模糊综合指数评价模型

### 1) 评价模型的确立

公路运输调查数据中，调查项目种类繁多，直接考虑如此繁多的项目对 GDP 的影响过于复杂，且各自权重难以确定。由题意及现有资料可知，公路运输对 GDP 的影响可分为 4 个部分：直接贡献，波及效果，对相关行业的直接影响，创造就业。而各类调查项目可归为这 4 个部分，同时，难以判定某些调查项目对 GDP 的影响，对这些调查项目可予以忽略。具体划分如下图表：



选取第 3 级的 9 个指标（交通运输基础设施经营、交通运输自身的产值、加工制造业、贸易、金融业、第三产业（服务业）、修理业、直接就业机会、间接就业机会）作为模糊综合评价因子集，理由如下：

- ① 对于第 2 级的 4 个指标，难以直接求出其影响值；
- ② 直接考虑繁多的调查项目对 GDP 的影响过于复杂，且各自权重难以确定。

9 个指标对应的调查项目为：

## 参赛队号#1248

调查项目 评价指标	I	II	III	IV
交通运输基础设施经营	购置金额	过路过桥费所占运输费用比例		
交通运输自身的价值	年运输收入	客座的利用率	每年实际工作天数	
加工制造业	燃油消耗	车辆折旧费		
贸易	平均载货量	全年运货量		
金融业	结算形式	现金结算比例		
第三产业（服务业）	途中住宿花销	途中餐饮花销	途中通讯费	车辆保险
修理业	更换润滑油	更换轮胎	更换零部件	正常保养费
直接因素	产权形式	驾驶人数		
间接因素	收益盈余合伙人分成	罚款支出		

## 2) 模糊综合评价模型的求解

## (1) 说明几个概念[6]:

①因素集： $U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7, u_8, u_9\} = \{\text{交通运输基础设施经营、交通运输自身的产值、加工制造业、贸易、金融业、第三产业（服务业）、修理业、直接就业机会、间接就业机会}\}$

②评价集：因素集的不同评价等级组成的集合。

$V = \{v_1, v_2, v_3, v_4\} = \{I, II, III, IV\}$  I, II, III, IV 分别代表对 GDP 影响很显著、较为显著、显著性一般、显著性较弱。

③i 对 j 的隶属度：第 i 个评价指标可以被评第 j 类的可能。例如， $r_{11}$  表示对 I 类的隶属程度。

④模糊关系矩阵：由隶属度组成的矩阵： $R = (r_{ij})_{N \times 4}$  (N 表示各类地区样本数)

⑤权重向量，由各因素所占权重组成：

$$W = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9\}$$

⑥模糊综合评价向量：

$$B = W \bullet R \quad (\text{选用 } M(\bullet, +) \text{ 算子, 即加权平均型, 对所有因素综合均衡考虑})$$

⑦计算综合指数 FCI

$$FCI = B \cdot T \quad (\text{其中 } T = [t_1, t_2, t_3, t_4]^T)。$$

## (2) 基于 AHP 法确定各模糊因素权重

AHP（层次分析法）在模型二中已详细叙述，直接给出最终求解结果。

①先考虑第 2 级对第一级影响（对 GDP 影响：直接贡献 > 相关行业消费 = 创造就业机会 > 波及效果）

## 参赛队号#1248

对比矩阵为：

$$A = \begin{bmatrix} 1 & 4 & 3 & 4 \\ \frac{1}{4} & 1 & \frac{1}{2} & 1 \\ \frac{1}{3} & 2 & 1 & 2 \\ \frac{1}{4} & 1 & \frac{1}{2} & 1 \end{bmatrix}$$

$$\lambda_{\max} = 4.1107$$

$CI = 0.0369 < 0.90$  (通过一致性检验)

$$w = [w_A \quad w_B \quad w_C \quad w_D] = [0.5319 \quad 0.1197 \quad 0.2179 \quad 0.1305]$$

(其中  $w_A$ 、 $w_B$ 、 $w_C$ 、 $w_D$  分别表示客货运输、波及效果、对相关行业的直接消费，创造就业机会对 GDP 影响所占权重)。

②确定第三级对第二级的影响

依旧运用 AHP 法，这里直接给出最终权重（对比矩阵及一致性检验见附录）

$$w_{A1} = w_{A2} = 0.5$$

( $w_{A1}$ 、 $w_{A2}$  分别表示购置金额、过路过桥费对交通基础设施经营的影响权重)

同理可算出其余第 3 级对各自第 2 级的影响权重，结果为：

$$w_{B1} = w_{B2} = 0.5$$

$$[w_{C1} \quad w_{C2} \quad w_{C3}] = [0.1634 \quad 0.2970 \quad 0.5376]$$

$$[w_{D1} \quad w_{D2}] = [0.6667 \quad 0.3313]$$

③9 个指标对 GDP 影响权重的确定

9 个指标对 GDP 影响权重的矩阵  $\tilde{w}$  为

$$\begin{aligned} \tilde{w} &= [w_A * [w_{A1} \quad w_{A2}] \quad w_B * [w_{B1} \quad w_{B2}] \quad w_C * [w_{C1} \quad w_{C2} \quad w_{C3}] \quad w_D * [w_{D1} \quad w_{D2}]] \\ &= [0.2660 \quad 0.2660 \quad 0.0599 \quad 0.0599 \quad 0.0356 \quad 0.0674 \quad 0.1169 \quad 0.0870 \quad 0.0435]; \end{aligned}$$

④各评价指标等级的确定

很显然 9 个评价指标与待评价量成正相关性，从各类指标数据（数值从相关样本统计量线性加权平均后得到）分析可知，某些指标很大部分值与均值（或中位值）偏离程

## 参赛队号#1248

度很大，某些指标很小部分的值与均值（或中位值）偏离程度也很大，故选出各类指标的上限值 max2（大于 90%同类指标值）及下限值 min2（小于 90%同类指标值），然后将上限值、下限值均分五类，间距  $d = (\max 2 - \min 2) / 5$ ，则第 I 级的临界值为  $\max 2 - d$ ，第 II 级临界值为  $\max 2 - 2 * d$ ，第 III 级的临界值为  $\max 2 - 3 * d$ ，第 IV 级的临界值为  $\max 2 - 4 * d$ 。

得到各指标分级标准如下：

分级指标 \ 分级		所占权重%	I 级	II 级	III 级	IV 级
直接贡献(53.19%)	交通运输基础设备	26.5950	0.4974	0.3852	0.2731	0.1610
	交通运输自身产值	26.5950	0.4192	0.3186	0.2179	0.1173
波及效果(11.97%)	加工制造业	5.9850	0.5069	0.3848	0.2627	0.1406
	贸易	5.9850	0.2371	0.1778	0.1186	0.0593
对相关行业的直接消费(21.73%)	金融业	3.5605	0.8000	0.6000	0.4000	0.2000
	第三产业	6.4716	0.5071	0.3858	0.2645	0.1432
	修理业	11.6947	0.5444	0.4129	0.2814	0.1499
创造就业机会(13.11%)	直接就业机会	8.7000	0.7367	0.6400	0.5433	0.4467
	间接就业机会	4.3500	0.2462	0.1846	0.1231	0.0615

## ⑤模糊关系矩阵的确定

各类地区 9 个指标的数值对应的矩阵为 A，则 i 个样本指标值为

$$X_i = A(i,:) = (x_1, x_2, \dots, x_9)$$

$1 \leq i \leq 9$  时

设  $z(i, j)$  表示第 i 个指标对应第 j 级数值

对 I 级隶属函数：

$$r_{i1} = \begin{cases} 1, & x_i > zb(i,1) \\ \frac{x_i - zb(i,2)}{zb(i,1) - zb(i,2)}, & zb(i,2) < x_i < zb(i,1) \\ 0, & x_i < zb(i,2) \end{cases}$$

## 参赛队号#1248

对 II 级隶属函数：

$$r_{i2} = \begin{cases} \frac{zb(i,1) - x_i}{zb(i,1) - zb(i,2)}, & zb(i,2) \leq x_i \leq zb(i,1) \\ \frac{x_i - zb(i,3)}{zb(i,2) - zb(i,3)}, & zb(i,3) < x_i < zb(i,2) \\ 0, & x_i > zb(i,1) \text{ 或 } x_i < zb(i,3) \end{cases}$$

对 III 级隶属函数：

$$r_{i3} = \begin{cases} \frac{zb(i,2) - x_i}{zb(i,2) - zb(i,3)}, & zb(i,3) \leq x_i \leq zb(i,2) \\ \frac{x_i - zb(i,4)}{zb(i,3) - zb(i,4)}, & zb(i,4) < x_i < zb(i,3) \\ 0, & x_i > zb(i,2) \text{ 或 } x_i < zb(i,4) \end{cases}$$

对 IV 级隶属函数：

$$r_{i4} = \begin{cases} 0, & x_i \geq zb(i,3) \\ \frac{zb(i,3) - x_i}{zb(i,3) - zb(i,4)}, & zb(i,4) < x_i < zb(i,3) \\ 1, & x_i \leq zb(i,4) \end{cases}$$

### ⑥计算综合指数 FCI

根据日常经验取影响指数  $T=[4,3,2,1]^T$  (即影响显著模糊指数为 4, 影响较为显著为 3,

影响一般为 2, 影响较弱为 1)。

进而可计算出各类地区各样本的模糊指数值, 如较发达地区 1 的第 1 个样本,

其模糊综合评价向量为  $[0.3868 \ 0.0687 \ 0.0529 \ 0.4910]$ , 模糊指数为 2.35 (影响介于较显著和一般之间)。(所有样本的模糊指数见电子版附录)

各类城市的平均模数指数为：

各类城市	模糊指数
发达城市 (城市编号2)	2.85
较发达城市 (城市编号： 1, 3, 4, 8, 9, 25, 29)	3.01
中度发达城市 (城市编号：5, 6)	2.65
欠发达城市 (城市编号 7 22)	2.49

## 参赛队号#1248

由此可以看出，城市生产力水平越高，本质上越要求公路运输超前发展，公路运输对城市 GDP 的贡献也就越大（模糊指数：较发达城市>中度发达城市>欠发达城市），但随着生产力增长趋势饱和，公路运输对城市 GDP 的贡献也趋于饱和，甚至会小幅减小，这也是发达城市模糊指数小于较发达城市的重要原因。

此外，由各类地区的模糊指数可以看出，发达城市和欠发达城市客货运输对 GDP 的影响贴近于有较大显著性，中度发达和欠发达城市客货运输对 GDP 的影响介于有较大显著性和显著性一般之间，考虑到发达和较发达城市占了多数，且其 GDP 总量大，故可认为客货运输对 GDP 的影响较显著。

根据模型一的分析结果，该省交通建设对 GDP 的影响较显著（或很显著），从模型三中该省客货运输对 GDP 的影响为较显著，综合该省交通建设及客货运输对 GDP 的影响，可认为该省公路运输业对 GDP 的影响较为显著。

## 5.2 问题二

问题二要求为提高模型一的精度，对现有调查项目进行调整。模型一中公路运输业对 GDP 的影响评估主要是基于附件一及附件二，考虑到对交通建设的影响力评价模型较为成熟可靠，且附件二的调查数据内容难以改变，故以下只分析附件一中调查项目的调整。

## 5.2.1 模型四（相关性分析模型）

## &lt;1&gt;模型建立

问题一中已对客货运输对 GDP 的影响指数进行较为全面的评价，但实际中每类指标对 GDP 的影响程度可能不同，并且对生产力水平存在差异的各类城市，同类指标对 GDP 的影响程度也有可能不同，根据以上 2 种差异可对调查项目做更科学的调整。下对 9 类评价指标与影响指数进行相关性分析。

## 1) 数值分析：

记模糊评价指标变量  $u_j = (x_{1j}, x_{2j}, \dots, x_{nj})^T (j=1, 2, \dots, 4)$  [7]，城市模糊指数为 FCI

$$\text{根据公式 } r_{jk} = \frac{\sum_{i=1}^9 (u_{ij} - \bar{u}_j)(u_{ik} - \overline{FCI_k})}{\left[ \sum_{i=1}^9 (u_{ij} - \bar{u}_j)^2 \sum_{i=1}^9 (u_{ik} - \overline{FCI_k})^2 \right]^{\frac{1}{2}}}$$

针对相关系数  $r_{jk}$ ，我们一般认为其在 (0.75, 1) 内，相关性很强；在 (0.3, 0.75) 内，相关性较强；在 (0, 0.25) 内，相关性很弱。

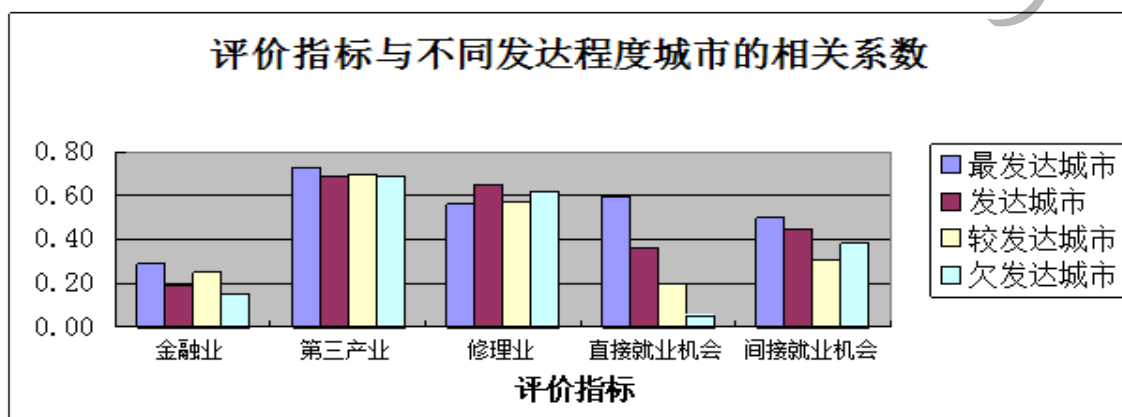
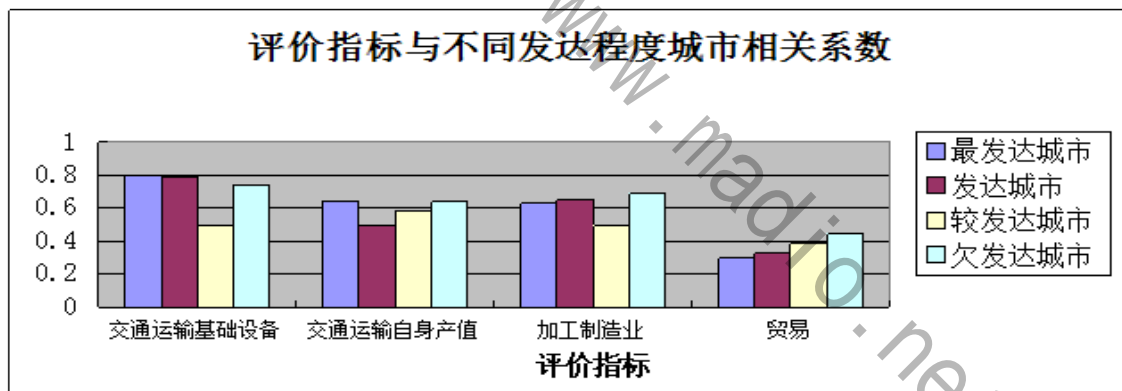
计算各分类组别与公路运输业相关的部分对不同发达程度城市影响的相关系数如下：

## 参赛队号#1248

类别 评价指标 \ 城市	最发达城市	发达城市	较发达城市	欠发达城市
交通运输基础设施	0.8010	0.7861	0.4902	0.7378
交通运输自身产值	0.6412	0.4981	0.5796	0.6387
加工制造业	0.6354	0.6456	0.4985	0.6938
贸易	0.2955	0.3287	0.3844	0.4468
金融业	0.2929	0.1860	0.2515	0.1493
第三产业	0.7312	0.6875	0.6963	0.6869
修理业	0.5613	0.6477	0.5717	0.6174
直接就业机会	0.5920	0.3589	0.1956	0.0584
间接就业机会	0.4999	0.4465	0.3097	0.3816

从表中我们可以看出大部分数据集中在(0.3, 0.5)之间, 相关性都较强, 而根据实际情况, 不同指标对城市 GDP 的影响是有差别的, 这种区分方法就屏蔽了一些信息。所以, 这里我们进一步划分, 认为相关系数  $r_{jk}$  在(0.75, 1)内, 相关性很强; 在(0.5, 0.75)内, 相关性较强; 在(0.3, 0.5)内, 相关性较弱。

## 2) 图表分析



## 参赛队号#1248

根据上述图表，我们可将 9 项模糊指标对 GDP 的影响分为 3 类，

①交通运输基础设施、交通运输自身产值、加工制造业、第三产业、修理业与不同发达城市的相关系数都很大，交通运输基础设施有两个相关系数已经超过 0.75，其他几乎都在 (0.5, 0.75) 内，说明与公路运输业相关的这些部门与城市发达程度联系最紧密，即它们对各自城市的 GDP 贡献最大，我们称为第一类。

交通运输基础设施、交通运输自身产值是公路运输业的最主要成分，其对 GDP 的贡献是很显然的；加工制造业的原材料、成品的运输都极大程度的依靠运输业，所以贡献也很大；运输业会推动人员的流动，促进第三产业的发展；运输业中的交通工具的维修必定会带动修理业发展。

②还可以看出与公路运输业相关的第三产业、修理业对各城市的相关系数基本持平，相关系数在 (0.5, 0.75) 之间，说明这两者对各城市的 GDP 贡献基本相同；而其他对各城市 GDP 贡献有一些差别；再者，贸易、金融类对各城市影响相对较小，毕竟，公路运输业涉及的是人员、货物的流通，与金融等相关的行业关联较小，我们称为第三类。

③此外公路运输业带来的直接和间接就业机会反映的是公路运输业带来的就业机会增加，对城市 GDP 的贡献一般，相关系数基本在 (0.3, 0.5) 内，我们称为第二类。但根据其针对不同发达程度的城市相关系数基本递减可知，城市越发达，公路运输业的发展所能提供的就业机会越多，说明这些评价指标包括的部门很有发展潜力，即随着城市发达水平的提高，这些部门会有较大发展，这也与我们的常识相一致。

## <2>调查项目的调整

①既然我们分析了不同类别指标与不同发达程度城市 GDP 的相关性不同，那在做数据调查的时候要针对不同发达程度城市设计不同问卷，对不同类别指标也应有所侧重点，这样才能达到提高模型精度的要求。具体如下：

第一类中的 5 项指标（分别为交通运输基础设施、交通运输自身产值、加工制造业、第三产业、修理业），对不同发达程度城市的 GDP 贡献都很大，而且很稳定，说明这些成分在公路运输业对城市 GDP 的贡献中占最主要地位，我们在制定调查问卷时应该予以充分考虑，尽量减少由于统计对这些指标的影响。我们可以对其进行细化，考虑尽可能多的影响因素，如在第三产业这一指标中加入“运输形式（本地或异地）”这一项，因为异地运输带来的服务业增值会多于本地运输带来的增值，通过调查分析这一项就可以更准确的得出运输业对服务业的贡献。

第二类（直接和间接就业机会）中的两项指标的影响也较大，但是对不同发达程度城市的影响差别也较大，我们可以根据不同发达程度城市对这一类指标调查情况进行调整。对该城市影响大的予以详细考虑，影响很小的可以粗略考虑。如直接就业机会对发达城市影响很大，就可以增加“提供的就业机会”等项的调查，而对于不发达城市，其影响很小，就可以适当缩减该项调查精度。

第三类（贸易、金融类）指标对各种发达程度的城市影响均很小，我们就可以将其中影响最小的一些项适当删除。在一定的样本容量的情况之下，这并不影响我们的分析结果。

②除此之外，我们还可以增加“满意度”这一项。因为人们对于公路运输业的满意程度也客观的反映了该行业的目前发展状况、未来发展前景、从事该行业的人员的收入



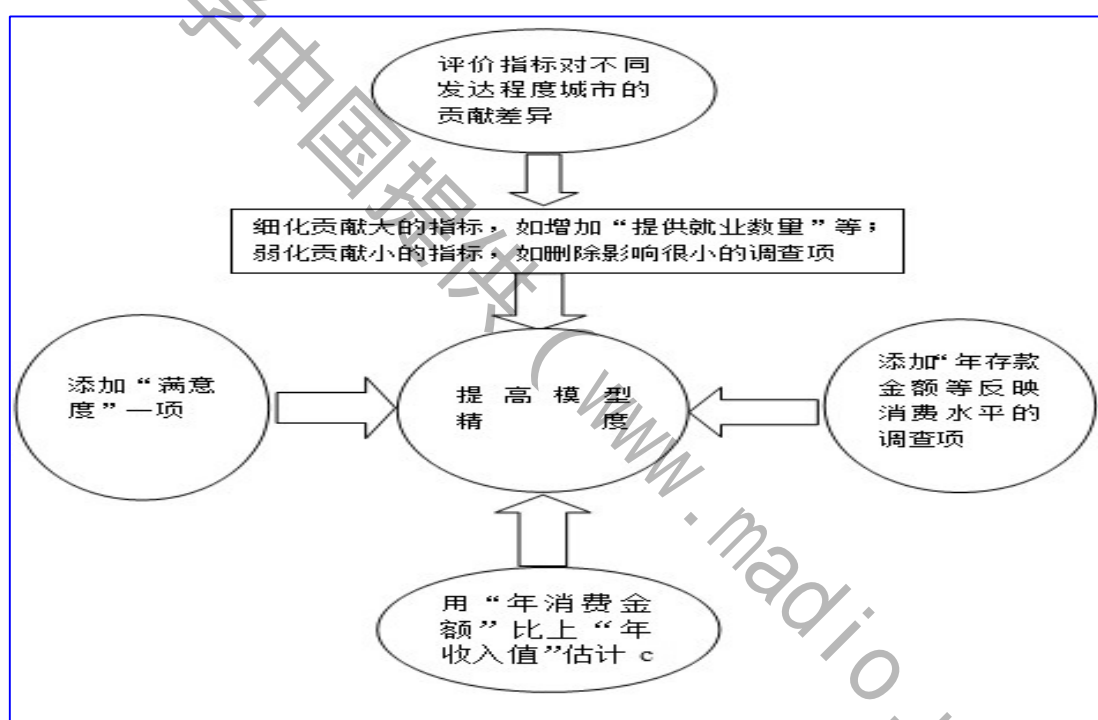
## 参赛队号#1248

状况、生活水平等，反过来，公路运输业的发展状况、人们的生活水平又会对 GDP 产生影响。我们可以认为满意度越高，对城市 GDP 的影响越大。

③我们还可以从调整模型参数上提供模型精度。通过增加调查一些项来估计模型参数，从而使参数更接近实际情况，也就提高了模型的精度。如在模型一中计算消费波及效果  $c_k$  时，人为的取了  $c=0.5$ ，我们可以在调查项目中增加“年消费金额”，通过“年消费金额”比上“年收入值”作为  $c$  的估计值来计算  $c_k$  会更精确。

④还可以增加“年存款金额”、“收入增长”等项目，依此来看出城市消费水平，可看成公路运输业带来的后续消费。因为消费水平，收入增长率都是反映 GDP 的重要指标。

近似调整模型如下：



## 6. 基于支持向量机的模型检验

### 6.1 支持向量机的模型检验的介绍

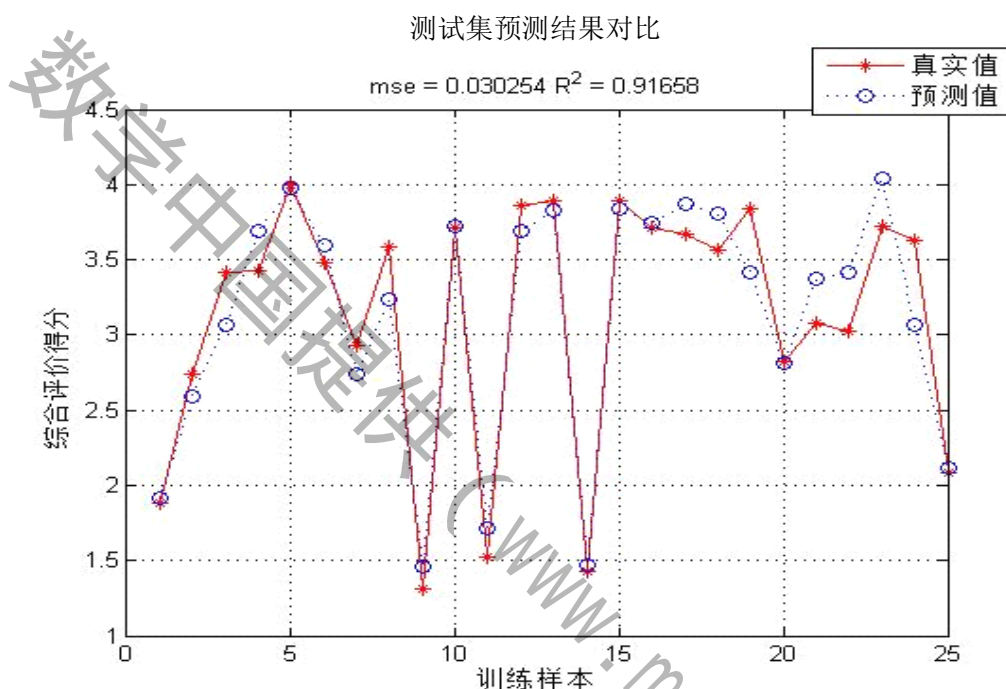
支持向量机[8]是一种新的机器学习算法，其基础是统计学习理论采用结构风险最小化原则。与传统的神经网络相比，有以下特点：

- (1) SVM 算法最终转化为二次规划问题，从理论上可以得到全局最优解，避免了神经网络陷于局部最优的问题。
- (2) 拓扑结构由支持向量决定，避免了传统神经网络需要反复试验确定网络结构的问题。
- (3) 利用非线性变换将原始变量映射到高维特征空间中构造线性分类函数，既保证了模型具有良好的泛化能力，又解决了“维数灾难”问题。

## 参赛队号#1248

## 6.2 模型三的检验

考虑到客货运输对 GDP 的影响含多方面，所以二者之间并非简单的线性关系。因而利用 SVM 强大的非线性回归拟合功能，采用默认的 RBF 核函数，利用交叉验证方法寻找最佳参数  $c$ （惩罚因子）和参数  $g$ （方差），检验问题一中基于 AHP 的模糊综合评价的正确性。随机选取 100 个样本（前 75 个样本训练，后 25 个样本预测），以 9 类模糊指标作为网络的输入，样品的模糊指数作为网络的输出，验证模型三的正确性。拟合结果如下：

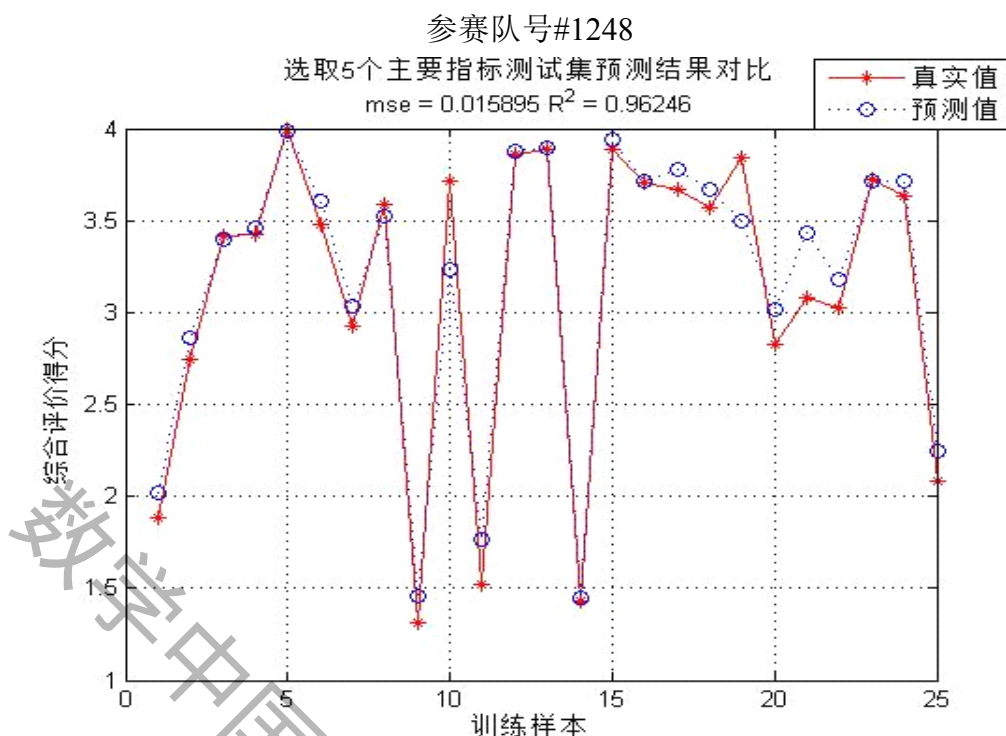


由训练结果可知，训练得到模型相关系数  $R = 91.658\%$  (接近于 1)，均方根误差  $mse=0.03$ ，以及预测值与实际值的对比，说明基于 AHP 的模糊综合指数评价方法性对可靠。

## 6.3 模型四的检验

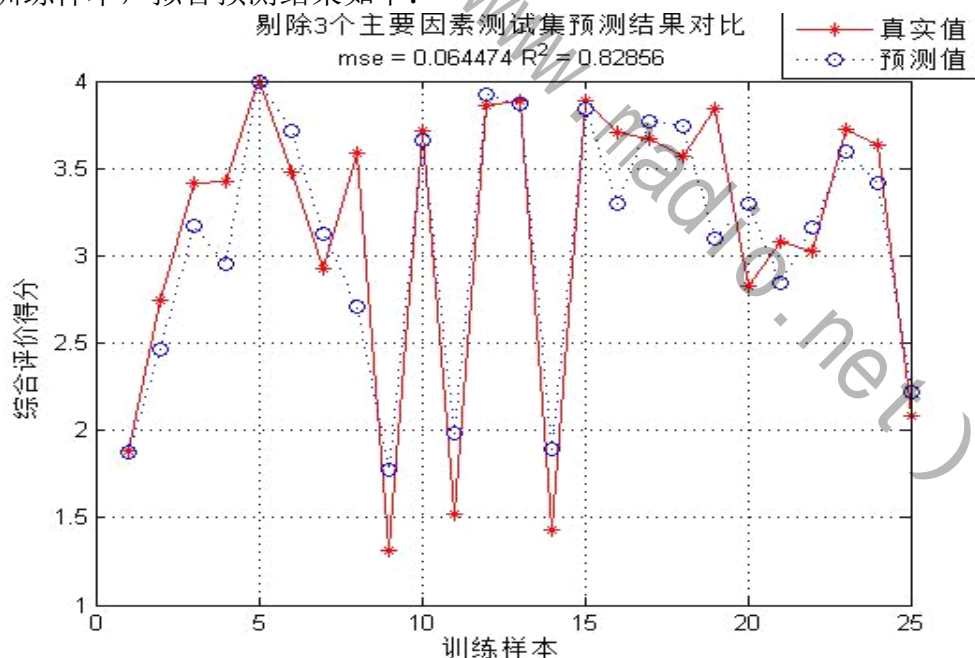
模型四通过计算各类城市 9 个模糊指标与模糊指数的相关系数矩阵，分析 9 个指标与客货运输对 GDP 影响的相关性，以分析各类指标的影响水平。但求得的相关系数仅仅是指二者间的线性相关性，客货运输对 GDP 影响包含很多方面，往往不是绝对线性那么简单（有时候在一定假设及误差允许范围内可近似为线性），各类指标与模糊指数的非线性相关性并未得到很好地验证，下采用 SVM 算法对模型四的相关性分析进行验证。

模型四中得到 5 项指标对 GDP 影响较大，考虑用这 5 项指标作为 SVM 输入，模糊指数作为 SVM 输出，选取与模型三检验的相同测试及训练样本，拟合预测结果如下：



将该图与上图进行比较分析，发现去除 4 个影响较小的因素后，相关系数及均方根误差明显变小，且拟合预测结果更加精确。说明了模型四选取的 5 个影响较大指标、4 个影响较小指标相对可靠。

其次，剔除 9 个指标影响较大的交通运输基础设施、第三产业、修理业 3 个指标，考虑用其余 6 项指标作为 SVM 输入，模糊指数作为 SVM 输出，选取与模型三检验的相同测试及训练样本，拟合预测结果如下：



继续将该图与上述两张图进行对比分析，发现拟合预测效果明显变差，原因主要在于影响因素较大指标的剔除及影响因素较小指标的加入，亦能反应模型四相关性分析的可靠性。

在自然界中，两类事物间不存在绝对的线性联系，特别是客货运输对于 GDP 影响包含众多方面，二者间联系更不会是线性那么简单。但在实际问题分析研究中，在一定假

### 参赛队号#1248

设及误差允许范围内，常常可将非线性联系转化为线性。SVM 虽能很好地验证二者间非线性联系，但难以挖掘具体的非线性联系，对于 9 个模糊指标与客货运输对 GDP 的影响间非线性联系，本文不予深入研究，但从 SVM 检验的结果来看，二者间的非线性联系与线性联系关联度较大，因此可认为模型四是正确可靠的。

## 7. 模型推广

当今社会是一个信息社会，统计学是一门学科，它是通过一定的数学模型，利用收集来的数据进行量化分析，综合分析，撰写统计分析报告，提出观点和看法，给决策者提供一定的理论依据。信息社会中统计分析应用广泛，主要体现在：为各部门提供相关数据、对水质进行评价、对生产产品进行质量分级等方面。所以，统计分析在我们现实生活中无处不在，该题中应用的统计分析方法较多，值得参考。

## 8. 模型的优缺点：

### 8.1 模型优点：

- ①考虑了该省各城市间的差异，分析了随着生产力水平的提高，客货运输对 GDP 影响的变化趋势。
- ②产业关联分析科学系统的评价了交通建设对 GDP 的影响力。
- ③基于 AHP 的模糊综合指数评价方法很好的将定量和定性相结合，综合了影响 GDP 的各个因素做出全面评价。
- ④相关性分析法很好的比较了各个因素对 GDP 影响差异及城市间的差异，为调查项目的修改提供了方向。

### 8.2 模型缺点：

- ①虽然层次分析法很好的将定量与定性相结合，也通过了一致性检验，但难以避免主观性的影响，对模型的评价精度存在一定影响。
- ②舍弃了一些调查项目，对数据的挖掘仍有待进一步加强。

## 9. 参考文献

- [1] 李伟, 交通运输投资与经济增长关系的研究, 天津大学硕士学位论文
- [2] 汪传旭, 交通运输业对国民经济贡献的衡量方法, 中国公路学报, 第 17 卷 第 1 期: 1, 2004 年 1 月
- [3] 百度百科网站: <http://baike.baidu.com/view/10147.htm>
- [4] 谢中华等,《MATLAB: 统计分析与应用: 40 个案例分析》, 北京: 北京航空航天大学出版社, 2010, P316-323
- [5] 汪晓银等,《数学建模与数学实验》, 北京: 科学出版社, 2010, P263-267
- [6] 李学文等,《数学建模优秀论文》, 北京: 清华大学出版社, 2011, P1-19
- [7] 司守奎等,《数学建模算法与应用》, 北京: 国防工业出版社, 2011, P200-201
- [8] 史峰等,《MATLAB 智能算法: 30 个案例分析》, 北京: 北京航空航天大学出版社, 2011, P269-289

## 参赛队号#1248

## 10. 附录

各城市聚类变量均值表

城市编号	购置金额 (千元)	全年运 货量 (吨)	路途堵塞停 驶天数(天)	产权形式 (1、国有 2、集体 3、 个体私营 4、承包租 赁)	收入 (十 万元)	途中花费 (万元)	维护费 用(十 万元)
城市 1	309.551	3781	35.2857	3.0638	5.634 3	1.3316	2.1477
城市 2	324.5193	27028	11.4335	2.4864	5.830 7	2.0738	2.594
城市 3	295.4587	1731	46.2273	3.0128	4.775 9	1.2235	2.2417
城市 4	212.1607	2868	6.9475	2.8284	5.891 4	1.4602	2.0586
城市 5	349.8533	1765	0	3.0946	2.536 2	0.4648	0.9578
城市 6	342.8302	6417	5.0629	3.3032	3.976 3	0.2934	1.2862
城市 7	141.6619	1754	4.7752	3.1286	2.205 4	0.3299	0.9693
城市 8	269.8697	5333	17.3176	3.006	4.748 1	2.0392	2.4828
城市 9	323.2667	19871	28.5464	3.0052	4.574 1	1.1793	2.2544
城市 22	175.1071	1222	0.7091	3.0571	2.250 3	0.7618	1.075
城市 25	261.0208	4944	12.0966	3.0769	4.416 9	0.9865	1.8979
城市 29	389.4302	8606	7.9535	3	5.450 3	1.8117	2.2951

部分程序（详见电子版）

K均值聚类

```
clc;
```

```
clear;
```

```
[X, textdata] = xlsread('AA.xls');
```

```
cityname = textdata(2:end,1);
```

```
X = zscore(X); % 数据标准化，即减去均值，然后除以标准差
```



## 参赛队号#1248

```

startdata = X([1,3,5,10],:);
idx = kmeans(X,4,'Start',startdata);
[S, H] = silhouette(X,idx); % 绘制轮廓图，并返回轮廓值向量s和图形句柄h
cityname(idx == 1)
cityname(idx == 2)
cityname(idx == 3)
cityname(idx == 4)

```

## 模糊综合评价

```

clear;
clc;
A=xlsread('各地区9个指标得分');
B=xlsread('各地区9个指标得分','较发达地区');
C=xlsread('各地区9个指标得分','中度发达');
D=xlsread('各地区9个指标得分','欠发达');
e=[A' B' C' D']';
[l,n]=size(e);
f=zeros(1,9);%排序指标
pj=zeros(4,9);%各级指标
m=zeros(1,9);
n=zeros(1,9);
jj=zeros(1,9);
for i=1:9
    px=sort(e(:,i+1));
    m(i)=px(floor(0.55*1));
    n(i)=px(floor(0.05*1));
    jj(i)=(m(i)-n(i))/5;
    pj(1,i)= m(i)- jj(i);
    pj(2,i)=m(i)- 2*jj(i);
    pj(3,i)=m(i)- 3*jj(i);
    pj(4,i)=m(i)-4*jj(i);
end
pj=pj';
pj(5,:)=[0.8 0.6 0.4 0.2];

r=zeros(9,4);
w=[ 26.595
26.595
5.985
5.985
3.5605
6.4716
11.6947
8.7

```

## 参赛队号#1248

```

4.35
]'/100;
[m,n]=size(A);
defen1=zeros(m,1);
for j=1:m
    x=A(j,2:end);
    for i=1:9
        zz=pj(i,:);
        xx=x(i);
        y(1)=(xx>=zz(1))+(zz(2)-xx)/(zz(2)-zz(1))*((xx>zz(2))&(xx<zz(1)));
        y(2)=(xx>zz(2)&xx<=zz(1))*(xx-zz(1))/(zz(2)-zz(1))+(xx>=zz(3)&xx<=zz(2))
        *(zz(3)-xx)/(zz(3)-zz(2));
        y(3)=(xx>zz(3)&xx<=zz(2))*(xx-zz(2))/(zz(3)-zz(2))+(xx>=zz(4)&xx<=zz(3))
        *(zz(4)-xx)/(zz(4)-zz(3));
        y(4)=(xx>zz(4)&xx<=zz(3))*(xx-zz(3))/(zz(4)-zz(3))+(xx<=zz(4));
        r(i,:)=y;
    end
    d=fuzzy_zhpj(3,w,r);
    r=zeros(9,4);
    defen1(j)=d*[4;3;2;1];
end
disp('A类影响得分为')
disp(sum(defen1)/m)

[m,n]=size(B);
defen2=zeros(m,1);
for j=1:m
    x=B(j,2:end);
    for i=1:9
        zz=pj(i,:);
        xx=x(i);
        y(1)=(xx>=zz(1))+(zz(2)-xx)/(zz(2)-zz(1))*((xx>zz(2))&(xx<zz(1)));
        y(2)=(xx>zz(2)&xx<=zz(1))*(xx-zz(1))/(zz(2)-zz(1))+(xx>=zz(3)&xx<=zz(2))
        *(zz(3)-xx)/(zz(3)-zz(2));
        y(3)=(xx>zz(3)&xx<=zz(2))*(xx-zz(2))/(zz(3)-zz(2))+(xx>=zz(4)&xx<=zz(3))
        *(zz(4)-xx)/(zz(4)-zz(3));
        y(4)=(xx>zz(4)&xx<=zz(3))*(xx-zz(3))/(zz(4)-zz(3))+(xx<=zz(4));
        r(i,:)=y;
    end
    d=fuzzy_zhpj(3,w,r);
    r=zeros(9,4);
    defen2(j)=d*[4;3;2;1];
end
disp('B类影响得分为')
disp(sum(defen2)/m)

```

## 参赛队号#1248

```

[m,n]=size(C);
defen3=zeros(m,1);
for j=1:m
    x=C(j,2:end);
    for i=1:9
        zz=pj(i,:);
        xx=x(i);
        y(1)=(xx>=zz(1))+(zz(2)-xx)/(zz(2)-zz(1))*((xx>zz(2))&(xx<zz(1)));
        y(2)=(xx>zz(2)&xx<=zz(1))*(xx-zz(1))/(zz(2)-zz(1))+(xx>=zz(3)&xx<=zz(2))
        *(zz(3)-xx)/(zz(3)-zz(2));
        y(3)=(xx>zz(3)&xx<=zz(2))*(xx-zz(2))/(zz(3)-zz(2))+(xx>=zz(4)&xx<=zz(3))
        *(zz(4)-xx)/(zz(4)-zz(3));
        y(4)=(xx>zz(4)&xx<=zz(3))*(xx-zz(3))/(zz(4)-zz(3))+(xx<=zz(4));
        r(i,:)=y;
    end
    d=fuzzy_zhpj(3,w,r);
    r=zeros(9,4);
    defen3(j)=d*[4;3;2;1];
end
disp('C类影响得分为')
disp(sum(defen3)/m)

[m,n]=size(D);
defen4=zeros(m,1);
for j=1:m
    x=D(j,2:end);
    for i=1:9
        zz=pj(i,:);
        xx=x(i);
        y(1)=(xx>=zz(1))+(zz(2)-xx)/(zz(2)-zz(1))*((xx>zz(2))&(xx<zz(1)));
        y(2)=(xx>zz(2)&xx<=zz(1))*(xx-zz(1))/(zz(2)-zz(1))+(xx>=zz(3)&xx<=zz(2))
        *(zz(3)-xx)/(zz(3)-zz(2));
        y(3)=(xx>zz(3)&xx<=zz(2))*(xx-zz(2))/(zz(3)-zz(2))+(xx>=zz(4)&xx<=zz(3))
        *(zz(4)-xx)/(zz(4)-zz(3));
        y(4)=(xx>zz(4)&xx<=zz(3))*(xx-zz(3))/(zz(4)-zz(3))+(xx<=zz(4));
        r(i,:)=y;
    end
    d=fuzzy_zhpj(3,w,r);
    r=zeros(9,4);
    defen4(j)=d*[4;3;2;1];
end
disp('D类影响得分为')
disp(sum(defen4)/m)

```



## 参赛队号#1248

## %相关性分析

```
r=zeros(10,4);  
x=xlsread('模糊综合评价得分','最发达地区','B2:K180');  
x=zscore(x);  
rr=corrcoef(x);  
r(:,1)=rr(:,end);  
  
x=xlsread('模糊综合评价得分','较发达地区','B2:K2008');  
x=zscore(x);  
rr=corrcoef(x);  
r(:,2)=rr(:,end);  
  
x=xlsread('模糊综合评价得分','中度发达','B2:K298');  
x=zscore(x);  
rr=corrcoef(x);  
r(:,3)=rr(:,end);  
  
x=xlsread('模糊综合评价得分','欠发达','B2:K564');  
x=zscore(x);  
rr=corrcoef(x);  
r(:,4)=rr(:,end);
```