

参赛队号 # 1370

第五届“认证杯”数学中国

数学建模网络挑战赛

承 诺 书

我们仔细阅读了第五届“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)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：

参赛队员（签名）：

队员 1：王宁

队员 2：史俊

队员 3：王夺

参赛队伍组别：本科组

参赛队号 # 1370

第五届“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）：

1370

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 # 1370

2012 年第五届“认证杯”数学中国 数学建模网络挑战赛

题目：白屋顶有用吗？

——基于偏最小二乘线性回归分析法的热岛效应研究

关键词：热岛效应 灰色相关度分析 偏最小二乘线性回归分析

摘 要：

本文以“白屋顶计划”为背景，利用北京市以及上海市的相关数据，选取影响城市热岛效应强度的不同因子的相关度这一角度，旨在定量研究“白屋顶计划”所对应的热岛效应影响因子对城市热岛效应强度的影响。

模型一主要分析了各影响因子与热岛效应的相关性。根据一定的合理假设，我们选择了 12 个不同的影响因子，分别是：降水量、平均气温、日照时间、平均风速、平均气压、相对湿度、雨日数、人口密度、机动车数量、城市绿化率、耗电量、太阳能反射指数。我们搜集并综合了北京市和上海市的不同的影响因子的数据，对于数据的处理，我们利用 C 语言编程工具，在灰色相关度分析方法的基础上，计算出了各因子的与热岛效应的相关度。根据不同的影响因子与热岛效应的相关度的差异可以定量地得出对热岛效应的决定性较大的影响因子，进而就可以对众多的影响因子做出合理的筛选。

模型二为最小二乘线性回归分析法模型。此模型中，我们通过对众多影响因子的筛选，决定除了六个与热岛效应强度关联度大的影响城市热岛效应的主要因子，分别是城市绿化率、日照时数、机动车数量、平均风速、太阳能反射率、平均气温。利用最小二乘线性回归分析法建立模型，通过 Matlab 软件拟合出热岛效应强度（ y ）与城市绿化

率、日照时数、机动车数量、平均风速、太阳能反射率、平均气温（ $x_0 \sim x_5$ ）的线性

回归方程。由于“白屋顶计划”只能决定 x_6 一个影响因子，因此我们通过带入不同的城市综合反射率即可得到“白屋顶计划”实施前后热岛效应强度的变化情况。鉴于不同因素对热岛效应强度的影响程度不同，我们又选取了另外一个可以改变的人为影响因子——机动车数量，作为单变量，进而求得机动车数量减少前后，城市热岛效应的变化情况。根据结果可以看出：“白屋顶计划”实施后对城市热岛效应的减弱有着明显作用，而且效果比机动车数量的减少更为显著。进一步突出了“白屋顶计划”对于减弱城市热岛效应的作用。

最后，我们利用灵敏度检验，对结果进行了分析；还对模型的优缺点进行了指出与评价，并提出了进一步完善模型的大体方向。

参赛队号 1370 所选题目 B

参赛密码 _____
(由组委会填写)

Abstract

This paper is based on the "White Roofs Plan". With the data of Beijing and Shanghai, we studied quantitatively the factor of the heat island effect corresponding to the "White Roofs Plan" from the point of view of the different factors that affect the urban heat island intensity.

The Model I was used to analyze the correlation of each impact factor and the heat island effect. According to certain reasonable assumptions, we selected 12 different impact factors are: water, average temperature, duration of sunshine, average wind speed, atmospheric pressure, relative humidity, rain days, population density, the number of motor vehicles, urban greening rate, power consumption, solar reflective index. After collecting, synthesizing and processing the data of Beijing and Shanghai about the different impact factors, we use Gray correlation analysis method and calculate the correlation of each factor and the heat island effect. Due to the difference of the correlation of each impact factor and the heat island effect, we can obtain quantitatively which factors are more decisive on the heat island effect and furthermore make a reasonable screening a large number of impact factors.

Model II is a least squares method, linear regression analysis model. By screening a number of impact factors, We have come to the six most important factors, which are associated with heat island intensity. They are Rate of Urban Greening, Sunshine Hours, Number of Motor Vehicles, Average wind speed, Solar reflectance, Average temperature. Using the model II, Linear regression equation is fitted by Matlab software, which is the rate of heat island intensity, urban greening, sunshine hours, the number of motor vehicles, the average wind speed, solar reflectivity, the average temperature relationship. "White Roofs Plan" can only decide on a factor x_6 , so we can obtain the changes of the heat island intensity before and after Implementation of the "White Roofs Plan" by making the reflectivities of the different cities into the regression equation. The different factors have different influence of the heat island intensity, we also select a human impact factor, which can be changed, Number of motor vehicles. As a single variable, we can conclude that the changes of the urban heat island effect before and after the reduced number of motor vehicles. As can be seen based on the results, Urban heat island effect is significantly reduced after the implementation of the plan of "the white roof", and the effect is more significant than the change of the number of motor vehicles. The effect of "White Roofs Plan" is highlighted, while is to weaken the urban heat island effect.

At last, with the test of sensitivity, the results of model were analyzed; Based on this, we also point out the advantages and disadvantages of the model and give direction of further improvement of the model.

Key Words: Urban heat island effect; Gray correlation analysis; Least squares linear regression analysis

目录

问题重述	6
问题分析	6
模型假设	6
符号说明	7
模型的建立与求解	7
模型 1 的建立	7
模型 1 的求解	8
模型 2 的建立	9
模型 2 的求解	14
模型的检验	18
模型的评价及优化	18
参考文献	19
附录	19

1. 问题重述

城市热岛效应(Urban heat island effect)是指城市中的气温明显高于外围郊区的现象。由于城市地区绝大多数建筑都是有水泥、沥青等所构成的，下垫面导热率高；加之空气污染严重，人口密度大，温室效应较严重，致使吸收较多的太阳能热量，而且还有大量的人为热量进入空气中；另一方面又因建筑物密集，不利于热量的扩散，形成局部高温中心，并由此向外围郊区递减。

为了缓解城市热岛效应，有专家提出，可以将城市建筑物的屋顶涂成白色，减少对太阳光的吸收率，可以使城市的气温降低，进而达到节能和环保的效果。

包括诺贝尔物理学奖获得者、现任美国能源部长朱棣文在内的许多学者都赞同这种方法，但是也有一些人提出反对的观点，认为该方法“不能在短时间内得到推广”，“对减轻热岛效应的贡献不会太大”，或是“在视觉上让人无法承受”等。

本文主要解决以下问题：

建立一个合适的模型计算出各个因子与城市热岛效应的相关度；

对于“白屋顶计划”决定的因素，建立适当的模型预测“白屋顶计划”实施后热岛效应强度是变化趋势；

设计一种方法检验模型的灵敏度，并对模型的优缺点讨论分析。

2. 问题分析

热岛效应是现代城市特有的一种气温异常现象，城市的热岛效应强度可以用城市中心区域郊区同时期气温差值来表示。影响城市热岛效应的因子有很多，既有自然因素，也有人为因素的影响。

要想评估“白屋顶计划”对城市热岛效应的影响，首先要选取一个指标分析各因子与热岛效应的相关度，根据对以前文献的分析我们选取降水量、平均气温、日照时间、平均风速、平均气压、相对湿度、雨日数、人口密度、机动车数量、城市绿化率、耗电量、太阳能反射指数 12 项作为影响城市热岛效应的因子，热岛效应的强度作为分析指标。然后根据不同因子对热岛效应的不同的相关度，既可排序筛选出相关度较高的影响因子。

根据筛选出来的六个主要影响因子：城市绿化率、日照时数、机动车数量、平均风速、太阳能反射率、平均气温的数据，利用偏最小二乘线性回归分析法拟合出热岛效应强度与影响因子的线性回归方程。接着既可利用次方程对“白屋顶计划”实施后对城市热岛效应的影响进行预测。

最后根据灵敏度的计算对模型的结果进行检验。

3. 模型假设

假设与热岛效应相关度小的次要因子对其影响可以忽略不计；

假设城市每年气象指标无异常的变化；

假设搜集到的数据全部真实可信。

4. 符号说明

SRI ：太阳能反射指数

r^2 ：线性相关系数

F ： F 检验参数

\mathcal{V} ：城市热岛效应强度

5. 模型的建立与求解

首先我们引入太阳能反射指数与城市综合反射指数的概念，太阳能反射指数(Solar Reflectance Index, 简称 SRI) 显示了材料表面抗拒太阳热能的能力，SRI 越高，这种材料在太阳照射下的升温幅度越小。

根据定义，标准黑色（太阳能反射率 0.05，发射率 0.9）的 SRI 为 0，而标准白色（太阳能反射率 0.80，发射率 0.9）的 SRI 为 100。

一些常见的材料的 SRI 列如下表：

材料	SRI
沥青	0
灰色水泥瓦	25
白色矿物颗粒沥青	28
红色陶土屋瓦	36
铝	56
白色 EPDM 橡胶	84
白色水泥瓦	90
白色 PVC	104
白色涂料	100-107

表 1 各种材料的 SRI

而城市综合反射指数是在考虑城市建筑面积率，道路面积率等因素之后得到的一个城市对太阳光的综合反射指数。

① 模型 1 的建立

灰色关联度分析

利用灰色关联度对影响城市热岛效应的因子群进行贡献测度分析。所谓关联度是指其他数列与参考数列之间的接近程度，它描述了系统发展过程中因素间的相对变化情

参赛队号 #1370

况，即变化大小、方向与速度的相对性。需要指出的是，关联度的大小只是因子间相互作用、相互影响的外在表现，关联度的排序才能体现众多因子对参考因子的相对影响程度。

设有 m 个评价指标的 n 个评价样本组成的原始时间数列为：

$$X_i = \{X_i(k) | k=1,2,\dots,n, i=0,1,2,\dots,m-1\}$$

式中： X_0 为参考数列，其他为比较数列，以下若无说明，均满足 $i \neq 0$ 。对原始数列进行均值生成处理，得到数列 X_i 。记第 k 时刻 X_0 与 X_i 的绝对值差为

$$\Delta_i(k) = |X_0(k) - X_i(k)| \quad (i=1,2,\dots,m-1),$$

则 $X_i(k)$ 与 $X_0(k)$ 的关联系数 $\xi_i(k)$ 可由公式 (1) 求得：

$$\xi_i(k) = \frac{\min_i \min_k \Delta_i(k) + \rho \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \rho \max_i \max_k \Delta_i(k)} \quad (1)$$

式中： $\min_i \min_k$ 称为两级最小差，其中 \min_i 为第一级最小差，表示在 X_i 曲线上各点与 X_0 中各相应点的距离的最小值； $\min_i \min_k$ 表示在各曲线找出的最小差 \min_k 的基础上，再按 $i=1,2,\dots,m-1$ 找出的所有曲线中最小差的最小值； $\max_i \max_k$ 为两级最大差，意义与两级最小差相似； $\rho \in [0,1]$ ，称为分辨系数， ρ 值越小，分辨能力越大，一般取 $\rho=0.5$ 。由于获得的关联系数过多，信息过于分散，不便于比较，因此有必要将每一个比较数列各时刻的关联系数集中体现在一个值上，即灰色关联度。灰色关联度的计算如公式 (2)（一般采用平均值法）：

$$\gamma_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (2)$$

将灰色关联度按大小进行排序，若灰色关联序为 $\gamma_1 > \gamma_2 > \dots > \gamma_{m-1}$ ，则表明 X_1 对 X_0 的影响最大， X_2 次之，以此类推。

② 模型 1 的求解

在确定城市热岛效应影响因子的关联度时，本文综合了北京和上海两个城市近近十年间的各项影响因子数据（参见附录），计算历年个城市热岛因子的基础上，利用城市热岛效应影响因子的各项指标的无量纲结果，进行分析计算。

参赛队号 # 1370

针对热岛强度和 12 项影响因子，设热岛强度为参考数列，其余 12 项为比较数列，通过 C 语言编程计算可得到各个因子与热岛强度的灰色相关度如表 1 所示。相关 C 程序参见附录。

影响因子	关联度	影响因子	关联度
降水量	0.3547	雨日数	0.4604
平均气温	0.4881	人口密度	0.4343
日照时数	0.6344	机动车数量	0.6074
平均风速	0.4882	城市绿化率	0.6462
平均气压	0.2415	耗电量	0.4348
相对湿度	0.4825	太阳能反射指数	0.5797

表 2 城市热岛效应与各影响因子间关联度

由结果可以看出，城市绿化率、日照时数与机动车数量是影响城市热岛效应的主要因素，平均风速、太阳能反射指数、平均气温以及相对湿度其次，雨日数、耗电量、人口密度再次之，降水量与平均气压是最次要的影响因子。

由此结果还可以看出人为影响因子相对于自然影响因子占主导作用。

③ 模型 2 的建立 偏最小二乘回归分析

在一般的多元线性回归模型中，有一组因变量 $Y = \{y_1, y_2, \dots, y_q\}$ (q 为因变量个数) 和自变量 $X = \{x_1, x_2, \dots, x_m\}$ (m 为自变量个数)，当数据总体满足高斯-马尔科夫定理时，由最小二乘法有：

$$B = X(X^T X)^{-1} X^T Y \quad (3)$$

式中 B 为估计的回归系数。

当 X 中的变量存在严重的多重相关性（变量本身物理意义决定了它们之间的相关性，或由样本点数量不足造成），式 (3) 中行列式 $(X^T X)$ 几乎接近于零，求解 $(X^T X)^{-1}$ 时会含有严重的舍入误差，使回归系数估计值的抽样变异性显著增加。更有甚者，当 X 中的变量完全相关时， $(X^T X)$ 是不可逆矩阵，无法求解回归系数。此时，若仍沿用最小二乘法拟合回归模型，回归结果将会出现许多反常现象，致使其精度、可靠性得不到保证。在实际工作中，变量的多重相关性是普遍存在的。偏最小二乘法就能较好地解决这类问题。

在本文中，由于影响城市热岛效应的各个因子的相关度不尽相同，在借鉴前人的研究成果的基础上根据前面模型的结果忽略次要因子的影响，选取影响热岛效应的六个主要因子作为回归分析的自变量 $X_0 \sim X_5$ ，分别是：城市绿化率、日照时数、机动车数

量、平均风速、太阳能反射指数、平均气温。因变量选为城市的热岛效应强度（ Y ）。

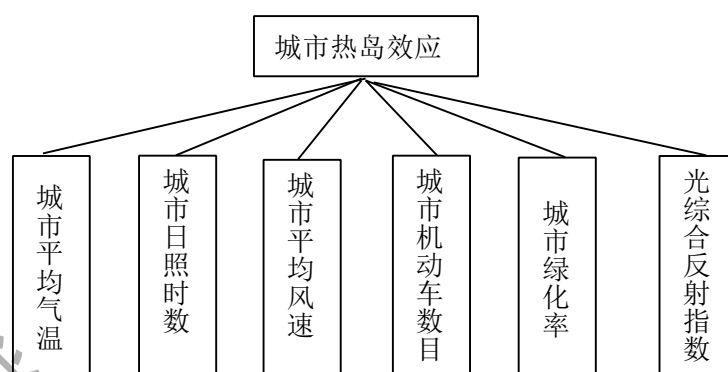


图1 影响城市热岛效应的因子

偏最小二乘法是多元线性回归、典型相关分析和主成分分析的集成和发展。其思路是：首先，从自变量集合 X 中提取成分 $t_h (h=1,2,\dots)$ ，各成分相互独立；然后，建立这些成分与自变量 X 的回归方程，其关键在于成分的提取。与主成分回归不同的是，偏最小二乘回归所提取的成分既能很好地概括自变量系统中的信息，又能最好地解释因变量，并排除系统中的噪声干扰。因而有效地解决了自变量间多重相关性情况下的回归建模问题。

当 $q=1$ 时，为单变量偏最小二乘回归模型（记为 PLS1）；当 $q>1$ 时，为多变量偏最小二乘回归模型。由于本文讨论的热岛强度问题只有一个因变量，故属于 PLS1 模型。
数据标准化处理
标准化的目的是使样本点的集合重心与坐标原点重合。

$$\begin{cases} F_0 = (F_{0y})_n \\ F_{0y} = [y - E(y)] / S_y \end{cases} \quad (4)$$

$$\begin{cases} E_0 = (E_{01}, E_{02}, \wedge, E_{0m})_{n \times m} \\ E_{0i} = x_i^* = [x_i - E(x_i)] / S_{x_i}, (i=1,2,\wedge, m) \end{cases} \quad (5)$$

式中 F_0 ， E_0 分别为 y ， X 的标准化矩阵； $E(y)$ ， $E(x_i)$ (i 同上) 分别为 y ， X 的均值； S_y ， S_{x_i} 分别为 y ， X 的均方差； n 为样本用量。

第一成分 t_1 的提取

已知 F_0, E_0 ，可以从 E_0 中提取的第一个成分 t_1 ，

$$t_1 = E_0 W_1$$

式中 W_1 为 E_0 的第一个轴，为组合系数。 $\|W_1\|=1$ ； t_1 是标准化变量 $x_1^*, x_2^*, \wedge, x_m^*$ 的线性组合，为原信息的重新调整。

从 F_0 中提取第一个成分 u_1 ，

$$u_1 = F_0 C_1$$

式中 C_1 为 F_0 的第一个轴， $\|C_1\|=1$ 。在此，要求 t_1 ， u_1 能分别很好地代表 X 与 Y 中的数据变异信息，且 t_1 对 u_1 有最大的解释能力。根据主成分分析原理和典型的相关分析的思路，实际上是要求 t_1 与 u_1 的协方差最大，这是一个最优化问题。经推导可得

$$\begin{cases} E_0^T F_0 F_0^T E_0 W_1 = \theta_1^2 W_1 \\ F_0^T E_0 E_0^T F_0 C_1 = \theta_1^2 C_1 \end{cases} \quad (6)$$

式中 θ_1 为优化问题的目标函数； W_1 为 $E_0^T F_0 F_0^T E_0$ 的特征向量， θ_1^2 为对应的特征值。 C_1 为对应于矩阵 $F_0^T E_0 E_0^T F_0$ 最大特征值 θ_1^2 的单位特征向量。

要 θ_1 取最大值，则 W_1 为 $E_0^T F_0 F_0^T E_0$ 矩阵最大特征值的单位特征向量，本文讨论的热岛效应问题中， $C_1=1$ ，则 $u_1 = F_0$ 。

$$W_1 = \frac{E_0^T F_0}{\|E_0^T F_0\|} = \frac{1}{\sqrt{\sum_{i=1}^m r^2(x_i, y)}} \begin{bmatrix} r(x_{1,y}) \\ \wedge \\ r(x_m, y) \end{bmatrix}$$

$$t_1 = E_0 W_1 = \frac{1}{\sum_{i=1}^m r^2(x_i, y)} [r(x_1, y)E_{01} + r(x_2, y)E_{02} + \wedge + r(x_m, y)E_{0m}]$$

式中 $r(x_i, y)$ 为 x_i 与 y 的相关系数。

从 t_1 中可以看出, t_1 不仅与 X 有关, 而且与 Y 有关; 另外, 若 x_i 与 Y 的相关程度越强, 则 x_i 的组合系数越大, 其解释性就越明显。

求得轴 W_1 后, 可得成分 t_1 。分别求 F_0 , E_0 对 t_1 的回归方程为

$$E_0 = t_1 P_1^T + E_1, \quad F_0 = t_1 r_1 + F_1 \quad (7)$$

式中 $P_1 = E_0^T t_1 / \|t_1\|^2$, 为回归系数, 向量; $r_1 = F_0^T t_1 / \|t_1\|^2$, 为回归系数, 标量; E_1, F_1

分别为回归方程的残差矩阵, $E_1 = [E_{11}, E_{12}, \wedge, E_{1m}]$; $F_1 = F_0 - t_1 r_1$ 。

第二成分 t_2 的提取

以 E_1 取代 E_0 , F_1 取代 F_0 , 用上面的方法求第 2 个轴 W_2 和第 2 个成分 t_2 , 有

$$W_2 = \frac{E_1^T F_1}{\|E_1^T F_1\|} = \frac{1}{\sqrt{\sum_{i=1}^m C_{0v}^2(E_{1i}, F_1)}} \begin{bmatrix} C_{0v}(E_{11}, F_1) \\ \wedge \wedge \\ C_{0v}(E_{1m}, F_1) \end{bmatrix}$$

$$t_2 = E_1 W_2$$

式中 $C_{0v}(\cdot)$ 表示协方差。

施行 E_1 , F_1 对 t_2 的回归, 有

$$\begin{cases} E_1 = t_2 P_2^T + E_2 \\ F_1 = t_2 r_2 + F_2 \end{cases} \quad (8)$$

式中 $P_2 = E_1^T t_2 / \|t_2\|^2$, $r_2 = F_1^T t_2 / \|t_2\|^2$ 。

第 h 成分 t_h 的提取

同理, 可推求出第 h 成分 t_h 。H 可用交叉有效性原则进行识别。H 小于 X 的秩。

推求偏最小二乘回归模型

F_0 关于 t_1, t_2, \dots, t_h 的最小二乘回归方程为

$$\hat{F}_0 = r_1 t_1 + r_2 t_2 + \wedge + r_h t_h \quad (9)$$

由于 t_1, t_2, \dots, t_h 均是 E_0 的线性组合, 有偏最小二乘回归的性质有

$$t_i = E_{i-1} W_i = E_0 W_i^* (i=1, 2, \wedge, h) \quad (10)$$

式中 $W_i^* = \prod_{k=1}^{i-1} (I - W_k P_k^T) W_i$ 。

将式 (10) 带入式 (9) 得

$$\hat{F}_0 = r_1 E_0 W_1^* + r_2 E_0 W_2^* + \wedge + r_h E_0 W_h^* = E_0 (r_1 W_1^* + \wedge + r_h W_h^*) \quad (11)$$

记 $y^* = F_0, x_i^* = E_{0i}, \alpha_i = \sum_{k=1}^h r_k W_{ki}^* (i=1, 2, \wedge, m)$, 则式 (11) 可还原成标准化变量的回归方程为

$$\hat{y}^* = \alpha_1 x_1^* + \alpha_2 x_2^* + \wedge + \alpha_m x_m^* \quad (12)$$

式 (12) 还可以进一步写成原变量的偏最小二乘回归方程为

$$\hat{y} = [E(y) - \sum_{i=1}^m \alpha_i \frac{S_y}{S_{x_i}} E(x_i)] + \alpha_1 \frac{S_y}{S_{x_1}} x_1 + \wedge + \alpha_m \frac{S_y}{S_{x_m}} x_m \quad (13)$$

交叉有效性原则

记 y_i 为原始数据, t_1, t_2, \dots, t_m 是在偏最小二乘回归过程中提取的成分。 \hat{y}_{hi} 是使用全部样本点并取 t_1, t_2, \dots, t_h 个成分回归建模后, 第 i 个样本点的拟合值。 $\hat{y}_{h(-i)}$ 是在建模时删去 i 个样本点, 取 t_1, t_2, \dots, t_h 个成分回归建模后, 再用此模型计算的 y_i 的拟合值。记

$$\left\{ \begin{array}{l} S_{Sh} = \sum_{i=1}^n (y_i - \hat{y}_{hi})^2 \\ P_{RESSh} = \sum_{i=1}^n (y_i - \hat{y}_{h(-i)})^2 \\ Q_h^2 = 1 - \frac{P_{RESSh}}{S_{Sh-1}} \end{array} \right.$$

当 $Q_h^2 \geq 0.0975$ 时，引进的新的成分 t_h 会对模型的预测能力有明显的改善作用。这就是交叉有效性原则。

④ 模型 2 的求解

利用 Matlab 软件对北京以及上海的影响因子数据（参见附录）进行处理：
在 Matlab 中键入如下数据：

(1) 回归分析及检验

X=[

1	12.70	2418.70	2.6	62.18	33.24	18.59
1	13.10	2596.50	2.5	78.43	34.22	18.59
1	13.10	2420.70	2.3	89.85	35.60	18.59
1	13.10	2594.00	2.4	95.14	36.30	18.59
1	12.80	2667.20	2.5	104.12	36.50	18.59
1	12.90	2611.70	2.4	114.47	38.78	18.59
1	13.20	2588.40	2.3	133.93	40.57	18.59
1	12.90	2260.20	2.5	163.07	40.87	18.59
1	13.50	2515.40	2.4	182.42	41.91	18.59
1	13.20	2576.10	2.4	209.73	42.00	18.59
1	13.40	2192.70	2.2	239.12	42.50	18.59
1	14.00	2351.10	2.2	273.36	43.00	18.59
1	13.40	2391.40	2.2	313.68	43.50	18.59
1	13.30	2511.80	2.2	368.11	44.40	18.59
1	12.60	2382.90	2.3	449.72	45.00	18.59
1	17.18	1649.50	3.0	49.19	22.19	18.59
1	17.21	1914.90	3.1	5.01	23.82	18.59
1	17.50	1780.00	2.9	62.30	30.01	18.60
1	17.00	1714.00	3.2	71.90	35.20	18.60
1	17.50	1965.00	3.4	83.51	36.03	18.59
1	17.10	1829.00	3.1	95.16	37.00	18.59
1	17.90	1678.10	2.8	107.04	37.30	18.59
1	18.15	1616.90	2.9	119.70	37.60	18.59
1	17.20	1635.70	3.3	132.12	38.00	18.59

参赛队号 #1370

```

1 17.40    1680.90    2.9    147.11    38.10    18.59
1 17.20    1662.00    3.1    175.51    38.20    18.60];
Y=[1.00  0.97  0.92  1.3  1.52  1.8  2  1.9  2.2  2.1  2.4  2.5  2.3  2.6
2.7 1.00  1.25  1.10  1.30  1.58  1.42  1.32  1.44  1.52  1.62  1.60 ]';
[b,bint,r,rint,stats]=regress(Y,X);
b,bint,stats

```

得出结果：b =

132.2876

0.0824

0.0004

-0.0031

0.0038

0.0275

-7.2281

bint =

-484.9867 749.5619

-0.0789 0.2438

-0.0004 0.0012

-0.5856 0.5794

0.0021 0.0054

-0.0010 0.0559

-40.4141 25.9579

stats =

0.8572 19.0031 0.0000 0.0542

由以上数据可得：

	β_0	β_1	β_2	β_3	β_4	β_5	β_6
数值	132.2876	0.0824	0.0004	-0.0031	0.0038	0.0275	-7.2281
置信区间	[-484.9867, 749.5619]	[-0.0789, 0.2438]	[-0.0004, 0.0012]	[-0.5856, 0.5794]	[0.0021, 0.0054]	[-0.0010, 0.0559]	[-40.4141, 25.9579]

表3 拟合方程系数及置信区间

还可以得到：

相关系数 $r^2 = 0.8572$ $F = 19.0031$ 概率 $P = 0.0000$ ， $P < 0.0542$

故而有以上数据可知偏最小二乘法的原始变量回归模型为：

$$y = 132.2876 + 0.0824x_1 + 0.0004x_2 - 0.0031x_3 + 0.0038x_4 + 0.0275x_5 - 7.2281x_6$$

残差分析，作残差图

在 matlab 中键入如下命令：

`rcoplot(r,rint)`

得到回归模型的残差图如下

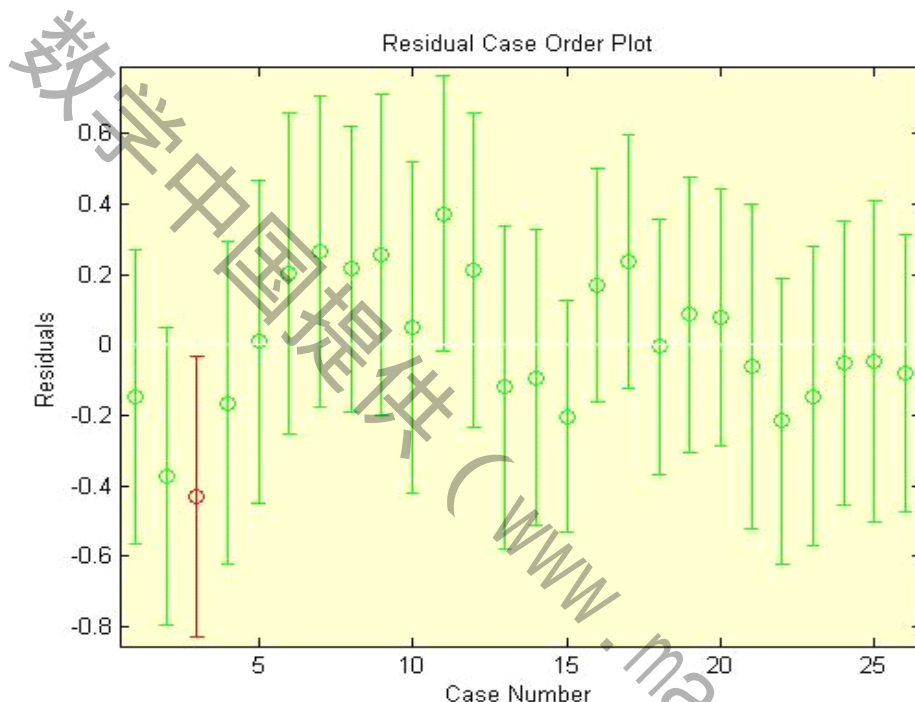


图2 回归模型残插图

从残差图中可以看出，除了第三个数据外，其余数据的残差都离原点较近，且残差的置信区间均包含原点，这说明回归模型

$$y = 132.2876 + 0.0824x_1 + 0.0004x_2 - 0.0031x_3 + 0.0038x_4 + 0.0275x_5 - 7.2281x_6$$

能较好的符合原始数据，而第三个点可视为异常点。

利用建立的该回归模型可以定量的预测“白屋顶计划”实施后对热岛效应强度产生的影响。根据假设，将影响城市热岛效应的其他五个因子暂时看成是每年变化不大的，因此可以将 $x_1 \sim x_5$ 看成常数，只需将城市综合反射指数— x_6 分别取灰色屋顶与白色屋顶时的值计算出相对应的城市热岛效应强度。

取 2010 年城市的其他影响因子数据，带入回归方程计算，

以灰色屋顶情况作为基准，取灰色水泥砖瓦的 SRI=25，则城市综合反射指数

$x_6=18.57$ 则有热岛效应强度为

参赛队号 #1370

$$y_1 = 1.38$$

保持其他影响因子不变，增加屋顶白色屋顶的数量使得城市综合阳光反射率

$x_6 = 18.68$ 情况下：

$$y_2 = 0.67$$

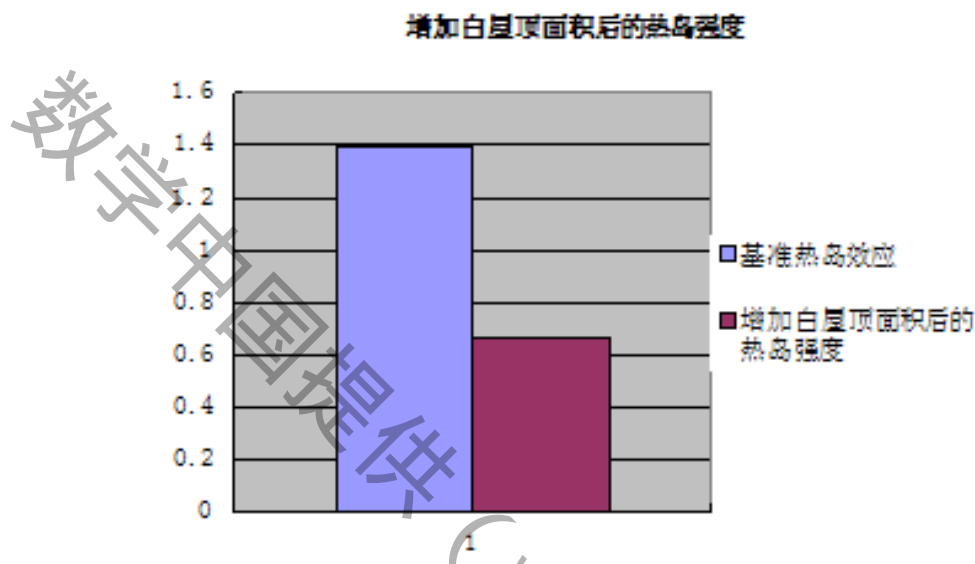


图4 增加白屋顶后热岛效应强度对比

由此可以看出，在其他影响因子不变的情况下，白色屋顶可以有效地降低城市热岛效应的强度，而且效果比较明显。

为了比较其他因子跟白色屋顶对热岛效应减小的强度的差别，我们再选取机动车数量作为单变量，分别计算热岛效应强度。结果分别为：

$$y_1^* = y_1 = 1.38 \quad y_2^* = 1.04$$

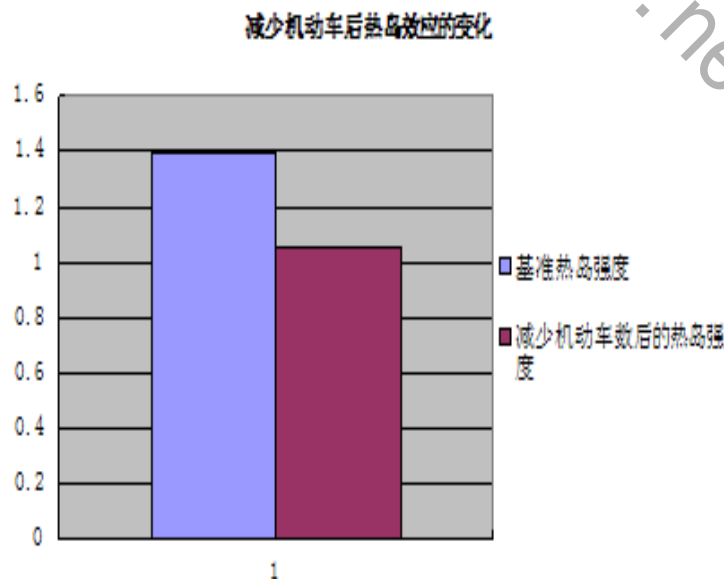


图 5 减少机动车数量后热岛效应强度变化对比

6. 模型的检验

灵敏度分析

由于本问题中对热岛效应影响的因素有很多，并且随机性很大，所以我们对模型的相关参数进行灵敏度分析。模型对这些参数的敏感性反映了各种因素影响结果的显著程度；反之，通过对模型参数的稳定性和敏感性分析，又可以反映和检验模型的实际合理性。

但我们发现本问题是情况复杂的多因素问题。其指标主要包括平均气温，日照时数，平均风速，机动车数量，绿化率，屋顶材料阳光反射率，平均气压，相对湿度，雨日数，人口密度，耗电量等，但由于其中情况复杂多变，部分指标之间互相影响，因此我们只对最主要并且相互独立的六个指标进行灵敏度分析，包括平均气温，日照时数，平均风速，机动车数量，绿化率，屋顶材料阳光反射率。

对以上六个指标的每个指标单独增大 10 个百分点，而其他指标保持不变，观察其对热岛强度的影响程度。为了保证数据的广泛性，我们选取了近十年来上海和北京的数据作为研究参考对象。以下是各指标增大 10 个百分点，其他保持不变，得出模型结果变化，如下表所示：

平均气温(℃)	日照时数(时)	平均风速 (m/s)	机动车数 (万辆)	城市绿化率 (%)	屋顶材料阳光反射率
0.03784	0.01775	0.00026	0.01781	0.02805	0.08540

表 4 各影响因子灵敏度

通过对六个最主要影响因子的灵敏度分析，我们可以看出屋顶材料阳光反射率对于热岛效应强度的影响比其他因子要大得多。可见“白屋顶计划”是有一定的理论根据和可行性。

7. 模型的评价及优化

模型的优点：

- 1) 我们引用了灰色相关度分析方法计算相关度，不仅研究了各个因子对热岛效应强度的影响，还考虑到了各个因子的内在的互相影响，相关度结果较为客观。
- 2) 对影响因子进行量化分析，使因子变化导致的热岛效应强度变化显得更直观。
- 3) 引入 Matlab 计算，大大的减少了人为计算量。
- 4) 利用灵敏度对模型进行检验，使得“白屋顶计划”的影响程度更具可比性。

模型的缺点：

- 1) 在数据处理时，对一些次要因子的处理不够充分，会使结果存在一定的误差。
- 2) 对影响因子的量化方法必然会使得数据失去一部分真实性，容易导致误差的积累。
- 3) 对次要因子的影响的忽略必然会导致拟合出的方程系数会与实际有偏差。

模型的改进：

针对模型中的假设，我们拟合方程时忽略了很多次要影响因子，这方面导致的误差会相对较大，因此可以在时间与精力允许的范围内有选择地增加一定量的影响因子数据，以提高拟合曲线的相关系数，以及模型预测的准确度。但是增加因子一定要有数据支持或是理论依据，因此就需要更深入的研究各因子之间的互相作用来选取合适的影响因子。所以本文的改进方向主要是选取更为合适的影响因子，使得拟合的曲线更准确。

8. 参考文献

- 【1】刘转年，阴秀菊。西安城市热岛效应及气象因素分析[J]。干旱区资源与环境，2008，22（2）；88~90。
- 【2】王郁，胡非。近10年来北京夏季热岛的变化及环境效应的分析研究[J]。地球物理学报，2006，49（1）；61~68。
- 【3】林学椿，于淑秋，唐国利。北京城市化进程与热岛强度的研究[J]。自然科学进展，2005，15（7）；882~886。
- 【4】曹爱丽，张浩，张艳，马蔚纯。上海近50年气温变化与城市化发展的关系[J]。地球物理学报，2008，51（6）；1663~1669。
- 【5】何萍，陈辉，李宏波，席武俊。云南高原楚雄市热岛效应因子的灰色分析[J]。地理科学进展，2009，28（1），26~32。
- 【6】马敏洋，于钦赞，许舒婷。基于数理统计分析对影响广州市热岛效应因素的研究[J]。科技传播，2010，63~65。
- 【7】李宏波，何萍，张杰兵。楚雄市2008年城市热岛效应实测数据分析[J]。云南地理环境研究，2010，22（4），82~86。
- 【8】丁金才，张志凯，奚红，周红妹。上海地区盛夏高温分布和热岛效应的初步研究[J]。大气科学，2002，26（3），412~420。
- 【9】田喆，朱能，刘俊杰。城市气温与人为影响因素的关系[J]。天津大学学报，2005，38（9），830~833。
- 【10】陈志，俞炳丰，胡汪洋，罗昔联，秦临香。城市热岛效应的灰色评价与预测[J]。西安交通大学学报，2004，38（9），986~988。
- 【11】李学文，李炳照，王宏州。数学建模优秀论文精选与点评。清华大学出版社，2011，210~229。

9. 附录

- 1) 北京市与上海市近十年各项影响因子数据：

参赛队号 # 1370

城市	年份	降水量 (mm) X1	平均 气温 (°C) X 2	日照时数 (时) X3	平均风速 (m/s) X4	平均气压 (百帕) X5	相对湿度 X6	雨日数 (日) X7	人口密度 (千人/平 方公里) X8	机动车 数(万辆) X9	城市绿化 率(%) X10	耗电量(亿 千瓦时) X11	阳光反 射率 X12	热岛强 度Y
北京	1996	700.9	12.7	2418.7	2.6	1011.0	51.00	103	767.20	62.18	33.24	244.37	0.25	1.00
北京	1997	430.9	13.1	2596.5	2.5	1012.9	56.00	76	755.60	78.43	34.22	263.61	0.25	0.97
北京	1998	731.7	13.1	2420.7	2.3	1012.5	62.00	93	759.30	89.85	35.60	276.21	0.25	0.92
北京	1999	266.9	13.1	2594.0	2.4	1012.5	53.00	86	766.00	95.14	36.30	297.26	0.25	1.30
北京	2000	371.1	12.8	2667.2	2.5	1012.7	54.25	83	842.10	104.12	36.50	384.43	0.25	1.52
北京	2001	338.9	12.9	2611.7	2.4	1012.9	56.25	78	842.80	114.47	38.78	399.94	0.25	1.80
北京	2002	370.4	13.2	2588.4	2.3	1012.7	49.00	84	867.10	133.93	40.57	439.96	0.25	2.00
北京	2003	444.9	12.9	2260.2	2.5	1013.3	54.00	93	887.50	163.07	40.87	467.61	0.25	1.90
北京	2004	483.5	13.5	2515.4	2.4	1012.6	49.17	94	909.80	182.42	41.91	513.18	0.25	2.20
北京	2005	410.7	13.2	2576.1	2.4	1012.8	49.00	79	937.20	209.73	42.00	570.54	0.25	2.10
北京	2006	318.0	13.4	2192.7	2.2	1012.5	53.00	86	963.40	239.12	42.50	611.57	0.25	2.40
北京	2007	483.9	14.0	2351.1	2.2	1012.6	54.00	78	995.10	273.36	43.00	667.01	0.25	2.50
北京	2008	626.3	13.4	2391.4	2.2	1012.6	52.00	100	1032.90	313.68	43.50	689.72	0.25	2.30
北京	2009	480.6	13.3	2511.8	2.2	1011.9	51.00	86	1069.40	368.11	44.40	739.15	0.25	2.60
北京	2010	522.5	12.6	2382.9	2.3	1012.2	51.00	88	1195.50	449.72	45.00	809.90	0.25	2.70
上海	2000	1331.7	17.2	1649.5	3.0	1026.1	76.33	128	1674.00	49.19	22.19	559.42	0.25	1.00
上海	2001	1276.8	17.2	1914.9	3.1	1024.2	75.83	127	1614.00	55.01	23.82	592.99	0.25	1.25
上海	2002	1435.4	17.5	1780.0	2.9	1020.2	73.00	134	1625.00	62.30	30.01	645.71	0.25	1.10
上海	2003	756.9	17.0	1714.0	3.2	1015.2	75.00	120	1711.00	71.90	35.20	745.97	0.25	1.30
上海	2004	1061.0	17.5	1965.0	3.4	1010.9	72.33	114	1742.00	83.51	36.03	821.44	0.25	1.58
上海	2005	1059.8	17.1	1829.0	3.1	1006.5	70.00	125	1778.00	95.16	37.00	921.97	0.25	1.42
上海	2006	1149.7	17.9	1678.1	2.8	1004.9	70.00	125	1815.00	107.04	37.30	990.15	0.25	1.32
上海	2007	1290.4	18.2	1616.9	2.9	1006.4	68.83	136	1858.00	119.70	37.60	1072.80	0.25	1.44
上海	2008	1086.5	17.2	1635.7	3.3	1012.3	70.00	136	1888.46	132.12	38.00	1138.22	0.25	1.52
上海	2009	1289.4	17.4	1680.9	2.9	1019.3	70.00	132	1921.00	147.11	38.10	1153.38	0.25	1.62
上海	2010	1128.9	17.2	1662.0	3.1	1026.5	68.67	137	2302.66	175.51	38.20	1295.87	0.25	1.60

2) 相关度计算C语言程序:

#include<stdio.h>

#include<math.h>

#define ru 0.20

void main()

{

int n, p, i, j;

n=26;p=12;

double data[26][13]=

{

1.00, 1331.70, 17.18, 1649.50, 3.0, 10261.00, 76.33, 128.00, 1674.00, 49.19, 22.19, 55
9.42, 0.25,1.25, 1276.80, 17.21, 1914.90, 3.1, 10242.00, 75.83, 127.00, 1614.00, 55.01, 23.82, 59
2.99, 0.25,1.10, 1435.40, 17.50, 1780.00, 2.9, 10202.00, 73.00, 134.00, 1625.00, 62.30, 30.01, 64
5.71, 0.25,

1.30, 756.90, 17.00, 1714.00, 3.2, 10152.00, 75.00, 120.00, 1711.00, 71.90, 35.20, 745.

参赛队号 #1370

```

97, 0.25,
1.58, 1061.00, 17.50, 1965.00, 3.4, 10109.00, 72.33, 114.00, 1742.00, 83.51, 6.03, 821.
44, 0.25,
1.42, 1059.80, 17.10, 1829.00, 3.1, 10065.00, 70.00, 125.00, 1778.00, 95.16, 37.00, 92
1.97, 0.25,
1.32, 1149.70, 17.90, 1678.10, 2.8, 10049.00, 70.00, 125.00, 1815.00, 107.04, 37.30, 9
90.15, 0.25,
1.44, 1290.40, 18.15, 1616.90, 2.9, 10064.00, 68.83, 136.00, 1858.00, 119.70, 37.60, 1
072.80, 0.25,
1.52, 1086.50, 17.20, 1635.70, 3.3, 10123.00, 70.00, 136.00, 1888.46, 132.12, 38.00, 1
138.22, 0.25,
1.62, 1289.40, 17.40, 1680.90, 2.9, 10193.00, 70.00, 132.00, 1921.00, 147.11, 38.10, 1
153.38, 0.25,
1.60, 1128.90, 17.20, 1662.00, 3.1, 10265.00, 68.67, 137.00, 2302.66, 175.51, 38.20, 1
295.87, 0.25,
1, 700.9, 12.7, 2418.7, 2.6, 1011.0, 51.0, 103.0, 767.2, 62.2, 33.2, 244.37, 0.25,
0.97, 430.9, 13.1, 2596.5, 2.5, 1012.9, 56.0, 76.0, 755.6, 78.4, 34.2, 263.60, 0.25,
0.92, 731.7, 13.1, 2420.7, 2.3, 1012.5, 62.0, 93.0, 759.3, 89.9, 35.6, 276.20, 0.25,
1.3, 266.9, 13.1, 2594.0, 2.4, 1012.5, 53.0, 86.0, 766.0, 95.1, 36.3, 297.26, 0.25,
1.52, 371.1, 12.8, 2667.2, 2.5, 1012.7, 54.3, 83.0, 842.1, 104.1, 36.5, 384.42, 0.25,
1.8, 338.9, 12.9, 2611.7, 2.4, 1012.9, 56.3, 78.0, 842.8, 114.5, 38.8, 399.94, 0.25,
2, 370.4, 13.2, 2588.4, 2.3, 1012.7, 49.0, 84.0, 867.1, 133.9, 40.6, 439.96, 0.25,
1.9, 444.9, 12.9, 2260.2, 2.5, 1013.3, 54.0, 93.0, 887.5, 163.1, 40.9, 467.60, 0.25,
2.2, 483.5, 13.5, 2515.4, 2.4, 1012.6, 49.2, 94.0, 909.8, 182.4, 41.9, 513.18, 0.25,
2.1, 410.7, 13.2, 2576.1, 2.4, 1012.8, 49.0, 79.0, 937.2, 209.7, 42.0, 570.53, 0.25,
2.4, 318.0, 13.4, 2192.7, 2.2, 1012.5, 53.0, 86.0, 963.4, 239.1, 42.5, 611.57, 0.25,
2.5, 483.9, 14.0, 2351.1, 2.2, 1012.6, 54.0, 78.0, 995.1, 273.4, 43.0, 667.00, 0.25,
2.3, 626.3, 13.4, 2391.4, 2.2, 1012.6, 52.0, 100.0, 1032.9, 313.7, 43.5, 689.71, 0.25,
2.6, 480.6, 13.3, 2511.8, 2.2, 1011.9, 51.0, 86.0, 1069.4, 368.1, 44.4, 739.14, 0.25,
2.7, 522.5, 12.6, 2382.9, 2.3, 1012.2, 51.0, 88.0, 1195.5, 449.7, 45.0, 809.90, 0.25
};
double avg[13], data1[26][13], sum; //初始化
for(j=0; j<p+1; j++)
{
    sum=0;
    for(i=0; i<n; i++)
    {
        sum=sum+data[i][j];
    }
    avg[j]=sum/n;
    printf("%lf ", avg[j]);
}

```

参赛队号 #1370

```

printf("\n");
printf("以上为比较数列的平均值\n");
for(j=0;j<p+1;j++)
{
    for(i=0;i<n;i++)
    {
        data1[i][j]=data[i][j]/avg[j];
    }
}
//均值化处理
for(i=0;i<n;i++)
{
    for(j=0;j<p+1;j++)
    {
        printf("%lf ",data1[i][j]);
    }
    printf("\n");
}
printf("以上为均值化数列\n");
//输出均值化数列

double delta_data[26][13];
for(i=0;i<n;i++)
{
    for(j=1;j<p+1;j++)
    {
        delta_data[i][j]=fabs(data1[i][j]-data1[i][0]);
    }
}
//计算各比较数列同参考数列在同一时期
的绝对差
for(i=0;i<n;i++)
{
    for(j=1;j<p+1;j++)
    {
        printf("%lf ",delta_data[i][j]);
    }
    printf("\n");
}
printf("以上为各比较数列同参考数列在同一时期的绝对差\n");
//输出各比较数
列同参考数列在同一时期的绝对差
double min=10000000,max=0;
for(i=0;i<n;i++)
{
    for(j=1;j<p+1;j++)

```

参赛队号 #1370

```

    {
        if(delta_data[i][j]>max) max=delta_data[i][j];
        if(delta_data[i][j]<min) min=delta_data[i][j];
    }
}
printf("绝对差最大值为%lf\n", max);
printf("绝对差最小值为%lf\n", min); //求出绝对差最大值
和最小值
double association[26][13];
for(i=0;i<n;i++)
{
    for(j=1;j<p+1;j++)
    {
        association[i][j]=(min+ru*max)/(delta_data[i][j]+ru*max);
    }
} //求元素的关联系数
for(i=0;i<n;i++)
{
    for(j=1;j<p+1;j++)
    {
        printf("%lf ", association[i][j]);
    }
    printf("\n");
}
printf("以上为各元素的关联系数\n"); //输出各元素的关联系数
double avg_association[13];
for(j=1;j<p+1;j++)
{
    sum=0;
    for(i=0;i<n;i++)
    {
        sum=sum+association[i][j];
    }
    avg_association[j]=sum/n;
    printf("%lf ", avg_association[j]);
}
printf("\n");
printf("以上为各指标关联度\n");
}

```

参赛队号 # 1370

3) Matlab 计算数据:

X=[

1331.70	17.18	1649.50	3.0	10261.00	76.33	128.00	1674.00	49.19
22.19	559.42	0.25						
1276.80	17.21	1914.90	3.1	10242.00	75.83	127.00	1614.00	55.01
23.82	592.99	0.25						
1435.40	17.50	1780.00	2.9	10202.00	73.00	134.00	1625.00	62.30
30.01	645.71	0.25						
756.90	17.00	1714.00	3.2	10152.00	75.00	120.00	1711.00	71.90
35.20	745.97	0.25						
1061.00	17.50	1965.00	3.4	10109.00	72.33	114.00	1742.00	83.51
36.03	821.44	0.25						
1059.80	17.10	1829.00	3.1	10065.00	70.00	125.00	1778.00	95.16
37.00	921.97	0.25						
1149.70	17.90	1678.10	2.8	10049.00	70.00	125.00	1815.00	107.04
37.30	990.15	0.25						
1290.40	18.15	1616.90	2.9	10064.00	68.83	136.00	1858.00	119.70
37.60	107.21	0.25						
1086.50	17.20	1635.70	3.3	10123.00	70.00	136.00	1888.46	132.12
38.00	113.82	0.25						
1289.40	17.40	1680.90	2.9	10193.00	70.00	132.00	1921.00	147.11
38.10	115.34	0.25						
1128.90	17.20	1662.00	3.1	10265.00	68.67	137.00	2302.66	175.51
38.20	129.59	0.25						
700.90	12.70	2418.70	2.60	1011.00	51.00	103.00	767.20	62.20
33.20	244.37	0.25						
430.90	13.10	2596.50	2.50	1012.90	56.00	76.00	755.60	78.40
34.20	263.61	0.25						
731.70	13.10	2420.70	2.30	1012.50	62.00	93.00	759.30	89.90
35.60	276.21	0.25						
266.90	13.10	2594.00	2.40	1012.50	53.00	86.00	766.00	95.10
36.30	297.26	0.25						
371.10	12.80	2667.20	2.50	1012.70	54.30	83.00	842.10	104.10
36.50	384.43	0.25						
338.90	12.90	2611.70	2.40	1012.90	56.30	78.00	842.80	114.50
38.80	399.94	0.25						
370.40	13.20	2588.40	2.30	1012.70	49.00	84.00	867.10	133.90
40.60	439.96	0.25						
444.90	12.90	2260.20	2.50	1013.30	54.00	93.00	887.50	163.10
40.90	467.61	0.25						
483.50	13.50	2515.40	2.40	1012.60	49.20	94.00	909.80	182.40

参赛队号 # 1370

```

41.90 513.18 0.25
410.70 13.20 2576.10 2.40 1012.80 49.00 79.00 937.20 209.70
42.00 570.54 0.25
318.00 13.40 2192.70 2.20 1012.50 53.00 86.00 963.40 239.10
42.50 611.57 0.25
483.90 14.00 2351.10 2.20 1012.60 54.00 78.00 995.10 273.40
43.00 667.01 0.25
626.30 13.40 2391.40 2.20 1012.60 52.00 100.00 1032.90 313.70
43.50 689.72 0.25
480.60 13.30 2511.80 2.20 1011.90 51.00 86.00 1069.40 368.10
44.40 739.15 0.25
522.50 12.60 2382.90 2.30 1012.20 51.00 88.00 1195.50 449.70
45.00 809.90 0.25 ];
y=[1.00 1.25 1.10 1.30 1.58 1.42 1.32 1.44 1.52 1.62 1.60 1.00 0.97 0.92 1.30
1.52 1.80 2.00 1.90 2.20 2.10 2.40 2.50 2.30 2.60 2.70]';
beta0=[2.0 1.0 -0.5 0.7 2.1 5.4 1.2 3.8 6.7 1.4 2.4 4.5];
betafit=nlinfit(X,y,'model',beta0)];
beta0=[2.0 1.0 -0.5 0.7 2.1 5.4 1.2 3.8 6.7 1.4 2.4 4.5];
betafit=nlinfit(X,y,'model',beta0)

```