五一数学建模竞赛

承　诺　书

我们仔细阅读了五一数学建模竞赛的竞赛规则。

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）与本队以外的任何人（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它公开的资料（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，我们愿意承担由此引起的一切后果。

我们授权五一数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

参赛题号（从 A/B/C 中选择一项填写）：B

参赛队号：T20764386196480

参赛组别（研究生、本科、专科、高中）：本科

所属学校（学校全称）：四川大学锦城学院

参赛队员：队员 1 姓名：杨海

队员 2 姓名：潘攀

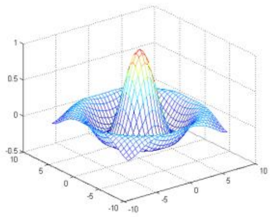
队员 3 姓名：阳滢滢

联系方式：Email：yanghai12721@163.com 联系电话：15528469959

日期：2021 年 5 月 01 日

（除本页外不允许出现学校及个人信息）

五 一 数 学 建 模 竞 赛



题 目： 消防救援问题

关键词：时间序列模型 最小二乘法 曲线拟合模型 图论模型 Floyd 算法

摘 要：

随着社会的不断进步，人们的消防意识逐渐提高，消防问题早已成为社会各界高度重视的问题。本文对某地区的消防救援问题进行了分析。

对于问题一，依据附件数据统计出各时间段的出警次数，根据各时间段出警次数的权值将 30 个值班人员规划到不同时间段值班，考虑以下约束：每个时间段值班人数不少于 5 人、全天值班人数不多于 30 人，根据权值算出某个时间段需要安排的值班人数，不足 5 人的时间段按 5 人分配，剩余 25 人再依据权值分配值班，最后得出每年 2 月、5 月、8 月、11 月中第一天三个时间段的值班人数。

对于问题二，统计各月份出警数据，用 SPSS 软件分析出异常值，并对异常值进行替换处理，可视化处理后发现数据具有周期性，且周期值为 12，用 Daniel 检验法检验时间序列的平稳性，建立消防救援出警次数的预测模型，将 2020 年的数据代入该模型，预测数据较为准确，将 2020 年各月真实数据代入模型得出 2021 年各月出警次数的预测值。

对于问题三，统计各类事件在各月发生次数，建立了多种关系模型，通过比较拟合度最终采用最小二乘法拟合数据，确定了最优模型。

对于问题四，计算各类事件密度，用 MATLAB 工具箱建立出事件密度与区域相关性的三维图像模型，得出各区域相关性最强的事件类型为⑦类事件。

对于问题五，在问题四的基础上，先求出各区域的人口密度，再将人口密度进行升序排列，得出人口密度与事件密度的散点图，先剔除 P 点的事件密度分析散点图，得出各类事件密度随人口密度在小范围内增长时的变化情况，再用曲线拟合模型对原散点图进行曲线拟合，分别得出各类事件密度随人口密度在大范围内增长时的变化情况，分别综合对比所有散点图和拟合图，得出了事件密度与人口密度的变化规律。

对于问题六，首先依据附件 1 和附件 2 结合选址类问题分析出影响最终决策的因子有：已有消防站对各区域的救援程度、人口密度、事件发生次数、最短路径，分别计算出各种影响因子的权值，利用图论模型的 Floyd 算法求出各区域的最短距离后得出距离影响因子，综合所有影响因子的权值，分别计算出消防站建立在不同区域后对其它所有区域的贡献值，绘制各区域贡献值的直方图，经对比得出消防站建在 D 点最佳，用相同模型得出 2021 年-2029 年消防站应依次建于 D 区、K 区、F 区。

一、问题重述

随着我国经济的高速发展，城市空间环境复杂性急剧上升，各种事故灾害频发，安全风险不断增大，消防救援队承担的任务也呈现多样化、复杂化的趋势，探究各类消防事件在不同时间、空间的发生规律对消防救灾具有重要意义。

问题如下：

（1）将每天分为三个时间段（0:00-8:00 为时段Ⅰ，8:00-16:00 为时段Ⅱ，16:00-24:00 为时段Ⅲ），每个时间段安排不少于 5 人值班。假设消防队每天有 30 人可安排值班，根据附件数据建立数学模型确定消防队在每年 2 月、5 月、8 月、11 月中第一天的三个时间段各应安排多少人值班。

（2）以该地 2016 年 1 月 1 日至 2019 年 12 月 31 日的数据为基础，以月份为单位，建立消防救援出警次数的预测模型；以 2020 年 1 月 1 日至 2020 年 12 月 31 日的数据作为模型的验证数据集，评价模型的准确性和稳定性，并对 2021 年各月份的消防救援出警次数进行预测，完成表 1。

（3）依据 7 种类别事件的发生时间，建立各类事件发生次数与月份关系的多种数学模型，以拟合度最优为评价标准，确定每类事件发生次数的最优模型。

（4）根据图 1，请建立数学模型，分析该地区 2016-2020 年各类事件密度在空间上的相关性，并且给出不同区域相关性最强的事件类别（事件密度指每周每平方公里内的事件发生次数）。

（5）依据附件 2，请建立数学模型，分析该地各类事件密度与人口密度之间的关系（人口密度指每平方公里内的人口数量）。

（6）目前该地有两个消防站，分别位于区域 J 和区域 N，请依据附件 1 和附件 2，综合考虑各种因素，建立数学模型，确定如果新建 1 个消防站，应该建在哪个区域，如果在 2021-2029 年每隔 3 年新建 1 个消防站，则应依次建在哪些区域

二、 问题分析

2.1问题一的分析

该问题要求确定消防队在每年2月、5月、8月、11月中第一天的三个时间段各应安排多少人值班，查阅国家统计局发布的全国火灾数据以及附件中提供的原始数据，可以判断出接警量在Ⅱ、Ⅲ时段上有聚集现象，计算5年以来这4天各时段发生事件的总和，分别计算每天每个时段接警量在该天总接警量上的权值，通过权值按比例分配30个名额，不足5个名额按5个分配，剩余名额再按其余时段的权值来分配。

2.2问题二的分析

该问题属于预测类问题，要求以月份为单位，通过2016年-2019年的出警数据建立出警次数的预测模型。用2020年的数据检验模型的准确性和稳定性并对2021年各月份的出警次数进行预测。接警量具有一定的随机性，初步预测消防事件与气候有关，且气候随月份呈周期性变化，首先制作出前四年各月份与出警量的关系图，观察发现接警量随月份呈周期性变化，且周期值为12，对于周期性预测问题采用时间序列模型来解决，首先检验序列平稳性，观察序列的总体变化趋势，调整模型以达到最优拟合。

2.3 问题三的分析

该问题要求依据7种类别事件的发生时间，建立各类事件发生次数与月份关系的多种数学模型，以拟合度最优为评价标准，确定每类事件发生次数的最优模型。首先整理数据，得到各月份各类事件发生的次数，通过最小二乘法拟合数据，用R^2、RMSE判断模型拟合效果，通过MATLAB程序求解参数，最后得出拟合度最优的模型。

2.4问题四的分析

该问题要求分析该地区2016-2020年各类事件密度在空间上的相关性，并且给出不同区域相关性最强的事件类别，首先需要整理数据计算出各区域各类事件密度，通过 MATLAB建立模型，分析各类事件在空间上的相关性，最后通过图像可以直观的判断各区域相关性最强的事件类型。

2.5问题五的分析

该问题要求分析该地各类事件密度与人口密度之间的关系，该问题可在问题四的基础上求解，首先计算出各区域的人口密度，通过数据分析判断人口密度与事件密度的关系，绘制图像判断判断两者的相关性，通过散点图确定二者关系与二次拟合曲线比较相似，即建立二次关系表达式，同时利用MATLAB进一步统计计算得到各类事件与人口密度的关系拟合曲线。

2.6问题六的分析

该问题要求依据附件1和附件2，综合考虑各种因素，建立数学模型选出1个区域来建立消防站，首先依据附件1和附件2我们进行综合分析得到消防站的建立应考虑各区域以及各区域之间的最短距离、各地域的事件发生的概率及人口密度、各地域附近是否已有消防站、甚至是交通是否通畅等因素。然后我们可以运用图论模型中的Floyd算法求出各区域间的最短距离，再用问题五求出的各区域人口密度，和待求的各区域事件发生权重和已有消防站对各区域救援程度的权重，根据求出的权重计算出区域需要建立消防站的重要程度后再进行比较，得出最适合建立消防站的区域。2021年-2029年用相同模型进行计算。

三、 模型假设

1. 假设短期内没有外部因素对该地区的接警量产生巨大影响。
2. 假设各地区人口数量短时间内没有较大的迁入或迁出。
3. 假设接警量只受附件所提供指标的影响。
4. 假设把每个区域当作一个点，不考虑事件在区域内发生的位置。
5. 假设消防车可以以恒定的速度行驶至各区域。

四、 符号说明

符号	符号说明
Y	指标数值的最终变动
T	长期趋势变动
S	季节变动
C	循环变动
I	不规则变动
a	显著水平
a_t	截距
b_t	斜率
L	损失函数
R^2	拟合优度（可决系数）
SST	总体平方和
SSE	误差平方和
SSR	回归平方和

五、 模型的建立与求解

5.1 问题一的模型建立与求解

5.1.1 建模的前期准备

1.数据可视化

首先从 EXCEL 读取数据到 MATALAB 中，为了更加直观地感受数据，对数据进行了可视化处理，并以图片的形式将数据特征表现出来。观察图像，我们可以发现 8 点-24 点为接警高峰期。

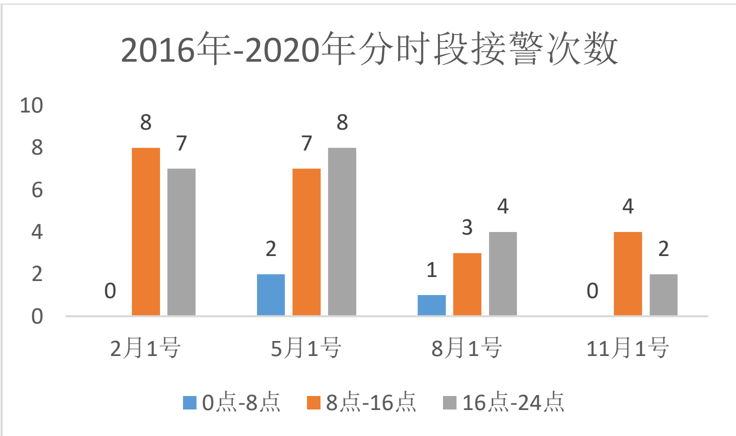


图 1　2016-2020 年分时段接警次数分析图

通过查阅资料得到全国火灾发生数量与时段的关系图：

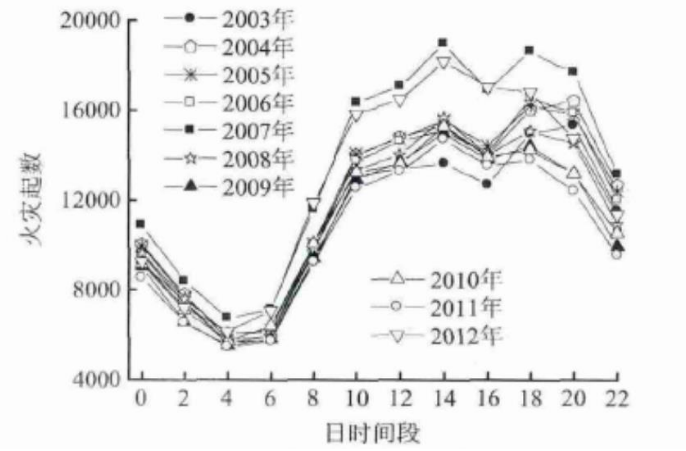


图 2 全国火灾起数与时段的关系

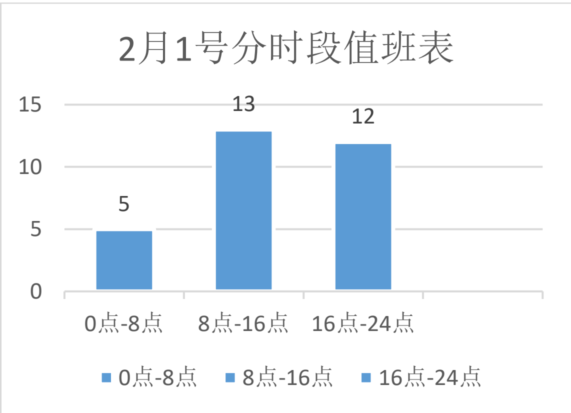
观察上图，我们发现 8-24 点火灾事故高发，结合我们从上图得出的结论可以初步断定接警现象在Ⅱ、Ⅲ时段上有聚集现象

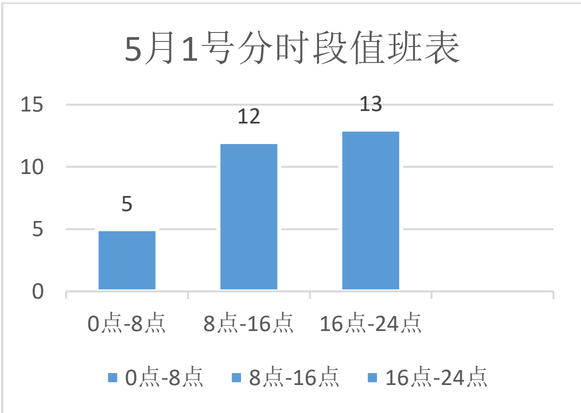
5.1.2 模型的建立

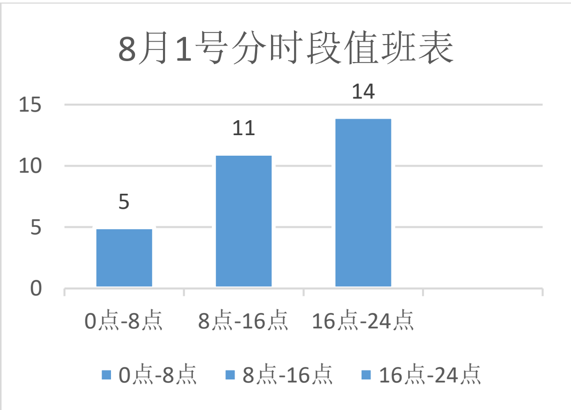
不同时段的值班分配与该时段的接警次数有关，利用各时段接警量在全天总接警量中所占权重确定各时段的值班人数。

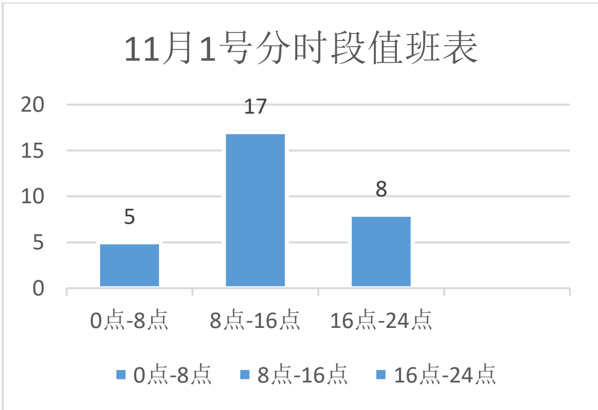
5.1.3 模型的求解

结合各时间段所占权重，考虑约束条件：每个时段的值班人数不低于5人、总人数不超过30人，得出以下结果。









5.2 问题二模型建立与求解

5.2.1 模型的前期准备

1.数据预处理

剔除异常值，异常值指样本中的个别值，其数值明显偏离它们所属样本的其余观测值,也称异常数据，离群值。如果存在异常值但不剔除，会影响结果的准确性，通过客观的观察分析,发现数据中存在少量不合理的数据;我们采用 SPSS 软件对数据进行分析,用 Z 分数识别异常值，将 Z 分数低于-3 或高于 3 的数据看成是异常值，复查错误值、极端值，予以剔除，提高数据的准确性[1]。

2. 数据整理

以月份为单位，整理 2016 年到 2019 年每月的接警数量，关系如下图。

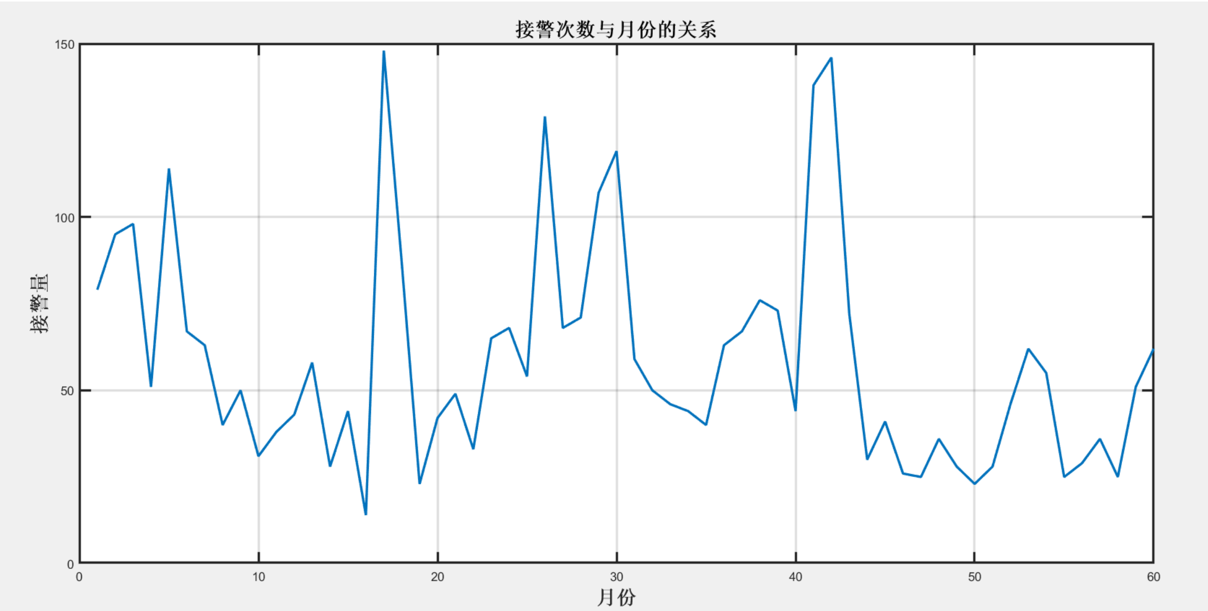


图 3 2016 年-2020 年每月接警次数图

观察到救援接警次数与月份之间有明显的周期性关系，我们可以采用时间序列模型进行建模预测。

2. 利用时间序列进行分析

时间序列也称动态序列，将指标数值按照时间顺序排列而成的数值序列，时间序列分析分成三大部分，分别是描述过去、分析规律和预测未来，数据的变化规律一般包含长期变动趋势，季节变动规律，周期变动规律，不规则变动，且四种变动相互影响，具有乘积关系[2]，即：

$$Y = T \times S \times C \times I$$

本文结合 Spss 软件处理序列数据，并对时间序列数据进行建模。

（1）时间序列分析步骤：

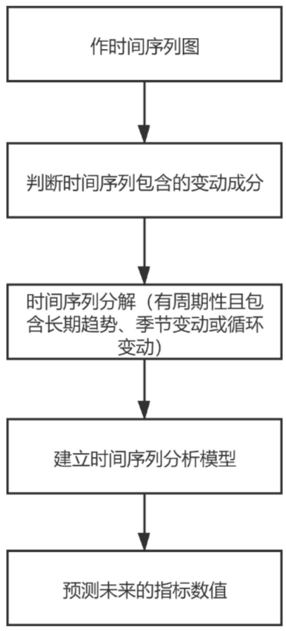


图 4 时间序列分析步骤

（2）Daniel 检验法检验平稳性[3]：

观察序列是否存在着趋势，不检测自相关。该方法建立在 Spearman 相关系数基础之上，利用非参数方法中 Spearman 秩相关系数检验两变量是否相关的原理来检验 yt 与时间 t 是否存在着同时增加或减少的趋势。对 n 对（yt,t）计算 Spearman 相关系数，然后对小样本时采用附表或大样本时采用正太近似所确定的临界点检验其显著性：

基本步骤如下：

H_0：序列没有趋势

H_1：序列存在(向上或向下)趋势

检验统计是:

(1) 小样本 :$n \leqslant 30$

$$r_s = 1 - 6\sum_{t=1}^{n} \sim \mathrm{d}t^2 / n(n^2-1)$$

其中 $\mathrm{d}t = t - y_t$ 的秩 $= t - R(y_t)$

(2) 大样本，$n > 30$

$$Z = \frac{r_s - \mu_r}{\sigma_{rs}}\left(\mu_{rs} = 0, \sigma_{rs} = \frac{1}{\sqrt{n-1}}\right)$$

判决规则 :给定显著性水平 α, 当 $n \leqslant 30$ 时, 如果 $|r,| > r_{\{}\alpha/2$， 则拒绝 H_0

当 $n > 30$ 时, 如果 $|z| > z_{a/2}$ 则拒绝 H_0

结论：$n > 30$，$|z| > z_{a/2}$ 拒绝 H_0，以 $(1-a)\times 100\%$ 的置信度认为序列有趋势，经计算 r_s 为负, 趋势为向下。

5.2.2 模型的建立

（1）移动平均(moving average)预测模型 [4]

由于时间序列的数值受周期变动和不规则变动的影响，起伏较大，不易显示出发展趋势时，所以我们选择使用移动平均法，消除这些因素的影响，经过观察分析数据存在直线趋势与周期波动，用简单移动平均法和加权移动平均法来预测就会出现滞后偏差。因此，需要使用趋势移动平均法修正，利用移动平均滞后偏差的规律来建立直线趋势的预测模型。

一次移动的平均数：

$$M_t^{(1)}=\frac{1}{N}(y_t+y_{t-1}+\cdots+y_{t-N+1})$$

在一次移动平均的基础上再进行一次移动平均就是二次移动平均，其计算公式为

$$M_t^{(2)}=\frac{1}{N}(M_t^{(1)}+\cdots+M_{t-N+1}^{(1)})=M_{t-1}^{(2)}+\frac{1}{N}(M_t^{(1)}-M_{t-N}^{(1)})$$

利用移动平均的滞后偏差建立直线趋势预测模型。，设时间序列 $\{y_t\}$ 从某时期开始具有直线趋势，且认为未来时期也按此直线趋势变化，则可设此直线趋势预测模型为

$$\hat{y}_{t+T}=a_t+b_tT,\quad T=1,2,\cdots \qquad (7)$$

其中 t 为当前时期数；T 为由 t 至预测期的时期数；a_t 为截距；b_t 为斜率。两者又称为平滑系数。现在，我们根据移动平均值来确定平滑系数，由模型（7）可知

$$\begin{aligned}
&a_t=y_t\\
&y_{t-1}=y_t-b_t\\
&y_{t-2}=y_t-2b_t\\
&\quad\cdots\\
&y_{t-N+1}=y_t-(N-1)b_t
\end{aligned}$$

所以

$$\begin{aligned}
M_t^{(1)}&=\frac{y_t+y_{t-1}+\cdots+y_{t-N+1}}{N}=\frac{y_t+(y_t-b_t)+\cdots+[y_t-(N-1)b_t]}{N}\\
&=\frac{Ny_t-[1+2+\cdots+(N-1)]b_t}{N}=y_t-\frac{N-1}{2}b_t
\end{aligned}$$

因此

$$y_t-M_t^{(1)}=\frac{N-1}{2}b_t \qquad (8)$$

由式（7)，类似式（8）的推导，可得

$$y_{t-1}-M_{t-1}^{(1)}=\frac{N-1}{2}b_t$$

所以

$$y_t-y_{t-1}=M_t^{(1)}-M_{t-1}^{(1)}=b_t$$

类似式（8）的推导，可得

$$M_t^{(1)}-M_t^{(2)}=\frac{N-1}{2}b_t \qquad (11)$$

于是，由式（8）和式（ 11 ）可得平滑系数的计算公式为

$$\begin{cases}a_t=2M_t^{(1)}-M_t^{(2)}\\ b_t=\dfrac{2}{N-1}(M_t^{(1)}-M_t^{(2)})\end{cases} \qquad (12)$$

5.2.3 模型的求解

（1）根据计算公式，取N值，分别计算一次和二次移动平均值，$M_{21}^{(1)}$，$M_{21}^{(2)}$，再由公式（12），求 a21 和 b21：

$$a_{21}=2M_{21}^{(1)}-M_{21}^{(2)}$$

$$b_{21}=\frac{2}{6-1}(M_{21}^{(1)}-M_{21}^{(2)})$$

于是，得到相应t值时直线趋势预测模型，并通过时间序列预测 2020 接警量，通过 MATLAB 程序得到下图（程序见附录）：

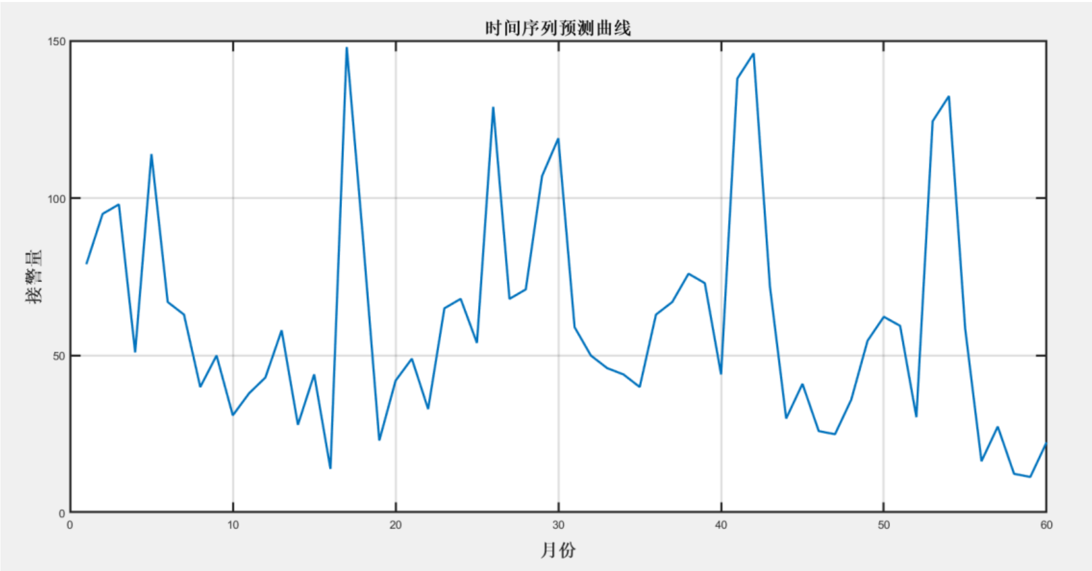


图 5 时间序列预测模型预测 2020 年接警量

（2）根据 2020 年的数据，用 matlab 的 nlinfit 和 lsqcurvefit 函数[5]来进行参数优化，适当调整参数，进行回归拟合，然后对 2021 年的数据进行预测分析得到下图：

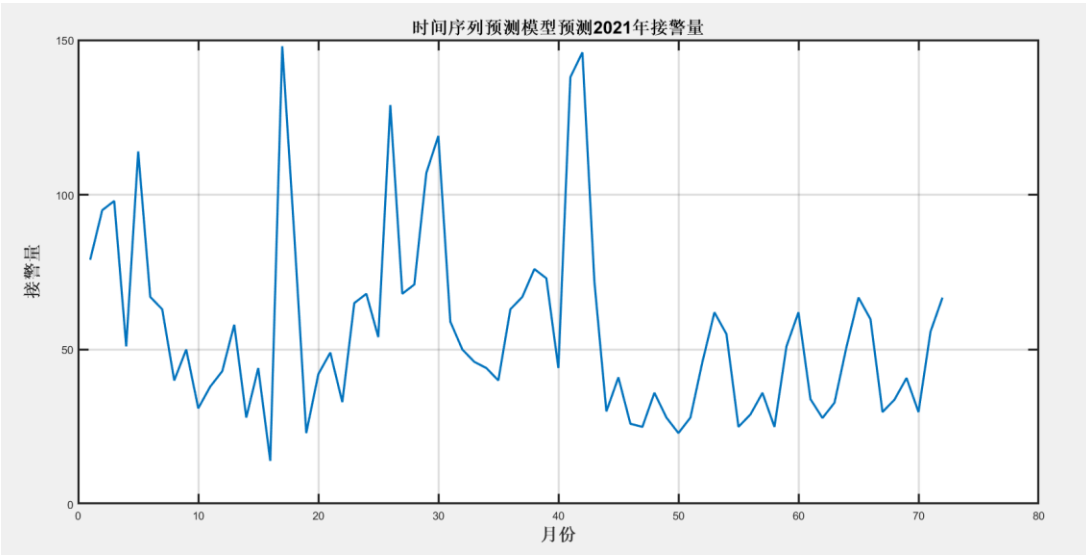


图 6 时间序列预测模型预测 2021 年接警量

图 6 时间序列预测模型预测 2021 年接警量

编写 MATLAB 程序（见附录）计算得到 2021 年每月接警量，小数部分进行四舍五入，详细数据见下表：

表 1　预测 2021 年每月接警量

月份	预测值（次）
2021 年 1 月	46
2021 年 2 月	44
2021 年 3 月	48
2021 年 4 月	66
2021 年 5 月	82
2021 年 6 月	75
2021 年 7 月	45
2021 年 8 月	49
2021 年 9 月	56
2021 年 10 月	45
2021 年 11 月	71
2021 年 12 月	82

（3）评价模型的稳定性与准确性：

时间序列模型预测是根据过去的变化趋势预测未来的发展，理论基础是客观事物发展的连续规律性，因此预测结果不会发生突然跳跃式变化，而是渐进变化的，因此具有相对较强的稳定性，但未来发展变化规律和发展水平，不一定与其历史和现在的发展变化规律完全一致，时间序列预测法侧重突出时间的影响因素，外界因素影响没办法考虑到，因此存在一定的预测误差的缺陷，当遇到外界发生较大变化，往往出现较大偏差导致准确性不足，本文模型对于中短期预测较为准确，长期预测效果较差。

5.3 问题三的模型建立与求解

5.3.1 模型的前期准备

1．整理数据

与问题二类似，通过分析处理原始数据整理出各类事件发生次数与月份的大致关系。

2．利用最小二乘法进行数据拟合

最小二乘法的定义：$\hat{y}_i = kx_i + b$，$\hat{k}, \hat{b} = \arg\min_{k,b}\left(\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2\right)$

5.3.2 模型的建立

采用拟合模型反映各类事件发生次数与月份关系。

设样本点为 $(x_i, y_i), i = 1, 2, \cdots, n$，我们设置的拟合曲线为 $y = kx + b$

令拟合值 $\hat{y}_i = kx_i + b$

那么 $\hat{k},\hat{b} = \underset{k,b}{\operatorname{argmin}}\left(\sum^{n}(y_i-\hat{y}_i)^2\right) = \underset{k,b}{\operatorname{argmin}}\left(\sum_{i=1}^{n}(y_i-kx_i-b)^2\right)$

令 $L=\sum_{i=1}^{n}(y_i-kx_i-b)^2$，现要找 k,b 使得 L 最小

（L 在机器学习中被称为损失函数，在回归中也常被称为残差平方和）

$$\begin{cases}\dfrac{\partial L}{\partial k}=-2\sum_{i=1}^{n}x_i(y_i-kx_i-b)=0\\ \dfrac{\partial L}{\partial b}=-2\sum_{i=1}^{n}(y_i-kx_i-b)=0\end{cases} \Rightarrow \begin{cases}\sum_{i=1}^{n}x_iy_i=k\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}y_i=k\sum_{i=1}^{n}x_i+bn\end{cases} \Rightarrow \begin{cases}n\sum_{i=1}^{n}x_iy_i=kn\sum_{i=1}^{n}x_i^2+bn\sum_{i=1}^{n}x_i\\ \sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i=k\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i+bn\sum_{i=1}^{n}x_i\end{cases}$$

$$n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i=kn\sum_{i=1}^{n}x_i^2-k\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i \Rightarrow \hat{k}=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}y_i\sum_{i=1}^{n}x_i}{n\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i}$$

同理我们可得到：$\hat{b}=\dfrac{\sum_{i=1}^{n}x_i^2\sum_{i=1}^{n}y_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_iy_i}{n\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}x_i}$

5.3.3 模型的求解

建立事件类型与月份之间关系的数学模型，运用 MATLAB 利用原始数据每类事件在每月的接警数量进行拟合，

（1）评价拟合度的好坏：

拟合优度（可决系数） R^2

总体平方和 SST: $SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$

误差平方和 SSE: $SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$

回归平方和 SSR: $SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$

可以证明：$SST=SSE+SSR$（ 要用到我们求导得到的两个等式 ）拟合优度：$0\le R^2=\dfrac{SSR}{SST}=\dfrac{SST-SSE}{SST}=1-\dfrac{SSE}{SST}\le 1$

R^2 越接近 1 ，说明误差平方和越接近 0 ，误差越小说明拟合的越好。

（注意：R^2 只能用于拟合函数是线性函数时，拟合结果的评价）

表 2 各类事件的最优拟合度和均方根误差表

事件类型	R^2	RMSE
第一类	0.9604	13.17
第二类	0.607	5.63
第三类	0.9105	58.5
第四类	0.8011	9.299
第五类	0.837	4.923
第六类	0.9328	8.105
第七类	0.9846	37.27

5.4 问题四的模型建立与求解

5.4.1 模型的前期准备

1.数据整理

对原始数据进行整理，计算出各区域各类事件发生的总数。

2.指标含义

我们不妨用各区域平均每周事件发生总数与各区域面积的比值来代表事件密度，由于周数是一个常数，我们可以直接用事件发生总数来代替它。

即：各区域各类事件密度=各区域各类事件发生总数/各区域面积 ①

5.4.2 模型的建立与求解

（1）通过公式①计算出各区域各类事件密度，由于结果较小，我们将数据扩大 100 倍使用，结果如下表：

表 3 各区域各类事件密度表

事件密度	事件 1	事件 2	事件 3	事件 4	事件 5	事件 6	事件 7
地区 A	21.1111	6.6667	24.4444	21.1111	1.1111	3.3333	96.6667
地区 B	15.8333	2.5000	13.3333	5.8333	0	1.6667	58.3333
地区 C	13.6364	1.1364	15.9091	10.2273	2.2727	0	61.3636
地区 D	25.3333	12	33.3333	17.3333	2.6667	2.6667	110.6667
地区 E	22.5225	3.6036	56.7568	8.1081	1.8018	2.7027	125.2252
地区 F	30.5882	1.1765	29.4118	10.5882	2.3529	1.1765	104.7059
地区 G	17.6991	1.7699	29.2035	16.8142	2.6549	1.7699	103.5398
地区 H	16.1290	6.4516	11.8280	11.8280	2.1505	0	64.5161
地区 I	4.8000	0.8000	8	4.8000	0	0.8000	39.2000
地区 J	27.0270	17.5676	37.8378	16.2162	1.3514	10.8108	98.6486
地区 K	15.1515	2.2727	24.2424	7.5758	0	1.5152	85.6061
地区 L	3.9063	1.5625	17.9688	10.9375	2.3438	4.6875	37.5000
地区 M	12.6050	5.0420	42.0168	10.0840	2.5210	5.0420	104.2017
地区 N	14.6067	5.6180	33.7079	6.7416	0	1.1236	49.4382
地区 P	1750	500	2350	1280	190	2330	8010

（2）通过表 4 制作出各类事件密度在空间上的分布图：

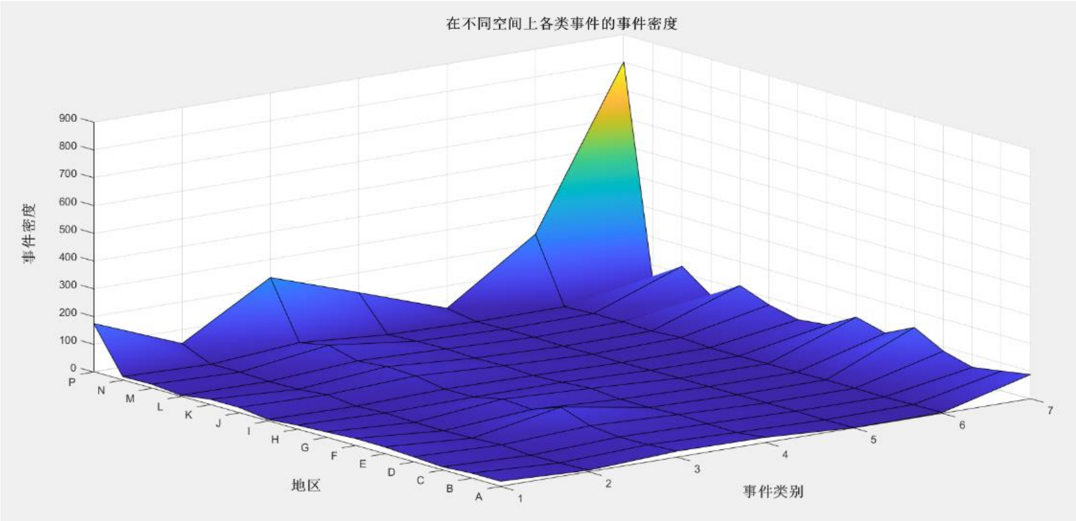


图 7　各类事件密度在空间上的分布图

（3)观察上图我们可以得出结论：P 区的各类事件密度在所有区域中都是最高的，且各区域事件密度最高的事件类型均为第 7 类事件，所以各区域相关性最强的事件也均为第 7 类事件。

5.5 问题五的模型建立与求解

5.5.1 模型的前期准备

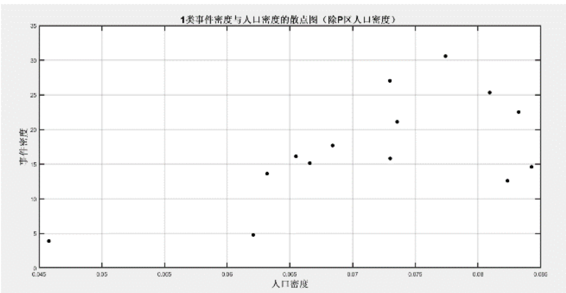
1.数据整理

对原始数据利进行整理，结合区域面积计算出各区域人口密度。

（1）指标含义

人口密度：每平方公里内的人口数量

（2）观察原始数据不直观，不容易发现其规律，利用散点图来观察各事件密度与人口密度之间的关系，7 类事件的事件密度随人口密度的变化情况分别如下：



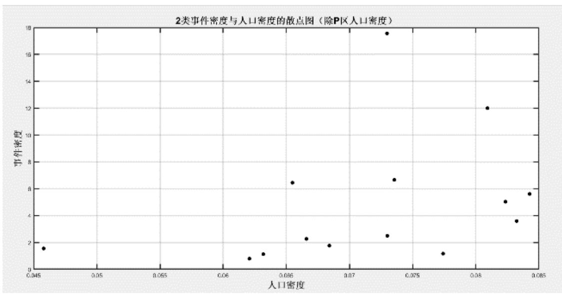
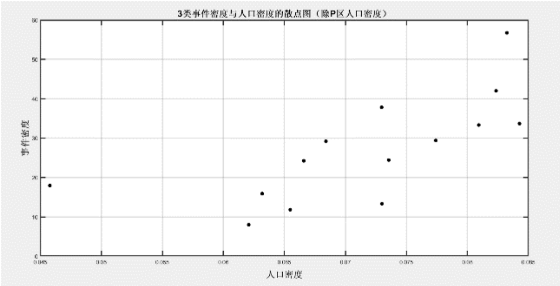


图 8　1 类事件密度与人口密度的散点图　　图 9　2 类事件密度与人口密度的散点图



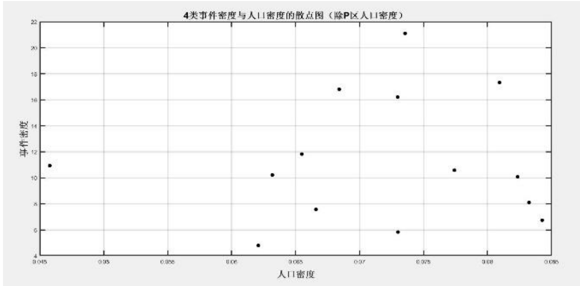
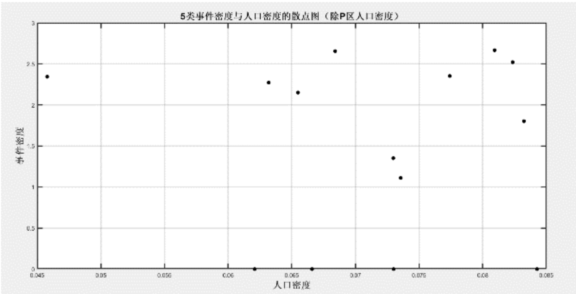


图 10　3 类事件密度与人口密度的散点图　　图 11　4 类事件密度与人口密度的散点图



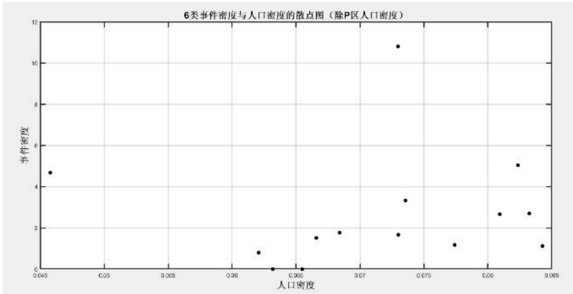


图 12　5 类事件密度与人口密度的散点图　　图 13　6 类事件密度与人口密度的散点图

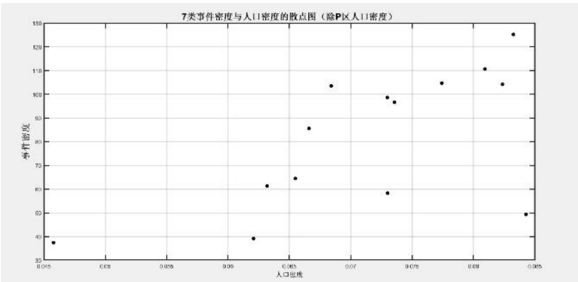


图 14　7 类事件密度与人口密度的散点图

2.结论

观察上图，我们可以得出以下结论：

（1）对于第 1 类事件，事件密度先随着人口密度上升而上升，事件密度达到峰值时，事件密度随人口密度上升而下降；

（2）对于第 2、3、6、7 类事件，总体上事件密度随着人口密度上升而上升，2、6 类上升较为缓慢，且 2 类事件具有较强随机性；

（3）对于第 4、5 类事件，事件密度和人口密度没有特别明显的相关性

5.5.2 模型的建立与求解

（1）采用二次函数拟合算法，通过 Matlab 建立拟合模型，在给定一组数据序列 $(x_i, y_i), \mathrm{i}=0,1,2\ldots\mathrm{m}$，用二次多项式拟合这组数据时，设 $\mathrm{p(x)} = a_0 + a_1 x + a_2 x^2$，拟合函数与数据序列的均方误差如下：

$$Q(a_0, a_1, a_2) = \sum_{i=1}^{m} (p(x_i) - y_i)^2 = \sum_{i=1}^{m} (a_0 + a_1 x_i + a_2 x_i^2 - y_i)^2$$

由多元函数的极值原理， $Q(a_0, a_1, a_2)$ 的极小值满足

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = 2\sum_{i=1}^{m}(a_0 + a_1 x_i + a_2 x_i^2 - y_i) = 0 \\ \dfrac{\partial Q}{\partial a_1} = 2\sum_{i=1}^{m}(a_0 + a_1 x_i + a_2 x_i^2 - y_i)x_i = 0 \\ \dfrac{\partial Q}{\partial a_2} = 2\sum_{i=1}^{m}(a_0 + a_1 x_i + a_2 x_i^2 - y_i)x_i^2 = 0 \end{cases}$$

整理得出二次多项式函数拟合方程：

$$\begin{pmatrix} m & \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 \\ \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i^3 & \sum_{i=1}^{m} x_i^4 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} x_i^2 y_i \end{pmatrix}$$

（3）根据散点图的观察确定 x 和 y 的关系与二次曲线比较相似，所以采用二次函数来拟合各事件密度与人口密度关系，其中 1 类事件密度与人口密度关系的拟合图如下：

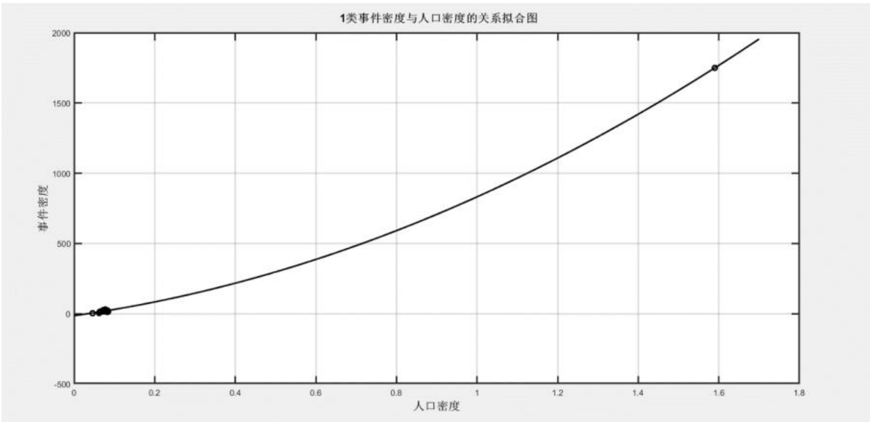


图 15 1 类事件密度与人口密度关系的拟合图

根据上文建立的模型，利用 Matlab 软件进一步计算出其余各类事件密度与人口密度关系的拟合曲线，将所有曲线绘制在一张三维图中，如下图所示：

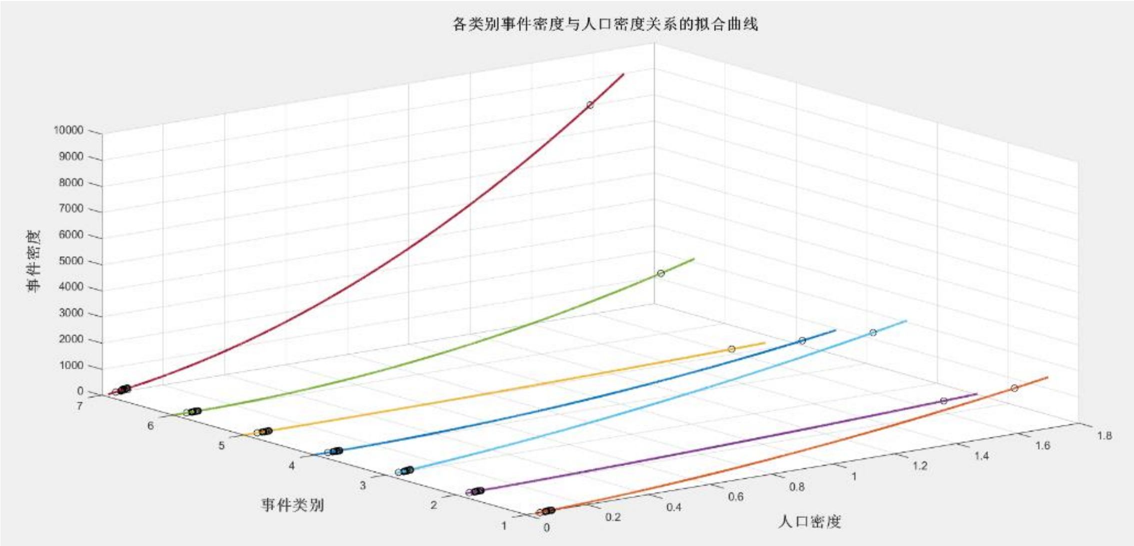


图 16　各类事件密度与人口密度关系的拟合曲线

我们根据该图可以得到以下结论：

（1）对于所有类别事件，当人口密度在总体上持续增加时，事件密度也会持续增加。

（2）对于第 7 类事件，事件密度随人口密度上升得较为剧烈；

（3）对于 2、5 类事件，事件密度随人口密度上升得较为缓慢；

5.6 问题六的模型建立与求解

5.6.1 模型的建立

（1）图论模型：

Floyd 算法的基本原理：Floyd 算法是寻找加权图中任意两点之间最短路径的重要算法。其基本思想是任意节点 A 到任意节点 B 的最短路径不外乎 2 种可能，从 A 直接到 B，或从 A 经过若干个节点 X 到 B。假设 Dis(AB）为节点 A 到节点 B 的最短路距离，对于每一个节点 X，检查 Dis(AX)+Dis(XB)< Dis(AB)是否成立，若成立，证明从 A 到 X 再到 B 的路径比 A 的路径短，则更新 Dis(AB)=Dis(AX)+Dis(XB），遍历所有节点 X 后，Dis(AB）中记录的便是 A 到 B 的最短路距离。[6]

（2）算法设计和符号说明：

n：共有 n=i，j(i，j=1,2,3⋯,15)个区域来表示 A~P 区域

m：已知消防站的区域

D_{ij}：i 区域到 j 区域的最短路径

$\mathrm{A_i}$：i 区域面积

$\mathrm{P_i}$：i 区域人口总数

$\mathrm{T_i}$：i 区域事件发生总次数

$\mathrm{H_{mj}}$：m 区域对 j 区域的救援程度权值

$\mathrm{PP_i}$：人口密度$\mathrm{PP_i}=\frac{\mathrm{Pi}}{\mathrm{Ai}}$

$\mathrm{TH_i}$：i 区域发生事件的权值$\mathrm{TH_i}=\frac{\mathrm{Ti}}{\sum_{\mathrm{i=1}}^{\mathrm{n}}\mathrm{Ti}}$

$\mathrm{C_i}$：i 区域对其他区域的总贡献程度$\mathrm{C_i}=\sum_{\mathrm{j=1}}^{\mathrm{n}}\frac{1}{D_{ij}}\times\mathrm{H_{mj}}\times\mathrm{PP_i}\times\mathrm{TH_i}$

总贡献程度即该区域对其它所有区域的救援帮助能力之和。由于距离越远救援能力越差，所以我们用距离的倒数$\frac{1}{D_{ij}}$乘以各地受已有消防站的影响因子H_{mj}乘以人口密度因

子PP_i乘事件密度因子TH_i求和，由于结果较小，我们不妨将结果扩大 10000 倍作为总贡献值，虽然 P 区域 5 年来发生的事件总和为 1641 起，远远超过其它区域，但考虑到 P 区域面积较小，且人口密度是其它区域的 20 倍左右，我们不考虑在 P 点建立消防站。经计算，得到分别在所有区域（已有消防站区域除外）新建消防站的总贡献值矩阵C_i

5.6.2 模型的求解

（1）使用图论模型的 Floyd 算法求出各区域间的最短路径（代码见附录）结果如下表：

表 4 各区域之间的最短路径

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	P
A	0	11.1	19.1	11.4	25.7	18.3	30.7	39.9	48.9	17.8	23.2	27.6	8.2	21.4	19.9
B	11.1	0	8.2	12.8	19.3	17.6	33.4	51	60	25.5	34.3	38.7	19.3	22.8	21.3
C	19.1	8.2	0	7.7	11.1	9.4	28.3	46.6	55.6	20.4	29.9	34.3	22	17.7	16.2
D	11.4	12.8	7.7	0	14.3	6.9	20.6	38.9	47.9	12.7	22.2	26.6	14.3	10	8.5
E	25.7	19.3	11.1	14.3	0	7.4	29.2	51.7	60.7	25.5	35	39.4	28.6	18.6	22.8
F	18.3	17.6	9.4	6.9	7.4	0	21.8	44.3	53.3	18.1	27.6	32	21.2	11.2	15.4
G	30.7	33.4	28.3	20.6	29.2	21.8	0	26.8	35.8	12.9	13.4	14.5	22.5	10.6	16.5
H	39.9	51	46.6	38.9	51.7	44.3	26.8	0	9	26.2	16.7	12.3	31.7	33.1	30.4
I	48.9	60	55.6	47.9	60.7	53.3	35.8	9	0	35.2	25.7	21.3	40.7	42.1	39.4
J	17.8	25.5	20.4	12.7	25.5	18.1	12.9	26.2	35.2	0	9.5	13.9	9.6	6.9	4.2
K	23.2	34.3	29.9	22.2	35	27.6	13.4	16.7	25.7	9.5	0	4.4	15	16.4	13.7
L	27.6	38.7	34.3	26.6	39.4	32	14.5	12.3	21.3	13.9	4.4	0	19.4	20.8	18.1
M	8.2	19.3	22	14.3	28.6	21.2	22.5	31.7	40.7	9.6	15	19.4	0	16.5	13.8
N	21.4	22.8	17.7	10	18.6	11.2	10.6	33.1	42.1	6.9	16.4	20.8	16.5	0	5.9
P	19.9	21.3	16.2	8.5	22.8	15.4	16.5	30.4	39.4	4.2	13.7	18.1	13.8	5.9	0

（2）分别计算在各区域（已建消防站区域除外）新建消防站后的总贡献值矩阵，建立如下直方图：

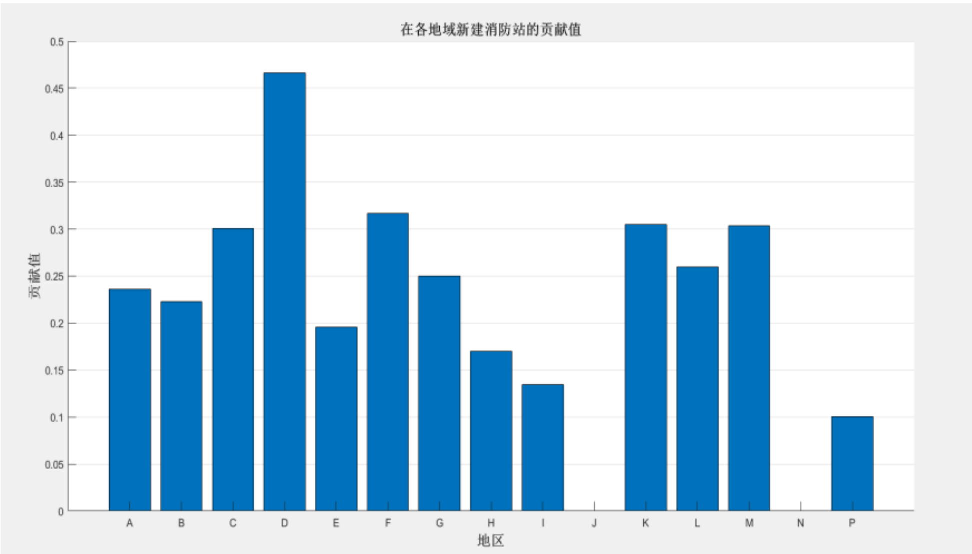


图 17 2021 年需要建立消防站的重要程度区域

结论：可以看出在 D 区域新建矩阵的贡献程度最高，所以我们选定 D 区域为第一个消防站的选址地，同理，添加 D 区域对其它区域的影响因子，得到新的所有区域（已建消防站区域除外）新建消防站的总贡献值矩阵，建立如下直方图：

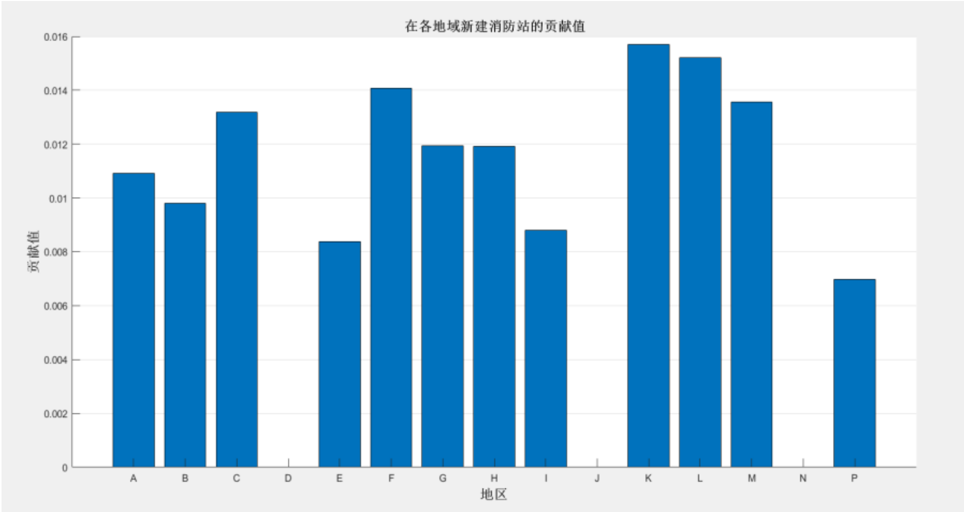


图 18　2024 年需要建立消防站的重要程度区域

结论：可以看出在 K 区域新建矩阵的贡献值最高，所以我们选定 K 区域为第二个消防站的选址地，同理，添加 K 区域对其它区域的影响因子，得到新的所有区域（已建消防站区域除外）新建消防站的总贡献值矩阵，建立如下直方图：

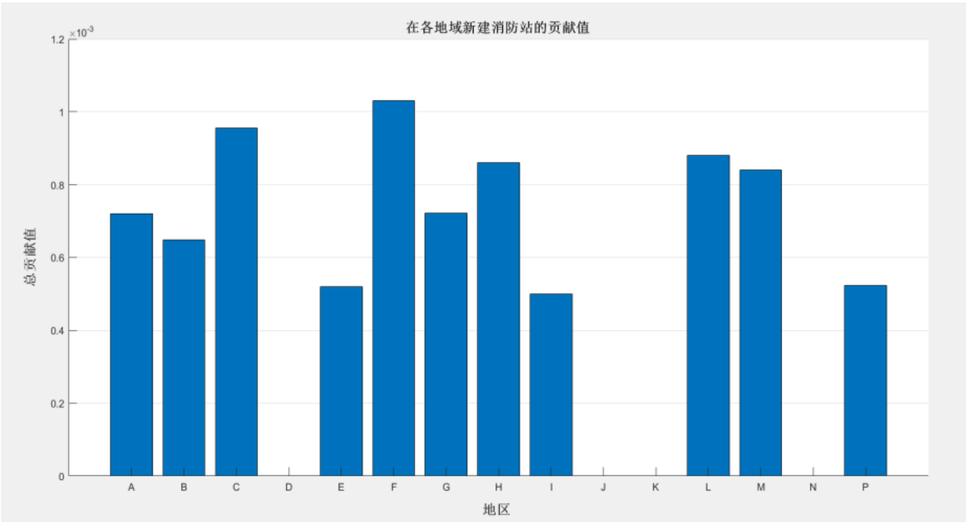


图 19　2027 年需要建立消防站的重要程度区域

结论：可以看出在 F 区域新建矩阵的贡献值最高，所以我们选定 F 区域为第三个消防站的选址地。

六、模型的评价与推广

一、模型的优点

1.第 2 问的预测模型基于时间序列，时间序列预测是根据过去的变化趋势预测未来的发展，理论基础是客观事物发展的连续规律性，因此预测结果不会发生突然跳跃式变化，而是渐进变化的，具有相对较强的稳定性。

2.第 5 问分别考虑了人口密度在小范围内变化和在大范围内变化对事件密度的影响。

3.第 6 问综合考虑了已有消防站对各区域的救援程度、人口密度、事件发生次数、最短路径等影响因子，分别计算出各种影响因子的权值，综合所有影响因子的权值后计算出了贡献值，模型结果具有较高的准确度和可信度。

二、模型的缺点

由于未给出区域的形状及事件在区域内发生的具体位置，所以最短路径的计算结果与现实中的路径长度有一定差别。

三、模型的推广

问题 1 建立的规划模型可推广于解决各机构的人员分配问题。

由于天气呈周期性变化的，且周期也是 12 个月，所以问题 2 建立的时间序列模型可用于预测未来时间段的天气状况。

问题 6 建立的模型也可以决策警察局、医院的建立区域。

参考文献

[1]茆诗松，程依明，濮晓龙，概率论与数理统计教程，北京：高等教育出版社，2011.2.

[2]常星花. 时间序列的建模、预报和应用研究[D].烟台大学,2013.

[3]刘娟,陈涛涛,迟道才.基于 Daniel 及 Mann-kendall 检验的辽西北地区降雨量趋势分析[J].沈阳农业大学学报,2014,45(05):599-603.

[4]司守奎，孙兆亮.数学建模算法与应用[M].2 版.北京:国防工业出版社，2015.

[5]张志涌，杨祖撄等，MATLAB 教程：R2012a，北京：北京航空航天大学出版社，2010.8

[6]刘芳,李思凡,张超平.基于 Floyd 算法的景区游历路线规划问题[J].农村经济与科技,2021,32(03):83-84+126.

附录：

时间序列模型

```
%% 第 2 问源码
clear;
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "Sheet1";
opts.DataRange = "B2:C3684";
opts.VariableNames = ["VarName2", "VarName3"];
opts.VariableTypes = ["string", "double"];
opts = setvaropts(opts, "VarName2", "WhitespaceRule", "preserve");
opts = setvaropts(opts, "VarName2", "EmptyFieldRule", "auto");
test = readtable("附件 2：某地消防救援出警数据.xlsx", opts, "UseExcel",
false);
clear opts
data=zeros(3682,2);
a=zeros(5,12);
n=height(test);
for i=1:n
    s=split(test{i,1},"-");
    data(i,1)=str2double(s(1));
    data(i,2)=str2double(s(2));
    a(data(i,1)-2015,data(i,2))=a(data(i,1)-2015,data(i,2))+1;
end
clear i n s test data;
a(1,5)=114;
x=a'; x=x(:);  %按照时间的先后次序，把数据变成列向量
s=12;  %周期 s=12
n=12;  %预报数据的个数
m1=length(x);   %原始数据的个数
for i=s+1:m1
    y(i-s)=x(i)-x(i-s); %进行周期差分变换
end
w=diff(y);   %消除趋势性的差分运算
m2=length(w); %计算最终差分后数据的个数
k=0; %初始化试探模型的个数
for i=0:3
    for j=0:3
        if i==0 & j==0
            continue
        elseif i==0
            ToEstMd=arima('MALags',1:j,'Constant',0); %指定模型的结构
        elseif j==0
            ToEstMd=arima('ARLags',1:i,'Constant',0); %指定模型的结构
```

```
        else
            ToEstMd=arima('ARLags',1:i,'MALags',1:j,'Constant',0);  %指定
模型的结构
        end
        k=k+1; R(k)=i; M(k)=j;
        [EstMd,EstParamCov,logL,info]=estimate(ToEstMd,w'); %模型拟合
        numParams = sum(any(EstParamCov)); %计算拟合参数的个数
        %compute Akaike and Bayesian Information Criteria
        [aic(k),bic(k)]=aicbic(logL,numParams,m2);
    end
end
fprintf('R,M,AIC,BIC 的对应值如下\n %f');  %显示计算结果
check=[R',M',aic',bic']
r=input('输入阶数 R=');m=input('输入阶数 M=');
ToEstMd=arima('ARLags',1:r,'MALags',1:m,'Constant',0); %指定模型的结构
[EstMd,EstParamCov,logL,info]=estimate(ToEstMd,w'); %模型拟合
w_Forecast=forecast(EstMd,n,'Y0',w') %计算 12 步预报值,注意已知数据是列向量
yhat=y(end)+cumsum(w_Forecast)    %求一阶差分的还原值
for j=1:n
    x(m1+j)=yhat(j)+x(m1+j-s); %求 x 的预测值
end
xhat=x(m1+1:end)   %截取 n 个预报值
aa=[a(1:5,:);xhat'];
subplot(1,2,1);
plot(1:72,[aa(1,:),aa(2,:),aa(3,:),aa(4,:),aa(5,:),aa(6,:)],'LineWidth',2
);
grid on
xlabel('月份', 'fontsize',12)
ylabel('接警量','fontsize',12)
set(gca,'LineWidth',2);
subplot(1,2,2);
plot(1:60,[a(1,:),a(2,:),a(3,:),a(4,:),a(5,:),],'LineWidth',2);
grid on
xlabel('月份', 'fontsize',12)
ylabel('接警量','fontsize',12)
set(gca,'LineWidth',2);
```

在不同空间上各类事件的事件密度

```
%% 在不同空间上各类事件的事件密度
clear;
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "Sheet1";
opts.DataRange = "D2:E3684";
opts.VariableNames = ["area", "cla"];
opts.VariableTypes = ["string", "categorical"];
opts = setvaropts(opts, "area", "WhitespaceRule", "preserve");
opts = setvaropts(opts, ["area", "cla"], "EmptyFieldRule", "auto");
tab = readtable("附件 2：某地消防救援出警数据.xlsx", opts, "UseExcel",
false);
clear opts;
a=zeros(15,7);
h=height(tab);
disp(h)
for i=1:h
    r=char(tab{i,1})-64;
    if r==16
        r=r-1;
    end
    c=char(tab{i,2})-9311;
    a(r,c)=a(r,c)+1;
end
clear tab r c i h;
p=zeros(15,7);
km=[90,120,88,75,111,85,113,93,125,74,132,128,119,89,10];
for i=1:15
    for j=1:7
      p(i,j)=a(i,j)*100/km(i);
    end
end
clear i j;
surf(1:7,[1:15]',a);
xlabel('事件类别');
ylabel('地区');
zlabel('事件密度');
```

各类事件密度与人口密度的散点图

```
clear;
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "Sheet1";
opts.DataRange = "D2:E3684";
opts.VariableNames = ["area", "cla"];
opts.VariableTypes = ["string", "categorical"];
opts = setvaropts(opts, "area", "WhitespaceRule", "preserve");
opts = setvaropts(opts, ["area", "cla"], "EmptyFieldRule", "auto");
tab = readtable("附件 2：某地消防救援出警数据.xlsx", opts, "UseExcel",
false);
clear opts;
a=zeros(15,7);
h=height(tab);
disp(h)
for i=1:h
    r=char(tab{i,1})-64;
    if r==16
        r=r-1;
    end
    c=char(tab{i,2})-9311;
    a(r,c)=a(r,c)+1;
end
clear tab r c i h;
p=zeros(15,7);
km=[90,120,88,75,111,85,113,93,125,74,132,128,119,89,10];

% p 代表各类事件密度，未除总周数，扩大了 100 倍
for i=1:15
    for j=1:7
       p(i,j)=a(i,j)*100/km(i);
    end
end
clear i j;

pe=[6.62,8.76,5.56,6.07,9.24,6.58,7.73,6.09,7.76,5.40,8.79,5.86,9.80,
7.50,15.90];
format long;
% pp 代表人口密度
pp=pe./km;
plot(pp(1:14),p(1:14,1)','ko');
```

```
xlabel('人口密度', 'fontsize',12)
ylabel('事件密度','fontsize',12)
grid on
set(gca,'LineWidth',2);
```

各类事件密度与人口密度关系的拟合曲线

```
%% 绘制三维各类事件密度与人口密度关系的拟合曲线
clear;
opts = spreadsheetImportOptions("NumVariables", 2);
opts.Sheet = "Sheet1";
opts.DataRange = "D2:E3684";
opts.VariableNames = ["area", "cla"];
opts.VariableTypes = ["string", "categorical"];
opts = setvaropts(opts, "area", "WhitespaceRule", "preserve");
opts = setvaropts(opts, ["area", "cla"], "EmptyFieldRule", "auto");
tab = readtable("附件 2：某地消防救援出警数据.xlsx", opts, "UseExcel",
false);
clear opts;
a=zeros(15,7);
h=height(tab);
disp(h)
for i=1:h
    r=char(tab{i,1})-64;
    if r==16
        r=r-1;
    end
    c=char(tab{i,2})-9311;
    a(r,c)=a(r,c)+1;
end
clear tab r c i h;
format long;
p=zeros(15,7);
km=[90,120,88,75,111,85,113,93,125,74,132,128,119,89,10];
for i=1:15
    for j=1:7
      p(i,j)=a(i,j)*100/km(i);
    end
end
clear i j;
surf(1:7,[1:14]',a(1:14,:));
pe=[6.62,8.76,5.56,6.07,9.24,6.58,7.73,6.09,7.76,5.40,8.79,5.86,9.80,
7.50,15.90];
pp=pe./km;
t1=pp;
t2=linspace(0,1.7,10000);
x=polyfit(t1,p(:,7)',2);
```

```
surf([],[],[]);
ylabel('事件类别');
xlabel('人口密度');
zlabel('事件密度');
hold on;
y=ones(1,10000);
yy=ones(1,15);
for i=1:7
    x=polyfit(t1,p(:,i)',2);
    plot3(t1,yy.*i,p(:,i)','ko')
    plot3(t2,y.*i,polyval(x,t2),'LineWidth',2);
end
```

各类事件密度与人口密度关系的拟合曲线

```
%%
% 计算各区事件数量的权值pr
clear,clc;
format long;
areasum=[157 117 92 153 245 153 196 14 105 73 155 180 101 216 99 1641];
asum=sum(areasum);
TH=zeros(1,15);
for i=1:15
    TH(i)=areasum(i)/asum;
end
clear ans asum i pr;
% 计算在各区域建立消防站后对各区的影响力，越小代表影响力越强
name='邻接矩阵.xlsx';
% sp代表各区域之间的最短距离
sp=xlsread(name,1);
clear name;
% S代表各区域分别到其它各区域的距离之和
S=sum(sp);
H=zeros(15,15);
for i=1:15
    test=sp(:,i)./S(i);
    H(i,:)=test';

end
sir=zeros(15,15);
for i=1:15
    for j=1:15
        if sp(i,j)~=0
            sir(i,j)=1/sp(i,j);
        end
    end
end
clear i test;
% 计算人口密度pp
pe=[6.62,8.76,5.56,6.07,9.24,6.58,7.73,6.09,7.76,5.40,8.79,5.86,9.80,
7.50,15.90];
km=[90,120,88,75,111,85,113,93,125,74,132,128,119,89,10];
pp=pe./km;
clear pe km;
% 计算第三个消防站贡献值
```

```
C=zeros(1,15);
for i=1:15
C(1,i)=sum(sir(i,:).*H(10,:).*H(14,:).*H(4,:).*H(11,:).*pp.*TH)*10000
;
end
C(1,10)=0;
C(1,14)=0;
C(1,4)=0;
C(1,11)=0;
i=1:15;
bar(i,C);
title('各地区贡献值');
xlabel('地区');
ylabel('贡献值');
clear i;
```

Floyd 算法

```
%% 弗洛依德算法
clc,clear,a=zeros(15);
a(1,[2:15])=[11.1,0,11.4,0,0,0,0,0,0,0,0,8.2,0,0];
a(2,[3:15])=[8.2,12.8,0,0,0,0,0,0,0,0,0,0,0];
a(3,[4:15])=[7.7,11.1,9.4,0,0,0,0,0,0,0,0,0];
a(4,[5:15])=[0,6.9,0,0,0,12.7,0,0,14.3,10,8.5];
a(5,[6:15])=[7.4,0,0,0,0,0,0,0,0,0];
a(6,[7:15])=[0,0,0,0,0,0,0,11.2,0];
a(7,[8:15])=[0,0,12.9,13.4,14.5,0,10.6,0];
a(8,[9:15])=[9,0,0,12.3,0,0,0];
a(9,[10:15])=[0,0,0,0,0,0];
a(10,[11:15])=[9.5,0,9.6,6.9,4.2];
a(11,[12:15])=[4.4,15.0,0,0];
a(12,[13:15])=[0,0,0];
a(13,[14:15])=[0,0];
a(14,[15:15])=[5.9];
n=length(a);b=a+a';
b(b==0)=inf;
b(1:n+1:end)=0;
for k=1:n
    for i=1:n
        for j=1:n
            if b(i,k)+b(k,j)<b(i,j)
                b(i,j)=b(i,k)+b(k,j);
            end
        end
    end
end
```