

摘要

针对问题一，首先对附件中的数据进行处理，正确导出每年 2 月、5 月、8 月、11 月中第一天的出警情况，之后对三个不同时段进行详细分析，画出四个月份三个不同时段柱状对比图，再利用 echarts 绘制出每年这四个月份不同时段接警次数的占比饼图，最后按照比例来安排每年 2 月、5 月、8 月、11 月中第一天的三个时间段各应多少人值班。

针对问题二，以附件中 2016 年 1 月 1 日至 2019 年 12 月 31 日的数据为基础，对附件中的数据进行处理后，利用 Holt-Winters 季节性预测模型建立消防救援出警次数的预测模型，利用模型模拟出 2020 年的数据，再与附件中的 2020 年的数据与模型模拟出的数据进行对比验证，通过验证结果评价模型的准确性和稳定性，并完成对 2021 年各月份的消防救援出警次数的预测，得出预测结果后，将结果填入表 1。

针对问题三，应题目要求对附件中的数据按照 7 种类别进行分组处理和深度处理，首先我们利用并置拟合方法，所得拟合优度 R^2 远小于 1；随后我们更换思路换为全局拟合法，以拟合度最优为评价标准，针对 7 种不同类别分别选用五次项拟合、ECS 拟合和 Gauss 拟合方法，从而建立最优模型确定出每类事件发生次数。

针对问题四，根据题目中所给图 1 中 15 个地点的距离相关图，我们采用了 MDS 多维尺度变换的方法建立数学模型，在数据处理及编程后得到具体的数据表，根据所得的坐标绘制出区域空间位置的坐标图，得出以 A-I 15 个地点为 X 轴，事件①-⑦为 Y 轴，事件①-⑦在各地的发生次数为 Z 轴的 3D 柱状图，最终根据图表数据得到不同区域相关性最强的事件类别。

关键词：python echarts 季节预测模型 多项式拟合 Gauss 拟合 ECS 拟合

1.问题的重述

1.1 背景知识

1.1.1 引言部分

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，消防救援队承担的任务也呈现多样化、复杂化的趋势。对于每一起出警事件，消防救援队都会对其进行详细的记录。

1.1.2 研究意义

在我国经济的高速发展的情况下，城市空间环境复杂性上升趋势迅速加快，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。对于每一起出警事件，消防救援队都会对其进行详细的记录。而对于这些记录，我们可以进行数据分析，并且在不同的情况下进行模拟假设，从而研究得出在不同情况下的最优出警方案，最大化的提高效率。

1.2 具体问题

1. 问题一

利用附件中的数据，对三个不同的时间段进行数据处理，数据处理后再分析每年2月、5月、8月、11月中第一天的三个时段的出警情况，最后得出三个时间段各应安排多少人值班。

2. 问题二

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. 问题三

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. 问题四

根据图 1，建立适当的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。

2.问题的分析

2.1 研究现象综述

近几年，由于我国经济的高速发展，导致城市的空间环境复杂性也急剧上升，不仅各种事故灾害频发，安全风险也在不断增大，消防救援队承担的任务也呈现多样化、复杂化的趋势。对于每一起出警事件，消防救援队都会对其进行详细的记录。

2.2 对问题的具体分析和对策

1.对问题一的分析

首先对附件中的数据进行处理，正确导出每年 2 月、5 月、8 月、11 月中第一天的出警情况数据，之后画出四个月份三个不同时段柱状对比图，再绘制出每年这四个月份不同时段接警次数的占比饼图，最后按照比例来安排每年 2 月、5 月、8 月、11 月中第一天的三个时间段各应多少人值班。

2.对问题二的分析

以附件中 2016 年 1 月 1 日至 2019 年 12 月 31 日的数据为基础，对附件中的数据进行处理后，利用 Holt-Winters 季节性预测模型建立消防救援出警次数的预测模型，利用模型模拟出的数据，再与附件中的数据与进行对比验证，通过验证结果评价模型的准确性和稳定性，并完成对 2021 年各月份的消防救援出警次数的预测，得出预测结果后，将结果填入表 1。

3.对问题三的分析

首先对附件中的数据按照 7 种类别进行分组处理和深度处理,利用并置拟合方法,所得拟合优度 $R^2=0.23$ 远小于 1; 随后我们更换思路换为全局拟合法,以拟合度最优为评价标准,针对 7 种不同类别分别选用五次项拟合、ECS 拟合和 Guess 拟合方法,从而建立最优模型确定出每类事件发生次数。

4.对问题四的分析

根据题目中所给图 1 中 15 个地点的距离相关图,利用了 MDS 多维尺度变换的方法建立数学模型,在数据处理及编程后得到具体的 数据表,根据所得的坐标绘制出区域空间位置的坐标图,得出以 A-I 15 个地点为 X 轴,事件①-⑦为 Y 轴,事件①-⑦在各地的发生次数为 Z 轴的 3D 柱状图,最终根据图表数据得到不同区域相关性最强的事件类别。

3.模型的假设

1.在求解得出最后的 2、5、8、11 四个月份第一天的三个时间段安排多少人员值班占比时,假设小数位的比例向上取整。

2. 在建立模型对 2021 年各月份的消防救援出警次数进行预测时,预测次数不足一次的,按一次出警计算。

3.假设

4.名词解释和符号说明

符号	说明
L_t	季节性预测模型中预测函数的水平函数
b_t	季节性预测模型中预测函数的趋势函数

S_t	季节性预测模型中预测函数的季节分量
α 、 β 和 γ	指数平滑参数
s	季节性预测模型中循环的长度

5.模型的建立与求解

5.1 问题一的分析与求解

1.对问题一的分析

首先对附件中的数据进行处理，正确导出每年 2 月、5 月、8 月、11 月中第一天的出警情况的数据，之后绘制出每年这四个月份不同时段接警次数的占比饼图，最后按照比例来安排每年 2 月、5 月、8 月、11 月中第一天的三个时间段各应多少人值班。

2.对问题一的求解

利用 PyCharm Community Edition、Visual Studio Code 等软件，导入附件中的数据后，结合 python、JavaScript 语言进行程序编写，最终绘制出每年四个月份不同时段接警次数的占比饼图。

在对附件中的数据进行处理后，我们绘制出了每年 2 月、5 月、8 月、11 月中第一天中的接警时间在 I 时间段（即每天 0:00-8:00）的柱状图，如下图 5.1 所示：

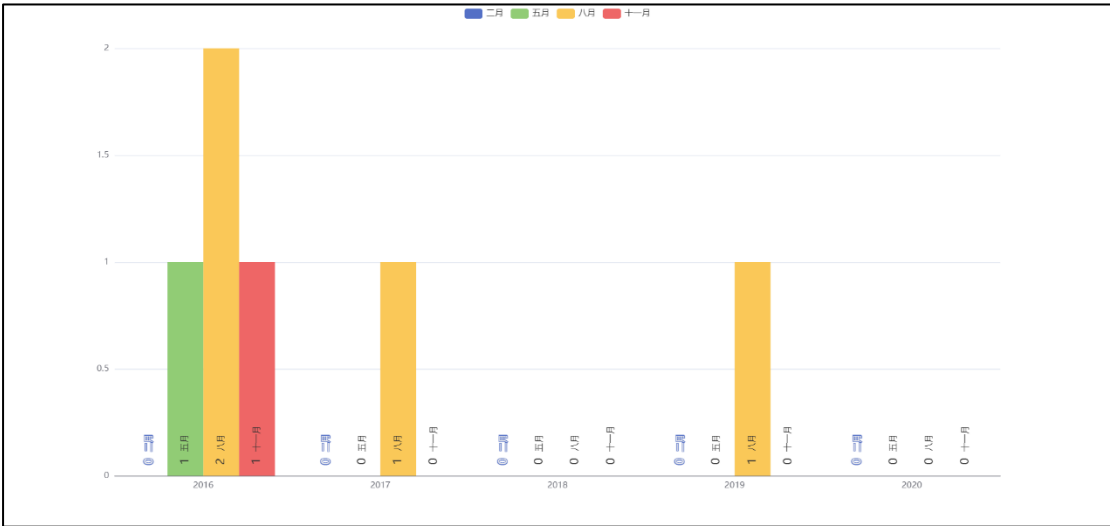


图 5.1 每年 2、5、8、11 月中第一天中的接警时间在 I 时间段柱状图

在图 5.1 中，我们可以直观的看出，在 2016 年—2020 年中，2 月份接警时间在 I 时间段的次数为 0，5 月份和 11 月份仅有一次接警时间在 I 时间段，8 月份最多，共有 4 次接警时间在 I 时间段。对此，我们可以分析出若在 I 时间段安排警力，可着重安排人数在 8 月份，其次是 5 月份和 11 月份，2 月份安排基础警力即可。

之后，我们又绘制出了每年 2 月、5 月、8 月、11 月中第一天中的接警时间在 II 时间段（即每天 8:00-16:00）的柱状图，如下图 5.2 所示：

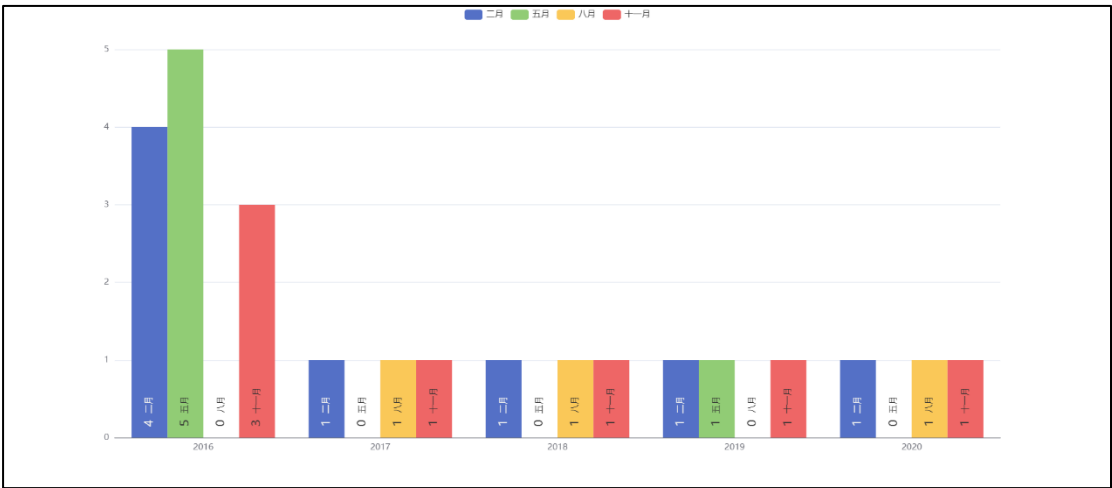


图 5.2 每年 2、5、8、11 月中第一天中的接警时间在 II 时间段柱状图

在图 5.2 中，我们观察后可知，在 2016 年—2020 年中，2 月份接警时间在 II 时间段的次数为 8 次，但其中 4 次在 2016 年；5 月份有 6 次接警时间在 II 时

间段，并且其中 5 次在 2016 年，8 月份共有 3 次接警时间在Ⅱ时间段，11 月份则有最多的 7 次接警记录。对此，我们可以分析出若在Ⅱ时间段安排警力，可着重安排人数在 2 月份和 5 月份，其次是 8 月份和 11 月份。

最后，我们又利用 echarts 绘制出了每年 2 月、5 月、8 月、11 月中第一天中的接警时间在Ⅲ时间段（即每天 16:00-24:00）的柱状图，如下图 5.3 所示：

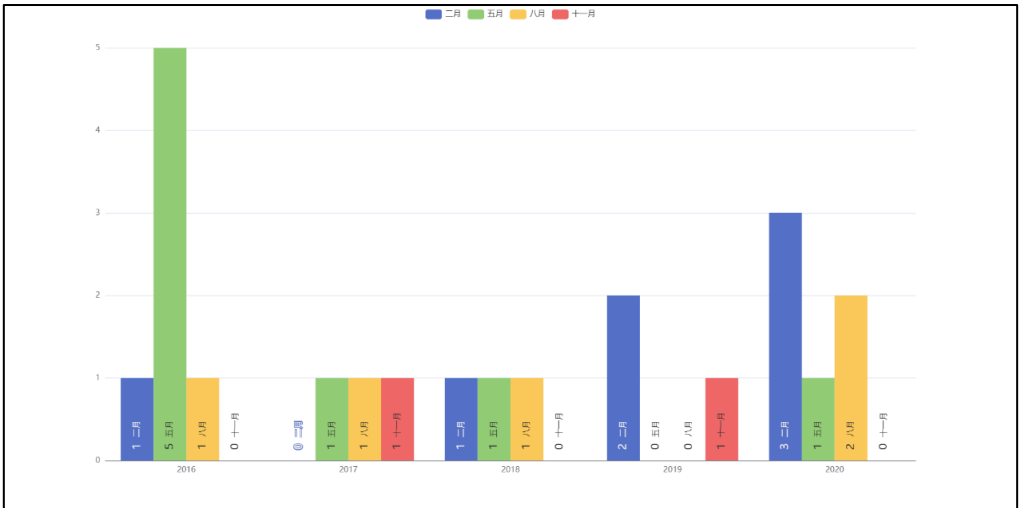


图 5.3 每年 2、5、8、11 月中第一天中的接警时间在Ⅲ时间段柱状图

在图 5.3 中，我们分析柱状图可知，在 2016 年—2020 年中，2 月份接警时间在Ⅲ时间段的次数为 7 次；5 月份有 8 次接警时间在Ⅲ时间段，并且其中 5 次在 2016 年；8 月份共有 5 次接警时间在Ⅲ时间段；11 月份则有最少的 2 次接警记录。对此，我们可以分析出若在Ⅲ时间段安排警力，可着重安排人数在 2 月份和 5 月份，其次是 8 月份，在 11 月份安排基础警力即可。

随后，我们又按照利用 python 语言编程处理后的 2、5、8、11 四个月份不同时间接警次数的数据，利用 echarts 绘制出了不同月份接警次数的占比饼图，利用饼图所占比例我们可以更加直观清晰的看出接警数据，从而可以最大化最优化的安排警力。

我们以不同颜色标记三个不同时段，得出在 2016 年—2020 年中每年二月份不同时段接警次数占比饼图如下图 5.4 所示：

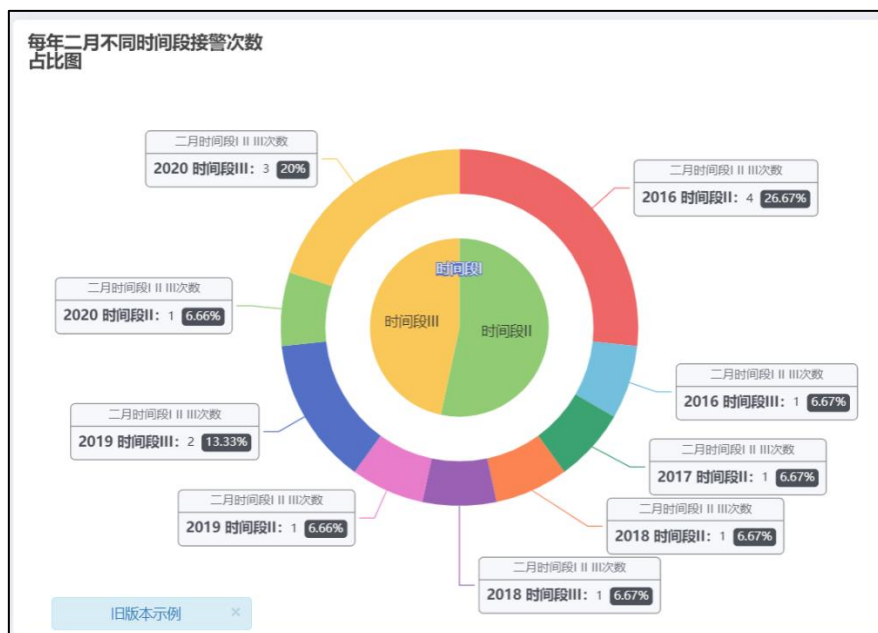


图 5.4 每年二月份不同时段接警次数占比饼图

在 2016 年—2020 年中每年五月份不同时段接警次数占比饼图如下图 5.5 所示：

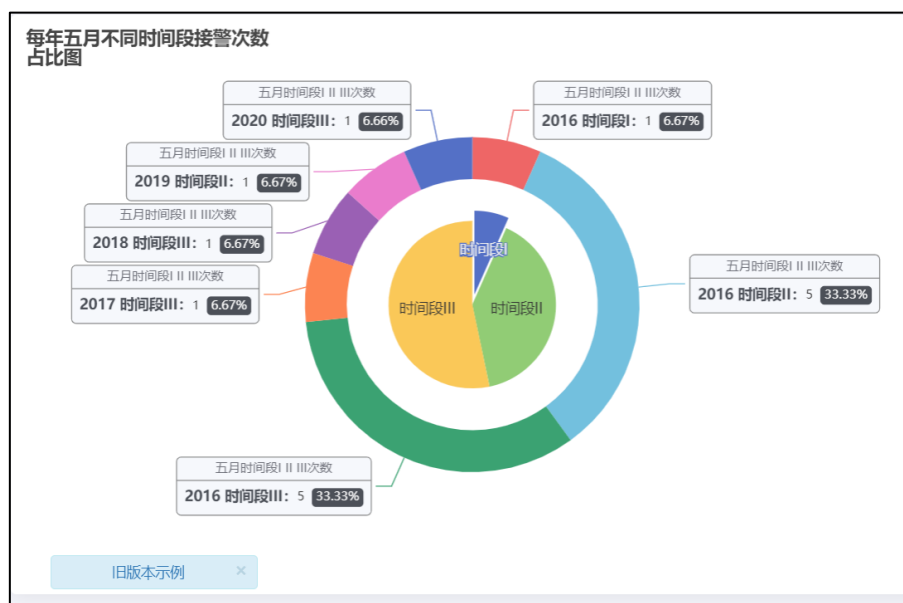


图 5.5 每年五月份不同时段接警次数占比饼图

在 2016 年—2020 年中每年八月份不同时段接警次数占比饼图如下图 5.6 所示：

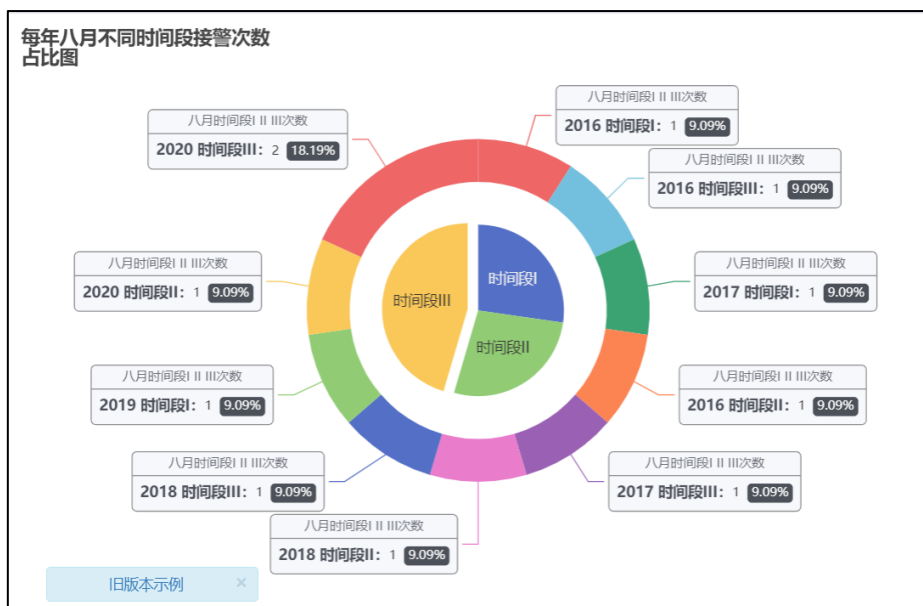


图 5.6 每年八月份不同时段接警次数占比饼图

在 2016 年—2020 年中每年十一月份不同时段接警次数占比饼图如下图 5.7 所示：

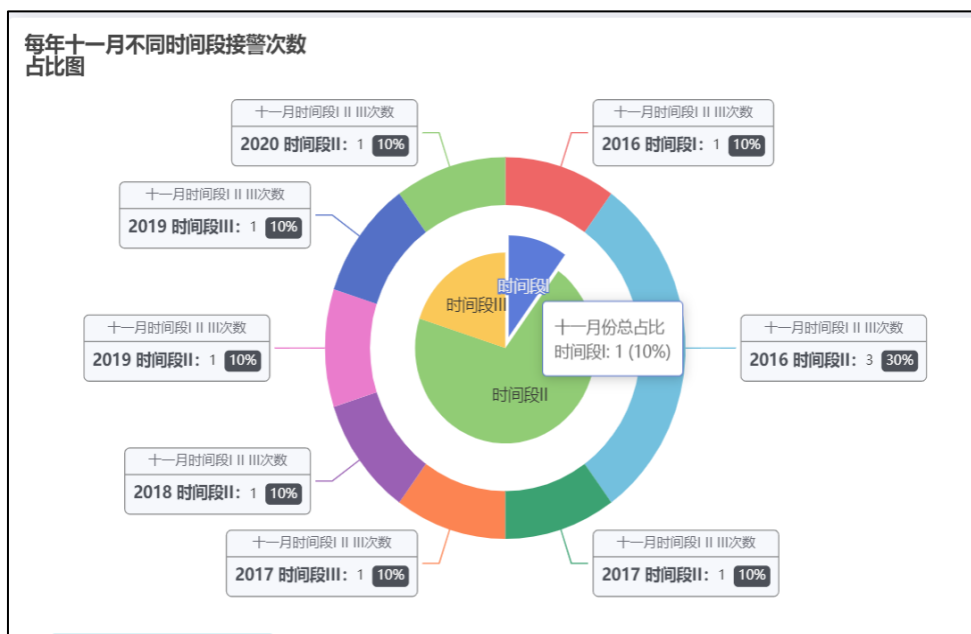


图 5.7 每年十一月份不同时段接警次数占比饼图

由上面所得的饼图可得：

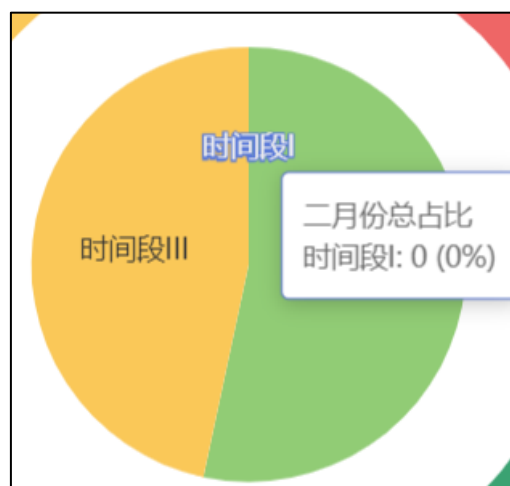


图 5.4-1 时间段 I 二月总占比为 0%

对于每年二月份：

2016-2020 年，对于时间段 I，因为每个时间段至少要 5 人值守，而 2016-2020 年二月的第一天总可用人数为 150，而为确保每个时间段都至少有 5 人，2016-2020 至少要 $5 * 5 = 25$ 人，也就是说总占比不足 16% ($30/150 = 0.16$) 的要按照 16% 处理。因此每年 2 月份时间段 I 虽然占比为 0%，但应当按照占比 16% 算，因此每年二月第一天时间段 I 分配 5 人。

2016-2020 年，对于时间段 II 与时间段 III，还剩下总 125 人，

时间段 II 是由图 5.4 外环占比图 2016 年的时间段 II (26.67%)，2017、2018、2019、2020 时间段 II (6.67%) 共同组成的总占比 (53.33%)，

根据占比 $125 \times 0.5333 \div 5 = 13.25$ ，每年二月份时间段 II 应当分配 13 人。

2016-2020 年，同理对于时间段 III $125 * 0.4667 \div 5 = 11.625$ ，也即每年二月份时间段应当分配 12 人。

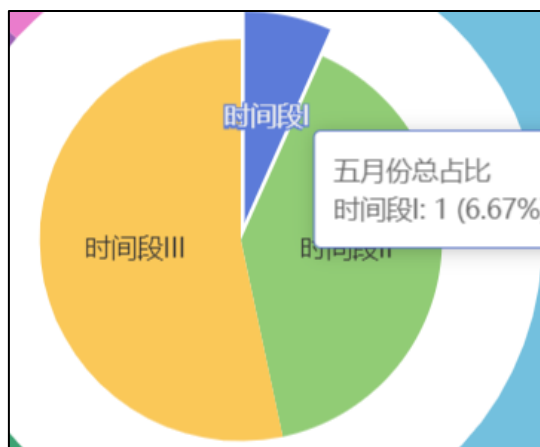


图 5.5-1 时间段 I 总占比为 6.67%

对于每年五月份:

同二月份一样, 虽然时间段 I 占比 6.67%, 但面对每个时间段至少有 5 人的原则, 我们按照占比 16%来处理。

每年五月份时间段 I 应该分配 5 人, 时间段 II 分配 $125 * 0.44 \div 5 = 11$ 人, 时间段 III 应分配 $125 * 0.56 \div 5 = 14$ 人。

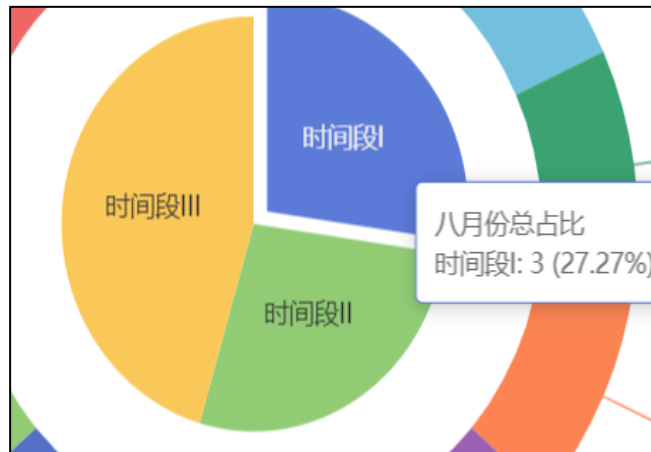


图 5.6-1 每年八月份时间段 I 总占比为 27.27%

对于每年八月份:

时间段 I 应当分配 $150 * 27.27\% \div 5 = 8$ 人,

时间段 II 应当分配 $150 * 0.2727 \div 5 = 8$ 人,

时间段 III 应当分配 $150 * 0.46 \div 5 = 14$ 人。

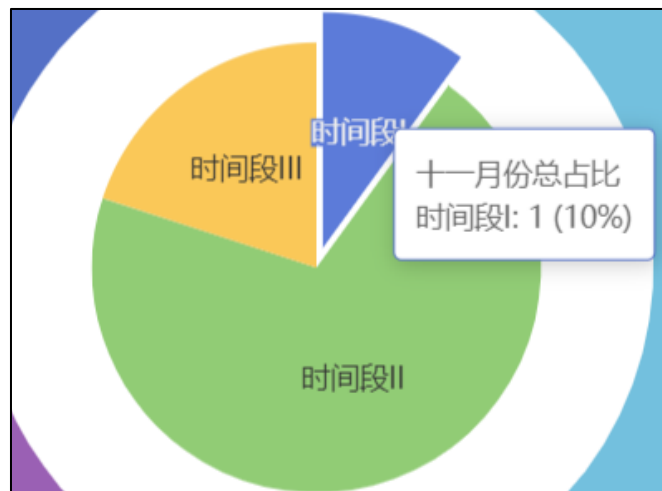


图 5.7-1 每年十一月份时间段 I 总占比为 10%

对于每年十一月份：

对于时间段 I，虽然占比 10%，但不足 16%，我们按照 16% 处理，即每年十一月份第一天时间段 I 应当分配 5 人。

时间段 II 应当分配 $125 * 0.75 \div 5 = 19$ 人

时间段 III 应当分配 $125 * 0.25 \div 5 = 6$ 人

综上所述，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的人数值班如下表所示：

	I 时间段	II 时间段	III 时间段
2 月	5	13	12
5 月	5	11	14
8 月	8	8	14
11 月	5	19	6

5.2 问题二的分析与求解

1. 对问题二的分析

首先我们以附件中 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，在对附件中的数据进行处理后，利用 Holt-Winters 季节性预测模型建立消防救援出警次数的预测模型，在模型建立之后，利用模型模拟出 2020 年 1 月 1 日至 2020 年 12 月 31 日的数据，再与附件中的 2020 年 1 月 1 日至 2020 年 12 月 31 日的数据与模型模拟出的数据进行对比验证，通过验证结果评价模型的准确性和稳定性，并完成对 2021 年各月份的消防救援出警次数的预测，得出预测结果后，将结果填入表 1。

2. 对问题二的求解

由于问题二需要我们利用地 2016 年 1 月 1 日至 2019 年 12 月 31 日的基础数据，以月份为单位，建立数学模型，并与 2020 年 1 月 1 日至 2020 年 12 月 31 日的数据进行对比验证，从而预测出 2021 年各月份的消防救援出警次数。

在对附件中数据进行处理后，并且结合问题一我们可以观察发现，在某些时段接警次数较多，但是在某些时段接警次数又相对较少，并且每年的接警情况又相对类似，我们称这种重复现象为“季节性”（Seasonality），如果数据集在一定时间段内的固定区间内呈现相似的模式，那么该数据集就具有季节性，因此，我们选择利用 Holt-Winters 季节性预测来建立数学模型。

Holt-Winters 季节性预测模型，是一种三次指数平滑预测，其背后的理念就是除了水平和趋势外，还将指数平滑应用到季节分量上。Holt-Winters 季节性预测模型由预测函数和三次平滑函数组成，一个是水平函数 L_t ，一个是趋势函数 b_t ，一个是季节分量 S_t ，以及平滑参数 α 、 β 和 γ 。对应关系为：

$$L_t = \alpha(y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$$

$$S_t = \gamma(y_t - L_t) + (1 - \gamma)S_{t-s}$$

$$F_{t+k} = L_t + kb_t + S_{t+k-s}$$

其中 s 为季节循环的长度， $0 \leq \alpha \leq 1$ ， $0 \leq \beta \leq 1$ ， $0 \leq \gamma \leq 1$ 。水平函数为季节性调整的观测值和时间点 t 处非季节预测之间的加权平均值。趋势函数和霍尔特线性方法中的含义相同。季节函数为当前季节指数和去年同一季节的季节性指数之间的加权平均值。在本算法，我们同样可以用相加和相乘的方法。当季节性变化大致相同时，优先选择相加方法，而当季节变化的幅度与各时间段的水平成正比时，优先选择相乘的方法。

我们可以看到趋势和季节性的预测准确度都很高。我们选择了 seasonal_period=3 作为每月份重复的数据，当然，也可以调整其它其它参数。最终利用代码实现了预测，如图 5.8。

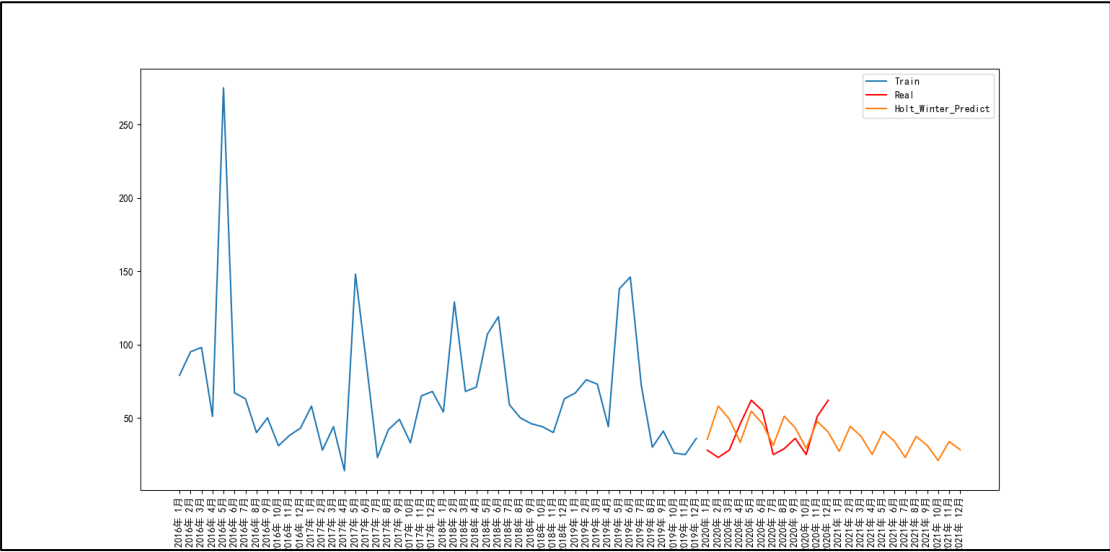


图 5.8 Holt-Winters 季节性预测模型

Train 是 2016 年—2019 年每年 1-12 月份的报警数据变化, Real 是 2020 年实际的各个月份的报警数据, Holt-Winter-Predict 是我们利用 Holt-Winters 季节性预测模型拟合出的

为了验证模型可靠性, 我们将附件中 2020 年的数据真实值与 2020 Holt-Winters 季节性预测模型的预测值进行对比验证, 得出真实值与预测值之间的方差约为 16.1065。

2020-2021 预测值:

[35.41189496 58.09152508 49.2148561 33.344634 54.63298158 46.22546947
31.27737304 51.17443807 43.23608283 29.21011207 47.71589456 40.24669619
27.14285111 44.25735106 37.25730955 25.07559015 40.79880755 34.26792291
23.00832919 37.34026405 31.27853628 20.94106823 33.88172054 28.28914964]

表 1 问题 2 的结果

月份	预测值（次）
2021 年 1 月	28
2021 年 2 月	45
2021 年 3 月	38
2021 年 4 月	26
2021 年 5 月	41
2021 年 6 月	35
2021 年 7 月	24

2021 年 8 月	38
2021 年 9 月	32
2021 年 10 月	21
2021 年 11 月	34
2021 年 12 月	29

5.3 问题三的分析与求解

1.对问题三的分析

首先我们针对附件所给的 7 种类别事件的发生时间，利用 python 对附件中的数据进行分组处理和深度分类处理，我们利用并置拟合方法所得的拟合优度并不理想，在此之后，我们更换思路将拟合方法换为全局拟合法，并且针对 7 种不同类别分别选用五次项拟合、ECS 拟合和 Gauss 拟合等不同的方法，以拟合度最优为评价标准，建立最优模型，以确定出每类事件发生次数。

2.对问题三的求解

在对问题进行分析讨论后，我们对附件中的数据按照年限、月份、7 种类别和发生次数进行预处理分组，由于数据处理结果较多，我们选取部分处理结果进行展示，完整的数据处理将放在附录中，处理结果（部分）如下表 5.2 所示：

表 5.2 预处理数据结果（部分）

year	month	type	count
2016	1	①	29
2016	1	②	1
2016	1	③	14
2016	1	④	11
2016	1	⑥	10
2016	1	⑦	14
2016	2	①	20
2016	2	②	5
2016	2	③	18
2016	2	④	9
2016	2	⑤	1
2016	2	⑥	7
2016	2	⑦	35
2016	3	①	7
2016	3	②	1

随后，我们把 7 种类型进行分组处理，即把某种类型在 1-12 个月内 2016-2020 年的发生次数进行统计处理，方便后续进行数据拟合，由于数据处理结果较多，

我们选取部分处理结果进行展示，完整的数据处理将放在附录中，分组合并处理结果（部分）如下表 5.3 所示：

表 5.3 分组合并处理结果（部分）

Type	Month	2016/Count	2017/Count	2018/Count	2019/Count	2020/Count
1						
	1	29	6	0	11	7
	2	20	5	2	13	6
	3	7	16	2	6	6
	4	7	7	2	5	8
	5	10	15	0	8	11
	6	9	9	3	9	13
	7	5	3	2	10	5
	8	5	1	1	3	7
	9	2	1	2	7	9
	10	3	1	1	4	5
	11	13	6	1	5	7
	12	4	5	12	9	18

我们利用 OriginLab 对数据进行分析处理后，在通过并置拟合方法对数据进行拟合后，所得拟合结果如下图 5.9 所示：

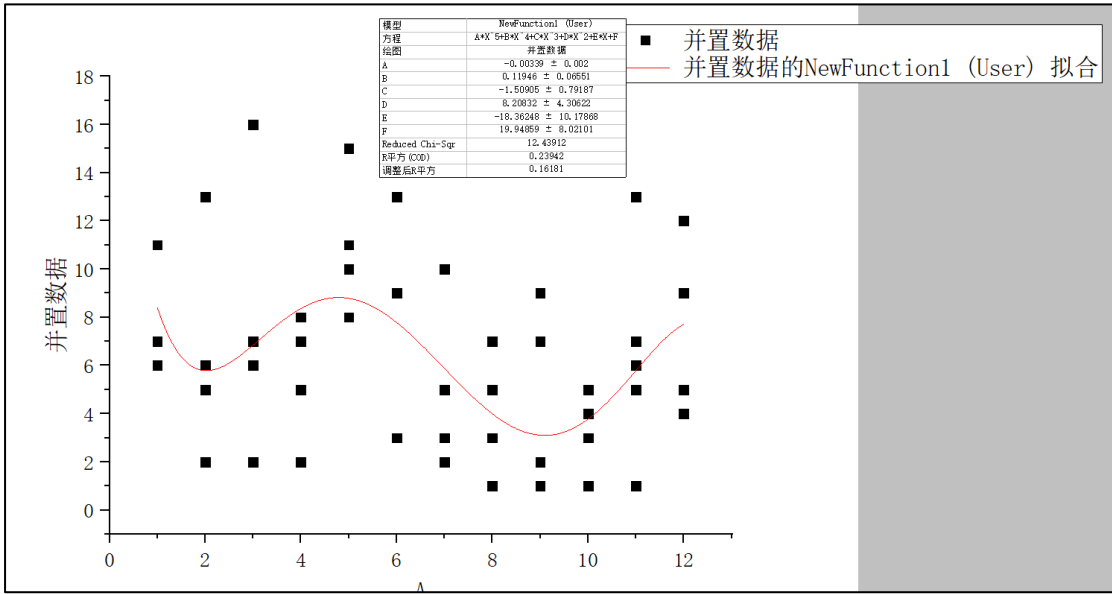


图 5.9 并置拟合结果

由图 3 可以看出，并置拟合所得曲线散点偏离曲线较多，并且在我们进一步拟合计算得出，最为直观的能体现出回归直线对观测值的拟合程度的度量拟合优度 $R^2 = 0.23$ ，远远小于 1，说明回归直线对观测值的拟合程度较差，精确度较低。

随后我们更换拟合方法为全局拟合法，对 2016 年—2020 年这 5 年中每年、每种类型的数据都进行拟合，以拟合度最优为评价标准，选择一条拟合优度最接近 1 的，即最为精准的拟合结果作为最优数学模型。

我们对类型①进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.4 所示：

表 5.4 类型①每年的五次项拟合曲线

A	B1	B2	B3	B4	B5
1 模型	[Book1]FitNL1!Notes.Model 事件1				
2 方程	$y = A0 + A1*x + A2*x^2 + A3*x^3 + A4*x^4 + A5*x^5$				
3 绘图	B	C	D	E	F
4 A0	68.27272 ± 13.00672	13.90909 ± 13.00672	-10.68182 ± 13.00672	12.13636 ± 13.00672	12.77273 ± 13.00672
5 A1	-53.81121 ± 17.2405	-16.6136 ± 17.2405	16.86106 ± 17.2405	1.79274 ± 17.2405	-8.76806 ± 17.2405
6 A2	19.08345 ± 7.43867	10.47171 ± 7.43867	-7.74946 ± 7.43867	-2.61571 ± 7.43867	3.42442 ± 7.43867
7 A3	-3.21916 ± 1.38433	-2.31983 ± 1.38433	1.55839 ± 1.38433	0.72258 ± 1.38433	-0.44197 ± 1.38433
8 A4	0.25288 ± 0.11545	0.20728 ± 0.11545	-0.14106 ± 0.11545	-0.07603 ± 0.11545	0.01609 ± 0.11545
9 A5	-0.00742 ± 0.00354	-0.00646 ± 0.00354	0.00469 ± 0.00354	0.00273 ± 0.00354	1.88537E-4 ± 0.00354
10 Reduced Chi-Sqr*	8.86154				
11 R平方(COD)	0.86837	0.71212	0.92902	0.53949	0.74044
12 R平方(COD)*	0.84408				
13 调整后R平方*	0.69335				

由表 5.4 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型①的拟合曲线，最佳拟合曲线为下图 5.10 所示：

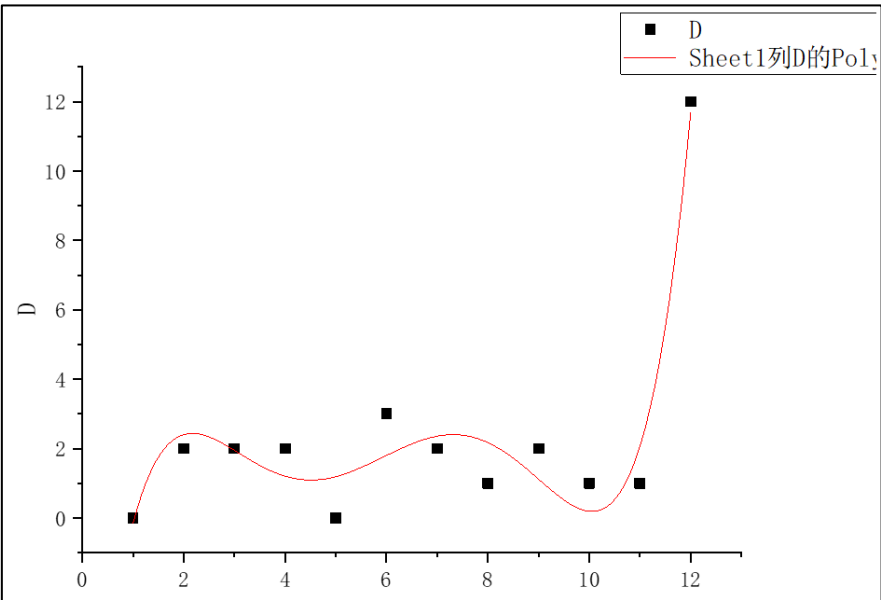


图 5.10 类型①事件发生次数的最优模型曲线

根据图 5.10 所示，以拟合度最优为评价标准，①类型事件发生次数的最优模型为：

$$Y_1 = 4.7 * 10^{-3} * x^5 - 1.4 * 10^{-1} * x^4 + 1.6 * x^3 - 7.75 * x^2 + 16.86 * x - 10.7$$

我们对类型②进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.5 所示：

表 5.5 类型②每年的五次项拟合曲线

A	B1	B2	B3	B4	B5
1 模型	事件二				
2 方程	$y = A0 + A1*x + A2*x^2 + A3*x^3 + A4*x^4 + A5*x^5$				
3 绘图	B	C	D	E	F
4 A0	-1.09091 ± 6.6209	1.72727 ± 6.6209	-3 ± 6.6209	12.77273 ± 6.6209	7.18182 ± 6.6209
5 A1	4.74859 ± 8.77604	-1.00666 ± 8.77604	5.76039 ± 8.77604	-12.87125 ± 8.77604	-10.51148 ± 8.77604
6 A2	-2.22576 ± 3.78655	1.2607 ± 3.78655	-2.24719 ± 3.78655	5.04383 ± 3.78655	5.18331 ± 3.78655
7 A3	0.44379 ± 0.70467	-0.33814 ± 0.70467	0.37612 ± 0.70467	-0.8634 ± 0.70467	-1.01162 ± 0.70467
8 A4	-0.03868 ± 0.05877	0.03326 ± 0.05877	-0.02811 ± 0.05877	0.06585 ± 0.05877	0.08474 ± 0.05877
9 A5	0.00122 ± 0.0018	-0.0011 ± 0.0018	7.73001E-4 ± 0.0018	-0.00183 ± 0.0018	-0.00255 ± 0.0018
10 Reduced Chi-Sqr*	2.29618				
11 R平方(COD)	0.05508	0.31984	0.15107	0.74961	0.44198
12 R平方(COD)*	0.40065				
13 调整后R平方*	-0.17873				

由表 5.5 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型②的拟合曲线，最佳拟合曲线为下图 5.11 所示：

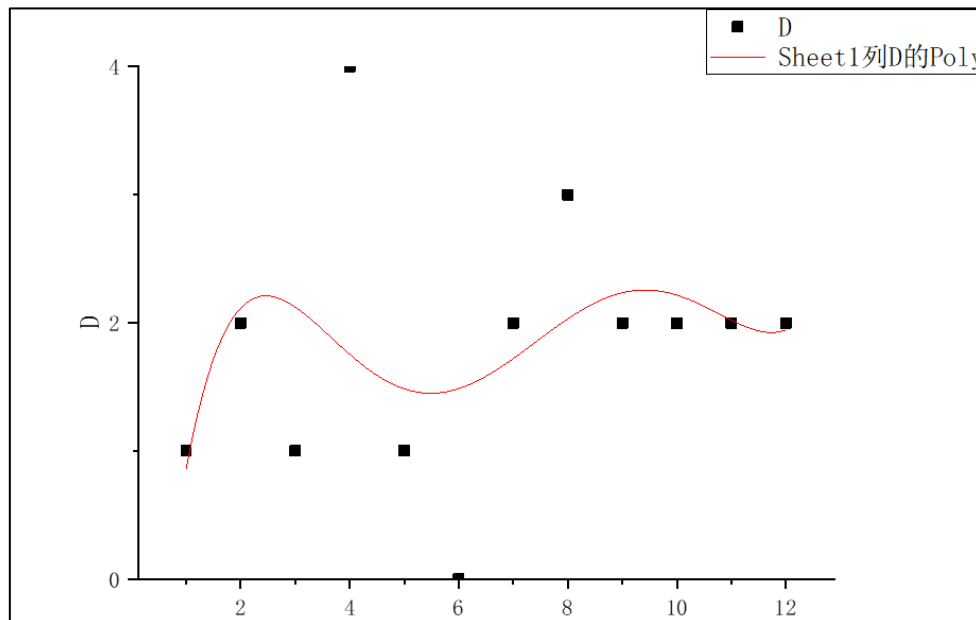


图 5.11 类型②事件发生次数的最优模型曲线

根据图 5.11 所示，以拟合度最优为评价标准，②类型事件发生次数的最优模型为：

$$Y_2 = -1.8 * 10^{-3} * x^5 + 6.6 * 10^{-2} * x^4 - 8.6 * 10^{-1} * x^3 + 5.0 * x^2 - 12.9 * x + 12.8$$

我们对类型③行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.6 所示：

表 5.6 类型③每年的五次项拟合曲线

1	A	B1	B2	B3	B4	B5
2	模型	事件三				
3	方程	$y = A0 + A1*x + A2*x^2 + A3*x^3 + A4*x^4 + A5*x^5$				
4	绘图	B	C	D	E	F
5	A0	13.90909 ± 74.10919	157.63633 ± 74.10919	0.81818 ± 74.10919	47.49999 ± 74.10919	14.18182 ± 74.10919
6	A1	6.79806 ± 98.23222	-241.53257 ± 98.23222	6.97958 ± 98.23222	-34.55179 ± 98.23222	-19.38344 ± 98.23222
7	A2	-6.16266 ± 42.38374	116.73487 ± 42.38374	2.87091 ± 42.38374	17.55992 ± 42.38374	9.80694 ± 42.38374
8	A3	1.33938 ± 7.88759	-22.31174 ± 7.88759	-1.13745 ± 7.88759	-3.62017 ± 7.88759	-2.04165 ± 7.88759
9	A4	-0.11592 ± 0.65782	1.83159 ± 0.65782	0.11125 ± 0.65782	0.30788 ± 0.65782	0.18184 ± 0.65782
10	A5	0.00354 ± 0.02017	-0.05419 ± 0.02017	-0.00333 ± 0.02017	-0.00918 ± 0.02017	-0.00569 ± 0.02017
11	Reduced Chi-Sqr*	287.68518				
12	R平方(COD)	0.77315	0.4615	0.37307	0.65767	0.80073
13	R平方(COD)*	0.53941				
14	调整后R平方*	0.09418				

由表 5.6 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型③的拟合曲线，最佳拟合曲线为下图 5.12 所示：

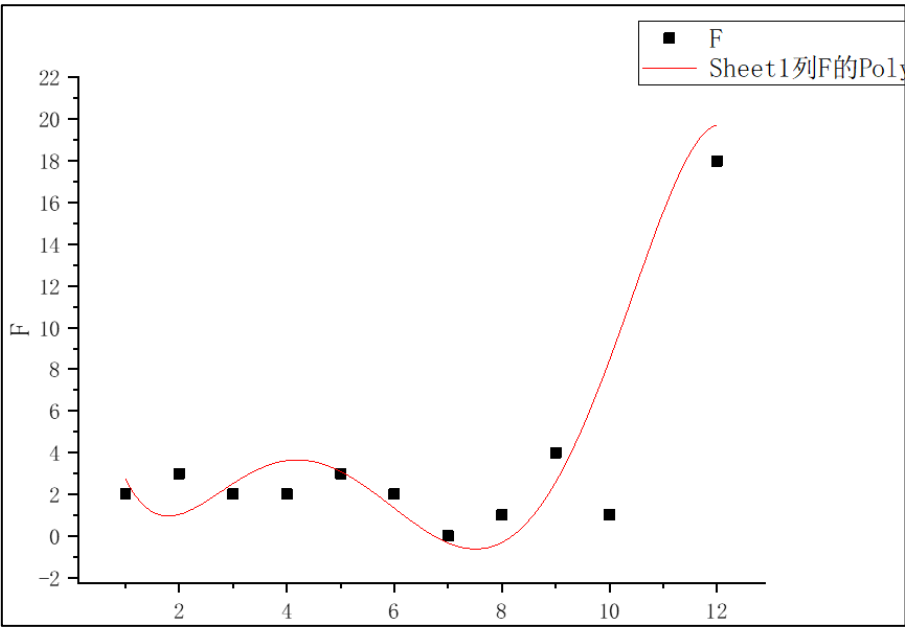


图 5.12 类型③事件发生次数的最优模型曲线

根据图 5.12 所示，以拟合度最优为评价标准，③类型事件发生次数的最优模型为：

$$Y_3 = -5.7 * 10^{-3} * x^5 + 1.8 * 10^{-1} * x^4 - 2.04 * x^3 + 9.8 * x^2 - 19.4 * x + 14.2$$

我们对类型④进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.7 所示：

表 5.7 类型④每年的五次项拟合曲线

	A	B1	B2	B3	B4	B5
1 模型	事件四					
2 方程	$y = A0 + A1*x + A2*x^2 + A3*x^3 + A4*x^4 + A5*x^5$					
3 绘图						
4 A0		B	C	D	E	F
5 A1		21.31818 ± 12.22481	23.63636 ± 12.22481	-4.77273 ± 12.22481	12.81818 ± 12.22481	10.40909 ± 12.22481
6 A2		-12.30731 ± 16.20407	-19.69617 ± 16.20407	10.60665 ± 16.20407	-15.31397 ± 16.20407	-7.51828 ± 16.20407
7 A3		3.06633 ± 6.99148	6.31127 ± 6.99148	-3.66097 ± 6.99148	7.18728 ± 6.99148	3.56206 ± 6.99148
8 A4		-0.3459 ± 1.30111	-0.9633 ± 1.30111	0.59978 ± 1.30111	-1.33778 ± 1.30111	-0.65399 ± 1.30111
9 A5		0.0179 ± 0.10851	0.07257 ± 0.10851	-0.05046 ± 0.10851	0.10879 ± 0.10851	0.05096 ± 0.10851
10 Reduced Chi-Sqr*		-3.2994E-4 ± 0.00333	-0.00214 ± 0.00333	0.00171 ± 0.00333	-0.00322 ± 0.00333	-0.00142 ± 0.00333
11 R平方(COD)		7.82812				
12 R平方(COD)*		0.74912	0.67132	0.87599	0.28293	0.09154
13 调整后R平方*		0.54611				
		0.10735				

由表 5.7 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型 1 的拟合曲线，最佳拟合曲线为下图 5.13 所示：

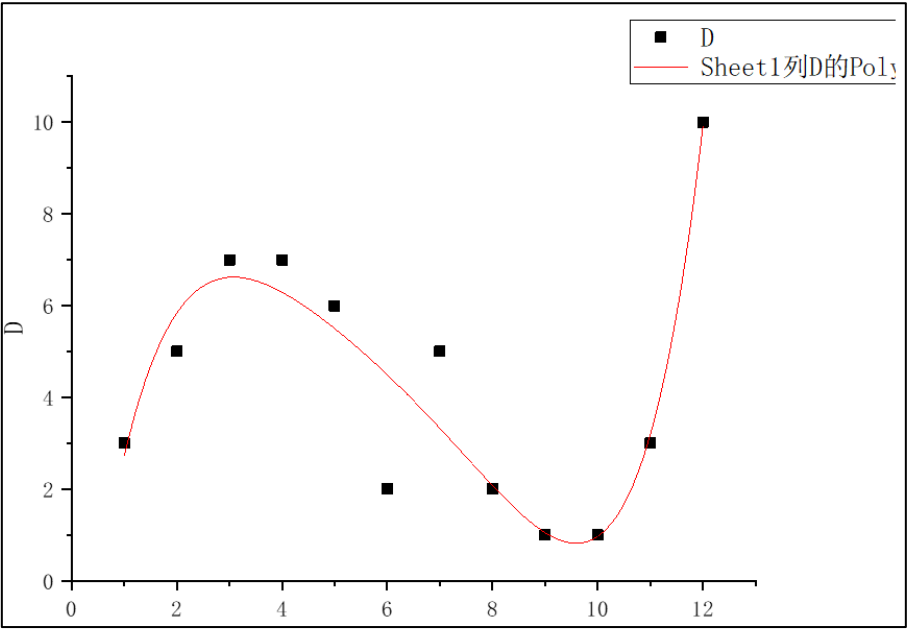


图 5.13 类型④事件发生次数的最优模型曲线

根据图 5.13 所示，以拟合度最优为评价标准，④类型事件发生次数的最优模型为：

$$Y_4 = -3.3 * 10^{-4} * x^5 + 1.8 * 10^{-2} * x^4 - 3.5 * 10^{-1} * x^3 + 3.1 * x^2 - 12.3 * x + 21.3$$

我们对类型⑤进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.8 所示：

表 5.8 类型⑤每年的五次项拟合曲线

	A	B1	B2	B3	B4	B5
1 模型	ECS 事件五					
2 方程	y=y0+A/(w*sqrt(2*pi))*(exp(-0.5*((x-xc)/w)^2)*(1+(a3/(3*2*1))*((x-xc)/w)*(((x-xc)/w)^2-3)+(a4/(4*3*2*1))*(((x-xc)/w)^4-6*((x-xc)/w)^3+3)+((10*a3^2)/(6*5*4*3*2*1))*(((x-xc)/w)^6-15*((x-xc)/w)^4+45*((x-xc)/w)^2-15)))					
3 绘图	B	C	D	E	F	
4 y0	0.91667 ± 0.33912	0.11695 ± 0.52824	0.5 ± 0.39159	0.66667 ± 0.33912	0.01544 ± 0.42547	
5 xc	7.5207 ± 0	3.12518 ± 1.05488	7.6781 ± 1.56465E6	3 ± 0	6.11585 ± 3.05477	
6 A	0.96185 ± 0	5.92763 ± 5.31353	7.93545 ± --	0 ± 0	2.88114 ± 3.45813	
7 w	0.01148 ± 0	1.88135 ± 0.762	0.20418 ± --	3.54167 ± 0	0.93968 ± 1.30747	
8 a3	93.35972 ± 0	-1.83751 ± 1.04922	1.39099 ± 3.90652E7	1 ± 0	-2.08819 ± 1.3371	
9 a4	49.72057 ± 0	-2.98976 ± 3.66223	-0.19371 ± 2.88627E7	1 ± 0	-0.09176 ± 4.63403	
10 Reduced Chi-Sqr*	1.38006					
11 R平方(COD)	-4.44089E-16	0.81432	0.82292	1.11022E-16	0.99557	
12 R平方(COD)*	0.56571					
13 调整后R平方*	0.1459					

由表 5.8 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型⑤的拟合曲线，最佳拟合曲线为下图 5.14 所示：

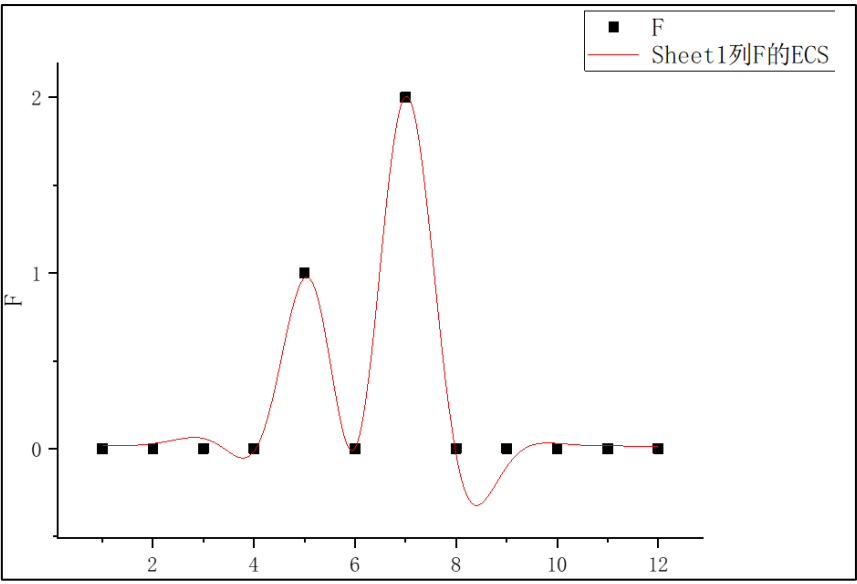


图 5.14 类型⑤事件发生次数的最优模型曲线

根据图 5.14 所示，以拟合度最优为评价标准，⑤类型事件发生次数的最优模型为：

$$y_5 = y_0 + A / (w \cdot \sqrt{2 \cdot \pi}) \cdot (\exp(-0.5 \cdot ((x - x_c) / w)^2) \cdot (1 + (a_3 / (3 \cdot 2 \cdot 1)) \cdot ((x - x_c) / w) \cdot ((x - x_c) / w)^2 - 3) + (a_4 / (4 \cdot 3 \cdot 2 \cdot 1)) \cdot (((x - x_c) / w)^4 - 6 \cdot ((x - x_c) / w)^3 + 3) + ((10 \cdot a_3^2) / (6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)) \cdot (((x - x_c) / w)^6 - 15 \cdot ((x - x_c) / w)^4 + 45 \cdot ((x - x_c) / w)^2 - 15)))$$

其中 $y_0 = 0.01544$

$x_c = 6.11585$

$A = 2.88114$

$w = 0.93968$

$a_3 = -2.08819$

$a_4 = -0.09176$

我们对类型⑥进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.9 所示：

表 5.9 类型⑥每年的五次项拟合曲线

	A	B1	B2	B3	B4	B5
1 模型	事件6					
2 方程	$y = A0 + A1 \cdot x + A2 \cdot x^2 + A3 \cdot x^3 + A4 \cdot x^4 + A5 \cdot x^5$					
3 绘图	B	C	D	E	F	
4 A0	10.40909 ± 12.20501	17.36364 ± 12.20501	-2.95454 ± 12.20501	-9.36363 ± 12.20501	8.68182 ± 12.20501	
5 A1	2.36892 ± 16.17782	-14.85909 ± 16.17782	8.347 ± 16.17782	17.98548 ± 16.17782	-10.06107 ± 16.17782	
6 A2	-3.72508 ± 6.98016	4.62837 ± 6.98016	-4.36533 ± 6.98016	-8.00697 ± 6.98016	4.05297 ± 6.98016	
7 A3	1.00247 ± 1.299	-0.64439 ± 1.299	1.00024 ± 1.299	1.67976 ± 1.299	-0.59944 ± 1.299	
8 A4	-0.09947 ± 0.10834	0.04193 ± 0.10834	-0.09868 ± 0.10834	-0.15489 ± 0.10834	0.03743 ± 0.10834	
9 A5	0.00335 ± 0.00332	-0.00104 ± 0.00332	0.00344 ± 0.00332	0.00505 ± 0.00332	-8.38989E-4 ± 0.00332	
10 Reduced Chi-Sqr*	7.80279					
11 R平方(COD)	0.55851	0.37425	0.5519	0.65867	0.85567	
12 R平方(COD)*	0.69718					
13 调整后R平方*	0.40444					

由表 5.9 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型 1 的拟合曲线，最佳拟合曲线为下图 5.15 所示：

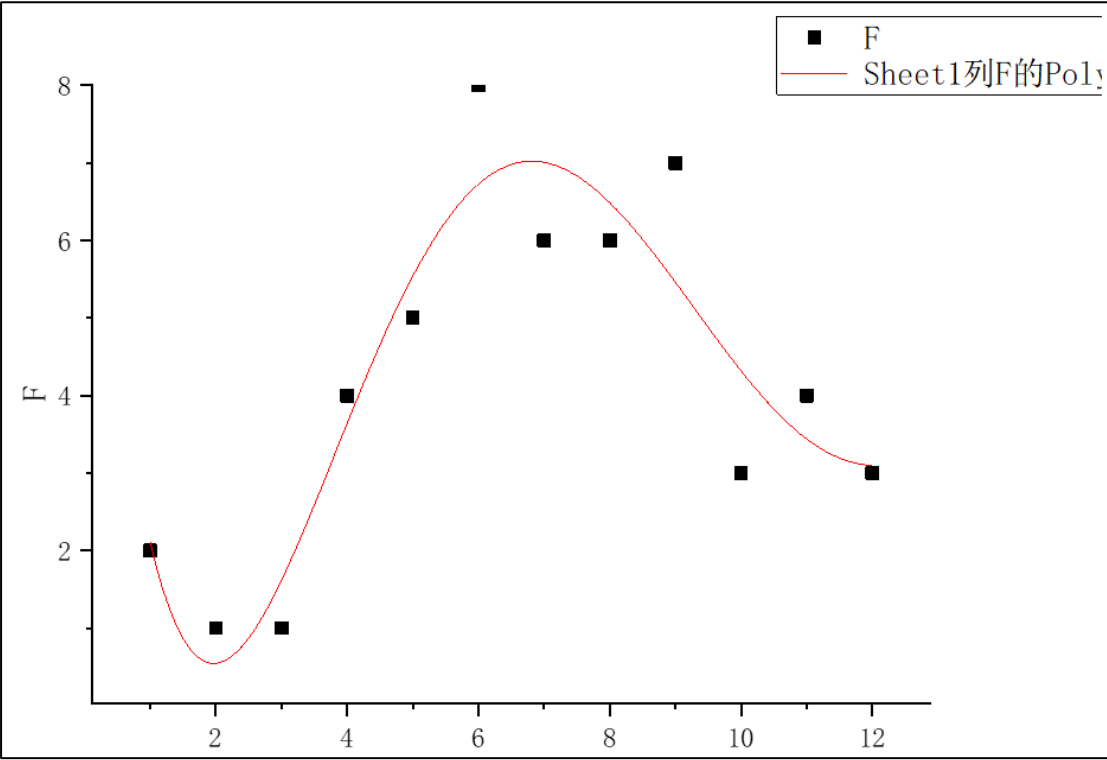


图 5.15 类型⑥事件发生次数的最优模型曲线

根据图 5.15 所示，以拟合度最优为评价标准，⑥类型事件发生次数的最优模型为：

$$Y_4 = -8.4 * 10^{-4} * x^5 + 3.7 * 10^{-2} * x^4 - 6.0 * 10^{-1} * x^3 + 4.1 * x^2 - 10 * x + 8.7$$

我们对类型⑦进行分析处理，利用 OriginLab 和五次项拟合处理，得出 2016 年—2020 年的拟合曲线方程，如下表 5.10 所示：

表 5.10 类型⑦每年的五次项拟合曲线

A	B1	B2	B3	B4	B5
1 模型	Gauss 事件7				
2 方程	$y=y_0 + (A/(w*\sqrt{\pi/2}))*\exp(-2*((x-xc)/w)^2)$				
3 绘图	B	C	D	E	F
4 y0	31.55368 ± 4.43262	33.68482 ± 5.95119	40.20003 ± 4.43238	15.57237 ± 4.69624	10.68868 ± 4.59833
5 xc	5.06476 ± 0.16787	3.5519 ± 0.55535	5.66877 ± 1143.16497	5.79733 ± 0.15458	4.95657 ± 0.54318
6 w	0.80201 ± 0.15807	3.03828 ± 1.34127	0.50833 ± 1721.65805	1.51303 ± 0.31598	1.42381 ± 0.92215
7 A	220.42041 ± 39.42082	-120.67661 ± 59.60904	77.15011 ± 634713.90995	154.13041 ± 32.47265	41.73221 ± 30.41187
8 Reduced Chi-Sqr*	176.79862				
9 R平方(COD)	0.92098	0.74975	0.49377	0.95145	0.89097
0 R平方(COD)*	0.90194				
1 调整后R平方*	0.85536				

由表 5.10 可以看出，根据方程所求出每年曲线方程的拟合优度 R^2 ，进行对比后选择 R^2 最为接近 1 的曲线，作为类型 1 的拟合曲线，最佳拟合曲线为下图 5.16 所示：

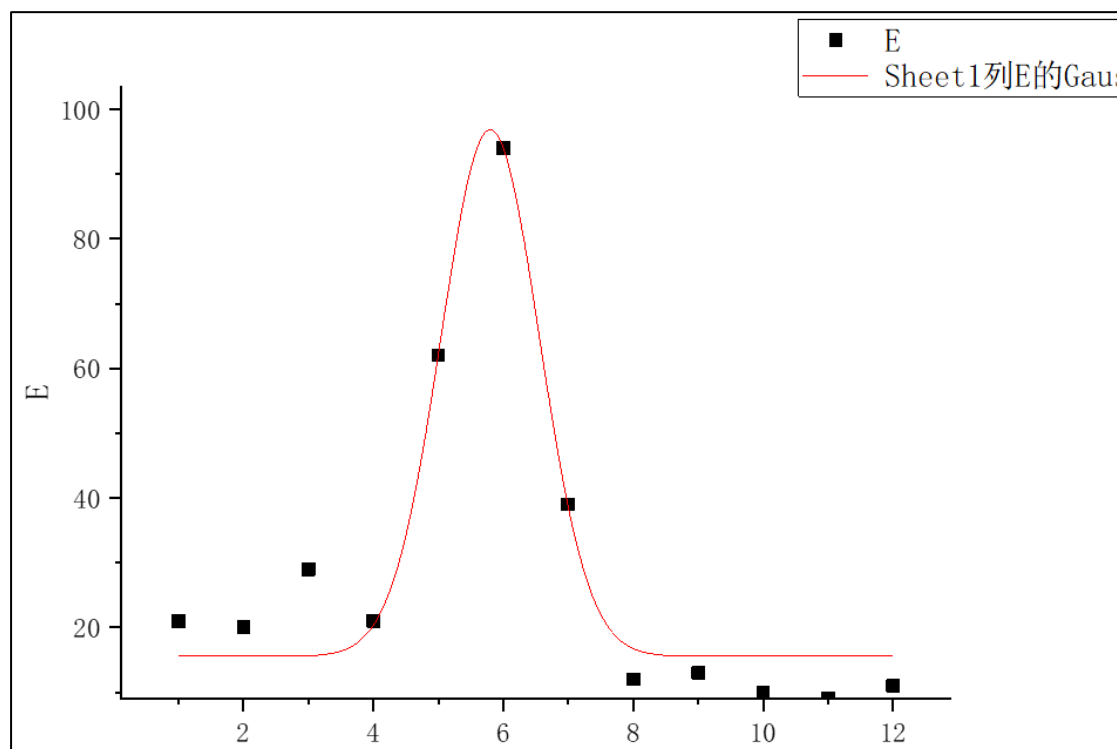


图 5.16 类型⑦事件发生次数的最优模型曲线

根据图 5.16 所示，以拟合度最优为评价标准，⑦类型事件发生次数的最优模型为：

$$y_7 = y_0 + \left(\frac{A}{w \cdot \sqrt{\pi/2}} \right) \cdot \exp(-2 \cdot ((x - x_c)/w)^2)$$

$$\text{其中 } w_0 = 15.57237 \quad x_c = 5.79733 \quad w = 1.51303 \quad A = 154.13041$$

在拟合过程中，针对 7 种不同类别分别选用五次项拟合、ECS 拟合和 Guess 拟合方法，其中类型①、②、③、④、⑥使用的是五次项拟合方法，类型⑤使用的是 ECS 拟合，类型⑦使用的是 Guess 拟合方法，针对不同的类型选取最合适的拟合方法，从而建立最优模型，确定出每类事件发生次数。

5.4 问题四的分析与求解

1.对问题四的分析

根据题目中所给图 1 中的距离相关图，采用 MDS 多维尺度变换的方法来建立数学模型，在数据处理及编程后得到具体的 数据表，根据所得的坐标数据绘制出区域空间位置的坐标图，得出以 A-I 15 个地点为 X 轴，事件①-⑦为 Y 轴，事件①-⑦在各地的发生次数为 Z 轴的 3D 柱状图，最终可以根据图表数据得到不同区域相关性最强的事件类别。

2.对问题四的求解

在对问题进行分析讨论及数据处理后，我们最终选择 MDS 多维尺度变换的最优方法来建立数学模型。多维尺度变换(multidimensional scaling, MDS)是在低维空间去展示高维多元数据的一种可视化方法。该方法的基本目标是将原始数据“拟合”到一个低维坐标系中，使得由降维所引起的任何变形最小。多维尺度变换涉及到的问题可以描述为：当 n 个项目中各对项目之间的相似性(或距离)给定时，求这些项目在低维空间的表示，并使得项目间的接近程度与原先的相似性(或距离)“大体匹配”，较为适合此题中解决问题四。

按照题目要求，我们对附件中数据处理后得到 2016 年—2020 年中每周在某个地方发生多少次某种事件的表格，由于数据较多，详细数据将放在附录中，数据的部分截图如下图所示：

A	B	C	D	E
Year	Weekend	Type	Area	Count
2016	1	①	F	1
2016	1	①	G	1
2016	1	①	P	2
2016	1	③	P	1
2016	1	④	G	1
2016	1	④	P	1
2016	1	⑥	P	1
2016	2	①	P	4
2016	2	②	A	1
2016	2	④	M	1
2016	2	④	P	1
2016	2	⑥	P	3
2016	2	⑦	D	1
2016	2	⑦	L	1
2016	3	①	A	1
2016	3	①	B	1
2016	3	①	D	3
2016	3	①	F	1
2016	3	①	J	2
2016	3	①	M	1

图 1 数据处理

由图 1 各区域之间的邻接关系及距离可知各区域间的距离，因此，我们以矩阵来表示距离，方便后续处理，任意两个区域的距离可表示为：

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & & d_{2n} \\ \vdots & & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix}$$

D 对应我们已经处理好的任意两区域的距离矩阵为下图：

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

图 任意两区域之间的距离

可以利用这些距离数据来计算出各个区域在二维空间上的地理分布, 即重构出区域的位置坐标 \bar{x} (原始真实的位置坐标记为 x), 使得第 i 个和第 j 个区域的距离 $\|\bar{x}_i - \bar{x}_j\| \approx d_{ij}$, 这样多维尺度变换问题可以看成是一个优化问题, 优化的目标函数为:

$$\min \left(\sum_{i < j} (\|\bar{x}_i - \bar{x}_j\| - d_{ij})^2 \right)^{1/2}$$

在通过目标函数进行求解后, 我们可以得到重构出的位置坐标, 按照所得坐标绘制二维空间地理分布图, 如下图所示:

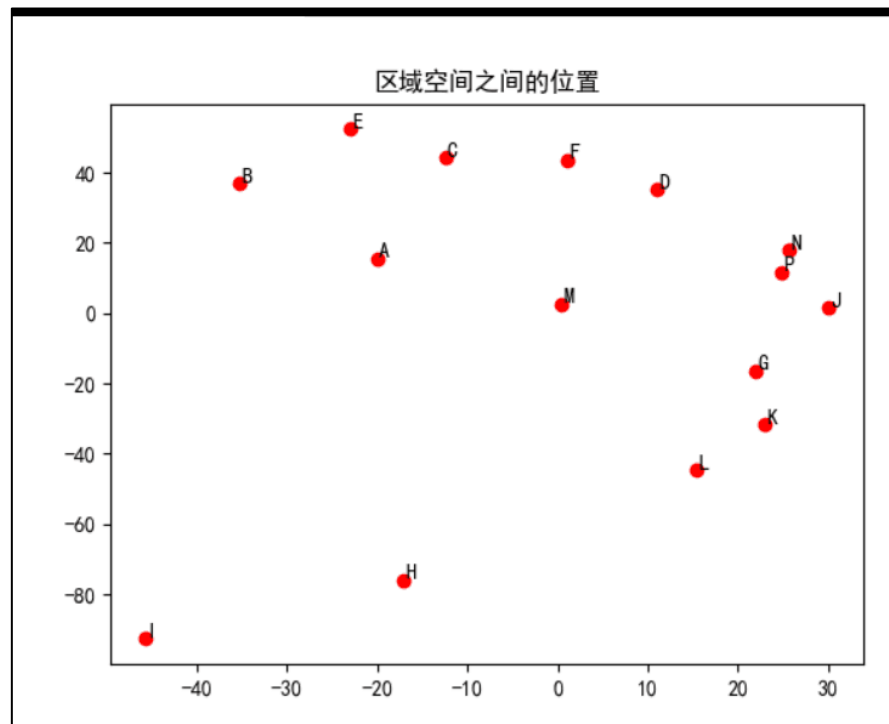


图 2 各区域二维空间的位置

我们再接着对事件发生的密度进行分析, 得到了如图。: 最强事件类别⑦

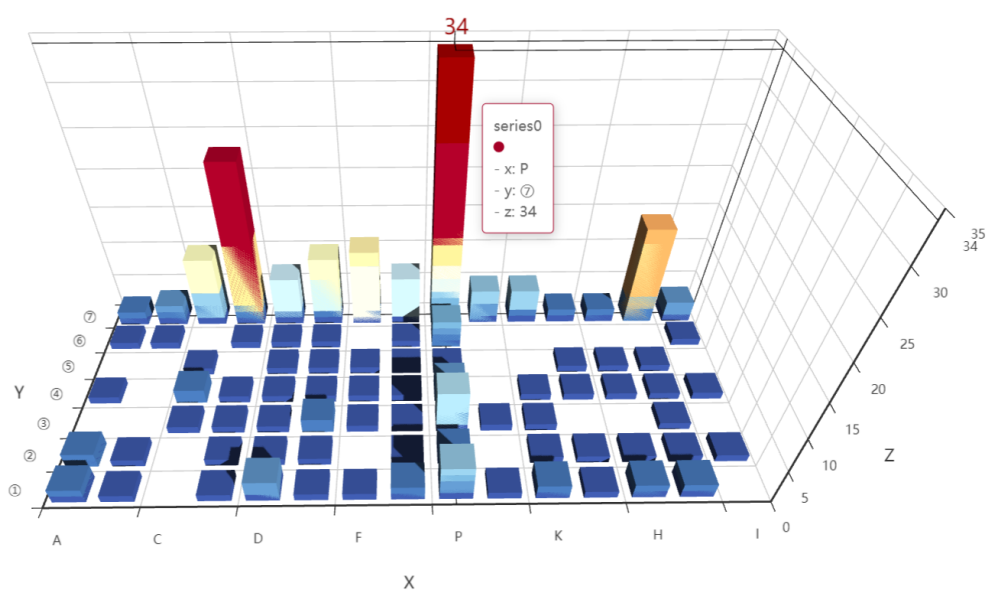


图 3 每周 不同区域各类事件发生的次数

6.模型改进及推广

本文在数据处理的过程中，由于时间及能力有限，在某些环节会有数据差异或者误差等等，可能会造成一定程度的结果失真，但从实际结果来看，也可较为完美的反映所得到的一般规律，可侧面反映出模型的实用价值。

7.参考文献

- [1]姜启源、谢金星、叶俊， 《数学模型》，北京:高等教育出版社，2005 年： .
- [2]司守奎、孙兆亮，《数学建模算法与应用》(第二版)北京:国防工业出版社，2016 年：
- [3]盛骤、谢式千、潘承毅， 《概率论与数理统计》，北京:高等教育出版社，2015 年。