For office use only	Team Control Number	For office use only
T1 _____	**123456789**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**C**	F4 _____

2020
MCM/ICM
Summary Sheet

The Paper Title

Abstract

Keywords: keyword1; keyword2

The Paper Title

January 15, 2020

Abstract

Keywords: keyword1; keyword2

Contents

1 Introduction

1.1 Problem Background

As we know, the Opioid is a double-edged sword.It can help patients reduce their pain and also makes people addicted to the Opioid.The reason is that Opioids can bring them short pleasure.However,becuase each people has different tolerance to drugs,when the intake of the drug gets more and more,people will have a stronger sense of plesure.This is the root cause of the abuse of opioids.Today, the United States is experiencing a crisis caused by the abuse of opioids.We will always face unprecedented challenges.There are some of the consequences of the abuse of opioids.

- **Neonatal Abstinence Syndrome**
  Neonatal Abstinence Syndrome usually appears 48 to 72 hours after the birth of an infant.The reason is that pregnant women take too much opioids during pregnancy to reduce their pain.

- **Abuse of opioid lead to an increase in crime rates.**
  The opioid is consist of natural, synthesis and semi-synthetic.Synthetic opioid include Fentanyl and Methadone.Otherwise, Semi-synthetic Opioid include Heroin.The former is easier to transport than the latter.Therefore, this is more likely to lead to an increase in crime rate.

- **A huge impact on the demographic structure of the United States.**
  In the US, there is a tendency to abuse opioids for people of all ages.And it will increase the mortality rate of the American people.

Figure 1: Drugs

2 Problem Analysis

2.1 Commonly Abused Drugs Charts

Drug	Description	Type
Morphine	an opioid analgesic drug	Natural
Codeine	an alkaloid present in opium	Natural
Heroin	a potent analgesic drug	Semi-Synthetic
Hydrocodone	a semi-synthetic opioid drug	Semi-Synthetic
Fentanyl	a potent analgesic drug	Synthetic
Methadone	an opioid drug	Synthetic
Pethidine	an artificial anesthetic drug	Synthetic
Endorphin	an endogenous opioid drug	Endogenous
Dynorphins	an endogenous opioid drug	Endogenous

Table 1: Commonly Abused Opioid Drugs Charts

2.2 Restatement of The Problem

- **Needing to solve the first question.**

  - According the original datasets,We need to build a mathematical model to describe the relationship between every year and the county which using opioids.

  - According the mathematical model,We can get the earliest location reported to be using opioid in five states.

  - If you continue to follow the increase in the year.Abuse of opioids will must bring irreversible harm to American society.It will become the United States governments concerns about this questions.

  - We use the mathematical model to predict what will happen in the future.

- **Needing to solve the second question.**

  - Need to explain the situation in which the opioids are abused up until now.

  - Need to find out what causes the social crisis result in by the abuse of opioid.

- **Needing to solve the third question.**

  - Optimize the model created by the first question and the second question.

2.3 Overview of Our Work

- **The idea of solving the first problem.**
  First of all, We need to use the eight years from 2010 to 2017 as the time variable and the total number of drug reports county as independent variables. And then we perform data preprocessing on the total number of drug reports county.Selecting the counties that rank the top 4 position in five states.At the same time, We will mark these counties on the map.According to these pictures,We can observe the earliest location reported to be using opioid in five states from these eight pictures.

- **The idea of solving the second problem.**

- **The idea of solving the third problem.**

3 Assumptions and Notations

3.1 Assumptions

- **Assuming all the latitude and longitude in the US is valid,** but not in the US are invalid data.
- **Assuming the earliest use of opioids in five states**is gradually spreanding to the other contry.
- **Assuming each states only related with the county it administers.**

3.2 Notations

Here is we need to use the symbols description in this paper.

Symbols	Description
n	each small area of a polygon
k	predicted object
p	predicted standard
$G_i(x_i, y_i)$	the geometric center of gravity of a small triangle
$G_j(x_j, y_j)$	the geometric center of gravity of the polygon
σ_i	size of each small area
σ_j	size of each polygon
Δ_{ij}	the reference sequence
δ_{ij}	the relationship between predicted object and predicted standard
$g(x)$	a sigmod activation function
f_m	common factor
ε_p	special factor
F	common factor vector
ε	pecial factor vector
A	load factor matrix

4 Data Processing

For the datasets provided by the title,We tried to preprocess the datasets to use the Geopy-Geocoding-Toolkit in the Python.The Geopy-Geocoding-Toolkit converts geocoding into geographic coordinates – latitude and longitude.Because Geopy is calling other third-party Map APIs to get geocoded information,then it converts to exact latitude and longitude coordinates.In essence,This is the mathematical principle used to calculate the two-point distance of the Earth's elliptical model.

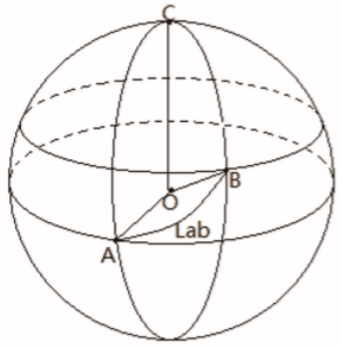


Figure 2: Earth Model

Thus, a simple form of the equation representing the geometric state of the Earth's

space.We use the **WGS-84** ellipsoidal standard.

The Geospatial Geometry for the earth equation is :
Since, a means the radius of the equator , b is the polar axis radius,

$$\frac{x^2+y^2}{a^2}+\frac{z^2}{b^2}=1 \tag{1}$$

for some, a = 6378137.000 m, b = 6356752.314 m.

4.1 Data Cleaning

After data preprocessing,We observed a small amount of abnormal data appeared in the datasets.These anomalous data have an obvious feature, their geocoded information is not in the United States.This also causes the latitude and longitude after the conversion is not within the United States.

- **Sort the Original Datasets**
  Firstly, We use Microsoft Excel to sort the five states and counties of the original datasets.
- **Abnormal Value Processing**
  Secondly,We observed a small amount of abnormal data appearing,these anomalous datas have an obvious feature. Its geocoded information is not in the USA, then we will delete thiese datas.
- **Get Location Latitude and Longitude**
  After deleting the abnormal datas, We use the Geopy-Geocoding-Toolkit in the Python to output latitude and longitude coordinates in the datasets.

4.2 Data Visualization

5 Model Design

5.1 Model 1 - The Mathematical Model of the Relationship between County Location and Opioids

- **The Polygonal Geometric Center of Gravity Model**
  Our team think that choose the counties that rank the top 4 about using opioids.So, we can infer the location of the counties where the opioids was the earliest used before this year. This location spreads around some counties and states in the form of origin. Therefore,We can put forward to buliding a Polygonal Geometric Center

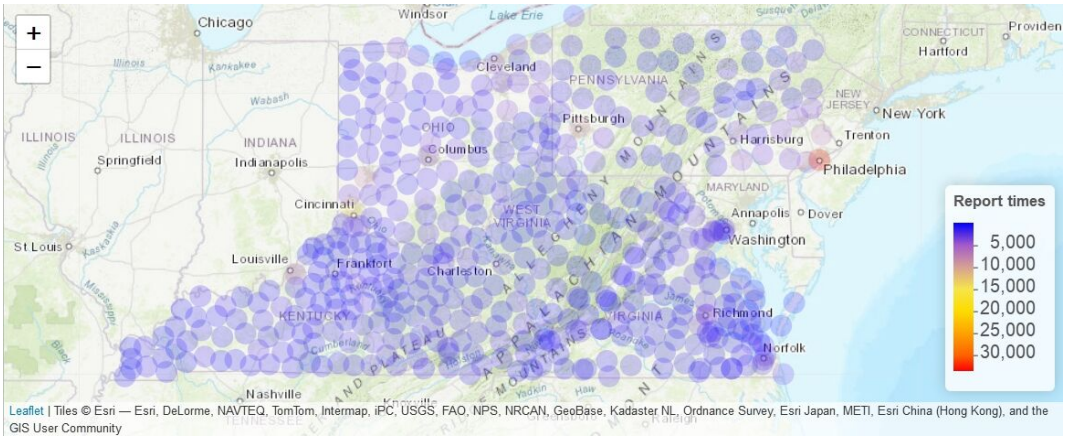


Figure 3: Using The Geopy-Geocoding-Toolkit deal with Data

of Gravity Model.

And then, we need to choose these four origins from the map of each year.Meanwhile, we also need to connect the latitude and longitude coordinates of these four origins with lines to get a polygonal area.In this area, finding out its geometric center of gravity on the map.This geometric center of gravity indicates where the drug was the first reported.It is worth noting that the small areas can be defined as small triangles.

The Geometric Center of Gravity equation for the small triangle is :

$$x_i = \frac{D\iint y d\sigma}{S} = \frac{\sum_{i=1}^{n} y_i \sigma_i}{\sum_{i=1}^{n} \sigma_i} \tag{2}$$

$$y_i = \frac{D\iint y d\sigma}{S} = \frac{\sum_{i=1}^{n} y_i \sigma_i}{\sum_{i=1}^{n} \sigma_i} \tag{3}$$

In this equation, $\boldsymbol{n}$ means each small area of a polygon. $\boldsymbol{\sigma_i}$ means size of each small area, $\boldsymbol{G_i}(x_i, y_i)$ means the geometric center of gravity of a small triangle.

The Geometric Center of Gravity equation for the polygon is :

$$x_j = \frac{\sum_{i=2}^{n-1}(x_1 + x_i + x_{i+1}) \begin{bmatrix} x_1 & y_1 \\ x_i & y_i \\ x_{i+1} & y_{i+1} \end{bmatrix}}{3\sum_{i-2}^{n-1} \begin{matrix} x_1 & y_1 \\ x_i & y_i \\ x_{i+1} & y_{i+1} \end{matrix}} \tag{4}$$

$$y_j = \frac{\sum_{i=2}^{n-1}(y_1 + y_i + y_{i+1}) \begin{bmatrix} x_1 & y_1 \\ x_i & y_i \\ x_{i+1} & y_{i+1} \end{bmatrix}}{3\sum_{i-2}^{n-1} \begin{matrix} x_1 & y_1 \\ x_i & y_i \\ x_{i+1} & y_{i+1} \end{matrix}} \tag{5}$$

In this equation, $\boldsymbol{n}$ means each small area of a polygon. $\boldsymbol{\sigma_j}$ means size of each polygon, $\boldsymbol{G_j}(x_j, y_j)$ means geometric center of gravity of the polygon.

5.2 Model 2 - The Grey Models for Short-term Predicts the Total Drug Reports in the future

- **The Grey Models for Short-term Predicts**
  According to Model 1,we have already got the location of the counties where the earliest using the opioids.At the same time,our team also observed a phenomenon about that the total number of drug reports county is increasing.But in other counties.The total number of drug reports county has not changed.

  Abuse of opioids is still growing,The total drugs reports as predictive sample is too single.We will intend to use the gray model to predict the sample.

  There are some steps about the Gray Model :

  - *Step 1: Determine the Reference Sequence*
    According to economic implications of each predicted standard.Selecting the optimal standard value among n predicted object to consist the reference sequence.Δ_{ij} means **the reference sequence**.

    $$\Delta_{ij}\left|x_{ij}\right|, i = 1, 2, ..., n; j = 1, 2, ..., p \tag{6}$$

    k means **predicted object**, p means **predicted standard**.

    $$x_k = \{x_i + x_{i+1} + ... + x_n\}(i = 1, 2, ..., n) \tag{7}$$

- *Step 2: Data Preprocessing*
  We observed a small amount of missing values and abnormal values.We intend to use *lagrangian* interpolation to preprocess the miss values and delete completely the abnormal values.

- *Step 3: Maximum Difference and Minimum Difference*
  The **maximum difference** and **minimum difference** equations is :

$$\Delta_{\max} = \max_{1\le i\le o} \max_{1\le i\le 0}(\Delta_{ij}) \tag{8}$$

$$\Delta_{\min} = \min_{1\le i\le o} \min_{1\le i\le 0}(\Delta_{ij}) \tag{9}$$

- *Step 4: Calculate the Gray Correlation*
  The gray correlation equation is :

$$\delta_{ij} = \frac{\Delta(\min) + \sigma\Delta(\max)}{\Delta ij + \sigma\Delta(\max)} \tag{10}$$

  In this equation, δ_{ij} means **the relationship** between predicted object and predicted standard. σ means **distinguish coefficients**.

- *Step 5: Calculating Prediction Coefficients*
  The prediction coefficients equation :

$$E_i = r_i * 100 \tag{11}$$

  In fact, the prediction coefficients and the grey correlation have the same meaning.

- *Step 6: Comparing and Sortng the Predicted Object*
  The feature of using the gray model to predict is about that we can change σ to improve predicted distinction.

5.3 Model 3 - The Logistic Regression for Medium-term Predicts the Total Drug Reports in the future

- **The Logistic Regression for Medium-term Predicts**
  We observe use the gray model to predict is always changed due to the single sample.This will lead to the gray model is unstable.So we will use the logistic regression for the second medium-term forecast of the total drugs reports.
  The Logistic Regression equation is :

$$\mathrm{Y} = \mathrm{f}(X_1, X_2, X_3, ..., X_P, \theta_1, \theta_2, \theta_3, ..., \theta_k) + \varepsilon \tag{12}$$

  for $\varepsilon \sim N(0, \sigma^2)$
  The sigmoid function is :

$$g(x) = \frac{1}{1+e^{-x}} \tag{13}$$

In this function, $g(x)$ is between 0 and 1.
The decision function is :

$$P(y=1|x;0) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T * x}} \tag{14}$$

In this function, $g(h)$ means the decision function associated with sigmoid function.
The assumption of the logistic regression is :

$$y^* = 1, P(y=1|x) > 0.5 \tag{15}$$

In the assumption,needing to set a prediction threshold $P(t) \geq 0.5$.
If you need higher prediction accuracy, then you can also choose higher prediction threshold.

5.4 Model 4 - The BP Artificial Neural Network predicts the next new disaster location

- **The BP Artificial Neural Network predicts the next position**
  More and more examples causes of death by abuse of opioids in the United States.We still need to predict the next new disaster location about use opioids excessively.The BP artificial neural network has a function of self-adaptation and self-learning,as long as the input data is enough, it can achieve very accurate prediction results.

  So our team will use the BP artificial neural network to predicts the next position.First of all,we will randomly select the location of state which with the most drugs reports data.The data is several input training samples,and the next new location is several output forecast result.

The Parameters of the BP Artificial Neural Network is :

- *Step 1: Determine the Time Sequence*

$$\Delta_{mn} \left| x_{ij} \right|, i = 1,2,...,n; j = 1,2,...,p \tag{16}$$

k means **predicted object**, p means **predicted standard**.

$$x_p = \{x_i + x_{i+1} + ... + x_n\}(i = 1,2,...,n) \tag{17}$$

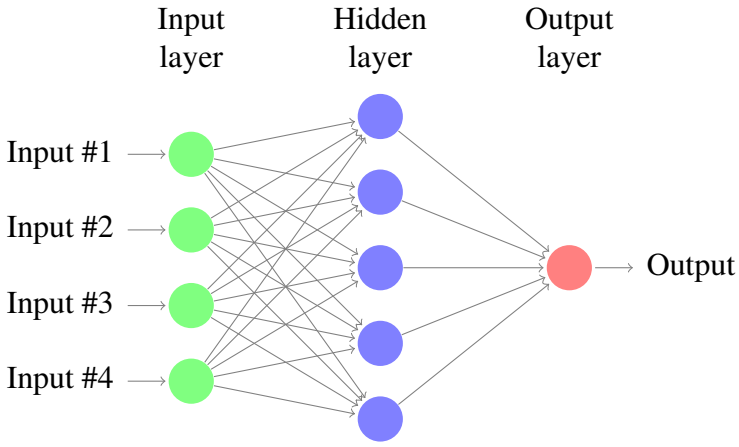


Figure 4: The BP Artificial Neural Network Model

- *Step 2: Set the sigmoid as a activation function*
  We use the **sigmod** as a activation function to optimize neural network performance and so on.

$$g(x) = \frac{1}{1+e^{-x}} \tag{18}$$

- *Step 3: Set a evaluation function*
  How to evaluate the predictive training model is very important.So we should set a evaluation function (MAE) to evaluate predictive model.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - y_i| \tag{19}$$

  In this equation, f_i means **actual value**, y_i means **predictive value**, $\bar{y}$ means **average predictive value**, N means the **input drugs reports** data.

- *Step 4: Other parameters*
  According to a general academic standards. Defining $0.75n$ input neural network nodes and t means training period.n is input training sample etc.

5.5 Model 5 - The Factor Analysis processing the multidimensional data

- **The Factor Analysis processing the multidimensional data**

  The first,we must find out null values and value with *.If the column has less than 10 such data,then using the Lagrangian interpolation to process.If the column has more than 10 such data,then deleting this data item.In order to better explain the relationship between the various factors, we will delete all the data representing

the percentage except estimate value.We also need to convert the data matrix into correlation matrix and load factor for orthogonal rotation.

$$X = (X_1, X_2, ..., X_p)^T \tag{20}$$

$$E(X) = \mu = (\mu_1, \mu_2, ..., \mu_p)^T, VAR(X) = \sum = (\sigma_{ij})_{p*p} \tag{21}$$

Thus, the factor analysis model is :

$$\left\{ \begin{array}{l} X_1 - \mu_1 = a_{11} f_1 + a_{12} f_2 + \cdots a_{1m} f_m + \varepsilon_1 \\ X_2 - \mu_2 = a_{21} f_1 + a_{21} f_2 + \cdots a_{2m} f_m + \varepsilon_2 \\ \vdots \\ X_p - \mu_p = a_{p1} f_1 + a_{p2} f_2 + \cdots a_{pm} f_m + \varepsilon_p \end{array} \right\} \tag{22}$$

for $f_1, f_2, ..., f_m, (m \prec p)$ means **common factor**, $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ means **special factor**.They are random variables that cannot be observed.

Thus, the above equation can be written as a matrix representation.

$$X = \mu + AF + \varepsilon \tag{23}$$

Here is the general assumption :

$$E(F) = 0, Var(F) = I_m$$

$$E(\varepsilon) = 0, Var(\varepsilon) = D = diag(\sigma_1^2 \sigma_2^2 ..., \sigma_p^2)$$

$$Cov(F, \varepsilon) = 0$$

6 Model Solve

6.1 Model 1 - The Polygonal Geometric Center of Gravity Model Solution

- **Model 1 - The Polygonal Geometric Center of Gravity Model**
  In the previous data preprocessing stage, we choose several counties that rank the top to used opioids.In order to easily analyze, we only select the counties that rank the top 4 position in five states to calculate latitude and longitude coordinates.
  According to The Geometric Center of Gravity Model which is be proposed by us.The four vertices of a polygon are consist of four geographic coordinates to infer to geometric center of gravity.So we can know where the earliest counties to use opioids were reported in five states.

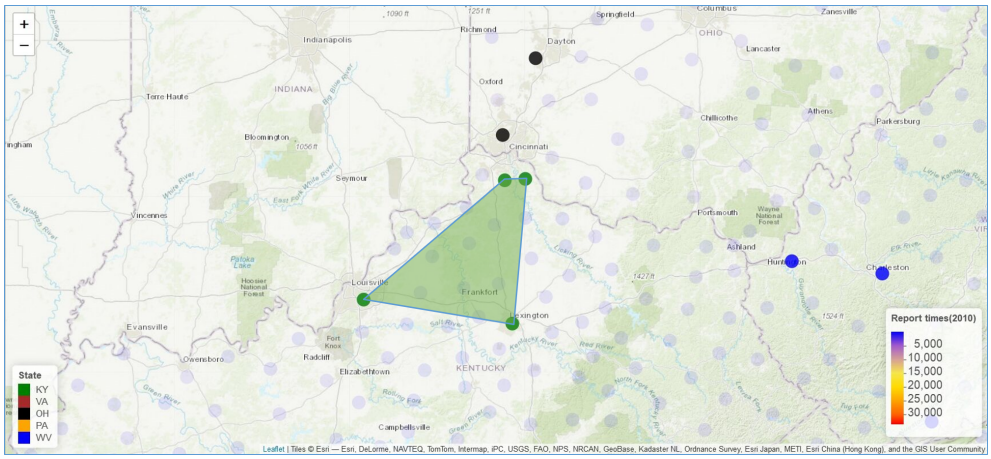
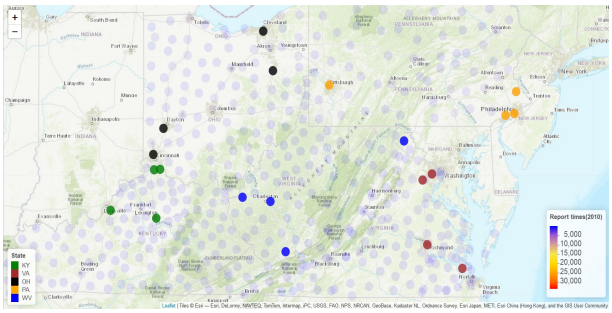


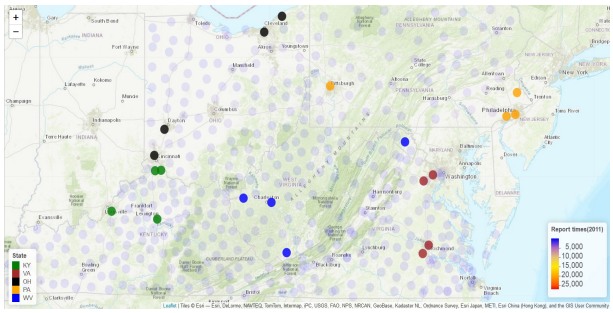
Figure 5: The Geometric Center of Gravity of the Polygon

Finally,we need to use the 8 years from 2010 to 2017 as the time variable and the total number of drug reports county as independent variables. Marking 8 geometric centers of gravity on the map.The areas where the earliest use of opioids in these 8 years are covered by these 8 geometric centers of gravity.Otherwise stated, The geometric center of gravity will always be moved on the map.

Here is the ranking of the total drug reports country in 2010 and 2011.

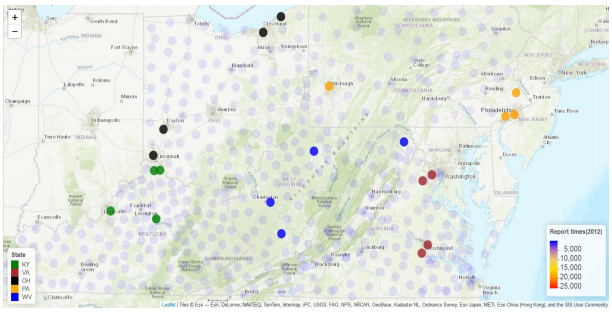


(a) The top 4 counties of the total drug reports datas in five states in 2010.

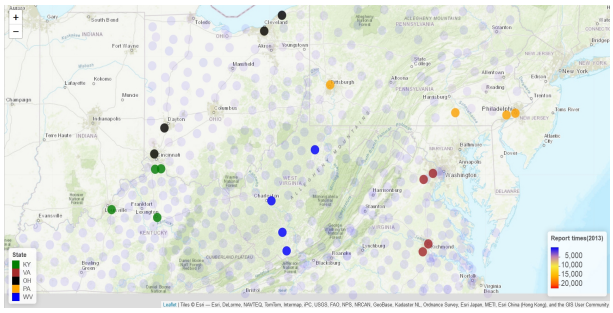


(b) The top 4 counties of the total drug reports datas in five states in 2011.

Here is the ranking of the total drug reports country in 2012 and 2013.

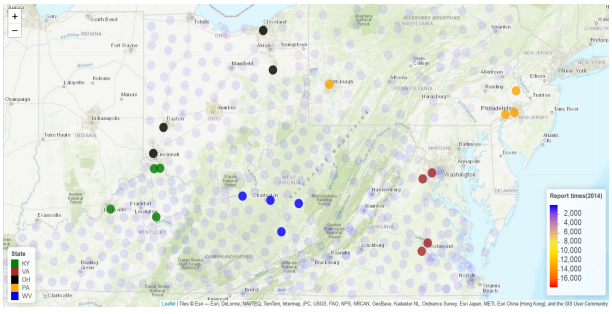


(c) The top 4 counties of the total drug reports datas in five states in 2012.

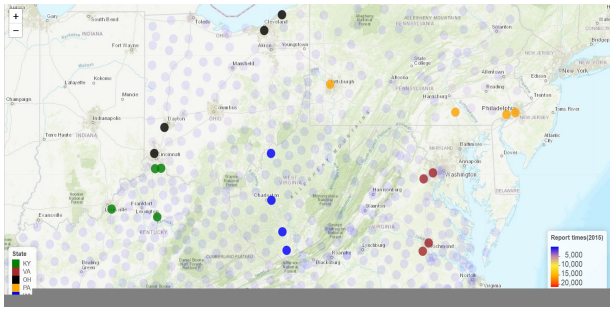


(d) The top 4 counties of the total drug reports datas in five states in 2013.

Here is the ranking of the total drug reports country in 2014 and 2015.

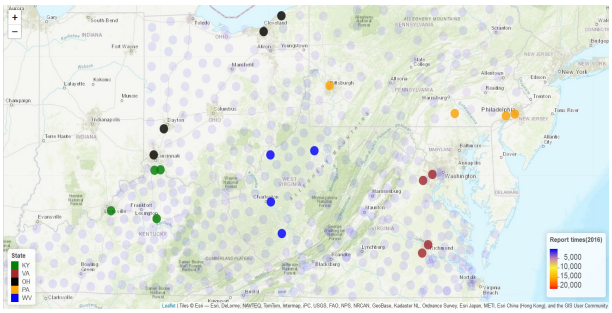


(e) The top 4 counties of the total drug reports datas in five states in 2014.

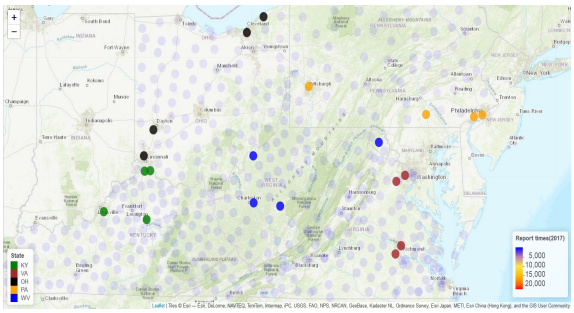


(f) The top 4 counties of the total drug reports datas in five states in 2015.

Here is the ranking of the total drug reports country in 2016 and 2017.



(g) The top 4 counties of the total drug reports datas in five states in 2016.



(h) The top 4 counties of the total drug reports datas in five states in 2017.

State	Longitude	Latitude
KY	84.893 907 18°W	38.342 108 78°N
VA	77.543 588 72°W	38.260 264 87°N
WV	80.125 077 97°W	38.596 155 50°N
OH	83.229 989 74°W	40.216 172 05°N
PA	77.015 097 42°W	40.187 376 16°N

Table 2: The location of the state that was the earliest reported to use opioids

According to the polygonal geometric center of gravity model,we can observe clearly the top 4 counties associate with 4 vertexs of the total drug reports datas in five states every year.And the five states have five geometric centers of gravity on map.So there are some longitude and latitude coordinates in five states.
At the end,we add and average the eight years of the total drugs reports data to get longitude and latitude coordinates.

Here are the latitude and longitude of the four geometric centers.The geometric centers stand for **the location which the earliest use of opioids**.

6.2 Model 2 - The Grey Models for Short-term Predicts Solution

According to the gray model,we intend to sum up the total drug data for eight years.Here is the total drug data table for **eight years**.

Table 1 The total drug data about all years.

State \ Year	2010	2011	2012	2013	2014	2015	2016	2017
KY	10453	10289	10722	11148	11081	9865	9093	9394
OH	19707	20330	23145	26846	30860	37127	42470	46104
PA	19814	19987	19959	20409	24904	25651	26164	27894
VA	8685	6749	7831	11675	9037	8810	10195	10448
WV	2890	3271	3376	4046	3280	2571	2548	1614

We use the gray model to predict the tatal drugs data about all years.As long as we have already gotten the Grey-correlation of each data.

State	Grey-correlation	Precision ≤ 0.35
PA	excellent	0.178
OH	excellent	0.059
WV	average	0.457
KY	fair	0.518
VA	poor	0.665

Table 3: Gray-correlation of each state

In this table.If the gray correlation is more lower,Then the predictive value will get more acurrate.We can draw some important conclusions.

- The total drug reports is more higher in the Pennsylvania and Ohio
- The total drug reports is more lower in the Kentucky
- The total drug reports is general in the West Virginia.
- The total drug reports is the lowest in the Virginia.

There are some charts about the ratio between predictive value and original value.

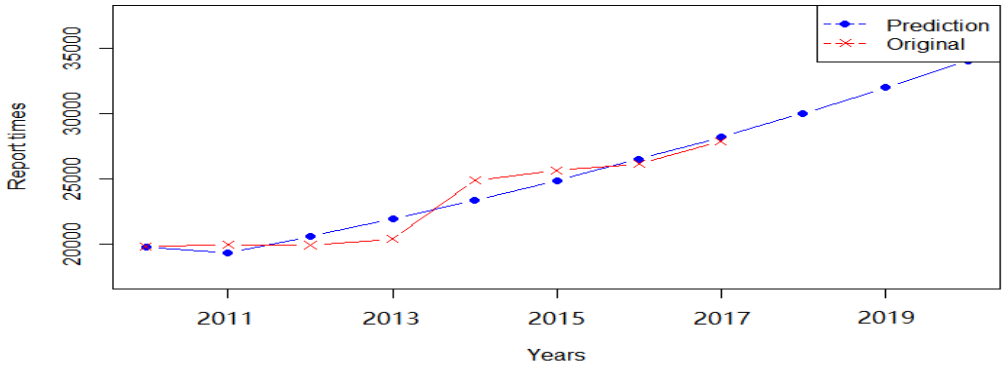
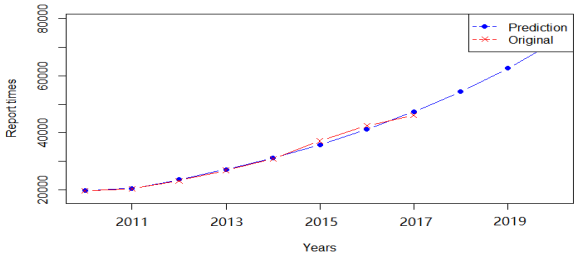
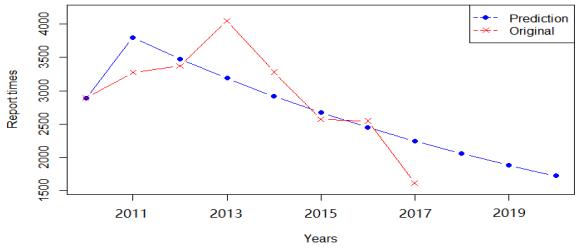


Figure 6: PA



(a) OH

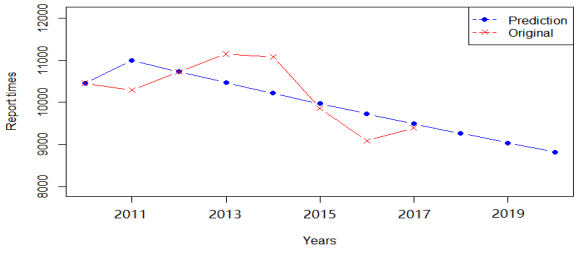


(b) VA

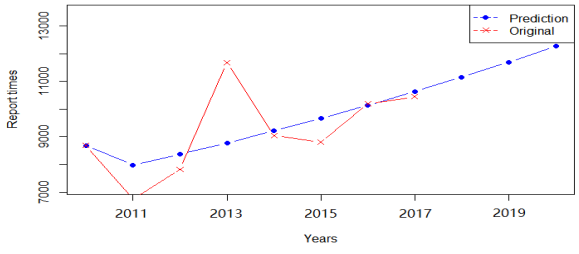
6.3 Model 3 - The Logistic Regression for Medium-term Predicts Solution

According to the sample data is single,We found that the prediction using the GM model is not very good.So we dicided to use the Logistic Regression predicted the total drug reports situation. In this table,we can draw some important conclusion.

- The small dots is raw data before 2017.And the small dots is predictive data in 2017 to 2020.



(c) KY



(d) WV

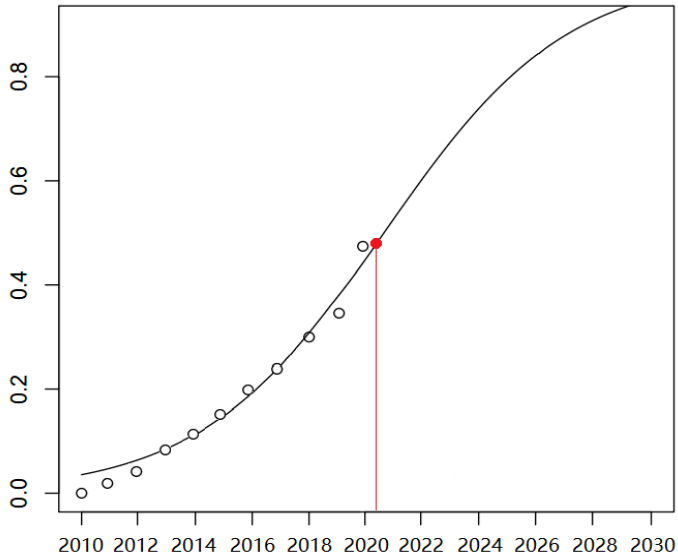


Figure 7: The logistic regression predicted total drug reports situation in 2010 to 2020.

- If the X-axis is 2020,Then the Y-axis means that is the highest drug identification threshold levels.
- This table approximate to gray model prediction results in the Pennsylvania.

6.4 Model 4 - The BP Artificial Neural Network predicts the next new disaster location

- For each states, h means input variable.
  - h_1 means the coordinates of the top 20 counties in 2010 to 2014.
  - h_2 means the coordinates of the top 20 counties in 2011 to 2015.
  - h_3 means the coordinates of the top 20 counties in 2012 to 2016.
- If $epochs$ = 16, the next most likely location for an opioid crisis will appear.
- From the data in the table, we can see that the next geographical location where drug abuse occurs is Kentucky.

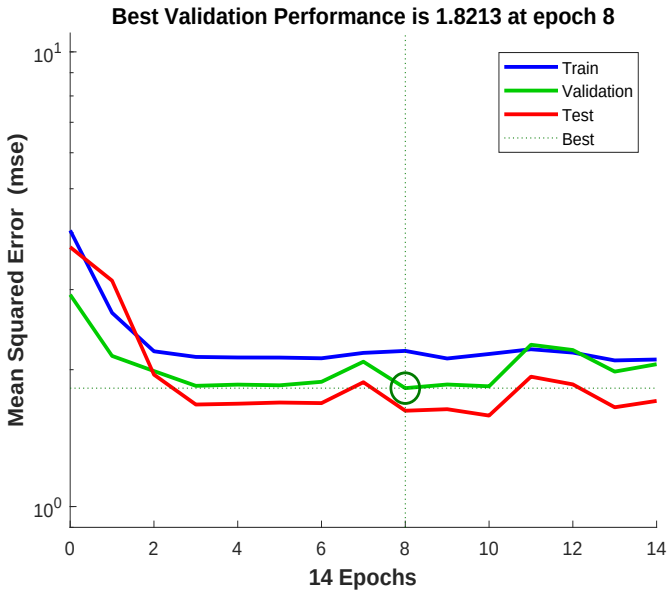


Figure 8: The BP artificial neural network predicts the geographical position of 20 counties.

longitude	85.604°W	84.386°W	84.535°W	84.424°W	83.557°W
latitude	38.085°N	38.074°N	38.941°N	38.926°N	36.746°N

Table 4: Predicting the next geographic location where abuse of opioids occurs.

6.5 Model 5 - The Factor Analysis process the multidimensional data

For question two, we need to use the socio-economic data from the US Census to analyze the combined causes of the crisis of opioid abuse.Our team observed that this is a very large amount of data, so we will use a factor analysis method to reduce it.

Take the socioeconomic data of the 2010 American Census as an example,we cleaned this data.When choosing the number of load factors, we find that it is feasible to use two

longitude	85.959°W	84.3°W	87.046°W	84.141°W	86.516°W
latitude	37.611°N	37.755°N	37.667°N	37.198°N	37.058°N

Table 5: Predicting the next geographic location where abuse of opioids occurs.

longitude	84.633°W	88.791°W	85.507°W	84.667°W	83.243°W
latitude	39.018°N	36.986°N	37.844°N	37.157°N	37.251°N

Table 6: Predicting the next geographic location where abuse of opioids occurs.

longitude	82.696°W	87.518°W	83.149°W	82.725°W	85.768°W
latitude	38.417°N	37.381°N	36.808°N	37.487°N	38.024°N

Table 7: Predicting the next geographic location where abuse of opioids occurs.

factors to indicate the degree of correlation between data.

	Factor1	Factor2	Factor3	Factor4	Factor5
SS loadings	137.677	73.618	10.505	6.19	10.784
Proportion Var	0.499	0.267	0.038	0.022	0.039
Cumulative Var	0.499	0.766	0.804	0.826	0.865

Table 8: Five load factors contribution

We can draw an important conclusion that Factor 1 and Factor 2 can respectively distinguish the US population data into two dimensions.At the same time, their cumulative contribution reached 80 percent.Factor 1 represents a group of people with poor marital status and parenting.but Factor 2 represents unmarried and children without children.

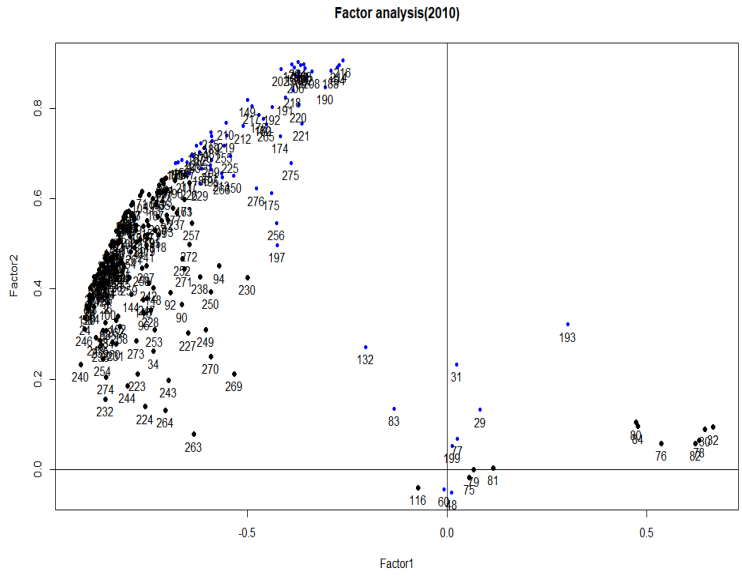


Figure 9: Distribution of population data in two load factors

7 Strengths and weaknesses

References

[1] D. E. KNUTH The TEXbook the American Mathematical Society and Addison-Wesley Publishing Company , 1984-1986.

[2] Lamport, Leslie, LATEX: " A Document Preparation System ", Addison-Wesley Publishing Company, 1986.

[3] http://www.latexstudio.net/

[4] http://www.chinatex.org/

Appendices

Appendix A First appendix

Here are simulation programmes we used in our model as follow.

Input matlab source:

```
function [t,seat,aisle]=OI6Sim(n,target,seated)
pab=rand(1,n);
for i=1:n
    if pab(i)<0.4
       aisleTime(i)=0;
    else
        aisleTime(i)=trirnd(3.2,7.1,38.7);
    end
end
```

Appendix B Second appendix

some more text **Input C++ source:**

```
//============================================================================
// Name        : Sudoku.cpp
// Author      : wzlf11
// Version     : a.0
// Copyright   : Your copyright notice
// Description : Sudoku in C++.
//============================================================================
```

```cpp
#include <iostream>
#include <cstdlib>
#include <ctime>

using namespace std;

int table[9][9];

int main() {

    for(int i = 0; i < 9; i++){
        table[0][i] = i + 1;
    }

    srand((unsigned int)time(NULL));

    shuffle((int *)&table[0], 9);

    while(!put_line(1))
    {
        shuffle((int *)&table[0], 9);
    }

    for(int x = 0; x < 9; x++){
        for(int y = 0; y < 9; y++){
            cout << table[x][y] << " ";
        }

        cout << endl;
    }

    return 0;
}
```