An Opioid Crisis Characterization and Prediction Model Based on Factor Analysis

Jialin Chen
518071910001

Abstract

Opioid abuse is a major concern in the United States. In order to solve this harm, it is of great significance to study the law of development and spread of drugs for the mitigation and control of drug harm. In this paper, we obtain the general theory of drug growth and spread by analyzing the drug reports and socio-economic data of WV, OH, KY, VA, and PA. Our conclusions can help and guide the control of opioid epidemics.

We first creat a visual chart of the states' annual drug report data, and summarize the overall development trend. We divide the causes of changes in the number of drug reports into natural growth and inter-regional migration. Moreover, we find the invariant in the process of movement: **Mass Center of Drug**. Refer to **Bass Model** and **Linear Model**, We build a comprehensive model to explain the two causes of increase in drug use. At the same time, the **Force Directed Model** is established to describe the spread effect on the basis of **Field Theory** in Physics. Then we try to find the most comprehensive and concise evaluation index from 608 social and economic evaluation indexes. We carry out **Cluster Analysis, Decision Tree, Pearson Correlation Coefficient Analysis**, which provide a reliable guarantee for **Exploratory Factor Analysis**. After Factor Rotation, we find the most interpretable 9 factors. By using the methods of **Multiple Linear Regression** and Stepwise Regression, we extract four highly representative indicators: Female Unhappy Marriage Indicator, Age Structure Indicator, Male Unhappiness Indicator, and Career Indicator.They are important factors affecting the proportion of drug users. According to our conclusion, we optimize the previous model and give an evaluation method for the effectiveness of anti-drug policy.

We finally conduct sensitivity analysis, dissect pros and cons of our model and present a memo of our work to the chief administrator.

1 Introduction

1.1 Restatement of Problems

Opioids are substances that act on opioid receptors and produce morphine effects, which are mainly used for anesthesia in medicine. But an overdose of opioids can lead to addiction, health damage and even death. Since the 1990s, a national epidemic of opioid use has ignited the deadliest drug crisis in American history. From 1999-2017, almost 400,000 people died from an overdose involving any opioid, including prescription and illicit opioids.Today, the opioid epidemic is still a serious problem in the United States, which will not only has a negative impact on people's physical health, but also lead to the lack of talents in important areas of the country and hinder economic development.

In the United States, there are five states: Ohio (OH), Kentucky (KY), West Virginia (WV), Virginia (VA), and Pennsyivania (PA) that suffer from a severe opioid epidemic. We performed statistical processing on the given data and made a graph (Figure 1 to show the number changes of drug reports in the five states in 2010-2017, where we find a steady increase in the number of drug reports, especially in OH.

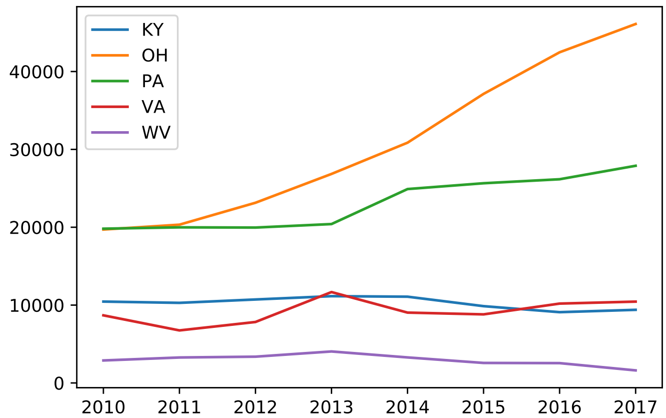


Figure 1: The figure of drug reports number change in 2010-2017

Many models have been proposed to predict similar phenomena. The Bass Diffusion Model [1] proposed by Frank m. Bass are often used as market analysis tools to predict the demand for new products. It inspired us to model the growth of drug use. The factor analysis [2] method proposed by C.E. spielman can find the hidden representative factors among many variables, which helped us to select the right factors from the high-dimensional, complex data to describe changes in the drug population. Therefore, we will build a comprehensive model to analyze and predict its trend.

1.2 Overview of Our Work

As required by the question, we solved the problem by the following steps.

- Build a model to describe the spread and characteristics of the opioids between the regions and locate the positions of origin. We are going to go down to a specific county.
- Predict the development of the opioid crisis and propose the future's potential concern about it.
- Identify socio-economic factors that influence the use of opioid and optimize our model by adjusting the parameters.

- Propose a response for the opioid crisis and test its validity.

Our process is shown in Figure2.

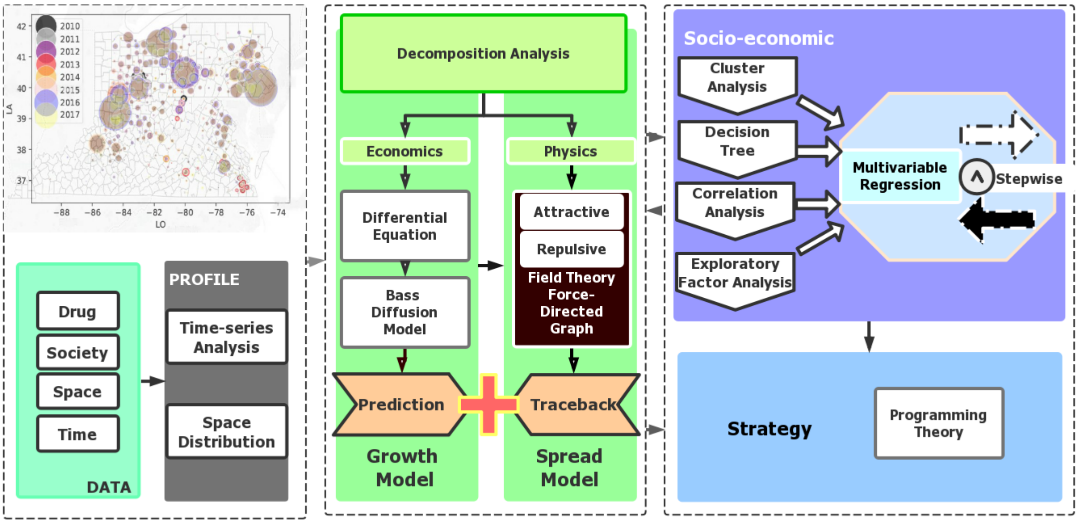


Figure 2: The process of model work

2 Assumptions and Notations

2.1 Assumptions

In our model, we make the following basic assumption to simplify the problem. Based on the fact, we explained the rationality of our assumptions.

Assumption1. *Each state has roughly equal enforcement of opioid laws and regulations. Therefore the number of the reported synthetic opioid and heroin incidents (cases) was linearly correlated with the number of opioid addicts, namely the severity of the opioid crisis in this state.*

Because we are not allowed to use other necessary data to quantify the state's ability to enforce the law, we assume this to simplify our model.

Assumption2. *We ignore the state's characteristics of land and sea.*

Drugs can be obtained from both legal (prescription use) and illegal (non-prescription use) sources. Actually, coastal states have easier access to the illegal drug trade, while inland states mostly get opioids through prescription drugs from hospitals. Due to the lack of necessary data, we make this reasonable assumption.

Assumption3. *The change of overall number of cases can reflect change trend of the prevalence of opioids better than specific opioid [3].*

It is proved that there has been substantial variability with which specific drugs have become dominant in varying populations and geographic locales. This variability all but negates the possibility of confident predictions about the future role of specific drugs.

Assumption4. *In our model, we do not consider the personal factor [4].*

The rate at which people develop tolerance and addiction depends on factors such as genetics. However, those factors might fade into the averages if researchers looked at a population as a whole.

Table 1: Notations

Symbol	Definition
m	The Number of Potential Drug Users
$T_i(t)$	The True Number of SOHI in County(i) at Time t
$N_i(t)$	The Natural Increase in Drug Users in County(i) at Time t
$H_i(t)$	Changes in Drug Users Caused by Inter-regional Migration in County(i) at Time t
$DDF_{ij}(t)$	Drug Diffusion Force Between County(i) and County(j) at Time t
d_{ij}	The Distance Between County(i) and County(j)
$RR(t)$	The Value of ReportRate at Time t
P	The Total Number of People in a Certain County
MC	Mass Center of SOHI
$SOHI$	The of Reported Synthetic Opioid and Heroin Incidents
LA	Latitude
LO	Longitude

2.2 Notations

The primary notations used in this paper are listed in **Table 1**. There can be some other notations to be described in other parts of the paper.

3 Data preprocessing

For data-analysis problem, there are some redundant and useless data in the large amount of row data. Therefore, we need to perform data preprocessing by cleaning, selecting, and integrating the data. Thus, we can conduct the follow-up studies.

Data Cleaning

We deleted the variables with a large amount of data missing, for they cannot provide us enough information. Moreover, we filtered out invalid data and removed the variables which are completely linearly dependent to simplify our dataset.

Data Normalization

Since we will use indicators with different units in part 2, it's significant to normalize the indicators and scale all the values in the range of [0,1]. What's more, normalization reduces the order of magnitude, thereby reducing the nature linear correlation between variables. Formula(1) gives a general form of the normalization.

$$x_{new} = \frac{x}{\Sigma x_i} \tag{1}$$

Where x_i are respectively values of the indicators of the same class.

4 Model Construction

4.1 Simple Model With the NFLIS Data

4.1.1 Opioids Spread and Characteristics

After data preprocessing, we add up all the values of DrugReports that belong to the same state. Based on our previous **Assumption3**, it is an variable that indicates the size of total number of opioid addicts in a state, namely the severity of the opioid crisis in this state. Based on the data analysis, we found that the choice of synthetic opioids or heroin is largely influenced by price factors. There is a competitive relationship between the two and they are substitutes for each other. Due to the loss of data such as prices, we only discuss the change in the sum over time.

In order to create a model to describe the spread and characteristic of SOHI between regions, we clarified two methods of changing the number of opioids addicts in a state. One is natural growth of drug users in closed areas, the other is inter-regional migration. There are significant differences between these two methods. The former causes an change in number, while the latter causes a change in distribution.

Therefore, we proposed a novel model to describe the characteristics of drugs when they spread in different methods.

Growth of drug users in closed areas

The spread of drugs is consistent with economic principles partly. In addition, drug addiction is a kind of contagious disease. In view of the above characteristics, we proposed a novel model, combining **BASS Diffusion Model** [6]and **Linear Model**.

To determine the growth pattern of SOHI in closed areas, we delved into the characteristics of opioid epidemics [5]. There are some people in the population who are impossible to become drug users, such as those without economic ability or those who have a very clear understanding of the harm of drugs. Therefore, we did not take the total population as a parameter, but the number of potential drug users m as a parameter.

Since the drug itself is a kind of commodity, its spread is affected by both internal and external causes. Internal cause refers to the influence of interpersonal communication, such as the imitation of the susceptible person to the addict. We use q to describe this effect. External cause refers to the publicity of mass media and the spontaneous craving for opioids of the susceptible which can be describe by parameter p.

Considering that drug addiction is a kind of physiological disease and that the number of drug reports is decreasing in some states as shown in Figure1, we add a cure factor $R(t)$ to the equation. This is the result of the government's anti-drug campaign. When the number of drug cases far exceeds the government's ability to control, the cure effect reaches saturation and the cure capacity remains constant. Therefore, $R(t)$ can be expressed as $R(t) = kt$. A **Differential Equation** as Equation 2 can be used to reveal the change of the total number.

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = p[m - N(t)] + q\frac{N(t)}{m}[m - N(t)] - k \tag{2}$$

After integrating the equation, we can get the following formulas.

$$N(t) = m\left[\frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}\right] - kt \quad ; \quad n(t) = m\left[\frac{p(p+q)^2 e^{(p+q)t}}{[p + qe^{-(p+q)t}]^2}\right] - k \tag{3}$$

where $N(t)$ and $n(t)$ are respectively the total number of drug addicts and its increase at time t.

We randomly chose KY and OH to validate our model. The program solved the optimal function and output the value of parameters as shown in Table 2. The code of the Bass model algorithm is shown in Appendix. Figure 3, Figure 4 showed the fitting curve and actual curve of the two states. The X-axis is the year, and the Y-axis is the total number of drug addicts. According to our observation, both upward trend and downward trend suits our model well. Bass model is suitable for the sales forecast of new products. Therefore, this model will produce errors when used for long-term prediction.

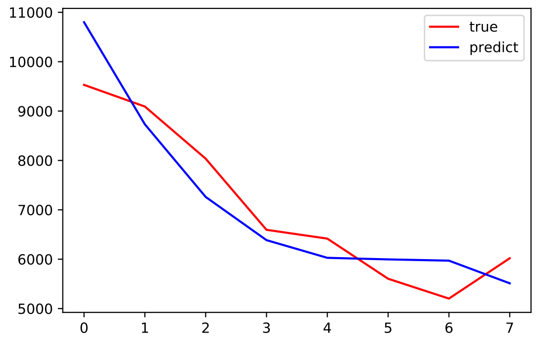


Figure 3: The fitting figure of KY

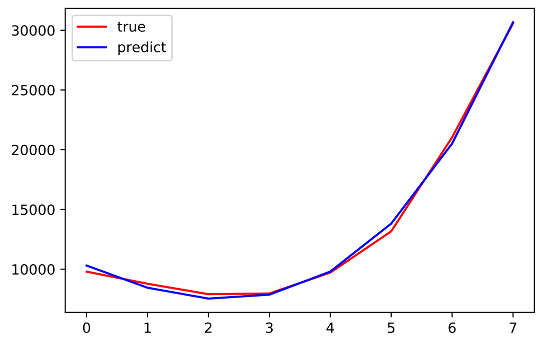


Figure 4: The fitting figure of OH

Table 2: The parameters'value of $N(t)$ in KY and OH

	p	q	m	k
KY	0.0175	0.3555	629077.8098	5129.8628
OH	0.0042	0.4162	2600345.8826	5079.6194

Inter-regional Migration

The difference between the fitted value and actual value is considered as the effect of inter-regional migration, thus we define the difference between the two value as

$$H(t) = T(t) - N(t),$$

where $T(t)$ means the true number of the drug addicts.

A county's drug addicts rate, geographic distance and other factors can attract or repel drug users. In view of such characteristics, we built a **Force Directed Model** based on **Field Theory** in physics to describe their interaction force which is the difference between repulsion and gravity. We named it of **Drug Diffusion Force (DDF)**, and the net DDF of $county(i)$ to $county(j)$ is defined as Formula 4.

$$DDF_{ij}(t) = \frac{T_i(t) - T_j(t)}{d_{ij}^2} \tag{4}$$

This equation reveals that the **Drug Diffusion Force** is proportional to the absolute value of the difference in the number of drug cases between the two places, and inversely proportional to the square of the distance between the two places. The direction is from the areas where drugs are more epidemic to less epidemic. Then take a region as the scope of consideration, one county in the region is affected by the drug diffusion force generated by other counties. The greater DDF is, then the greater rate of growth do county's drug cases have due to inter-regional migration.

Below, we randomly selected a square region with a longitude between -86 and -84 and a latitude between 37 and 39 in 2016 and 2017 as samples to verify our model. Because the counties located at the regional boundaries are greatly affected by the counties outside the region, we took counties in the center as samples and calculated DDF value and $H(t)$ of each county in 2016-2017. We plot out data as Figure 5 shown.

We can arrive at the conclusion that difference is linearly related to the sum of DDF it is subjected to. Here the difference is exactly $H(t)$. After calculating, we confirmed that the

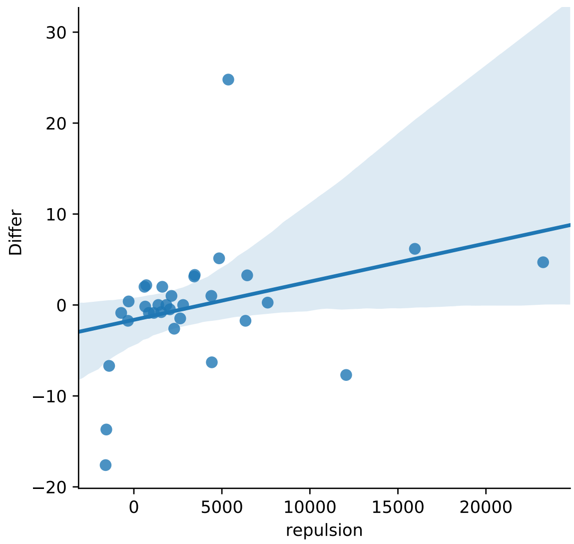


Figure 5: The relation between difference and the sum of DDF

function has the following form.

$$H_j(t) = k \cdot \sum_{i \neq j} DDF_{ij}(t) + b \tag{5}$$

where $k = 0.0004204$ and $b = -1.6301372$. The Force Directed Model measures the impact of a drug epidemic county on other surrounding counties and explained the problem of inter-regional migration to some degree.

4.1.2 Original Location of Specific Opioid

The location where specific opioid was first used has important guiding significance for controlling drug epidemic. We plot a **Bubble Chart** (Figure 6) to show the development and spread of SOHI over time. According to Figure 6, we found that the aggregation degree and locations

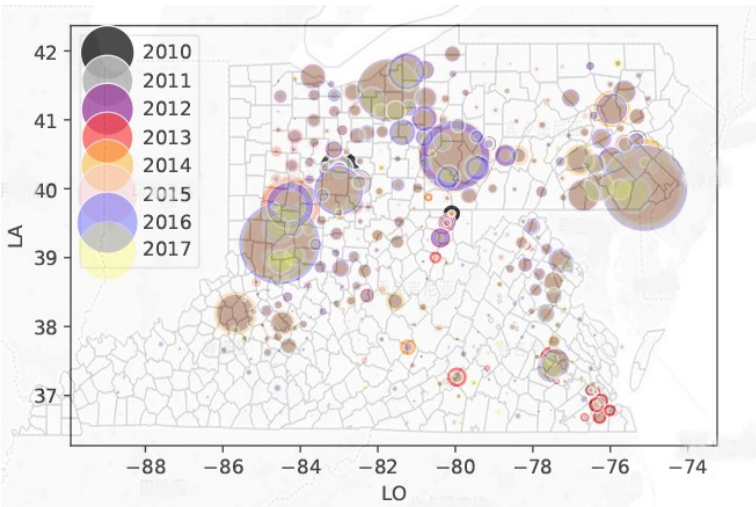


Figure 6: Development and spread of SOHI over time. Here circles of different colors represent different years, and the area of the circles indicates the value of SOHI in that year

of circles vary with the years. Therefore, we introduce "**Mass Center of SOHI**" to describe the cluster center of SOHI. Its position is defined as Equation 6.

$$\begin{cases} LA(t) = \frac{\sum_i T_i(t) \cdot LA_i}{\sum_i T_i(t)} \\ \\ LO(t) = \frac{\sum_i T_i(t) \cdot LO_i}{\sum_i T_i(t)} \end{cases} \tag{6}$$

where LA_i and LO_i mean latitude and longitude of $county(i)$.

SOHI spread evenly around its origin position. According to **Assumption2**, even if the external force is suppressed in the process of propagation, such as government's drug enforcement, the location of the suppressed is completely random. Therefore, the overall trend remains uniform diffusion. With the NFLIS data provided, we can figure out the positions of "**Mass Center of SOHI**" (hereinafter called the **'MC'**) every year. Put the position of MC on the longitude and latitude chart, and the change of the position shows the spread path of MC. Its origin position can be located by tracing back the trace.

Randomly, we choose oxymorphone in WV and propoxyphene in KY as our object of study and analyzed the spread of them. We visualized its trajectory as Figure 7 shown.

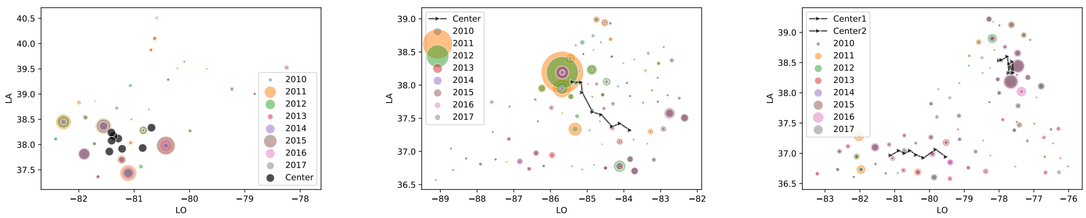


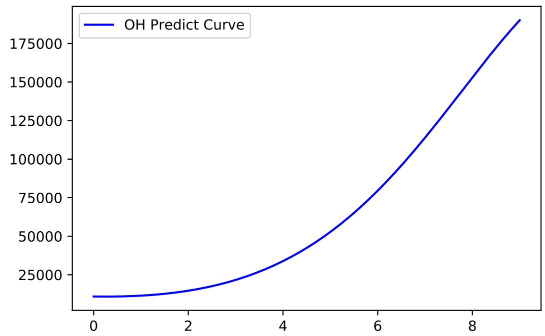
Figure 7: Examples of MC movement

- In Figure 7(a), we can conclude that the position of MC is nearly unchanged, which is concentrated in the center of the state. Therefore, it is most likely the source of oxymorphone. After checking the FIPScode-Coordinate Table, we locate it as KANAWHA.
- As Figure 7(b) showed, the original position where propoxyphene was first used is at the border of KY and Indiana (IN). There is a certain proportion of the drug use diffused into IN in the process of spread. Due to the loss of statistics of IN, the data in KY has more weight, as caused the movement of MC. According to our observation, MC moved approximately in the normal direction of the boundary, which suits our model perfectly. Trace back the trace, we can locate the position where the drug was first used.
- Since the geographical span of the state is very large, some drugs may originate from multiple places at the same time, thus the MC of different sources will move in a combined way of the above two way. An example is provided by Figure 7(c).

4.1.3 Prediction of the Opioids Spread

Derivative can reflect the spread of drugs in the state. When the growth rate reaches the maximum, the drugs in the state are most likely to lose control, triggering large-scale drug abuse, and it is difficult to be completely controlled. Therefore, it is reasonable to define that drug identification threshold is the value of the $T(t)$ when the time derivative of the function reaches the maximum value.

For example, prediction and its derivative of the number of drug cases in OH were figured out by our model. As can be seen from Figure 9, the derivative will reach the maximum value in 2018 and the government will be faced with the crisis of losing control of drugs completely. The threshold value is the the corresponding $T(t) = 14589$.



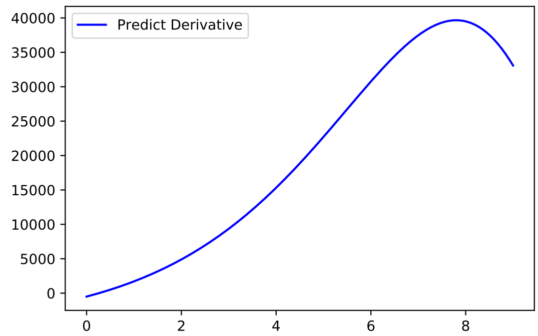


Figure 8: The prediction figure of T(t) in OH Figure 9: The derivative figure of T(t) in OH

At the early stage of spread, the growth rate of drugs is slow. However, it is so concealed that is difficult to catch. However, once the threshold value is approached or reached, a large amount of growth will occur, which is hard to be suppressed. It is hoped that government will not ignore the small data growth and make necessary measures in time, such as strengthening drug enforcement. In countries where the growth curve is declining, drug control is good, such as KY. conversely, drugs tend to get out of control in states like OH. Future drug changes can be predicted based on state or county calendar data. When there is an upward trend in the curve, there may be a drug outbreak in the future, and we should take timely measures to deal with it.

4.2 The Optimization Model With Socio-economic Factors

In the Section4.1.1, we used parameter m to represent the potential drug users in the population. Intuitively, the proportion of susceptible people is related to the socio-economic situation of the region. Therefore, we will analyze the influence of socio-economic factors on the proportion of potential drug users in the population so as to optimize our previous model and better predict the opioid epidemic.

As Section3.1 has mentioned, we normalized the given data by selecting the proportions and scaled all the values in the range of [0,1]. According to the number of people with certain characteristics and its corresponding proportion, we can calculate the total number of people (P) in each county and each state. Then the proportion of potential drug users is $\alpha = \frac{m}{P}$. Moreover, we introduced $ReportRate(RR)$ to linearly express the proportion of drug users. RR is defined as

$$RR(t) = \frac{T(t)}{P}$$

where $T(t)$ is the number of SOHI at time t.

4.2.1 Exploration of Correlation Factors

Cluster Algorithms

To demonstrate that regional $ReportRate(RR)$ is related to socio-economic factors, we first performed **K-means Algorithms**[7] to cluster counties, which is done by minimizing squared errors as Equation 7.

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{7}$$

where $\mu_i = \frac{1}{|C_i|}\sum_{x\in C_i}$, x is the mean vector of cluster C_i and $\|x-\mu_i\|^2$ represents the distance between the sample x and the mean point.

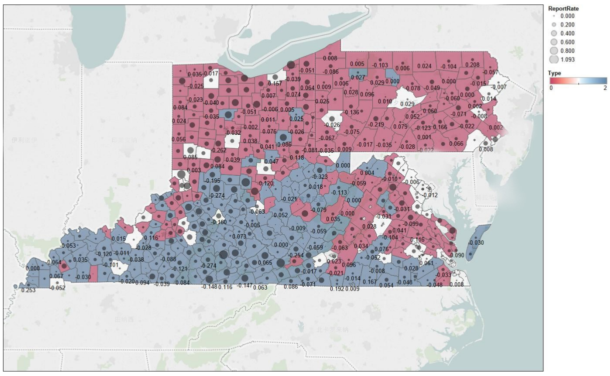


Figure 10: The cluster of counties. Here different colors means different clusters and the area of black circles represents the degree of drug epidemic

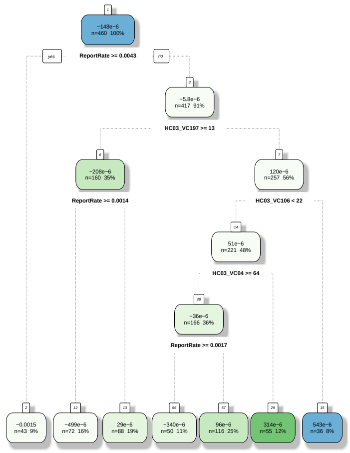


Figure 11: The Decision Tree. Each node represents a decision index, and the samples are classified by successive decisions

Figure 10 proved that there is a certain relationship between the drug use and socio-economic factors, thus we conducted a series of operations to gradually identify the factors significantly related to the number of drug cases.

Decision Tree

In order to find out the indexes affecting the number of drug cases, we presented a Decision Tree Model with Report Rate (RR), Household Rate (HR) and Delta Rate (DR) as the evaluation criteria (as shown in the Figure 11). After verification, we found that there was a high correlation between variables, which means Decision Tree could not fit perfectly.

Pearson Correlation Coefficient (PCC)

Then we use the **Pearson Correlation Coefficient (PCC)** to measure the correlation between the potential influential factors. The Pearson Correlation Coefficient (PCC) is defined as Equation 8.

$$\rho_{(x,y)} = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y} = \frac{E(XY)-E(X)E(Y)}{\sqrt{E(X^2)-E^2(X)}\cdot\sqrt{E(Y^2)-E^2(Y)}} \tag{8}$$

Where the numerator is the covariance and the denominator is the product of the standard deviations of the two indicators.

Correlation Matrix showed the sizing, plus or minus, of the correlation coefficients. Plus means the positive correlation and minus means the negative correlation. The closer the absolute value is to 1, the stronger correlation they share. Figure12 revealed that most variables are highly correlated.

Factor Analysis

How to identify the appropriate representative factors from a bunch of highly correlated variables to describe the number of drug users? We therefore tried **Exploratory Factor**

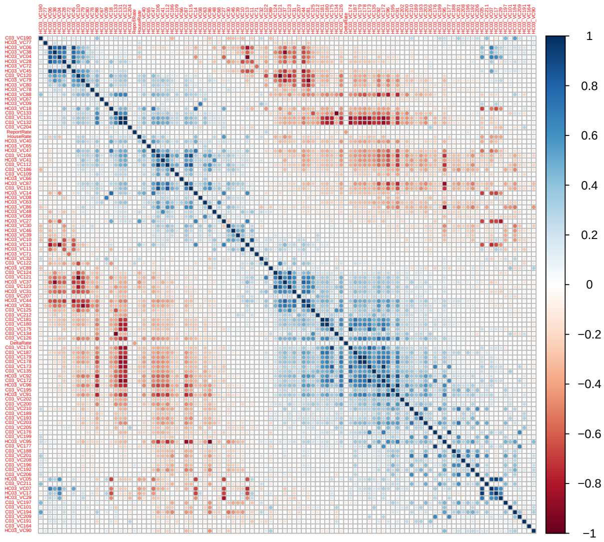


Figure 12: Pearson Correlation Coefficient Matrix

Analysis[8] and determined the number of common factors. Before performing Factor Analysis, we need to perform Kaiser-Meyer-Olkin Factor Adequacy (KMO) and Bartlett Test of Homogeneity of Variances to prove that the partial correlation between factors is good. The test result is shown in Table 3, and the result showed that the samples are suitable for Factor Analysis.

Table 3: Test result of KMO and Bartlett Test

KMO	value	0.75
Bartlett test	K-squared	77027
	df	103
	p-value	2.2e-16

We have a set of p observable ranbom variables, $x_1, \cdots, x_p$ with means $\mu_1, \cdots,$ μ_p. Suppose for some unknown constants ℓ_{ij} and k unobserved random variables F_j (called common factors because they can influence all the observed random variables), where $i \in 1, \cdots, p$ and $j \in 1, \cdots, k$, we have

$$x_i - \mu_i = \ell_{i1}F_1 + \cdots + \ell_{ik}F_k + \varepsilon_i \tag{9}$$

$$\mathbf{X} - \mu = \mathbf{LF} + \varepsilon \tag{10}$$

Here, the ε_i are unobserved stochastic error terms with ze ro mean and finite variance. Matrix **L** is the Loading Matrix.

Then We sort variables by their eigenvalues and plot **the Scree Test** (Figure 13). According to **Kaiser-Harris Principle**, we can retain the factors with eigenvalues greater than 1 as common factors. Figure 13 proved that we need at least 9 common factors to describe the total number. Then we performed **Principal Axis Iteration** to find the common factors. In order to make the meanings of these nine common factors reasonable. We carried out **Factor Rotation** to maximize the variance difference of each factor on the basis that each factor remains orthogonal. The processed nine common factors are named of $PAi(i = 1, \cdots, 9)$ as Figure 14 shown. According to the Loading Matrix, we selected the factors with the largest weight to the common factors. They do a good job of explaining what common factors mean. Therefore, we utilized these components to describe the effects of their corresponding common factors.

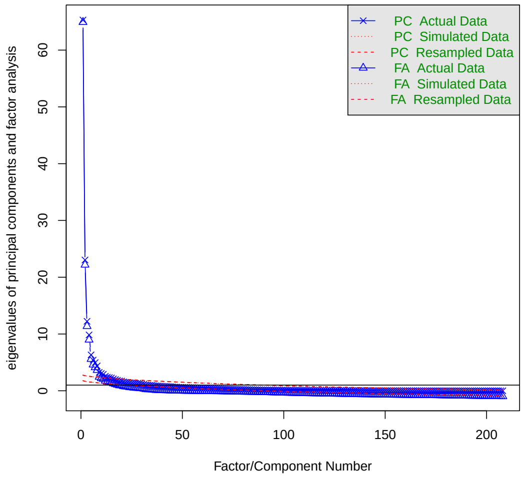


Figure 13: The scree test of factors analysis

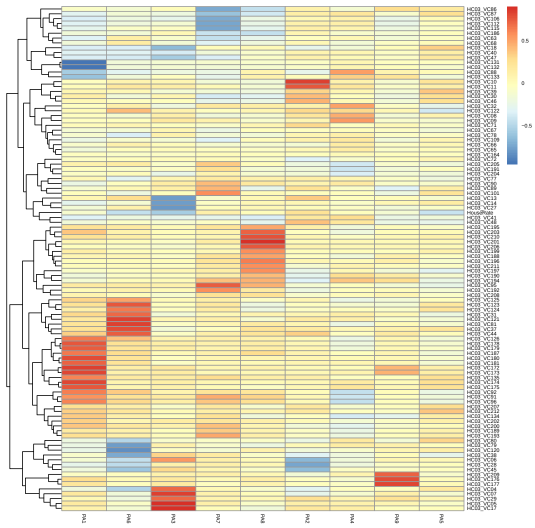


Figure 14: The heatmap of loading matrix

4.2.2 Model Modification and Parameter Adjustment

After Factor Analysis, we selected nine common factors. By introducing them into the model for testing, we finally determined several important factors through **Stepwise Regression Method**. They are PA2, PA3, PA4, and PA7, which are important without significant multicollinearity. Each common factor describes the social economy from a particular aspect so we named them according to their characteristics. The main relevant variables and detailed meanings are given in the Table 4, where P means positively related factors and N means negatively related factors.

Table 4: Meaning of the important indicators

Factor	Correlation	Name of Indicators
PA2	P: Female single parent, divorce rate	Female Unhappy Marriage
	N: Proportion of married women	
PA3	P: Proportion of children in the family	Age Structure
	N: Proportion of living alone and widowhood	
PA4	P: Proportion of divorced, poorly educated man	Male Unhappiness
	N: Proportion of married, highly educated man	
PA7	P: Proportion of veterans, high-education population	Career
	N: Non-disability unemployment rate	

We put the data into program and performed **Multiple Linear Regression**. The regression result is shown in Table 5. From the table, we can see that p-value¡0.05 and the result is desirable.

Therefore, RR can be expressed as Equation 11.

$$RR = 2.049e-05 \cdot PA2 + 2.661e-05 \cdot PA3 + 2.552e-06 \cdot PA4 - 6.985e-0 \cdot PA7 + 0.2601 \quad (11)$$

Table 5: The regression coefficient

PA2	PA3	PA4	PA7	Intercept	p-value
2.049e-05	2.661e-05	2.552e-06	-6.985e-06	2.601e-01	5.827e-05

where $PA2, PA3, PA4, PA7$ are F_2, F_3, F_4, F_7 in Equation 9.

It reveals that drug addiction rate is positive related to Female Marriage Unhappiness Indicator, Age Structure Indicator, and Male Unhappiness Indicator, and it is negative related to Career Indicator.

According to our previous discussion, we can see that ReportRate is positively related to the number of potential drug users in the population. Therefore, the influence of socio-economic factors on the proportion of potential drug users α is similar to their effect on ReportRate. In Section4.1.1, the value of m is fitted, but it can be calculated using the proportion α now, based on analyzing the socio-economic data.

4.2.3 Interpretation of Drug Abuse

According to our conversation, PA2, PA3, and PA4 are positively related to ReportRate, which indicates that the majority of drug users are more likely to be young people. This is exactly the same as the actual situation[9]. Moreover, the unhappiness and stress of men and women will lead to more people choosing drugs. At the same time, PA7 is negatively correlated with ReportRate, indicating that people with low education and unemployment are more likely to use drugs. Based on the above findings, it is reasonable to deduce that the main role of drugs is to relieve stress.

The divorce caused by bachelorism or different view of love, the excessive employment pressure, personal hedonism, spiritual emptiness,etc. all contributes to the growth in opioid use and addiction. Although opioid abuse is known to be dangerous, young people have insufficient ability to resist pressure and temptation. When they suffer setbacks in love and career, they tend to choose drugs to anesthetize themselves and obtain temporary pleasure. According to the data, the physiological dependence of opioids is extremely strong. Once started to use, it is difficult to get rid of addiction, simply relying on individual will. As a result, the opioid epidemic continues to grow.

4.3 Possible Strategy to Control the Opioid Epidemic

To control the Opioid Epidemic, we should try to keep the number of SOHI steady. According to Equation 3, we can figure out the value of m when the total number of SOHI $(N(t))$ basically stays steady over time. We call the corresponding proportion of potential drug users α_0. Take the VA's data in 2016 as an example. With our model, we figured out the value of α_0, which is 0.05617195. However, the true value of α is the ratio of the number of potential drug users to the total population, and we call it α_r. In VA, 2016, the value is 0.23061803.

With Equation 11, we have obtained the ReportRate of $county(i)$, which is named of RR_i. RR of a state is defined as Equation 12, and we call it RR_r.

$$RR_r = \frac{\sum_i RR_i \cdot P_i}{\sum P_i} \tag{12}$$

where RR_i and P_i are ReportRate and total population of $county(i)$. The value of VA in 2016 is 14.84621869.

Since ReportRate is positively related to the number of potential drug users in the population, namely α. In order to keep the number of SOHI steady, we have the constraint Equation

13.

$$0 < \frac{RR'}{RR_r} < \frac{0.05617195}{0.23061803} = 0.24357137 \tag{13}$$

where RR' is ReportRate after the implementation of the strategy. Therefore, we have Equation 14.

$$\begin{aligned} RR' =& 2.049e-05 \cdot PA2 + 2.661e-05 \cdot PA3 + 2.552e-06 \cdot PA4 \\ & -6.985e-0 \cdot PA7 + 2.601e-01 < 0.24357137 RR_r = 3.61611382 \end{aligned} \tag{14}$$

Because it is very difficult to change PA3 (age structure indicator). Moreover, the effects will take a long time to show. Therefore, we treat PA3 as a constant quantity for a period of time in the future. We use x, y, z axis to represent $PA2, PA4, PA7$. On the basis that factors cannot be infinitely negative, Equation 14 appears as a bounded plane in three dimensions as Figure 15 shown. The boundary coordinate is (81754.47, -6855.70, 78005.04).

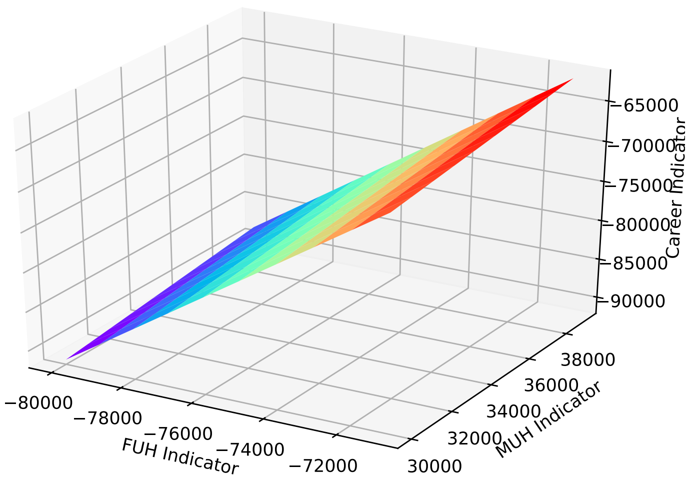


Figure 15: The threshold plane in three dimensions

Table 6: The programming result

x	y	z	objective function value
-81754.0	37149.41 5	-94081.111	-49476.52099251366

When the position of RR' is above the plane, the number of drug users can be effectively controlled under the corresponding strategy. When the position of RR' is below the plane, the opioid crisis cannot be countered well. For the sake of clarity, let's equate the cost of implementing the policy with the value changes of the indicators. Therefore, the goal is to minimize objective function, which is $PA7 - PA2 - PA4$. According to this rule, we can find the optimal strategy with the least cost. The programming result is shown in Table 6.

The results are instructive for the government. There are two aspects from which decisions can be made to effectively control the opioid crisis. One is trying to reduce the value of m, the other is reducing rhe value of q, p, which are discussed in Section4.1.1.

- m: Government can set policies that would increase Career Indicator, or decrease Female Marriage Unhappiness Indicator and Male Unhappiness Indicator. At the same time, before the government makes a decision, it is beneficial to estimate the changes in these indicators, thus estimating the changes in the potential drug users. When a number of policies are introduced, linear planning can be carried out to incorporate the impact of drugs into decision-making.For example, policies to build new institutions of higher learning in the local area and encourage local residents to receive education will significantly promote the improvement of the local population's educational level, thus improving the career factor and reducing the number of potential drug users [10].

- p, q: If the government makes a policy directly from the aspect of drug control, the effectiveness of the policy can be estimated by estimating the change of p and q.

5 Sentivity Analysis

We probe into the sensitivity of our Factor Model. We put the 2015 data into our model and obtain a predicted value. After regression, we found that the predicted value was highly linearly correlated with the true value. Since the factor itself is subject to time interference, we believe that the error is caused by the time factor. The regression result is shown in Table 7.

Table 7: The regression result of predicted value and true value

Intercept	Regression Coefficient	Multiple R-squared	P-value
0.01629	0.94629	0.08531	1.474e-10

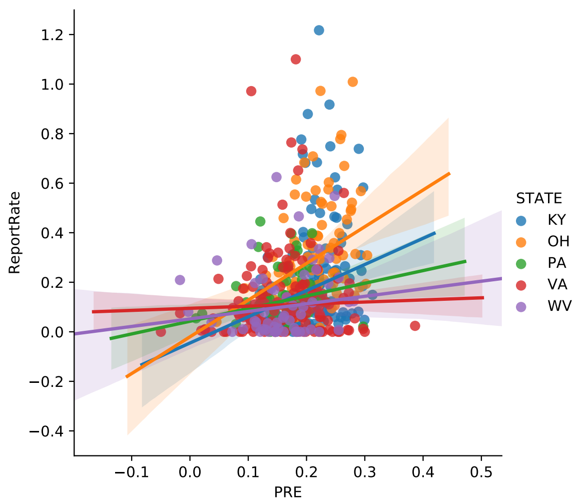


Figure 16: The prediction figure of different states in 2015

Similarly, we put the data of different states into our model and also obtained desirable regression results. Figure 16 reveals that our model is sensitive to both time and space. Hence, our model has good sensitivity.

6 Strengths and Weaknesses

6.1 Strengths

- **Efficient Data Processing:** We applied Factor Analysis Method to reduce dimensionality of data with multiple linear correlation and captureed almost all the influential factors.

- **Multi-angle Description:** We capture the essence of the opioid epidemic. The characteristics of drug transmission are considered from the perspectives of economics and biology. Therefore, our model has a high degree of applicability and comprehensiveness.
- **Good interpretability and Accuracy:** All of our models are based on the given data. We analyzed the correlation in the multidimensional data and selected the most explanatory common factors by the method of Factor Rotation. It makes our model more realistic.
- **Reliability and Stability of the Model:** According to the sensitivity analysis, the parameters of our model are sensitive to time and space, which ensures the reliability and good significance of our model.

6.2 Weaknesses

- The Bass model is suitful for predicting sales of new products. Our model may have errors when used for long-term predictions. But the model error can be reduced due to the novelty of synthetic opioids.
- The Mass Center Invariance Theorem in the propagation process is based on the assumption of uniform diffusion of drugs, so our model may have errors in predicting the drug diffusion in coastal cities.
- We did not consideration of county to county differences. We treat each county as an undifferentiated unit, ignoring the impact caused by the difference of culture, policy, geography,etc.

7 Conclusion

In this paper, we started by mapping the distribution of drugs and trying to make some connections between different counties. Then we propose a novel framework to characterize the drugs reports, analyze the similarity of the epidemic of drugs, found the range of drugs, and forecast the future evolution of drugs. Moreover, we propose a series of suggestion about how to evaluate drug policy. Finally, we conduct sensitivity analysis of some parameters in our model and discuss the strengths and weakness of our work.

References

[1] Mahajan, Vijay, Eitan Muller, and Frank M. Bass. "New Product Diffusion Models in Marketing: A Review and Directions for Research." Journal of Marketing 54.1 (1990): 1-26.

[2] Browne M W . Robustness of statistical inference in factor analysis and related models[J]. Biometrika, 1987, 74(2):375-384.

[3] Jalal H , Buchanich J M , Roberts M S , et al. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016[J]. Science, 2018, 361(6408).

[4] Sara R . How digital drug users could help to halt the US opioid epidemic[J]. Nature, 2018, 560(7718):295-297.

[5] Manchikanti L, Nd H S, Fellows B, et al. Opioid epidemic in the United States.[J]. Pain Physician, 2012, 15(3 Suppl):9-38.

[6] https://en.wikipedia.org/wiki/Bass_diffusion_model

[7] Rousseeuw P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis[J]. Journal of Computational Applied Mathematics, 1999, 20(20):53-65.

[8] https://en.wikipedia.org/wiki/Factor_analysis

[9] https://www.drugabuse.gov/drugs-abuse/opioids/opioid-overdose-crisis

[10] JohnsonK,JonesC,ComptonW,etal.FederalResponsetotheOpioidCrisis[J] .CurrentHiv/aidsReports,2018(4):1-9.