For office use only	Team Control Number	For office use only
T1 ________________	**1906623**	F1 ________________
T2 ________________	Problem Chosen	F2 ________________
T3 ________________		F3 ________________
T4 ________________	**C**	F4 ________________

2019
MCM/ICM
Summary Sheet
The Solution of the Opioid Crisis in the United States

The United States is experiencing a national crisis regarding the use of synthetic. We need to analyze the drug's transmission characteristics in five states and give a reasonable solution based on the established model.

First, based on the problem, we classify the data according to different states. And then analyze the characteristics of each state. Because opioids and heroin are addictive, and easy to rely on it again after withdrawal. It's similar to the spread of the virus. Thus we choose the Susceptible Infected Susceptible Model (SIS) and improved it. Through the model, we analyzed the changing characteristics of the number of people in each state. Due to the limitations of this model that it did not take into account changes in the total number of people throughout the state. We improve on the basis of the original and introduce constant growth rate of normal people δ . Further analysis leads to conclusions: The number of drug dependent people will gradually stabilize after several years.

Second, we need to consider the impact of socio-economic factors on the number of people dependent on drugs. So we separately counted the estimated value of each factor (HC01). Because we need to analyze the influence of all diversified factors on the independent variables and study the most influential factors, we use the multiple linear regression(MLR) analysis method. Several factors contributing to the constituency contribution rate are used for regression prediction.

Third, part3 is based on the analysis of Model 1 and Model 2. Analysis of changes in the number of drug dependence by changing the corresponding factors affecting the number of people dependent on drugs. For model two, we can make assumptions. After the government making specific measures, the parameters of the corresponding factors change, and the predicted values are calculated by the regression equation, and then compare with the original values to analyze the validity of the model.

In summary, our model completely answers the questions and proposes a reasonable solution.

Key Words: multiple linear regression(MLR), Susceptible Infected Susceptible Model (SIS), SPSS

Contents

I. Introduction

1.1 Background

1.1.1 Opioid[1]

Opioids are substances that act on opioid receptors to produce morphine-like effects. Medically they are primarily used for pain relief, including anesthesia. Other medical uses include suppression of diarrhea, replacement therapy for opioid use disorder, reversing opioid overdose, suppressing cough, suppressing opioid induced constipation. Extremely potent opioids such as carfentanil are only approved for veterinary use. Opioids are also frequently used non-medically for their euphoric effects or to prevent withdrawal.

Side effects of opioids may include itchiness, sedation, nausea, respiratory depression, constipation, and euphoria. Tolerance and dependence will develop with continuous use, requiring increasing doses and leading to a withdrawal syndrome upon abrupt discontinuation. The euphoria attracts recreational use and frequent, escalating recreational use of opioids typically results in addiction. An overdose or concurrent use with other depressant drugs commonly results in death from respiratory depression.

1.1.2 Situation

Because opioids are addictive and may result in fatal overdose, most are controlled substances. The Centers for Disease Control and Prevention (CDC) reports that deaths from prescription opioid usage has more than quadrupled since 1999 (CDC, Understanding the Epidemic, 2017). The number of prescription drug overdose is surpassing the combined number of people who overdosed during the crack cocaine epidemic of the 1980s and the black tar heroin epidemic of the 1970s, and is becoming the fastest growing drug problem in the US.

In 2011, an estimated 4 million people in the United States used opioids recreationally or were dependent on them. And it was reported by CDC that over 40 deaths every day in America because of prescription opioid usage[2]. Proof of fact, the United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational purposes (illegal, non-prescription use).

II. The Description of the Problem

2.1 Problem statement

Opioids are alkaloids and derivatives extracted from opioids (poppy) that relieve pain but are addictive for long-term use.In the United States, opioid addiction and abuse have evolved into “the most serious public health crisis”and even the media have called it the “most deadly drug epidemic” in American history. The United States is experiencing a national crisis regarding the use of synthetic and non-synthetic opioids, either for the treatment and management of pain (legal, prescription use) or for recreational purposes (illegal, non-prescription use).

In response to the above background, we need to solve the following problems:

● Using the NFLIS data provided, build a mathematical model to describe the spread and characteristics of the synthetic opioid and heroin incidents in and between the five states over time.

● Identify a drug threshold level which means the start of abnormal use. When overtake this level,what specific issue should the U.S government focus on.

● Using the model,in each of five states, identify any possible place where the specific opioid use might have started and speculate where and when it will occur in the future.

● Despite its known dangers，the opioid use still persists.Using the U.S. Census socio-economic data provided,explore whether the use of opioids reaches the current level is related to socio-economic factors in the census. And make corresponding modifications to the model in the part 1.

● Identify a possible strategy to fight with opioid crisis. Use the model from part 1 and part 2 to test the effectiveness of the strategy. Identify the significant parameters bounds which determines the success or failure of the predict model.

2.2 Analysis of Specific Issues

2.2.1 Analysis of Problem 1

As for the first problem, the volume of the NFLIS data is large and of different types.According to the distribution characteristics of the data, We first divide the NFLIS data by each state. Then in the same state from 2010 to 2017, count the ratio of drugs to the total number of people in this state . Next, establish a basic diffusion model (SIS model) based on statistical data. The model uses the year as the independent variable, and the proportion of heroin and opioids use is the dependent variable , which fits the trend of the ratio with the year. Finally, compare trends in five statesdescribe the spread and characteristics of the synthetic opioid and heroin incidents in five states.

2.2.2 Analysis of Problem 2

As for the second problem， according to the model of problem 1, the approximation of the trend of the proportion of each state can be fitted, and then the threshold of the abnormal use of the drug in the state can be obtained. Through the comparison of the thresholds, the characteristics of the spread of abnormal drug use can be analyzed, and the problems that the government should pay attention to are proposed for each state’s different characteristics.

2.2.3 Analysis of Problem 3

In terms of problem 3, we need to analyzed the diffusion characteristics of the second problem. According to the characteristics of the model, the proportion of a drug user in a state will gradually approach a value. When the proportion of drug users in the state begins to rise , the drug in this place can be considered that abnormal use has begun. We use regression analysis, combined with thresholds, to predict when opioids will occur in the future.

2.2.4 Analysis of Problem 4

In the process of processing U.S. Census socio-economic data, the main methods used are regression analysis, correlation analysis and factor analysis. The basic idea is: firstly use regression analysis, correlation analysis and other methods to study the relationship between various socio-economic factors, and then use factor analysis to extract common factors for all socio-economic factors. Next use some SPSS data processing techniques to calculate the contribution rate of each factor to the drug reports according to the extracted common factors. Several factors that have a more prominent contribution rate were selected to discuss whether the use of opioids is related to these social factors.

III. Basic assumption

1. Assume that the total number of people in the study area remains unchanged in a short period of time, and the population is divided into people who use heroin, people who use opioids, and normal populations.
2. Assumed that people who use heroin and opioid drugs are evenly mixed with the normal population, that is, they have the same influence on the surrounding people.

3. It is assumed that a drug dependent person may take the drug again after abstaining from dependence on the drug.

IV. Glossary & Symbols

4.1 Glossary

1.analgesic : pain relieving medication

2.heroin : an illegal, euphoria producing, highly addictive analgesic drug processed from morphine (a naturally occurring substance extracted from the seed pods of certain varieties of poppy plants).

3.non-synthetic opioids : a class of drugs made from extracting chemicals in opium leaves, e.g. morphine, codeine, heroin.

4.opioids : pain relieving drugs that are often highly addictive

5.socio-economic factors : factors within a society that describe the relationship between social and economic status and class such as education, income, occupation, and employment.

6.synthetic opioid : man-made opioids

7.Regression Analysis : In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships among variables.

4.2 Symbols

Symbols	***Definition***
t	The index of development
k	Transmission rate

h	Recovery cure rate
I_0	Initial drug abusers
σ	The average number of patients connect with normal each year (number of contacts)
x_i	Independent variable numbered i (social and economic factors)
$\tilde{x}$	Standardized independent variable numbered I (social and economic factors)
μ	Sample mean
s	Sample standard deviation

V. Models

5.1 Analysis and Solving of Part One

5.1.1 Model Preparation

(1) Data Processing

The amount of raw data is large, so we should first do data screening according to the integrity and usefulness of the information. First,we classify all data of NFLIS by 5 states, calculate the sum of annual drug reports and total drug reports in each state. Then use EXCEL to calculate the ratio of the two sums and make a trend line , fitting the trend line equation(The highest order of the independent variable is 5th order).

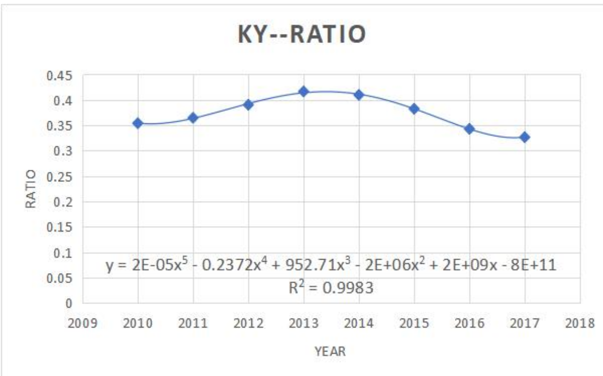


Fig1: the ratio in Kentucky

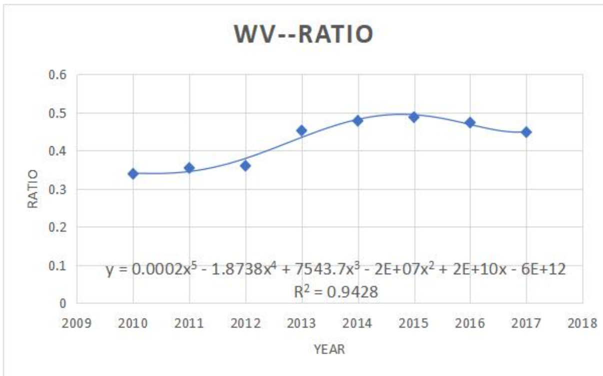


Fig2: the ratio in West Virginia

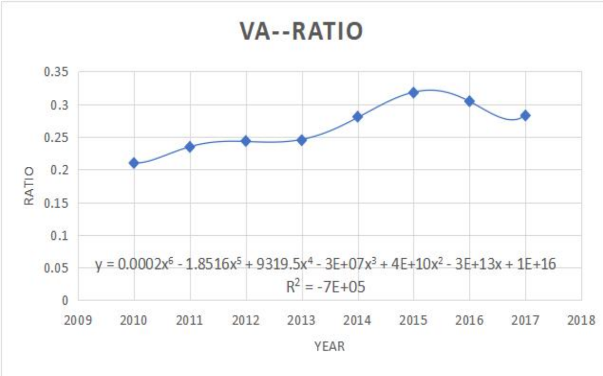


Fig3: the ratio in Virginia

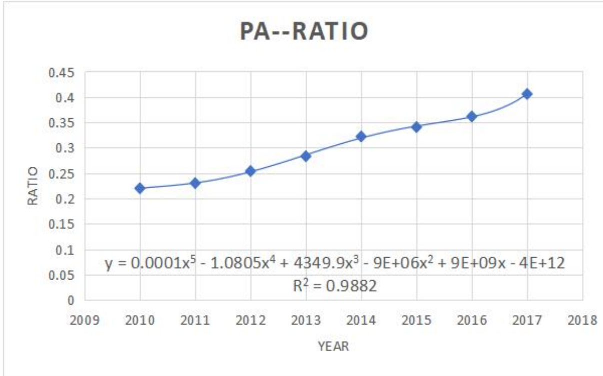


Fig4: the ratio in Pennsylvania

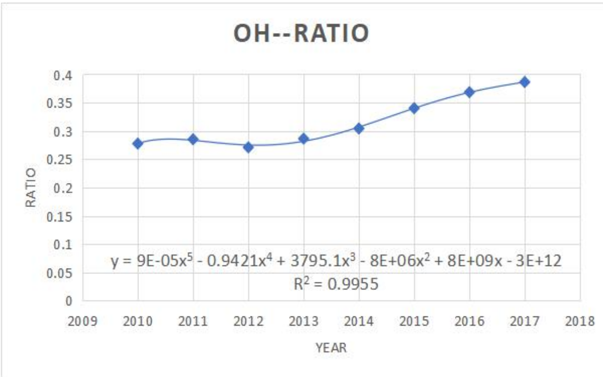


Fig5: the ratio in Ohio

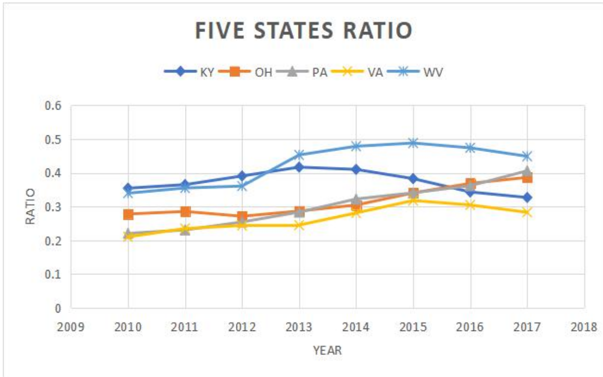


Fig6: the ratio in five states

(3) The Assumption of Model

1.Assume that the total number of people in the study area remains unchanged in a short period of time, and the population is divided into people who use heroin, people who use opioids, and normal populations.

2. Assumed that people who use heroin and opioid drugs are evenly mixed with the normal population, that is, they have the same influence on the surrounding people.

3. It is assumed that a drug dependent person may take the drug again after abstaining from dependence on the drug.

5.1.2 Model Establishment

The spread of addictive drugs relates to the number of dependents in a city, the impact of the number of people on normal people, the abstinence rate of the number of addicted people, and the relevant economic indicators of the city. In the first part we only discuss the former three factors.

Set the year to t'

$$t = t' - 2010 \tag{1}$$

At time t

$$I'(t) + S'(t) = N \tag{2}$$

$$I(t)\text{+S(t)=N} \tag{3}$$

Where $I'(t)$ is the number of opioids and heroin used in time t . Where $S'(t)$ is the number of normal people in time t . N is the total number of people. If the total number of people remains the same, normalize the population . So $I(t)$, $S(t)$ indicate the ratio of the number of people using the drug and the normal number of people to the total number of people. Due to the large number of people, it can be considered that $I(t)$ and $S(t)$ are continuously changing with respect to time.

Suppose there are reasons for the increase in the number of people using drugs during this time period $[t, t+\Delta t]$:

1.Affected by the people who had already used drugs.

2.People who quit dependence but use drugs again.

Assume that the propagation coefficient is k , release dependency rate is h , in a short time, the change of $I(t)$ obeys:

$$\begin{cases} \dfrac{dI}{dt} = kI(1-I) - hI \\ I(0) = I_0 \end{cases} \tag{4}$$

I_0 is the proportion of drug use in 2010.

The solution is obtained by the separation variable method:

$$I(t) = \begin{cases} \left[\left(\dfrac{1}{I_0} - \dfrac{1}{1-\sigma^{-1}}\right) \cdot e^{-(k-h)t} + \dfrac{1}{1-\sigma^{-1}} \right]^{-1} & k \neq h \\ \dfrac{I_0}{ktI_0 + 1} & k = h \end{cases} \tag{5}$$

Among them

$$\sigma = k / h \tag{6}$$

It indicates the average number of normal people who are contacted to the people using the opioids and heroin. Fig7(1), 7(2) is the model curve. It's easy to understand:

$$\lim_{t \to \infty} I(t) = \begin{cases} 1 - \dfrac{1}{\sigma} & \sigma > 1 \\ 0 & \sigma \leq 1 \end{cases} \tag{7}$$

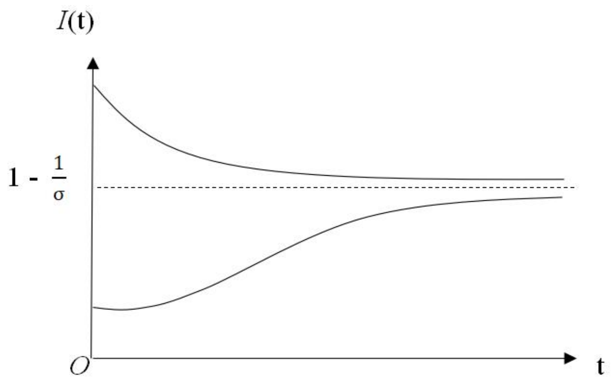


Fig7(1):the model curve when $\sigma>1$

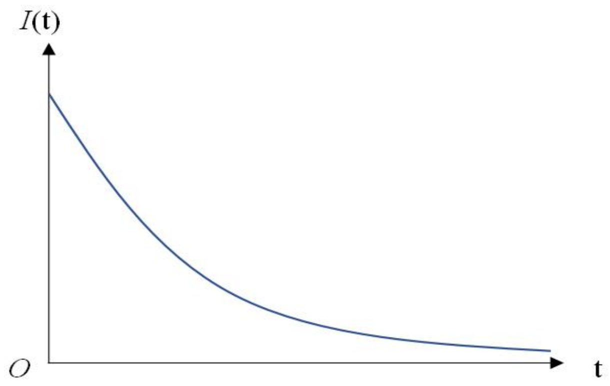


Fig7(2):the model curve when $\sigma \le 1$

As can be seen, $\sigma=1$ is a threshold. When $\sigma \le 1$, Proportion of drug use $I(t)$ gradually become lower and eventually tends to zero. This situation corresponds to, the number of normal people who are affected by drug users and become drug users is no more than the number of original drug dependence. When $\sigma>1$, the increase or decrease of $I(t)$ is depend on I_0. If $I_0 < 1-\sigma^{-1}$, $I(t)$ monotonically increases.If $I_0 > 1-\sigma^{-1}$, $I(t)$ monotonically decreases. In both cases, $I(t)$ has a non-zero limit value.

5.1.3 Model optimization

In order to make the above model more accurate, we introduced a new factor: Constant growth rate of normal people δ (including birth and immigration). Set $R(t)$ to the number of deaths and emigration. We have established the following optimized models.

$$\begin{cases} \dfrac{dS}{dt} = -kIS + \delta \\ \dfrac{dI}{dt} = kIS - hI \\ \dfrac{dR}{dt} = hI \\ S(0) = n\,, I(0) = 1\,, R(0) = 0 \end{cases} \tag{8}$$

Here $I(0)$ is different from the model above. That means at the initial moment, no one used or relied on this drug.

After simplified:

$$\begin{cases} \dfrac{dS}{dt} = -kIS + \delta \\ \dfrac{dI}{dt} = kIS - hI \\ S(0) = n\,, I(0) = 1 \end{cases} \tag{9}$$

The only equilibrium point of the equations is

$$S^* = \frac{h}{k} \qquad I^* = \frac{\delta}{h} \tag{10}$$

The first approximation system of the equations at the equilibrium point is

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{k\delta}{h} S - hI \\ \dfrac{dI}{dt} = \dfrac{k\delta}{h} S \end{cases} \tag{11}$$

In order to get the trend of the number of people in various categories, we need to study the characteristic equation of the system:

$$\det(\lambda I - A) = \begin{vmatrix} \lambda + \dfrac{k\delta}{h} & \mathrm{h} \\ -\dfrac{k\delta}{h} & \lambda \end{vmatrix} = \lambda^2 + \frac{k\delta}{h}\lambda + k\delta = 0 \tag{12}$$

Where A is the coefficient matrix of the system.

Here, we introduce ***Routh －Hurwitz stability criterion***[4]: In control system theory, the Routh－Hurwitz stability criterion is a mathematical test that is a necessary and sufficient condition for the stability of a linear time invariant (LTI) control system. The importance of the criterion is that the roots p of the characteristic equation of a linear system with negative real parts represent solutions e^{pt} of the system that are stable (bounded). Thus the criterion provides a way to determine if the equations of motion of a linear system have only stable solutions, without solving the system directly. In the real number system quadratic equation, $\lambda^2 + a_1\lambda + a_2 = 0$. The necessary and sufficient condition for both roots to have a negative real part is $a_1 > 0, a_2 > 0$.

Based on ***Routh －Hurwitz stability criterion***, We can judge that the real parts of both roots are negative, which is $Re\,\lambda_{1,2} < 0$. Therefore, (S^*, I^*) is gradually stable.

It can be seen that in the case of stable population growth δ, no matter what value k and $R(t)$ take, the number of drug dependence increases to a certain extent and tends to be stable. Because $t = 0$ in the equation represents the initial state(at the initial moment, no one used or relied on the drug) , instead of 2010. So we think that the number of drug dependent people will eventually stabilize from 2010 to 2016.

5.1.4 Results

According to relevant information[6], the value of the transmission rate k is approximately 0.19%.

Suppose the value of h is approximately 0.06%. So we estimate the threshold(σ) $2.63>1$.

Compare I_0 and $1-\frac{1}{\sigma}$ in five states. we can see their characteristics and the analysis results are as follows:

Table 1

state	I_0	characteristics (Proportion of opioid users)	The time it takes to reach the steady state from 2010
Kentucky	0.353	Generally stay still	1.003
Ohio	0.277	↑	18
Pennsylvania	0.220	↑	24.57
Virginia	0.209	↑	25.571
West Virginia	0.333	↑	7.38

5.1.5 Analysis of the Result

According to the results shown in the chart, it can be concluded that the number of opioids users in Kentucky in the five states remains basically the same and will reach the threshold after 1.003 year. The other four states maintain this rising trend. But the rate of growth is different. The Ohio state reached the threshold after 18 years, and the Pennsylvania reached the threshold after 24.57 years. Virginia states reach a threshold after 25.571, and West Virginia reaches a threshold after 7.38 years.

5.2 Analysis and Solving of Part Two

5.2.1 Model Preparation

(1) Data Processing

In the given ACS data, the number of people of specific county corresponding to all socioe-conomic factors from 2010 to 2016. We need to explore whether the use of opioids reaches the current level is related to socio-economic factors in the census. So we need to map all the economic factors of the specific county to the use of opioids in the county in order to analyze them. The following are the details of the step.

- **Step1:** Classify ACS data for each year by state.
- **Step2:** In statistics, missing data occur when no data value is stored for the variable in an observation. Missing data are a common occurrence and can have a significant effect on the conclusions that can be drawn from the data. So we delete the specific factor which is no data. If the

percentage of missing data exceeds 50%, the imputation will result in great error, which we treat as a threshold for missing data[3].

- **Step3:** Add the number of people of specific factors for each county from 2010 to 2016.
- **Step4:** Use statistical tools SPSS, import all socio-economic factors, set the "total drug reports county" as dependent variable. Analyze which the factors contribute the most to the dependent variable and the most obvious impact.

(2) The Foundation of Model

Set the dependent variable is y, independent variable is $x_1, x_2, \ldots\ldots, x_n$. Assume that n sets of independent data $(y_i, x_{i1}, x_{i2}, \ldots\ldots, x_{ip})$, $i = 1,2,3\ldots\ldots, n$ have been obtained. And set the following linear relationship between them:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots\ldots + \beta_p x_{ip} + \varepsilon_i \, , i = 1,2,\ldots\ldots, n \tag{8}$$

Among them, ε_{i} are random variables. They are independent of each other and fit this relationship: $E(\varepsilon) = 0$, $D(\varepsilon) = \sigma^2$. The above relationship are called "***Multiple linear regression model***".

$\beta_{\mathrm{j}}(j = 0,1,2,\ldots\ldots, p)$ is called "***Regression coefficient***".

Multiple linear regression estimates the estimation $\hat{\beta}_j$ of regression coefficient β_{j} by least squares. Thus obtain the following regression equation :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots\ldots + \hat{\beta}_p x_{ip} \, , i = 1,2,\ldots\ldots, n \tag{9}$$

5.2.2 Model Establishment

To quantitatively describe the impact of each state's indicators on the use of opioids and heroin, we use the Stepwise method to construct the MLR models for each state and find the relevant indices for each state for each relevant factor. Multiple linear regression analysis was performed on each factor using statistical tool spss. However, due to the large number of counties and the limitation of competition time, in our model, the analysis will be conducted in ***Kentucky***.

Let each factor be an independent variable x_i, $i = 1,2,\ldots\ldots, m$. Let the total drug reports in each county be the independent variable y_j , $j = 1,2,\ldots\ldots, n$. That also means there are m factors and n counties. Observed data of independent variables are recorded as matrix :

$$A=(a_{ij})_{n\times m}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \tag{10}$$

The dependent variable observation data is recorded as a matrix :

$$B=\begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}^T \tag{11}$$

Convert each indicator a_{ij} to a standardized indicator $\tilde{a}_{ij}$:

$$\tilde{a}_{ij}=\frac{a_{ij}-\mu_j}{s_j}\ ,\ i=1,2,\cdots n; j=1,2,\cdots m; \tag{12}$$

Among them

$$\mu_j=\frac{1}{n}\sum_{i=1}^{n}a_{ij};\ s_j=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a_{ij}-\mu_j)^2} \tag{13}$$

μ_i, s_i is the sample mean and variance of independent variable $x_{\rm i}$

Similarly,convert each indicator b_i to a standardized indicator $\tilde{b}_i$:

$$\tilde{b}_i=\frac{b_i-\mu}{s}\ ,\ i=1,2,\cdots n; \tag{14}$$

Among them

$$\mu=\frac{1}{n}\sum_{i=1}^{n}b_i;\ s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(b_i-\mu)^2} \tag{15}$$

μ, s is the sample mean and variance of independent variable y.

Model information:

Coefficieents

	Unstandardized B	Coefficients Std.Error	Standardized Coefficients Beta	t	Sig

(constant)	-83.956	56.906		-1.475	0.143
HC01_VC15	0.191	0.031	0.574	6.097	0.000
HC01_VC194	0.045	0.006	0.703	7.733	0.000
HC01_VC78	-0.061	0.018	-0.536	-3.379	0.001
HC01_VC123	0.042	0.010	0.292	4.267	0.000
HC01_VC132	-0.010	0.003	-0.558	-3.487	0.001
HC01_VC69	0.715	0.115	0.553	6.226	0.000

Use statistical tools SPSS, calculate and select the largest model of R, R Square, and Adjusted R Square in all models.

Model Summary[i]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics: R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.980[a]	.961	.961	572.537	.961	2937.372	1	118	.000	
2	.983[b]	.966	.965	541.341	.004	14.992	1	117	.000	
3	.984[c]	.968	.968	522.603	.003	9.540	1	116	.003	
4	.986[d]	.972	.971	496.631	.003	13.450	1	115	.000	
5	.987[e]	.974	.973	479.022	.002	9.610	1	114	.002	
6	.988[f]	.976	.974	464.781	.002	8.093	1	113	.005	
7	.988[g]	.977	.975	454.868	.001	5.979	1	112	.016	
8	.988[h]	.977	.976	453.811	.000	.475	1	112	.492	2.235

a. Predictors: (Constant), HC01_VC61
b. Predictors: (Constant), HC01_VC61, HC01_VC15
c. Predictors: (Constant), HC01_VC61, HC01_VC15, HC01_VC194
d. Predictors: (Constant), HC01_VC61, HC01_VC15, HC01_VC194, HC01_VC78
e. Predictors: (Constant), HC01_VC61, HC01_VC15, HC01_VC194, HC01_VC78, HC01_VC123
f. Predictors: (Constant), HC01_VC61, HC01_VC15, HC01_VC194, HC01_VC78, HC01_VC123, HC01_VC132
g. Predictors: (Constant), HC01_VC61, HC01_VC15, HC01_VC194, HC01_VC78, HC01_VC123, HC01_VC132, HC01_VC69
h. Predictors: (Constant), HC01_VC15, HC01_VC194, HC01_VC78, HC01_VC123, HC01_VC132, HC01_VC69
i. Dependent Variable: TotalDrugReportsCounty

Fig8:the result of SPSS

After processing and analysis, we chose model 8.

5.2.3 Results

After processing, the regression equation between the standardized factor variables is obtained:

$$\tilde{y} = 0.574\tilde{x}_{15} + 0.703\tilde{x}_{194} - 0.536\tilde{x}_{78} - 0.292\tilde{x}_{123} - 0.558\tilde{x}_{132} + 0.553\tilde{x}_{69} \tag{16}$$

The regression equation for reducing the normalized variables to the original variables is:

$$y = -83.956 + 0.191x_{15} + 0.045x_{194} - 0.061x_{78} + 0.042x_{123} - 0.010x_{132} + 0.715x_{69} \tag{17}$$

According to the results of the model analysis, among the hundreds of basic information (independent variables) surveyed in Kentucky, *the number of single-person households over 65 years old (non-family households)*, *language mastery*, *enrolled population aged three years and older*,*the number of people who have lived for more than one year in different counties* and *the the number of people whose birthplace is in the United States* are the major factors in the change in the number of opioids and heroin use.

5.2.4 Analysis of the Result

This multivariate linear relationship can be used to predict the number of opioids and heroin use in a given year. The 2016 statistics are used as predictive data, and the 2016 relevant variable index can be brought into the regression equation to obtain the 2016 forecast: 147462, the measured value is 163305, the relative error is 9.701%. Similarly, the above model is established, and multiple linear regression models of other four states can be established. The equation is as follows:

$$
\begin{aligned}
&y_{\mathrm{OH}}=134.664+0.023x_{191}-0.170x_{188}+0.326x_{38}-0.641x_{96}-0.095x_{87}+0.037x_{134}\\
&-0.263x_{85}+0.674x_{207}-0.726x_{207}-0.199x_{187}\\
&y_{VA}=134.664+0.037x_{118}-0.113x_{204}-0.282x_{76}+0.042x_{85}-0.254x_{197}+1.216x_{96}\\
&y_{WV}=-199.983+0.135x_{191}+0.745x_{191}-0.094x_{96}\\
&y_{PA}=-652.823+1.956x_{04}-0.49x_{183}-1.66x_{133}+0.147x_{147}-0.211x_{92}+0.030x_{96}
\end{aligned}
\tag{18}
$$

The biggest factor influencing in each of the states is the school enrollment and education.

5.3 Analysis and Solving of Part Three

5.3.1 Model Preperation

During 2010-2017, "Drug Reports" ratio of each kind of drugs of the five major states is showed below.

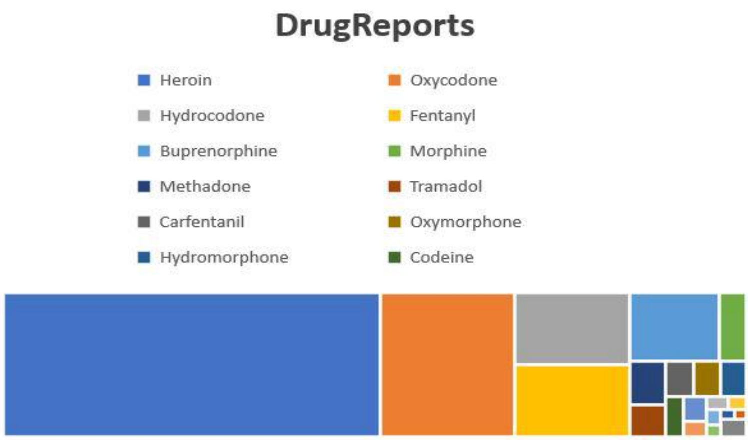
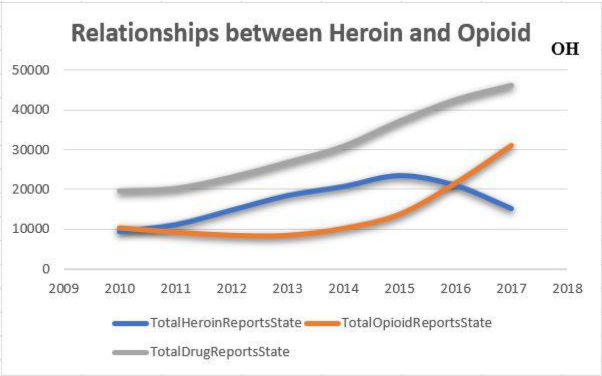


Fig9: Drug Reports ratio chart

It can be seen from the figure that the number of people taking heroin is a large proportion. Among all types of Drug, heroin is an illegal, highly addictive opioid. The relationship between the number of heroin and opioid A in each state is shown in the figure below:



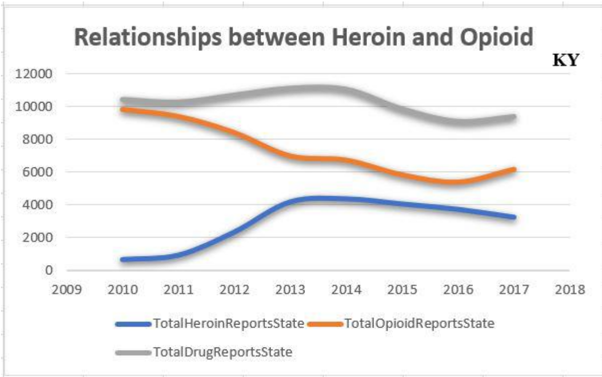
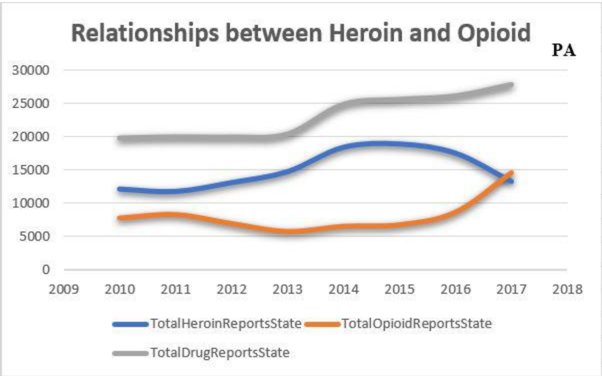


Fig10(1):the relationship in Ohio Fig10(2):the relationship in Kentucky



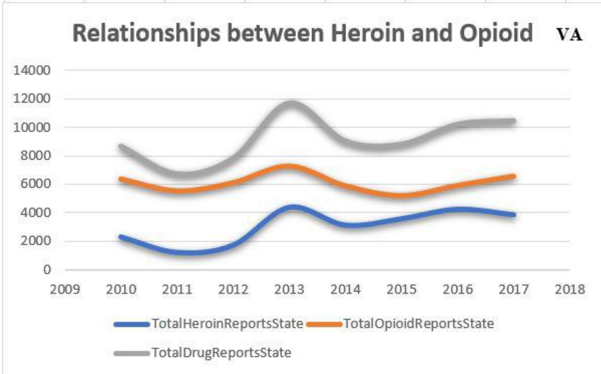


Fig10(3): the relationship in Pennsylvania Fig10(4):the relationship in Virginia

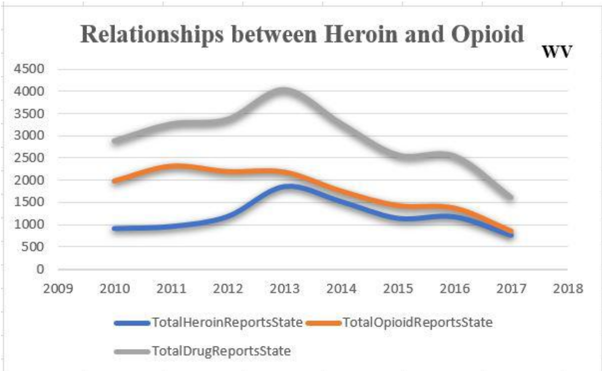


Fig10(5):the relationship in West Virginia

As can be seen from the figure, there is a close relationship between opioids and heroin dependence. And for governments and organizations, opioid prescription drugs are easier to control than illegal opioids(such as heroin). Therefore, based on the models established in Part 1, Part 2, we propose the following suggestions on how to reduce the abuse of opioid analgesics:

- Using Model 1 and through use of existing **surveillance systems**[5] (such as *medical examiner data*, *emergency medical services data*, *near-real time emergency department data*, and *poison center data*) to predict the timing and location of outbreaks of opioid overdose in the United States. Identify, intervene and prevent the drug abuse and metastasis.

- Collect data on socioeconomic factors in different regions, with the help of model 2, find risk factors by analyzing the extent of excessive intake of opioids and the correlation of socioeconomic factors.And use government forces to intervene in time.

- Strengthen the control of prescription drugs in the region and limit the distribution and use of opioids in different time periods.

5.3.2 Results

For the socio-economic factors of Model 2, it can be assumed that the government has intervened on some of the most important indicators. We can modify the values of relevant parameters and predict the number of opioids and heroin used in a certain state in a certain year.

Take Kentucky as an example, the reason why non-family households over 65 years old in 2016 will become an important factor may be related to the physical condition of the elderly. The incidence of cancer in the elderly has increased, so the number of people using opioids has increased. The government can indirectly reduce the number of elderly people with cancer by introducing relevant policies to protect the lives of non-family elderly people.

Assume that after the introduction of the policy, the number of elder non-family households has decreased by 20. Similarly, assuming a 10% increase in the number of people enrolled in school over the age of three, controlling the number of migrants indirectly increases the number of births in the United States by 10%, using Kentucky's regression equation to find its value:

$$y' = -83.956 + 0.191x'_{15} + 0.045x'_{194} - 0.061x'_{78} + 0.042x'_{123} - 0.010x'_{132} + 0.715x'_{69} = 126545 \tag{19}$$

$$x'_{15} = (1-20\%)x_{15}, x'_{194=}x_{194}, x'_{78} = (1+10\%)x_{78}, x'_{123} = x_{123}, x'_{132} = (1+10\%)x_{132}, x'_{69} = x_{69} \tag{20}$$

It can be analyzed that the number of people using drugs after taking relevant measures has decreased by 12.337% compared with the previous forecast. Theoretically, it has certain benefits, but it remains to be seen before implementation.

VI. Error Analysis and Model Checking

6.1 Model One

It is possible to fit the trend graphs of the opioids and heroin use ratios in the five states in the future, and compare the results expected from our model to see if the effect reaches the approaching value in the corresponding time.

6.2 Model Two

In order to more intuitively and quickly observe the marginal effect of each independent variable in interpretation, a residual analysis histogram can be drawn, as shown in Figure 11:

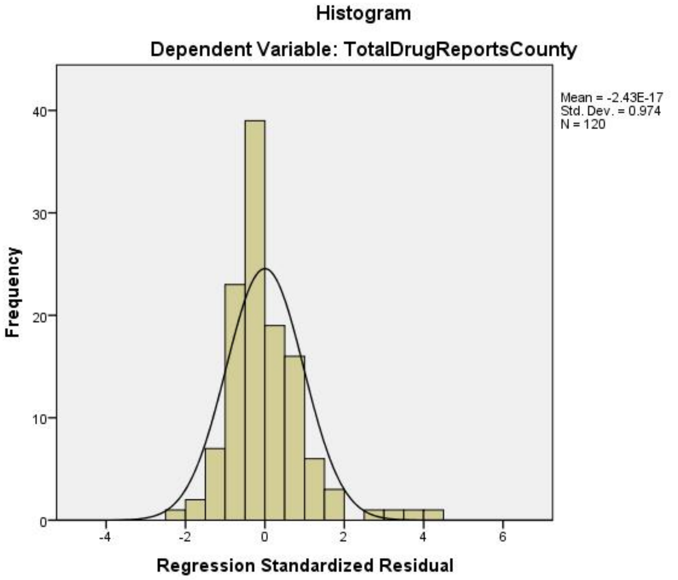


Fig11:Dependent Variable of Total Drug Reports County

Comparing the residual analysis histogram with the normal distribution curve distributed above, it can be considered that the residual of the regression equation roughly obeys the normal distribution, and the regression hypothesis that the residual distribution is normally distributed is satisfied.

In order to examine the model accuracy of this regression equation, $(\hat{y}_i, y_i)$ can be used as a coordinate value. And draw a prediction map for all sample points. $\hat{y}_i$ is the predicted value at the ith sample point.In this prediction graph, if all the graphs can be evenly distributed around the diagonal of the graph, the fitting value of the equation is very different from the original value, and the fitting effect of this equation is satisfactory.

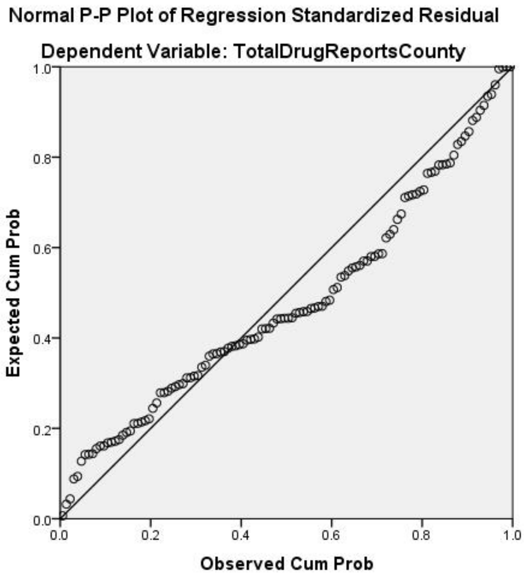


Fig12:P-P plot in Kentucky

VII. Evaluation of Model

7.1 Strength and Weakness

7.1.1 Strength

1. Model 1 uses a diffusion model of infectious diseases, especially the re-addictiveness of drugs, which can determine the number of drug users to reach a stable value over time.
2. The improved infection model can explain the spread characteristics of some drugs or infectious viruses over a long period of time.
3. Principal component analysis, canonical correlation analysis and linear regression analysis are used in Model 2, so in addition to providing a reasonable model in the analysis results, some related content similar to principal component analysis and canonical correlation analysis can be completed.
4. Using multivariate linear regression models, through multiple sets of data, the linear relationship between the three can be analyzed intuitively and quickly. Regression analysis can improve the correlation between the various factors and the degree of fitting, and improve the prediction equation.

7.1.2 Weakness:

1.Model 1 may have a certain deviation for analysis in a short period of time.
2.Model 2 may ignore interaction effects and nonlinear causality.

VIII. Memo

Dear Chief Administrator,

Upon hearing that the United States is experiencing a national crisis regarding the use of synthetic, my fellow colleagues and I are very concerned about this issue. We seek to mitigate the crisis by establishing mathematical models , processing and analyzing the data given. This letter will analyze the drug's transmission characteristics in five states and give our reasonable solution based on the established model.

As we all know, opioids and heroin are addictive, and easy to rely on them again after withdrawal. It’s similar to the spread of the virus. Thus we choose the diffusion model (SIS) and improved it. Though the model, we found that the number of drug dependent people will gradually stabilize after several years.

Obviously, the number of people dependent on drugs has a close relationship with socio-economic factors. Using the multiple linear regression(MLR) analysis method, we found the risk factors for five states(just use ***Kentucky*** as an example):

	Unstandardized B	Coefficients Std.Error	Standardized Coefficients Beta	t	Sig
(constant)	-83.956	56.906		-1.475	0.143
HC01_VC15	0.191	0.031	0.574	6.097	0.000
HC01_VC194	0.045	0.006	0.703	7.733	0.000
HC01_VC78	-0.061	0.018	-0.536	-3.379	0.001
HC01_VC123	0.042	0.010	0.292	4.267	0.000
HC01_VC132	-0.010	0.003	-0.558	-3.487	0.001
HC01_VC69	0.715	0.115	0.553	6.226	0.000

Based on the models established, we propose the following suggestions on how to reduce the abuse of opioid analgesics:

- Through use of existing **surveillance systems**[5] (such as *medical examiner data*, *emergency medical services data*, *near-real time emergency department data*, and *poison center data*) to predict the timing and location of outbreaks of opioid overdose in the United States. Identify, intervene and prevent the drug abuse and metastasis.

- Collect data on socioeconomic factors in different regions, with the help of our model, find risk factors by analyzing the extent of excessive intake of opioids and the correlation of socioeconomic factors. And use government forces to intervene in time.

- Strengthen the control of prescription drugs in the region and limit the distribution and use of opioids in different time periods.

To conclude, we hope that our studies will be helpful for you to make the strategy of fighting with the crisis of heroin and opioids. We sincerely hope that with our research and your efforts, the day of the eradication of crisis of heroin and opioids to come as soon as possible!

IX. References

[1] Wiki-opioid https://en.wikipedia.org/wiki/Opioid#Society_and_culture
[2] CDC https://www.cdc.gov/
[3] Wu Sen, Feng Xiaodong, Shan Zhiguang. Missing Data Imputation Approach Based on Imcomplete Data Clustering. Chinese Journal of Computers,Vol.35(2012)
[4] https://en.wikipedia.org/wiki/Routh%E2%80%93Hurwitz_stability_criterion
[5] https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6426a3.htm?s_cid=mm6426a3_w
[6] https://emergency.cdc.gov/han/han00384.asp