

For office use only

Team Control Number

For office use only

T1 _____

1919828

F1 _____

T2 _____

F2 _____

T3 _____

Problem Chosen

F3 _____

T4 _____

C

F4 _____

2019

MCM/ICM

Summary Sheet

The L^AT_EX Template for MCM Version v6.2.1

Summary

America is suffering from a national opioid crisis, bringing huge impact on the health and safety of society. We attempted to build a Time-Space model to describe the opioid usage and transportation traits in America, and giving advice to the government on drug control.

First, we build a relationship graph containing every county in these 5 states, consider the autoregressive internal factor and nearby counties' external affecting factor of each county, and build the equation that characterizes the variation of the opioid incident cases. We get the rule describing drug crime cases count by analyzing the equations, and get the source of drug propagation by the minimum spanning tree.

Then, we clarify the internal floating factor by considering the population and economic investigation report. We classify various factors into three major types, and determine their relevance and significance by AHP. Subsuming these economic and social factors into the equations, we get a much better fitting model and less deviations.

Finally, we successfully describe national drug criminals traits with the modified model, and identify regional drug centers by the degree of centrality in the graph. We provide advice to the government by referring to the drug measures in other countries and regions.

Overall, our model adapts well in the problem of describing the the U.S. opioid crisis, and is able to give out a practical solution. Besides, we believe that this model can also be applied to those scenarios of interactions and overall analysis of local areas in other disciplines, such as sales pricing in different regions, news dissemination and public opinion.

Keywords: Time-Space Model, Drug Control, AHP, Centrality Degree, Minimum Spanning Tree

The L^AT_EX Template for MCM Version v6.2.1

January 28, 2019

Summary

America is suffering from a national opioid crisis, bringing huge impact on the health and safety of society. We attempted to build a Time-Space model to describe the opioid usage and transportation traits in America, and giving advice to the government on drug control.

First, we build a relationship graph containing every county in these 5 states, consider the autoregressive internal factor and nearby counties' external affecting factor of each county, and build the equation that characterizes the variation of the opioid incident cases. We get the rule describing drug crime cases count by analyzing the equations, and get the source of drug propagation by the minimum spanning tree.

Then, we clarify the internal floating factor by considering the population and economic investigation report. We classify various factors into three major types, and determine their relevance and significance by AHP. Subsuming these economic and social factors into the equations, we get a much better fitting model and less deviations.

Finally, we successfully describe national drug criminals traits with the modified model, and identify regional drug centers by the degree of centrality in the graph. We provide advice to the government by referring to the drug measures in other countries and regions.

Overall, our model adapts well in the problem of describing the U.S. opioid crisis, and is able to give out a practical solution. Besides, we believe that this model can also be applied to those scenarios of interactions and overall analysis of local areas in other disciplines, such as sales pricing in different regions, news dissemination and public opinion.

Keywords: Time-Space Model, Drug Control, AHP, Centrality Degree, Minimum Spanning Tree

Contents

| | | |
|----------|---|----------|
| 1 | Introduction | 1 |
| 1.1 | Background | 1 |
| 1.2 | Our Work | 1 |
| 2 | Assumptions | 2 |
| 3 | Related Knowledge | 3 |
| 4 | Data Analysis | 4 |
| 4.1 | Data Origin | 4 |
| 4.2 | Obvious Factors | 4 |
| 4.3 | Indirect Factors | 4 |
| 4.4 | Resulting Dataset | 5 |
| 4.5 | Error Handling | 5 |
| 4.6 | Spreading Index | 5 |
| 4.7 | Considering Eco-Social Factors | 5 |
| 5 | Model Construction and Simulation Analysis | 6 |
| 5.1 | Problem 1 | 6 |
| 5.1.1 | Construction Process | 6 |
| 5.1.2 | Model Simplification | 7 |
| 5.1.3 | Problem Solving | 9 |
| 5.1.4 | Sensitivity Analysis | 9 |
| 5.2 | Problem 2 | 9 |
| 5.2.1 | Construction Process | 9 |
| 5.3 | Problem 3 | 11 |
| 5.3.1 | Problem Solving | 11 |
| 5.3.2 | Evaluating | 11 |
| 5.4 | Strengths | 11 |
| 5.5 | Weaknesses | 12 |

| | |
|-----------------------------------|-----------|
| 6 Conclusions | 12 |
| Appendices | 15 |
| Appendix A First appendix | 15 |
| Appendix B Second appendix | 15 |

1 Introduction

1.1 Background

About 275 million people worldwide, which is roughly 5.6 per cent of the global population aged 15-64 years, used drugs at least once during 2016. Some 31 million of people who use drugs suffer from drug use disorders, meaning that their drug use is harmful to the point where they may need treatment. Initial estimations suggest that, globally, 13.8 million young people aged 15-16 years used cannabis in the past year, equivalent to a rate of 5.6 per cent. Roughly 450,000 people died as a result of drug use in 2015, according to WHO. Of those deaths, 167,750 were directly associated with drug use disorders (mainly overdoses). The rest were indirectly attributable to drug use and included deaths related to HIV and hepatitis C acquired through unsafe injecting practices.

Opioids continued to cause the most harm, accounting for 76 per cent of deaths where drug use disorders were implicated. PWID some 10.6 million worldwide in 2016 endure the greatest health risks. More than half of them live with hepatitis C, and one in eight live with HIV. The headline figures for drug users have changed little in recent years, but this stability masks the striking ongoing changes in drug markets. Drugs such as heroin and cocaine that have been available for a long time increasingly coexist with NPS and there has been an increase in the non-medical use of prescription drugs (either diverted from licit channels or illicitly manufactured). The use of substances of unclear origin supplied through illicit channels that are sold as purported medicines but are destined for non-medical use is also on the increase. The range of substances and combinations available to users has never been wider.

In 2015 and 2016, for the first time in half a century, life expectancy in the United States of America declined for two consecutive years. A key factor was the increase in unintentional injuries, which includes overdose deaths. In 2016, 63,632 people died from a drug overdose in the United States, the highest number on record and a 21 per cent increase from the previous year. This was largely due to a rise in deaths associated with pharmaceutical opioids, including fentanyl and fentanyl analogues. This group of opioids, excluding methadone, was implicated in 19,413 deaths in the country, more than double the number in 2015. It is necessary for us to study the law of drug spread and take corresponding measures to curb the trend of drug spread.[1]

1.2 Our Work

First, we consider the influence of time and space on the number of drug use cases in counties, and combine the knowledge of graph theory to construct Time-Space Equation. We study the spread of drug use, identify the locations where specific drug use started, give the drug identification threshold levels at

what drug use will become harder to control, and test the effectiveness of our model.

Secondly, we take the influence of some important factors into consideration, improve our model, construct Time-Space & Social-Economic Equation(TSSE Equation) and compare the fitting result with the model before.

Finally, we come up with some solutions for the spread of drugs, test their effectiveness, and write a memo to the Chief Administrator, DEA/NFLIS Database. We use Figure [1] as a flowchart for our work.

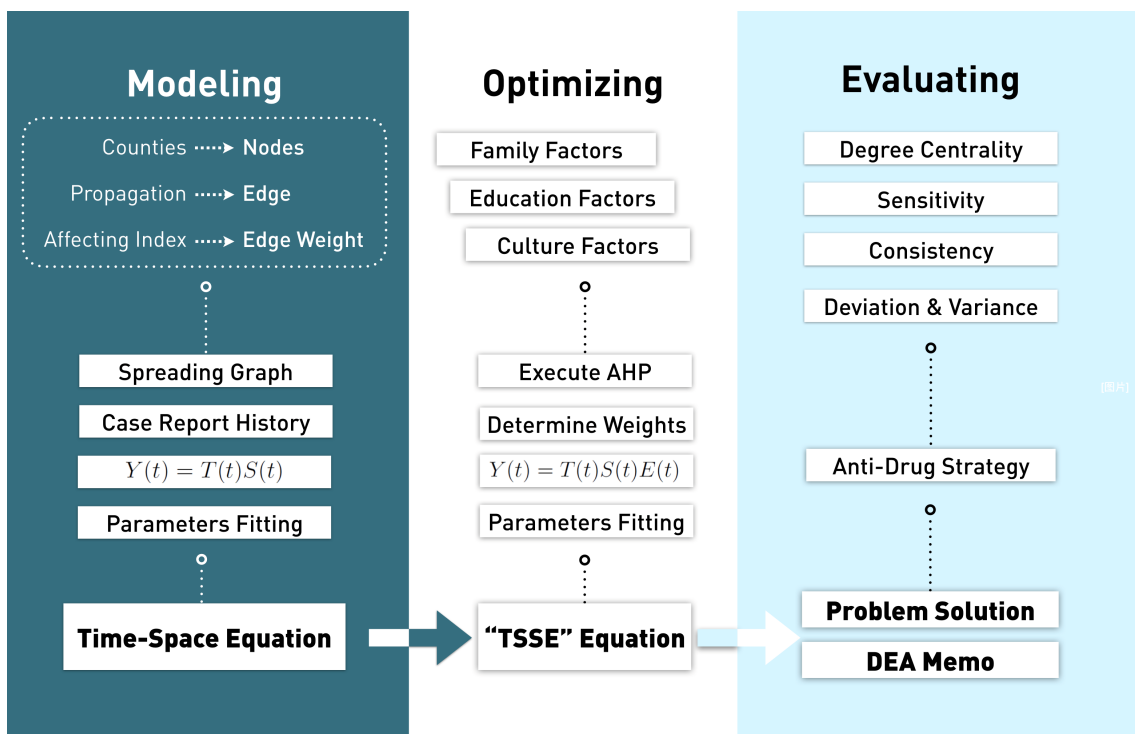


Figure 1: The Flow Chart of the Modeling

2 Assumptions

We make the following basic assumptions in order to simplify the problem. Each of our assumptions is justified and is consistent with the basic fact.

- **The reported counts contains all the cases of drug use in the states and counties.** There is no unreported drug use case and the report will not lead to a reduction in drug use.
- **The total population of the states and counties remains essentially stable in these years.** The effect of population density on drug use cases is constant.

- **The rest of America and the rest of the world have a constant impact on the five states.** We assume that the external influence on these five states remains the same.
- **States and counties have stable policies on drug use.** We assume that policy does not change during the study period.

3 Related Knowledge

Graph Theory

In **Part 1**, we need to describe the spread of the reported cases and identify possible locations where specific opioid use started. From article[2], we learn that weighted directed graph can solve this problem well. The advantage of using graph theory analysis is that it can directly reflect that interaction between the counties and it helps to analyze the path of transmission of certain drugs. Based on the linkages between the counties and counties, it is easier to derive strategies for controlling drug spread.

The following is the weighted directed graph we constructed. In this diagram [2], each node represents a county and the directed edge $W_{C,D}$ between two nodes C and D represents the influence coefficient of county C on county D is $W_{C,D}$ (The larger the influence coefficient, the more likely the county to be in county c, and the more likely the drug use in county D is to be transmitted from county C). Thus, a graph is constructed for all counties in the five states.

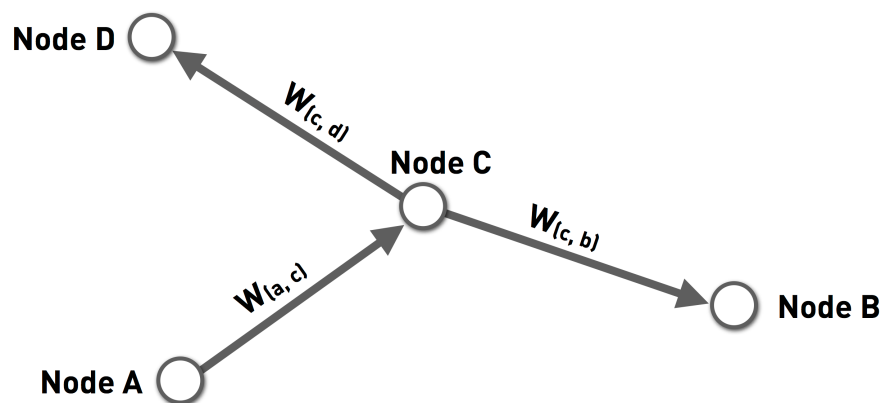


Figure 2: An Example of the Weighted Directed Graph

4 Data Analysis

4.1 Data Origin

The major data source is the “MCM_NFLIS_Data.xlsx” file. It contains all incidents involved with narcotic analgesics and heroin occurring from 2010 to 2017. Hopefully it can figure out the drug crisis spreading amount the northeast U.S.

4.2 Obvious Factors

Some factors are provided directly and can be grabbed at the first time:

- **Substances of Drugs**

The main substances of all incidents were provided, hence we may separately analyze them to validate the model.

- **Case Count**

Many drugs involved in all these incidents, but they are not equally significant. The reported count recording the troubles they made can be a convincing factor to determine how influential one drug is.

- **Geography Location**

The specific county name in OH, KY, WV, VA, and PA are provided in detail. Thanks to the data on the Internet[3], we could get accurate longitude and latitude to locate a single county. That would be a great help in visualizing our analysis.

4.3 Indirect Factors

There are several important factors that affects the modeling procedure but not provided directly.

- **Case History**

There are accurate data recording how many cases involving with drugs occurred in each county since 2010 until 2017. Putting individual case counts altogether would allow them make more sense.

- **Neighbor Counties**

Besides grabbing the location of each county, their several nearest neighbor counties and their case history is also recorded to build a drug crisis spreading model.

- **Population Density** The annual population of each county is also provided. Combining them with the area information of each county, we can divide each county into several approximate density levels. That would be helpful in solving Problem 2.

4.4 Resulting Dataset

- 451 Unique Counties
- 69 Kinds of Substances
- 24062 Case Recordings

4.5 Error Handling

Thanks to the high quality database, there is no much difficulty finding out informations above. But it should be noticed that some names of counties are shared by two different states (for example, Adams County in Ohio and Adams County in Pennsylvania). To avoid conflicting analyzing result, both state name and county name should be used to identify a county. Hence the county data can be ensured unique.

4.6 Spreading Index

To describe the spreading process quantitatively, opioid and heroin case report data during these years and the location data of all counties are introduced to fit the equation mentioned above. According to general intuition, two counties at long distance might not lead to major spreadings. So for one specific county, only those data from the nearest 'k' counties can be used in model fitting. Considering the fitting efficiency, 'k' should be no more than 10. The fitting result should go beyond a threshold value to make sense. Any lower spreading index should be treated as disturbance and ignored. By judging the value of spreading index, a weighted directed graph can be constructed to analyze the origin of any single kind of drugs.

4.7 Considering Eco-Social Factors

In Problem 1 solving phase, only opioid and heroin related case data are applied in model building, but the total amount of drug abusing cases are provided separately. Hence those spreading factor above cant be directly used here, and another spreading index fitting for general drug abusing cases is necessary. Performing AHP needs data series of various factors. Eco-Social Factor Data in 2017 were not provided, so the fitting would be limited between 2010 and 2016

under this phase. In the modeling part of Problem 1 solution, we have determined a best 'k' value that optimizes the fitting procedure. It is assumed that the same 'k' should also provide the optimized result, so as to reduce excessive fitting calculations.

5 Model Construction and Simulation Analysis

5.1 Problem 1

5.1.1 Construction Process

We analyze and process the data of "MCM_NFLIS_Data.xlsx" and establish databases and a flow chart[3] for the total number of synthetic opioid and heroin cases in 451 counties of the five states from 2010 to 2017.

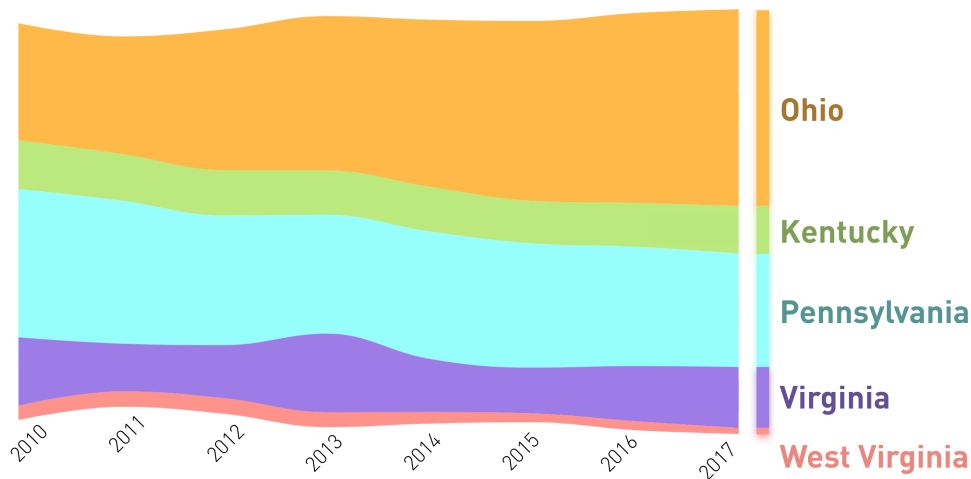


Figure 3: Total Drug Use In Five States in Years 2010-2017

Through the overall analysis of the flow chart, we find that the number of cases has a tendency to rise slowly. We take the following steps to build this model. First of all, we think that the change in the number of cases in a county $Y(t)$ depends on two factors: time $T(t)$ and space $S(t)$ and these two factors have equally important influence on the change of the case number. Based on this, the space-time model of $Y(t) = T(t)S(t)$ type for the county case number is established:

1. **The impact of the number of local cases in the past on the present.** According to the knowledge of time series, we need $T(t)$ to reflect the trend and periodicity of case number changes. Therefore, $T(t)$ should be a component that fluctuates slowly with time. We assume that $T(t) = C_{base}e^{\sin(\omega t + \varphi)}$,

where C_{base} is the base level of the number of cases in the county, ω in e is the angular velocity of the number of cases in the county, and φ is the modified initial phase.

2. **The influence of the number of cases in the current space around on the number of local cases.** According to the figure we establish, we need $S(t)$ to reflect the randomness and correlation of the number of cases, so $S(t)$ should be a random component monitoring the influence of the number of cases in the surrounding space. We believe that the impact on the number of the cases in the local counties is positively correlated with the number of the adjacent county cases and negatively correlated with the distance between the two places. At the same time, we note that distance and influence are not strictly inversely proportional. For example, at a distance of enough, a difference of 1 or 2 kilometers does not make a significant difference. We think that taking the logarithm of the distance is going to be a good solution.

Based on the above, we get Equation (1).

$$P_j(t) = C_{base} e^{\sin(\omega t + \varphi)} \left(1 + \sum_{i=1}^k \frac{C_{i,j} P_i(t)}{\ln(D_{i,j} + 1)} \right) \quad (1)$$

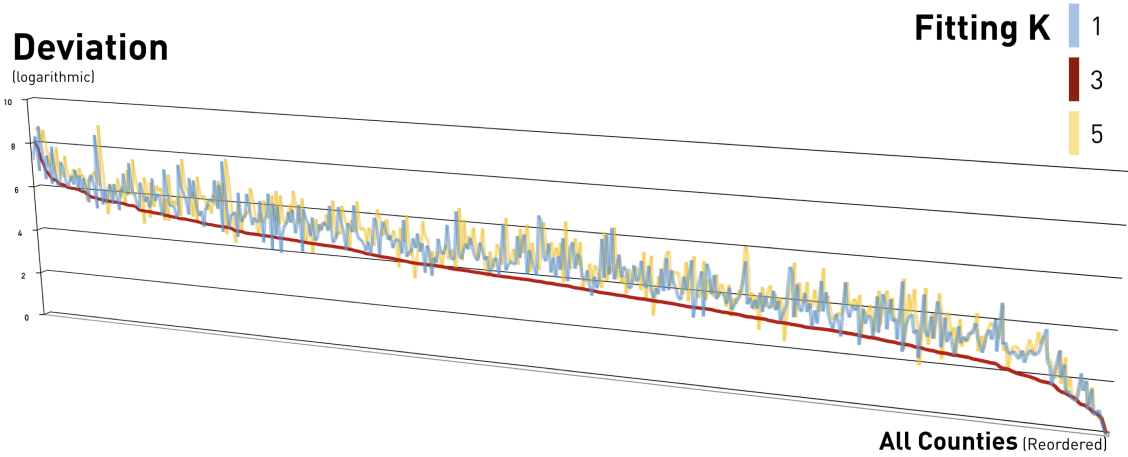
In the equation, P_j represents the number of cases in the target county, P_i represents the number of cases in the surrounding county, $(1 + \sum_{i=1}^k \frac{C_{i,j} P_i(t)}{\ln(D_{i,j} + 1)})$ represents the effect of k surrounding counties on target counties (We consider that this parameter is going to be the value of fluctuating on a base of 1), $C_{i,j}$ represents the influence coefficient of county i on county j , and $D_{i,j}$ represents the distance between county i and county j .

5.1.2 Model Simplification

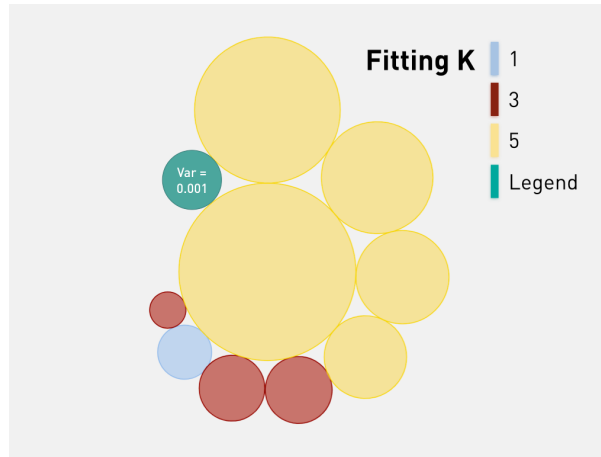
In the model, if we take into account the relationship among all counties, then k will be 451, and we will calculate 451 correlation coefficients when we analyze each county. Considering that there is the data of only eight years from 2010 to 2017 in the file "MCM_NFLIS_Data.xlsx", we cannot specifically calculate the specific value of each parameter. And when the geographical location of two counties is far from each other, the relationship between them is relatively small, and then its influence can be ignored. In other words, we can only take adjacent counties that are close to the target county as the analysis target. How do we define proximity? Based on the map information, we collect the longitude and latitude information of each county. Since the latitude and longitude of the five states do not vary much, we used the longitude and latitude of the two locations to calculate their relative distances(2).

$$D_{i,j} = \sqrt{(long_i - long_j)^2 + (lat_i - lat_j)^2} \quad (2)$$

In the equation, $long_i$ represents the longitude of county i , $long_j$ represents the longitude of county j , lat_i represents the latitude of county i , and lat_j represents the latitude of county j . For each county, we rank the surrounding counties from smallest to largest by distance. How many surrounding counties should we choose? From article[], we learned that using variance and deviation analysis can help us get the right parameter k . We get Figure 4 as a result. From the figure, when $k = 3$, we can get the best fit.



(a) Deviation Analysis



(b) Variance Analysis

Figure 4: Deviation And Variance Analysis

Finally, we can get the simplified equation(3):

$$P_j(t) = C_{base} e^{\sin(\omega t + \varphi)} \left(\frac{C_1 P_1(t)}{\ln(D_1 + 1)} + \frac{C_2 P_2(t)}{\ln(D_2 + 1)} + \frac{C_3 P_3(t)}{\ln(D_3 + 1)} \right) \quad (3)$$

5.1.3 Problem Solving

Based on Equation(3), we put the data of eight years into the formula to fit it, and we get a graph of each county based on the total amount of drugs⁵.



Figure 5:

5.1.4 Sensitivity Analysis

5.2 Problem 2

5.2.1 Construction Process

In this problem, we develop the model in Problem 1, and use the modified model as a tool to probe into the problem with the analytic hierarchy process. First, we consider a problem from the bottom up. We analyze and process the data in the "ACS_1x_5YR_DP02" folder, and find that they divide the county's demographics and family structure by different criteria. Then we establish a structural proportion database for each county from 2010 to 2017. These indicators are grouped into three categories: family factors, educational factors, and cultural factors and the structure of each indicator is divided into influential factors and non-influential factors⁶. For example, in the education level indicator, we consider that people with high school education and below account for the majority of cases, while those with higher levels of education account for the minority. After this division, we do the geometric average on the indicator to get the criterion of a category. Inside criterion, each indicator has the same weight.

Secondly, we fit the model from the top down. We use the control variable method to study the importance of family factors to educational factors, family

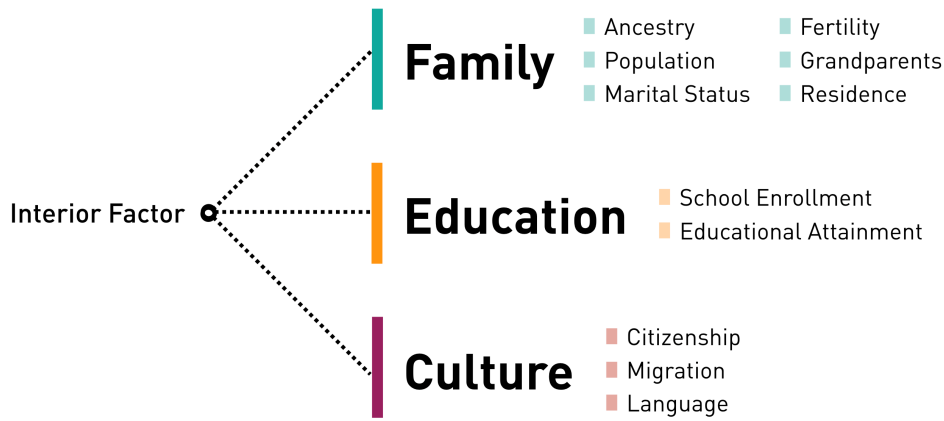


Figure 6: AHP of Factors Affecting Drug Use

factors to cultural factors, and family factors to cultural factors. The following is an example of the importance of the education factor to the family factor of a county. We assume that family factor is F , educational factor is E , the degree of importance is x . We add $(Fx + E)$ as a coefficient to model one and get Equation (4). After fitting, we can get the value of $x_{F,E}$.

$$P_j(t) = C_{base} e^{\sin(\omega t + \varphi)} \left(\frac{C_1 P_1(t)}{\ln(D_1 + 1)} + \frac{C_2 P_2(t)}{\ln(D_2 + 1)} + \frac{C_3 P_3(t)}{\ln(D_3 + 1)} \right) (Fx + E) \quad (4)$$

And so on we can get three important levels ($x_{F,E}$, $x_{F,C}$, $x_{E,C}$) for each county. And then we can get the following matrix.

$$\begin{pmatrix} 1 & x_{F,E} & x_{F,C} \\ \frac{1}{x_{F,E}} & 1 & x_{E,C} \\ \frac{1}{x_{F,C}} & \frac{1}{x_{E,C}} & 1 \end{pmatrix}$$

A new matrix is obtained by normalizing each column of the matrix. Then we figure out the average value of each row, and they are influence coefficients of family factor, education factor and cultural factor of each county (a, b , and d). Finally taking the logarithm of the population of each county, we take the weighted average of the three coefficients and get The general coefficients of the three influence factors. We get the final equation: Equation(5).

$$P_j(t) = C_{base} e^{\sin(\omega t + \varphi)} \left(\frac{C_1 P_1(t)}{\ln(D_1 + 1)} + \frac{C_2 P_2(t)}{\ln(D_2 + 1)} + \frac{C_3 P_3(t)}{\ln(D_3 + 1)} \right) (aF + bE + dC) \quad (5)$$

5.3 Problem 3

5.3.1 Problem Solving

In model 1, we consider that case number changes are influenced by both time and space factors and obtain the fitting result with small deviation. In model 2, we combine the results of the county population survey, focus on the internal factors, and identify the important degree of the influence of different population factors through the analytic hierarchy process. Based on all the data provided and the county geographic information that we have looked up, we can get a complete equation(5) of how the county case numbers are changing.

We have taken into account the response to the drug crisis in many regions and find that they take into account both the whole and the parts. In other words, the states control the trend by making laws, strengthen safety education, and monitoring food safety. In the meanwhile, the counties monitor the region where drug cases are most frequent. The above measure takes both efficiency and effect into consideration, with little investment to achieve the purpose of stable drug control.

Degree Centrality can reflect the degree of association between a county and other neighboring counties. In other words, in counties with high clustering coefficient, traffic is always developed and drug trade is active. We decide to use the Degree Centrality as the measurement standard. By sorting the county Degree Centrality, we can get the list of key areas.

Based on the above results, we get strategies to deal with the drug crisis:

1. For counties with a high degree of centrality, we should strictly control their cargo links with surrounding counties and reduce drug trafficking between them.
2. We should make efforts to raise the educational level of the people.
3. The community should pay more attention to the family relations and mediate the family contradiction in time.

5.3.2 Evaluating

5.4 Strengths

- **Simplification**

When building the model, we find the influence of the three surrounding counties account for a significant proportion, so we selectively ignore the influence of other counties. This greatly simplifies the process of model building and analysis.

- **Exactness**

We fully consider the various influencing factors that are already known, and get a good fit after detailed discussion for the factors.

- **Stabilitys**

Our models have some undetermined factors. We carry out Sensitivity analysis, consistency test and deviation analysis on these parameters, and obtain expected results.

- **Stabilitys**

We have produced a general algorithm to solve the opioid crisis problem. Our model can not only fit the description of the opioid crisis in the United States, but also be used in other disciplines to analyze the interaction between local regions and the overall planning, such as sales pricing in different regions, news communication, and so on.

5.5 Weaknesses

- When calculating distance between counties, the curvature of the earth isn't considered. These 5 states latitude varies between 36° N and 42° N. That causes notable bias when the latitude difference can't be ignored.
- Not all social and economic factor data is accessible, so the model can't be complete as lacking traffic, politics, and laws information.
- Only 5 states data of the U.S. were provided, so these counties lying in the borderline between the 5 states and the rest would not be well described.

6 Conclusions

MEMORANDUM

To: The Chief Administrator, DEA/NFLIS Database

From: Team #1919828

Date: Jan 28th, 2019

Subject: Drug Crime Control Measures for The Opioid Crisis

Honorable the Chief Administrator, DEA/NFLIS Database,

Situation: Our research based on east-northern 5 states reveals that American opioid and heroin related crime frequency is strongly regional and spatial. The total incident amount of each county floats around a reference value, at the same time, significantly affected by nearby counties and border geographical environment. Specific kinds of drug's propagation indicates the existence of drug selling chain and regional crime gang. Drug control faces a complex situation and great pressure.

Influencing Factor: According to economic research reports from NFLIS, the crime frequency is 52.16% related to culture factors, including nationality ratio, first language proportion, and ancestral combination of its population. It is also 31.82% related to education factors, as the education level and dropout rate of its population. It is merely 16.02% related to family factors, e.g., family people numbers, relatives' health status, and martial status.

Prediction: In the next decade, American opioid and heroin crime cases would keep increasing spirally. It is predicted that the growth-recession cycle is distributed between 3.5 years and 12.5 years, along with an approximately 5.6% increase by each cycle. Regional criminal gangs would develop closer communication, leading to the synchronized periodical floating of crime cases in a specific region. Consequently the managing of drugs would be much harder.

Advice: Ordered by importance, as following:

- Raise the attention of drug crime, and implement strict drug managing laws. Control the planting and import of drug crops with the help of U.S. Department of Agriculture and Commerce.
- Combat the drug criminals according to its growth-recession cycle, deploy police force in advance, and maintain a low level of stability in the number of criminals.
- Pay more attention to crimes around the inter-state borders, cut drug trafficking channel, and promote interstate cooperation in drug controlling system.
- Legislate for aggravated punishment.

The above is the summary of our study. We sincerely hope that it will provide you with useful information. Thanks!

References

- [1] "WORLD DRUG REPORT 2018", <https://www.unodc.org/wdr2018/prelaunch/>
- [2] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D U. "Complex networks: Structure and dynamics". Physics reports, 2006, 424(4): 175-308.
- [3] <https://www.census.gov/geo/maps-data/>
- [4] <http://www.chinatex.org/>

Appendices

Appendix A First appendix

Aliquam lectus. Vivamus leo. Quisque ornare tellus ullamcorper nulla. Mauris porttitor pharetra tortor. Sed fringilla justo sed mauris. Mauris tellus. Sed non leo. Nullam elementum, magna in cursus sodales, augue est scelerisque sapien, venenatis congue nulla arcu et pede. Ut suscipit enim vel sapien. Donec congue. Maecenas urna mi, suscipit in, placerat ut, vestibulum ut, massa. Fusce ultrices nulla et nisl.

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

Appendix B Second appendix