**Identify risk factors of using opioid in Indiana**

**Introduction**

The opioid crisis is a big problem in the United States. According to the DEA, 128 people died after overdosing the opioid in 2018 every day, so opioid misuse is a very series issue. Recently, the Washington Post released opioid sell data from 2006-2012. I believe that it is important to look at the data and find out what proportions of the population are hit the hardest by the opioid abuse.

I found that the most 2 important factors that contribute to opioid use in Indiana are unemployment rate and white proportion. Counties that have higher unemployment rate consume higher opioid on average, and counties that have higher white population consume less opioid on average.

**Dataset**

The opioid data by Washington post has the opioid sell number in each store location, the amount of opioid sold, and the time that the opioid was sold. The data time range is from 2006-2012.

The demographic data is gathered using US census bureau. The demographic factors are:

* Poverty rate for each year from 2006-2012
* Unemployment rate for each year from 2006-2012
* Median household income from 2006-2012
* Female proportion in 2010
* White proportion in 2010
* Proportion of people who never earned a diploma in 2010
* Proportion of people who earned a diploma in 2010
* Proportion of people who went to college but not get a bachelor in 2010
* Proportion of people who got a bachelor degree in 2010
* Proportion of people in different age group in 2010

**data processing and data analysis**

Using the original opioid data, I gather the amount of opioid is sold each year for each county in Indiana from 2006-2012. The amount of opioid sold is in mg. I divided the amount of opioid sold by the population data gathered from US census bureau to offset the population effect.

By looking at the heatmap of opioid sold per person in each county in 2006-2012 (Figure 1), we see that there is some spatial correlation between the nearby counties.

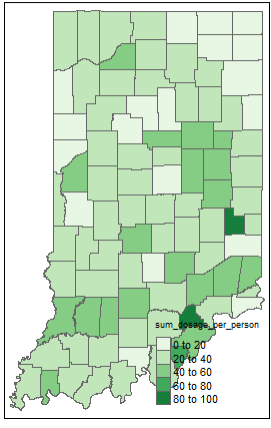


Figure 1: opioid sold per person in Indiana in 2006(check appendix for other years)

By looking at the data in each county, I notice that there is an increasing trend of opioid sold per person with each year. So there is a temporal correlation for the data.

For the demographic data, US census only collects demographic data once every 10 years. Because our data spans from 2006-2012. I feel like using demographic data from 2010 is the best.

The data doesn’t report any opioid sold in Ohio and Warren. Since its not likely the case because Ohio and Warren both have significant population (around 5000 and 8000 respectively). I believe the data is missing.

For race, because the Indiana is predominantly white (83%). I believe that it is best to just have white proportion vs non white proportion.

For age group, I decide to split the age into 3 major groups: young people (20-39), middle age(40-59), and old people (60+). The reason I don’t have a group for 20- because I believe people under 18 cannot purchase opioid.

Note that there can be significant collinearity between predictions. For example, race data, education data, and economy data will have significant correlation.

**Model**

Because there are significant spatial temporal factor in the data. I decided to use the spatial-temporal model. The spatial temporal model I use can be found the R package SpatioTemporal.

The model is in the form:

Where y(s,t) is the observation, is the structure mean field, and v(s,t) is the residual field.

The mean field is:

Where the is the spatial temporal coefficient, and is just the spatial coefficient. Beta\_i follows normal distribution with mean α\_i\*X. The model assumes that y(s,t) is normal distribution. For more information, you can check the SpatioTemporal package information.

We have 3 spatial temporal predictors: poverty rate, unemployment rate, and household median income. The rest are spatial predictors.

I will run the model for the data from 2006-2011. Then I will predict the result in 2012 and check for the mean squared error just to make sure the model gives a fair result

**Model result**

I first run the model with all the predictors. Looking at the result (table 1), we see that unemployment rate, white proportion, proportion of people who earn bachelor degree, proportion of young people, and proportion of people who are middle age are significantly different from 0. I run the model and I found that the MSE of predictions for opioid use in 2012 is 29.3, which is a fair prediction.

|  |  |  |  |
| --- | --- | --- | --- |
| Parameters: | Point estimate | Sd | tstat |
| Gamma(median HH income) | 1.1\*10^-4 | 1.55\*10^-4 | 0.71 |
| Gamma(poverty rate) | 0.17 | 2.44\*10^-1 | 0.70 |
| Gamma(unemployment rate) | 1.05 | 3.73\*10^-1 | 2.84 |
| Alpha(female proportion) | 3.80 | 2.18 | 1.74 |
| Alpha(white proportion) | -2.01 | 6.10\*10^-1 | -3.28 |
| Alpha(high school diploma) | -1.57 | 0.90 | -1.75 |
| Alpha(some college) | -0.353 | 0.93 | -0.38 |
| Alpha(Bachelor degree) | -2.25 | 0.72 | -3.13 |
| Alpha(young people proportion) | 4.74 | 1.74 | 2.7 |
| Alpha(middle age) | 5.01 | 1.86 | 2.68 |
| Alpha(old people) | 2.75 | 1.48 | 1.86 |

Table 1: model results for all predictors.

So I start with the model with all the significant predictors from above. I select the best model by dropping the least significant predictors. Then I try to add different predictors to see if they are significant. I also keep track of the MSE to see if the MSE increases or decreases drastically. And the final model I find is:

|  |  |  |  |
| --- | --- | --- | --- |
| Parameters: | Point estimate | Sd | tstat |
| Gamma(unemployment rate) | 1.21 | 0.37 | 3.28 |
| Alpha(white proportion) | -1.86 | 0.47 | -3.96 |

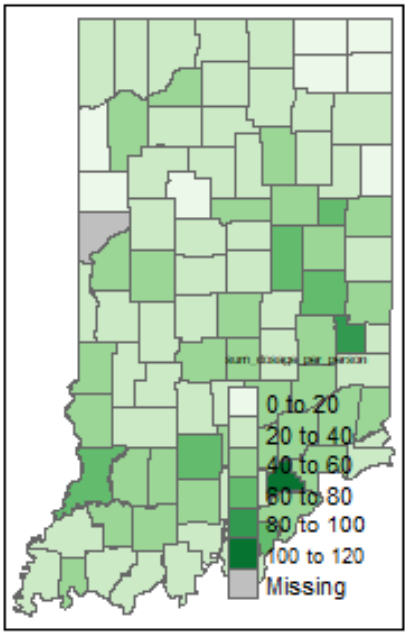
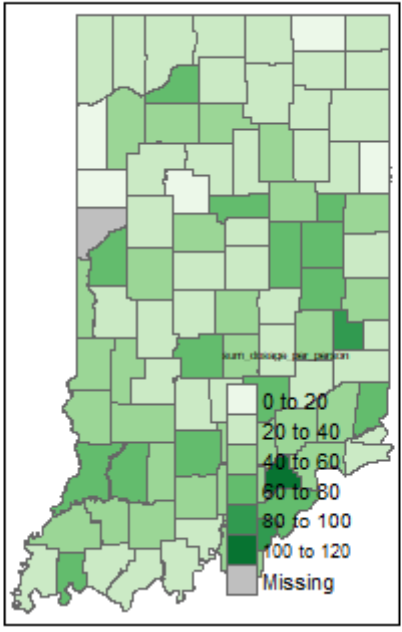
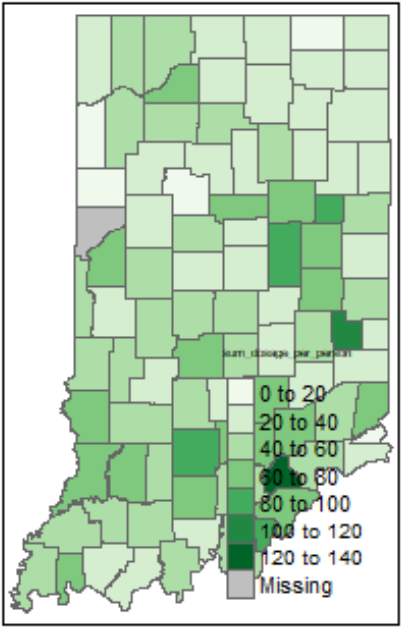
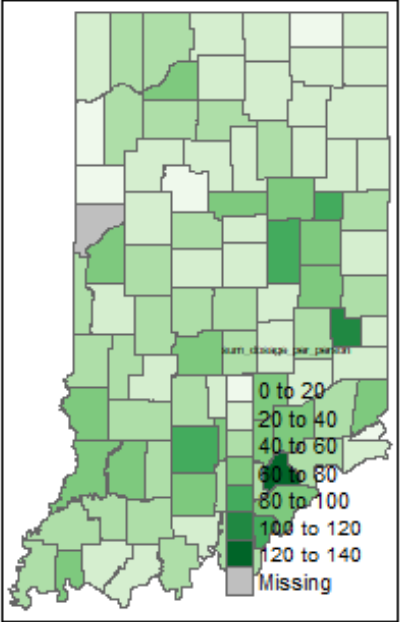
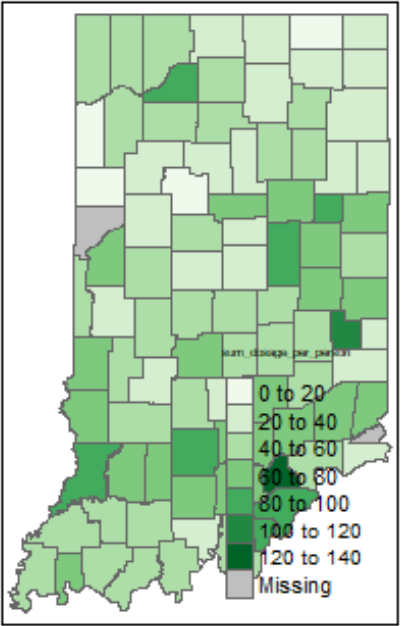
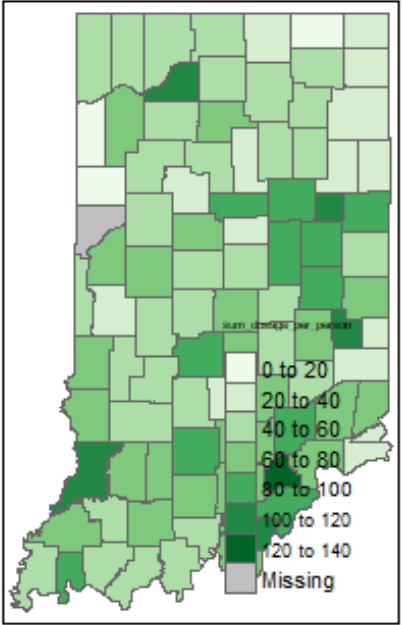
Table 2: model results for best predictors.

The model reduced model has MSE of 28.3. So this model actually performs better and is simpler.

I also run a model alone with female proportion and find that the female proportion is not a good predictor for opioid consumption. This is not consistent with some theory that doctors are less likely to prescribe women for opioid.

From the model, we see that counties in Indiana that have higher unemployment rate consume more opioid, and counties that have higher non-white population consume more opioid. I believe this is the case because unemployment can lead to depression that leads to opioid consumption. Opioid consumption doesn’t depend on sex. Other factors might contribute to the opioid use, but they are correlated with unemployment or race data.

Appendix:

Figures: opioid sold per person in Indiana in 2007-2012