Crime in New York City

**Abstract**

**Part I. Introduction**

Topic

What lead to a decrease in crime in NYC since the 1960s? After reading from Steven Levitt's book Freakonomics, we assumed that features listed below may contribute to the decrease and our aim is to examine the causal relation of these features to the crime rate in NYC.

* Abortion rates
* Capital punishment [x]
* Increased incarceration rates
* Growing number of police officers [x] - same
* Dow Jones industrial average
* Employment rate

Background

Abortion Laws

(approved in 1973 and how this affect our research)

Capital punishment in NYC

(Which isn’t apply to NYC since it doesn’t have death penalty)

The formula in Freakonomics

(Linear regression, and explain why we use these features)

III. Summary

**Part II. Data Processing**

I. Source

(describe the way to get data, easy part)

For abortion, we found historical abortion rate data of New York city from Centers for Disease Control and Prevention (CDC), who offers reproductive health report [1] annually, thus we got a list of abortion data from 1982 – 2015. Still, with some data missing, we try to find a better data source.

Another option is using Alan Guttmacher Institute (AGI) survey’s data which is 1971 – 2016, and still has many data missing. Finally, we found a historical abortion statistic of New York State [2], which gave data range from 1930 – 2018 and contains both CDC and AGI data.

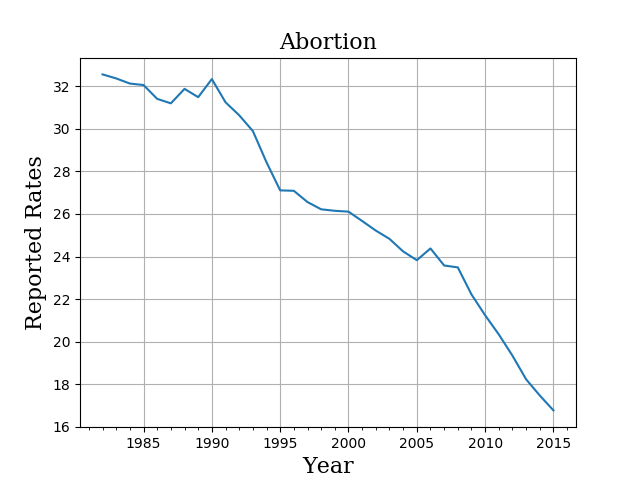
II. Missing data

(Use the paper we discussed to tell the way fit data in)

III. Data Stationary

Stationary data means the data has properties that its mean, variance and covariance is stable with time and do not have trend or seasonal effects.

Before Stationary



Log transformation (if data distribution skewed on the left or right side)

No need for abortion but will see how other features look like

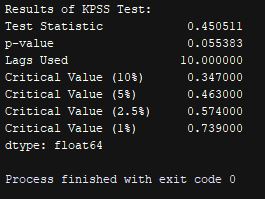
Often used for parametric statistical test

KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test

(for data with a trend, and in this case, abortion)

Null Hypothesis we made (H0): The data with trend is stationary

Alternative Hypothesis we made (Ha): The data series has a unit root(nonstationary)



(need convert to table…)

Analysis:

IV. Cointegration

**Part III. Granger Causality**

**Part IV. Counterfactual Analysis**

**Part V. Linear Regression Model**

**Part VI. Discussion**

**Reference**

Johnston, R. (2018, November). Historical abortion statistics, New York (USA). Retrieved November 2018, from <http://www.johnstonsarchive.net/policy/abortion/usa/ab-usa-NY.html>

Reproductive Health. (2017, November 16). Retrieved November 10, 2018, from <https://www.cdc.gov/reproductivehealth/data_stats/abortion.htm>

Brownlee, J. (2018, October 18). How to Check if Time Series Data is Stationary with Python. Retrieved November 21, 2018, from <https://machinelearningmastery.com/time-series-data-stationary-python/>

https://www.researchgate.net/post/Do\_we\_have\_any\_permission\_to\_test\_Granger\_causality\_or\_not

<https://www.statisticshowto.datasciencecentral.com/granger-causality/>

<https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python/>

<https://www.statisticshowto.datasciencecentral.com/kpss-test/>

<https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.kpss.html>

<https://freakonometrics.hypotheses.org/12729>

<https://www.zhihu.com/question/23680352>