Project 5505

Yifan Li, Simao Gao, Zijian Huang

December 4, 2018

Abstract

In this project, we analysis the patterns among time of crimes, number of crimes and victims' age and find that

1. people who are 35 years old are most likely to be attacked.

2. 7am-8am and 3am-4am is safe time zone comparatively while 4am-6am and 20pm-22pm is the most dangerous.

3. The expected number of crimes occurring in one month is among (526.3546, 536.7214).

1 Introduction and Description of Data

Here are three questions we want to analysis.

1. What is the distribution of the age of victims?
2. Do criminals have a preference for crime time? In other words, is there any association between daily time interval and the number of crime.
3. What is the distribution of the number of crimes occurring in one month?

Variables we used are: *area.id*, *Date.Occurred*, *Time.Occurred*, and *Victim.Age*.

2 Exploratory Data Analysis

First we plot the histogram of victims' age. From Figure 1, it is a good idea to fit a poisson distribution.

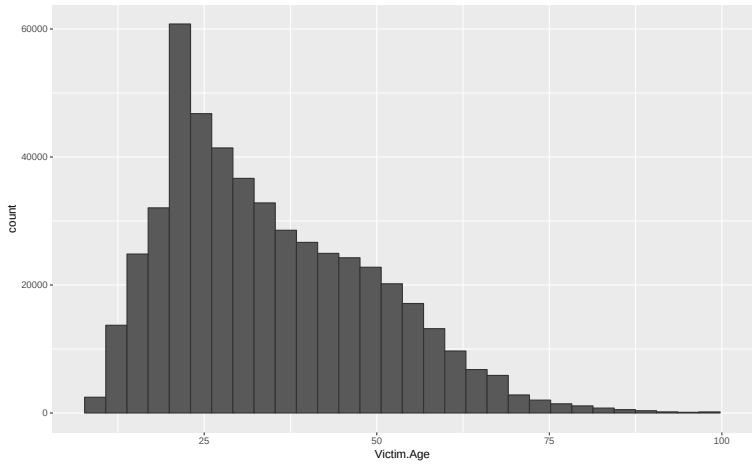

Figure 1: Histogram of Victim Age

For the purpose to analyze the relationship between the daily time interval with the number of crimes. Firstly, the daily time is cut from 00:00:00 to 24:00:00 into hourly basis, such as, 00:00:00-01:00:00 and 01:00:00-02:00:00. Thus, we will gain 24 intervals. Secondly, the corresponding frequency should be counted. Finally, we plot a histogram of crime in different time intervals.

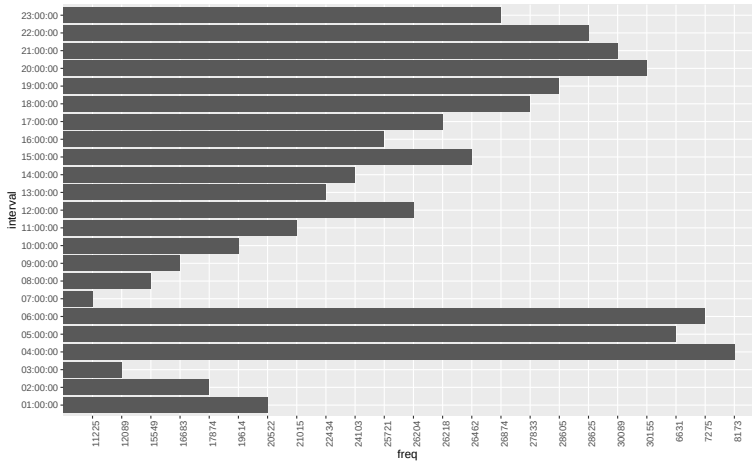

Figure 2: Histogram of Victim Time

From Figure 2, the number of the crime are largest. It also means it will be dangerous during the period during 4am-6am and 20pm-22pm. However, at 7am-8am and 3am-4am will be safe and almost no crime. They may be safe zone.

3 Statistical Inference

3.1 Distribution of the Age of Victims

- H_0: the age of victims follows poisson distribution

- Test used: CHi-square test

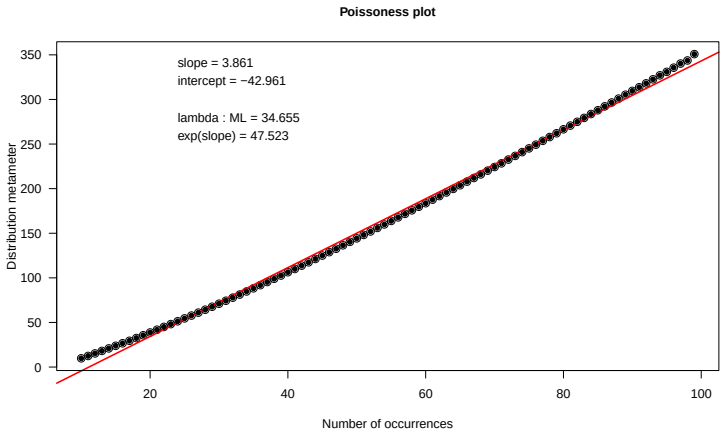

Figure 3: Poissoness Plot of Victims' Age

From Figure 3, the poisson distribution fits the data quite well. So we do a Chi-square test to test whether it follows the poisson distribution or not. And the p-value is 0.2492. So we don't reject H_0 and the age of victims follows poisson distribution and the MLE estimation of parameter is about 35. This indicates that people who are 35 should be careful. They are most likely to be attacked.

3.2 Distribution of the Time of Victims

- H_0: the time of victims follows uniform distribution

- Test used: CHi-square test

Set the null hypothesis as the frequency follow the union distribution. Actually, from Figure 2, the expected result should be not follow the uniform distribution. In other words, there exists relationship between the time interval and frequency. Suppose proportion of the occurrence of each frequency is equal: 1/24= 0.04166667. The expected value of the frequency will be equal to the proportion times the total number of crimes. After apply the chi-square test, we realized that the p-value=0 which will lead us to reject the null hypothesis

and state that the distribution of the frequency is not follow uniform distribution. In other words, there exist some relationship between the time interval and frequency.

3.3 Distribution of the Number of Crimes

- H_0: the number of crimes follows poisson distribution

- Test used: Poissoness Plot

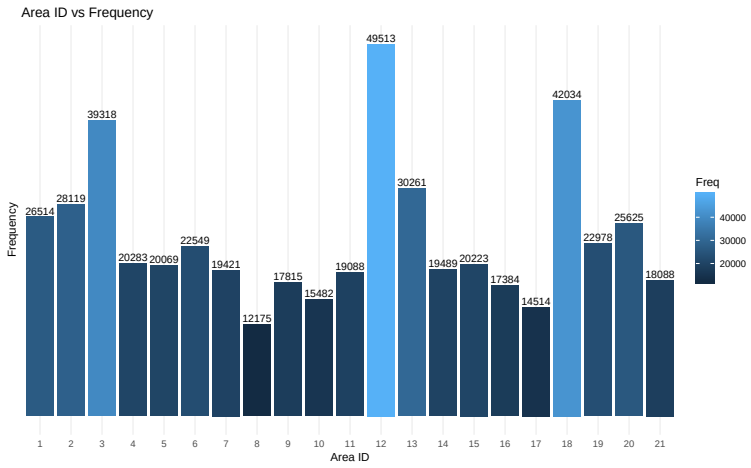

Figure 4: Area ID vs Frequency

From Figure 4, area with id "12" has the highest frequency of crimes. So we are curious that whether the number of crimes occurrence in each month for area 12 follows distribution or not, hence, we construct poissonness plot to find out.

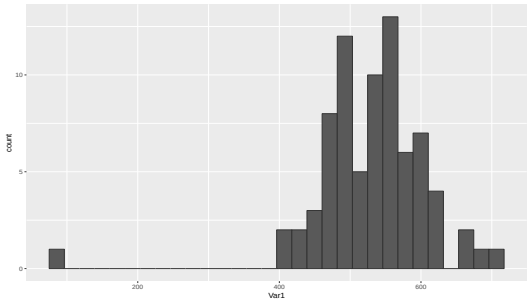

Figure 5: Histogram of Crimes' number

From Figure 5, "2017-10" is a outlier since it has only 80 crimes comparing to 500-800 crimes for other months. After removing the outlier, the poissoness plot is as follows.

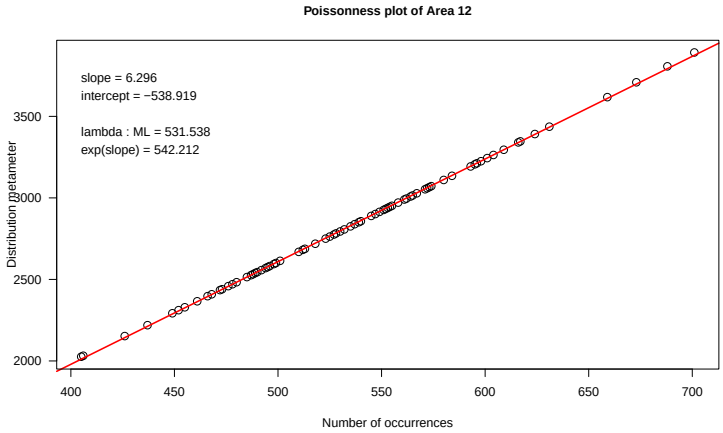

Figure 6: Poissoness Plot of Crimes' number

From Figure 6, it is apparent that the points approximated lay on the straight line. We can say that poisson distribution fits the data set.

Therefore, with MLE of λ, we can calculate the 95% confidence interval for this MLE. $Var(\hat{\lambda}) = \sqrt{\frac{\hat{\lambda}}{n}}$.Therefore, the 95% confidence interval for λ using MLE method is (526.3546, 536.7214).

4 Conclusion

So, our conclusion is:

1. The age of victims follow a poisson distribution with parameter equal to 35.

2. There exists association between daily time interval and the number of crime.

3. The number of crimes occurring in one month follows poisson distribution. 95% confidence interval for it is (526.3546, 536.7214).