UM-SJTU JOINT INSTITUTE
PROBABILISTIC METHODS IN ENGINEERING
(VE401)

PROJECT 2 REPORT

TEAM PROJECT GROUP 30

Authors' Name and ID
Chen Zhibo 515020910276
Pan Chongdan 516370910121
Shen Yuan 516370910122
Xiang Zhiyuan 516370910126
Zhan Yan 516370910206

Date: August 1, 2018

Contents

1 Synopsis

The mass shooting problem in the United States is increasingly severe in recent years, and many statistic group has collected the relevant data for several years. Our group believe that the knowledge of statistical methods can be applied in the area of public security. So we based on the data obtained from Gun Violence Archive to analyze the distribution of numbers of occurrence of mass shooting in one day in the United States.

Following the example of the article *London murders: a predictable pattern?*, at first we thought it may follow a Poisson distribution. However, the data shows from year 2013 to 2017, there is evidence to reject that the numbers of occurrence of mass shooting in one day follow a Poisson distribution, even if we test the data in each individual year, the hypotheses are all rejected.

But then, we find the most mass shooting occur on Saturday and Sunday, and we confirmed it depends on the weekdays.

Then we tested the hypotheses that the numbers of occurrence of mass shooting in one week follow a Poisson distribution, or the numbers of occurrence of mass shooting in one year follow a Poisson distribution. Unfortunately, there is evidence to reject these two hypotheses.

Then we narrowed down our test range, instead of testing the 5 years data at the same time, we just focused on the weekly data from January-June 2018. And at this time, we found there is no evidence the data does not follow a Poisson distribution with parameter $k = 6.04$, that means the numbers of occurrence of mass shooting in one week may follow a Poisson distribution in a short period, like a single year.

Finally we got our conclusion, that is the occurrence of mass shootings depends on weekdays, and it is more likely to happen in weekends. And the numbers of occurrence of mass shooting in one week may follow a Poisson distribution in a short period, like a single year, but not in a long period, and the reason may be that the occurrence of mass shooting is not a random event. People may tend to shoot when the social security is not so good, it depends on many social factors, so in a short period, when the social and national conditions don't change a lot, the the numbers of occurrence of mass shooting in one week may follow a Poisson distribution, but in a long period, since the society and country are changing, which may cause the parameter k of the Poisson distribution to change, so we can't simply use Poisson distribution to test the data in a long period.

2 Project Introduction

The article *London murders: a predictable pattern*?, written by David Spiegelhalter and Arthur Barnett [4], discusses the pattern of murders in London between April 2004 and September 2007 based on data of the London Metropolitan Police, obtained from the British Home Office. And in this project, we followed the example of David and Arthur, and instead of basing on the London murders, we focused on the social issue in the United States, that is Shooting Accident, especially the Mass Shooting problem. The mass shooting problem is increasingly severe in recent years, and our group believe that the knowledge of statistical methods can be applied in the area of public security. So we tried to base on the data obtained from Gun Violence Archive[1] between January 2013 and December 2017 to analyze the mass shooting, hoping to find any pattern behind the data, so that we may be able to predict the occurrence of mass shooting in a period of time, and it may be helpful to prevent these issues to happen in the future.

3 Exercise

3.1 Definitions of mass shooting and mass murder[5]

So, what is mass shooting? How we define it?

A mass shooting is an incident involving multiple victims of firearms-related violence, and there is not a broadly accepted definition, for example, the United States'Congressional Research Service defines a "public mass shooting" as one in which four or more people selected indiscriminately, not including the perpetrator, are killed, and another unofficial definition of a mass shooting is an event involving the shooting(not necessarily resulting in death) of five or more people(sometimes four) with no "cooling-off period". And a mass murder is the act of murdering a number of people, typically simultaneously or over a relatively short period of time and in close geographic proximity, and FBI defines mass murder as murdering four or more persons during an event with no "cooling-off period" between murders.

Since there is not a broadly accepted definition of mass shooting, in our project, we decide to use the definition of Gun Violence Archive(GVA), and actually, if you comparing with other statistical data, you will find the GVA mass shooting numbers are higher than some other sources. Because GVA uses a purely statistical threshold to define mass shooting based ONLY on the numeric value of 4 or more shot or killed, not including the shooter. And GVA does not parse the definition to remove any subcategory of shooting, that means GVA does not exclude, set apart, caveat, or differentiate victims based upon the circumstances in which they were shot.

3.2 Mass shooting data from 2013 to 2017

To begin our project, first we collected numbers of mass shootings in one day from 2013 to 2017, and created a figure below. Based on the figure we found that the number of days with only one mass shooting happened is greatest, and then is the days with 2,3,4,5 mass shootings happened, and the number of days with 6 mass shootings happened is the smallest. Since in the article *London murders*: *a predictable pattern*?, David Spiegelhalter and Arthur Barnett estimated that the London homicides follow a Poisson distribution with parameter $k = 0.44$, and our figure showed there is a chance the mass shootings in the United States follow a Poisson distribution. But we needed to do some tests to confirm our hypothesis.

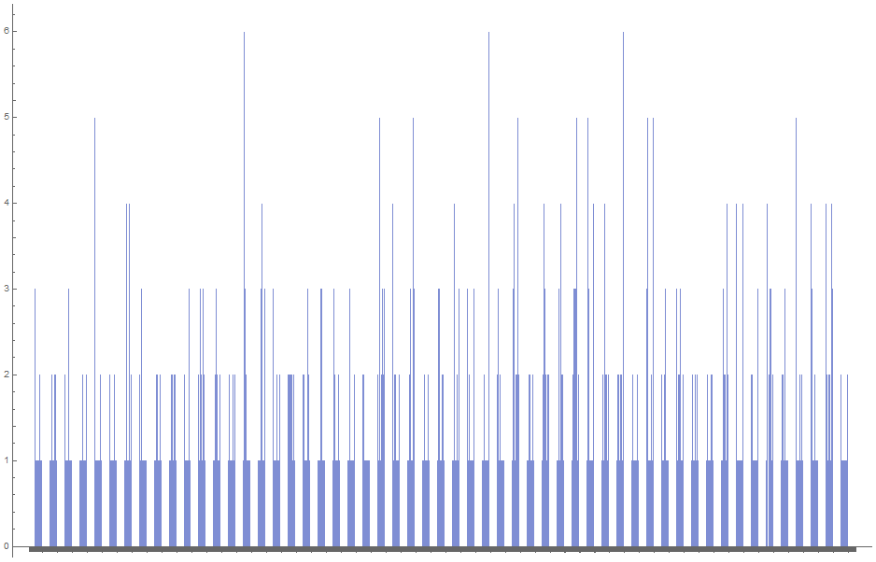


Figure 1: The data is from 2013 to 2017

3.3 Test of Poisson distribution of mass shooting in US

Here we decided to use Pearson's test to test our hypothesis, that is the mass shootings in the United States follow a Poisson distribution.

We use the data from GVA and get the number of mass shootings that happened in each day between January 1st, 2013 and December 31st, 2017:

The total mass shootings number is $540+240\times2+102\times3+37\times4+15\times5+6\times6 = 1585$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{1585}{1826} \approx 0.868$$

Then we first calculate:

$$P[X = 0] = \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.41981$$
$$P[X = 1] = \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.36442$$

Numbers of mass shootings each day	numbers of days
0	886
1	540
2	240
3	102
4	37
5	15
6	6

Table 1: Observed number of days with number of mass shootings from years 2013 to 2017

$$
\begin{aligned}
P[X = 2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.15810 \\
P[X = 3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.04581 \\
P[X = 4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.00993 \\
P[X = 5] &= \frac{e^{\hat{k}}\hat{k}^5}{5!} \approx 0.00172
\end{aligned}
$$

$$
\begin{gathered}
P[X \geq 6] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4] - P[X = 5] \\
= 0.00025
\end{gathered}
$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5, p_6) = (0.41981, 0.36442, 0.15810, 0.04581, 0.00993, 0.00172, 0.00025)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 1826$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	766.57	886
1	665.43	540
2	288.69	240
3	83.65	102
4	18.13	37
5	3.14	15
6	0.4565	6

Table 2: Observed number of days and expected number of days with number of mass shootings from years 2013 to 2017

However, if we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

The problem can be solved by combining the last two categories: Then let us test:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	766.57	886
1	665.43	540
2	288.69	240
3	83.65	102
4	18.13	37
5	3.60	21

Table 3: Observed number of days and expected number of days with number of mass shootings from years 2013 to 2017

H_0: The number of mass shootings happened in one day in United States follows a Poisson distribution with parameter $k = 0.868$.

For $n = 6$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 4$ degrees of freedom. We testing the hypothesis at $\alpha = 0.05$ and we will reject H_0 if $X^2 \geq x_{0.05,4} = 9.49$. Since we have:

$$X^2 = \frac{(886 - 766.57)^2}{766.57} + \frac{(540 - 665.43)^2}{665.43} + \frac{(240 - 288.69)^2}{288.69} + \frac{(102 - 83.65)^2}{83.65} + \frac{(37 - 18.13)^2}{18.13} + \frac{(21 - 3.60)^2}{3.60} \approx 158.23$$

Since $X^2 \geq x_{0.05,4}$, we can reject H_0, which means there is no evidence that the number of mass shootings happened in one day in United States follows a Poisson distribution with parameter $k = 0.868$.

Now, if we test the individual years for adherence to a Poisson distribution, what will we get?

First, we test the data for year 2013:

The total mass shootings number is $105 + 47 \times 2 + 8 \times 3 + 6 \times 4 + 1 \times 5 = 252$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{252}{365} \approx 0.6904$$

Numbers of mass shooting each day	numbers of days
0	198
1	105
2	47
3	8
4	6
5	1

Table 4: Observed number of days with number of mass shootings in year 2013

Then we first calculate:

$$\begin{aligned} P[X=0] &= \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.50137 \\ P[X=1] &= \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.34615 \\ P[X=2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.11949 \\ P[X=3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.02750 \\ P[X=4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.00475 \end{aligned}$$

$$\begin{gathered} P[X \geq 5] = 1 - P[X=0] - P[X=1] - P[X=2] - P[X=3] - P[X=4] \\ = 0.00074 \end{gathered}$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5) = (0.50137, 0.34615, 0.11949, 0.02750, 0.00475, 0.00074)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 365$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	183.00	198
1	126.34	105
2	43.61	47
3	10.04	8
4	1.73	6
5	0.27	1

Table 5: Observed number of days and expected number of days with number of mass shootings in year 2013

However, if we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

The problem can be solved by combining the last two categories:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	183.00	198
1	126.34	105
2	43.61	47
3	10.04	8
4	2.00	7

Table 6: Observed number of days and expected number of days with number of mass shootings in year 2013

For $n = 5$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 3$ degrees of freedom.

$$X^2 = \frac{(198 - 183.00)^2}{183.00} + \frac{(105 - 126.34)^2}{126.34} + \frac{(47 - 43.61)^2}{43.61} + \frac{(8 - 10.04)^2}{10.04} + \frac{(7 - 2)^2}{2} \approx 18.01$$

So the P value of our test for year 2013 is 0.000438, it's quite small.

Then, let's test the data for year 2014:

Numbers of mass shooting each day	numbers of days
0	194
1	103
2	45
3	18
4	3
5	1
6	1

Table 7: Observed number of days with number of mass shootings in year 2014

The total mass shootings number is $103 + 45 \times 2 + 18 \times 3 + 3 \times 4 + 1 \times 5 + 1 \times 6 = 270$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{270}{365} \approx 0.7397$$

Then we first calculate:

$$\begin{aligned}
P[X = 0] &= \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.47724 \\
P[X = 1] &= \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.35303 \\
P[X = 2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.13057 \\
P[X = 3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.03220 \\
P[X = 4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.00595 \\
P[X = 5] &= \frac{e^{\hat{k}}\hat{k}^5}{5!} \approx 0.00088
\end{aligned}$$

$$\begin{gathered}
P[X \geq 6] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4] - P[X = 5] \\
= 0.00013
\end{gathered}$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5, p_6) = (0.47724, 0.35303, 0.13057, 0.03220, 0.00595, 0.00088, 0.00013)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 365$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	174.19	194
1	128.86	103
2	47.66	45
3	11.75	18
4	2.17	3
5	0.32	1
6	0.05	1

Table 8: Observed number of days and expected number of days with number of mass shootings in year 2014

However, if we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	174.19	194
1	128.86	103
2	47.66	45
3	11.75	18
4	2.54	5

Table 9: Observed number of days and expected number of days with number of mass shootings in year 2014

The problem can be solved by combining the last three categories:
For $n = 5$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 3$ degrees of freedom.

$$X^2 = \frac{(194 - 174.19)^2}{174.19} + \frac{(103 - 128.86)^2}{128.86} + \frac{(45 - 47.66)^2}{47.66} + \frac{(18 - 11.75)^2}{11.75} + \frac{(5 - 2.54)^2}{2.54} \approx 13.30$$

The P value of our test for year 2014 is 0.004031, it's also quite small.

Then, let's test the data for year 2015:

Numbers of mass shooting each day	numbers of days
0	164
1	124
2	42
3	21
4	7
5	6
6	1

Table 10: Observed number of days with number of mass shootings in year 2015

The total mass shootings number is $124 + 42 \times 2 + 21 \times 3 + 7 \times 4 + 6 \times 5 + 1 \times 6 = 335$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{335}{365} \approx 0.9178$$

Then we first calculate:

$$\begin{aligned}
P[X = 0] &= \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.39939 \\
P[X = 1] &= \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.36657 \\
P[X = 2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.16822 \\
P[X = 3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.05146 \\
P[X = 4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.01181 \\
P[X = 5] &= \frac{e^{\hat{k}}\hat{k}^5}{5!} \approx 0.00217
\end{aligned}$$

$$\begin{gathered}
P[X \geq 6] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4] - P[X = 5] \\
= 0.00038
\end{gathered}$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5, p_6) = (0.39939, 0.36657, 0.16822, 0.05146, 0.01181, 0.00217, 0.00038)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 365$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	145.78	164
1	133.80	124
2	61.40	42
3	18.78	21
4	4.31	7
5	0.79	6
6	0.14	1

Table 11: Observed number of days and expected number of days with number of mass shootings in year 2015

However, if we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

The problem can be solved by combining the last three categories:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	145.78	164
1	133.80	124
2	61.40	42
3	18.78	21
4	5.24	14

Table 12: Observed number of days and expected number of days with number of mass shootings in year 2015

For $n = 5$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 3$ degrees of freedom.

$$X^2 = \frac{(164 - 145.78)^2}{145.78} + \frac{(124 - 133.80)^2}{133.80} + \frac{(42 - 61.40)^2}{61.40} + \frac{(21 - 18.78)^2}{18.78} + \frac{(14 - 5.24)^2}{5.24} \approx 24.03$$

The P value of our test for year 2015 is 2.46×10^{-5}, it's too small.

Then, let's test the data for year 2016:

Numbers of mass shooting each day	numbers of days
0	165
1	98
2	55
3	28
4	12
5	6
6	2

Table 13: Observed number of days with number of mass shootings in year 2016

The total mass shootings number is $98 + 55 \times 2 + 28 \times 3 + 12 \times 4 + 6 \times 5 + 2 \times 6 = 382$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{382}{365} \approx 1.047$$

Then we first calculate:

$$\begin{aligned}
P[X = 0] &= \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.35114 \\
P[X = 1] &= \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.36749 \\
P[X = 2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.19230 \\
P[X = 3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.06709 \\
P[X = 4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.01755 \\
P[X = 5] &= \frac{e^{\hat{k}}\hat{k}^5}{5!} \approx 0.00367
\end{aligned}$$

$$\begin{gathered}
P[X \geq 6] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4] - P[X = 5] \\
= 0.00076
\end{gathered}$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5, p_6) = (0.35114, 0.36749, 0.198230, 0.06709, 0.01755, 0.00367, 0.00076)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 365$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	128.17	165
1	134.13	98
2	70.19	55
3	24.49	28
4	6.41	12
5	1.34	6
6	0.28	2

Table 14: Observed number of days and expected number of days with number of mass shootings in year 2016

However, if we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

The problem can be solved by combining the last two categories:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	128.17	165
1	134.13	98
2	70.19	55
3	24.49	28
4	6.41	12
5	1.62	8

Table 15: Observed number of days and expected number of days with number of mass shootings in year 2016

For $n = 6$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 4$ degrees of freedom.

$$X^2 = \frac{(165 - 128.17)^2}{128.17} + \frac{(98 - 134.13)^2}{134.13} + \frac{(55 - 70.19)^2}{70.19}$$
$$+\frac{(28 - 24.49)^2}{24.49} + \frac{(12 - 6.41)^2}{6.41} + \frac{(8 - 1.64)^2}{1.64} \approx 53.65$$

The P value of our test for year 2016 is 6.23×10^{-11}. It's too small.

Then, let's test the data for year 2017:

Numbers of mass shooting each day	numbers of days
0	165
1	110
2	51
3	27
4	9
5	1
6	2

Table 16: Observed number of days with number of mass shootings in year 2017

The total mass shootings number is $110 + 51 \times 2 + 27 \times 3 + 9 \times 4 + 1 \times 5 + 2 \times 6 = 346$. Now if the number of mass shootings follows poisson distribution, then the estimator for k can be represent as below:

$$\hat{k} = \overline{X} = \frac{346}{365} \approx 0.9479$$

Then we first calculate:

$$\begin{aligned}
P[X = 0] &= \frac{e^{\hat{k}}\hat{k}^0}{0!} \approx 0.38754 \\
P[X = 1] &= \frac{e^{\hat{k}}\hat{k}^1}{1!} \approx 0.36736 \\
P[X = 2] &= \frac{e^{\hat{k}}\hat{k}^2}{2!} \approx 0.17412 \\
P[X = 3] &= \frac{e^{\hat{k}}\hat{k}^3}{3!} \approx 0.05502 \\
P[X = 4] &= \frac{e^{\hat{k}}\hat{k}^4}{4!} \approx 0.01304 \\
P[X = 5] &= \frac{e^{\hat{k}}\hat{k}^5}{5!} \approx 0.00247
\end{aligned}$$

$$\begin{aligned}
P[X \geq 6] = 1 - P[X = 0] - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4] - P[X = 5] \\
= 0.00045
\end{aligned}$$

We can then replace the distribution of X with that of a categorical random variable with parameters:

$$(p_0, p_1, p_2, p_3, p_5, p_6) = (0.38754, 0.36736, 0.17412, 0.05502, 0.01304, 0.00247, 0.00045)$$

We then can calculate the expected frequencies $E_i = np_i$ with $n = 365$ as follows:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	141.45	165
1	134.09	110
2	63.55	51
3	20.08	27
4	4.76	9
5	0.90	1
6	0.16	2

Table 17: Observed number of days and expected number of days with number of mass shootings in year 2017

However, we cannot apply Pearson's test here, since our data doesn't satisfy the first criteria: $E[X_i] = np_i \geq 1$ for all $i = 1, ..., k$.

The problem can be solved by combining the last two categories:

Category i	Expected Frequency E_i	Observed Frequency Q_i
0	141.45	165
1	134.09	110
2	63.55	51
3	20.08	27
4	4.76	9
5	1.06	3

Table 18: Observed number of days and expected number of days with number of mass shootings in year 2017

For $n = 6$ categories, the statistic:

$$X^2 = \sum_{i=0}^{n-1} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $n - 1 - m = 4$ degrees of freedom.

$$X^2 = \frac{(165 - 141.45)^2}{141.45} + \frac{(110 - 134.09)^2}{134.09} + \frac{(51 - 63.55)^2}{63.55} + \frac{(27 - 20.08)^2}{20.08} + \frac{(9 - 4.76)^2}{4.76} + \frac{(3 - 1.06)^2}{1.06} \approx 20.44$$

The P value of our test for year 2017 is 0.00041, it's quite small.

Since the P values of test for year 2013, 2014, 2015, 2016 and 2017 are 0.000438, 0.004031, 2.46×10^{-5}, 6.23×10^{-11} and 0.00041, so actually there is evidence that even for individual year, the numbers of mass shootings happened each day may not follows poisson distribution.

So, should we then change our direction of analysis, that means may be we should focus on the numbers of deaths or injuries caused by mass shootings in one day. However, after discussion, our group denied it. But from the data, we found the numbers of deaths or injuries caused by mass shootings in one day vary from 4 to 500! And according to the GVA's definition of mass shooting, this number cannot go down to 0,1,2,3. So obviously it's unreasonable to test the numbers of deaths or injuries caused by mass shootings. Also, through our common sense, the numbers of deaths or injuries depend on the circumstances of the crime scene.

But, should we draw a conclusion that the occurrence of mass shooting absolutely has nothing to do with the Poisson distribution? The answer is no.

Because we found the P value of individual year is much larger than the P value of years from 2013 to 2017. So we guessed that the numbers of mass shootings happened in one day follow a poisson distribution only in a short time period, like a month, or a week, that means the parameter k of the poisson distribution may be different from year to year, month to month, or week to week, so for a longer time period, since the parameter k of the poisson distribution is changing, it's impossible for us to determine a single parameter k for a long period. To see this, we continued our analysis.

3.4 Distribution of mass shooting on weekdays and months

This time, we analyzed the occurrence of a mass shooting on weekdays and months, and we got the figure below:

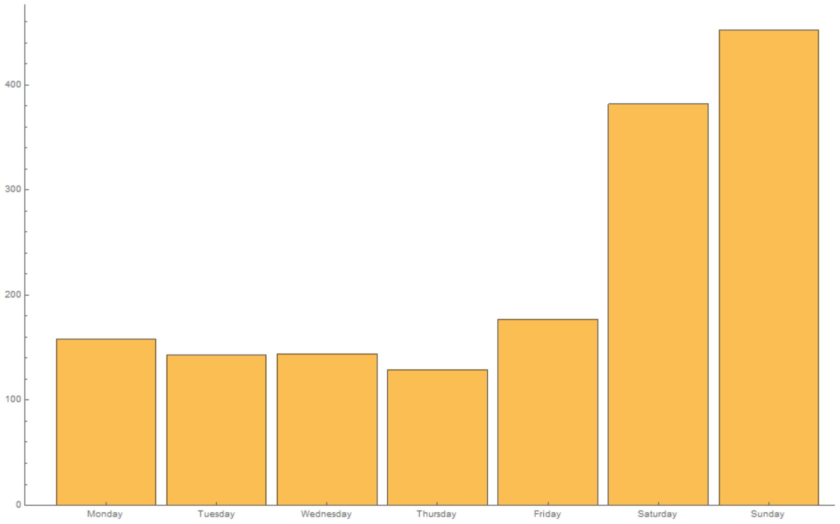


Figure 2: Occurrence of a mass shooting on weekdays

It's obvious that most mass shooting occur on Saturday and Sunday, so it depends on the weekdays.

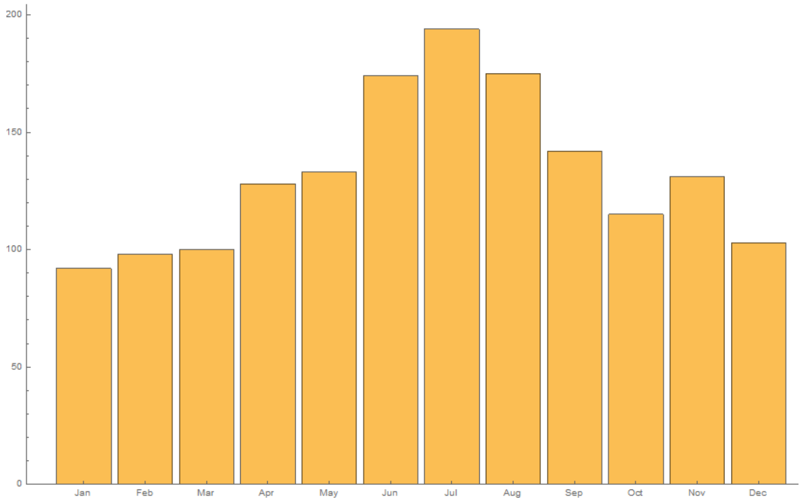


Figure 3: Occurrence of a mass shooting on months

It's also obvious that more mass shooting occur in summer, so it also depends on the months.

3.5 Analysis of mass shooting distribution on weekdays

We were interested in this phenomenon, and we wanted to do some more precise calculations to prove the occurrence of a mass shooting does depend on the weekdays. And we wanted to figure out whether this phenomenon could be reason why we failed to accept that the mass shootings follow a Poisson distribution in the previous analysis.

3.5.1 Whether Mass shootings follow Poisson Distribution

Data does not Follow Poisson Distribution Using the software we get the average number of shootings for each day in the week as shown below:

Day	Average Number of Shooting
Sunday	1.73
Monday	0.61
Tuesday	0.55
Wednesday	0.55
Thursday	0.49
Friday	0.68
Saturday	1.46

Table 19: Average Number of Shooting for Each Day

We find that the average number of shootings on weekend is much larger than those on the weekdays. However, if it follows a Poisson distribution, it should be the same for weekend and weekdays. In the following part, we would use FisherâĂŹs null hypothesis testing to reject the idea that it has same probability to have mass shootings on weekday and weekend. Because our data is so big, we can regard it as a normal distribution.

$$H_0 : \mu_1 = \mu_2$$

Type	Data Number n	Average Number of Shooting $\overline{X}$	Standard Deviation S
Weekday	1304	0.58	0.8089
Weekend	522	1.60	1.3740

Table 20: Average Number of Shooting for Weekday and Weekend

As the variances have a big difference, we would use the pooled T-Test with unequal variances. Apply Smith-Satterthwaite to get the γ.

$$\gamma = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1/n_1)^2}{n_1-1} + \frac{(S_2/n_2)^2}{n_2-1}} = 1247.7$$

we round it down to $\gamma = 1247$,so

$$T_\gamma = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = -8.805$$

From the table we get $t_{0.025,300} = 1.968$ so we can see $T_\gamma >> t_{0.025,300} > t_{0.025,1247}$ so we can reject H_0 at the 0.05 level of significance.

By the FisherâĂŹs null hypothesis testing, we can see the possibility for mass shootings in weekday and weekend is not same, so it would reject the idea that the number of shootings daily follows a Poisson distribution. Because if it follows a Poisson distribution, it would have same probabity to have mass shootings on weekday and weekend

Whether Weekly Data Follow Poisson Distribution Now we had proved the occurrence of a mass shooting depends on weekdays, then it's no meaning for us to analyze the data daily. So the best way was to analyze the data weekly.

If we consider the data weekly, because the week starts on Sunday,we can get 260 full weeks form the date, we can get the results as below.

Data Number n	Average Number of Shooting $\overline{X}$	Standard Deviation S
260	6.07	5.4448

Table 21: Number of Shooting for Weeks

To test if it follows Poisson Distribution, we would use PearsonâĂŹs chi-squared goodness-of-fit test. We rearrange our observed and expected number of mass shooting by combining into three adjacent categories. We would use the Poisson distribution with k= 6.07.

H_0: the weekly number of mass shootings follows a Poisson distribution with the parameter $k = 6.04$.

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	53	37.7
$[4, 8)$	136	153.2
$[8, \infty)$	71	69.1

Table 22: Categorie for Number of Shooting for Weeks

Then we can get

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 8.17$$

From the table we get $\chi^2_{0.05,1} = 3.84$ so we can see $X^2 >> \chi^2$ so we should reject H_0 at the 0.05 level of significance.

If we only consider the individual year, we let 52 weeks to become a year and get the following results; (note: here the year is not same as the year we usually discussed)

Year	Data Number n	Average Number of Shooting $\overline{X}$	Standard Deviation S
1	52	4.81	2.6645
2	52	5.19	2.6423
3	52	6.42	3.0828
4	52	7.27	3.7632
5	52	6.65	2.8414

Table 23: Average Number of Shooting for Week by Year

If we see it yearly, we would find the standard deviation in each individual year is much smaller than we consider it as a whole. Here we would test the first year as an example to see whether it following a Poisson distribution in each individual year.

H_0: the weekly number of mass shootings follows a Poisson distribution with the parameter $k = 4.81$.

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	16	15.2
$[4, 8)$	29	30.8
$[8, \infty)$	7	6.0

Table 24: Categorie for Number of Shooting for First 52 Weeks

Then we can get

$$X_1^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 0.33$$

From the table we get $\chi^2_{0.05,1} = 3.84$ so we can see $X_1^2 < \chi^2$ so we should not reject H_0 at the 0.05 level of significance.

Similarly, we can also not reject the idea the number of mass shootings for week follows a Poisson distribution in each individual year except for fourth 52 Weeks which is 2016.

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	14	12.5
$[4, 8)$	29	31.5
$[8, \infty)$	9	8.0

Table 25: Categorie for Number of Shooting for Second 52 Weeks

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	6	6.1
$[4, 8)$	31	29.5
$[8, \infty)$	15	16.4

Table 26: Categorie for Number of Shooting for Third 52 Weeks

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	9	3.6
$[4, 8)$	23	25.5
$[8, \infty)$	20	22.9

Table 27: Categorie for Number of Shooting for Fourth 52 Weeks

Categorie	Actuall Number of Weeks	Expected Number of Weeks
$[0, 4)$	8	5.3
$[4, 8)$	24	28.5
$[8, \infty)$	20	18.2

Table 28: Categorie for Number of Shooting for Fifth 52 Weeks

Then we can get

$$X_2^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 0.52$$

$$X_3^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 0.20$$

$$X_4^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 8.85$$

$$X_5^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i} = 2.28$$

To conclude, we would say the number of mass shootings for weeks on a large time scale does not follow a Poisson distribution. But for a relative small time period, like one year, the number appear to follow the poisson distribution

It may because the k for the Poisson distribution is always changing, it may be affected by the economic or social security, it needs further analysis to give a more concrete result. But for a short time, the k for Poisson is relative stable.

3.6 Weeks with n Shootings

3.6.1 Predicted and Actual Number of Weeks with n Shootings

By the data above we can draw the actual graph like below.

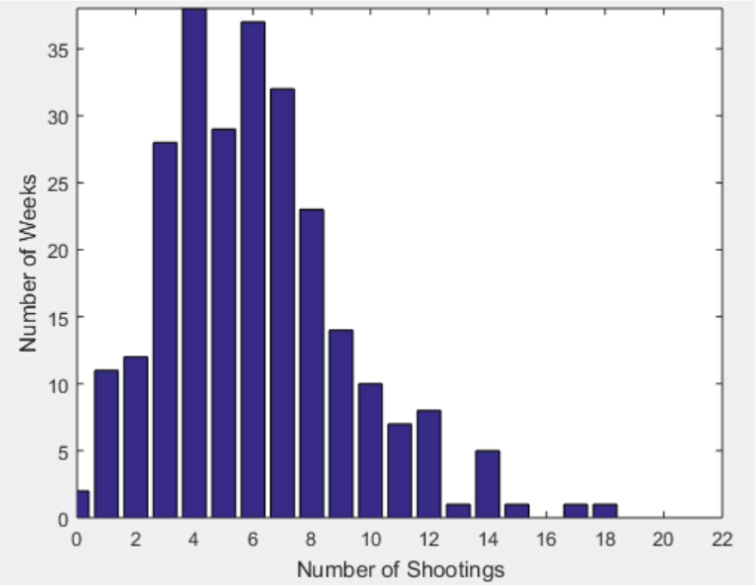


Figure 4: Actual Number of Weeks with n Shootings

Since we believe the number for mass shooting for long time period does not follow the Poisson distribution, here we would calculate the five years separately then add it together, and we get the following graph for the predicted number.

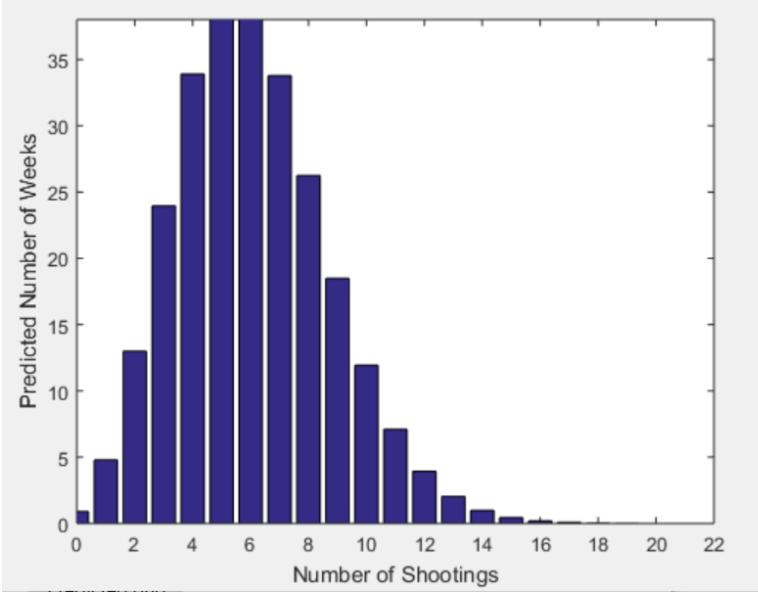


Figure 5: Predicted Number of Weeks with n Shootings with Poisson Distribution

3.7 Study of confidence interval for k

Then, we wanted to find the confidence interval for the parameter k based on the weekly data of the years 2013 to 2017.

At first, we need to get the expression of the confidence interval[3].

Based on the method of maximum likelihood, we use $\bar{X}$ to serve as the estimator $\hat{k}$ for k, which is proved in the lecture slide. As one essential assumption for a Poisson distribution is that "the number of 'arrivals' during non-overlapping time intervals are independent", we can know that

$$E[\bar{X}] = k, Var\bar{X} = k/n.$$

In our case, the sample size is large enough for us to reasonably assume that $\bar{X}$ follows a normal distribution with $E[\bar{X}] = k, Var\bar{X} = k/n$. Accordingly, the statistic $Z = \frac{\bar{X}-k}{\sqrt{k/n}}$ follows a standard normal distribution. It follows that a $100(1-\alpha)\%$ confidence interval for k is

$$\bar{X} \pm z_{\alpha/2}\sqrt{k/n}$$

which is identical to the form that

$$\hat{k} \pm z_{\alpha/2}\sqrt{k/n}$$

Nevertheless, the confidence interval depends on the parameter k we hope to estimate. We can solve this problem by replacing k by $\hat{k}$ and the interval becomes

$$\hat{k} \pm z_{\alpha/2}\sqrt{\hat{k}/n}$$

In this way, we actually replace σ by S, so we need to change $z_{\alpha/2}$ to $t_{\alpha/2}$. Notwithstanding, considering that our sample size is large enough for the central limit theorem to hold, it is rational for us to neglect the very subtle difference between $z_{\alpha/2}$ and $t_{\alpha/2}$ so that our solution can be simplified.

Using the weekly data of the years 2013 to 2017, we have can know that there are 262 weeks and the total number of mass shootings is 1585. Thus, we have:

$$\hat{k} \pm z_{\alpha/2}\sqrt{\hat{k}/n} = [5.75, 6.35].$$

□

3.8 Study of weekly data of Jan-Jun 2018 for Poisson distribution

Now we know that the weekly data of the years 2013 to 2017 follow a Poisson distribution. And we wanted to test the weekly data of year 2018 to further confirm our analysis.

Using the weekly data of January to June 2018, we can know that there are 26 weeks and the total number of mass shootings is 157 in that period. Hence, the estimate for k is $\hat{k}_{2018} = \bar{X} = 157/26 = 6.04$, which falls into the confidence interval calculated above. The data are shown in the following table:

# Mass Shootings	0	1	2	3	4	5	6	7	8
Observed Frequency	0	1	3	1	4	7	2	0	3
# Mass Shootings	9	10	11	12	13	14	15	16	17
Observed Frequency	0	2	1	1	0	0	0	0	1

We can calculate the corresponding expected frequencies by $E_x = nP[X = x] = n\frac{e^{\hat{k}}\hat{k}^x}{x!}$:

# Mass Shootings	0	1	2	3	4	5	6	7	8
Expected Frequency	0.06	0.38	1.13	2.28	3.44	4.15	4.18	3.60	2.72
# Mass Shootings	9	10	11	12	13	14	15	16	≥ 17
Expected Frequency	1.82	1.10	0.60	0.30	0.14	0.06	0.02	0.01	0.01

To satisfy the two criteria of the Pearson statistic, we need to rearrange our observed and expected frequencies by combining adjacent categories. However, the maximum number of categories is only 3 (if there are 4 categories, no way of division will be appropriate). The revised table is shown as follows:

Mass-Shooting Category i	[0,4]	[5,7]	[8,$+\infty$)
Observed Frequency O_i	9	9	8
Expected Frequency E_i	7.29	11.93	6.78

We need to test the null hypothesis H_0: the weekly number of mass shootings follows a Poisson distribution with the parameter $k = 6.04$.

This is in equivalence to the test H_0: the weekly number of mass shootings follows a categorical distribution with parameters $(\frac{7.29}{26}, \frac{11.93}{26}, \frac{6.78}{26})$. For $N = 3$ categories, the statistic

$$X^2 = \sum_{i=1}^{N} \frac{(O_i - E_i)^2}{E_i}$$

follows a chi-squared distribution with $N - 1 - m = 1$ degree of freedom. Let $\alpha = 0.05$, and we will reject H_0 if $X^2 > \chi^2_{0.05,1} = 3.84$. We have:

$$X^2 = \frac{(9-7.29)^2}{7.29} + \frac{(9-11.93)^2}{11.93} + \frac{(8-6.78)^2}{6.78} = 1.34 < 3.84 = \chi^2_{0.05,1}$$

Hence, we fail to reject H_0 at the 5% level of significance. There is no evidence that the weekly data of January to June 2018 do not follow a Poisson distribution.

□

3.9 Data plot 2018 in a single graph along with the prediction intervals

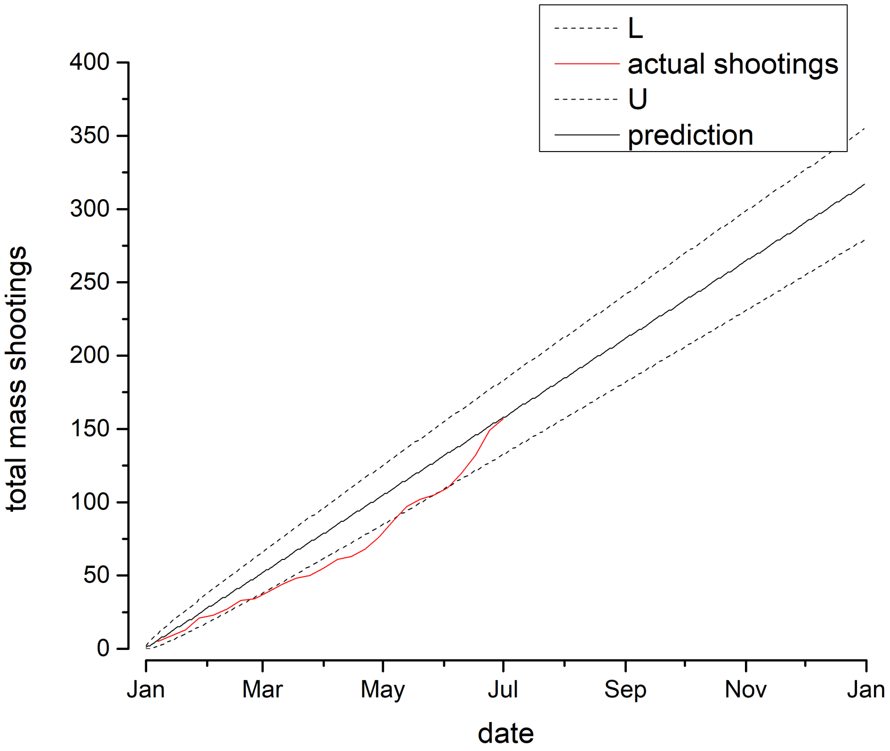


Figure 6: Data 2018 with the prediction intervals

[2]Suppose we have two discrete random variables X,Y. Suppose X counts a certain event in a sample of total size n from a Poisson distribution with mean λ. Then X $\sim$ Poisson(nλ). Suppose Y is the future counts from the same Poisson distribution from a sample of total size m. Then Y $\sim$ Poisson(mλ). Let $\widehat{\lambda}$ = X/n, $\widehat{Y}$ = m$\widehat{\lambda}$ ($\widehat{Y}$=0.5m/n when X = 0)and $\widehat{var}$($\widehat{Y}$-Y) = m$\widehat{Y}$(1/n+1/m). Then ($\widehat{Y}$ - Y)/$\sqrt{\widehat{var}(\widehat{Y}-Y)}$ follows a standard normal distribution asymptotically.

To get the prediction interval [L,U] where L,U satisfy $P_{X,Y}$(L$\leq$Y$\leq$U)$\geq$1-2α. Since ($\widehat{Y}$ - Y)/$\sqrt{\widehat{var}(\widehat{Y}-Y)}$ $\sim$ N(0,1), a 100(1-2α)% two-sided confidence interval for Y is given by $\widehat{Y}$ $\pm$ $z_{1-\alpha}\sqrt{\widehat{var}(\widehat{Y}-Y)}$, where $z_{1-\alpha}$ is the value that P[Z$\leq z_{1-\alpha}$]]=1-α, Z$\sim$N(0,1). Since we are only considering integer values of Y, we can see that [L,U] is given by

[$\lfloor\widehat{Y}$- $z_{1-\alpha}\sqrt{\widehat{var}(\widehat{Y}-Y)}\rfloor$,$\lceil\widehat{Y}$+$z_{1-\alpha}\sqrt{\widehat{var}(\widehat{Y}-Y)}\rceil$].

4 Conclusion

After all these tests, our group finally got a conclusion.

The occurrence of mass shootings in the United States depends on weekdays, and it is more likely to happen in weekends. This makes sense, since in weekends people have more free time, and the shooter may have more time to prepare the assault. Since it depends on weekdays, we cannot analyze the daily data. After analyzing the weekly data, we found the numbers of occurrence of mass shooting in one week may follow a Poisson distribution in a short period, like a single year, but not in a long period.

And the reason may be that the occurrence of mass shooting is not a random event. People may tend to shoot when the social security is not so good, it depends on many social factors, so in a short period, when the social and national conditions don't change a lot, the the numbers of occurrence of mass shooting in one week may follow a Poisson distribution, but in a long period, since the society and country are changing, which may cause the parameter k of the Poisson distribution to change, that means at different years, the numbers of occurrence of mass shooting in one week may follow Poisson distributions with different parameter k,so we can't simply use Poisson distribution to test the data in a long period.

Our conclusion is helpful to prevent mass shooting. Because the different parameter k at different years in some way may reflect the social security at that year, lower k reflects better social security, while higher k reflects the president should try to strengthen the social security. Also, the higher probability of mass shooting happening in weekends may give the government some hints to prevent mass shootings.

5 Reference

1 Gun Violence Archive. Mass shootings. http://www.shootingtracker.com/

2 K. Krishnamoorthy and Jie Peng. Improved closed-form prediction intervals for binomial and poisson distributions. Journal of Statistical Planning and Inference, 141(5):1709 âĂŞ 1718, 2011.

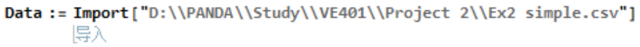
3 VV Patil and HV Kulkarni. Comparison of confidence intervals for the poisson mean: some new aspects. REVSTATâĂŞStatistical Journal, 10(2):211âĂŞ227, 2012.

4 D. Spiegelhalter and A. Barnett. London murders: a predictable pattern? Significance, 6(1):5âĂŞ8, 2009. onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2009.00334.x/abstract [Online; accessed 5-July-2015]

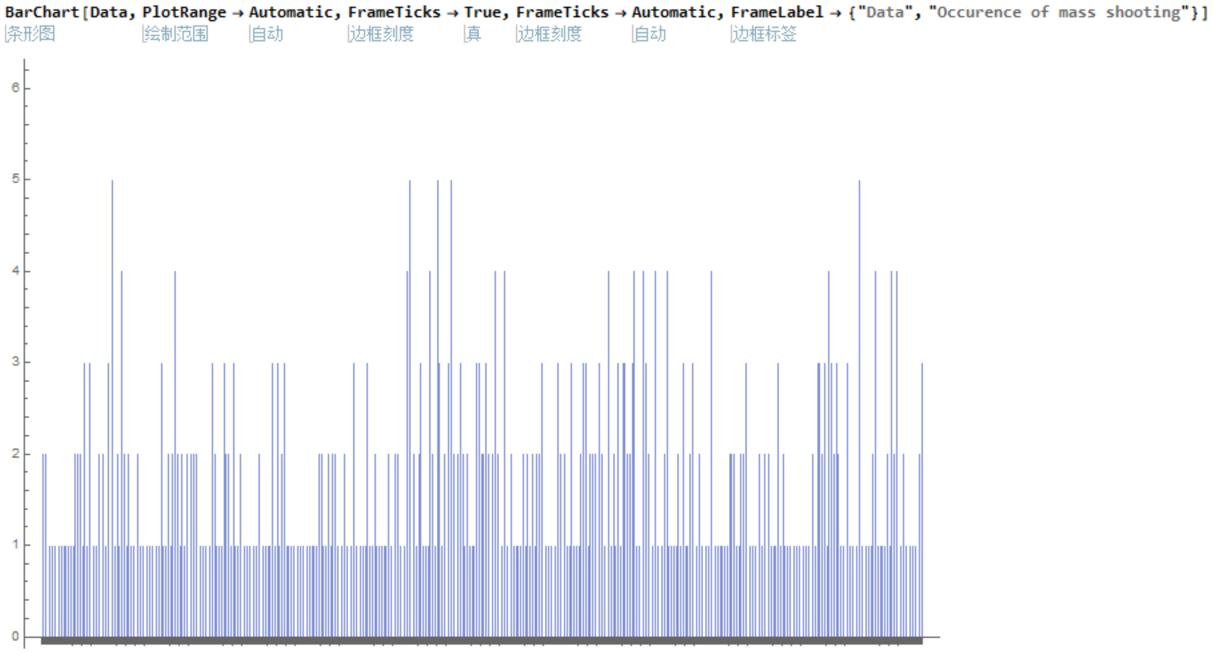
5 General Methodology of GVA. http://www.gunviolencearchive.org/methodology

6 Appendix

- Mathematica code for Ex2



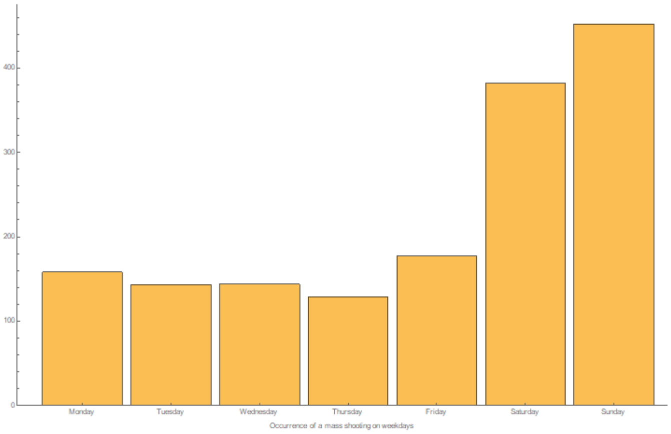
```
Data := Import["D:\\PANDA\\Study\\VE401\\Project 2\\Ex2 simple.csv"]
```



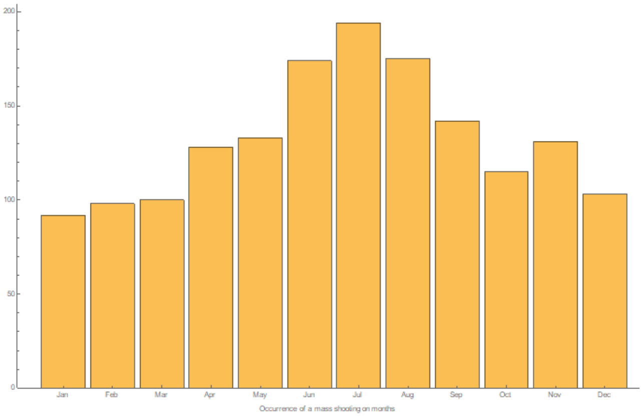
```
BarChart[Data, PlotRange → Automatic, FrameTicks → True, FrameTicks → Automatic, FrameLabel → {"Data", "Occurence of mass shooting"}]
```

- Mathematica code for Ex4

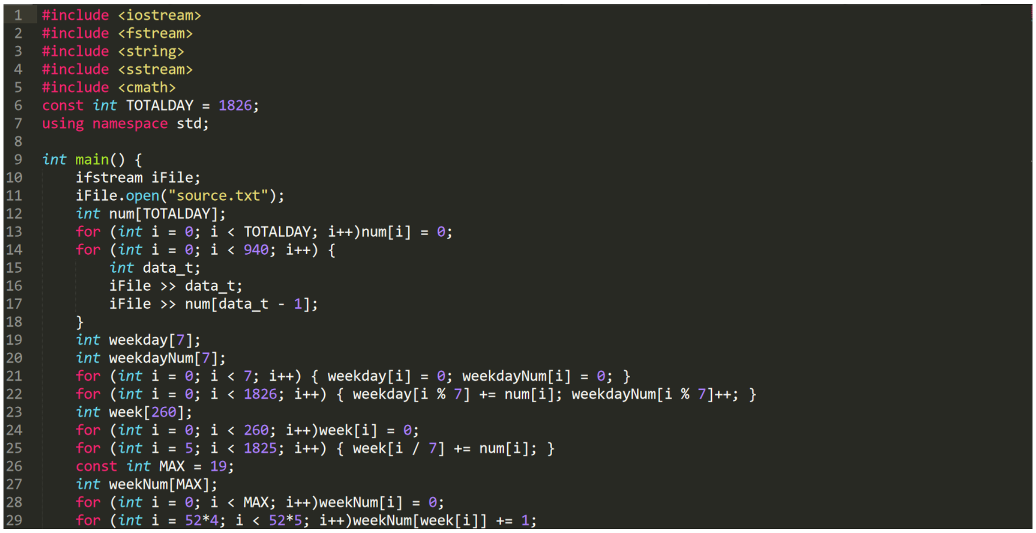
```
BarChart[{158, 143, 144, 129, 177, 382, 452}, ChartLabels → {"Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"}, PlotRange → Automatic, Frame → {True, True, False, False}, FrameTicks → Automatic, FrameLabel → {"Occurrence of mass shooting on weekdays"}]
条形图                                        图表标签                                                                                        绘制范围   自动      边框    真   真   假   假      边框刻度    自动      边框标签
```



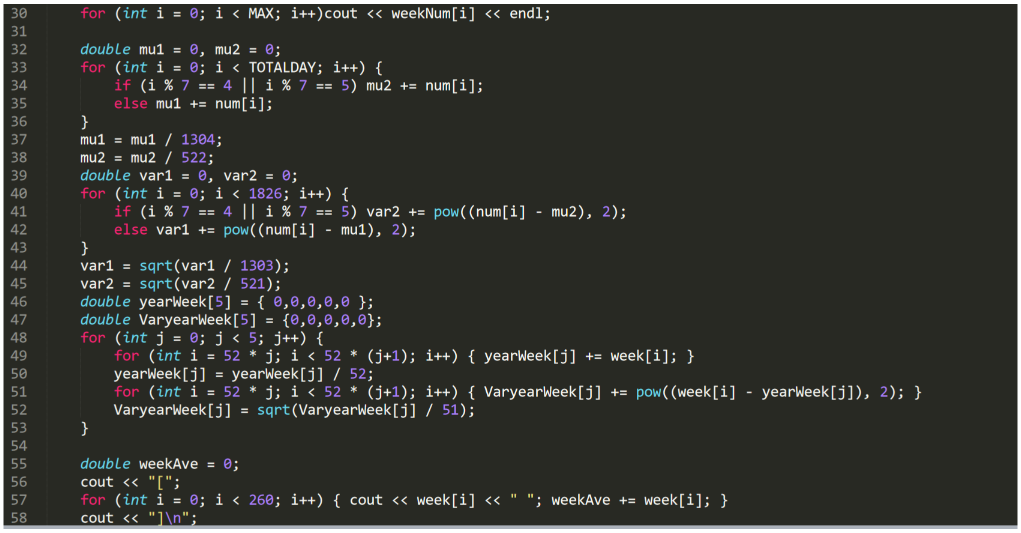
```
BarChart[{92, 98, 100, 128, 133, 174, 194, 175, 142, 115, 131, 103}, ChartLabels → {"Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"}, PlotRange → Automatic, Frame → {True, True, False, False}, FrameTicks → Automatic, FrameLabel → {"Occurrence of  mass shooting on months"}]
条形图                                                                  图表标签                                                                                       绘制范围   自动      边框    真   真   假   假      边框刻度    自动      边框标签
```



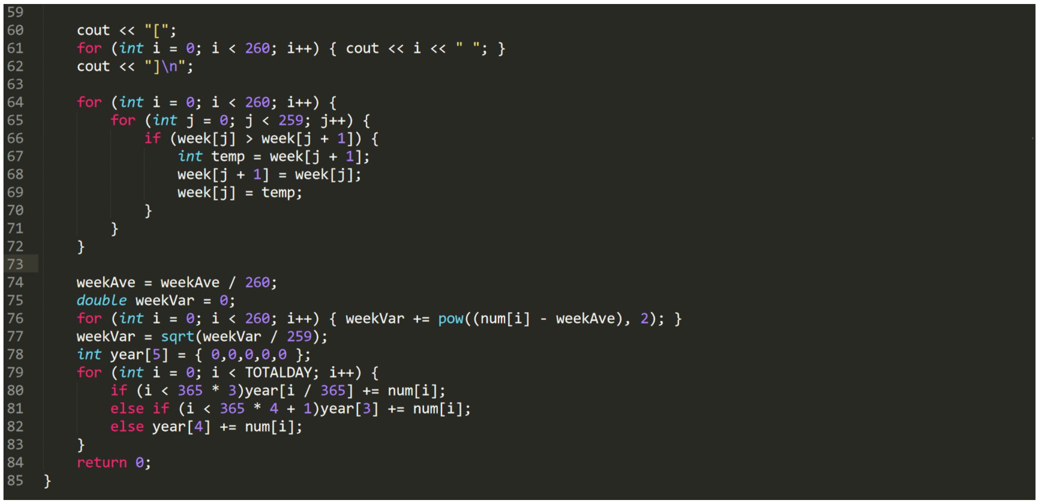
- C++ code for Ex5



```cpp
#include <iostream>
#include <fstream>
#include <string>
#include <sstream>
#include <cmath>
const int TOTALDAY = 1826;
using namespace std;

int main() {
    ifstream iFile;
    iFile.open("source.txt");
    int num[TOTALDAY];
    for (int i = 0; i < TOTALDAY; i++)num[i] = 0;
    for (int i = 0; i < 940; i++) {
        int data_t;
        iFile >> data_t;
        iFile >> num[data_t - 1];
    }
    int weekday[7];
    int weekdayNum[7];
    for (int i = 0; i < 7; i++) { weekday[i] = 0; weekdayNum[i] = 0; }
    for (int i = 0; i < 1826; i++) { weekday[i % 7] += num[i]; weekdayNum[i % 7]++; }
    int week[260];
    for (int i = 0; i < 260; i++)week[i] = 0;
    for (int i = 5; i < 1825; i++) { week[i / 7] += num[i]; }
    const int MAX = 19;
    int weekNum[MAX];
    for (int i = 0; i < MAX; i++)weekNum[i] = 0;
    for (int i = 52*4; i < 52*5; i++)weekNum[week[i]] += 1;
```

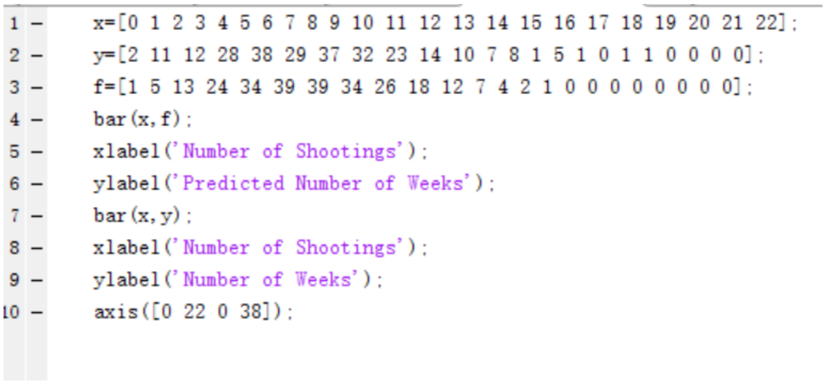


```
    for (int i = 0; i < MAX; i++)cout << weekNum[i] << endl;

    double mu1 = 0, mu2 = 0;
    for (int i = 0; i < TOTALDAY; i++) {
        if (i % 7 == 4 || i % 7 == 5) mu2 += num[i];
        else mu1 += num[i];
    }
    mu1 = mu1 / 1304;
    mu2 = mu2 / 522;
    double var1 = 0, var2 = 0;
    for (int i = 0; i < 1826; i++) {
        if (i % 7 == 4 || i % 7 == 5) var2 += pow((num[i] - mu2), 2);
        else var1 += pow((num[i] - mu1), 2);
    }
    var1 = sqrt(var1 / 1303);
    var2 = sqrt(var2 / 521);
    double yearWeek[5] = { 0,0,0,0,0 };
    double VaryearWeek[5] = {0,0,0,0,0};
    for (int j = 0; j < 5; j++) {
        for (int i = 52 * j; i < 52 * (j+1); i++) { yearWeek[j] += week[i]; }
        yearWeek[j] = yearWeek[j] / 52;
        for (int i = 52 * j; i < 52 * (j+1); i++) { VaryearWeek[j] += pow((week[i] - yearWeek[j]), 2); }
        VaryearWeek[j] = sqrt(VaryearWeek[j] / 51);
    }

    double weekAve = 0;
    cout << "[";
    for (int i = 0; i < 260; i++) { cout << week[i] << " "; weekAve += week[i]; }
    cout << "]\n";
```



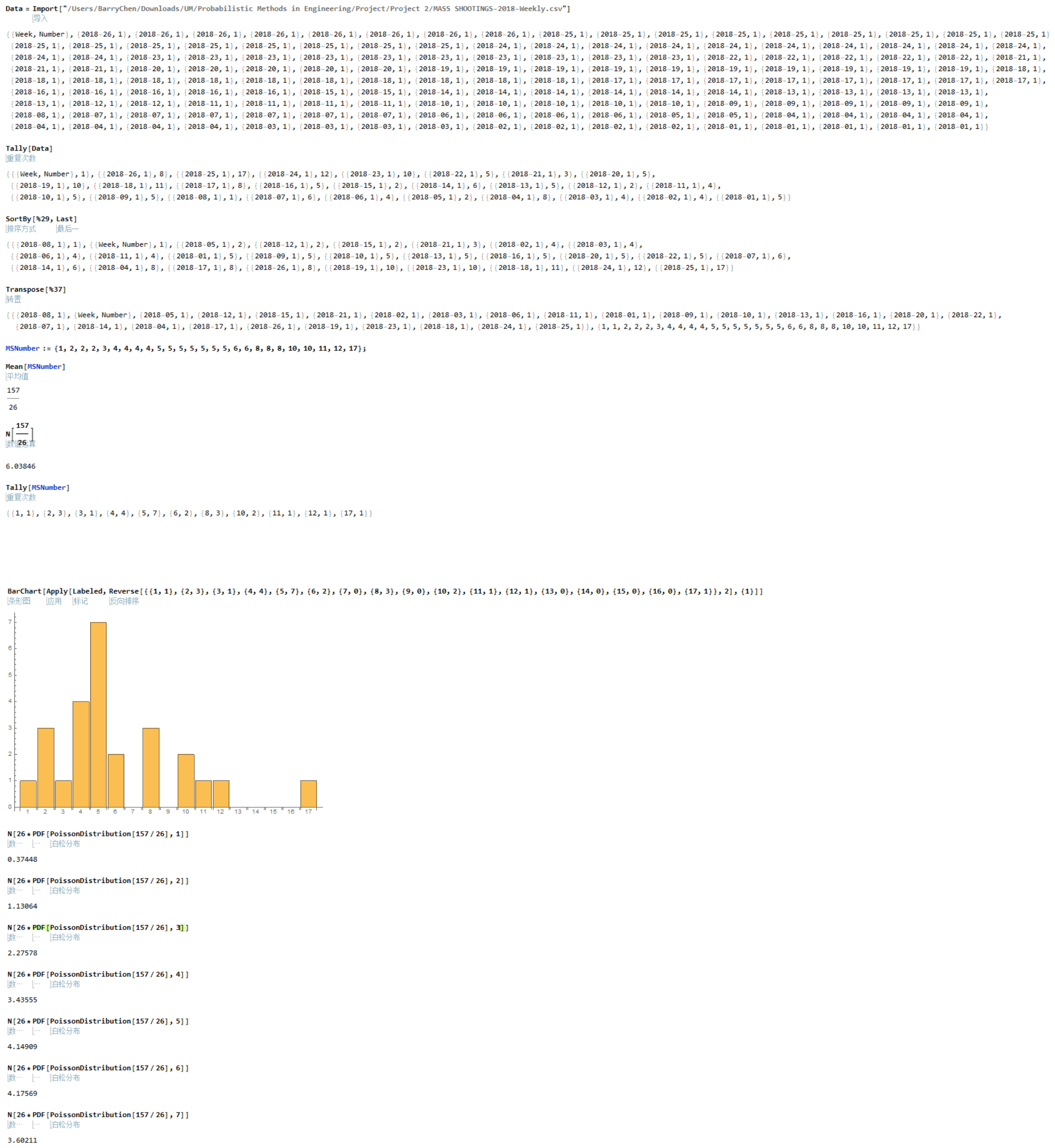
```

    cout << "[";
    for (int i = 0; i < 260; i++) { cout << i << " "; }
    cout << "]\n";

    for (int i = 0; i < 260; i++) {
        for (int j = 0; j < 259; j++) {
            if (week[j] > week[j + 1]) {
                int temp = week[j + 1];
                week[j + 1] = week[j];
                week[j] = temp;
            }
        }
    }

    weekAve = weekAve / 260;
    double weekVar = 0;
    for (int i = 0; i < 260; i++) { weekVar += pow((num[i] - weekAve), 2); }
    weekVar = sqrt(weekVar / 259);
    int year[5] = { 0,0,0,0,0 };
    for (int i = 0; i < TOTALDAY; i++) {
        if (i < 365 * 3)year[i / 365] += num[i];
        else if (i < 365 * 4 + 1)year[3] += num[i];
        else year[4] += num[i];
    }
    return 0;
}
```

- Matlab code for Ex6



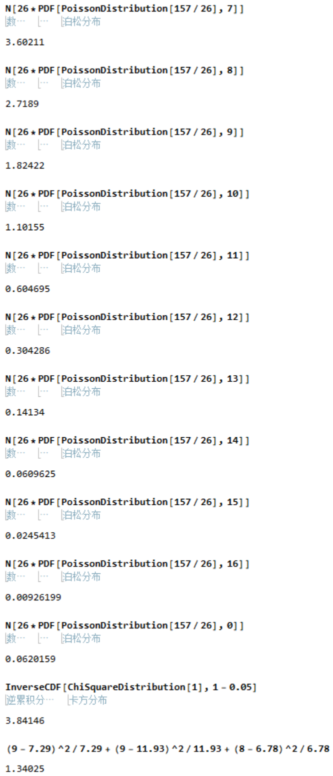
```
x=[0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22];
y=[2 11 12 28 38 29 37 32 23 14 10 7 8 1 5 1 0 1 1 0 0 0 0];
f=[1 5 13 24 34 39 39 34 26 18 12 7 4 2 1 0 0 0 0 0 0 0 0];
bar(x,f);
xlabel('Number of Shootings');
ylabel('Predicted Number of Weeks');
bar(x,y);
xlabel('Number of Shootings');
ylabel('Number of Weeks');
axis([0 22 0 38]);
```

- Mathematica code for Ex7 and Ex8



```
Data = Import["/Users/BarryChen/Downloads/UM/Probabilistic Methods in Engineering/Project/Project 2/MASS SHOOTINGS-2018-Weekly.csv"]
       导入

{{Week, Number}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-26, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}
 {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-25, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1}, {2018-24, 1},
 {2018-24, 1}, {2018-24, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-23, 1}, {2018-22, 1}, {2018-22, 1}, {2018-22, 1}, {2018-22, 1}, {2018-22, 1}, {2018-21, 1},
 {2018-21, 1}, {2018-21, 1}, {2018-20, 1}, {2018-20, 1}, {2018-20, 1}, {2018-20, 1}, {2018-20, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-19, 1}, {2018-18, 1},
 {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-18, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1}, {2018-17, 1},
 {2018-16, 1}, {2018-16, 1}, {2018-16, 1}, {2018-16, 1}, {2018-16, 1}, {2018-15, 1}, {2018-15, 1}, {2018-14, 1}, {2018-14, 1}, {2018-14, 1}, {2018-14, 1}, {2018-14, 1}, {2018-14, 1}, {2018-13, 1}, {2018-13, 1}, {2018-13, 1}, {2018-13, 1},
 {2018-13, 1}, {2018-12, 1}, {2018-12, 1}, {2018-11, 1}, {2018-11, 1}, {2018-11, 1}, {2018-11, 1}, {2018-10, 1}, {2018-10, 1}, {2018-10, 1}, {2018-10, 1}, {2018-10, 1}, {2018-09, 1}, {2018-09, 1}, {2018-09, 1}, {2018-09, 1}, {2018-09, 1},
 {2018-08, 1}, {2018-07, 1}, {2018-07, 1}, {2018-07, 1}, {2018-07, 1}, {2018-07, 1}, {2018-07, 1}, {2018-06, 1}, {2018-06, 1}, {2018-06, 1}, {2018-06, 1}, {2018-05, 1}, {2018-05, 1}, {2018-04, 1}, {2018-04, 1}, {2018-04, 1}, {2018-04, 1},
 {2018-04, 1}, {2018-04, 1}, {2018-04, 1}, {2018-04, 1}, {2018-03, 1}, {2018-03, 1}, {2018-03, 1}, {2018-03, 1}, {2018-02, 1}, {2018-02, 1}, {2018-02, 1}, {2018-02, 1}, {2018-01, 1}, {2018-01, 1}, {2018-01, 1}, {2018-01, 1}, {2018-01, 1}}

Tally[Data]
重复次数

{{{Week, Number}, 1}, {{2018-26, 1}, 8}, {{2018-25, 1}, 17}, {{2018-24, 1}, 12}, {{2018-23, 1}, 10}, {{2018-22, 1}, 5}, {{2018-21, 1}, 3}, {{2018-20, 1}, 5},
 {{2018-19, 1}, 10}, {{2018-18, 1}, 11}, {{2018-17, 1}, 8}, {{2018-16, 1}, 5}, {{2018-15, 1}, 2}, {{2018-14, 1}, 6}, {{2018-13, 1}, 5}, {{2018-12, 1}, 2}, {{2018-11, 1}, 4},
 {{2018-10, 1}, 5}, {{2018-09, 1}, 5}, {{2018-08, 1}, 1}, {{2018-07, 1}, 6}, {{2018-06, 1}, 4}, {{2018-05, 1}, 2}, {{2018-04, 1}, 8}, {{2018-03, 1}, 4}, {{2018-02, 1}, 4}, {{2018-01, 1}, 5}}

SortBy[%29, Last]
排序方式      最后一

{{{2018-08, 1}, 1}, {{Week, Number}, 1}, {{2018-05, 1}, 2}, {{2018-12, 1}, 2}, {{2018-15, 1}, 2}, {{2018-21, 1}, 3}, {{2018-02, 1}, 4}, {{2018-03, 1}, 4},
 {{2018-06, 1}, 4}, {{2018-11, 1}, 4}, {{2018-01, 1}, 5}, {{2018-09, 1}, 5}, {{2018-10, 1}, 5}, {{2018-13, 1}, 5}, {{2018-16, 1}, 5}, {{2018-20, 1}, 5}, {{2018-22, 1}, 5}, {{2018-07, 1}, 6},
 {{2018-14, 1}, 6}, {{2018-04, 1}, 8}, {{2018-17, 1}, 8}, {{2018-26, 1}, 8}, {{2018-19, 1}, 10}, {{2018-23, 1}, 10}, {{2018-18, 1}, 11}, {{2018-24, 1}, 12}, {{2018-25, 1}, 17}}

Transpose[%37]
转置

{{{2018-08, 1}, {Week, Number}, {2018-05, 1}, {2018-12, 1}, {2018-15, 1}, {2018-21, 1}, {2018-02, 1}, {2018-03, 1}, {2018-06, 1}, {2018-11, 1}, {2018-01, 1}, {2018-09, 1}, {2018-10, 1}, {2018-13, 1}, {2018-16, 1}, {2018-20, 1}, {2018-22, 1},
  {2018-07, 1}, {2018-14, 1}, {2018-04, 1}, {2018-17, 1}, {2018-26, 1}, {2018-19, 1}, {2018-23, 1}, {2018-18, 1}, {2018-24, 1}, {2018-25, 1}}, {1, 1, 2, 2, 2, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 8, 8, 8, 10, 10, 11, 12, 17}}

MSNumber := {1, 2, 2, 2, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 6, 6, 8, 8, 8, 10, 10, 11, 12, 17};

Mean[MSNumber]
平均值

157
---
26

  157
N[---]
  26
数值运算

6.03846

Tally[MSNumber]
重复次数

{{1, 1}, {2, 3}, {3, 1}, {4, 4}, {5, 7}, {6, 2}, {8, 3}, {10, 2}, {11, 1}, {12, 1}, {17, 1}}


BarChart[Apply[Labeled, Reverse[{{1, 1}, {2, 3}, {3, 1}, {4, 4}, {5, 7}, {6, 2}, {7, 0}, {8, 3}, {9, 0}, {10, 2}, {11, 1}, {12, 1}, {13, 0}, {14, 0}, {15, 0}, {16, 0}, {17, 1}}, 2], {1}]]
条形图      应用   标记      反向排序
```

```
N[26 * PDF[PoissonDistribution[157 / 26], 1]]
数⋯  [⋯   泊松分布

0.37448

N[26 * PDF[PoissonDistribution[157 / 26], 2]]
数⋯  [⋯   泊松分布

1.13064

N[26 * PDF[PoissonDistribution[157 / 26], 3]]
数⋯  [⋯   泊松分布

2.27578

N[26 * PDF[PoissonDistribution[157 / 26], 4]]
数⋯  [⋯   泊松分布

3.43555

N[26 * PDF[PoissonDistribution[157 / 26], 5]]
数⋯  [⋯   泊松分布

4.14909

N[26 * PDF[PoissonDistribution[157 / 26], 6]]
数⋯  [⋯   泊松分布

4.17569

N[26 * PDF[PoissonDistribution[157 / 26], 7]]
数⋯  [⋯   泊松分布

3.60211
```



```
N[26 * PDF[PoissonDistribution[157 / 26], 7]]
数…   …   泊松分布

3.60211

N[26 * PDF[PoissonDistribution[157 / 26], 8]]
数…   …   泊松分布

2.7189

N[26 * PDF[PoissonDistribution[157 / 26], 9]]
数…   …   泊松分布

1.82422

N[26 * PDF[PoissonDistribution[157 / 26], 10]]
数…   …   泊松分布

1.10155

N[26 * PDF[PoissonDistribution[157 / 26], 11]]
数…   …   泊松分布

0.604695

N[26 * PDF[PoissonDistribution[157 / 26], 12]]
数…   …   泊松分布

0.304286

N[26 * PDF[PoissonDistribution[157 / 26], 13]]
数…   …   泊松分布

0.14134

N[26 * PDF[PoissonDistribution[157 / 26], 14]]
数…   …   泊松分布

0.0609625

N[26 * PDF[PoissonDistribution[157 / 26], 15]]
数…   …   泊松分布

0.0245413

N[26 * PDF[PoissonDistribution[157 / 26], 16]]
数…   …   泊松分布

0.00926199

N[26 * PDF[PoissonDistribution[157 / 26], 0]]
数…   …   泊松分布

0.0620159

InverseCDF[ChiSquareDistribution[1], 1 - 0.05]
逆累积分…   卡方分布

3.84146

(9 - 7.29) ^2 / 7.29 + (9 - 11.93) ^2 / 11.93 + (8 - 6.78) ^2 / 6.78

1.34025
```