

ARE FATAL POLICE SHOOTINGS IN THE UNITED STATES GETTING WORSE?

Tanrui Wu 518370910221

Taoyue Xia 518370910087

Xingyu Zhu 518370910023

2021-08-04

Introduction

On May 26, 2020, a 46-year-old African American man called George Floyd was killed during an arrest after Derek Chauvin, a Minneapolis Police Department officer, who knelt on Floyd's neck for 9 minutes and 29 seconds as three other officers looked on and prevented passers-by from intervening [1]. The case that shocked the United States, questioned the behavior of the US police and triggered an unprecedented scale of demonstrations across the country in the summer of 2020. It also arouse a national-wide discussion about fatal cases during enforcement of police. People questioned about whether there were more and more such fatal cases caused by police in recent years, and called for fair and impartial enforcement. In our project, we consider whether these shocking incidents in the media are evidence that things are really getting worse, and we decide to observe it from a certain angle: number of fatal police shootings in US each year.

Each year, There are thousands of fatal police shootings happening in the United States. So do they follow a predictable pattern? Does the number of shootings increase every year? In our project, we analyze the data of those shootings from 2015 to 2020 and try to find some rules behind them. Since the shootings are independent; happening one does not change the probability of when the next one will happen, and the shootings occur with an almost constant rate within a fixed interval of time. We make our assumption that the number of shootings per day follows a Possion distribution and try to test it during this project.

What means "fatal police shooting"? From the detailed information provided on the website, The Washington Post. Fatal force [2]. We can have a general idea of the meaning of the term "fatal police shooting". Among

the term, the "police" stands for not only on-duty police officers, but it also can be off-duty officers or deputies of the County sheriff. In the cases that fatal police shooting occur, most of the suspects were shot and killed immediately, but there are also people shocked by stun gun first and shot later. Also, they all show threat to the "police" to some extent, before being shot. And however the process was, they died in the end, which corresponds to the word, "fatal".

In this project, we first explain the meaning of "fatal", summarize the data, and visualize the data from 2015 to 2020. Then we test whether the number of shootings per day in the last 6 years follows a Poisson distribution through the hypothesis test, and calculate the confidence interval of the parameter k through bootstrap method, which is also the expected shooting number per day.

Data

The data we download from the website [3] records the information of every fatal police shooting in the U.S each day from January of 2015 to July of 2021. It includes the name, gender, age and race of the suspects, and record how he/she is killed, and the condition when the fatal shooting happened, for example, the threat level, whether he/she was armed, whether he/she showed sign of mental illness and so on. And most importantly, it contains the location and date of the cases.

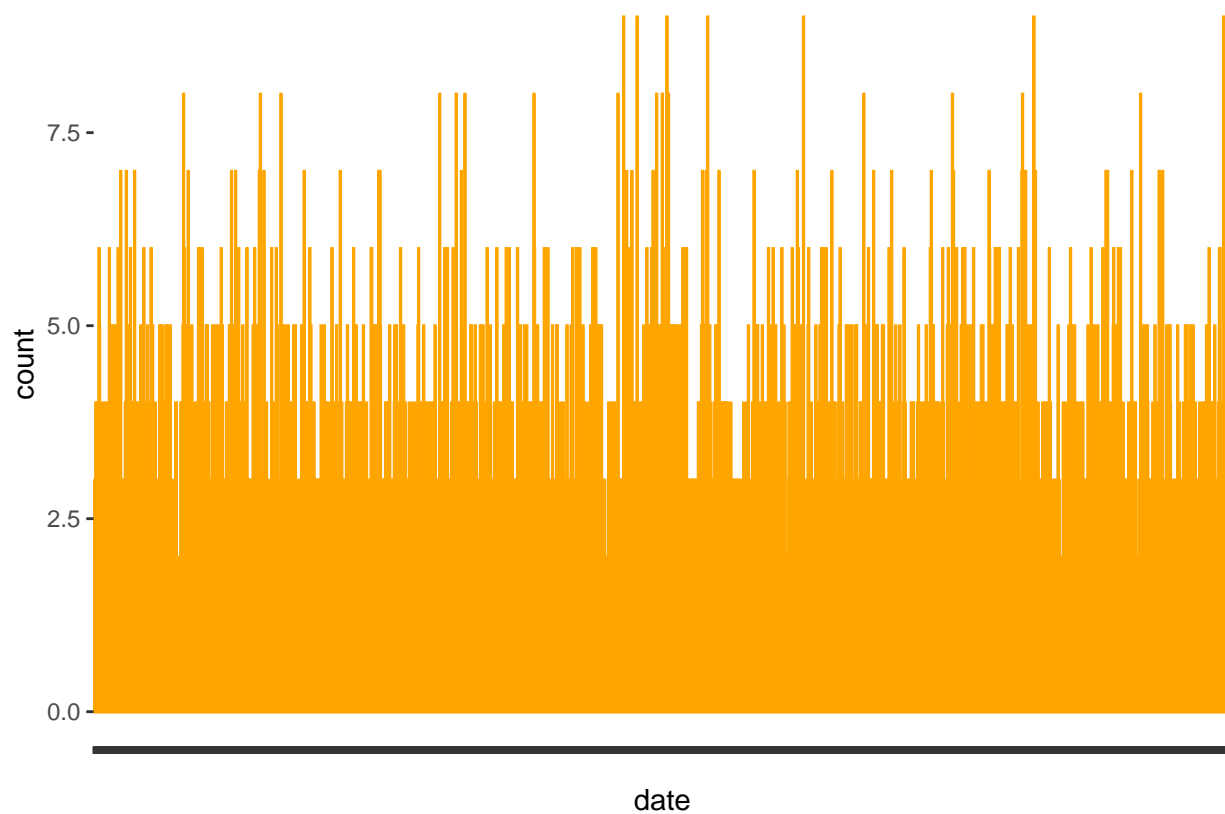
The file `fatal-police-shootings-data.csv` contains data about each fatal shooting in CSV format. Each row has the following variables:

- id: a unique identifier for each victim
- name: the name of the victim
- date: the date of the fatal shooting in YYYY-MM-DD format
- manner of death:
 - shot
 - shot and Tasered
- armed: indicates that the victim was armed with some sort of implement that a police officer believed could
 - undetermined: it is not known whether or not the victim had a weapon
 - unknown: the victim was armed, but it is not known what the object was

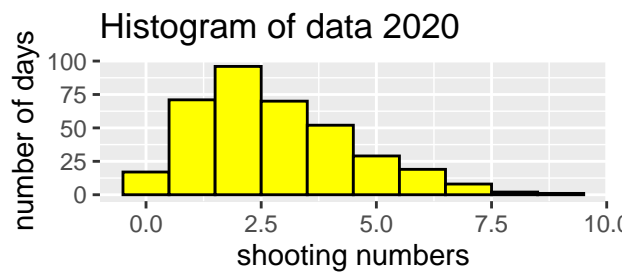
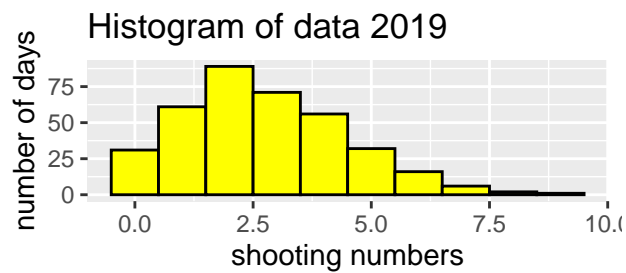
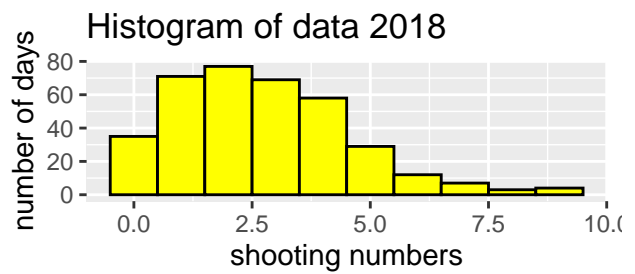
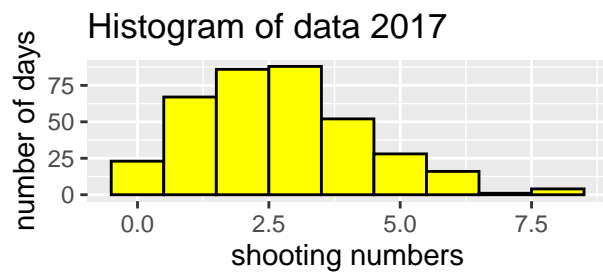
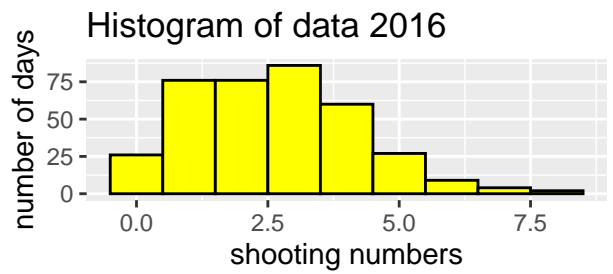
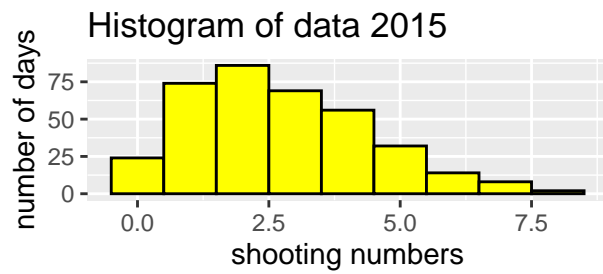
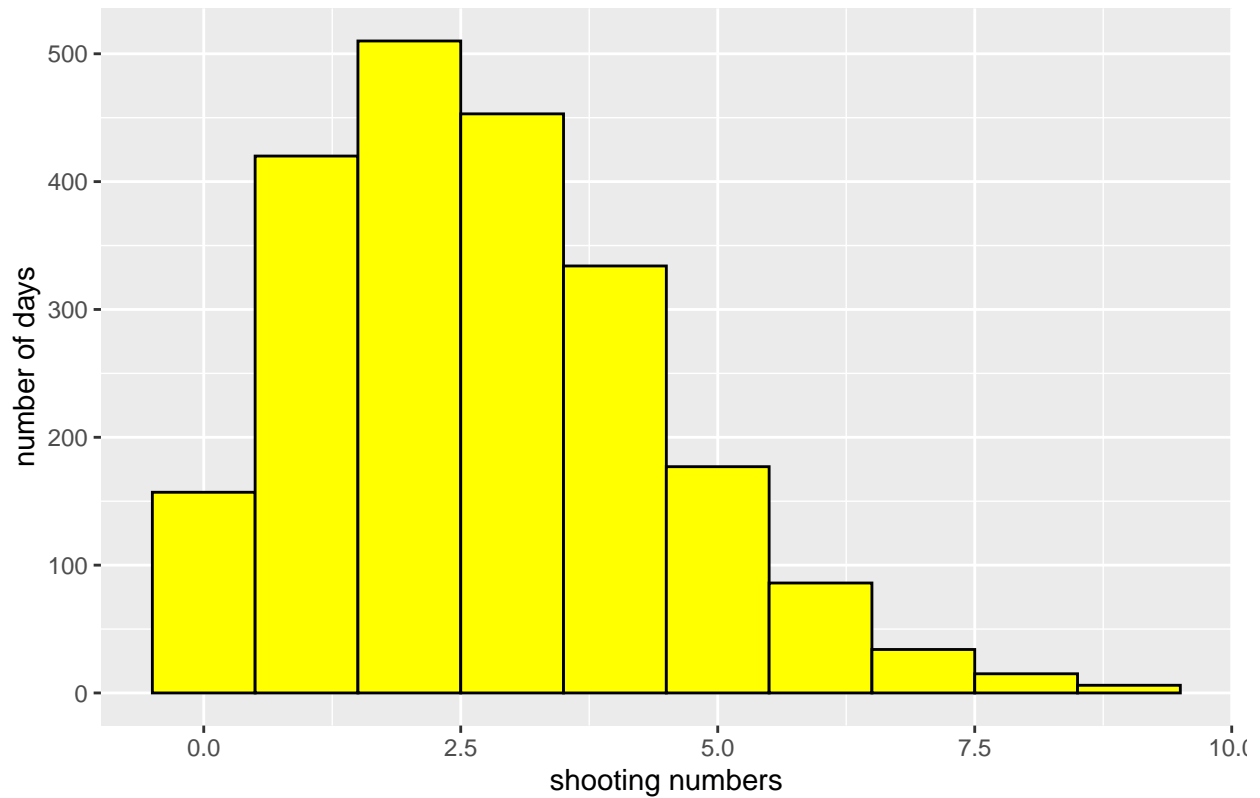
- unarmed: the victim was not armed
- age: the age of the victim
- gender: the gender of the victim. The Post identifies victims by the gender they identify with if reports indicate that it differs from their biological sex.
 - M: Male
 - F: Female
 - None: unknown
- race:
 - W: White, non-Hispanic
 - B: Black, non-Hispanic
 - A: Asian
 - N: Native American
 - H: Hispanic
 - O: Other
 - None: unknown
- city: the municipality where the fatal shooting took place. Note that in some cases this field may contain a county name if a more specific municipality is unavailable or unknown.
- state: two-letter postal code abbreviation
- signs of mental illness: News reports have indicated the victim had a history of mental health issues, expressed suicidal intentions or was experiencing mental distress at the time of the shooting.
- threat level: The threat level column was used to flag incidents for the story by Amy Brittain in October 2015.
- flee: News reports have indicated the victim was moving away from officers
 - Foot
 - Car
 - Not fleeing

- body camera: News reports have indicated an officer was wearing a body camera and it may have recorded some portion of the incident.
- latitude and longitude: the location of the shooting expressed as WGS84 coordinates, geocoded from addresses.
- is geocoding exact: reflects the accuracy of the coordinates.

To have a general impression of the data from 2015 to 2020, we first plot the number of fatal police shooting per day from 2015 to 2020,



Histogram of all data from 2015 to 2020



From the diagram, we can see that the number of fatal police shooting per day varies from 0 to 9. There are only few days that no shooting happened. The shape of “Histogram of data” is very similar to Poisson distribution. However, it is hard to find any obvious issues directly from the diagram, so we need to do further analysis in detail.

Methods

Denote X as the number of fatal shootings happened each day. First we want to estimate the value of average fatal shootings per day \bar{X} . Since sample mean is an unbiased estimator, we use following estimator:

$$\hat{k} = \bar{X} = \frac{1}{n} \sum_{i=1}^n (X_i),$$

where n is the number of days in a year. If the value of \hat{k} didn't grow significantly in 2020, there is no evidence that fatal shooting is more frequent. We will use bootstrap instead of Monte Carlo method to give a 95% confidence interval for \hat{k} . This is because it is not realistic to let time go back and draw multiple samples for the number of shooting per day, making it hard for us to use Monte Carlo methods. Using bootstrap, we can estimate the sampling distribution of \hat{k} by drawing multiple samples from the original sample.

What we want to study after that is the distribution of X . If the distribution of X in 2020 didn't differ significantly from distributions in past few years, there is no evidence that the police had changed their gun-using habits. From Fig.2, we find that the data shown by the histogram fit a Poisson distribution to some extent, which motivate us to use hypothesis test to prove our view.

Now we give the null hypothesis that the number of everyday fatal shootings follows a Poisson distribution with parameter \hat{k} . The alternative hypothesis is that the number of everyday fatal shootings follows a Poisson distribution with parameter $\hat{k} + 0.5$.

$$H_0 : X \sim \text{Poisson}(k), k = \hat{k}$$

$$H_1 : X \sim \text{Poisson}(k), k \geq \hat{k} + 0.5$$

After that, referring to Goodness-of-Fit Test for a discrete distribution, we can use the Person statistic X_{k-1}^2 such that:

$$T(X)_1 = \sum_{i=0}^k \frac{(E_i - O_i)^2}{E_i},$$

to test whether certain data follows a Poisson distribution. Here, E_i is the expected number of days at which

i fatal shootings happened, while O_i is the observed number of days at which i fatal shootings happened. The value of E_i can be expressed as $E_i = n \cdot P[X = i]$. Monte Carlo methods will be applied to obtain a rejection region for this test statistic and to calculate the power of the test. We can decide whether to reject the null hypothesis by comparing the computed $T(X)_1$ and the critical value we get by using Monte Carlo method.

We are also interested in the ratio of mean to variance of the data,

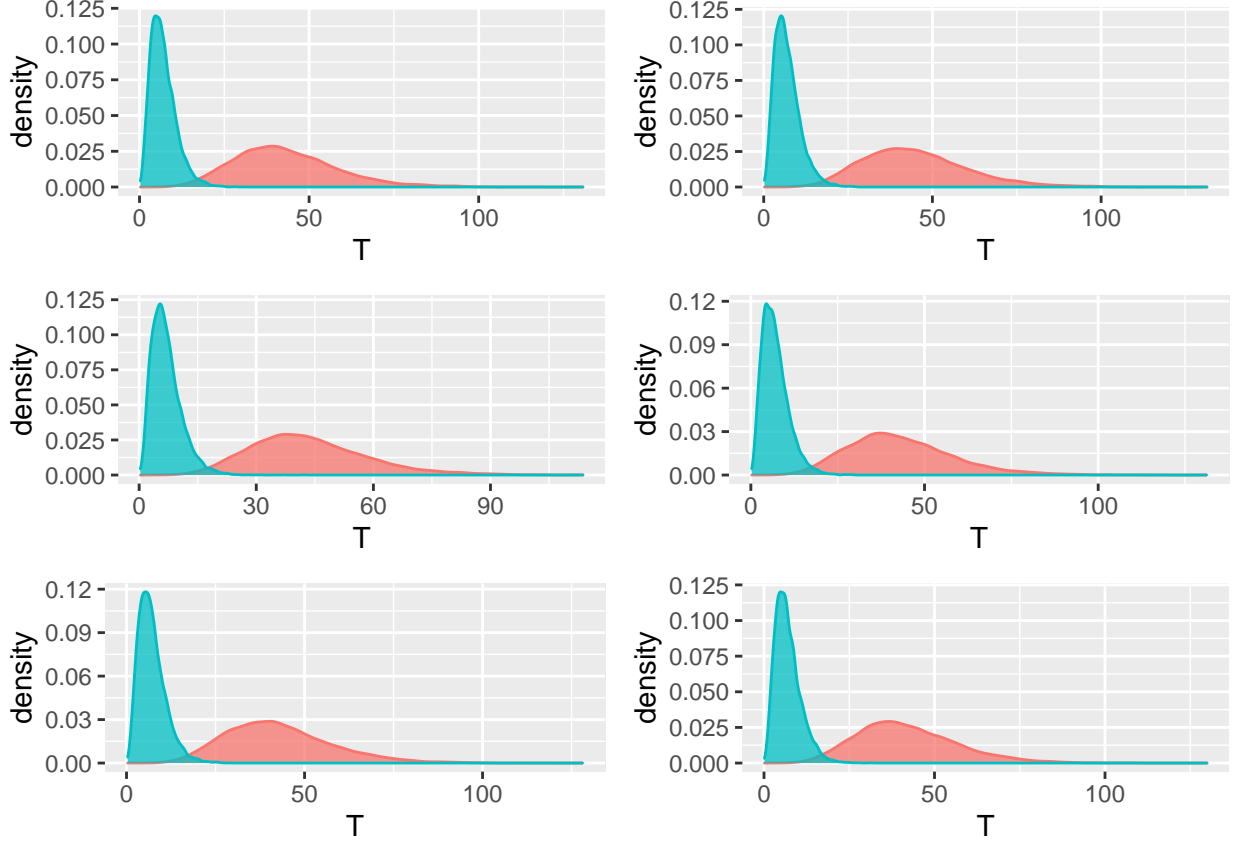
$$T(X)_2 = \frac{\bar{X}}{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2},$$

since the mean equals the variance for data following Poisson distribution, which will be proved by Monte Carlo method. We will estimate the ratio of mean to variance of our data and provide a 95% confidence interval for this value using bootstrap. We expect such confidence interval to include 1, which indicates evidences that our fatal shooting data may follow a Poisson distribution.

Simulation

We will randomly generate 10000 samples. Each sample is drawn from Poisson distribution with parameter \hat{k} . Each sample size equals to number of days one year. For each sample, the X_{k-1}^2 statistic will be calculated and the resulting 10000 values will form a sample distribution of X_{k-1}^2 . The 2.5% and 97.5% quantile will be the critical values for our test.

After that, we draw the six density plots corresponding to the years from 2015 to 2020, which is shown below.



After that, we calculate the critical regions as the interval between 2.5% and 97.5% quantile of each data set.

We also generate another 10000 samples drawn from Poisson distribution with parameter $k^* = \hat{k} + 0.5$.

Following the same procedure, we can similarly calculate 10000 values of statistic X_{k-1}^{*2} .

Finally, we can get the power of our test as the probability of the 10000 X_{k-1}^{*2} falling in the rejection region $(-\infty, 2.5\% \text{ quantile}) \cup (97.5\% \text{ quantile}, \infty)$.

The rejection region and corresponding power is shown in the table below:

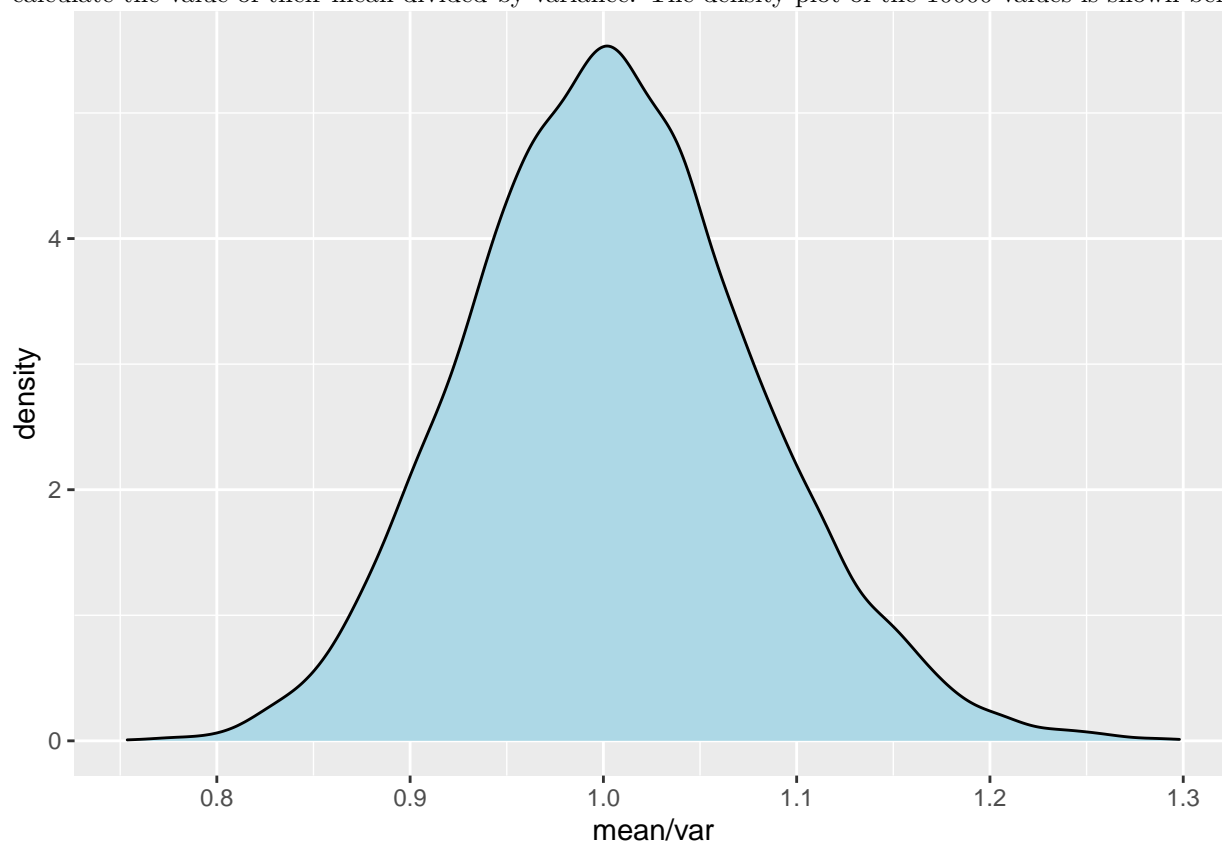
Year	Rejection Region	Power
2015	$(-\infty, 1.715) \cup (15.976, +\infty)$	0.9874
2016	$(-\infty, 1.639) \cup (16.042, +\infty)$	0.9922
2017	$(-\infty, 1.756) \cup (16.270, +\infty)$	0.9893
2018	$(-\infty, 1.660) \cup (16.129, +\infty)$	0.9879
2019	$(-\infty, 1.748) \cup (16.215, +\infty)$	0.9887
2020	$(-\infty, 1.678) \cup (15.865, +\infty)$	0.9877

Table 1: Results for Monte Carlo Hypothesis Test

From table 1, we find that the power of our test is extremely close to 1, which indicates that the simulation of our test is powerful and successful.

We then continue to prove that the mean and variance of a Poisson distribution are equal, and taking one sample into account is enough to show the relation between the two.

To ensure the consistency, we reuse the sample generated to simulate data in 2015. For the 10000 samples, we calculate the value of their mean divided by variance. The density plot of the 10000 values is shown below:



From the plot, we can obviously see that it follows a normal distribution and reaches the peak at 1. Still, to get the 95% confidence interval, we calculated the two-tailed quantile from 2.5% to 97.5%, which is (0.8699, 1.1630). We can simply find that 1 lies in the confidence interval, thus we have confidence that the value of mean divided by variance is 1.

Therefore, in the analysis part, we can calculate the real mean and variance. If their division is close to 1, then we will have great confidence to believe that the pattern of shootings follows a Poisson distribution.

Analysis

As discussed in previous part, we are interested in the average number of fatal shootings per day, k . After estimating the value of the parameter k by \hat{k} , we also want to give a confidence interval for \hat{k} . In previous part, we introduce one estimator for parameter k :

$$\hat{k} = \bar{X} = \frac{1}{n} \sum_{i=1}^n (X_i).$$

Suppose that $X_1^*, X_2^*, \dots, X_n^*$ is a sample randomly picked with replacement from the original sample which has n data. We calculate the statistic

$$\hat{k}^* = \bar{X}^* = \frac{1}{n} \sum_{i=1}^n (X_i^*)$$

with the newly-picked sample $X_1^*, X_2^*, \dots, X_n^*$.

We do such re-sampling 10000 times. Then the resulted 10000 \hat{k}^* values form the sampling distribution of \hat{k} .

We will then take the basic bootstrap confidence intervals, which is calculated as follows:

$$[2\hat{k} - \hat{k}_{0.975}^*, 2\hat{k} - \hat{k}_{0.025}^*],$$

the normal bootstrap confidence intervals, which is calculated as follows:

$$T \pm z_\alpha \sqrt{\frac{1}{B-1} \sum_{i=1}^B (T_i^* - \bar{T}^*)^2},$$

and the percentile bootstrap confidence intervals:

$$[T_{\alpha/2}^*, T_{1-\alpha/2}^*]$$

The result of the bootstrap confidence interval is shown below in Table 2.

Years	T_1 statistic	Normal	Basic	Percentile
2015	2.721	(2.552, 2.896)	(2.543, 2.893)	(2.548, 2.899)
2016	2.620	(2.448, 2.787)	(2.445, 2.795)	(2.445, 2.795)
2017	2.701	(2.537, 2.864)	(2.543, 2.863)	(2.540, 2.860)
2018	2.712	(2.523, 2.905)	(2.532, 2.912)	(2.512, 2.893)
2019	2.737	(2.562, 2.914)	(2.556, 2.915)	(2.559, 2.918)
2020	2.797	(2.630, 2.965)	(2.625, 2.962)	(2.633, 2.970)

Table 2: Bootstrap for mean of each year

Judging from the result, we can find that there is an increasing trend in the average number of fatal shootings per day. However, such increase is every slow. Last year (year 2020), the estimated average number increased by 2% and last year’s estimated value fell in the confidence interval of last few years’ estimated values. Our possible explanation for the increase is that due to the economic recession caused by coronavirus, the crime rate increased in 2020, thus leading to a higher fatal shooting number. Therefore, we believe the change in the number of fatal shootings in the past few years are still acceptable, and there is no need to be panic about the current situation and go parading on the street. But the increasing trend is still worth mentioning. We suggest that the number of fatal shootings should be paid attention in the coming years. If the increasing trend is speeding up more rapidly, maybe gun-using of the Police need further discussion.

Now we calculate the T_1 statistics from the fatal-police-shootings data and try to compare them with the critical regions we calculate in the “Simulation” part by using 10000 randomly generated samples. And we find that the real statistic of every year from 2015 to 2020 do not fall into the rejection region, which means we cannot reject $H_0 : X \sim \text{Poisson}(k), k = \hat{k}$.

To gather more evidence for our conclusion, we calculate the T_2 statistics from our data. The result is shown below in Table 3. The result indicates that although year 2018 is a little abnormal, the mean of our data is closely equal to the variance of our data, which is a major characteristic of Poisson distribution. Therefore, we tend to believe that American police’s shooting habit do not change much in the past few years.

Years	T_2 statistics	Normal	Basic	Percentile
2015	0.9364	(0.8031, 1.0544)	(0.7983, 1.0455)	(0.8274, 1.0746)
2016	1.0348	(0.887, 1.173)	(0.875, 1.161)	(0.908, 1.195)
2017	1.0323	(0.873, 1.178)	(0.860, 1.168)	(0.897, 1.205)
2018	0.7843	(0.6557, 0.8948)	(0.6514, 0.8871)	(0.6815, 0.9172)
2019	0.9084	(0.7816, 1.0212)	(0.7752, 1.0127)	(0.8040, 1.0415)
2020	0.9507	(0.8108, 1.0813)	(0.7964, 1.0708)	(0.8306, 1.1050)

Table 3: Bootstrap for mean/var of each year

Up to now, we have shown that every year’s thousands of fatal police shootings happening in the United States do follow a predictable pattern: the Poisson distribution. This finding can answer the question that whether the pattern of police shooting change in recent years. By seeing the evidence that every year’s data all follow a Poisson distribution, we can say that American police’s habit of shooting, at least from the angle of number of shootings, do not change much.

Discussion

In this research project, we calculate a confidence interval for the mean fatal shooting numbers per day using bootstrap. Judging from the value and the confidence interval, we conclude that the current police shooting situation is still acceptable. Using Goodness-of-fit test, we perform Monte Carlo hypothesis test and accept that our fatal-police-shootings data followed Poisson distribution every year. The rejection region is given by simulating samples drawn from Poisson distribution. We also calculate the ratio of mean to variance of our data and gather more evidence that the pattern of shooting numbers remained unchanged in the past years. Therefore, we believe American police's habit of shooting do not change much so far, but gun-using of the Police is still a problem worth mentioning in the further.

So what's the meaning of doing this project or why should anyone care about this? Actually, by making full use of the kind of data collected by The Washington Post, both the public and police forces can be alert to real changes in the incidence of fatal police shootings, but not be either unnecessarily alarmed or falsely reassured by apparent variability.

What's more, from a statistical point of view, for the data to be most useful, we need to think carefully about the most appropriate way to present the data for different audiences. For the media and public consumption, our comments must remain firmly based on the data and should always seek to avoid either unnecessary alarm or unwanted complacency [4]. For policymakers, the analysis should rather be focused to support proportionate risk-based decision-making, avoiding unnecessary new initiatives and making best use of existing resources.

This project only analyzed a certain angle of fatal-police-shootings data: the number of shootings per day, which means there are many angles left for us to do in the future research. For example, we can divide the data according to the races to see whether there is serious racial discrimination of the fatal police shooting. From roughly observing the data, we can see that the number of white Americans is larger than black Americans. However, when considering the proportion of a race in the national population, the rate at which black Americans are killed by police is probably higher than the rate for white Americans. Besides, we can put the elements of gender, age and position into consideration. After analyzing all these parameters, we can have a more general idea about American police's habits of shooting.

Reference

- 1 "George Floyd Death Homicide, Official Post-Mortem Declares". BBC News, 2021, <https://www.bbc.com/news/world-us-canada-52886593>.
- 2 The Washington Post. Fatal force. <https://www.washingtonpost.com/graphics/investigations/police-shootings-database/>. Web.
- 3 "washingtonpost/data-police-shootings", GitHub, 2021. [Online]. Available: <https://github.com/washingtonpost/data-police-shootings>.
- 4 D. Spiegelhalter and A. Barnett. London murders: a predictable pattern? *Significance*, 6(1):5–8, 2009. <http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2009.00334.x/abstract>.