Fatal Police Shootings in the U.S.

Ali Malenchik

The 2014 killing of 18-year old Michael Brown by police in Ferguson, MO caused national outcry against racial inequality and bias in the criminal justice system. Brown's death led to an investigation on the reporting of fatal police shootings, finding that the "FBI undercounted fatal police shootings by more than half" due to the voluntary nature of reporting (Washington Post). In 2015, the Washington Post began gathering data from news accounts, social media postings, and police reports. This data includes victim age, demographic data, circumstances of the shooting, and more. In this analysis, my main focus is a question that arose from the killing of Michael Brown: *are fatal police shooting victims disproportionately young, black men?* In addition, using the available city demographic data, I will explore whether such shootings are correlated to an area's education level, poverty rate, median income, and diversity.

As part of exploratory data analysis, I generated histograms, bar plots, box plots, descriptive statistics, and outlier analysis for the variables in question. My first observation relates to the **age** of the shooting victim. Though the victim's age ranges from 6 to 91, the mean, mode, and median are on the lower end (37, 36, and 35, respectively). The interquartile range is 27 to 45, indicating that the middle 50% of victims are on the younger side of the age range. The distribution appears to be asymmetric, with a tail that extends farther to the left than to the right. Finally, using the IQR I determined that the cutoff for outliers is <0 and >72. As such, all 50 outliers are on the older side (ranging from 72 to 91).

When analyzing the distribution of the **gender** attribute, I observed that 96% of the shooting victims were male. 5197 of the 5426 recorded values were male, and only 229 female. When analyzing **race** attribute, I found that of the 4806 recorded values, the majority were White, Black, and Hispanic. White made up 2406 or 50.1% of the races, Black made up 1255 or 26.1%, and Hispanic made up 938 or 19.5%. This is significant because when compared to the overall demographic distribution in the U.S., one would expect the percentage of White victims to be higher as they make up approximately 77% of the population (census.gov). Conversely, the proportion of Black shooting victims is higher than expected, as they make up approximately 13% of the population (census.gov). Using both **age and race** attributes, I plotted the age of the victim versus his or her age. Based on this analysis, it can be observed that the median age for Black victims is lower than other races (with the exception of Native-Americans). Additionally, the IQR and range for age of White victims is considerably wider than other races.

My next observations relate to the demographic data for the city in which the shooting took place. The **median income** ranges widely from $12,083 - $178,765. There is a significant tail to the right, and the middle 50% of values is a far narrower range, from $38,554.50 to

$54,222.50. The mean, mode, and median all sit near or just under $50K. Only one outlier was found on the lower end (<$15,052.50), whereas a whopping 283 outliers fall on the higher end (>$77,724.50). This is evidence of the skew in median income distribution. Poverty rate exhibits a similar distribution with a slightly less noticeable tail to the right. Poverty rate ranges from 0-76.4%, with the mean, mode, and median around 20%. Despite the wide range, the cutoff for outliers is relatively low, and again all 83 outliers are on the higher end (>39.3%). The mean graduation rate is 84.28%. The distribution is negatively skewed, with the data heavily condensed in the IQR of 80.5-89.7%. There are 144 values below the outlier cutoff (<66.7%).

In the next step of analysis, I created a Probability Mass Function. In order to address the main focus question, I first removed records for female victims to remove gender as a confounding variable. I think split the data frame into two groups - race Black and race Other - using the race attribute. After plotting the PMF for both groups, the age for Black victims is very negatively skewed toward the younger age. Age for other races is slightly negatively skewed as well, though it appears that more values are present for higher ages. Using the same separated data frames, I created a Cumulative Distribution Function. Based on the results, we can see that the CDF for Black male victims approaches 1 at a faster rate than males of other races. The Exponential CDF does not appear to be a good representation of the data as the line is not approximately straight.

In scatterplot and correlation analysis, I evaluated some of the demographic data for cities in which fatal police shootings took place. Both the Pearson (-0.74) and Spearman (-0.78) correlation coefficients between poverty rate and median income are significant, and the scatterplot indicates that as median income increases, poverty rate decreases (which inherently makes sense). However, the relationship does not appear to be linear based on the plot. The Pearson and Spearman correlation coefficients for graduation rate and median income are more moderate (0.42 and 0.45, respectively), and the relationship appears that it might be logarithmic. As the median income increases, graduation rate increases dramatically and then tapers off. Finally, the coefficients for graduation rate and poverty rate (-0.5, -0.58) are also moderate. The relationship appears to be more linear, and as poverty rate increases, graduation rate decreases. I decided to conduct linear regression analysis on this relationship and plotted a linear fit with an intercept of 94.44 and slope of -0.53. However, based on the coefficient of determination, only 25% of the data fit the regression model, indicating a poor fit.

The primary question I wanted to explore in this analysis was whether fatal police shooting victims are typically young, black men. In order to gain insight on this I performed a hypothesis test, comparing the difference of mean age for the two groups used earlier (Black men and men of other races). The null hypothesis states that there is no difference in mean age between the two groups. After calculating the p-value to be 0, it can be concluded that the difference is significant and unlikely to have occurred by chance.

During my analysis I ran into some limitations that I believe hinder the reliability of my results. Firstly, I intentionally removed confounding variables such as whether the individual was

armed, whether they exhibited signs of mental illness, etc. I believe these variables may provide some insight into the discrepancies found in my analysis, however they were difficult to work with as they are categorical. Another limitation was that city demographic data was collected in 2015 and may not be the best representation as it does not reflect the demographics at the time of the shooting. I also would have liked to plot fatal shootings in a geographical map to demonstrate the cities and states with the most instances of fatal force, though I ran out of time. Additionally, it could have been beneficial to have more quantitative data related to the shooting victim, e.g. his or her income, prior arrest records, etc. It also could have been useful if similar data were available for non-shooting deaths in police custody (e.g. strangulation, blunt force, etc.). Finally, it would be interesting to see data related to the individual officer such as age and race, to determine if any trends exist.

I did make one assumption prior to starting analysis, that most fatal shootings would occur in low education/high poverty areas. Based on the distributions, this is not necessarily true. During analysis I struggled with understanding how to best represent categorical variables. Since most of my analysis relied on quantitative data, I was restricted to using age and city demographic data. In retrospect it would have been more useful to have quantitative data related to the shooting incident itself.

References

Fatal Force: Police shootings database. (2020, January 22). Retrieved from https://www.washingtonpost.com/graphics/investigations/police-shootings-database/

U.S. Census Bureau QuickFacts: United States. (n.d.). Retrieved from https://www.census.gov/quickfacts/fact/table/US/RHI125218#RHI125218