Analysis of U.S. Police Shootings

Kathryn Puczkowskyj

12/9/2019

## Executive Summary

Using observations on fatal police shootings in the U.S., we ran a chi-squared test to find out whether threat level and race were independent. Our results indicate that the greatest association is likely between the Hispanic race and either of the identified threat levels (but much less so with the undetermined level).

## Introduction

Police shootings and the potential influence race plays in those shootings have become increasingly hot topics in the past several years. Entire cultural events like the #BlackLivesMatter movement have emerged to draw attention to the apparent association between race and police shootings. After a short time Googling, I found this data set. When I reviewed the variables collected, I became interested in whether the race of the person shot could have been associated with the threat level that the police reported.

## Data

The data set examined in this assignment was downloaded from a Github repository maintained by the Washington Post. They have compiled a database of every fatal shooting in the U.S. by a police officer in the line of duty since January 1st, 2015. There are 4,763 observations and 14 variables. The two variables that we are interested in here are race and threat level. See the attached Appendix for a description of the levels in these two variables.

## Methods

Because these are two categorical variables with multiple levels, we will perform a chi squared test for independence. Our null hypothesis would be that there is no relationship between race and threat level of the individual killed. In other words, these two variables are independent. After this test, we will calculate residuals to see which levels of each categorical variable are more likely to have an influence.

## Results

The contingency table, Table 1 below, makes it seem like there could be an association between threat level and race. In order to see if there really is an association, we next look at Table 2, the expected values. As we can see, the expected values are different than the observed values. Below, we will take a look at how different they are.

Table 1 - Contingency Table: a cross-tabulation of the factors

|  |  |  |  |
| --- | --- | --- | --- |
|  | attack | other | undetermined |
| A | 43 | 34 | 2 |
| B | 735 | 320 | 58 |
| H | 460 | 304 | 48 |
| N | 39 | 30 | 3 |
| O | 26 | 15 | 0 |
| Unknown | 343 | 141 | 39 |
| W | 1393 | 641 | 89 |

Table 2 - Expected values by race and threat level, assuming independence.

|  |  |  |  |
| --- | --- | --- | --- |
|  | attack | other | undetermined |
| A | 50.40542 | 24.63049 | 3.964098 |
| B | 710.14214 | 347.00924 | 55.848625 |
| H | 518.09112 | 253.16397 | 40.744909 |
| N | 45.93911 | 22.44804 | 3.612849 |
| O | 26.15977 | 12.78291 | 2.057317 |
| Unknown | 333.69662 | 163.06005 | 26.243334 |
| W | 1354.56582 | 661.90531 | 106.528868 |

The Pearson’s chi-squared test on the race and threat level of recorded shootings results in a chi-squared test statistic of 46.91 with 12 degrees of freedom, and a p-value less than 0.00001. This tells us that there is some association between threat level and race, but does not tell us which levels are most influential. The association plot shows positive residuals in blue and negative residuals in red. The height of the box indicates the size of the residual, so this hispanics with a determined threat level had the largest variation from expected values. Black and Native American individuals with an undetermined threat status had the lowest variation from expected values.

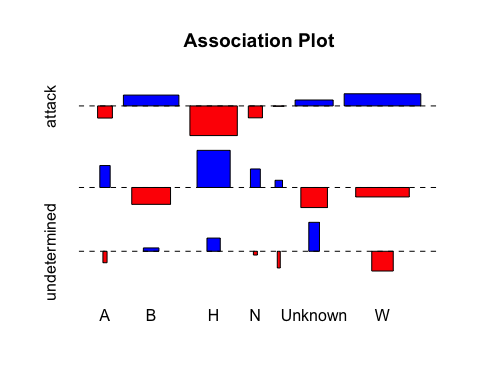


Figure - Association plot of residuals for threat level (vertical) and race (horizontal)

## 

## Conclusion & Take-Aways

Our analysis shows that there is an association between the race and threat level for people who are fatally shot by police in the U.S. The biggest residuals correspond to Hispanics with “other” or “attack” threat levels. Given these results, we would likely reject the null hypothesis of independence.

If we were to expand upon this analysis, it would be interesting to incorporate all police shootings instead of just fatal ones. Having a more inclusive data set would give a larger picture of how subconscious prejudice might relate to these shootings. I would also like to dive deeper into showing whether or not being a threat to others has more influence on whether or not a person gets shot by police than their race. It would likely be difficult to determine since it is likely that the police officer is self-reporting whether they perceived an emminent threat or not. Either way, I think it’s a topic that could be explored a lot more.

## Appendix A: Variable Levels

Levels of Race: White non-Hispanic, Black non-Hispanic, Asian, Native American, Hispanic, Other, None (unknown)

Levels of Threat Level: Attack (direct & immediate threat to life), Other (officers or others faced significant threat), Undetermined

## Appendix B: Code

#setup   
require(readr)  
require(knitr)  
require(mosaic)  
require(RColorBrewer)  
require(corrplot)  
#read in data  
mydata <- read.csv("PoliceShootingData\_20191202.csv")  
#create contingency table  
kable(table(mydata$race,mydata$threat\_level),caption = "Contingency Table: a cross-tabulation of the factors")  
#get expected values and put into table  
shooting.count <- nrow(mydata)  
race.split <- table(mydata$race)/shooting.count  
threat.split <- table(mydata$threat\_level)/shooting.count  
expected.tbl <- outer(race.split,threat.split)\*shooting.count  
knitr::kable(expected.tbl,caption="Expected values by race and   
 threat level, assuming independence.")  
#chi-squared test  
chisq.test(table(mydata$race,mydata$threat\_level))  
#association plot to visualize of potential association  
assocplot(table(mydata$race,mydata$threat\_level),col=c("blue","red"),main="Association Plot")