Section 2: Week 4: Analyze Statistical Output

Nate Bachmeier

TIM-7101: Statistics for Technology Leaders

August 23rd, 2020

North Central University

# Analyze Statistical Output

NCU-Cares (NCU-C) is a politically neutral non-profit that seeks to improve the work through targeted lobbying efforts. The death of George Floyd has risen the debate of police violence and reform to the national stage (Crary & Morrison, 2020). While the topic rests on American’s hearts and minds, it has also become highly partisan with many efforts to undermind the conversation (McCaskill, 2020). It can be challenging to find complete records as federal regulators do not mandate data collection policies. The Washington Post attempts to fill this gap with its dataset based on news events, exposing the location, context, and several demographic markers for each fatality (Washington Post, 2020). Researchers can use these results to help determine if the solution to police violence comes from racial issues or another aspect better explains the national issues.

# Section I: Describe the Problem and Hypothesis

Annually nearly one thousand citizens die from police violence, which raises the central question, why? The collective perspective of the Black Lives Matter movement is that police exert disproportionate force against people of color (BLM, 2020). This perspective often comes with the quote that “Black civilians were more than twice as likely as White civilians to be unarmed (Nix, Campbell, Byers, & Alpert, 2017)” during the fatality. Nevertheless, others argue the brutality victims are experiencing a mental health crisis, and this is the actual reason (Lamb, Weinberger, & DeCuir, 2014). While these perspectives effectively drive media headlines, are they both missing the forest among the trees? Does another factor more accurately explain the challenges that are occurring? Instead, NCU-C hypothesizes that neither sanity nor race is the driving cause of police violence. Alternatively, provocation might better explain the need for forceful escalations that result in death.

# Section II: Describe the Data Set

The Washington Post provides demographic and contextual information about victims from January of 2015 to the present day. Each entry captures the threat level, flee status, any weapons, age, gender, race, and city. NCU-C enhanced these 5489 records to include several nominal features, such as ‘has a projectile,’ to simplify analysis on the free form weapons column. The inclusion of an ‘age group’ property also exists for smoothing visualization charts by partitioning into five-year windows. Aside from these transformations, no alterings of the original data set are present.

## Exploring by Race

America’s racial make-up is approximately 63% white, 15% Hispanic, 13% black, and 9% other (Census Bureau, 2019). If all things are equal, then looking at the raw victim statistics should convey a similar breakdown. These initial expectations are comparable though slightly skewed in Washington Posts’ data set when grouping by *race* (see Figure 1). After adding a second level of grouping by *year*, it also raises an observation that the number of victims is relatively stable across time. From January 2015 to December 2019, the mean death rate is 905, with a standard deviation of 35. While the situation is not getting any better, it is also not becoming worse.

Figure 1: Victims by Race



## Exploring by Age

A normal distribution exists for the victim’s age around the mean of 37 with a standard deviation of 13 years. After grouping by race, the data shows that minorities encounter deadly confrontations with the police roughly seven years younger. From these initial range values, it is possible to calculate the statistical effect of a person’s *race* and *age* relative to a similar group. An effect size is a measurement in z-scores with values typically between zero to one. This two-level comparison conveys that a medium-level effect exists for Whites, and a minimal difference exists between Blacks and Hispanics (see Table 1). These results roughly align with the exploration of race, which suggests that a skew exists in the data, but its not the smoking gun.

Table 1: Influence of Age

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Race | Mean | Standard Deviation | Effect vs. White | Effect vs. Black | Effect vs. Hispanic |
| All | 37.12 | 13.12 | -0.11 | 0.38 | 0.30 |
| White | 39.95 | 13.37 | -- | 0.61 | 0.54 |
| Black | 32.47 | 11.33 | -0.38 | -- | -0.10 |
| Hispanic | 33.54 | 10.87 | -0.30 | 0.10 | -- |

## Exploring by Sanity

An argument exists that the solution to police violence is defunding the police, and using those resources for drug rehabilitation and civil service programs (BLM, 2020). Assuming those changes went into effect, would it make a difference? Washington’s data suggests that 22% of all fatalities are people experiencing a mental health crisis. Next, comparing the effect of *sanity* against the age distribution concludes that a small size exists (0.20). Based on these results, it does not seem that trading funding would produce the desired outcome. There are likely other potential benefits that come from changing funding levels of police and civil services. However, those are outside the scope of this research project. NCU-C needs to continue its searching into other variables of the dataset before concluding on the best lobbying action.

Table 2: Effect of Mental Health

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mental Illness | Mean (Age) | Standard Deviation | Effect vs. All | Effect vs. Sane |
| Yes | 39.74 | 13.87 | 0.20 | 0.25 |
| No | 36.35 | 12.80 | -0.06 | -- |

## Exploring by Weapon Type

The data set includes the ‘armed’ column that contains free-form text describing any weapons on the victim. One of the challenges with analyzing this field comes from the various subtle differences in its values (e.g., *baseball bat* versus *baseball bat and bottle*). Enhancements of each record include categorical-features that bucket the weapons by genre. These buckets are named projectiles, sharp/blunt instruments, tool/small objects, explosive, unspecified, vehicles, and unarmed. When the suspect has multiple weapons, such as both gun and knife, the higher risk object dictates the category. According to these categorical-aggregations, roughly 58% of victims had a firearm, plus another 18% had a sharp/blunt instrument (see Figure 2).

There are nationally fifty-three million people who have an interaction with a law enforcement officer each year (BJS, 2015). Of this population, annually, approximately one thousand dies. These figures suggest that roughly 0.0019% of all interactions end with the officer killing the suspect. Further removing situations with guns and knives (76%) reduces the figure to 0.00046% of interactions result in death!

Figure 2: Victim Weapon Category



## Exploring Unarmed

One particular subset that gains much attention is unarmed citizens that die during police encounters. According to the dataset since 2015, this represents 349 of the 5489 (6.36%) of fatal incidents. These victims are nearly universally 33 years old, with a standard deviation of 11, regardless of race (see Table 3). An alternative statistic directly compares the counts of entries marked as unarmed and finds that there is no substantial difference within the data. While Nix’s frequently cited paper is accurate in stating that “unarmed Blacks die at nearly twice the rate of Whites,” the differences are within the margin of error (see Table 4). Further expanding the definition of unarmed to include non-weapons, such as staplers and pens, makes the low-risk situations approximately equal regardless of race.

Table 3: Effect of Race, Age, and Unarmed Status

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | STD | Count | Effect versus White |
| All | 32.37 | 11.04 | 349 | -0.03 |
| White | 33.18 | 11.34 | 146 |  |
| Black | 32.17 | 10.92 | 121 | -0.02 |
| Hispanic | 31.10 | 11.11 | 63 | 0.10 |

Table 4: Unarmed or Low-Risk Status by Race

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Race | Unarmed  Rate | Unarmed  Total | Low-Risk Situation Rate | Low-Risk Situation Total |
| Black | 9.50% | 124 | 12.95% | 169 |
| Whites | 5.83% | 146 | 10.11% | 253 |
| Hispanics | 6.92% | 63 | 11.32% | 103 |

# Section III: Limitations and Findings

## Risks and Limitations

Despite the best efforts and intentions, there are risks to the completeness and validity of these results. For instance, all results derive from the Washington Post police shooting data and do not include other sources. This design decision can cause missing or incorrect observations to feed into the analysis process. It is also possible for others to dispute specific categorization decisions while grouping the free-form text of the ‘armed’ column. A third major area of potential risk comes from the project’s rapid timeline. These quality threats could manifest as duplicate counts or erroneous filters. There is also minimal consideration of outcomes statistical significance, due to a heightened focus on descriptive analysis during this phase.

## Observations

Assuming all things equal, the number of deaths per race should be roughly equal to the racial distribution in America. Based on that ratio, there should be more White deaths and fewer Blacks. Despite a data skew existing, it is not significant enough to conclude that police are arbitrarily killing one group. That is not to suggest that friction does not exist between civil rights and law enforcement, only that racism is not the root cause of police brutalities. Another frequent position states that mental health is too blame for these untimely demises. However, a more in-depth investigation into the deaths of victims experiencing a mental health crisis versus other sane fatalities concludes a minimal effect size. While there are many valid reasons for communities to invest additional resources into mental health services, police brutality is not among them.

Police brutality is an onion, and it stinks. Ideally, no one should ever die, but that does not mean the officers are committing some light form of genocide. Nor can one entirely blame any mental illnesses in these situations. Peeling the next layer places the focus on weapons at the scene of the crime. According to simple aggregation counts, nearly 72% of all victims possessed a gun or knife during their altercation. There needs to be additional research into this cluster as it presents the highest safety risk and greatest reward. Of the remaining people either unarmed or defending themselves with low-risk weapons (6-12%), these also warrant further thought.