Racial Divide: Analyzing Police Shooting Data for Age Disparities

Authors: Veda Sahaja Bandi, Sindhuja Baikadi, Mythri Krpet Jayaram

The Issues:

The dataset under scrutiny comprises police shootings resulting in fatalities across the United States, encompassing critical details such as the age and racial demographics of individuals involved in these tragic encounters.

Key questions prompting analysis include:

- What disparities exist in the average ages of black individuals versus white individuals shot and killed by law enforcement officers?
- Does the data reveal a difference in average ages between Black and White individuals shot and killed by police? If a difference exists, what is its magnitude and direction?
- Is the observed difference in average ages between black and white individuals subjected to police shootings statistically significant?

Understanding these disparities and their statistical significance is crucial for evaluating systemic issues within law enforcement practices and fostering equitable policing standards.

Findings:

Upon analyzing the police shooting data concerning individuals in the United States, it is evident that there exists a considerable disparity in the average age between black and white victims. Specifically, black individuals shot by police tend to be approximately 7 years younger on average than their white counterparts. This discrepancy highlights potential systemic issues within law enforcement practices and necessitates a thorough examination of community-police relations. The substantial effect size of 0.57 underscores the notable difference observed between the age of black and white victims of Police Shootings, as outlined by Cohen and elaborated upon by Sawilowsky.

Furthermore, statistical analysis reveals that the observed difference in average ages between black and white victims is highly unlikely to have arisen by random chance. The statistical significance of this observation is akin to the improbability of obtaining 23 consecutive heads from flipping a fair coin. These findings underscore the critical need to address racial disparities and advocate for equitable policing standards to uphold justice and fairness in law enforcement interactions.

Discussion:

Our analysis revealed a significant age gap between black and white individuals involved in police shootings, with a substantial effect size observed between the average ages, exceeding 7 years. This notable effect size highlights the pronounced differences in age demographics among victims of police shootings, emphasizing the urgency of understanding the underlying factors contributing to this divergence. Further investigation and the acquisition of verifiable data are crucial to uncover the true factors at play.

Potential avenues for exploration include disparities in police interactions based on race or ethnicity, demographic variations in age composition among racial groups, and socioeconomic disparities. By delving deeper into these potential explanations, we can gain a more comprehensive understanding of the origins of this troubling age gap.

Exploring the implications of these findings for broader issues like police practices, community relations, and public safety is equally important. This knowledge can inform decision-making and policy development as we strive for a more equitable future. Only through rigorous investigation and a commitment to evidence-based solutions can we address this complex issue effectively.

Appendix A: Method

Data was downloaded as a comma-separated (.csv) file and imported into Jupyter Notebook. The analysis was done by isolating the age and race variables, leading to the creation of two distinct subgroups: ages of black individuals and ages of white individuals.

The smooth approximations provided a visual and quantitative way to compare the age distribution of black and white people shot by the Police. Cohen's d value is used to estimate the effect size between the mean age of black people shot by police and the mean age of white people shot by police.

Average ages for black and white individuals were computed to assess the normality of age distributions in preparation for a t-test examining the difference in average ages. In cases where the age distributions by race deviated significantly from normality, a Monte Carlo procedure was employed to calculate a p-value. This value served as an estimate of the probability of observing the observed age differences if the null hypothesis of no statistical difference was true. Relevant Python code can be found in Appendix C.

Appendix B: Results

From a database comprising 6054 data points, contains both age and race factors. Upon extracting the age and race data, 2078 cases pertained to black individuals, while 3079 cases corresponded to white individuals. The age distribution of the Black people shot is larger than the age distribution of White people shot.

Effect Size

Smooth histogram approximations comparing the age distributions of black and white individuals reveal a notable leftward shift, indicating a lower average age for black people.

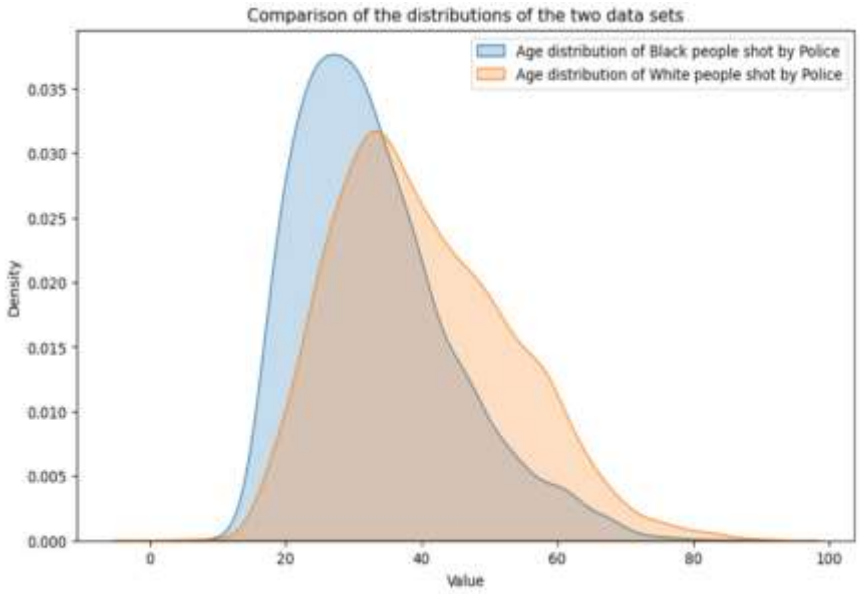

Figure 1: Age distributions of black and white people shot by Police

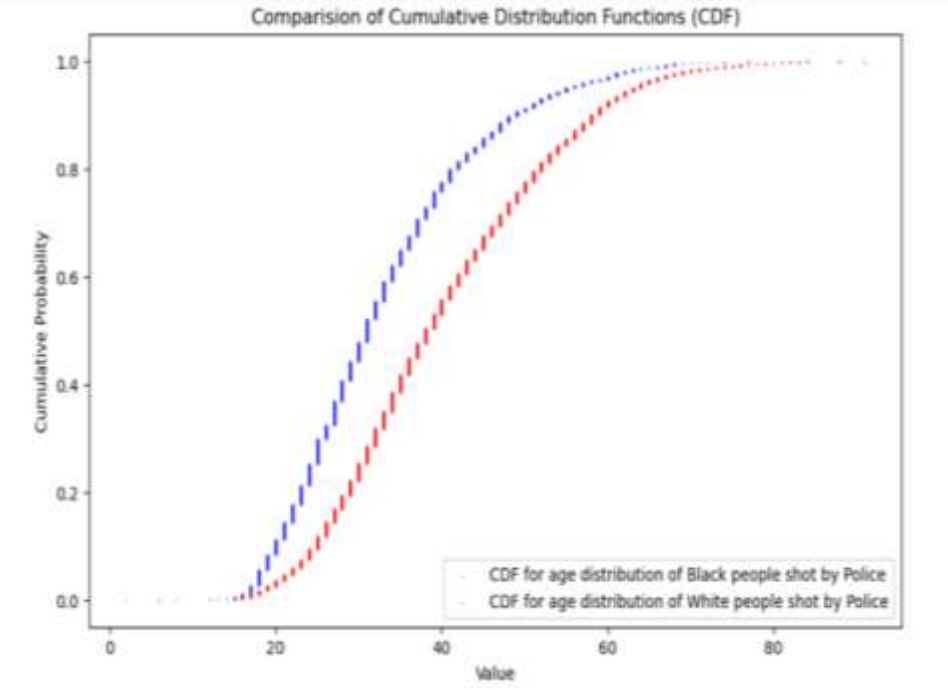

Figure 2: Comparison of Cumulative Distribution Functions (CDF)

The average age for black individuals was 33.05 years, while for white individuals, it was 40.27 years, reflecting a notable 7.22-year difference in average age.

Quantile plots for the age distributions of black and white individuals shot by the Police suggest significant deviations from normality.

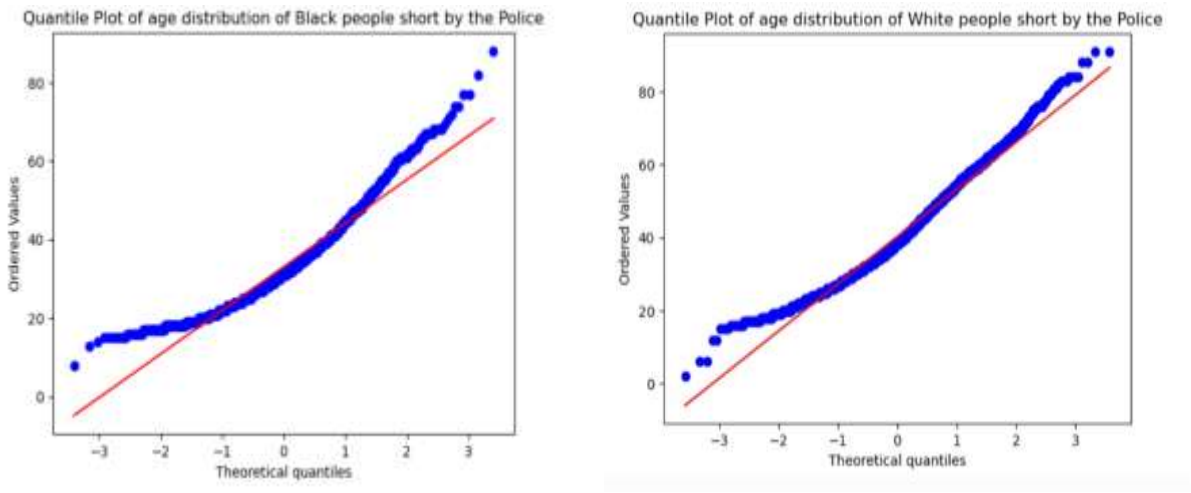

Figure 3: Quantile plots of the age distribution of Black and White people killed by the Police

The Cohen's d value of 0.57, obtained from the analysis of police shootings, signifies a considerable effect size. This implies a substantial and consistent discrepancy in the average ages of individuals, particularly between black and white victims of police shootings. These results highlight the importance of understanding age-related dynamics in such incidents and emphasize the need for deeper exploration into the factors influencing this divergence.

Statistical Significance:

The t-test yielded a highly significant p-value of less than 0.05 ($p < 0.0001$) in the analysis of police shootings, indicating that the observed average age difference is statistically significant. This robust statistical evidence suggests that the disparity in average ages between black and white individuals in police shootings is not attributable to chance.

As our data exhibits a slight positive skewness, deviating from the assumption of normal distribution required by the T-test, we opt for calculating the p-value through the Monte Carlo method. This approach acknowledges and addresses the non-normal distribution of our data, ensuring a more appropriate statistical analysis.

Due to notable deviations from normality in the size distributions, the Monte Carlo method was preferred over traditional t-tests. With a maximum of 10,000,000 simulations, pooled sizes of Black and White individuals shot by Police were computed. The observed difference in means, indicating the age contrast between the two groups, was then assessed. The resulting p-value from the Monte Carlo simulation was 9.999×10^{-8}, suggesting significant departures from normality and advocating against a t-test-derived p-value.

A rigorous analysis involving 10 million random samplings revealed no instances where the average age difference matched the observed gap between Black and White individuals shot by police. Using the method advocated by Davison & Hinkley (1997), and described in North, Curtis & Sham (2002, p. 439), the calculated p-value was 10^{-7} which is equivalent to the probability of flipping heads 23 consecutive times on a fair coin. This extremely small p-value provides strong evidence against the possibility that the observed difference could be due to chance alone, suggesting a statistically significant difference in the ages of Black and White individuals involved in police shootings.

Appendix C: Code

In this appendix, we document the Python code for the comparison of cumulative distributions of ages of black and white people shot by Police, Cohen's d, T-test, and Monte Carlo procedure to estimate the effect size and statistical significance for the Police Shooting dataset.

```
# Plotting the distribution of crab sizes
plt.figure(figsize=(10, 6))
sns.kdeplot(black_data, fill=True, label='Age distribution of Black people shot by Police')
sns.kdeplot(white_data, fill=True, label='Age distribution of White people shot by Police')
plt.xlabel('Value')
plt.ylabel('Density')
plt.title('Comparison of the distributions of the two data sets')
plt.legend()
plt.show()
```

```
# Plotting the Cumulative Distributions of Crab sizes
sorted_black_data = np.sort(black_data)
sorted_white_data = np.sort(white_data)
cumulative_prob1 = np.arange(1, len(sorted_black_data) + 1) / len(sorted_black_data)
cumulative_prob2 = np.arange(1, len(sorted_white_data) + 1) / len(sorted_white_data)
plt.figure(figsize=(10, 6))
plt.scatter(sorted_black_data, cumulative_prob1, label='CDF for age distribution of Black people shot by Police', s=0.05, color='blue')
plt.scatter(sorted_white_data, cumulative_prob2, label='CDF for age distribution of White people shot by Police', s=0.05, color='red')
plt.xlabel('Value')
plt.ylabel('Cumulative Probability')
plt.title('Comparison of Cumulative Distribution Functions (CDF)')
plt.legend()
plt.show()
```

```
# Cohen's d for Effect Size
def cohend(d1, d2):
    n1, n2 = len(d1), len(d2)
```

```
    s1, s2 = np.var(d1, ddof=1), np.var(d2, ddof=1)
    s = np.sqrt(((n1 - 1) * s1 + (n2 - 1) * s2) / (n1 + n2 - 2))
    u1, u2 = np.mean(d1), np.mean(d2)
    return (u2 - u1) / s
cohend_value = cohend(black_data, white_data)
print(f"Cohen's d: {cohend_value}")
```

Output

Cohen's d: 0.5754223321506542

```
# T-test for statistical significance
# T-test for statistical significance
t_stat, p_value_t = ttest_ind(black_data, white_data)
print(f"t-test: {t_stat}")
print(f"p-value (T-Test): {p_value_t}")
```

Output:

t-test: -21.257464575572655

p-value (T-Test): 9.020231053936359e-97

```
# Monte Carlo Method
max_simulations = 10000000
# Pooled ages, total, and lengths
PooledAges = np.concatenate((black_data, white_data))
total = np.sum(PooledAges)
w = len(white_data)
b = len(black_data)
# Observed difference in means
diff = np.mean(white_data) - np.mean(black_data)
# Monte Carlo simulation
r = 0
for n in range(1, max_simulations + 1):
    A = np.random.choice(PooledAges, size=w, replace=False)  # Random sample without replacement
    x = np.mean(A)
    y = (total - np.sum(A)) / b
    if x - y > diff:
```

```
        r += 1
# Calculate p-value
p_value_mc = (r + 1) / (n + 1)
print("p-value (Monte Carlo)=", p_value_mc) print("p-value (Monte Carlo)=", p_value_mc)
```

Output:

p-value (Monte Carlo)= 9.9999990000001e-08

Contributions:

Mythri Krpet Jayaram - 02127296: Worked on the Discussion, Method, and Results sections. Also self-plotted the graphs to analyze the data using the various methods discussed in the report.

Veda Sahaja Bandi - 02105111: Outlining key observations and insights, worked on the coding portion of the project to implement necessary functionalities and features.

Sindhuja Baikadi - 02128756: Worked on the Issues and Findings section of the report, addressing various issues and identifying insights