An Investigation of Racial Disparities in Traffic Stops and Citations in Durham County, North Carolina

Abstract

This paper investigates if there is any evidence of racial bias in traffic stops and citations in Durham County, North Carolina. We used data from the Stanford Open Policing Project to investigate the relationship between a subject's demographic attributes (primarily race, with some insight into sex) and the likelihood of being stopped by police in traffic or receiving a traffic citation in Durham. We hypothesize that race and the likelihood of being stopped by police in traffic or receiving a citation in Durham are associated, with black people disproportionately more stopped relative to their population proportion and more likely to receive a citation upon being stopped. Our conclusions are 1) black people are disproportionately stopped in traffic as compared to their demographic makeup within the Durham population; 2) black people are not the most likely to receive a citation upon being stopped in traffic—rather, Hispanic people are the most likely to receive a citation; 3) black females are more likely to be receive a citation upon being stopped than are black males.

Background and Significance

The US incarcerates more people than any other country, and people of color make up a disproportionate percent of the prison population. Police funding has grown significantly over the past four decades, and over-policing in communities of color is a pressing issue. In Michelle Alexander's book, *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, she discusses The War on Drugs as one of the biggest causes of contemporary mass incarceration, citing police pretext stops as one of its tactics. In pretext stops, cops can pull over a "suspicious" driver on the pretext of a very minor traffic violation (e.g., turning on red, going over the speed limit) and then do a drug sweep of the car, which may result in an arrest for drug-related charges. According to a Pew Research Center survey, "Black adults are about five times as likely as whites to say they've been unfairly stopped by police because of their race or ethnicity." We would like to investigate if similar elements of discrimination in traffic stops are evident in Durham County, North Carolina. We would further like to investigate if the financial penalties for such traffic stops (i.e., citations) are disproportionately issued based on race. We hypothesize that race and the likelihood of being stopped by police in traffic or receiving a citation in Durham are associated, with black people disproportionately more stopped relative to their population proportion and more likely to receive a citation upon being stopped.

Methods

Data Collection and Variables

Our data is an attempted census of individual police stops in Durham created by the Stanford Open Policing Project (SOPP), which collects data on law enforcement nationwide. The SOPP pulled the Durham data from Statewide, NC data, with 100% coverage rate for all variables except for 85.2% for time of the traffic stop and 96.4% for the basis of the search. There are 326,024 observations in the "Durham" dataset, each an individual police stop recorded in Durham between December 2001 and December 2015. There are 29 variables, relevant variables including the following: *subject_race*, which describes the race of the subject involved; *subject_sex*, which describes the sex of the subject involved; *subject_age*, which describes the age of the subject at the time of the traffic stop; and *citation_issued*, which describes whether a citation was issued during the stop. For demographic comparisons, we used 2010 Durham County census data. Consequently, we filtered our dataset to only have values from Durham County and to remove missing data on race and age, which left us with 323,147 observations to sample from.

Exploratory Data Analysis

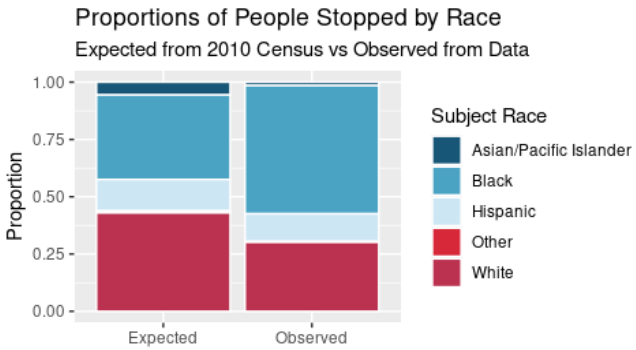


Figure 1:

Using the census data, we compared the expected proportion of people stopped by race to the observed proportion of people stopped by race. The segmented bar chart illustrates that a significantly greater proportion of black people were stopped compared to what was expected based on the proportion of black people in Durham County in the 2010 census.

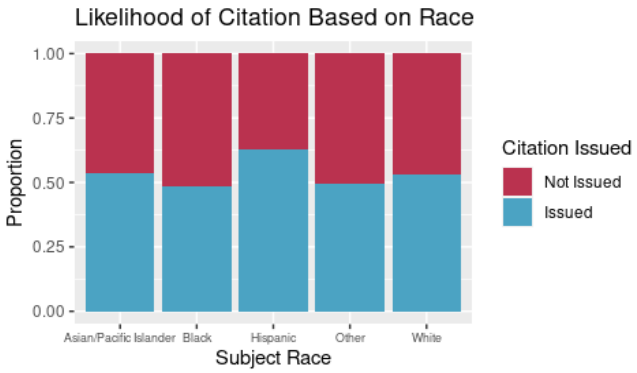


Figure 2:

Additionally, we visualized the proportion of citations given based on race. According to the visualization, Hispanic people have the highest proportion of citations issued. Contrary to our hypothesis, black people appear to be the race with the lowest proportion of citations issued.

Analytic Methods

Since the attempted census of 323,147 observations was too large to create a bootstrapped null distribution, we created a stratified proportional sample of 3,231 observations—roughly 1% of the original dataset. To further explore the difference in observed and expected proportion of black people stopped in traffic from the exploratory data analysis, we used a simulation-based one-proportion z-test to check if that difference—given Durham County race demographics—was statistically significant at the 1% significance level (Appendix 5).

Furthermore, as a preliminary check to decide if further investigation was warranted, we conducted a chi-squared test for independence between race and citation issued that provided sufficient evidence for association between the two (Appendix 6). We then checked the conditions for logistic regression, found that they were met, and continued with a logistic regression model (Appendix 2, 4). Our logistic regression model to quantifies the effect of race—adjusted for other demographic variables—on the log-odds of receiving a citation, and contains four predictors: the subject's race, the subject's age, the subject's sex, and an interaction variable between the subject's race and subject's sex (Appendix 7). The log-odds of receiving a citation for a white female was the referent. To determine the statistical significance of the coefficients, we set the alpha level to 0.01.

Results

First Research Question: Demographic Factors and Likelihood of Being Stopped

Our statistical analysis provides sufficient evidence at the 1% significance level that black people are disproportionately more likely to be stopped in Durham County relative to their proportion within the population (Appendix 5). This finding is both statistically and practically significant as indicated by the stop rate visualization in the exploratory data analysis.

Second Research Question: Demographic Factors and Likelihood of Citation

Our logistic regression model is laid out in the Appendix 7 table and yields a few relevant conclusions (Appendix 7):

1. Holding age and sex constant, we expect the odds that a Hispanic person will receive a citation upon being stopped by police in Durham County to be roughly 1.429 times the odds that a white person will receive a citation upon being stopped by police. The coefficient is statistically significant (p-value < 0.01), meaning there is less than a 1% chance such a coefficient or more extreme would be found in the data if race and the likelihood of receiving a citation were not associated.

2. Holding age and sex constant, we expect the odds that a black person will receive a citation upon being stopped by police in Durham County to be approximately 0.881 times the odds that a white person will receive a citation upon being stopped by police. The coefficient is also statistically significant (p-value < 0.01).

3. In most cases, the interaction variable between sex and race does not result in a statistically significant coefficient—the critical exception being the case of black people. Holding age and race constant, upon being stopped, a black man's odds of receiving a citation are expected to be approximately 0.915 times the odds that a black woman will receive a traffic citation upon being stopped. Thus, unlike nearly every other race listed, black people are the only race where women are statistically significantly more likely to receive a citation upon being stopped.

Discussion

Through our analysis, we have learned that black people are disproportionately more likely to be stopped for a traffic violation in Durham County. This stop rate is both statistically significant and practically relevant, with black people being stopped at a 2.5 times higher rate than white people (Appendix 1). When visualizing stop rates by race over time, we found that an even greater proportion of black people were stopped in traffic in Durham in 2015 than in 2001, which may indicate that discriminatory police stops are not an issue of the past, but rather, are still relevant today (Appendix 3). Contrary to our hypothesis on racial bias in issuance of citations, however, we have found that black people are disproportionately *less* likely than both white people and Hispanic people to receive a citation upon being stopped. Based on our data, we are unable to explain the difference in conclusions between our first and second research question; this may indicate a problem area for further research.

Additionally, we found that Hispanic people, particularly Hispanic males, were the most likely to receive a citation upon being stopped. Unfortunately, our observational data cannot create a causal association between a person's race and the likelihood of being stopped or receiving a citation. Nonetheless, our data is consistent with the discussion on racial profiling and pretext stops in Michelle Alexander's *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*, which indicates that black people are more likely to be stopped in traffic and that race does play some role in traffic penalties. This discovery has profound implications, as financial penalties of citations may impose a heavy burden on those who come from lower socio-economic classes.

For most races, there is very little change in data between citations given by sex. However, the odds that a black female will receive a citation are significantly greater than their male counterparts. This could possibly show unspoken discrimination against black females, an idea we believe necessitates further analysis. A statistical exploration of stop rates of black females compared to black males and white females could illuminate the unique intersectional experiences faced by black females within America.

A key limitation of our data analysis was that the SOPP separated races differently than the census data. Specifically, the SOPP data on race did not account for biracial people or people of other races; it also did not detail what category White, Black, or Asian Hispanics would fall under. As a result, our proportions are slightly skewed. In the future we would attempt to match the categories between the different datasets. If we were able to standardize the groupings, we would have less ambiguous data and more accurate proportions and conclusions. Another crucial limitation is that we are unable to establish if the trends we found were causally related. We are only able to take note of trends and cannot isolate causation; thus, our data does not provide conclusive evidence of racial discrimination by police in Durham County.

References

Alexander, Michelle. *The New Jim Crow: Mass Incarceration in the Age of Colourblindness*. Penguin Books, 2019.

Barghouty, Pheobe, et al. “The Stanford Open Policing Project.” Openpolicing.stanford.edu, 2020, openpolicing.stanford.edu/.

“Criminal Justice Facts.” The Sentencing Project, 2 Sept. 2020, www.sentencingproject.org/criminal-justice-facts/.

DeSilver, Drew, et al. “10 Things We Know about Race and Policing in the U.S.” Pew Research Center, Pew Research Center, 17 Aug. 2020, www.pewresearch.org/fact-tank/2020/06/03/10-things-we-know-about-race-and-policing-in-the-u-s/.

Rhodes, Nancy, et al. “Risky Driving among Young Male Drivers: The Effects of Mood and Passengers.” Transportation Research Part F: Traffic Psychology and Behaviour, Pergamon, 26 Dec. 2014, www.sciencedirect.com/science/article/pii/S1369847814001727.

“U.S. Census Bureau QuickFacts: Durham County, North Carolina.” Census Bureau QuickFacts, 2010, www.census.gov/quickfacts/fact/table/durhamcountynorthcarolina/RHI225219.

Appendices

Appendix 1

We quantified the stop rates of each race compared to their 2010 population proportion for the purpose of making interracial comparisons and to determine their practical significance. The outputted table is below.

Table 1: Summary statistics for stop rate based on census proportions by race in 2010:

Subject Race	n	Population	Stop Rate
Asian/Pacific Islander	471	15120	0.031
Black	17691	99630	0.178
Hispanic	4220	36990	0.114
Other	133	2430	0.055
White	8749	116100	0.075

It appears that black people have the highest stop rate among races in Durham County in 2010. This rate is roughly 2.5 times higher than the rate that white people get stopped in Durham County.

Appendix 2

This appendix is devoted to showing why we chose the logistic regression model that we did. Utilizing the demographic characteristics of sex, race, and age, we attempted to create the most robust and explanatory model possible from the data. Below are the calculated AIC and BIC values for each model we considered. The model with the lowest AIC and BIC values was the one we chose, as it followed the principle of Occam's Razor the most faithfully (it explained the most in the least complex manner).

Logistic regression of Race (BIC, then AIC):

```
## [1] 444902.9
## [1] 444849.5
```

Logistic regression of Race and Age:

```
## [1] 444211.7
## [1] 444147.6
```

Logistic Regression of Race, Sex, and Age:

```
## [1] 444197.4
## [1] 444122.6
```

Noting that the best logistic regression included all three variables, we tested interaction variables.

Logistic Regression of Race, Sex, Age, and Race * Age:

```
## [1] 444225.5
## [1] 444107.9
```

BIC value increased, so we eliminated the interaction variable.

Logistic Regression of Race, Sex, Age, and Age * Sex:

```
## [1] 444202.1
## [1] 444116.6
```

BIC again increased, so we eliminated the interaction variable.

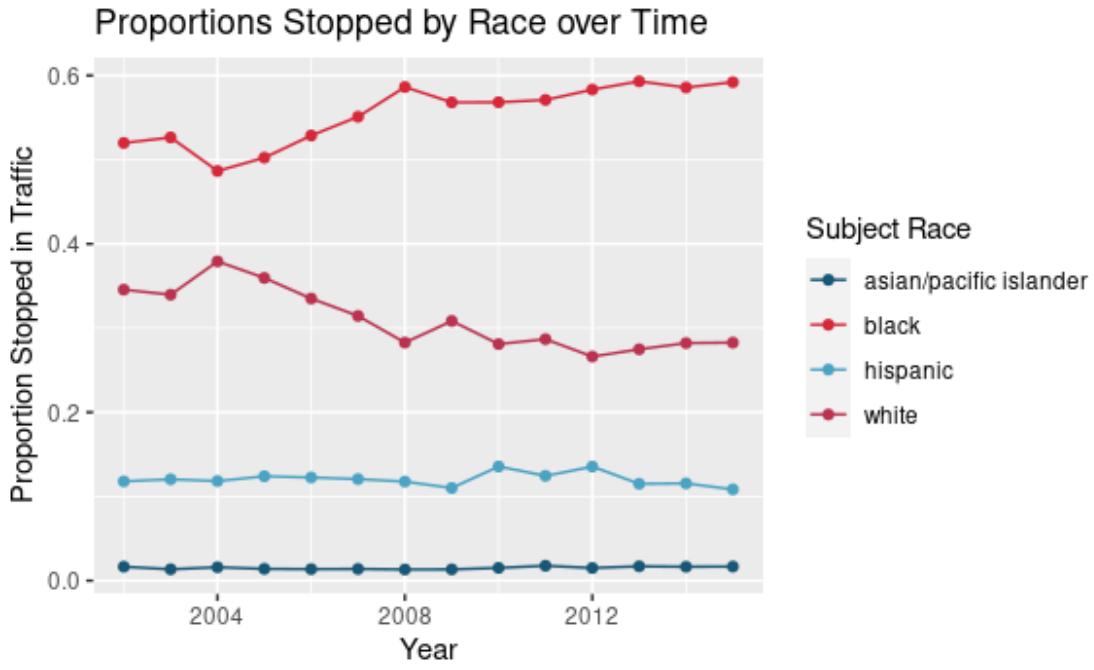
Logistic Regression of Race, Sex, Age, and Race * Sex:

```
## [1] 444182.6
## [1] 444065
```

We obtained both our lowest AIC value here and a lower BIC value, making this our most robust and explanatory logistic regression model.

Appendix 3

This visualization shows stop rates by race over the time period of the data collection. From 2001 to 2015, it does not appear that the proportions of people by race stopped in traffic have significantly shifted towards statistics that match the population proportion; in fact, the visualization seems to indicate that a *greater* proportion of black people were stopped in traffic in 2015 than in 2001 or 2002. This may allude to enduring relevance of our conclusions today.

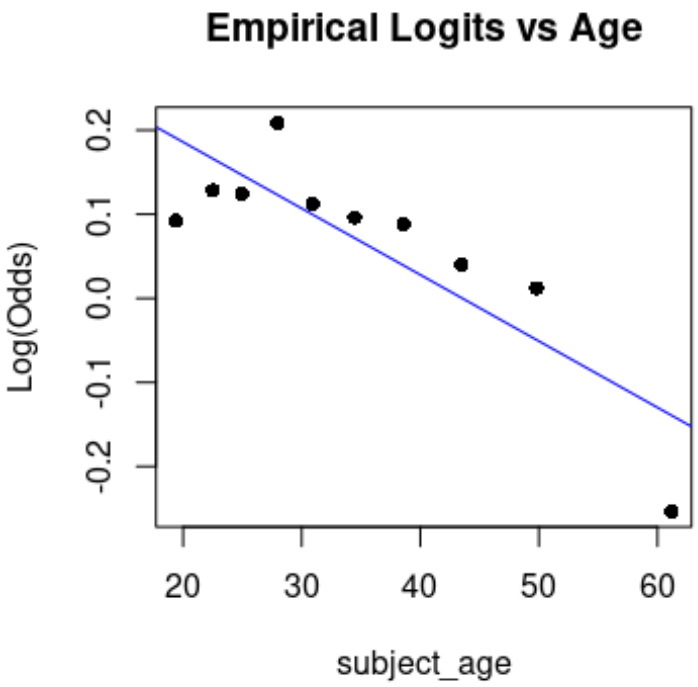


Appendix 4

Conditions of Logistic Regression:

1. Independence - Each traffic stop is independent of other traffic stops; one traffic stop resulting in a citation does not affect the likelihood that other traffic stops result in citations.
2. Linearity - Below, we have depicted a scatterplot of the relationship between age and the empirical log-odds of receiving a citation. The Linearity Assumption is met because there is

a roughly linear relationship between the age of a subject (the quantitative predictor) and the log-odds of receiving a citation.



3. Randomness - We do not have reason to believe that the attempted census of police stops would have results that differ substantially with a full census of police stops.

Conditions met. Proceed with a logistic regression model.

Appendix 5

Let ρ equal the true proportion of stopped drivers who were black within Durham County.

$H_0: \rho = 0.369$. The true proportion of stopped drivers who were black within Durham County is equal to the true proportion of black people within Durham County (0.369).

$H_A: \rho > 0.369$. The true proportion of stopped drivers who were black within Durham County is greater than the true proportion of black people within Durham County.

α = 0.01

```
## [1] 0
```

Because our p-value of 0 is less than our α of 0.01, we reject the null hypothesis. There is sufficient evidence to indicate that the true proportion of people who are stopped within Durham County that are black is greater than the proportion of black people within the Durham County population (0.369). This indicates that black people are disproportionately stopped at a higher rate.

Appendix 6

We conducted a chi-squared test of independence to determine if a person's race is associated with a higher chance of receiving a citation upon being stopped.

H_0: Race and the likelihood of receiving a citation upon being stopped are not associated.

H_A: Race and the likelihood of receiving a citation upon being stopped are associated.

α = 0.01.

The chi-squared test for independence outputted a statistic of 2785.354. The distribution of the test statistic is a chi-squared distribution, which is unimodal and right-skewed with 4 degrees of freedom.

Since our p-value of 0 is less than our α of 0.01, we reject the null hypothesis. There is sufficient evidence to indicate that race and the likelihood of receiving a citation upon being stopped are associated.

Appendix 7

This logistic regression model uses the log-odds of receiving a citation for a white female as the referent, and a series of four equations (one for each race) can be created based on the coefficients from the table.

Table 2: Logistic regression model for the likelihood of receiving a citation

term	estimate	std.error	statistic	p.value
(Intercept)	0.373	0.015	25.484	0.000
subject_raceasian/pacific islander	0.031	0.048	0.634	0.526
subject_raceblack	-0.127	0.013	-9.899	0.000
subject_racehispanic	0.357	0.025	14.458	0.000
subject_raceother	-0.253	0.105	-2.402	0.016
subject_age	-0.007	0.000	-26.459	0.000
subject_sexmale	0.032	0.013	2.443	0.015
subject_raceasian/pacific islander:subject_sexmale	-0.043	0.061	-0.715	0.474
subject_raceblack:subject_sexmale	-0.121	0.016	-7.364	0.000
subject_racehispanic:subject_sexmale	-0.004	0.029	-0.136	0.892
subject_raceother:subject_sexmale	0.151	0.125	1.206	0.228