### 

### The Impact of Racial Bias of Stop and Frisk Tactic

### In The New York Police Department

S. Alexander Zaman  
Melih Elibol  
Roberto Brooks

Final Project - Fall 2015

Stat E-139

### 

### 

### 

### 

### 

### 

### 

### 

### 

### Table of Contents

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| I. | Background | |  |  |  | 3 |
|  | A. | Introduction | | |  | 3 |
|  | B. | Existing Research | | |  | 3 |
| II. | Strategy | | | |  | 3 |
|  | A. | Research Questions | | |  | 3 |
|  | B. | Dataset | | |  | 4 |
| III. | Hypothesis Testing | | | |  | 5 |
|  | A. | Question 1 | | |  | 5 |
|  | B. | Question 2 | | |  | 6 |
|  | C. | Question 3 | | |  | 7 |
|  | D. | Question 4 | | |  | 9 |
| IV. | Conclusion and Bias Analysis | | | |  | 10 |
|  | A. | Potential Bias | | |  | 10 |
|  | B. | Conclusion | | |  | 11 |
|  |  |  | | |  |  |
|  | References | | | |  | i |

# BACKGROUND

### A. Introduction

Recently there has been a lot of press around racism in America, particularly racism in the form of excessive force used by police. Cases of police using excessive force are all over the news and protests such as #BlackLivesMatter have come up in response.

One controversial policing tactic is the “Stop and Frisk” tactic, where police can stop and investigate pedestrians. Law enforcement argues that this tactic is effective at preventing crime by getting guns and contraband off the street, particularly in high-risk neighborhoods. However, many civil liberties advocates argue this practice is racially biased and many of these stops were not based on reasonable suspicion as defined by the law[[1]](#footnote-0), even though it is constitutionally required by the fourth amendment.

### B. Existing Research

There are numerous studies that have addressed how the “Stop and Frisk” tactic justifies violence[[2]](#footnote-1). While these studies are informative on data, methodology, and factor, none of these studies examined whether the use of force is significantly different for individuals of different race.

With our paper, we hope to add to the existing body of knowledge on the topic of stop and frisk, which already covers issues including: stop and frisk usage on minority precincts, racial difference in the total number of people stopped, and the focus of stop and frisk in impoverished neighborhoods.[[3]](#footnote-2),[[4]](#footnote-3)

# STRATEGY

### A. Research Questions

The NYPD has been practicing “Stop and Frisk” for over ten years. They have made public data collected during these investigations.[[5]](#footnote-4) Racial disparities from analyzing this data set are heavily documented, but is also unsurprising since the tactic is used in poorer, riskier neighborhoods, which have a lot more underrepresented minorities to begin with. Instead of looking at this disparity, our project intends to use this data to address the following research questions:

1. Do NYPD police officers use more/less force and make more/fewer arrests for any particular race?
2. Are police more likely to make stops near the end of the month versus the rest of the month, which may suggest pressure to meet a number or informal quota?

### B. Dataset

Our data set uses the NYPD stop and frisk database. This database offers a complete collection of police reports filed after stopping people and dates back from 2003. The data is very large and has a wide array of columns. We have combined certain binary variables of interest to accurately represent the information in which we are interested. The following derived columns were made from the data set:

A ‘Force Used’ column is generated from other columns indicating use of force by law enforcement officers. If any force was used for a given observation, this column is set to true for that observation, otherwise it is set to false. Following are the types of force provided by the NYPD:

* pf\_hands - Hands
* pf\_wall - Suspect against wall
* pf\_grnd - Suspect on Ground
* pf\_drwep - Weapon Drawn
* pf\_ptwep - Weapon Pointed
* pf\_baton - Baton
* pf\_hcuff - Handcuffs
* pf\_pepsp - Pepper Spray
* pf\_other - Other

A ‘Had Weapon’ column was created. This variable indicates whether the suspect had any kind of weapon. If any of the following indicator columns are set to true for a given observation, this column is set to true, otherwise it is set to false:

* pistol - Pistol
* riflshot - Rifle / Shotgun
* asltweap - Assault Weapon
* knifcuti - Knife or cutting instrument
* machgun - Machine Gun
* othrweap - Other Weapons

Race is an important categorical variable we use in the data set. Each category is converted to a binary variable:

|  |  |  |  |
| --- | --- | --- | --- |
| **Code** | **Race** | **Code** | **Race** |
| **A** | Asian / Pacific Islander | **Q** | White - Hispanic |
| **B** | Black | **W** | White |
| **I** | American Indian / Alaskan native | **U / X\*** | Unknown |
| **P** | Black - Hispanic | **Z** | Other |

\* The codebook states Unknown race as X, but there is no data with X, just U.

Date data was adjusted to test our hypothesis of whether the last week of the month had more stops. We created a new column that extracted the day of the month from the date. Using the 22nd of the month as a cutoff, we created an indicator variable for the end of the month. Days after the 22nd are “end of the month,” and days before are not “end of the month.”

# III. Hypothesis Testing

## Question 1: Is there a difference between the amount of force used between races?

Previous studies show individuals of different races are not equally likely to be stopped. We account for this bias in our experimental design by comparing the **odds** of force being used on individuals stopped. The binary response variable (“Force Used”) naturally led us to consider logistic regression to examine the relationship between race and force used. Our test setup is as follows:

Let

Using R, we constructed a logistic regression model with the dependent variable of the derived field ‘force used’, and each race category converted to binary indicator variables. We ended up with the following fit equation:

(The intercept represents , Asian. See Section II.B “Data Sets” for the race code legend. See Appendix for details on the removal of unknown race.)

Test for significant difference between White and Black Hispanic:

All predictor variables were found to be statistically significant. Our findings suggest a significant difference between the amount of force used for different races. The percent chance of force used during a given stop is as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Asian / P.I. | Black | American Indian | Black Hisp. | White Hisp. | White | Other |
| 17.344% | 22.085% | 18.360% | 25.185% | 23.799% | 16.368% | 21.418% |

Looking at the confidence intervals, we see that there exists a significant difference. Particularly, you can see how the confidence bands for whites and/or asians do not overlap with those of blacks or hispanics. The confidence bands are displayed below:

0.312 % 99.688 %

(Intercept) 17.02384 17.66844

raceB 21.31178 22.87897

raceI 17.07308 19.71985

raceP 24.27503 26.11806

raceQ 22.97356 24.64554

raceW 15.71329 17.04538

raceZ 20.50025 22.36450

\* Data in Percentages

\*\* Confidence interval is Bonferroni Corrected

The confidence band significance was further verified using a t-test of linear combinations, which yield significant p-values under Bonferroni corrections.[[6]](#footnote-5) Thus, we reject the null hypothesis that force used is the same for all races. Further investigation into these results are discussed in the conclusion and analysis.

## Question 2: Are there more or less arrests for any particular race?

The following test is very similar to the previous one. Instead of ‘force used’ as the response variable, we use ‘arrest made.’

Let

A logistic regression model was constructed for data between 2008 and 2014. We look at how likely a stop leads to an arrest depending on race. This provides the following model:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Asian / P.I. | Black | American Indian | Black Hisp. | White Hisp. | White | Other |
| 6.665% | 6.288% | 5.266% | 7.103% | 6.525% | 6.630% | 5.111% |

We find some cases of non-overlapping confidence intervals. People classified as ‘Other’ seem particularly low in their arrest rates when stopped. Moreover, ‘American Indians’ have a confidence band that does not overlap with Whites, Hispanics, or Asians. Following are the confidence bands:

0.312 % 99.688 %

(Intercept) 6.455600 6.880267

raceB 5.891559 6.709396

raceI 4.581102 6.047683

raceP 6.620174 7.617773

raceQ 6.108983 6.969116

raceW 6.189148 7.101550

raceZ 4.690094 5.567876

\* Data in Percentages

\*\* Confidence interval is Bonferroni Corrected

As in the previous analysis, we confirmed the significance further using linear combination t-tests.[[7]](#footnote-6) We reject the null hypothesis that different races are arrested at the same rate when stopped.

## Question 3: Are police more likely to make more stops near the end of the month versus the rest of the month?

To address this question, we created an indicator variable for “end of month”, which is true when it’s the end of the month, and false when it’s not the end of the month. “end of month” is false for the first ¾ of the month, and true for the last ¼ of the month. Since the end of the month was about ¼ of the month and the rest of the month was about ¾ of the month, we multiply the average for the end of the month by 3 to balance counts.

The test statistic of the mean and the setup of comparison made this a good fit for a two sample t-test comparison with the following setup.

Let

We performed a Welch two-sample t-test, and found no difference in the mean number of stops at the beginning of the month and at the end of the month (t = 0.0288, p = 0.977). We performed this test because: (1) We have random, independent samples of individuals stops in different boroughs; (2) We are comparing a dependent continuous variable, number of stops, to an independent categorical variable (stops~month) (3) We have enough observations in each category for the Central Limit Theorem to hold; (4) and the variances of each category are different.

**Welch Two Sample t-test**

data: test.month$BegMonth and test.month$EndMonth

t = 0.028884, df = 93.923, p-value = 0.977

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-3848.652 3962.277

sample estimates:

mean of x mean of y

20984.35 20927.54

# IV. Conclusions and Bias Analysis

## A. Potential Bias

### Analyzing race for force used and arrests

Our statistical analysis found a statistically significant difference between the races analyzed. Black and Black Hispanics had a notably larger chance that force would be used on them if they were stopped. Before producing conclusions, we investigate whether there were other factors that may have biased this.

One thing we investigated was whether any particular race was more likely to have a weapon. Such correlations may indicate force used due to finding weapons. Having a weapon was indeed a significant predictor of force being used. Moreover, the data suggested that those with weapons were over twice as likely to have force used on them.[[8]](#footnote-7) This begs the question, maybe races with more force used against them were more likely to be found with a weapon.

Our data found that there was a significant difference in the likelihood of any particular race having a weapon. We found that whites had a significantly higher probability than any other race to carry a weapon when they were stopped. If we were to follow our initial supposition that those with weapons were more likely to have force used upon them, then this data would have suggested that whites stopped should have more force used against them than the other group. Surprisingly, we instead found that they are one of the least likely to have force used against them.[[9]](#footnote-8)

This potential bias source also applies for the arrests made analysis. However, even though we find statistically significant treatment for the ‘Other’ and ‘American Indian’ race, they don’t stand out in any way with regards to arrests made.

## B. Conclusion

Our analysis set out to find statistical trends in the NYPD Stop and Frisk data set that would give us better insight to the following questions:

1. Do NYPD police officers use more/less force and make more/fewer arrests for any particular race?
2. Are police more likely to make stops near the end of the month versus the rest of the month?

We believe this offers an important contribution as it could allow for a better understanding of the fairness of NYPD practices.

It is important to note that our findings are based on an existing data set of police reports and is by nature, an observational study. Thus, no causational conclusions can be made from the data, only correlational. Moreover, there are so many social, economic, and other factors that contribute to the data that our models, even backed with strong statistical significance, should be accepted as insight to further investigation, not conclusive evidence of any trends.

### Uneven use of force between races

Our statistical analysis of Force used for each race shows that, in the data set, those identified in the reports as Black and Hispanic are significantly more likely to have force used against them when stopped than those identified as ‘White’ and ‘Asian’.

What makes this data more interesting is that although having a weapon is associated with a much larger probability of force being used, only the ‘White’ group were statistically more likely to have a weapon in the 2008 - 2014 data reviewed.

The fact that although whites are significantly more likely to have a weapon than blacks or hispanics but significantly less likely to have force used against them offers evidence that there may be bias in how police officers judge whether the use of force is needed.

Our data also looked at arrests and found that reports of those identified as ‘American Indians’ and ‘Other’ in the reports were statistically less likely to be associated with an arrest being made. Although, this is interesting, it is not clear what potential reasons may be associated with this. However, it is important to note, though, that they are the two least represented racial groups in the data in terms of number of reports between 2008 - 2014.

### There is no evidence of more stops at the end of the month

We aggregated the data and compared the number of stops made in the last week of the month versus the rest of the month. This was with the intent to see if police were responding to any pressures like a ‘monthly stop quota’.

Our t-test did not reject the null hypothesis. We found no difference in the mean number of number of stops at the beginning of the month and at the end of the month (t = 0.0288, p = 0.977). Thus, we found no evidence that there was a significant different between the odds of police making a stop at the beginning of the month versus the rest of the month.

# IV. APPENDIX

## Excluding Unknown Race entries for race related questions

Looking at question one, we plotted the original data and found the following regression:

(See section II.c “Data Sets” for the race code legend. The intercept represents (Asian) )

Call:

glm(formula = FORCE\_USED ~ race, family = binomial(), data = fulltable)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.7618 -0.7065 -0.7065 -0.5979 1.9025

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.561457 0.008222 -189.909 < 2e-16 \*\*\*

raceB 0.300761 0.008432 35.670 < 2e-16 \*\*\*

raceI 0.069280 0.024063 2.879 0.00399 \*\*

raceP 0.472706 0.009665 48.911 < 2e-16 \*\*\*

raceQ 0.397751 0.008637 46.052 < 2e-16 \*\*\*

raceU -0.001128 0.019577 -0.058 0.95404

raceW -0.069610 0.009570 -7.273 3.5e-13 \*\*\*

raceZ 0.261523 0.012032 21.735 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3345680 on 3179027 degrees of freedom

Residual deviance: 3335214 on 3179020 degrees of freedom

AIC: 3335230

Number of Fisher Scoring iterations: 4

The regression summary also found that all the statistics were statistically significant except for the unknown race category.

From 2008 - 2014, Unknown race stops accounted for only 1% of the massive dataset and the data does not serve to help us understand our question nor would excluding them bias the race data.

race N

1: B 1665022

2: Q 788476

3: P 205679

4: W 304519

5: Z 76988

6: A 103184

7: I 13045

8: U 22115

Thus, we opted to refit the data without these confusing points. Refitting the data, the new equation ends up being exactly the same as the old but without the Unknown race predictor and all the coefficients are significant:

Call:

glm(formula = FORCE\_USED ~ race, family = binomial(), data = table)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.7618 -0.7065 -0.7065 -0.5979 1.9025

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.561457 0.008222 -189.909 < 2e-16 \*\*\*

raceB 0.300761 0.008432 35.670 < 2e-16 \*\*\*

raceI 0.069280 0.024063 2.879 0.00399 \*\*

raceP 0.472706 0.009665 48.911 < 2e-16 \*\*\*

raceQ 0.397751 0.008637 46.052 < 2e-16 \*\*\*

raceW -0.069610 0.009570 -7.273 3.5e-13 \*\*\*

raceZ 0.261523 0.012032 21.735 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3324995 on 3156912 degrees of freedom

Residual deviance: 3314822 on 3156906 degrees of freedom

AIC: 3314836

Number of Fisher Scoring iterations: 4

## Beginning vs End of Month Data

The table below depicts the number of stops per month/year (for question 3)

|  |  |
| --- | --- |
| year month BegMonth EndMonth  1: 2010 1 24926 29696  2: 2010 2 24010 22456  3: 2010 3 23795 24679  4: 2010 4 27494 33258  5: 2010 5 31292 29813  6: 2010 6 24354 23876  7: 2010 7 24179 21471  8: 2010 8 21497 24270  9: 2010 9 22486 23247  10: 2010 10 26009 29329  11: 2010 11 26018 23929  12: 2010 12 22508 16693  13: 2011 1 26436 32089  14: 2011 2 32829 28080  15: 2011 3 31159 32679  16: 2011 4 31926 30132  17: 2011 5 30462 29710  18: 2011 6 27638 29091  19: 2011 7 27284 25889  20: 2011 8 23456 25903  21: 2011 9 22883 28121  22: 2011 10 27660 31461  23: 2011 11 32596 25493  24: 2011 12 27713 25034 | year month BegMonth EndMonth  25: 2012 1 31810 37021  26: 2012 2 34715 31197  27: 2012 3 35323 33536  28: 2012 4 28504 26996  29: 2012 5 24350 21842  30: 2012 6 17075 15041  31: 2012 7 14448 18581  32: 2012 8 18045 18194  33: 2012 9 17796 18817  34: 2012 10 19238 17100  35: 2012 11 10884 14447  36: 2012 12 15648 12303  37: 2013 1 17441 19835  38: 2013 2 17965 16882  39: 2013 3 16728 10882  40: 2013 4 10063 10900  41: 2013 5 10348 10823  42: 2013 6 8595 7690  43: 2013 7 5559 5494  44: 2013 8 4001 2289  45: 2013 9 1759 2083  46: 2013 10 2132 2239  47: 2013 11 2197 2086  48: 2013 12 2015 1845 |

Regression Frisked ~ Race

The coefficients for the logistic regression of being frisked after being stopped:

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.0794159 0.2253893 0.352 0.72458

binaryendmonth -0.0217576 0.0037518 -5.799 6.66e-09 \*\*\*

raceB 0.5203530 0.0096064 54.167 < 2e-16 \*\*\*

raceI 0.0136102 0.0270490 0.503 0.61485

raceP 0.5052509 0.0119013 42.454 < 2e-16 \*\*\*

raceQ 0.4538757 0.0098554 46.053 < 2e-16 \*\*\*

raceU 0.0707770 0.0221869 3.190 0.00142 \*\*

raceZ 0.2642137 0.0168533 15.677 < 2e-16 \*\*\*

sexZ -0.2708425 0.0163157 -16.600 < 2e-16 \*\*\*

month 0.0078648 0.0005926 13.271 < 2e-16 \*\*\*

cityBRONX 0.1553748 0.2251750 0.690 0.49018

cityBROOKLYN -0.3948219 0.2251604 -1.754 0.07951 .

cityQUEENS 0.0206330 0.2251715 0.092 0.92699

citySTATEN IS -0.5382043 0.2253487 -2.388 0.01693 \*

citySTATEN ISLAND -0.0870866 0.2389788 -0.364 0.71555

## Analysis of Weapons Found vs Race

### Force ~ Weapon

Call:

glm(formula = FORCE\_USED ~ HAD\_WEAPON, family = binomial(), data = table)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.1825 -0.6977 -0.6977 -0.6977 1.7505

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -1.288756 0.001376 -936.3 <2e-16 \*\*\*

HAD\_WEAPONTRUE 1.300612 0.009994 130.1 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3324995 on 3156912 degrees of freedom

Residual deviance: 3309043 on 3156911 degrees of freedom

AIC: 3309047

Number of Fisher Scoring iterations: 4

> conf<-confint.lm(fit.forcetoweapon)

> conf[2,] = conf[2,] + conf[1,] # add intercept

> conf.probs<- exp(conf) / (exp(conf) + 1) \* 100

> conf.probs

2.5 % 97.5 %

(Intercept) 21.56069 21.65208

HAD\_WEAPONTRUE 49.73926 50.85346

### Weapon ~ Race

We used a logistical regression to see the likelihood of weapons being found in a stop with race as the predictors.

> # Weapon vs Race

> fit.weapon<-glm(HAD\_WEAPON~race, data=table, family=binomial())

> summary(fit.weapon)

Call:

glm(formula = HAD\_WEAPON ~ race, family = binomial(), data = table)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.1991 -0.1677 -0.1522 -0.1522 3.0813

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -4.51589 0.03010 -150.043 < 2e-16 \*\*\*

raceB 0.06275 0.03096 2.027 0.0427 \*

raceI -0.20648 0.09838 -2.099 0.0358 \*

raceP 0.22782 0.03563 6.393 1.62e-10 \*\*\*

raceQ 0.25810 0.03159 8.170 3.08e-16 \*\*\*

raceW 0.60438 0.03281 18.420 < 2e-16 \*\*\*

raceZ -0.22251 0.04915 -4.527 5.98e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 436130 on 3156912 degrees of freedom

Residual deviance: 434684 on 3156906 degrees of freedom

AIC: 434698

Number of Fisher Scoring iterations: 7

> getProbs(fit.weapon)

(Intercept) raceB raceI raceP raceQ raceW raceZ

0.010815630 0.011507956 0.008815638 0.013545379 0.013956037 0.019617824 0.008676677

> conf<-confint.lm(fit.weapon, level=(1-0.05/8) )

> intercept<-conf[1,1:2]

> conf[2:7,1] <- conf[2:7,1] + intercept[1]

> conf[2:7,2] <- conf[2:7,2] + intercept[2]

> conf.probs<- exp(conf) / (exp(conf) + 1) \* 100

> conf.probs

0.312 % 99.688 %

(Intercept) 0.9969695 1.173249

raceB 0.9755693 1.357064

raceI 0.6220436 1.247998

raceP 1.1342368 1.616928

raceQ 1.1815371 1.647807

raceW 1.6568907 2.321454

raceZ 0.6998071 1.075356

## t-test for Linear Combinations of Races

The following setup shows the t-test for the null hypothesis “force used is the same for all races.”

Simultaneous Tests for General Linear Hypotheses

Fit: glm(formula = FORCE\_USED ~ race, family = binomial(), data = fulltable)

Linear Hypotheses:

Estimate Std. Error z value Pr(>|z|)

raceP - raceW == 0 0.542315 0.007056 **76.86** **<2e-16** \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Adjusted p values reported -- single-step method)

Our null hypothesis was as follows: . Since we found a significant difference above, we can reject the null hypothesis.

1. “Stop and Frisk Data”. NYCLU - http://www.nyclu.org/content/stop-and-frisk-data [↑](#footnote-ref-0)
2. Simmons, Kami Chavis.“The Legacy of Stop and Frisk: Addressing the vestiges of a violent police culture”. 2014 [↑](#footnote-ref-1)
3. Goel, Rao, & Shroff. “Precinct or Prejudice”. 2015 [↑](#footnote-ref-2)
4. Coviello and Persico “An Economic Analysis of Black-White Disparities”. 2013. [↑](#footnote-ref-3)
5. “The Stop, Question and Frisk Data”. nyc.gov - <http://www.nyc.gov/html/nypd/html/analysis_and_planning/stop_question_and_frisk_report.shtml> [↑](#footnote-ref-4)
6. An example of this methodology can be found in the appendix, “t-test for Linear Combinations of Races”. [↑](#footnote-ref-5)
7. An example of this methodology can be found in the appendix, “t-test for Linear Combinations of Races”. [↑](#footnote-ref-6)
8. Data can be found in the appendix under “Analysis of Weapons Found with Race” [↑](#footnote-ref-7)
9. ibid. [↑](#footnote-ref-8)