Final Project: New York City Stop-Question-Frisk Analysis

Huey Fields, Nandeeka Nayak, William Meng, Aomsin Pongpiriyakarn

Introduction

Stop-Question-Frisk (SQF) is a New York City crime policy that allows officers to stop, question, and frisk potential suspects in order to remove illegal weapons and contraband and reduce crime in New York. All stops made using the policy are required to be recorded by the officer. The report includes the race, age, and gender of the suspect; when, where and for how long the stop occurred; and some information on the result of a stop—including whether or not the suspect was frisked, whether or not a weapon or contraband was found, and whether or not an arrest was made.

Through the first part of this report, we did some initial data analysis. We came to a number of both surprising and unsurprising conclusions. First, suspects whose race made up a significant proportion of those stopped, like African Americans and Hispanics, were more accurately reported in the data than those with low representation, like Asian/Pacific Islander and Native American and Alaska Native. Race, location, and the ID show had a large impact on whether or not a suspect was frisked. Gender had little to no effect on the distribution of ages. Finally, none of the binary factors, including whether or not an an arrest was made, weapons were found, and suspects were frisked, did a good job predicting the variability in the amount of time police stopped suspects.

In the second part, we chose to substantiate the claim made by the NYCLU that Black and Latino New Yorkers were more likely to be frisked than White New Yorkers, yet they were less likely to be found with a weapon. In doing so, we looked at a number of other claims. First, we showed that NYPD officers frequently frisked people even if there was little to no evidence that they had a weapon. Then, we proved that the proportion of African Americans and Hispanics stopped was greater than their proportion in the city. Finally, we showed that these minorities were less likely to be found with a weapon than their white counterparts ("Stop and Frisk During the Bloomberg Administration" 2014).

Through this project, we hoped to determine the effectiveness of the Stop-Question-Frisk policy and understand whether or not its use was racially biased. Stop-Question-Frisk was one of the many polarizing issues in the 2016 presidential election. One side argued that police officers using the policy unfairly targeted minorities like African Americans and Hispanics. The other maintained that the policy was effective in reducing crime. A thorough analysis of the data showed that the policy was often abused and minorities were often racially profiled, even though—as the NYCLU points out—crime decreased under Stop-Question-Frisk ("Stop and Frisk During the Bloomberg Administration" 2014).

Required Questions

Question 1

Due to inaccuracies in police officers' ability to identify suspects, we want to study how such misidentifications affect the collected data. In this question, we will look at the subset of the 2010 data for which suspects were identified as Asian/Pacific Islander or Hispanic to determine the probability that a stopped suspect is Hispanic, the probability that a stopped suspect is Asian/Pacific Islander, the probability that someone who is identified by the officer as Hispanic is actually Hispanic, and the probability that someone who is identified as Asian/Pacific Islander is actually Asian/Pacific Islander.

Given that a Hispanic person has a 95% chance of correctly being identified as Hispanic, and otherwise is misidentified as Asian/Pacific Islander, and as Asian/Pacific Islander has a 95% chance of correctly being identified as Asian/Pacific Islander, and otherwise is misidentified as Hispanic.

Table of the total number of stops for each race

```
##
## AMERICAN INDIAN/ALASKAN NATIVE         ASIAN/PACIFIC ISLANDER
##                           2579                          19732
##                          BLACK                 BLACK-HISPANIC
##                         315083                          38689
##                          OTHER                        UNKNOWN
##                          15360                           4395
##                          WHITE                 WHITE-HISPANIC
##                          54810                         150637
```

The number of Black Hispanic suspects is 38689.
The number of White Hispanic suspects is 150637.
The number of Asian/Pacific Islander suspects is 19732.
Then, the total number of Hispanic and Asian/Pacific Islander suspects is 38689 + 150637 + 19732 = 209058.

Let
Event I: a suspect is identified as Hispanic = 0.906
Event H: a suspect is actually Hispanic = 0.094
Event I^c: a suspect is identified as not Hispanic (i.e. suspect is identified as Asian/Pacific Islander) = 0.997
Event H^c: a suspect is not actually Hispanic (i.e. suspect is actually Asian/Pacific Islander) = 0.494

Then
The probability that a stopped suspect is Hispanic: $P(I)$
The probability that a stopped suspect is Asian/Pacific Islander: $P(I^c)$
The probability that someone who is identified by the officer as Hispanic is actually Hispanic: $P(H|I)$
The probability that someone who is identified as Asian/Pacific Islander is actually Asian/Pacific Islander: $P(H^c|I^c)$

$P(I) = \frac{38689+150637}{38689+150637+19732} = 0.906$
$P(I^c) = \frac{19732}{38689+150637+19732} = 0.094$

We know that a Hispanic person has a 95% chance of correctly being identified as Hispanic, and otherwise is misidentified as Asian/Pacific Islander, so
$P(I|H) = 0.95$

We know that an Asian/Pacific Islander has a 95% chance of correctly being identified as Asian/Pacific Islander, and otherwise is misidentified as Hispanic, so
$P(I^c|H^c) = 0.95$

From the Law of Total Probability, we get
$P(H) = P(H|I)P(I) + P(H|I^c)P(I^c)$

Using Bayes' Rule, we solve for
$P(H|I) = \frac{P(I|H)P(H)}{P(I)}$

Plug in P(H):
$P(H|I) = \frac{P(I|H)[P(H|I)P(I)+P(H|I^c)P(I^c)]}{P(I)}$
Plug in values:
$P(H|I) = \frac{0.95*[P(H|I)*0.906+P(H|I^c)*(0.094)]}{0.906}$ (1)

Using Bayes' Rule, we solve for
$P(H^c|I^c) = \frac{P(I^c|H^c)P(H^c)}{P(I^c)}$
$1 - P(H|I^c) = \frac{P(I^c|H^c)[1-P(H)]}{P(I^c)}$

Plug in P(H) and values:
$1 - P(H|I^c) = \frac{0.95*[1-(P(H|I)*0.906+P(H|I^c)*0.094)]}{0.094}$ (2)

Solving (1) and (2):
$P(H|I) = \frac{4066}{4077} = 0.997$
$P(H|I^c) = \frac{214}{423} = 0.506$
$P(H^c|I^c) = 1 - P(H|I^c) = 1 - 0.506 = 0.494$

In this problem, 99.7% of Hispanics were identified correctly, while only 49.4% of Asian/Pacific Islanders were identified correctly. These calculations show that the race data for races with very little representation in New York (Asian/Pacific Islanders, Native Americans/Alaska Natives, etc) could be highly inaccurate. However, when understanding police bias based on race, the true race of the potential suspect does not matter; instead, it is important to understand how police react to people they perceive to be of that race.

Question 2

By looking at the 2010 data, we will compare the rates at which suspects of each race and each borough were frisked.

Table comparing frisk rates for each race:

```
##
##     AMERICAN INDIAN/ALASKAN NATIVE ASIAN/PACIFIC ISLANDER  BLACK
##   0                           1369                  10626 134526
##   1                           1210                   9106 180557
##
##     BLACK-HISPANIC  OTHER UNKNOWN  WHITE WHITE-HISPANIC
##   0          14604   7512    2423  30738          62081
##   1          24085   7848    1972  24072          88556
```

Linear regression model to compare frisk rates between races (Asian/Pacific islander used as baseline):

```
##
## Call:
## lm(formula = frisked ~ r2, data = sqf2010)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6225 -0.5730  0.4121  0.4269  0.5608
##
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       0.461484   0.003516 131.263  < 2e-16 ***
## r2AMERICAN INDIAN/ALASKAN NATIVE  0.007690   0.010341   0.744    0.457
## r2BLACK                           0.111562   0.003624  30.783  < 2e-16 ***
## r2BLACK-HISPANIC                  0.161044   0.004320  37.277  < 2e-16 ***
## r2OTHER                           0.049454   0.005314   9.306  < 2e-16 ***
## r2UNKNOWN                        -0.012792   0.008237  -1.553    0.120
## r2WHITE                          -0.022294   0.004100  -5.438  5.4e-08 ***
## r2WHITE-HISPANIC                  0.126393   0.003739  33.805  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.4939 on 601277 degrees of freedom
## Multiple R-squared:  0.009626,   Adjusted R-squared:  0.009614
## F-statistic: 834.9 on 7 and 601277 DF,  p-value: < 2.2e-16
```

We chose to use Asian/Pacific Islander as a baseline to compare the frisked rates across all races. Looking at the results from the test above, which compared the frisked rates of Asian/Pacific Islanders to the frisked rates of all the other races, the differences between the Asian/Pacific Islander racial group and the racial groups Black, Black-Hispanic, Other, White, and White-Hispanic showed statistical significance due to p-values of below 0.05. On the other hand, the differences between the frisked rates of Asian/Pacific Islander compared to the frisked rates of American Indian/Alaskan Native and Unknown did not show statistical significance due to p-values of greater than 0.05. We can also conclude that the overall frisked rate was highest among Black-Hispanics since Black-Hispanics showed the greatest positive difference in frisked rates (estimate of positive 0.161). Whites were least likely to be frisked; the estimated difference was only -0.22.

Table comparing frisk rates for each borough:

```
##
##             BRONX BROOKLYN MANHATTAN QUEENS STATEN IS
##   0     30  39920    94311     58940  55905     14773
##   1     30  72495   100844     63142  88167     12728
```

Linear regression model to compare frisk rates between boroughs (Manhattan used as baseline):

```
##
## Call:
## lm(formula = frisked ~ city3, data = sqf2010)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6449 -0.5172  0.3551  0.4828  0.5372
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.5172097  0.0014104 366.703   <2e-16 ***
## city3          -0.0172097  0.0636369  -0.270    0.787
## city3BRONX      0.1276775  0.0020371  62.677   <2e-16 ***
## city3BROOKLYN  -0.0004718  0.0017983  -0.262    0.793
## city3QUEENS     0.0947551  0.0019170  49.428   <2e-16 ***
## city3STATEN IS -0.0543902  0.0032894 -16.535   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4928 on 601279 degrees of freedom
## Multiple R-squared:  0.01382,    Adjusted R-squared:  0.01382
## F-statistic:  1686 on 5 and 601279 DF,  p-value: < 2.2e-16
```

We chose to use Manhattan as a baseline to compare the frisked rates across all the other boroughs. Looking at the results from the test above, the differences between the Manhattan frisked rates and the frisked rates in Bronx, Queens, and Staten Island showed statistical significance due to p-values of below 0.05. The difference in frisked rates between Manhattan and Brooklyn were the least (an estimate difference of 0.00047) while the difference in frisked rates between Manhattan and Bronx were the greatest (an estimate of 0.128). We can also conclude that the overall frisked rate was highest in Bronx since Bronx showed the greatest positive difference in frisked rates (estimate of positive 0.128). The lowest overall rate is Staten Island since Staten Island showed the most negative difference in frisked rates (estimate of -0.054).

Question 3

In this question, we will look at the 2010 data and compare the age distribution for male suspects with the age distribution for female suspects, and for all suspects.

Histograms and QQ Plots for the distribution of the ages of males and ages of females in the 2010 data

Both histograms seem to be relatively bell-shaped, however, they are both skewed right, meaning that more young people were stopped. The Q-Q Plot shows this. The plot is off the line at the ends partly for this reason, and also partly because noone should be older than 100 or younger than 0.

QQ plots for the ages of male suspects compared to the ages of female suspects as well as the ages of males suspects compared to ages of the entire population:

The age distributions are relatively normally distributed and relatively similar in the center but because of the bounds on age, they do not follow the same pattern at the ends. When we compare ages of male suspects to those of the entire population, even these two distributions seem pretty similar.

The NYCLU concluded that young males were disproportionately targeted ("Stop and Frisk During the Bloomberg Administration" 2014). However, because the distributions of age for different genders had the same shape and their Q-Q plots were linear, we can conclude that there was little to no bias in the ages of people stopped based on their gender.

Question 4

In this problem, we want to estimate the probability that in 2015, a suspect was frisked for the entire population and for suspects who refused to provide identification, using 95% confidence intervals.

Table comparing frisk rates for different types of identification provided:

```
##
##     OTHER PHOTO REFUSED VERBAL
##   0   151  4349     249   2557
##   1   282  8629     390   5956
```

Entire population:

$\hat{p} = \frac{15257}{7306+15257} = 0.6761955$ $n = 22563$

95% confidence interval:
$(\hat{p} - 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}, \hat{p} + 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}})$
$(0.6701, 0.6823)$
We are 95% confident that p, the proportion of stops resulting in frisks for the entire population, is between 0.6701 and 0.6823.

Refused to give identification:

$\hat{p} = \frac{390}{249+390} = 0.6103286$ $n = 639$

95% confidence interval:
$(\hat{p} - 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}, \hat{p} + 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}})$
$(0.5725, 0.6482)$
We are 95% confident that p, the proportion of stops resulting in frisks for suspects who refused to provide identification, is between 0.5725 and 0.6482.

These confidence intervals seem to suggest that the proportion of stops resulting in frisks for suspects who refused identification is lower than the proportion of stops resulting in frisks for the whole population.

t-test for the difference between the proportion of suspects frisked for the entire population and the proportion of suspects frisked for suspects who refused to give their identification

```
##
##  Welch Two Sample t-test
##
## data:  friskedTotal and friskedRefusedID
## t = 3.368, df = 671.64, p-value = 0.0008005
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.02746674 0.10426706
## sample estimates:
## mean of x mean of y
## 0.6761955 0.6103286
```

The p-value for this test is 0.0008005 and $\alpha = 0.05$, so we reject the null hypothesis that the difference between the proportions is 0. A 95% confidence interval for the difference estimates that the proportion of suspects frisked for the entire population was between 0.02746674 and 0.10426706 greater than the proportion of suspects frisked for suspects who refused to give their identification. This suggests that subjects who refused to give their identification were actually frisked at a lower rate than the general population.

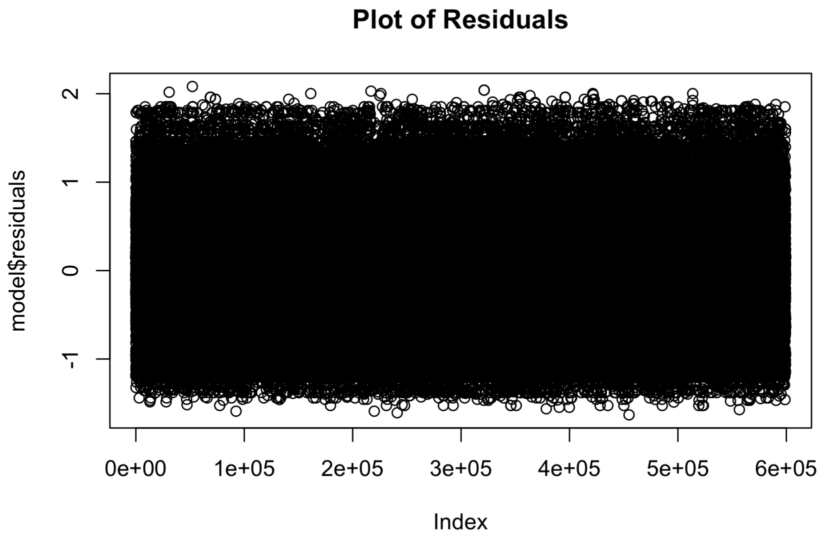
Question 5

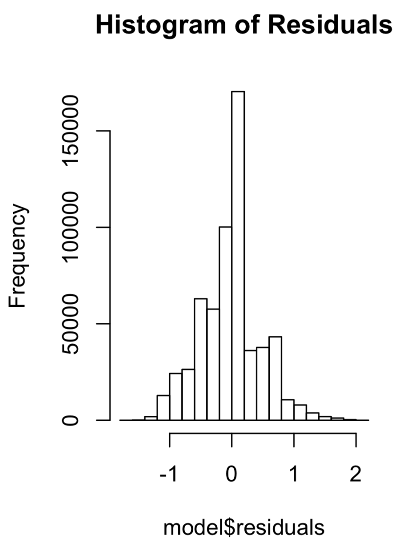
For the 2010 data, we will decide which of the following binary factors: `arstmade`, `searched`, `inside`, `sumissue`, `frisked`, `weap`, `contrabn`, `radio`, and `pf` had a significant effect on the length of the stop (`perstop`) by using linear regression.

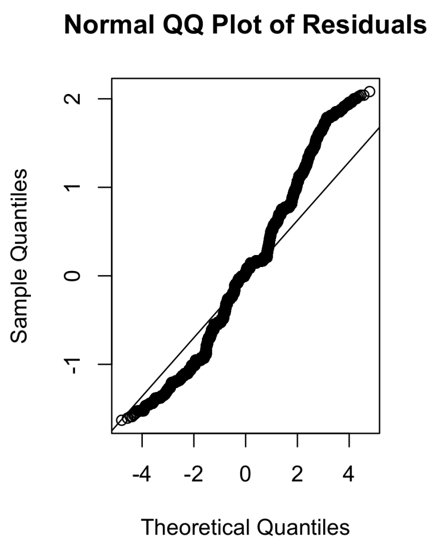
Linear regression for log(`perstop` + 1) against `arstmade`, `searched`, `inside`, `sumissue`, `frisked`, `weap`, `contrabn`, `radio`, and `pf`

```
##
## Call:
## lm(formula = perstopLog ~ arstmadeFiltered + searchedFiltered +
##     insideFiltered + sumissueFiltered + friskedFiltered + weapFiltered +
##     contrabnFiltered + radioFiltered + pfFiltered, data = sqf2010)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63161 -0.26033  0.03396  0.18651  2.08170
##
## Coefficients:
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)       1.623226   0.001167 1390.729  < 2e-16 ***
## arstmadeFiltered  0.105872   0.003188   33.211  < 2e-16 ***
## searchedFiltered  0.085714   0.002662   32.195  < 2e-16 ***
## insideFiltered   -0.041367   0.001556  -26.594  < 2e-16 ***
## sumissueFiltered  0.252071   0.002492  101.136  < 2e-16 ***
## friskedFiltered   0.023394   0.001446   16.176  < 2e-16 ***
## weapFiltered     -0.008253   0.005957   -1.385    0.166
## contrabnFiltered -0.026819   0.005151   -5.207 1.92e-07 ***
## radioFiltered     0.175943   0.001477  119.136  < 2e-16 ***
## pfFiltered        0.058535   0.001664   35.171  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.492 on 599201 degrees of freedom
## Multiple R-squared:  0.0524, Adjusted R-squared:  0.05238
## F-statistic:  3681 on 9 and 599201 DF,  p-value: < 2.2e-16
```

Plot, histogram, and Normal QQ-plot of residuals for regression line







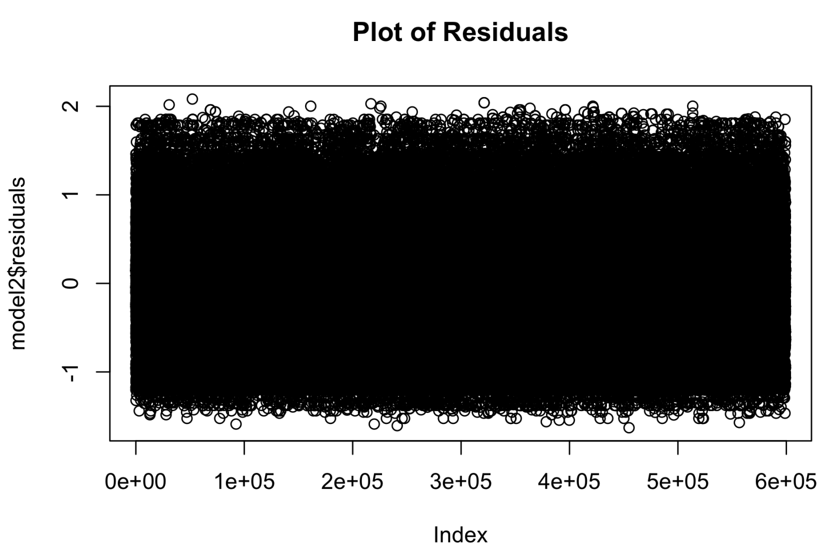
We began by performing linear regression with log(`perstop` + 1) as the dependent variable and 9 explanatory variables: `arstmade`, `searched`, `inside`, `sumissue`, `frisked`, `weap`, `contrabn`, `radio`, and `pf`. `perstop` was filtered to get rid of nonsensical values. To this end, only values of `perstop` greater than 0 and less than 40, and the values of the explanatory variables corresponding to this filtered set, were used to make the regression model.

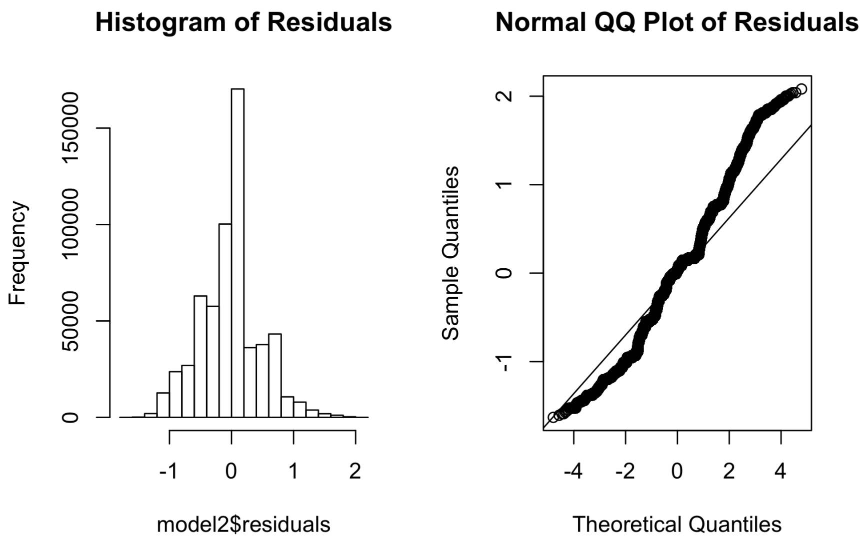
The residual plot, histogram, and QQ plot suggest that the residuals are approximately Normally distributed. With this, we can validate the p-values given in the regression model. While most of the p-values are exceptionally low (less than $2 * 10^{-16}$), the p-value for `weap`, 0.166, is insufficient to reject the null hypothesis that `weap` and log(`perstop` + 1) are unrelated. Because of this, we will create a new regression model with `weap` excluded.

Linear regression for log(perstop + 1) against `arstmade`, `searched`, `inside`, `sumissue`, `frisked`, `contrabn`, `radio`, and `pf`

```
##
## Call:
## lm(formula = perstopLog ~ arstmadeFiltered + searchedFiltered +
##     insideFiltered + sumissueFiltered + friskedFiltered + contrabnFiltered +
##     radioFiltered + pfFiltered, data = sqf2010)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.63035 -0.26030  0.03396  0.18657  2.08173
##
## Coefficients:
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)       1.623242   0.001167 1390.807  < 2e-16 ***
## arstmadeFiltered  0.105287   0.003160   33.321  < 2e-16 ***
## searchedFiltered  0.085158   0.002632   32.355  < 2e-16 ***
## insideFiltered   -0.041408   0.001555  -26.625  < 2e-16 ***
## sumissueFiltered  0.251968   0.002491  101.139  < 2e-16 ***
## friskedFiltered   0.023357   0.001446   16.153  < 2e-16 ***
## contrabnFiltered -0.026503   0.005146   -5.151  2.6e-07 ***
## radioFiltered     0.175965   0.001477  119.158  < 2e-16 ***
## pfFiltered        0.058520   0.001664   35.163  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.492 on 599202 degrees of freedom
## Multiple R-squared:  0.05239,    Adjusted R-squared:  0.05238
## F-statistic:  4141 on 8 and 599202 DF,  p-value: < 2.2e-16
```

Plot, histogram, and Normal QQ-plot of residuals for regression line





Our linear regression equation involved 8 explanatory variables and took the form:
$\hat{y} = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_4 * X_4 + \beta_5 * X_5 + \beta_6 * X_6 + \beta_7 * X_7 + \beta_8 * X_8$

With the coefficients we found in regression plugged in, we get:
$log(perstop + 1) = 1.623242 + 0.105287 * arstmade + 0.085158 * searched - 0.041408 * inside + 0.251968 * sumissue + 0.023357 * frisked - 0.026503 * contrabn + 0.175965 * radio + 0.058520 * pf$
All of the explanatory variables are binomial variables. According to the regression model intercept, if all of the explanatory variables are 0 (i.e. false), then a stop will take $e^{1.623242} - 1 = 4.069499$ minutes (in order to calculate `perstop` from log(`perstop` + 1), we simply raise e to the power of log(`perstop` + 1) and subtract 1). According to the coefficients of the regression model, if an arrest is made, the stop will last an extra 0.1110294 minutes, if the suspect is searched, the stop will take an extra 0.0888891 minutes, if the stop was made in an inside location, the stop will be 0.0422773 minutes shorter, if a summons was issued, the stop will take an extra 0.2865549 minutes, if the suspect is frisked, the stop will take an extra 0.0236319 minutes, if contraband is found on the suspect, the stop will be 0.0268573 minutes shorter, if there was a radio run during the stop, the stop will take an extra 0.1923963 minutes, and if physical force was used during the stop, the stop will take an extra 0.0602662 minutes.

Some of the coefficients make logical sense. For instance, it is reasonable to expect that a stop would take longer if an arrest were made or the subject was frisked during it. The coefficient for `contrabn`, however, suggests that if contraband is found on a suspect during a stop, the stop will be shorter. This is a little confusing, as it could logically be assumed that finding contraband on a suspect would require additional actions on the part of the officer that would increase the duration of the stop.

As in our first regression model, the residual plot, histogram, and QQ plot suggest that the residuals are approximately Normally distributed. With this, we can validate the p-values given in the regression model.

The R^2 value for our regression model is 0.05239. This tells us that 5.239% of the variability in `perstop` is due to our explanatory variables for the regression (`arstmade`, `searched`, `inside`, `sumissue`, `frisked`, `contrabn`, `radio`, and `pf`). The p-values for the relationship of each explanatory variable to `perstop` in the regression model were very low. The p-value for the relationship between `contrabn` and `perstop` was $2.6 * 10^{-7}$, while all other p-values were below $2 * 10^{-16}$. The p-values allow us to reject the null hypothesis that each explanatory variable is unrelated to log(`perstop` + 1) suggest that there is some relationship between log(`perstop` + 1) and all of the explanatory variables included in the regression model.

Our low R^2 value suggests that there a great deal of variance in the data. This makes sense, as the explanatory variables we used for the regression model are all binomial. Because all of the explanatory values only range from 0 to 1, with no values in between, and there is a very large sample size, there will naturally be a great deal of variance in the values of log(`perstop` + 1) corresponding to each of these variables.

Chosen Problem

Note: In this problem we make the assumption that a weapon found on a suspect in a stop constitutes a "successful" stop in which a crime was prevented.

The null hypotheses for the t-tests conducted in this problem are based off of the assumption that the data given by the NYCLU in their 2003-2013 report ("Stop and Frisk During the Bloomberg Administration" 2014) are representive of true population parameters.

Claim: Many of the frisks conducted were unjustified

NYCLU says that officers conducted frisks in 51.9 percent of all stops ("Stop and Frisk During the Bloomberg Administration" 2014). We tested this using a two-sided, one-sample t-test with a null hypothesis that p = .519.

t-test for frisk rate of all stops (2010):

```
##
##  One Sample t-test
##
## data:  frisked
## t = 65.849, df = 601280, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.519
## 95 percent confidence interval:
##  0.5598872 0.5623959
## sample estimates:
## mean of x
## 0.5611416
```

Given an alpha level of .05 and a p-value of 2.2e-16, we reject the null hypothesis that 51.9% of stops resulted in the suspect being frisked. However, because the 95% confidence interval is (0.5598872, 0.5623959), we can conclude that even more than 51.9% of people were frisked.

However, the NYCLU claims that weapons were found in only 2.0% of the stops in which frisks were conducted ("Stop and Frisk During the Bloomberg Administration" 2014).

t-test for weapon-find rate in all frisks (2010):

```
##
##  One Sample t-test
##
## data:  foundWithWeapon
## t = 4.361, df = 337400, p-value = 1.295e-05
## alternative hypothesis: true mean is not equal to 0.02
## 95 percent confidence interval:
##  0.02059377 0.02156316
## sample estimates:
##  mean of x
## 0.02107846
```

Given an alpha level of .05 and a p-value of 1.295e-05, we reject the null hypothesis that weapons were only found in 2.0% of frisks. However, because the 95% confidence interval is (0.02059377, 0.02156316), very close to the value given by the NYCLU, we can conclude that the NYCLU's claim that officers were unjustifiably frisking people still holds.

Question 4 of the required questions asked to look at the proportion of people frisked in 2015. In 2014, New York City placed new restrictions on the use of SQF and though the number of SQF incidents have decreased, we wanted to see if they succeeded in reducing the proportion of suspects frisked. Note that variable frisked refers to the number frisked in the 2010 data.

t-test for frisk rate in all stops (2015)

```
##
##  Welch Two Sample t-test
##
## data:  frisked and frisked2015
## t = -36.177, df = 24505, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1212875 -0.1088205
## sample estimates:
## mean of x mean of y
## 0.5611416 0.6761955
```

Given an alpha level of .05 and a p-value of 1.295e-05, we reject the null hypothesis that the number of people frisked in 2010 and 2015 was the same. Even after the new restrictions in 2014 on the use of SQF, the proportion of people frisked increased, meaning that the new restrictions were not effective in reducing abuse of the policy.

Claim: Blacks/Latinos were Disproportionately more likely to be "targeted" or stopped:

2010 Census ("Results from the 2010 Census- Population Growth and Race/Hispanic Composition" 2010, 6):
NYC population: 8,175,133
Black/African American nonhispanic: 1,861,295 = 22.8%
Hispanic Origin: 2,336,076 = 28.6%

2010 SQF Data:
Total Stopped: 601285
Black: 315083 = 52.4%
Hispanic Origin: 189326 = 31.5%

Null Hypothesis: Blacks stopped by the SQF program was 22.8%, which is the proportion of Blacks in the NYC population.
Alternative Hypothesis: Blacks stopped by the SQF program was higher or lower than the actual proportion of Blacks in NYC.

Hypothesis test:

```
##
##  One Sample t-test
##
## data:  stopsBlack2
## t = 459.61, df = 601280, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.228
```

```
## 95 percent confidence interval:
##  0.5227537 0.5252784
## sample estimates:
## mean of x
## 0.5240161
```

We reject the null hypothesis due to a p-value of less than 0.05. Therefore, the proportion of Blacks stopped does not represent the actual proportion of Blacks living in New York City. The 95% confidence interval is positive 52.3% to 52.5% and this indicates that Blacks were stopped 29.5% to 29.7% more compared to the actual population proportion of 22.8%.

Null Hypothesis: Hispanics stopped was 28.5%, which is the proportion of Hispanics in the NYC population. Alternative Hypothesis: Hispanics stopped by the SQF program was higher or lower than the actual proportion of Hispanics in NYC.

Hypothesis test:

```
##
##  One Sample t-test
##
## data:  stopsHisp2
## t = 49.866, df = 601280, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.285
## 95 percent confidence interval:
##  0.313695 0.316043
## sample estimates:
## mean of x
##  0.314869
```

We reject the null hypothesis due to a p-value of less than 0.05. Therefore, the proportion of Hispanics stopped does not represent the actual proportion of Hispanics living in New York City. The 95% confidence interval is positive 31.4% to 31.6% and this indicates that Hispanics were stopped 2.9% to 3.1% more compared to the actual population proportion of 28.5%.

Therefore, we can conclude that both Blacks and Hispanics were stopped disproportionately more compared to their actual population.

Claim: The proportion of stops in which Blacks and Latinos were found carrying weapons was less than the proportion of stops in which Whites were found carrying weapons

Table by race of suspects upon whom weapons were and were not found (0 corresponds to weapon not found, 1 corresponds to weapon found):

```
##
##                                      0      1
##   AMERICAN INDIAN/ALASKAN NATIVE   2563     16
##   ASIAN/PACIFIC ISLANDER          19538    194
##   BLACK                          311606   3477
##   BLACK-HISPANIC                  38137    552
##   OTHER                           15244    116
##   UNKNOWN                          4344     51
##   WHITE                           53692   1118
##   WHITE-HISPANIC                 148563   2074
```

Proportion of stops resulting in a weapon found for each race:

```
##                            races proportions
## 1 AMERICAN INDIAN/ALASKAN NATIVE 0.006203955
## 2         ASIAN/PACIFIC ISLANDER 0.009831745
## 3                          BLACK 0.011035188
## 4                 BLACK-HISPANIC 0.014267621
## 5                          OTHER 0.007552083
## 6                        UNKNOWN 0.011604096
## 7                          WHITE 0.020397738
## 8                 WHITE-HISPANIC 0.013768198
```

NYCLU 2003-2013 proportion of stops resulting in a weapon found for Blacks and Latinos = 0.019 ("Stop and Frisk During the Bloomberg Administration" 2014)
NYCLU 2003-2013 proportion of stops resulting in a weapon found for Whites = 0.033 ("Stop and Frisk During the Bloomberg Administration" 2014)

Testing for Blacks/Latinos proportion:

```
##
##  One Sample t-test
##
## data:  BlackLatinoWeapons
## t = -44.828, df = 504410, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.019
## 95 percent confidence interval:
##  0.01179759 0.01240102
## sample estimates:
##  mean of x
## 0.01209931
```

Testing for White proportion:

```
##
##  One Sample t-test
##
## data:  WhiteWeapons
## t = -20.872, df = 54809, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0.033
## 95 percent confidence interval:
##  0.01921429 0.02158118
## sample estimates:
##  mean of x
## 0.02039774
```

Testing to see if 2010 Black/Latino proportions are different from White proportions:

```
##
##  Welch Two Sample t-test
##
## data:  BlackLatinoWeapons and WhiteWeapons
## t = -13.318, df = 62137, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.009519726 -0.007077132
## sample estimates:
##  mean of x  mean of y
## 0.01209931 0.02039774
```

The proportions of stops resulting in a weapon found on both Whites and Blacks/Latinos in the 2010 data had a statistically significant difference from the figures presented in the NYCLU report. Given the p-values of >2.2e-16 and alpha-level of .05, we can reject the null hypothesis that the proportion of stops resulting in a weapon in 2010 was the same as the proportion from 2003 to 2013 cited in the NYCLU report. The 95% confidence intervals for both proportions suggest that this proportion is even lower than reported by the NYCLU for both Whites and Blacks/Latinos. Furthermore, given a p-value of >2.2e-16 and an alpha-level of .05, we can reject the null hypothesis and conclude that the proportion of Black/Latino stops resulting in weapons found was actually less than the proportion of White stops resulting in a weapon.

Conclusion and Recommendations

Our research concludes that, based on data from 2010 and 2015, the the New York Stop-Question-Frisk program is not very effective at finding weapons and contraband. Out of all the people stopped, more than 50% of the people were frisked, but only about 2% of the people were found with a weapon. The proportion of people frisked even increased from the year 2010 to 2015. We also found that, in 2010, while Blacks and Hispanics were stopped disproportionately more than any other racial group, they were less likely to possess weapon. The Stop-Question-Frisk program was relatively unbiased based on gender and age such that the amount of people stopped for both male and female populations by age showed similar distributions.

From these findings, we recommend that the New York City Police Department decrease the amount of frisks in order to improve the efficiency of the program, yielding a higher percentage of suspects found with weapons or contraband. In order to increase the efficiency, we ask the police department to reconsider their criteria for targeting suspects based on race. We recommend that the police stop disproportionately targeting suspects that are Black or Hispanic since we proved that they were actually less likely to have weapons. We hope that if the program were to continue, the distribution of proportions for race would become more evenly distributed based on the actual proportions while the distribution for gender and age remains fair as it is now. We hope that by making these changes, the program would become more effective in finding weapons and contraband and decreasing the overall violence in New York City.

Bibliography

"Results from the 2010 Census- Population Growth and Race/Hispanic Composition." 2010. *NYC Department of City Planning.*

"Stop and Frisk During the Bloomberg Administration." 2014. *New York Civil Liberties Union.* http://www.nyclu.org/files/publications/stopandfrisk_briefer_2002-2013_final.pdf.