Aria Coalson, Alessandro Maiuolo, Robert Fulton, Maria Croom

Math 035 – Probability and Statistics

Final Project

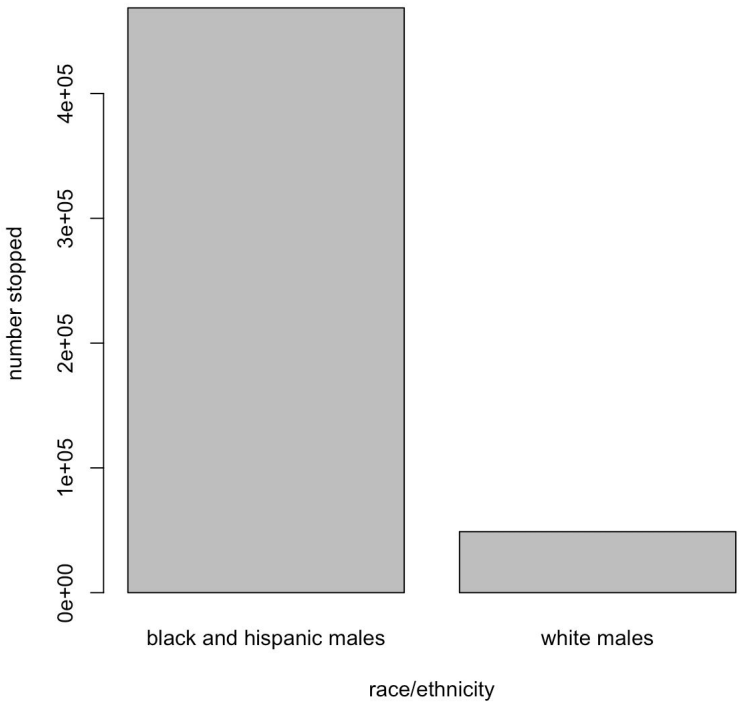
17 December 2018

Stop Question Frisk in New York

The Stop-Question-Frisk Policy, or SQF, was implemented in 2002 and carried through 2013. It continued after 2013 but on a smaller scale. SQF in New York City was particularly controversial. SQF allowed police officers to stop, question, and frisk any individual under suspicion, but grounds for suspicion were very loose to effectively nonexisting. NYCLU analyzed the data and made several claims, and we would like to verify some of those claims with the data we have been provided. Specifically, we wanted to look and see whether the NYCLU discriminates based on race by evaluating multiple factors such frequency of frisks and arrest rates per our groupings.

We conducted a two-sided proportion test to determine whether or not the article's claim of a 2% success rate for frisks was reasonable – success being defined as finding a weapon. The P-value was 0.001843, so we reject the article's claim. The 99% confidence interval (0.0001860155, 0.0019709112) suggested that the actual success rate was 0.02% to 0.20% higher than the article's claim of 2%. Since the article only gave its answer with one significant digit, and the difference is very small, it is possible that they calculated the same success rate as we did and then rounded it to 2%.

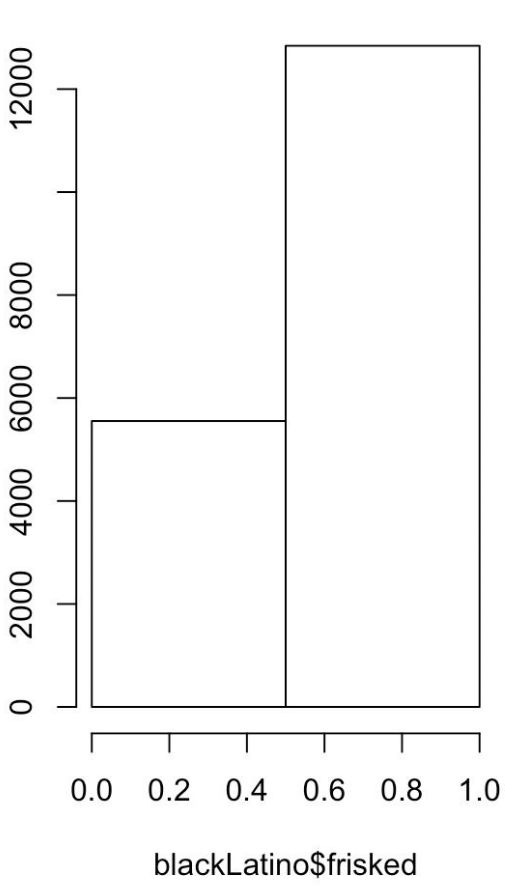
A barplot can help us compare the number of black and hispanic male suspects stopped with the number of white male suspects stopped:

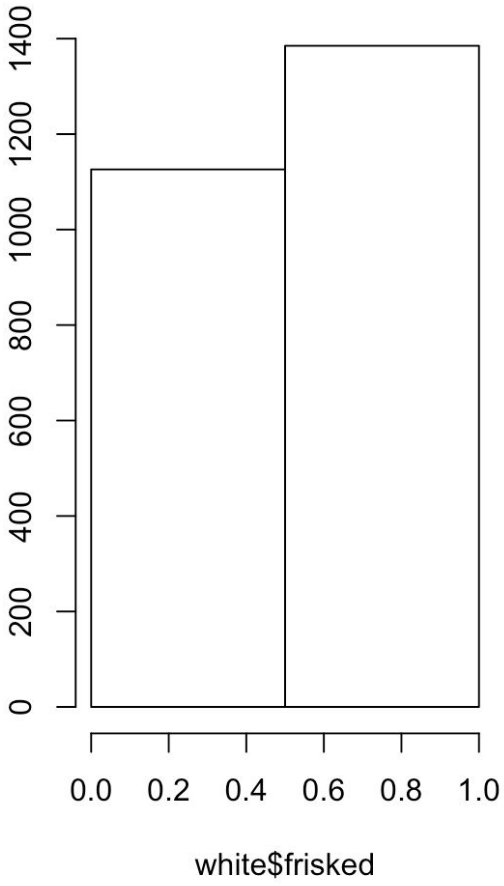


Based on the 2010 data, we'd expect the proportion of black and latino/hispanic males between the ages of 14-24 to be about 27 percent, with a 95 percent confidence interval of (0.2727, 0.2750). The probability that our sample's composition would be this different from the article's claim of 41 percent, assuming that it is accurate, is less than 2.2 x 10^(-16). Assuming that our 2010 sample is representative of the 2003- 2013 population, this would cause us to reject the article's claim with more than 99 percent confidence. This test was conducted using the prop.test function in R and is shown in the Rmd file.

The article claimed that in the years 2002-2013, black and latino people were more likely to be a target of SQF than white people. We wanted to see if this trend

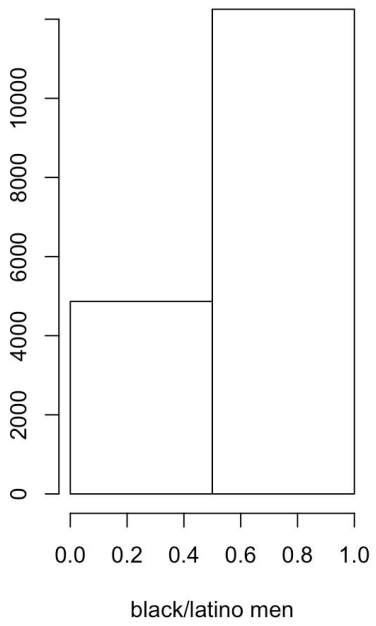
continued in 2015. We wanted to see if the proportion of black and latino individuals getting frisked was different than the proportion of white individuals getting frisked. The bar plots below suggest that the proportion of white people getting frisked was lower:

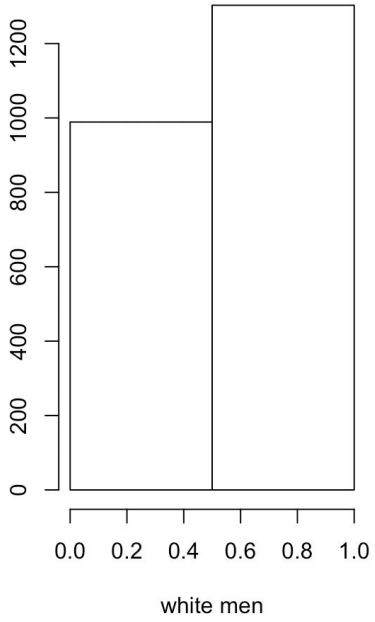


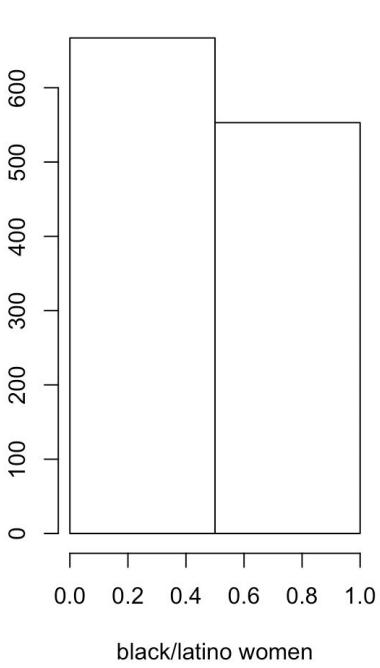


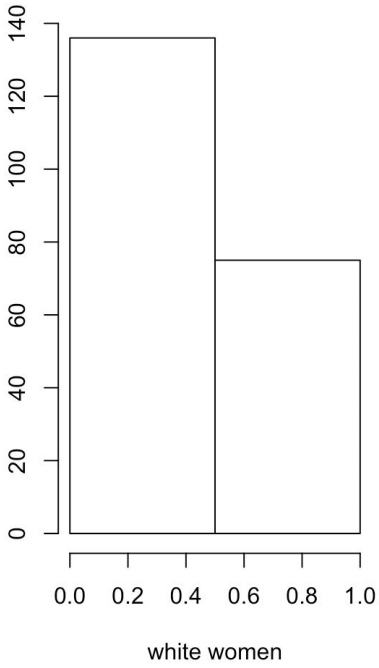
We conducted a two sample proportion test to see if there was a significant difference. The P-value was very small (< 2.2e-16), allowing us the reject the null hypothesis: the two proportions are the same. The 99% confidence interval (0.119298, 0.173772) suggests that the proportion of white individuals getting frisked was lower than the proportion of black and latino individuals getting frisked.

We decided to further explore our question, so we conducted the same test except we used male-only data and female-only data. Here are the bar plots of the four groups:



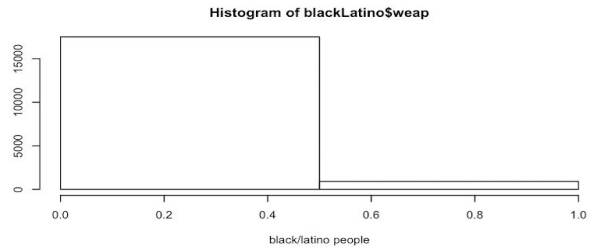


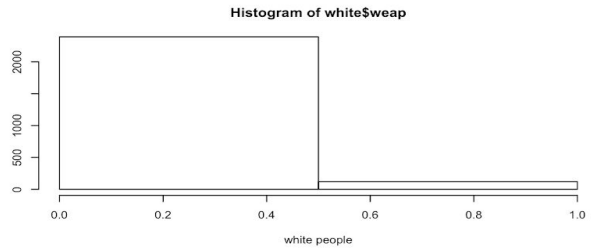


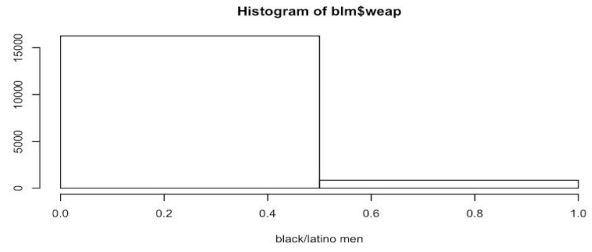


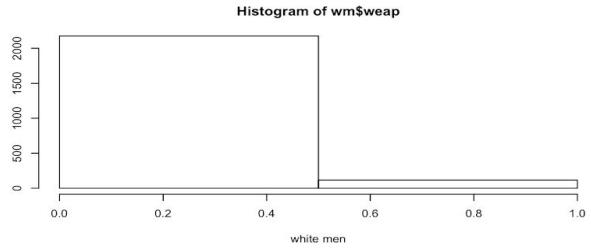
The plots show that black and latino men and women could have a higher likelihood of getting frisked than their white counterparts. After performing the two-sample proportion tests on the male and female groups, the P-values (< 2.2e-16 and 0.0102, respectively) were low enough to reject the null hypothesis: the proportions are independent of race. Both 99% confidence intervals – men: (0.1189444, 0.1756166), women: (0.01152897, 0.18412793) – suggest that black/latino men and women had a higher chance of being frisked than white men and women.

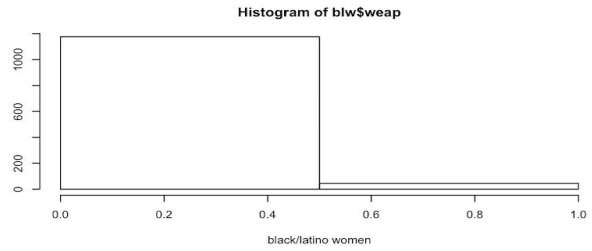
We wanted to see if the difference in frisk rate between white people and black/latino people could be justified by how often a weapon was found. We tested the black/latino and white populations as a whole. We also tested men and women separately. Here are the bar plots for the six groups:

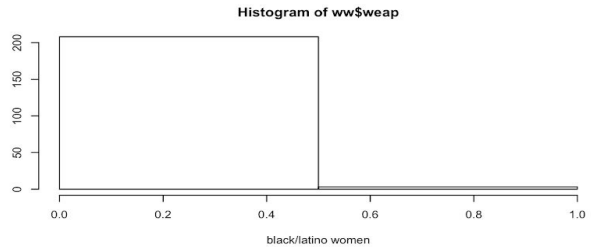






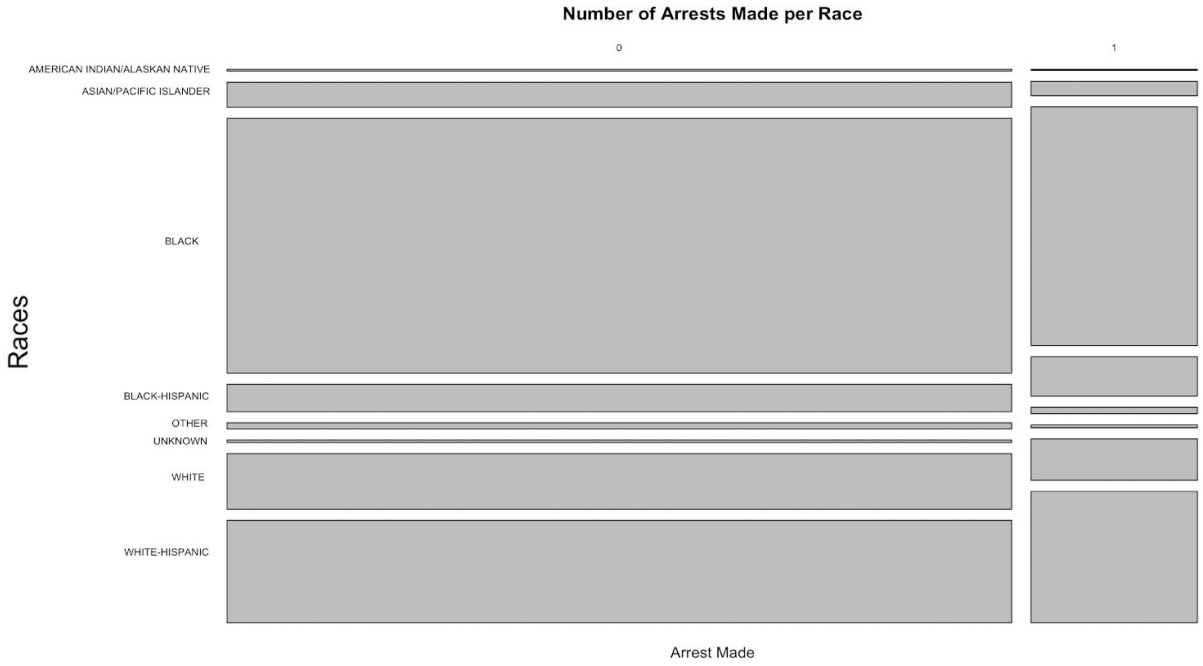






The proportions for the entire group and the male group look the same. The proportions for the female group look like they could be different. We conducted two sample proportion tests for each of the three groups (all genders, male, female). In each test, the P-value was large (all: 0.8238, men: 0.9513, women: 0.1513), so we failed to reject the null: the proportions are equal. This means that we cannot say that the rate at which black/latino people carry weapons and the rate at which white people carry weapons are any different, so the difference in frisk proportions is not justified by the rates at which weapons are found between the two groups.

Based on all of our previous data, it seems that black-hispanics are much more likely to to be targeted by the New York Police than other races, especially those who are white. With a mosaic plot, we can visualize the set numbers of people arrested based on their race.



However, this data does not allow us to see if Black-Hispanics or any other race face discrimination when arrested as it just gives us the numbers of how many people are arrested, which could look deceptively larger or smaller depending on how large the data set is per race.

In order to verify if Black-Hispanics are targeted without looking specifically at the claims made in the article, we decided to use logistic regression to conduct predictive analysis on how race effects arrests made using the number of “white” arrests as a baseline.

Race	Estimate	Z value	P(>|z|)
White (baseline)	-1.84721	-31.749	< 2e-16
American Indian	-0.30745	-0.813	0.415970
Asian/Pacific Islander	-0.25387	-2.249	0.024484
Black	0.23131	3.661	0.000251
Black-Hispanic	0.65282	7.592	**3.15e-14**
White-Hispanic	0.54442	8.066	**7.29e-16**
Other	0.36561	2.281	0.022570
Unknown	0.49138	2.121	0.033959

Out of all the races, only American Indian has a p-value > 0.05 meaning that data is not statistically significant. Data on Black-Hispanics, White-Hispanics, and Blacks have the lowest p values making the conclusions we draw from them the most statistically significant. The baseline of the log odds of being arrested is -1.84. The values in the second column labeled “Estimate” indicate how much the odds for being arrested

increase/decrease based on each race. We can those who are Black, Black-Hispanic, White-Hispanic, an Other Race, or an Unknown Race are more likely to be arrested than someone who is White. Most noticeably, Black-Hispanic people have the highest odds of being arrested with their odds being -1.19439 as opposed to the significantly lower number -1.84721 for White people.

Using this fitted model, we can predict the actual probability of each race being arrested. This yields the following Data:

RACE	White	Black	Black-Hispanic	White-Hispanic	Other	Unknown
$\mathbf{P_{arrested}}$	13.62%	16.58%	23.25%	21.37%	18.52 %	20.49%

The probability of a white person being arrested is 13.62%, the lowest arrest percentage out of all the races sampled. Black-Hispanics, as shown consistently throughout this report have a much higher rate, being arrested around 23.25% (with White-Hispanics' arrest rates closely following at 21.37%).

In conclusion, the Black-Hispanic race seems to be disproportionately targeted by the police. Weapons are found approximately 2% of the time subjects are frisked. One could argue that this is a good success rate, but it is not good enough to outweigh the discrimination present in this policy. While people in the NYC district who are Black and/or hispanic don't make up a significant portion of the total population, they are overrepresented in the 2010 sqf data. Specifically, about 27 percent of the 2010 stops happened to Black or hispanic men aged 14-24. Black and hispanic people are also

more likely to be frisked than the white people, indicating a disproportionate application of the SQF policy. Based on our predictive model, people in the NYC district who are white are much less likely to be arrested than those who are of Black, Black-Hispanic, White-Hispanic, Other, or Unknown race. From all of our data, it seems like Black-Hispanics, an ethnic minority, are more likely to be targeted by the police (in terms of being frisked/arrested) despite them being equally likely to be in possession of a weapon showing the double standards within the NYCLU police department.

Bibliography

https://www.nyclu.org/sites/default/files/publications/stopandfrisk_briefer_2002-2013_final.pdf