Propensity Scores in San Diego Policing

A LOOK INTO SAN DIEGO PD TRAFFIC STOPS & BIAS

Paul Gagliardi | Statistics & Probability 2 | 8/20/2021

Introduction

Policing in the United States, especially in the last few years, has been a topic of contention in politics. Race, gender, sexual orientation, and other factors are often at the core of the debates in policing.

This paper aims to explore the effects of discrimination around race and gender when a driver is pulled over in the San Diego Area, using the concept of Propensity Scores to determine who is most likely to be searched or arrested when a traffic stop is made. An individual with a higher propensity score around searches is more likely to be searched, so effectively the highest propensity score individuals indicate the highest level of police bias.

Data Source

I obtained this dataset from an aggregated police record database at https://www.policedatainitiative.org/, and from the extension there examining San Diego PD traffic stops https://data.sandiego.gov/datasets/police-vehicle-stops/. This data is from Oct 2017 – June 2018, and contains 16 variables:

- stopID, stop_cause, service_area, **subject_race, subject_sex, subject_age,** date_time, sd_resident, **arrested, searched,** obtained_consent, contraband_found, property_seized, action, search_type, search_basis

Above, the variables examined are in boldface, and they are all binomial in nature. Every row in the data represents a traffic stop, and the columns contain the variables above.

All subjects were stopped at a routine traffic stop. We are looking at their different treatment from that point given their information around Sex and Race.

DATA EXPLORATION

In a brief exploration of this data, I examined the relative likelihood of a subject being arrested given 2 binomial factor covariates: Male vs. female and White vs Non-White. This was done by simply comparing the sample populations of each; what percent of males/females/whites/non-whites were searched or arrested. Immediately we can see some intriguing data points

- 10.4% of stops on males ended in searches vs 4.5% of females
- 10.2% of non-whites were searched vs 5.9% of whites

With this info, we can assume that propensity scores for each group will be different. Going forward, we can consider and refer to 4 groups: White Males (WM), White Females (WF), Non-White Males (NWM), and Non-White Females (NWF).

Model and Addressing Research Questions

For this project, Propensity Score Matching was used, integrally the MatchIt package in R. Propensity scores estimate the probability that an individual would have received a particular treatment based on observed baseline characteristics. This is used often in education and healthcare to examine the effects of different schools or health treatment options. In this case, the "treatment" that the scores are looking at are subject searches and arrests. Put simply, how likely is a subject to be arrested or searched given their inherent traits.

MatchIt attempts to match samples of the searched/arrested and control (not searched or arrested) groups with similar covariate distributions. This simplifies things significantly in a group with only two binomial variables: there can only be four scores, one for each group (WM, WF, NWM, NWF). Matching these scores to events is then quite easy, as it become a probability function for each group. WM have a probability to get searched based on the data, and therefore have perfect matching when using the propensity score probability as that will determine precisely how likely given the factors that a WM will be searched. Exactly as many WM in the data set will be searched in a MatchIt function as were actually searched, because the propensity score represents that exact probability. See for more matching at the appendix.

Dataset Visualization and Results

Searches

Running the MatchIt package on the search data provided the following results (Abridged with ....., see code for full output) Model:

Summary of Balance for Matched Data:

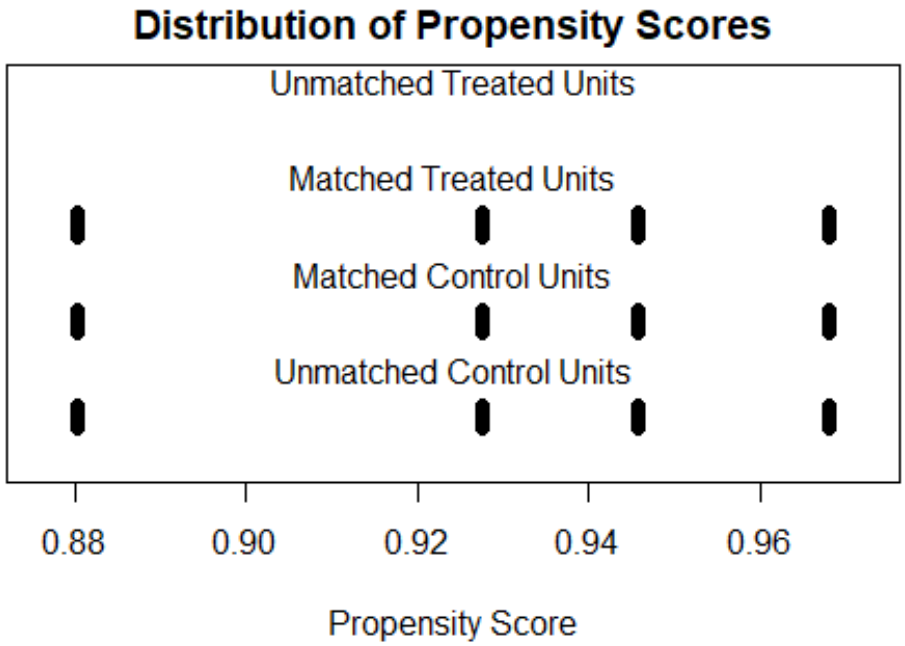
	Means Treated	Means Control	Std. Mean Diff.
distance	0.9003	0.9003	0
subject_raceW	0.2334	0.2334	0
subject_raceNW	0.7666	0.7666	0
subject_sexF	0.1514	0.1514	0
subject_sexM	0.8486	0.8486	0

…..

Sample Sizes:

	Control	Treated
All	31289	2986
Matched	2986	2986
Unmatched	28303	0
Discarded	0	0

Here the means treated represent the relative propensity scores of each group scaled*[1] onto a jitter plot, they are represented as such.



ANALYSIS

Females had the lowest score, followed by white subjects. They represent the lowest propensity score or lowest probability. Males are the most likely group, and non-whites the second most likely group. Because the Male score was highest, this is the most decisive factor in determining a high probability of search when pulled over, though being a non-white subject is nearly as significant.

Distance on the previous page represents some degree of confidence in the matching, however examining the perfect control/treated table and the jitter plot we can see every subject due to the binary nature of the study was matched perfectly to one of 4 scores.

[1] *Not sure mathematically how these are scaled, but the maximum score is taken to 1

Arrests

Summary of Balance for Matched Data:

	Means Treated	Means Control Std.	Mean Diff.
distance	0.9638	0.9638	0
subject_raceW	0.3076	0.3076	0
subject_raceNW	0.6924	0.6924	0
subject_sexF	0.1932	0.1932	0
subject_sexM	0.8068	0.8068	0

Sample Sizes:

	Control	Treated
All	33095	1180
Matched	1180	1180
Unmatched	31915	0
Discarded	0	0

ANALYSIS

Females had the lowest score again, and white subjects also had a low score but both were less significant than in the search determination (all scores grew toward 0.5) .Males are the most likely group, and non-whites the second most likely group. Because the Male score was highest, this is the most decisive factor in determining a high probability of arres when pulled over, though being a non-white subject is nearly as significant (not as significant as for searches).

Distance is a little bit better here, given a better degree of general confidence on the scores, the max being 1.

Conclusion

I derived two solid things from this study, first to name the study itself. Propensity score matching is a poor fit for this rather simple interaction, and though the results we received are interesting I would rather have studied propensity score matching through a different dataset.

Regarding the results of the scores, we can determine 4 significant answers given our solutions, and draw one general conclusion with regards to this dataset and timeframe:

1. Males are significantly more likely to be searched than females when pulled over
2. Non-whites are significantly more likely to be searched than whites when pulled over
3. Males are more likely to be arrested than females when pulled over, but this is less than a determination than in searches
4. Non-whites are more likely to be arrested than whites when pulled over, but this is less than a determination than in searches

We can conclude from here that it is fortunate indeed to be pulled over as a non-white male. By basis of your birth, you are likely to have a higher skepticism about you.

Appendix:

MATCHING

There are detailed and very descriptive methods of matching, but for the purpose and length of this paper they are best looked into here: https://cran.r-project.org/web/packages/MatchIt/vignettes/matching-methods.html.

The specific method used can be briefly gone over from https://cran.r-project.org/web/packages/MatchIt/vignettes/MatchIt.html: "Here, we begin by briefly demonstrating 1:1 nearest neighbor (NN) matching on the propensity score, which is appropriate for estimating the ATT. One by one, each treated unit is paired with an available control unit that has the closest propensity score to it. Any remaining control units are left unmatched and excluded from further analysis. Due to the theoretical balancing properties of the propensity score described by Rosenbaum and Rubin (1983), propensity score matching can be an effective way to achieve covariate balance in the treatment groups."