Racial Fairness in Police Traffic Stops

A Bayesian Regression Analysis of Discrimination Against Asians

Northeastern University

Master of Science in Computer Science – Align

Professor: Dr. Maryam Farahmand

CS 5002 - Discrete Structures

Student: Yue (Yelena) Yu

NUID: 002676233

April 2023

Summary

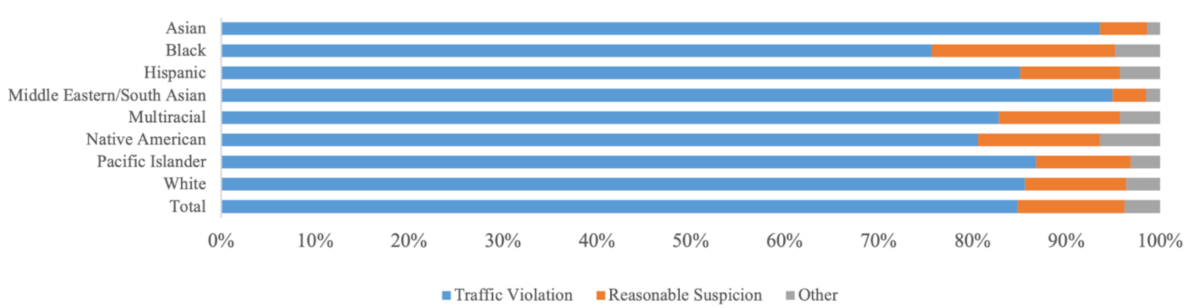
The primary objective of this essay is to enhance public awareness of the importance of promoting and upholding racial fairness for Asian group, with a specific focus on whether discrimination towards Asian individuals occurs during police pull-overs. We used a Bayesian regression model to analyze data collected from *The Stanford Open Policing Project* data to examine whether there is evidence to support such discrimination.

This study identifies several variables that could possibly impact the likelihood of a police officer discriminating against Asian individuals during a traffic stop. These variables include the driver's gender, the district of the pullover occurred, the driver's race. After analyzing the data using the Bayesian regression model, we found evidence to suggest that Asian individuals are subject to discrimination during police pull-overs. Specifically, the results show that Asian drivers are more likely to get a citation or being arrested than other racial groups.

Introduction

In recent years, the issue of racial inequalities in law enforcement procedures has gained significant public attention. "Asian driver" is used frequently as a stereotype and mock people of Asian as bad drivers. Research has indicated that certain racial groups are more likely to be subjected to police stops compared to others.

According to the 2020 Annual report from California Racial and Identity Profiling Advisory (RIPA) Board, traffic violations accounted for over 90% of the reasons for Asian drivers being pulled over, which is a substantially higher than that of any other racial group.



Resource: https://oag.ca.gov/sites/all/files/agweb/pdfs/ripa/ripa-board-report-2020.pdf

The use of traffic stops as a tool for racial profiling has been widely documented and debated. In this essay, we will explore the issue of racial bias in traffic stops from a Bayesian perspective, using data from *The Stanford Open Policing Project.* We will analyze the impact of race on police behavior after a traffic stop and evaluate the limitations of our model.

Data Resources

The Stanford Open Policing Project contains data of traffic across 88 cities located in 42 different states of the United States. We will be focusing on data from San Francisco, including statistics such as race, gender, arrest made, citation issued, warning issued, and outcome. After cleaning the data through eliminating incomplete data, we have a database of around 900,000 rows. While the accuracy, completeness, and relevance of the data have been verified, it is possible that some of the data provided by police officers may be falsified or intentionally misrepresented.

The original data of pullover information includes but not limited to date, district, subject age, subject race, subject gender, vehicular type, arrest made, citation issued, warning issued, outcome, contraband found, search conducted, search vehicle, search basis, reason for stop etc. The data used for our analysis include district, subject race, subject gender, arrest made, citation issued, and warning issued.

Assumptions & Parameters

One assumption made by our study is that the police officer did not know the race of the driver before stopping them and only stopped them when a traffic violation is observed. The study analyzed the penalty results from the police after the stop, which helped to create a Bayesian regression model that was used to evaluate the impact of race on police decision-making after a traffic stop. We combined results including an arrest made or a citation issued as an incident (actions taken by the police officer). A warning is not considered as a citation.

Model Design & Inputs & Output

Our analysis involved implementing Bayesian logistic regression using a Bayesian hierarchical model in R and RJAGS. The response variable represents the number of incidents, defined as either an arrest or a citation, and the inputs are the races involved in the traffic stops, the number of incidents, and the number of stops. The hyperparameter used is betarace, which is sampled from a normal distribution, with the racial group being treated as random effects.

Markov-Chain Monte-Carlo (MCMC) simulation is used to generate samples, with 1,000 iterations being used for adaptation and another 1,000 for burn-in. To ensure adequate coverage of the domain space, four chains with different initial states were employed. We also employed a thinning interval of four to mitigate correlation between consecutive samples resulting from the Markov properties of MCMC.

We define the prior for β to be a normal distribution, where each race has its own corresponding β value used as a logistic regression coefficient for determining if an incident happens to a subject. This value served as the logistic regression coefficient for determining if an incident occurred with a subject. It is important to note that a warning was not considered an incident.

To implement Bayesian analysis, we implemented the following equation:

$$P(Y|\beta) * P(\beta) = P(\beta|Y) * P(Y) \implies P(\beta|Y) = P(Y|\beta) * P(\beta) \ / \ P(Y)$$

P(Y), the normalizer, was difficult to compute but was built into the Markov-Chain Monte-Carlo simulation to ensure that the posterior density integrated to 1.

$P(\beta)$ is a prior distribution, with a probability density function provided for β. Through the training process, $P(\beta|Y)$ (i.e., the posterior) was estimated by fitting the dataset. In other words, an initial guess for β was made, and then our knowledge of β was updated by fitting our observations Y.

$P(\beta|Y)$ is the posterior distribution, which enables us to gain knowledge of β from the evidence in our dataset. We obtained samples of β by simulating our JAGS model, which we then used for comparison.

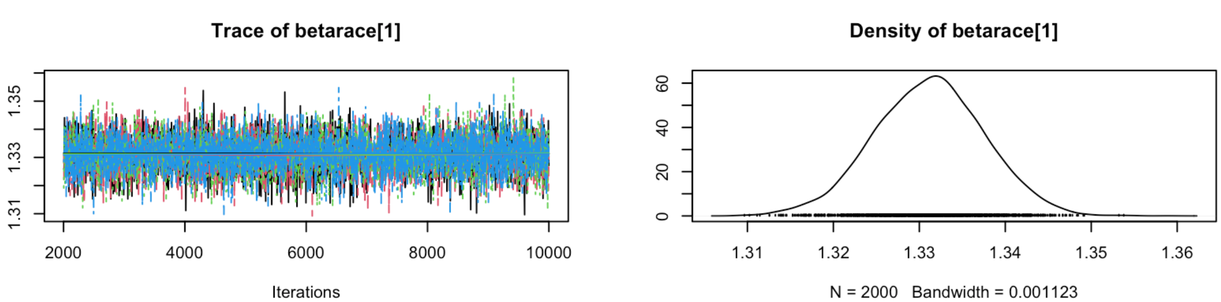
$P(Y|\beta)$ represents our observation, which is the data from our dataset used to fit our model.

Appendix A: R model, used to prepare data and simulation

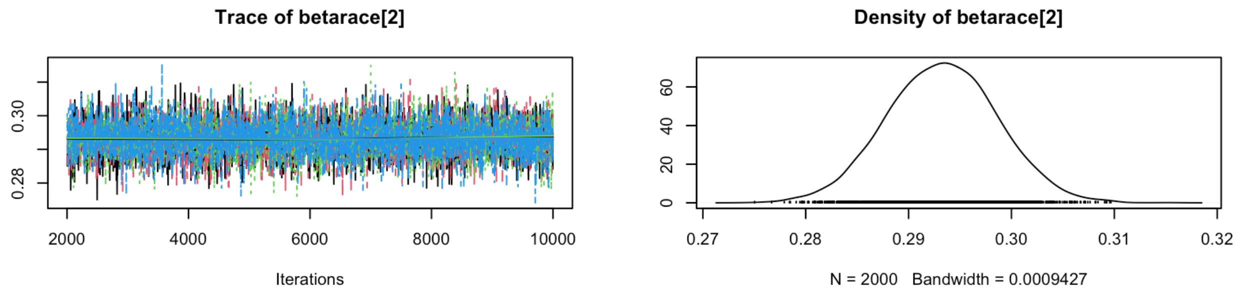
Appendix B: JAGS model, used to create an object representing a Bayesian graphical model, specified with a BUGS-language description of the prior distribution, and a set of data.

Model Result

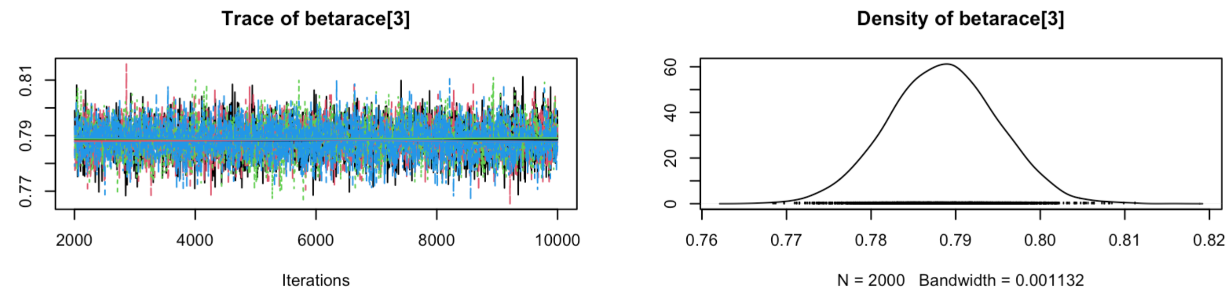
Asian and pacific islander



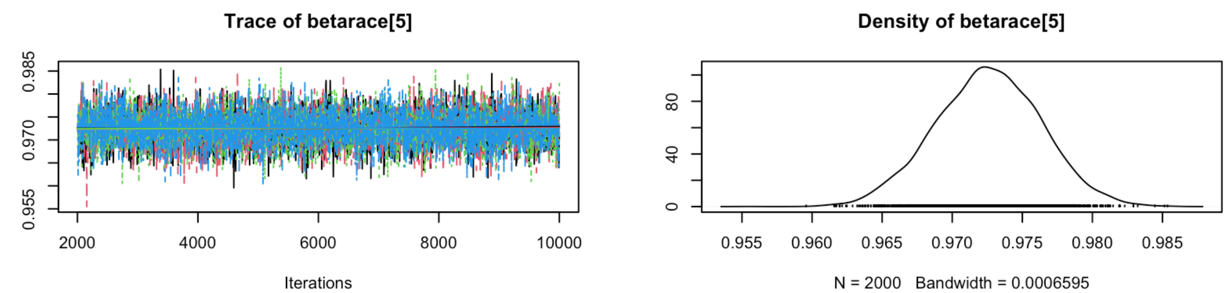
Black



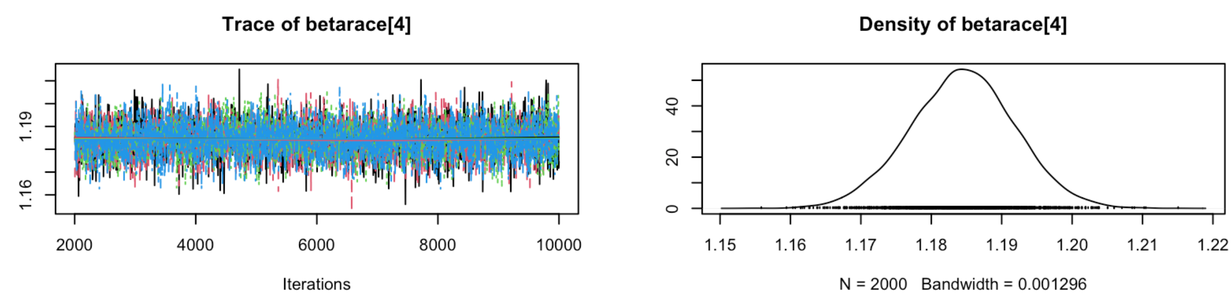
Hispanic



White



Other



After we generated traces and density distributions for each betarace, we can visually observe the differences in betarace distributions for different races. Specifically, the betarace for Asians is primarily situated between 1.31 to 1.35, while for Blacks it ranges between 0.27 to 0.31, for Hispanics it falls between 0.76 to 0.81, for Whites it spans from 0.96 to 0.99, and for Other races, it is between 1.15 to 1.21.

Next, we calculated the means for each race, which corresponds to the central point of the density function. The results revealed that the beta coefficient for Asians, at 1.331128, was significantly higher than that of the other racial groups, such as: β_other of 1.184448, β_white of 0.9725466, β_hispanic of 0.7885007, and β_black of 0.2932641. This suggests that during a police stop, individuals of Asian are more likely to experience an incident, such as receiving a citation or being arrested. This finding provides evidence for the existence of discrimination during police stops.

```
[1] "====================="
[1] white    api      hispanic black    other
Levels: api black hispanic other white
[1] 1.331128
[1] 0.2932641
[1] 0.7885007
[1] 1.184448
[1] 0.9725466
[1] "====================="
[1] "percentage samples where bete_api greater than beta_white"
[1] 1
[1] "====================="
```

Additionally, we determined the percentage of samples where the beta coefficient for Asians was greater than that of Whites, which turned out to be 100%. This indicates that in all cases within our sampled data, the beta coefficient for Whites was consistently lower than that of Asians. This pattern is also visually visible in the density figures presented earlier, where the beta coefficient for Asians is primarily concentrated between 1.31 to 1.35, while for Whites, it falls within the range of 0.96 to 0.99.

Model Evaluation & Model Limitations

We conducted an evaluation process over our model utilizing simple statistics:

Race	**Incidents**	**Stops**	**Incident/Stops**	**Rank by %**	**Beta from Simulation**	**Rank by Beta**
white	256729	353802	73%	3	0.9725837	3
api	116924	147813	79%	1	1.331179	1
hispanic	74980	109054	69%	4	0.7886193	4
black	82439	143923	57%	5	0.2933018	5
other	75266	98291	77%	2	1.184477	2

According to the table presented above, it is also notable that individuals of Asian have a higher probability (79%) experiencing an incident, which is in line with our Bayesian predictive model. Furthermore, the ranking of incidents as a percentage of stops corresponds with the ranking of beta coefficients.

The biggest limitation of our model is the input for Bayesian regression model. While a Bayesian regression model usually takes multiple inputs as parameters, we only take race as our single predictor, which generated a Bayesian simple linear regression model. However, our model's performance and accuracy may be limited by the simplicity of the model and the availability of additional relevant predictors. With only one predictor, our model may not capture the full complexity of the relationship between the predictors and the outcome variable. Incorporating additional independent variables into the model can increase its complexity and potentially improve its accuracy.

Although we initially considered district and gender as inputs in our model design and regression process, we found that these two variables did not have a significant impact on the outputs we were interested in predicting. Furthermore, by analyzing the Gelman-Rubin statistics, we found that we could not achieve convergence with the extra two predictors, indicating that these variables were not informative or relevant for our analysis. Hence, we decided to exclude them from the final version of our model.

As a result of excluding district and gender from our model, we recognized that our model may be susceptible to underfitting, failed to capture the nonlinear relationship between the predictors and the outcome variable.

However, the main goal of our project is to demonstrate this method of Bayesian analysis is feasible. Thus, future work may include increasing the complexity of the model by introducing more predictors.

Conclusion

After conducting a thorough analysis, our findings imply that there is significant racial discrimination prevalent in policing practices regarding traffic stops. These discriminations can potentially lead to instances of racial profiling and bias, which can have far-reaching implications for individuals and communities. By raising more attention and awareness of this kind of discrimination, together with addressing it, we can create a more just and inclusive society that ensures equal protection and treatment for all individuals, irrespective of their race or ethnicity.

Appendix A: R Code – prepare data and simulation

```
library(rjags)


police <- read.csv("ca_san_francisco_2020_04_01-cleaned_dataset.csv",
header=TRUE)
police_sub <- subset(police, district != 'N/A')
police_sub$incidents <- police_sub$arrest_made | police_sub$citation_issued


df <- data.frame(matrix(ncol = 4, nrow = 0))
names(df) <- c("race", "sex", "incidents", "stops")
for (race in unique(police_sub$subject_race)) {
  rows = police_sub[(police_sub$subject_race == race), ]
  stops = nrow(as.matrix(rows))

  rows = police_sub[(police_sub$subject_race == race &
                     police_sub$incidents == TRUE), ]
  incidents = nrow(as.matrix(rows))

  row = data.frame(race = race,
                   incidents = incidents,
                   stops = stops)
  df = rbind(df, row)
}


d1 <- list(stops=df$stops, incidents=df$incidents,
           race=unclass(factor(df$race)))


inits1 <- list(list(betarace=c(10,10,10,10,10)),
               list(betarace=c(10,-10,10,-10,10)),
               list(betarace=c(-10,10,-10,10,-10)),
               list(betarace=c(-10,-10,-10,-10,-10)))


m1<-jags.model("model.bug",d1,inits1,n.chains=4,n.adapt=1000)


update(m1, 1000)


x0 <- coda.samples(m1, c("betarace"), n.iter=8000, thin=4)
print(gelman.diag(x0, autoburnin = FALSE))
```

```
plot(x0)


print("=====================")
print(mean(as.matrix(x0[,"betarace[2]"]) > as.matrix(x0[,"betarace[1]"])))
print("=====================")


x1 <- coda.samples(m1, c("prob", "incidents"), n.iter=8000)


probs <- as.matrix(x1)[, paste("prob[", 1:nrow(df),"]", sep="")]
incidents <- as.matrix(x1)[, paste("incidents[", 1:nrow(df),"]", sep="")]


Tchi <- numeric(nrow(incidents))
Tchirep <- numeric(nrow(incidents))


for (s in 1:nrow(incidents)) {
  Tchi[s] <- sum((df$incidents     -
df$stops*probs[s,])^2/(df$stops*probs[s,]*(1-probs[s,])))
  Tchirep[s] <- sum((incidents[s,] -
df$stops*probs[s,])^2/(df$stops*probs[s,]*(1-probs[s,])))
}


print(mean(Tchirep))
print(mean(Tchi))
```

Appendix B: JAGS model

```
model {


  for (i in 1:length(stops)) {
    incidents[i] ~ dbin(prob[i], stops[i])
    logit(prob[i]) <- betarace[race[i]]

  }


  for (j in 1:max(race)) {
    betarace[j] ~ dnorm(0, 0.1)
  }

}
```

References

1. *Executive summary - 2020 RIPA Board report - racial and identity ...* (n.d.). Retrieved April 28, 2023, from https://oag.ca.gov/sites/all/files/agweb/pdfs/ripa/ripa-exec-summary-2020.pdf
2. *The Stanford Open Policing Project*. openpolicing.stanford.edu. (n.d.). Retrieved April 27, 2023, from https://openpolicing.stanford.edu/data/