Does Improving Diversity Among Police Officers Reduce Racial Profiling

Daniel Meza, California State University Monterey Bay,

Department of Mathematics and Statistic

Adolfo Huerta, California State University Monterey,

Department of Mathematics and Statistics

November 28, 2017

## Abstract

Over the past few years, there have been traffic stop events where the actions taken by police officers were questionable. Our objective is to analyze traffic stop data from a large south eastern city in the United Sates to identify possible racial profiling. Our main concern is that if racial profiling is present, then police officers could use this method as a way to further investigate targeted minorities with little to no probable cause. Using three logistic regression models we attempt to answer three research questions that address concern of present racial profiling. In our first model we will we will investigate if there is any relationship between the chance of getting stopped when the police officer was white and the driver was Hispanic. Next, we test to determine if the result of the stop when the driver is Hispanic is more or less severe when the police officer is of a minority race. We would also like to see if the results of being stopped, leading to the driver being searched, is higher for minorities. Based on our analysis, our results indicate that out of the five different race categories, white police officers were not the race that contributed the most to traffic stops when the driver was Hispanic. The police officer’s race that is the most influential to traffic stops when the driver is Hispanic is American Indian,Native Alaskan,and Hawaiian police officers. Also we were able to identify significant information to provide as evidence in favor that the traffic stop being more severe for Hispanic drivers when the police officer is not a minority. Finally, our last results indicate that minorities are indeed more likely to be searched than white drivers by police officers. We were able to produce similar results to the studies we investigated, except for the study stating that white police officers contributed the most to Hispanic traffic stops.

## Introduction

On August of 2014, an African American male, Michael Brown, at only eighteen years of age was killed by a police officer after Brown was accused of strong-arm robbing a convenience store. Michael Brown was confirmed as being unarmed, but was still fatally shot after the officer claimed that Brown had allegedly attacked him, in order to try and take control of his firearm, until it was fired. However, the officer shot a total of twelve rounds in the whole altercation and at the time he was the only witness. The officer that shot Brown was not charged with murder and was released after the U.S. Department of Justice concluded that the officer shot Brown in self defense (New York Times, 2017). This one of the many cases where police officers are accused of racially profiling African Americans, Hispanic and other minorities.

The incident that motivated us to take on this topic were the killing of four Hispanic males in Salinas, California by police in the span of five months, in the year of 2014. One of our main interests is to identify any type of unfairness between the police officers and minorities. According to Nancy Krieger, a professor of social epidemiology at Harvard T.H. Chan School of Public Health, “Black men, compared to white men, were from five to 19 times at greater risk of a law enforcement-related death over the past 50 years” (Krieger 2015). There is a significant disparity in the risk of death when comparing a minority and a white individual, giving us some evidence to why our research questions are important. The data we will analyze does not contain information about citizens being killed by police. We will only try to provide evidence to issues regarding police racial profiling from traffic stop data.

In recent years, research explains that about six percent more of the stops were between a white officer and black citizen over a white officer and white citizen being stopped(Criminology, 2012). Therefore, the first question is if the police officers race is important in determining if the driver who was stopped was a minority. This is a very important question as it should show which are the officers races that are more likely to stop a minority driver. It should also tell us if there is a significant difference in white officers stopping minority drivers over white officers stopping white drivers for our population, which is a large southeastern city.

In a 2011 report, the Leadership Conference on Civil Rights found evidence of widespread racial profiling, showing that African Americans and Hispanics are disproportionately likely to be stopped and searched by police, even though they’re less likely to be found possessing contraband or committing a criminal act. If this report is accurate, then it should be the case that minorities have a higher rate of their traffic stop being less severe. This motivates our second research question, we want to determine if the result of a traffic stop is less severe when the driver and police officer is a minority. This research question follows from the first, since if certain police officers are more likely to stop minority drivers then it is likely that the officer will provide more severe consequences to that minority driver. Although the studies and report we have mentioned have given important results, our analysis will be important in that our dataset is more recent, being from the year 2016 as compared to the others being from the years 2011, 2014 and 2015.

Finally, we would like to see the results of being stopped leading to the driver being searched, is higher for minorities. We chose this because the case studies and research we read have found that minorities, like Hispanics and Blacks, to be at higher risk of police racial profiling. This is seen in the study, “Black and Hispanic Men Perceived to Be Large Are at Increased Risk for Police Frisk, Search, and Force”, in which the researchers look at a similar scenario. The results from this study showed that indeed Black and Hispanic suspects were at a greater risk of being frisked or searched compared to whites. Furthermore, results from the Leadership Conference on Civil Rights report gives as an example that in Illinois, Hispanic and black drivers were twice as likely to be searched after being stopped by police compared to white drivers. However, the surprising thing was that white drivers were twice as likely to have contraband in their possession (Washington Post, 2014).

## Methods

The traffic stop data was downloaded from Data.Gov. The Records from all 79885 reported traffic stops was for a large Southeastern city in 2016. There is no information on how the data was collected. For all of our questions we will use logistic regression.

For our first question, we considered logistic regression to model and classify a binary response variable describing if the driver was a minority. There were five categories in the variable for the drivers race (DrvRace) but we combined all categories that are minorities such as Asian, Black, Native American and Other/Unknown into the true category for the variable drv\_minority. In this model our explanatory variable was only the categorical variable with the police officers race.

In our second research question, the response variable was the result of the stop categorized as severe or non-severe. We combined the categories No action, Verbal Warning, and Written Warning from the result stop (RsltStop) into the non severe category. In addition, the severe category was made up of the remaining categories Citation and Arrest from the result of stop variable in our original data. We also converted the police officers race into two categories, white or minority in the same way.

In our final research question, we consider the binary response variable of being searched during a traffic stop different if the driver was a minority (non-white). We also combined the five categories for the drivers race into either a white driver or a minority driver to test versus whites. We also improve our model by creating an adjusted model with the variable of whether the officer is male and the variable whether the driver is male is true. In this adjusted model, we must check for collinearity between confounding variables and all the other assumption before continuing. We check if this is a better model with the G-statistic and nested test.

#### The variables are defined in the following way:

Variables/Names Month 1-12 (Month) Reason for Stop (RsnStop) Officer Race (OffRace) Officer Male (OffMale) Officer Years Service (OffYrsSrv) Driver Race (DrvRace) Driver Hispanic (DrvHisp) Driver Male (DrvMale) Driver Age (DrvAge) Search Vehicle (SrchVhcl) Result of Stop (RsltStop)

## Results

In the traffic stop data set from a large southeastern city of the United States, we found that the mean age of the drivers was about 36 years old. Also, we calculated that the mean number of service years for the police officers in this data set was around eleven years. We found that the proportion of drivers who are males is about 58 percent while the percent of female drivers was about 42 percent. We find that a proportion of about 0.4016 of drivers get searched. Also the proportion of minority drivers overall is 0.413 and the proportion of white drivers is 0.587 in the dataset. However, the proportion of drivers that are minorities and were searched is about 0.0539 and the proportion of drivers that are white and were searched is about 0.0206.

Based on the results for our first question, we found that all of the categories for the police officers race are significant predictors of when the driver who was stopped was a minority. In the appendix, the output for the summary of the first model shows each of the specific p-value’s less than .05 using Wald Z test, as we can see in Figure 1. Also, the general test was using a likelihood ratio test(LRT) was found significant too since we got a p-value of zero being less than .05 and a G-statistic of 148 with degrees of freedom of 4. Since our explanatory variable was found significant for all categories, we analyzed the slopes in the context of the odds model to determine an answer for our question. Basically, we took all of the estimates for the first model and we exponentiated them to measure the slope in the context of an odds model. The estimate with the greatest odds ratio was category 1(American Indian/Native Alaskan/Hawaii) with an odds ratio of 3.155036.

Figure 1. Summary Of First Model

|  |  |
| --- | --- |
| Police Officer Race Categories | P-Value |
| AmericanIndian/NativeAlaska/Hawaii (1) | 2.43e-14 |
| Asian/PacificIslander (2) | 8.76e-12 |
| Black/AfricanAmerican (3) | 6.39e-06 |
| Hispanic/Latino (4) | 2.63e-06 |
| White (5) | 8.04e-08 |

The results for our second questions where that the adjusted model was significant for the police officers race and for the variable when the driver stopped was a minority. Using the Wald Z test to test for specific hypothesis, we found all betas to have a p-value less than .05, meaning that they are significant, as seen in Figure 2. For the general hypothesis, a Likelihood Ratio test was used where the G-Statistic was 528, degrees of freedom was 2, and p-value was 0. Next, we will interpret the results of the model in the context of the odds and probability model. When we condition for the driver being Hispanic, we find that the odds ratio for when the police officer is White is 1.39 and when the police officer’s race is a minority then the odds ratio is 0.6942591.

Figure 2. Summary Of Second Model

|  |  |
| --- | --- |
| Variable Name | P-Value |
| Police officer(Minority) | 2e-16 |
| Driver (Minority) | .000154 |

The results for our third research question were that the unadjusted model was significant for the driver being a minority and that this variable holds predictive power over the response of being searched. For the general hypothesis, we used a Likelihood Ratio Test, where our G-Statistic was 606.6, with degrees of freedom of 1, which gave us a p-value of 0. We then interpret the coefficients of the model in the context of the odds model and find that the odds of a driver being white and being searched is 0.0209. The odds of a minority being searched is 2.714 times more than the odds of white driver being searched, not adjusting for any confounding variables. In context of the probability model, we interpret, when the driver is white, the probability that the driver gets searched is 0.3678. Our adjusted model was also found to pass the general hypothesis. We calculate a G-statistic of 1680.7 with degrees of freedom of 3 and receive a p-value of 0, which leads us to reject our general hypothesis. Using the Wald Z test to test for specific hypothesis, we found all betas to have a p-value less than .05, meaning that they are significant as you can see from Figure 3. We interpret the coefficients in the context of the odds and find that when driver is white, officer is female and driver is female, the odds of being searched are 0.00496. Also, when the driver is a minority, the odds of being searched is 2.828 times more than when the driver is white, holding the officers and drivers gender fixed. When the officer is a male, the odds of being searched is 1.608 times more than when the officer is a female, holding the driver being minority and the drivers gender fixed. Finally, when the driver is a male, the odds of being searched is 3.91 times more than when the driver is female, when we hold the drivers race and officers gender fixed. Finally, our AUC for these two models was 0.6049 and 0.698, for our unadjusted and adjusted models, respectively. This ensures us that, the adjusted model preforms better in correctly classifying if a driver was searched or not searched and so we use the result from the adjusted model.

Figure 3. Summary of Third Model

|  |  |
| --- | --- |
| Variable Name | P-value |
| Driver Minority (True) | <2e-16 |
| Officer Male (True) | 8.37e-16 |
| Driver Male (True) | <2e-16 |

## Discussion

In question one, our goal was to determine if white police officers have the greatest effect in the number of traffic stops when the driver is Hispanic. Before the results, we assumed that White police officer might have the greatest effect in traffic stops when the driver was Hispanic. Our results indicate that out of the five different race categories, white police officers were the fourth race that contributed the most to traffic stops when the driver was Hispanic. The police officer’s race that is the most influential to traffic stops when the driver was Hispanic were American Indian,Native Alaskan,and Hawaiian police officers. Therefore, we can interpret the result as when the police officer is American Indian,Native Alaskan or Hawaiian, the odds that the person stopped is Hispanic is multiplied by 3.155036.In the context pf probability, we can say that the probability that a Hispanic driver is stopped by a police officer who is American Indian,Native Alaskan or Hawaiian is 0.7593282. Therefore, we did not find sufficient evidence to claim that White police officers had the greatest impact in the number of traffic stops when the driver was a Hispanic.

In question two, we focused on trying identifying if the result of a traffic stop is less severe when the driver and police officer is a minority. Based on the results, we were able to identify significant information to provide as evidence in favor of our research question. In the context of the odds model we can say that, when we condition on the driver being Hispanic then the odds that the result of the traffic stop is considered severe(Citation or Arrest) is multiplied by 1.396 when the police officer’s race is White. On the other hand, when we condition on the driver being Hispanic, the odds that the result of the stop is considered severe is multiplied by .694, when the police officer is a minority. The probability of the result stop being considered severe when the police officer is a minority is 0.4097715, when we condition on the driver being Hispanic. The probability of the result stop being considered severe when the police officer is a White is 0.582618, when we condition on the driver being Hispanic. Therefore, using the probability as a measure we can state that when a police officer is white the probability of a traffic stop resulting as an arrest or citation when the driver is Hispanic is higher than when the police officer’s race is a minority.

For our third research question we wanted to determine if there is a significant difference in drivers being searched depending on them being a minority or white. From the proportions that we calculated in our exploratory data analysis, we could see that the proportion of minorities that were searched was 0.54 which is larger than the proportion of drivers that are white that are searched of 0.0206. Our results indicate that minorities are more likely to be searched than white drivers, since when the driver is a minority, the odds of being searched is 2.828 times more than when the driver is white, holding the officers and drivers gender fixed. This in fact does agree with the study which showed that indeed Black and Hispanic suspects were at a greater risk of being frisked or searched compared to whites. Another supporting evidence to this study from our results, is that when the driver is a male, the odds of being searched is 3.91 times more than when the driver is female, when we hold the drivers race and officers gender fixed. Therefore, it appears that the combination of being a minority and being male, there are significantly greater odds of being searched by the police officer in a traffic stop. There is still possible ways to further investigate the results of the case studies we read, by exploring the variable of the result of the stop. With further analysis, we can possibly answer the question that, even if minorities more likely to be searched than whites, are they less likely to posses contraband or commit a crime than white drivers.

## Appendix

\*\*Levels for categorical variables

Reason for Stop: 1=CheckPoint, 2=DWI, 3=Investigation, 4=Other, 5=Safe Movement, 6=SeatBelt, 7=Speeding, 8=StopLight/Sign, 9=VehicleMovement, 10=VehicleRegistry

Officer race: Blank=NotSpecified, 1=AmericanIndian/NativeAlaska/Hawaii, 2=Asian/PacificIslander, 3=Black/AfricanAmerican, 4=Hispanic/Latino, 5=White

Driver Race: 1=Asian, 2=Black, 3=NativeAmerican,4=Other/Unknown,5=White

Result of Stop: 1=NoActionTaken, 2=VerbalWarning,3=WrittenWarning, 4=Citation, 5=Arrest

library(readr)  
# trafficstop <- read\_csv("~/R/Data Sets/trafficstop.csv")  
trafficstop <- read\_csv("~/Desktop/ucsc/trafficstop.csv")

## Parsed with column specification:  
## cols(  
## Month = col\_integer(),  
## RsnStop = col\_integer(),  
## OffRace = col\_integer(),  
## OffMale = col\_integer(),  
## OffYrsSrv = col\_integer(),  
## DrvRace = col\_integer(),  
## DrvHisp = col\_integer(),  
## DrvMale = col\_integer(),  
## DrvAge = col\_integer(),  
## SrchVhcl = col\_integer(),  
## RsltStop = col\_integer()  
## )

# View(trafficstop)  
attach(trafficstop)  
mean(DrvAge)

## [1] 35.88215

mean(OffYrsSrv)

## [1] 11.1789

men<-sum(DrvMale==1)  
men\_prop<-men/length(DrvAge)  
men\_prop

## [1] 0.5795153

women<-sum(DrvMale==0)  
women\_prop<-women/length(DrvAge)  
  
table(DrvMale)

## DrvMale  
## 0 1   
## 33590 46294

## Question 1  
## Do White police officers have the greatest effect in the number of traffic stops when the driver is Hispanic?  
  
library(car)  
drv\_minority=trafficstop$DrvRace!=5 ### when true the driver is a minority  
offrace12=factor(trafficstop$OffRace)  
h1=glm(drv\_minority~offrace12,family = binomial())  
summary(h1) ###Summary for first model

##   
## Call:  
## glm(formula = drv\_minority ~ offrace12, family = binomial())  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6878 -1.3244 0.9876 1.0372 1.1353   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 1.1490 0.1507 7.625 2.43e-14 \*\*\*  
## offrace122 -1.0492 0.1537 -6.825 8.76e-12 \*\*\*  
## offrace123 -0.6846 0.1517 -4.513 6.39e-06 \*\*\*  
## offrace124 -0.7295 0.1553 -4.698 2.63e-06 \*\*\*  
## offrace125 -0.8099 0.1509 -5.366 8.04e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 107319 on 79162 degrees of freedom  
## Residual deviance: 107171 on 79158 degrees of freedom  
## (721 observations deleted due to missingness)  
## AIC: 107181  
##   
## Number of Fisher Scoring iterations: 4

g100<-(107319 -107171)  
g100

## [1] 148

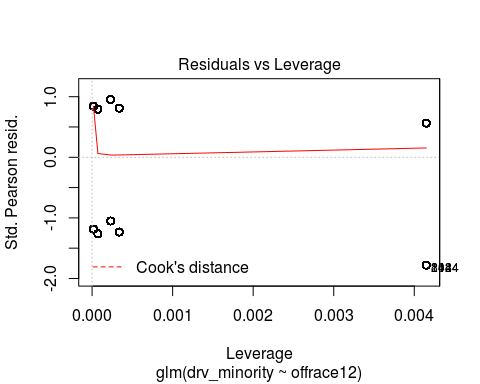
df100<-(79162-79158)  
df100

## [1] 4

1-pchisq(g100,df100)

## [1] 0

# large sample size of more than 79000 Assumption 1  
# Assumption 2 VIF not needed since we only have one explanantory variable  
plot(h1,5)## Assumption 3



# assumption 4 satisfied since we have a binary response variable in the model  
  
  
## comparing the odds between the different race of the police officers  
exp(1.1490)#American indian/Hawain/native Alaskan

## [1] 3.155036

exp(-1.0492)#Asian/Pacific Islander

## [1] 0.3502178

exp(-0.6846)#Black/African American

## [1] 0.5042919

exp(-0.7295 )#Hispanic

## [1] 0.48215

exp(-0.8099)## white

## [1] 0.4449026

### Converting the greatest odds ration into probability  
exp(1.1490)/(1+exp(1.1490))

## [1] 0.7593282

################################### #QUestion 2 IS the result of a traffic stop is less severe when the driver and police officer is a minority.   
  
offrace=trafficstop$OffRace  
minority\_off=offrace!=5 #### if the category is not five then the officer is a minority  
  
  
drv\_minority=trafficstop$DrvRace!=5 ### when true the driver is a minority  
  
  
Result\_Not\_Severe=trafficstop$RsltStop<=3 ###when true the result is not severe   
  
  
length(Result\_Not\_Severe)

## [1] 79884

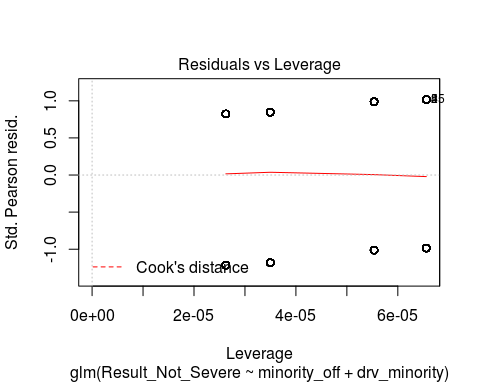
h2=glm(Result\_Not\_Severe~minority\_off+drv\_minority,family = binomial())  
summary(h2)

##   
## Call:  
## glm(formula = Result\_Not\_Severe ~ minority\_off + drv\_minority,   
## family = binomial())  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.346 -1.322 1.017 1.039 1.191   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.33353 0.01199 27.811 < 2e-16 \*\*\*  
## minority\_offTRUE -0.36491 0.01607 -22.706 < 2e-16 \*\*\*  
## drv\_minorityTRUE 0.05530 0.01461 3.785 0.000154 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 108358 on 79162 degrees of freedom  
## Residual deviance: 107830 on 79160 degrees of freedom  
## (721 observations deleted due to missingness)  
## AIC: 107836  
##   
## Number of Fisher Scoring iterations: 4

# large sample size of more than 79000 Assumption 1  
vif(h2)# Assumption 2 satisfied since the values are less tha 5

## minority\_off drv\_minority   
## 1.000176 1.000176

plot(h2,5)## Assumption 3



# assumption 4 satisfied since we have a binary response variable in the model  
  
g101<-(108358 -107830)  
g101

## [1] 528

df101<-(79162 -79160 )  
df101

## [1] 2

1-pchisq(g101,df101)

## [1] 0

#### Results for odds model  
  
exp(-0.36491)### odds ratio for when the police officer is a minority

## [1] 0.6942591

exp(-0.36491)/(1+exp(-0.36491))## prob when Police Officer is minority

## [1] 0.4097715

exp(0.33353 ) # 0dds ratio for when the police officer is not a minority

## [1] 1.395887

exp(0.33353)/(1+exp(0.33353 )) ## prob when police officer is not a minoirty

## [1] 0.582618

################################### #QUestion 3 The result of being stopped leading to the driver being searched, is higher for minorities?  
  
#Set variables that will be used in model and exploratory data analysis.   
drv\_minority=trafficstop$DrvRace!=5 ### when true the driver is a minority  
drv\_minority <- factor(drv\_minority)  
  
offrace=factor(trafficstop$OffRace)  
  
summary(SrchVhcl)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.00000 0.00000 0.00000 0.04016 0.00000 1.00000

#about .4016 of drivers stopped got searched  
  
summary(DrvRace)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 2.000 2.000 3.281 5.000 5.000

#Majority of drivers stopped are white and the least are Asians  
  
summary(drv\_minority)

## FALSE TRUE   
## 32969 46915

#There are 32,969 minority drivers and 46,915 white drivers  
#Which means the proportion of minority drivers is .413 and proportion of white drivers are .587  
  
summary(OffMale)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 1.000 1.000 0.916 1.000 1.000

#The proportion of male officers in our dataset is about 0.916  
  
summary(DrvMale)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 1.0000 0.5795 1.0000 1.0000

#The proportion of male drivers in our dataset is about 0.5795  
  
table(drv\_minority, SrchVhcl)

## SrchVhcl  
## drv\_minority 0 1  
## FALSE 32291 678  
## TRUE 44385 2530

#The proportion of drivers that are minorities and were searched is about 0.0539  
#2530 / ( 2530 + 44385)  
#The proportion of drivers that are white and were searched is about 0.0206  
#678 / ( 678 + 32291)  
  
  
#reduced model of search versus minority  
minor2Srch <- glm(SrchVhcl ~ drv\_minority, family = "binomial")  
summary(minor2Srch)

##   
## Call:  
## glm(formula = SrchVhcl ~ drv\_minority, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.3330 -0.3330 -0.3330 -0.2039 2.7872   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.86340 0.03880 -99.56 <2e-16 \*\*\*  
## drv\_minorityTRUE 0.99871 0.04386 22.77 <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: 26912 on 79883 degrees of freedom  
## Residual deviance: 26306 on 79882 degrees of freedom  
## AIC: 26310  
##   
## Number of Fisher Scoring iterations: 6

#Test General Hypothesis  
minorSummary <- summary(minor2Srch)  
G\_minor <- minorSummary$null.deviance - minorSummary$deviance  
G\_minor

## [1] 606.6462

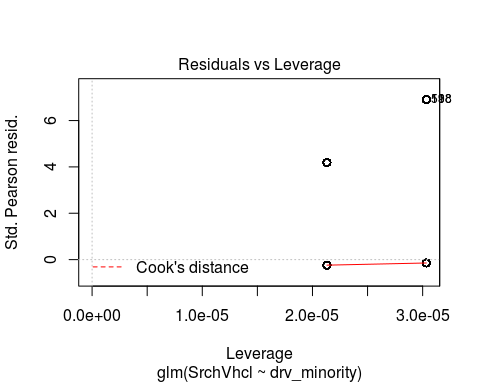
df\_minor <- minorSummary$df.null - minorSummary$df.residual  
df\_minor

## [1] 1

1 - pchisq(G\_minor, df\_minor)

## [1] 0

#Assumption one satisfied due to our large sample size of over 79000 observations  
#Assumption for VIF not needed since only one explanatory variable in the model  
  
#Check assumption three of Outliers  
plot(minor2Srch, 5)



#Assumption three satisfied since there is no observations with Cook's Distance larger than 0.5  
#Assumption four satisfied since our response is binary  
  
#When the driver is white, drv\_minority = 0, the odds that the driver  
#is searched is about 0.021   
exp(minorSummary$coefficients[1,1])

## [1] 0.02099656

#odds of white driver being searched  
  
exp(minorSummary$coefficients[2,1])

## [1] 2.714789

#odds of minority being searched is 2.714 times more than the odds of white driver being searched.   
  
#Interpret slope in context of probablity model  
exp(minorSummary$coefficients[1,1]) / exp(1 + minorSummary$coefficients[1,1])

## [1] 0.3678794

#When the driver is white, the probability that the driver gets searched is 0.3678  
  
#full model with varibles that were significant when modeled alone with SrchVhcl  
full2Srch <- glm(SrchVhcl ~ drv\_minority + OffMale + DrvMale, family = "binomial")  
summary(full2Srch)

##   
## Call:  
## glm(formula = SrchVhcl ~ drv\_minority + OffMale + DrvMale, family = "binomial")  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -0.4112 -0.4112 -0.2479 -0.2112 3.2592   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.30621 0.09284 -57.154 < 2e-16 \*\*\*  
## drv\_minorityTRUE 1.03947 0.04408 23.584 < 2e-16 \*\*\*  
## OffMale 0.47515 0.07741 6.138 8.37e-10 \*\*\*  
## DrvMale 1.36352 0.04818 28.298 < 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: 26912 on 79883 degrees of freedom  
## Residual deviance: 25232 on 79880 degrees of freedom  
## AIC: 25240  
##   
## Number of Fisher Scoring iterations: 7

#Test General Hypothesis  
full\_summ <- summary(full2Srch)  
full\_summ$coefficients

## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -5.3062067 0.09284065 -57.153919 0.000000e+00  
## drv\_minorityTRUE 1.0394691 0.04407578 23.583680 5.668060e-123  
## OffMale 0.4751475 0.07741457 6.137701 8.372427e-10  
## DrvMale 1.3635225 0.04818452 28.297935 3.664117e-176

G\_full <- full\_summ$null.deviance - full\_summ$deviance  
df\_full <- full\_summ$df.null - full\_summ$df.residual  
  
1 - pchisq(G\_full, df\_full)

## [1] 0

#Reject the Null Hypothesis to support that there is at least one significant predictor  
#in our model for the response variable  
  
#Test for colinearity between variables  
library(car)  
vif(full2Srch)

## drv\_minority OffMale DrvMale   
## 1.001537 1.001017 1.000539

#Assumption of collinearity satisfied since there are no VIF scores greater than 5.0  
  
#Interpret results of coefficients in terms of odds model.   
  
#Intercept coefficient  
exp(-5.30621) #When variables driver minority, officer male and driver male are zero, in other words

## [1] 0.004960692

#when driver is white, officer is female and driver is female, the odds of being searched are 0.00496  
  
#Interpret driver minority coefficient  
exp(1.03947)

## [1] 2.827718

#When variables officer male and driver male are held constant, as driver minority increases by one, meaning  
#when the driver is a minority, the odds of being searched is 2.828 times more than when the driver is white.  
  
#Interpret officer being male coefficient  
exp(0.47515)

## [1] 1.608255

#When variables driver minority and driver male are held constant, as officer male increases by one, meaning  
#when the officer is a male, the odds of being searched is 1.608 times more than when the officer is a female.   
  
#Interpret driver being male coefficient  
exp(1.36352)

## [1] 3.909932

#When variables driver minority and officer male are held constant, as driver male increases by one, meaning  
#when the driver is a male, the odds of being searched is 3.91 times more than when the driver is female.   
  
  
#Generate ROC to summarize classifier performance of adjusted vs. unadjusted model  
#Install pROC package  
# installed.packages(pROC)  
library(pROC)

## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

#Unadjusted reduced model  
coef(minor2Srch)

## (Intercept) drv\_minorityTRUE   
## -3.8633965 0.9987143

predprmin <- predict(minor2Srch, type = c("response"))  
rocurvemin <- roc(SrchVhcl ~ predprmin)  
rocurvemin #Reduced model gives us an AUC of 0.6049

##   
## Call:  
## roc.formula(formula = SrchVhcl ~ predprmin)  
##   
## Data: predprmin in 76676 controls (SrchVhcl 0) < 3208 cases (SrchVhcl 1).  
## Area under the curve: 0.6049

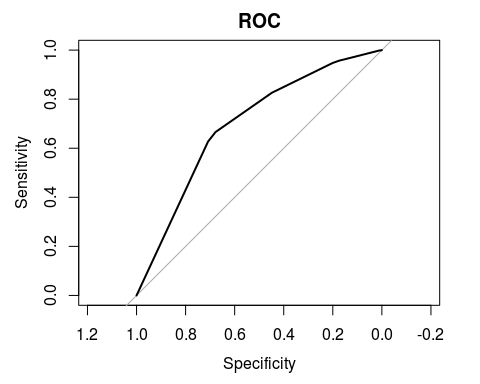
#Our model is doing a poor job judging from AUC  
  
#Adjusted full model  
coef(full2Srch)

## (Intercept) drv\_minorityTRUE OffMale DrvMale   
## -5.3062067 1.0394691 0.4751475 1.3635225

predprfull <- predict(full2Srch, type = c("response"))  
roccurvefull <- roc(SrchVhcl ~ predprfull)  
roccurvefull #Full model gives us an AUC of 0.698

##   
## Call:  
## roc.formula(formula = SrchVhcl ~ predprfull)  
##   
## Data: predprfull in 76676 controls (SrchVhcl 0) < 3208 cases (SrchVhcl 1).  
## Area under the curve: 0.698

#Since AUC of full model is larger, then we use full model.   
#Our model is doing an almost acceptable job judging from AUC  
  
plot(roccurvefull, main = "ROC")



## References

Krieger, Nancy, et al. “Police Killings and Police Deaths Are Public Health Data and Can Be Counted.” PLOS Medicine, Public Library of Science, 2015, journals.plos.org/plosmedicine/article?id=10.1371%2Fjournal.pmed.1001915.

Natarajan, Ranjana. “Racial Profiling Has Destroyed Public Trust in Police. Cops Are Exploiting Our Weak Laws against It.” The Washington Post, WP Company, 15 Dec. 2014, www.washingtonpost.com/posteverything/wp/2014/12/15/racial-profiling-has-destroyed-public-trust-in-police-cops-are-exploiting-our-weak-laws-against-it/?utm\_term=.beaecbc41c9c. ROJEK, J., ROSENFELD, R. and DECKER, S. (2012), POLICING RACE: THE RACIAL STRATIFICATION OF SEARCHES IN POLICE TRAFFIC STOPS\*. Criminology, 50: 993-1024. <doi:10.1111/j.1745-9125.2012.00285.x>

Smith, Mitch. “New Ferguson Video Adds Wrinkle to Michael Brown Case.” The New York Times, The New York Times, 11 Mar. 2017, www.nytimes.com/2017/03/11/us/michael-brown-ferguson-police-shooting-video.html. University, Stanford. “Stanford Researchers Develop New Statistical Test That Shows Racial Profiling in Police Traffic Stops.” Stanford News, 8 Sept. 2017, news.stanford.edu/2016/06/28/stanford-researchers-develop-new-statistical-test-shows-racial-profiling-police-traffic-stops/.