Cops and Stops: Racial Profiling and a Preliminary Statistical Analysis of Los Angeles Police Department Traffic Stops and Searches

California State Polytechnic University, Pomona

and

Loyola Marymount University

Department of Mathematics Technical Report

Megan Armentrout[*], Amber Goodrich[†], Jennifer Nguyen[‡], Lizette Ortega[§], Laura Smith[¶], Lily S. Khadjavi[‖]

Applied Mathematical Sciences Summer Institute
Department of Mathematics & Statistics
California State Polytechnic University Pomona
3801 W. Temple Ave.
Pomona, CA 91768

June 2007

[*]Whitworth University
[†]Central Washington University
[‡]California State Polytechnic University
[§]University of Arizona
[¶]University of California, Los Angeles
[‖]Loyola Marymount University, Los Angeles

Contents

Abstract

With data collection efforts underway in over 45 states, racial profiling in police practice is an issue of national concern. This study focuses on Los Angeles because of its diverse racial composition and the large quantity of data collected by the Los Angeles Police Department. Under a consent decree with the United States Department of Justice, the Los Angeles Police Department is required to make this data available. Using records for over 600,000 traffic stops, we analyze racial disparities found in stop and search rates. Logistic regression models are used to determine which variables are significantly related to disparities in search rates and other police practices. Based on our findings, the possibility of racial profiling cannot be ruled out.

1 Introduction

Racial profiling is a serious issue throughout the United States. A guide from the United States Department of Justice written by Deborah Ramirez defines racial profiling as "any police-initiated action that relies on the race, ethnicity or national origin rather than the behavior of an individual or information that leads the police to a particular individual who has been identified as being, or having been, engaged in criminal activity." (See [Ra].) According to a Gallup Poll conducted in Spring of 2001, 58% of Americans believe racial profiling is still occurs despite the fact that it is illegal. Statistical studies have been conducted in cities around the country that examine the relationship between race and police practice. (See [An], [Ge], [Kh] for recent examples.)

In the past, Los Angeles has been the center of several allegations related to police corruption. When accusations in the late 1990s proved to be true, a consent decree was established to monitor many aspects of the Los Angeles Police Department (LAPD). The Consent Decree is an agreement between the U.S. Department of Justice and the city of Los Angeles. The implementation of the Consent Decree has made it possible for us to obtain data on traffic stops. We now have the chance to statistically analyze whether the treatment of individuals differ according to race. The focus of our research is to determine whether or not racial profiling occurs in Los Angeles Police Department traffic stops and if it is related to the discovery rate, defined as the number of discoveries over the number of searches.

The Los Angeles Police Department is of interest to us for several reasons. First, Los Angeles is a very diverse city, and studying the data from the police department can provide us with understanding of police practice in such an environment. Second, previous work such as "Driving While Black" in *Chance* ([Kh]) was limited to analysis of aggregated LAPD traffic stop data. This study is one of the first being performed in Los Angeles using disaggregated data, as we are one of the few groups that have been able to obtain access to it. Third, the data we analyze are from traffic stops, and according to the Bureau of Justice Statistics, traffic stops are the most common interaction between police officers and civilians throughout the nation (see [Du]). Hence, analyzing traffic stops in particular would be an effective way of observing the treatment of civilians by police officers in general. Finally, in the wake of the allegations of misconduct which prompted the Consent Decree, the LAPD has a vested interest in understanding patterns in stop data to rebuild trust, especially in communities of color. Analyzing this data is a first step in that direction. To analyze the data we use chi-square tests and logistic regression models. We attempt to study the disparities in treatment of different racial groups across various areas of Los Angeles. To do this, we focus on search rates and discovery rates. Our results indicate that African American and Hispanic drivers are treated differently when compared to White and Asian drivers after being stopped.

A brief outline of our paper is as follows: First, we go into some detail about the background of the study. This section will discuss recent studies of racial profiling across the nation and mention some issues dealing with discrimination by the LAPD in the past. Also, the legal basis of a search is discussed. When is a search allowed, and what is a reasonable justification for a search? After the background, the process of obtaining the data is explained in detail. The geography of Los Angeles is discussed, as well as how the city is broken down by policing districts. Next, the methodology section introduces the statistical methods used: chi-square tests and logistic regression models. Our analysis covers our findings both citywide and in the different policing districts. We end with a conclusion of our research and some future work that we hope to pursue.

2 Background

2.1 Racial Profiling Studies in Other Areas

Racial profiling is an issue across the nation and has also played roles in court cases. We first discuss two influential court cases concerning racial profiling in which statistical analysis played a major role and then note two recent studies of interest.

2.1.1 Racial Profiling and Statistics in Court

In some states, studies have been conducted as a result of court cases where the plaintiff claimed to be unfairly treated during a traffic stop (see [La]). These cases cited racial profiling as the primary reason for unnecessary searches. Two seminal court cases using statistics to examine charges of racial profiling were *New Jersey v. Soto* and *Wilkins v. Maryland State Police*. In both cases, a statistical study of the driving population was conducted in order to create a benchmark for comparison to police stops and searches.

In New Jersey, Dr. John Lamberth and a team of researchers conducted two surveys of the highway driving population: stationary observations and a rolling survey. In the stationary observations, roadside observers counted the number of cars that drove by along with the race of the occupants. The rolling survey consisted of researchers driving 5 mph over the posted speed limit who counted the number of speeders, i.e., drivers who passed the observers, and drivers whom the observers passed. In both cases, the race of the driver was noted. In total, 2,096 cars were counted in the rolling survey.

Analyzing the data, the team of researchers found that in New Jersey 15% of speeders on the road were African American. They compared this benchmark to the New Jersey State Police stop data, for which 35% of the drivers stopped on the same portion of the turnpike were African American. In their study, they found that African Americans were 4.85 times more likely to be pulled over when compared to others and that this disparity was statistically significant, meaning, not due to chance (see [La]). These findings convinced Lamberth – and the judge – that racial profiling was indeed taking place along the New Jersey turnpike.

Similarly, in Maryland, Lamberth conducted a study to test whether police searched African-American drivers more often than expected along a portion of Interstate 95 in Maryland. In order to find the proportion of African-American drivers violating the law, Lamberth conducted a rolling survey similar to the one conducted in New Jersey. The survey enabled Lamberth to find the racial composition of the drivers on the highway along with the composition of the violators. Lamberth used the rolling survey as a benchmark to compare with the actual police data.

The results showed that 93.3% of all drivers were violating traffic laws, of which 17.5% were African-American drivers. He then obtained data from the Maryland State Police and noted that 72.9% of all drivers searched were African American. Through statistical analysis, Lamberth found that the difference in search rates versus the expected stop rates had virtually a zero percent probability of occurring by chance, meaning that African-American motorists were searched at a significantly higher rate than is probable by chance.

According to Lamberth, the ungrounded perception that African Americans and other minorities were frequent drug users fed the motivation to target African-American drivers. Also, racial profiling could have been fueled by the fact that police officers were partially rewarded based on the number of arrests they made, encouraging more searches of drivers (see [La]). The results of Lamberth's study were used in the court as evidence. This was a landmark case since the judge accepted the idea of statistical significance and ruled accordingly.

2.1.2 Racial Profiling in Pedestrian Stops in New York

Rolling surveys of highways to study police practice are not appropriate in all jurisdictions. In New York City, after minority communities became outraged by policing tactics that allegedly targeted racial and ethnic minorities, the New York Police Department's (NYPD) "stop-and-frisk" policy was evaluated. In a study targeting racial profiling of pedestrians, Jeffrey Fagan, Andrew Gelman and Alex Kiss analyzed the New York Police Department's stop-and-frisk policy. In order to investigate the problem, researchers analyzed reports of 125,000 pedestrian stops by the NYPD between January 1998 and March 1999.

The question was whether or not police officers disproportionately stop ethnic minorities. To answer this, the researchers separated the data based on police precinct and compared stop rates for different racial and ethnic groups while controlling for differences in arrest rates (see [Ge]). Comparing the crime rates of each group, it was found that the rate of minorities being stopped was much higher than the rate of Whites being stopped. This contradicted the NYPD claim that the high stop rates of minorities represented efficient police practice where "high crime areas" tend to have more minorities (see [Ge]).

In particular, a Columbia University research group found that the NYPD's policing strategy varied in its stop and searches of pedestrians. The researchers deduced that the likelihood of African Americans and Hispanics being stopped on the streets was much higher than the likelihood of Whites. While accounting for weapons and violent crime rates by race, White suspects were stopped only half as often as African Americans and Hispanics. However, African Americans and Hispanics were less likely to get arrested than Whites. One proposed interpretation of this is that African Americans and Hispanics are stopped more often than Whites without reason.

2.1.3 A Boston Study on Race of Officer

Taking a different approach to racial profiling and traffic stops, Kate Antonovics and Brian Knight studied "preference-based discrimination" in Boston by taking note of both the race of the officer and the race of the motorists (see [An]). Preference-based discrimination refers to the act of an officer searching drivers of a race different from their own more often than searching drivers of their own race. The study aimed to understand the reasons for observed differences in the search rates between African American, Hispanic, and White drivers during traffic stops. Antonovics and Knight found that officers are more likely to conduct a search if the driver is of a different race. It appears that preference-based discrimination plays a significant role in the differences in search rates of racial groups.

2.2 History of Police Corruption in Los Angeles

The Rampart "Community Resource Against Street Hoodlums" (CRASH) unit was an anti-gang division of the Los Angeles Police Department. In 1998, allegations were made against the division for several acts of misconduct including the planting of evidence such as guns and drugs at crime scenes. Finding that these allegations were true, the city of Los Angeles entered into a consent decree with the United States Department of Justice (see [Ri]).

The Consent Decree, which monitors the LAPD, covers several aspects of police practice ranging from interactions with civilians to internal audits, in an attempt to "promote police integrity and prevent conduct that deprives persons of rights, privileges, or immunities secured or protected by the Constitution or laws of the United States." A portion of the decree requires every officer to complete a written or electronic Field Data Report for every vehicle or pedestrian stop. These Field Data Reports provide the information which we will use about each stop. Details are discussed in the section regarding data.

2.3 Use of Benchmarks

We will be addressing whether police officers search minority drivers in Los Angeles at a disproportionate rate. We compare the search rates and discovery rates to determine which factors affect these rates. In addition, we discuss the rates within different districts of Los Angeles. It is of interest to see if location affects the outcome of the stop. Hence, we will look at search rates broken down by race for each policing district in Los Angeles.

There are multiple benchmarks to use for comparison when looking at stop rates, each having its own strengths and weaknesses. Some studies have used census data to determine if police are stopping people at a rate proportionate to the demographics of the area. This is not always accurate because the census data counts everyone in the area, not just the driving population. This implies that the population being captured is not representative of the population we are studying. Another option is using data provided by the Department for Motor Vehicles (DMV) for drivers license holders to determine who is on the road. This is problematic because not everyone who drives actually has a license. To avoid these difficulties, we focus on what happens after the driver has been stopped and what factors affect the outcome of the stop. Further, to compare different police districts, we use census data broken down by region to compare the search rates between separate policing districts, grouping districts by their demographics.

2.4 Legal Bases for Searches

There are Constitutional and other legal constraints governing when a police officer may conduct a search. The Fourth Amendment guarantees "the right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures," which includes vehicles. However, many conditions allow for a search. For example, a warrant is not necessary to search a vehicle as long as the officer has probable cause that the driver is engaging in illegal activity. Vehicles are mobile, allow for easier escape, and generally are not associated with the same level of privacy as a residence.

Reasons that are considered legal basis for a search include the following: plain sight of illegal or potentially dangerous paraphernalia such as drugs, alcohol or weapons; suspicious odor such as of marijuana or alcohol; or evidence that implicates the driver or passenger of committing a crime. If officers feel necessary, they can ask for consent to search the driver or vehicle and if given voluntarily, the consent eradicates the requirement for reasonable suspicion. Searches by consent are purely at the discretion of the officer, so we spend some time looking at this data in particular for use as an indicator to determine whether some searches are influenced by race.

Race itself is not a justifiable basis for search. The Consent Decree prohibits the use of race, color, ethnicity, or origin as the sole basis for stops or any action except when a particular suspect fits a specific description, where race may be a part of that description (see [LAPD]).

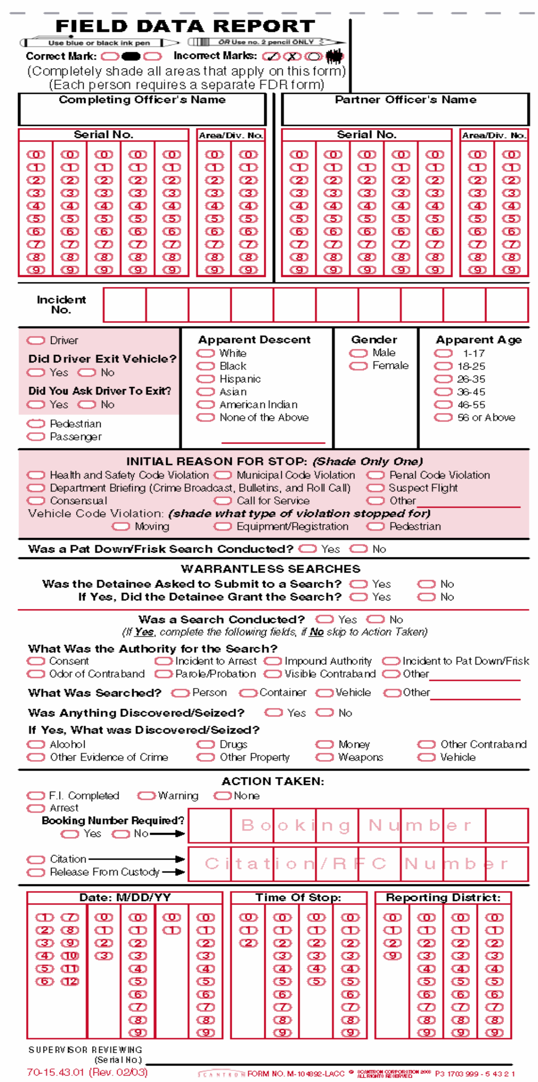
3 Data

3.1 Field Data Reports

The data we analyze comes from the Los Angeles Police Department. Following the terms of the Consent Decree, the LAPD is required to fill out a Field Data Report (Figure 1) at each traffic stop. Any time a driver, passenger, or pedestrian is detained, a Field Data Report (FDR) is filled out by the officer. When the Consent Decree was first established, police officers filled out Field Data Reports on written bubble sheets. In order to make the data more accurate, hand-held devices known as "Portable Officer Data Device Systems" (PODDS) have been developed. Since the electronic forms help eliminate misreading the forms, accuracy rates in data collection have improved. The most recent data is considered most accurate, so we focus our study on the one-year period beginning July 2004 and ending June 2005, which is the most recent period for which we have access to the data.

The Field Data Report includes information such as the race, age and gender of the detainee. The reports also include details of the stop such as: reason for stop, whether the detainee was asked to be searched, if a search was conducted, reasons for the search, if anything was discovered, what was discovered, and the outcome of the stop such as citation, arrest, no action, or warning. Also, the date, time and area of city in which the stop was made is recorded. The race noted is the perceived race of the driver. The options are White, Black, Hispanic, Asian, American Indian and None of the Above. The age is the perceived age range of the driver and includes (in years) 1-17, 18-25, 26-35, 36-45, 46-55, and 55 years or older. If the stop is because of a vehicle code violation, it is noted as one of three types: moving violation, equipment/registration violation, and pedestrian violation. There are eight choices, although more than one can be indicated, for authority for a search by the officer: consent, odor of contraband, incident to arrest, parole/probation, impound authority, visible contraband, incident to frisk, and other. There are also eight types of discoveries: vehicle, weapon, money, drug, alcohol, other contraband, other property and other evidence. A vehicle discovery implies that there is legal issue regarding the vehicle itself, e.g., a registration issue or stolen vehicle. We convert time of stop to daytime and nighttime, defining daytime as the hours from 6am to 6pm and nighttime from 6pm to 6am. The location of the stop, by policing district, is also noted, as discussed in the following section.

Figure 1: Field Data Report



Source: LAPD Training Manual [LAPD]

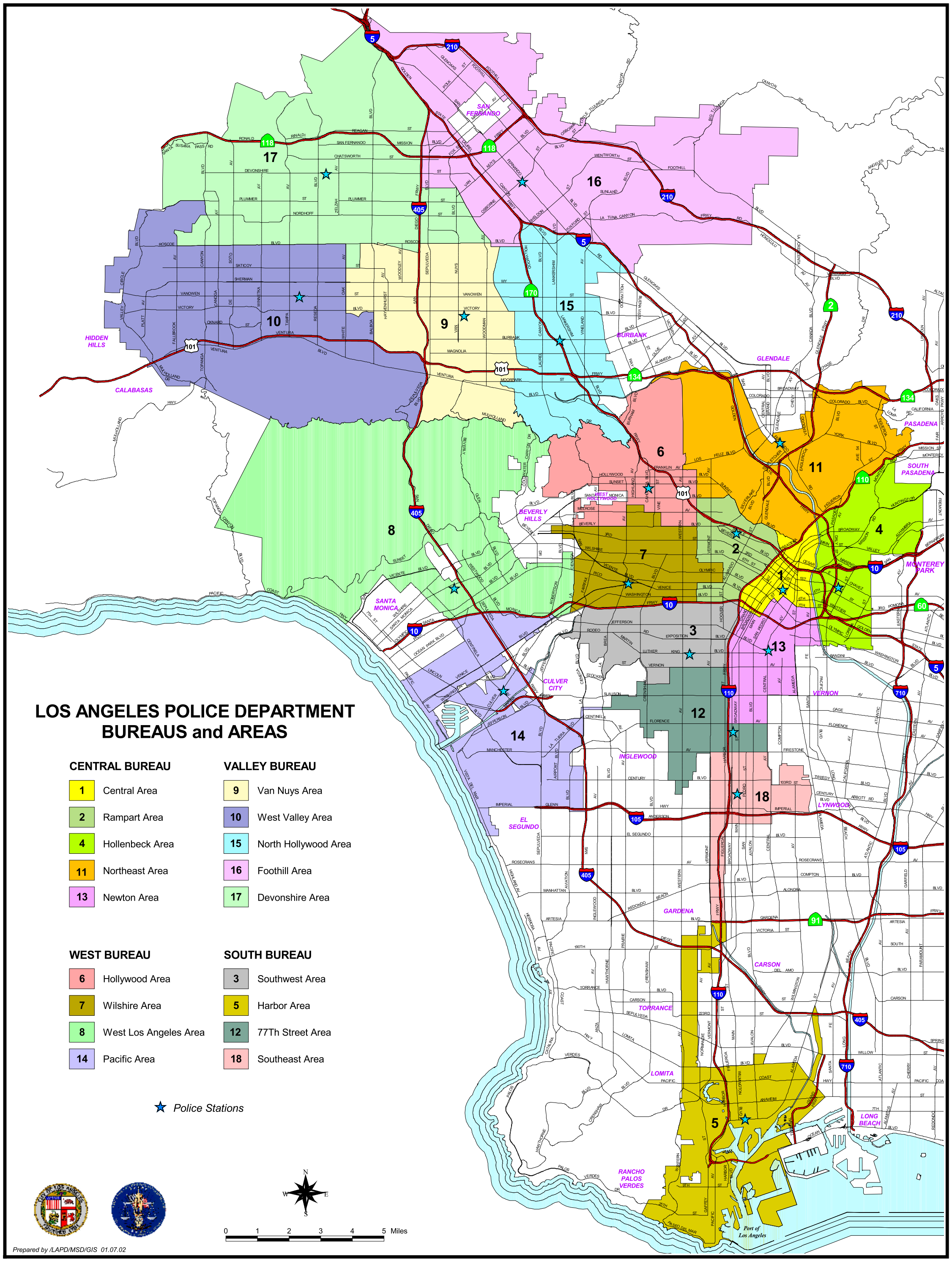
Data from the one year period includes over 950,000 driver, passenger and pedestrian observations. For this study, we focus only on the drivers. We are interested in determining how race affects what happens once the driver is stopped. A few records include incomplete data and were coded as “invalid” by the city; these 273 observations have been excluded from our study. The driver data set contains 638,732 driver records. Of the remaining number of valid drivers, 36.6% are Hispanic, 34.3% are White, 19.5% are African American, 7.9% are Asian, 1.7% are classified as Other. Although the police record Native American stops, the number of these cases is so small we have grouped them with the Other category. Within the Other category, 38.7% are classified as Middle Eastern and 10.8% are Native American. When the data is broken down by policing districts, 4,025 driver records were not coded, so they were not included in the analysis.

3.2 Geography

Los Angeles is a very diverse city made up of many communities, each of which has its own unique demographic. Some areas have a large percentage of Hispanics, some are primarily White, and others are more diverse. Much of our study takes a citywide look at the data. However, from the Field Data Reports we are aware of the general location where each stop was made. As seen in Figure 2, the city of Los Angeles is subdivided into 18 areas, called reporting districts, which we will refer to as policing districts. Each Field Data Report includes the district number in which the stop took place. This information allows us to compare the treatment of drivers between districts. Policing strategies and actions may differ depending on where the stop was conducted. In section 5, we analyze how the search, consensual search and discovery rates compare across the different districts of Los Angeles.

As shown below in Table 1 the demographics vary by large amounts according to policing district. The data was compiled using census data broken down into zip codes (see [**?**]). By comparing a zip code map with a policing district map, we approximated which zip codes fit into each policing district. From the census data, we were able to get an approximation of the demographics of each policing district. The percentages do not sum to 100 because only the four largest racial/ethnic groups are given. In section 5, we discuss the statistical methods used to determine whether being in different reporting districts change the likelihood of getting searched.

Figure 2: Los Angeles Policing Districts



Source: City of Los Angeles

Table 1: Approximate Demographics of LA by Policing District

Reporting District	Asian	African American	Hispanic	White
Central Area	27.69%	15.65%	43.01%	11.84%
Rampart Area	21.70%	3.93%	63.83%	8.71%
Southwest Area	5.11%	38.17%	47.92%	6.34%
Hollenbeck Area	7.16%	.94%	88.20%	2.80%
Harbor Area	8.18%	6.80%	56.54%	26.05%
Hollywood Area	5.99%	4.36%	27.28%	58.69%
Wilshire Area	11.88%	16.33%	25.51%	42.73%
W. Los Angeles Area	14.37%	1.89%	8.66%	71.42%
Van Nuys Area	7.41%	4.79%	52.76%	32.13%
W. Valley Area	10.19%	3.89%	30.16%	52.32%
Northeast Area	15.46%	2.53%	55.90%	23.14%
77th Street Area	.66%	50.09%	45.89%	1.54%
Newton Area	.78%	9.69%	87.10%	1.73%
Pacific Area	12.27%	10.63%	27.30%	46.07%
N. Hollywood Area	6.95%	5.17%	40.96%	43.40%
Foothill Area	4.93%	3.84%	68.68%	20.73%
Devonshire Area	16.97%	4.21%	31.20%	44.69%
Southwest Area	9.90%	27.17%	57.68%	3.82%

4 Methodology

To analyze the data, we use the chi-square test and binary logistic regression. We use chi-square tests to determine whether or not two variables are independent of one another. For example, in our analysis we test to see if search rates and race are independent of one another. In other words, does the probability of being searched vary significantly by race? The chi-square test only indicates whether the variables are independent. If the variables are found to be dependent, then we use logistic regression models to determine how the variables are related and to what extent specific variables contribute to another.

Although these methods are standard, we remind the reader of some of the details below. For more details, see [Ch] and [Wa]. SPSS was used for all logistic regression model computations.

4.1 Chi-Square

A probability distribution is a mathematical description of a random variable in terms of its values and the probability associated with each value. We will be using the chi-square distribution which is a special case of the gamma distribution. The chi-square distribution is not symmetrical and depends on the degrees of freedom, defined as follows.

Definition 1. The number of independent variables required for a statistical test is known as the number of *degrees of freedom*, *df*.

The number of degrees of freedom is used to determine the probability distribution, as follows.

A skewed probability distribution known as the gamma distribution, which has parameters $\alpha > 0$ and $\beta > 0$, is defined as:

$$f(y) = \begin{cases} \frac{y^{\alpha-1}e^{\frac{-y}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}, & \text{for } 0 \leq y < \infty \\ 0, & \text{elsewhere} \end{cases}$$

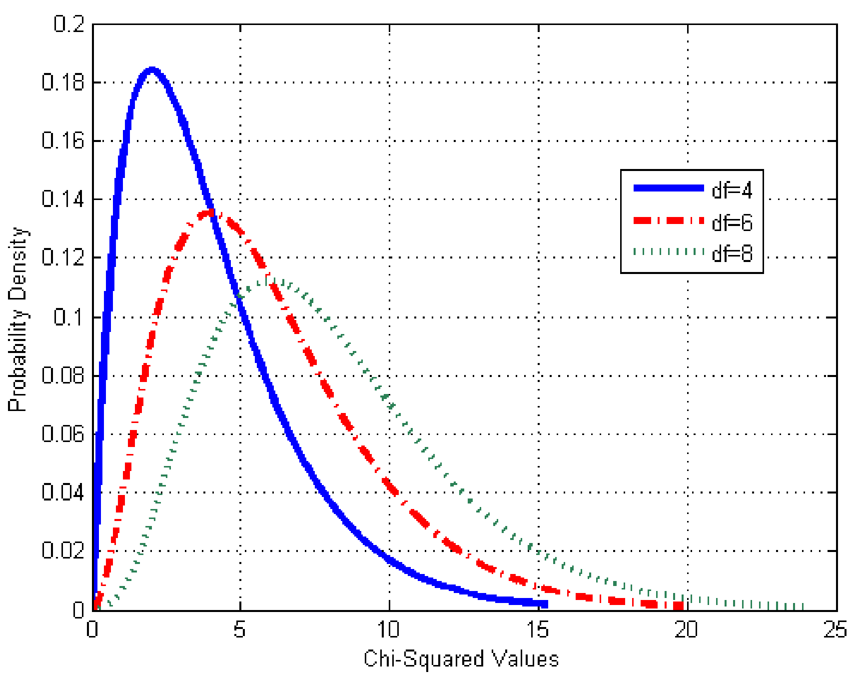
where $\Gamma(\alpha)$ is the gamma function given by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy.$$

Definition 2. A *chi-square distribution* is a special case of the gamma distribution with parameters $\alpha = \frac{v}{2}$, where v is the degrees of freedom, and $\beta = 2$.

In the chi-square distribution, the larger the degrees of freedom, the more bell-shaped the distribution begins to look.

Figure 3: Chi-Square Distribution



Using a chi-square distribution we must first calculate a chi-square statistic, defined as follows. Given categorical data in a table, each entry in the rows and columns represent a number of observations:

	Category 1a	Category 1b	Total
Category 2a	O_1	O_2	Row Total 2a
Category 2b	O_3	O_4	Row Total 2b
Total	Column Total 1a	Column Total 1b	Total

Let O_i denote the observed frequency in a category, which comes from the actual data under analysis. Let E_i denote the corresponding expected frequency, which is the theoretical value that we would expect if the data were distributed proportionally. In particular, given a table the expected frequencies can be calculated by:

$$E_i = \frac{(\text{Row Total}) \times (\text{Column Total})}{\text{Total}}$$

The degrees of freedom is simply

$$df = (\text{number of rows - 1})(\text{number of columns - 1})$$

not counting the total row or total column.

Definition 3. The *chi-square statistic*, χ^2, is given by the following formula:

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i},$$

where the sum is taken over the number of categories m.

Note especially that the greater the difference between the observed and expected values, the greater the χ^2 value.

We use the χ^2 distribution and a specific χ^2 value, say χ_0^2, for hypothesis testing. In hypothesis testing there is a null hypothesis and an alternate hypothesis.

Definition 4. The *null hypothesis*, denoted H_0, is the statement that is being tested. If the probability that the null hypothesis could have happened by chance is small, then we reject the null hypothesis. Otherwise, we fail to reject the null hypothesis.

Definition 5. The *alternate hypothesis*, denoted H_a, is the alternate to the null hypothesis. If the null hypothesis is rejected, then we will favor the alternate hypothesis.

When testing multiple variables, we use the χ^2 value to determine whether the variables are independent of each other. This is known as the Test of Independence. This test does not show how the variables relate to one another; it only shows if they are related. For a chi-square test, the following assumptions must hold: the samples are chosen at random, each outcome falls into one of m categories, and the sample size must be large enough so that the expected values are greater than or equal to five.

When using the Test of Independence, our null hypothesis is that the variables are independent. We reject the null hypothesis if it is unlikely to have happened by chance. Using a table or computer, one can find the probability of a chi-square being χ_0^2 or larger.

Definition 6. The *P-value*, $P(\chi^2 > \chi_0^2)$, is the probability of χ^2 being greater than χ_0^2.

In order to determine if a null hypothesis should be accepted, we compare the P-value to a threshold value of α which is determined by the researcher. Traditionally, social scientists often use $\alpha = 0.05$ while others may use $\alpha = 0.01$. If the chi-square value corresponds to a P-value less than α, then the null hypothesis is rejected. (Alternatively, the degrees of freedom and α may be used to find a critical χ^2 value with which to compare the actual χ_0^2 value, rather than computing P.)

Definition 7. The *level of significance*, denoted α, is a fixed probability of wrongly rejecting the null hypothesis when it is true.

4.2 Logistic Regression

The purpose of logistic regression is to determine the relationship between explanatory variables and a categorical response variable.

Definition 8. The *explanatory variables* are the independent variables, X_i.

Definition 9. The *response variable* is the dependent variable, Y.

Definition 10. A variable is *discrete* if it has a finite number of possible values.

Definition 11. A variable is *dichotomous*, also called binary, if it has only two possible values. For example, a dichotomous variable could have categories "Yes" and "No" or "1" and "0."

Under standard multiple linear regression, the response variable needs to be a continuous quantitative variable. The variables in our data, however, are discrete variables. In addition, our response variables, as mentioned in section 2.1, are dichotomous. Because this is the case, we cannot use multiple linear regression. Instead, it is appropriate to use logistic regression in our analysis to determine how the explanatory variables affect the response variable.

In our study we code our response variables as either 0 or 1. Although the response variable must be categorical, the explanatory variables have no such restriction in logistic regression models.

Definition 12. The *odds* is the ratio of the likelihood of success to the likelihood of failure. Let p denote the probability that $Y = 1$, which will constitute a success, i.e., $p = P(Y = 1)$. Then,

$$\text{log(odds)} = \text{odds} = \left(\frac{p}{1-p}\right) \tag{1}$$

For example, $p = \frac{2}{3}$ implies that for every 2 successes, there are 3 failures.

In equations (2) and (3), the β_i value is called a *partial regression coefficient*, or *parameter*. Suppose there are k explanatory variables $X_1, X_2, ..., X_k$ and a response variable, Y. The standard logistic regression model uses log odds, as follows:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k. \tag{2}$$

The transformation on the left is often referred to as the logit function:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right).$$

Note that as the X_i vary, logit(p) can take on any value, but $0 < p < 1$, as required for a probability value. Solving for p, we obtain

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)}}. \tag{3}$$

4.2.1 Parameter Estimates

Generally, in linear regression, the parameters β_i indicate how Y changes in response to one unit change in X. In logistic regression, we interpret β_i via an odds ratio explained below. Each β_i shows the relationship between X_i and the log(odds) of Y. The closer β_i is to 0, the less important X_i is in predicting the probability of Y being a success.

Definition 13. The *odds ratio* is defined as the ratio of the odds of one outcome compared to the odds of a second outcome. Given two possibilities of success, p_1 and p_2, we have

$$\text{odds ratio} = \frac{\left(\frac{p_1}{1-p_1}\right)}{\left(\frac{p_2}{1-p_2}\right)}. \tag{4}$$

From a logistic regression model, we can find the odds ratio using e^{β}. In particular, if $p_1 = P(Y = 1|X_i = 1)$ and $p_2 = P(Y = 1|X_i = 0)$, equation 2 shows that e^{β_i} is the odds ratio. Note that the closer the odds ratio is to 1, the more the explanatory variables are independent of the response variable. When the odds ratio = 1, the variables are completely independent. For example, let $\beta = 1.609$. Then $e^{\beta} = e^{1.609} = 5$. This means that when the explanatory variable increases by one unit, the odds that the response variable equals one increases by a factor of 5 when all other explanatory variables are held constant.

4.2.2 Accuracy of Model and R Square

When using logistic regression, in addition to parameter estimates statistical software outputs information useful for evaluating a model. Table 2 includes the likelihood ratio, -2LL, and two types of R Squares, Cox & Snell and Nagelkerke. We do not describe these in detail but make a few comments.

Table 2: SPSS Model Summary Output

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	448152.08	.096	.175

The likelihood ratio is close to a chi-square distribution for large sample sizes. It is used as a goodness of fit test for the model. As the model improves, -2LL will decrease. Along with -2LL, other measures of goodness of fit are given such as Cox & Snell R Square and Nagelkerke R Square. The R Square values typically range from 0 to 1 and the closer to one, the more accurate the model. These values all reveal similar information about the fit of the model.

4.2.3 Variable Selection

An important aspect of logistic regression is variable selection. From the Field Data Reports we have over 50 variables. However, to limit the number of variables, our goal is to determine which explanatory variables significantly affect the response variable and which do not. There are many ways to do this one of which is the chi-square test. By comparing the different chi-square values, we can take note of which variables lead to significant P-values. The explanatory variables that lead to significant values are the variables which will contribute the most to the prediction of the response variable. The explanatory variables with an insignificant P-value are the variables we can potentially eliminate from the model since they will not help much in predicting the response variable.

There are also methods implemented by SPSS that will select the most valuable variables. Forward Selection is a method in which variables are added one at a time in the order of their significance. For example, a variable with a P-value of 0.04 will be added after a variable with a P-value of 0.01. Once the model contains the most significant variables, as defined by the user, the selection process stops. Furthermore, Backward Elimination chooses variables in the opposite manner. It inserts all variables into the model and removes them in the order of least contribution. Lastly, there is a method known as Stepwise. This method combines Forward Selection with Backward Elimination. The process continually adds and removes variables based on the significance of each. This gives us a balance of Forward and Backward methods. If used correctly, each model should yield similar results, however, not always the same variables.

5 Findings/Analysis

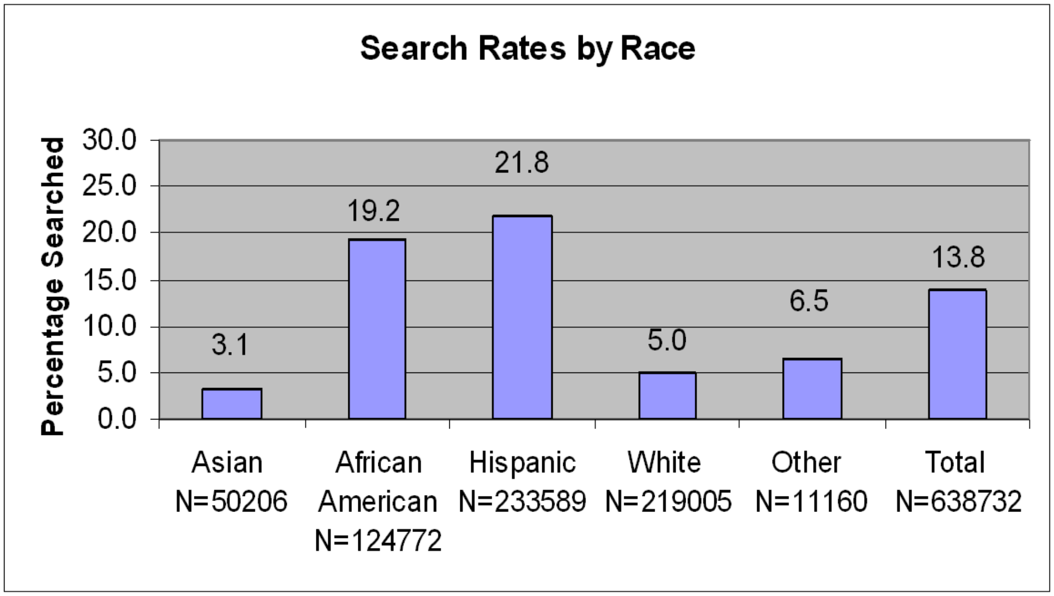
Within our findings, we first look into citywide data. We consider at search rates, purely consensual search rates, discovery rates, and discoveries from purely consensual searches. Also, discussion is included regarding which variables contribute to predicting whether the driver is asked for search. The data is then broken down by policing district to determine how the data changes as the geography changes. We look at search rates and discovery rates followed by models to analyze demographic differences throughout the city.

5.1 Analyzing Disparities Citywide Across Los Angeles

5.1.1 Search Rates

A search rate as it is used in this paper is defined as the number of drivers of a race or ethnicity searched over the number of drivers of that race stopped. If drivers of different races each had an equal chance of being searched, search rates for each group would be similar. Analyzing Los Angeles and the total number of drivers stopped and searched per race, we find striking disparities. As seen in Figure 3, search rates are highest for Hispanics and African Americans and lowest for Asians and Whites. There is a large disparity, African-American and Hispanic drivers are approximately 4 times more likely to be searched than White drivers. Also, Hispanic drivers are 7 times more likely to be searched than Asian drivers.

Figure 3: LAPD Search Rates, July '04- June '05



To see if these disparities are statistically significant, we used a chi-square test where there are 4 degrees of freedom with α=0.01. Here, our null hypothesis is that search rates are independent of race and our alternate hypothesis is that search rates are not independent of race. We obtained a chi-square value of over 35,000 from the observation data, Table 3. The area under the curve to the right of this chi-square value is effectively zero (and in fact, a χ^2 value such as this is literally off the chart). We reject the null hypothesis, and conclude search rate is not independent of race. The probability that these numbers could be distributed this way by chance is effectively zero.

Table 3: Number of Drivers Searched by Race

	White	African American	Latino	Asian	Other	Total
Searched	10944	23983	51017	1557	726	88227
Not Searched	208061	100789	182572	48649	10434	550505
Total	219005	124772	233589	50206	11160	638732

$$\chi^2 = 35346.38$$

5.1.2 Logistic Regression: Search Conducted

The χ^2 tests showed us that our variables were not independent of the search variable. To find how the explanatory variables affect the response variables, we used a logistic regression model. We tested how explanatory variables such as age, race, sex, reason for stop and time of stop affect the probability of being searched (Table 4). The parameters given by logistic regression are known as parameter estimates, $\hat{\beta}$, and refer to the values that SPSS finds through an iterative algorithm. These values tell how much the explanatory variables contribute to the prediction of the response variable. We test at the 1% level of significance. If the value in the significance column is less than 0.01, the explanatory variable is considered significant. That is, the explanatory variable contributes significantly to the model. Significance in this case refers to the probability that $\hat{\beta} = 0$, or that the variable does not contribute to the model. For the interested reader, the standard error for the parameter estimate and 95% confidence intervals (C.I.) for the odds ratio are included.

Table 4: Variables that may Affect Search Conducted

Explanatory Variable	$\hat{\beta}$	Standard Error	Significance	$e^{\hat{\beta}}$	95% C.I. Lower	95% C.I. Upper
Sex(Female = 1)	-1.133	.012	.000	.322	.314	.330
Race			.000			
Asian	-.507	.031	.000	.603	.567	.640
African American	1.471	.014	.000	4.355	4.238	4.238
Hispanic	1.488	.013	.000	4.429	4.321	4.541
Other	.051	.047	.276	1.052	.960	1.154
Age Range			.000			
1-17 years	2.242	.040	.000	9.408	8.692	10.182
18-25 years	1.578	.032	.000	4.847	4.551	5.162
26-35 years	1.302	.032	.000	3.678	3.453	3.918
36-45 years	1.012	.033	.000	2.752	2.581	2.936
46-55 years	.650	.035	.000	1.916	1.789	2.053
Moving	-.229	.008	.000	.795	.782	.809
Night	.672	.008	.000	1.957	1.925	1.990
Constant	-4.182	.033	.000	.015		

(Significance = 0.000 implies P-value < 0.001)

In this table, the odds ratio is interpreted with respect to a baseline value. The baseline for race is White, so each race is compared to it. This analysis shows that the odds of being searched for an African American is 4.355 times the odds of being searched for a White driver. Similarly, the odds of being searched for a Hispanic is 4.429 times the odds of being searched for a White driver.

In this model, female drivers are coded as 1 and male drivers as 0. The baseline is male, so female drivers are compared to male drivers when interpreting the data. The odds of a female being searched is 0.322 times the odds of a male being searched. The Age Range variable represents the age group of the driver. Recall from the Field Data Report that these are divided into drivers who are 1-17 years, 18-25, 26-35, 36-45, 46-55 and 55 and older. Here the baseline is the group, 55 and older. Each age group is significant in computing the probability the driver is searched. The younger the driver is, the higher the odds ratio. For example, the odds of a driver between the age of 1-17 being searched is 9.408 times the odds of a driver 55 years and older being searched.

The parameter estimate value for "Moving" is negative, which means when stopped for a moving violation, the odds of being searched is less than the odds of being searched when stopped for a non-moving violation. The variable Night represents which half of the day the driver was pulled over, day (6am to 6pm) or night (6pm to 6am). From this model, we know that when drivers are stopped at night, the odds of being searched are almost two times as much as if stopped during the day .

This test resulted in a -2LL of 382699.5 and a Cox and Snell R^2 value of 0.201, which implies there are other significant variables not accounted for in this model. However, we should not expect variables such as race to completely determine search practice; rather, we are testing to if they have a significant effect.

5.1.3 Discovery Rates

A discovery rate is defined as the number of discoveries over the number of searches conducted. This number tells us out of all the searches, what percent resulted in a discovery, of success. Some may argue that disparities in search rates can be attributed to previous findings from searches. However, when comparing the different types of discoveries from the Field Data Report, we find this is not always the case (Table 5).

Table 5: Discovery Rate Breakdown

	Asians	African Americans	Hispanics	Whites	Other	Total
Weapon Discovered	1.0%	1.4%	1.1%	1.6%	1.1%	1.2%
Money Discovered	1.1%	1.5%	1.2%	1.7%	2.3%	1.3%
Drug Discovered	6.6%	7.4%	4.6%	10.9%	8.3%	6.2%
Alcohol Discovered	0.6%	1.0%	1.3%	1.1%	.83%	1.2%
Other Contraband Discovered	1.9%	1.5%	1.0%	3.4%	3.6%	1.4%
Other Evidence Discovered	1.5%	1.5%	1.8%	2.6%	2.5%	1.8%
Vehicle Discovered	46.2%	30.1%	53.5%	35.3%	32.9%	44.6%
Other Property Discovered	4.5%	4.2%	5.4%	7.1%	6.5%	5.3%
Any Discovery/ Discoveries Made	59.4%	44.6%	66.2%	56.2%	51.5%	55.6%
No Discovery Made	40.6%	55.4%	33.8%	43.8%	48.5%	44.4%
Total Drivers	1557	23983	51017	10944	726	88227

Among the eight types of discoveries shown in the table, weapons, money and alcohol are the least likely to be found overall. When we compared discoveries of weapons, money and alcohol by race, we found that the percentage differences between the races are not significant. All other differences by race are statistically significant.

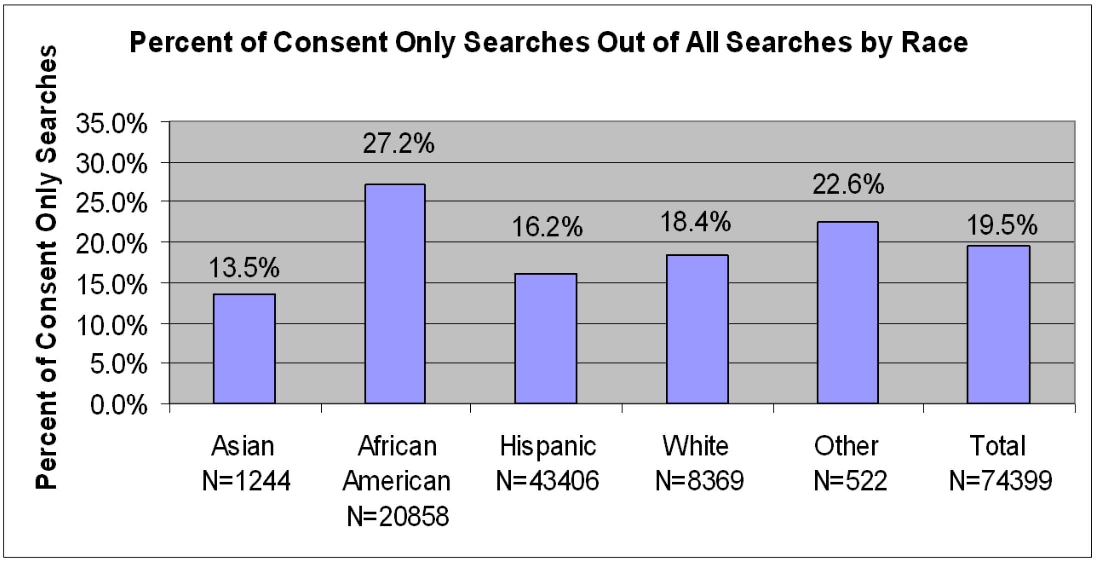
Notice that of all discovery types, a vehicle discovery is by far the most common, where a vehicle discovery is may be a stolen vehicle or other registration issue, for example. Breaking the vehicle discovery rate down by race, we see that Hispanic drivers have the highest discovery rate and African American drivers the lowest. Furthermore, White drivers are over 2 times more likely to have drugs discovered and 3 times more likely to have other contraband discovered than Hispanic drivers. Although vehicle discoveries make up the majority of discoveries, non-vehicle discoveries include drugs, money, and weapons, among other things. African Americans and Whites had more non-vehicle discoveries. Notably, Hispanics, who have more vehicle discoveries, have fewer non-vehicle discoveries than White and African American drivers.

5.1.4 Purely Consensual Search Rates

There are many reasons for an officer to conduct a search. At times, however, police officers may simply ask a driver for consent to search. For example, if there is no visible contraband or open alcohol present, the officer still has the authority to ask the driver to consent to a search. Moreover, drivers may not know that they have a legal right to refuse a consensual search. Hence, a consensual search is an example of a search for which an officer exercises much discretion. On the Field Data Report, the officer can indicate more than one reason for the search; here we analyze searches for which the only basis indicated was consent of driver, or purely consensual searches.

As we compare the purely consensual search rates across the races, we note statistically significant differences. When looking at the purely consensual search rate, Asians are searched the least (Figure 4). In addition, African Americans have the highest consensual search rate. A chi-square test shows these disparities to be statistically significant, meaning they could not have been distributed in such a way by chance.

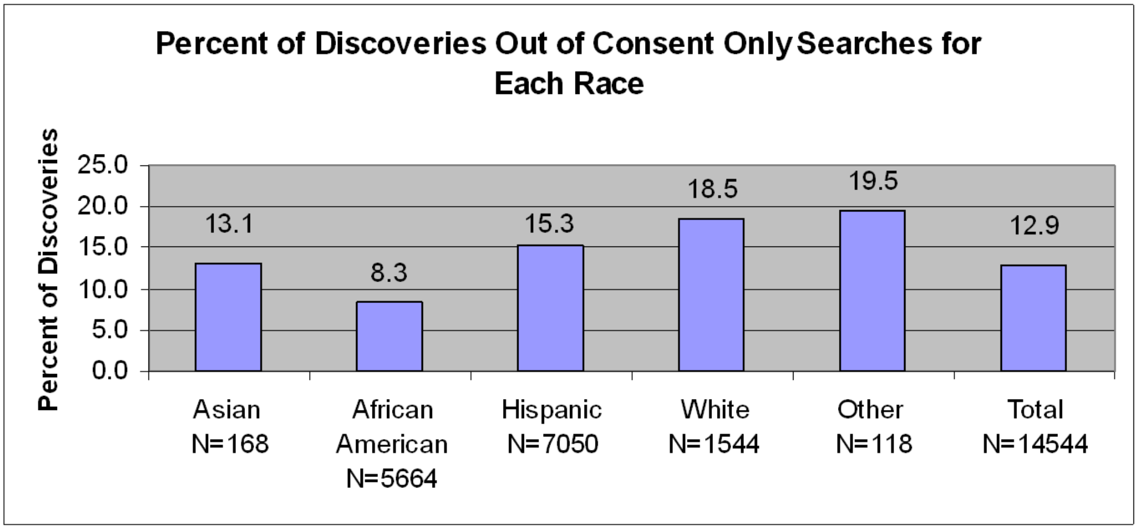
Figure 4: Consent Only Searches by Race



5.1.5 Discovery Rates from Consent Based Searches

Because of the high discretion officers have in conducting consent-based searches, we considered the discovery rates for purely consensual searches. As you can see in Figure 5, purely consensual searches of African-American drivers result in a discovery much less than for any other race, although the consensual search rate for African-Americans is the highest. White drivers have the second highest discovery rate from purely consensual searches. The differences in discovery rates of consensual searches by race are statistically significant. A chi-square test gives a value of about 189 which means the probability of the rates occurring in this way by chance is effectively zero.

Figure 5: Discovery Rates of Consensual Searches Only



5.1.6 Logistic Regression: Discoveries from Purely Consensual Searches

Now that the differences have been shown to be statistically significant, we can determine what variables play a part in determining if a discovery is made from a purely consent based search. We use logistic regression on purely consensual searches with the response variable as discoveries. Race, sex, age, reason for stop and time are included as explanatory variables.

The next table, Table 6, shows the resulting parameter estimates, standard error, significance, the odds ratio $e^{\hat{\beta}}$, and the 95% confidence interval for the odds ratio. Each race is being compared against the baseline, White drivers. The only race for whom the parameter estimate, $\hat{\beta}$, is significantly different from White drivers is African American. The odds of making a discovery during a consensual search of an African-American driver is just over half the odds of making a discovery when searching a White driver. Other than this, the race of the driver generally does not significantly affect the odds of making a discovery from a purely consensual search, as can be seen by the Significance column.

Table 6: Logistic Regression – Discoveries out of Consensual Only Searches

Explanatory Variable	$\hat{\beta}$	Standard Error	Significance	$e^{\hat{\beta}}$	95.0% C.I. Lower	95.0% C.I. Upper
Race			.000			
Asian	-.287	.239	.230	.750	.469	1.199
African American	-.836	.082	.000	.434	.369	.509
Hispanic	-.067	.077	.380	.935	.804	1.087
Other Race	.163	.243	.504	1.176	.730	1.896
Sex(Male = 1)	-.357	.103	.001	.700	.572	856
Age Range			.000			
46-55 years	.616	.184	.001	1.852	1.291	2.656
26-35 years	.407	.152	.007	1.503	1.115	2.024
36-45 years	.497	.160	.002	1.644	1.202	2.249
18-25 years	.057	.150	.701	1.059	.790	1.421
55 years and up	.655	.306	.032	1.925	1.057	3.505
Moving	.183	.050	.000	1.201	1.088	1.325
Night	-.097	.054	.073	.908	.817	1.009
Constant	-1.537	.189	.000	.215		

(Significance = 0.000 implies P-value < 0.001)

The baseline used to compare the age range is the drivers in the 1-17 years category. The drivers who are 26 years or older have significantly higher odds of having a discovery made through a consensual search than the odds of having a discovery made if 1-17 years, although the parameter estimate for age 55 and up is not significant at the $\alpha = 0.01$ level.

Also, the odds of having a discovery made from a consent search if pulled over for a moving violation is 1.201 times the odds of a discovery made from a consent search if pulled over for a non-moving violation. From this table, it can also be deduced that the odds of a discovery made for males during a purely consensual search is actually 30% less than the odds of making a discovery for females. The time of day in which the stop took place is not statistically significant to discovery rates through consensual searches only. Overall, the explanatory variables with the highest significance to the outcome of a discovery by purely consensual search in our logistic model are the age range of the driver and the the type of violation that resulted in the initial stop and African American race.

5.1.7 Logistic Regression: Driver Asked to be Searched

A portion of the FDR contains information about whether the driver was asked to be searched. We are trying to determine if all of our explanatory variables contribute a significant amount to this outcome. Table 7 shows the parameter estimates and how they change as new variables are added to the model. Age is added first, then reason for stop, then time. As the table shows, the parameter estimates do not change a large amount as each variable is added, but they do change. This means that the new variable contributes to the model and changes how much the old variables contribute, whether positive or negative.

Table 7: Parameter Estimates, Four Models: Driver Asked to be Searched

	Model 1 $\hat{\beta}$	Model 2 $\hat{\beta}$.	Model 3 $\hat{\beta}$	Model 4 $\hat{\beta}$
Sex(Female = 1)	-1.925	-1.897	-1.881	-1.801
Asian	-.672	-.707	-.727	-.752
African American	1.732	1.637	1.585	1.510
Hispanic	1.353	1.150	1.127	1.068
Other Race	.226	.176	.186	.139
18-25 years	-	-.366	-.392	-.425
26-35 years	-	-.842	-.866	-.860
36-45 years	-	-1.194	-1.209	-1.160
46-55 years	-	-1.613	-1.619	-1.536
55 years and up	-	-2.641	-2.621	-2.454
Moving	-	-	-.353	-.266
Night	-	-	-	.962

Table 8 gives the complete SPSS output when all the explanatory variables are added. From this table, we can see that all the variables except for Other as race are significant at the 1% level (and in fact all are significant at the 5% level), so the explanatory variables play a significant role in predicting whether the driver is asked to be searched. The highest magnitude of the parameter estimates are with the variables sex, African American, 46-55 years, and 55 years and up. This can be interpreted as these four explanatory variables contributing the most to predicting whether the driver is asked to be searched. Since all the variables are significant, they all contribute to the model, but the variables with the highest odds ratios contribute most.

Table 8: Logistic Regression - Driver Asked to be Searched as Response Variable

Explanatory Variable	$\hat{\beta}$	Standard Error	Significance	$e^{\hat{\beta}}$	95% C.I. Lower	95% C.I. Upper
Sex(Male = 1)	1.801	.025	.000	.165	5.767	7.367
Asian	-.752	.053	.000	.471	.425	.523
African American	1.510	.021	.000	4.528	4.350	4.714
Hispanic	1.068	.020	.000	2.910	2.800	3.023
Other Race	.139	.067	.039	1.150	1.007	1.312
18-25 years	-.425	.036	.000	.654	.609	.701
26-35 years	-.860	.036	.000	.423	.394	.454
36-45 years	-1.160	.038	.000	.313	.291	.338
46-55 years	-1.536	.043	.000	.215	.198	.234
55 yrs and up	-2.454	.070	.000	.086	.075	.099
Moving	-.266	.013	.000	.767	.748	.786
Night	.962	.013	.000	2.617	2.551	2.685

(Significance = 0.000 implies P-value < 0.001)

Notice the odds ratios for all variables, especially sex, where male is compared to the baseline, female, race, and night. The odds of an African-American driver being asked to be searched is 4.714 times the odds of a White driver being asked to be searched. Also, the odds of an Asian driver being asked to be searched is 0.523 times the odds of a White driver being asked to be searched. The odds of a male driver being asked to be searched is 7.367 times the odds of a female driver being asked to be searched.

5.2 Analyzing Disparities Across Policing Districts

5.2.1 Search Rates Across Policing Districts

As we have seen, search rates differ greatly by race. African Americans and Hispanics are searched at higher rates than Whites, so a question we want to answer is whether or not these rates also differ by policing district. The FDR designates which policing district they were stopped in.

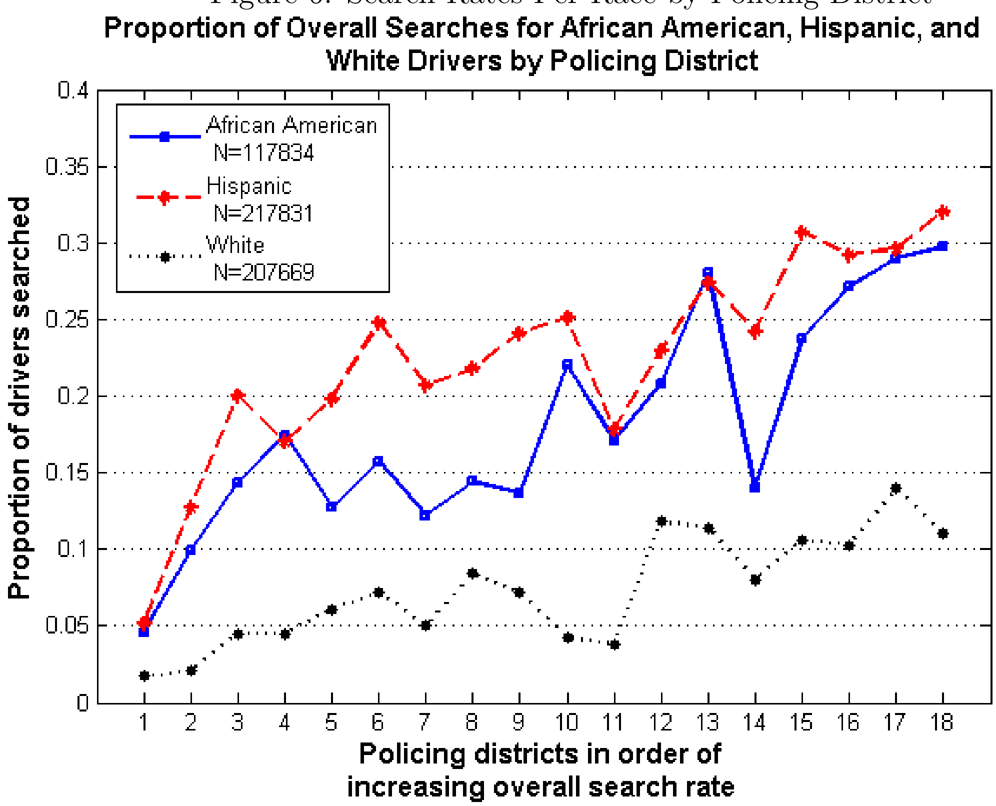
Analyzing the 18 districts (Table 9), we found the two districts with the highest search rates are 77th Street and Southeast Area. These two with lowest search rates are the Pacific Area and West Los Angeles. Comparing the extremes, the search rate in the top two districts is about 4-5 times that of the lowest two districts. This shows the great disparities within districts. Overall, the search rates of the districts range between 2.8% and 29.8%.

Table 9: Policing District Search Rates

Policing District	Search Rate		Policing District	Search Rate
RD 1: Central Area	12.1%		RD 10: West Valley Area	12.6%
RD 2: Rampart Area	23.5%		RD 11: Northeast Area	13.1%
RD 3: Southwest Area	16.2%		RD 12: 77th Street Area	27.3%
RD 4: Hollenbeck Area	22.1%		RD 13: Newton Street Area	29.8%
RD 5: Harbor Area	19.0%		RD 14: Pacific Area	2.8%
RD 6: Hollywood Area	13.5%		RD 15: North Hollywood Area	12.1%
RD 7: Wilshire Area	15.5%		RD 16: Foothill Area	21.8%
RD 8: West Los Angeles	4.6%		RD 17: Devonshire Area	13.4%
RD 9: Van Nuys Area	11.0%		RD 18: Southeast Area	28.7%

When each area is broken down by race, we are able to see if certain races are searched more depending on the policing district (Figure 6). For example, looking at Newton Street Area and Southeast Area, areas with the highest overall search rates, we observe that Hispanics are searched the most, followed by African Americans. Using the chi-square test, the differences by race are found to be statistically significant in all areas.

Figure 6: Search Rates Per Race by Policing District



5.2.2 Logistic Regression: Search Conducted Across Policing Districts

Previously, logistic regression was performed on the citywide data. Now, we take a closer look at each policing district to see which variables help to accurately predict the response variable. After performing a chi-square test, it was shown that search rates and policing districts were not independent of each other. In order to investigate which variables contributed the most to search rates by district, we performed logistic regression on each district. In these models, the only drivers considered were those stopped for a moving violation or an equipment/registration violation. Explanatory variables include race, sex, age, reason for stop, and time of day. The baseline comparisons are the following: for race, White; for sex, male; for age, 55 years and up; for reason for stop, non-moving; and for time of stop, day.

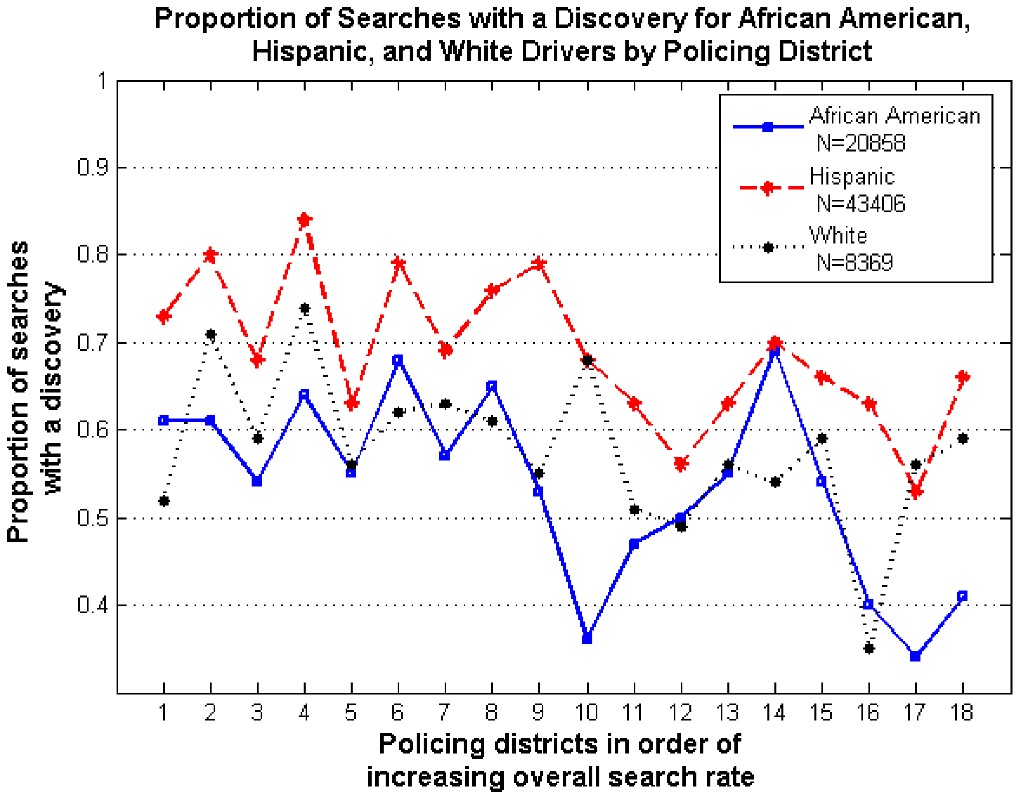
The logistic regression model outputs can be found in the appendix. The results show that in every district the odds of a male being searched is higher than the odds of a female being searched. In every policing district except the Pacific Area, the odds for being searched if pulled over for a non-moving violation is higher than the odds for being searched when pulled over for a moving violation.

The variables that contribute the most to the model in all districts are the first three age groups, ages 1-35, and the two races/ethnicities African American and Hispanic. In every district except the Van Nuys Area, the variable with the largest odds ratio is the youngest age group, 1-17 years. Finally, we note that for two areas, Foothill and Newton Street, all of the variables that are statistically significant.

5.2.3 Discovery Rates Across Policing Districts

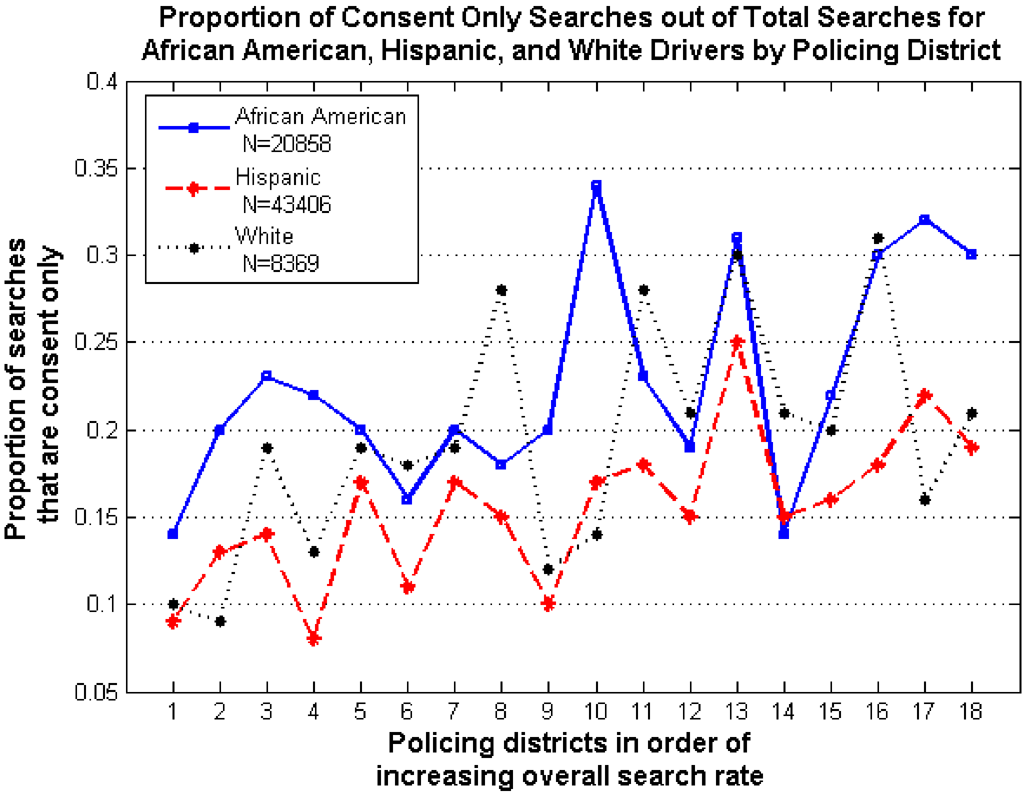
Next, we examine discovery rates and how they differ by the different geographic locations. In section 5.1, we saw that there are disparities in discovery rates citywide. Recall that African Americans are searched at a rate roughly four times higher than Whites, but these searches have a low discovery rate. Overall, discovery rates range from roughly 35% to a little over 80%. Hispanics have the highest overall discovery rates. In Figure 7, we can see that the discovery rates do indeed differ by policing district.

Figure 7: Discovery Rates by Policing District



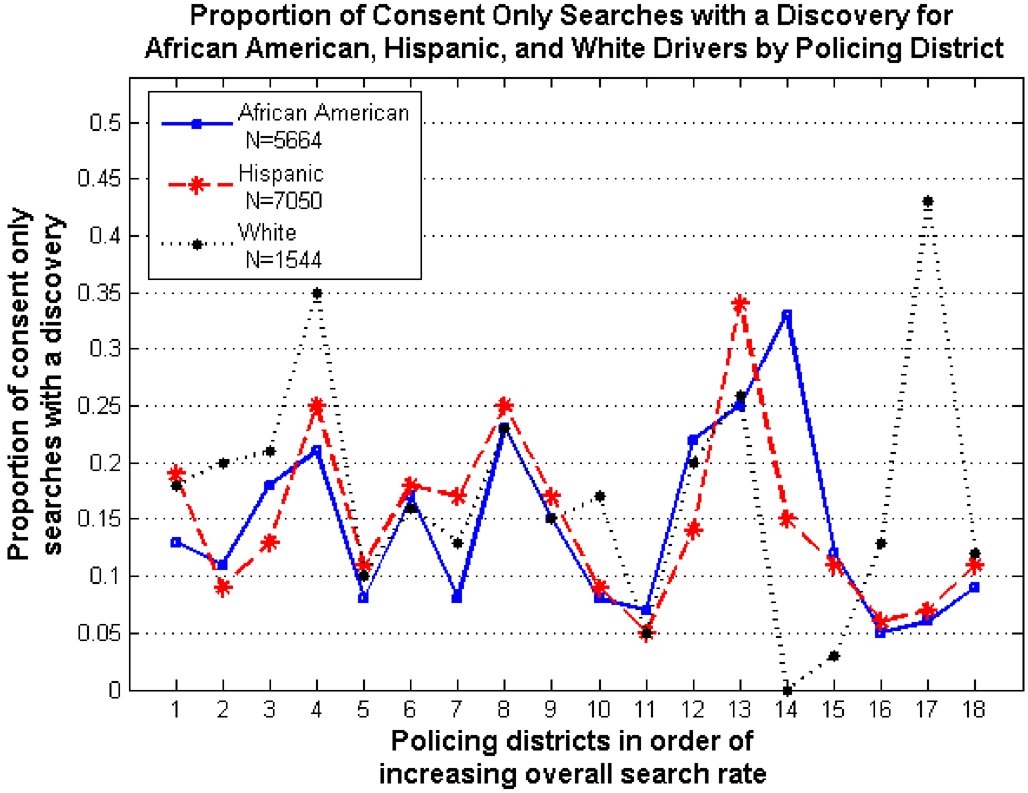
Next, we focus in particular on searches based purely on consent, for each policing districts. Figure 8 shows these percentages by race. African American consent search rates are higher than Hispanic and White consent rates most of the time. The consent search rates for Hispanics are never highest in any of the policing districts.

Figure 8: Purely Consensual Search Rates by Policing District



Next, we can compare the discovery rates from purely consensual searches to see if the high consent search rates for African Americans are accompanied by a high discovery rate within those searches, in contrast to the lower overall discovery rates. In fact, however, discovery rates for consensual searches are lower overall. Figure 9 shows the varying discovery rates from purely consensual searches across the policing districts. The range of discovery rates is from 5% to 35% for African-American and Hispanic drivers and 0% to roughly 40% for White drivers. This is in marked contrast to the discovery rates from all searches, which ranged from roughly 35% to a little over 80%. Thus from purely consensual searches, discoveries are much less likely than from searches with other bases.

Figure 9: Discoveries From Purely Consensual Search Rates by Policing District



5.2.4 Demographic Models

We have seen that search rates vary by policing district. Next, we consider how race plays a role in the make up of the policing district. We categorized policing districts by percentage of African American and Hispanic residents, using demographics by zip-code from the U.S. 2000 Census ([Zp]). The districts were then grouped into three categories; less than 50% African American and Hispanic, 50%-75%, and 75% and higher. This gives an indication of how policing may vary between communities with higher percentages of African Americans and Hispanics and the communities with lower percentages. In Figure 10, search rates by race/ethnicity are given for each category.

Table 10: Search Rates by Percentage of African Americans and Hispanics in Policing District

Percent African Americans and Hispanics in districts	Asian	African American	Hispanic	Other	White	Total
A. 0-50%	2.6	16.0	17.9	6.1	4.3	9.7
B. 50-75%	4.2	18.6	23.2	7.5	6.7	15.7
C. 75-100%	3.5	22.7	26.3	7.0	7.0	23.3

Comparing these demographic areas, we find that search rates do differ depending on the demographics of the area. We note two patterns which emerge. First, for the largest groups of drivers (African American, Hispanic, and White), as the percent of African American and Hispanic residents increase, so does the search rate. In category C, a driver is over 2 times more likely to be searched than in category A. Second, the absolute difference in search rates (e.g. 16%-4.3% for African American search rate-White search rate in category A) is lower in areas with lower percentages of African American and Hispanic residents. However, in each grouping, African American and Hispanics have much higher search rates than others.

5.2.5 Simpson's Diversity Index Trial

The coding of the demographics of policing districts allowed us to test the idea that although overall search rates increase with the percentage of minorities, a person of color could be more likely to be searched both in a majority White area and in a majority African American or Hispanic area, relative to a more diverse area. To do this, we used Simpson's Diversity Index, which is the sum of the squares of each of the demographic percentages, to compare or rank the disparity of each area in relation to one another. After ordering the districts by diversity index (but not combining them in categories) and comparing to the corresponding search rates for each race/ethnicity, there was only a weak correlation. In general, areas that are more diverse or are majority African American and Hispanic have higher search for all drivers rates, as seen in the previous section.

6 Conclusion

Initially, our concern was with whether or not racial profiling was being practiced in Los Angeles. Our results give sufficient evidence to support the possibility that racial profiling is indeed occurring. This is demonstrated by the fact that in every chi-square test and logistic regression model that we ran, the variable of race had a statistically significant effect. Specifically, chi-square tests and logistic regression models show that search rates and race are not independent in Los Angeles. Search rates have large disparities when compared by reporting districts, suggesting that in different areas of Los Angeles, officers policing the areas are using different tactics. However, it is important to note that although race is a significant factor as to whether one is searched or not, it is not the only factor. When testing search rates citywide, other explanatory variables were also significant.

Our study focuses on data between July 2004 and June 2005. From the data, we found that officers from the Los Angeles Police Department searched African-American and Hispanic motorists approximately four times more than White drivers and almost six times more than Asian motorists. The disparities between the search rates for African Americans, Hispanics, Whites and Asians are statistically significant and do not give reason to rule out that racial profiling may be occurring in Los Angeles.

Not only is search rate important, but so is the rate at which discoveries are made. If fewer discoveries are made in tandem with high search rates, it is hard to justify searching certain races more than others. In particular, discovery rates for White drivers are higher than the rates for African American drivers; at the same time, African-American drivers are subjected to searches more often. Also, Hispanics have the highest discovery rates, the majority of which are concern the vehicle which is being driven. It is interesting to note that African Americans are asked for consensual searches at high rates, but their discovery rates from consensual searches are always lower than Hispanics and often lower than Whites.

In each of the logistic regression models, the -2LL was very large, and the R^2 values were very small, which implies that the fit of the model is poor. However, this was to be expected, as we are considering the affects of various explanatory variables, and outcomes such whether or not a search is conducted should not depend purely on the variables to which we have access, such as gender. Our goal is not to see if these variables completely determine search practice but rather to test to see if they have significant effect. Moreover, the data set contains over 600,000 observations and the -2LL value is magnified by the size of the sample, all things being equal, and hence we expect a fairly large -2LL value for tests which include the entire set of drivers.

Because of the significance of the disparities found, it would be beneficial for the LAPD to consider what is contributing to these differences. The findings do not allay concerns that racial profiling may be happening in Los Angeles.

7 Future Work

Our work thus far is limited to the use of logistic regression models which only include binary categorical response variables. However, not all of our response variables need be binary. For the variables which are not, we would next consider an ordinal logistic regression model. This model accounts for variables that are still categorized but have more than two responses with a particular ordering. For example, we would like to see how outcomes of a stop relate to race. Traffic stop outcomes includes arrest, citation, warning, and no action. These can be ordered by severity, e.g., an arrest is more severe than a citation and so on. Preliminary analysis indicates that African Americans are cited less than other drivers although they often searched at higher rates.

Also, our models have not taken into account interaction between variables. This phenomenon occurs when one explanatory variable effects another explanatory variable which in turn effects the response variable. An example of interacting variables may be geography and race. From the demographic breakdown of the policing districts, we know that the racial composition of the neighborhood changes depending on the location. Therefore, if a certain outcome impacts one race more than another, in turn it impacts the areas that contain a high percentage of that particular race. The regression models included above do not simultaneously include race and policing district as explanatory variables. In order to account for interaction, a different type of regression model known as log-linear may be used. However, the variables race and policing district may also raise issues of multi-collinearity.

8 Appendix

Variable Descriptions for Logistic Regression Output in Figure 10

Race of Driver
RACE(1): Asian
RACE(2): African American
RACE(3): Hispanic
RACE(4): Other
BASELINE: White

Sex
Female = 1
Male = 0

Age of Driver
AGERNG(1): 1-17 years
AGERNG(2): 18-25 years
AGERNG(3): 26-35 years
AGERNG(4): 36-45 years
AGERNG(5): 46-55 years
BASELINE: 55 years and up

Reason for Stop
MOVREG: Moving violation = 1
Registration/equipment violation = 0

twlvhrblck
Night (6pm-6am) = 1
Day (6am-6pm) = 0

Constant
The value of β_0 in the regression model.

Figure 10: Logistic Regression: Search Conducted by Policing District (Sig. = 0.000 implies P-value < 0.001)

Policing District 1- Central Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	RACE			556.732	4	.000			
	RACE(1)	-.539	.150	12.974	1	.000	.583	.435	.782
	RACE(2)	1.510	.089	285.533	1	.000	4.525	3.798	5.391
	RACE(3)	1.410	.081	300.854	1	.000	4.098	3.494	4.806
	RACE(4)	-.383	.292	1.715	1	.190	.682	.384	1.210
	SEX(1)	-.750	.062	148.016	1	.000	.472	.419	.533
	AGERNG			280.918	5	.000			
	AGERNG(1)	2.526	.268	88.535	1	.000	12.507	7.389	21.168
	AGERNG(2)	1.739	.166	109.963	1	.000	5.694	4.114	7.881
	AGERNG(3)	1.502	.165	83.109	1	.000	4.489	3.250	6.199
	AGERNG(4)	1.086	.167	42.082	1	.000	2.963	2.134	4.113
	AGERNG(5)	.779	.176	19.588	1	.000	2.180	1.544	3.078
	MOVREG	-.141	.046	9.327	1	.002	.869	.794	.951
	twlvhrblck	.587	.046	159.164	1	.000	1.798	1.641	1.969
	Constant	-4.442	.178	624.144	1	.000	.012		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 2- Rampart Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	RACE			520.447	4	.000			
	RACE(1)	-.435	.113	14.727	1	.000	.647	.519	.808
	RACE(2)	.981	.098	99.928	1	.000	2.666	2.200	3.232
	RACE(3)	1.168	.079	216.387	1	.000	3.215	2.752	3.757
	RACE(4)	.276	.225	1.506	1	.220	1.318	.848	2.048
	SEX(1)	-.866	.063	189.717	1	.000	.420	.372	.476
	AGERNG			285.614	5	.000			
	AGERNG(1)	1.901	.216	77.729	1	.000	6.696	4.387	10.218
	AGERNG(2)	1.238	.139	79.825	1	.000	3.448	2.628	4.523
	AGERNG(3)	.982	.138	50.458	1	.000	2.669	2.036	3.499
	AGERNG(4)	.615	.142	18.795	1	.000	1.850	1.401	2.443
	AGERNG(5)	.217	.156	1.939	1	.164	1.242	.915	1.685
	MOVREG	-.186	.041	20.422	1	.000	.830	.766	.900
	twlvhrblck	.453	.042	117.615	1	.000	1.573	1.449	1.707
	Constant	-3.069	.155	393.694	1	.000	.046		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 3- Southwest Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	RACE			298.755	4	.000			
	RACE(1)	-1.429	.400	12.749	1	.000	.240	.109	.525
	RACE(2)	1.755	.126	194.429	1	.000	5.781	4.517	7.398
	RACE(3)	1.580	.127	153.986	1	.000	4.855	3.783	6.231
	RACE(4)	.067	.365	.034	1	.853	1.070	.523	2.188
	SEX(1)	-1.228	.042	866.458	1	.000	.293	.270	.318
	AGERNG			590.383	5	.000			
	AGERNG(1)	2.479	.141	310.404	1	.000	11.935	9.058	15.726
	AGERNG(2)	1.639	.112	215.398	1	.000	5.148	4.136	6.407
	AGERNG(3)	1.332	.112	140.530	1	.000	3.790	3.041	4.724
	AGERNG(4)	1.136	.115	98.430	1	.000	3.115	2.489	3.899
	AGERNG(5)	.791	.121	42.633	1	.000	2.206	1.740	2.797
	MOVREG	-.243	.031	61.237	1	.000	.784	.738	.834
	twlvhrblck	.380	.031	152.764	1	.000	1.462	1.376	1.553
	Constant	-4.450	.166	715.855	1	.000	.012		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 4- Hollenbeck Area

Variables in the Equation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	RACE			178.832	4	.000			
	RACE(1)	-.664	.213	9.766	1	.002	.515	.339	.781
	RACE(2)	.423	.182	5.374	1	.020	1.526	1.067	2.181
	RACE(3)	1.021	.123	69.158	1	.000	2.776	2.182	3.532
	RACE(4)	.196	.379	.268	1	.605	1.217	.579	2.555
	SEX(1)	-1.001	.054	339.583	1	.000	.368	.330	.409
	AGERNG			512.254	5	.000			
	AGERNG(1)	3.017	.197	235.186	1	.000	20.438	13.898	30.055
	AGERNG(2)	1.958	.167	137.148	1	.000	7.088	5.107	9.838
	AGERNG(3)	1.713	.168	104.295	1	.000	5.547	3.993	7.707
	AGERNG(4)	1.316	.171	59.041	1	.000	3.729	2.665	5.216
	AGERNG(5)	.872	.182	23.059	1	.000	2.393	1.676	3.416
	MOVREG	-.381	.038	97.748	1	.000	.683	.634	.737
	twlvhrblck	.335	.038	78.070	1	.000	1.398	1.298	1.505
	Constant	-3.654	.205	317.252	1	.000	.026		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 5- Harbor Area

Variables in the Equation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	RACE			203.549	4	.000			
	RACE(1)	-1.307	.211	38.467	1	.000	.271	.179	.409
	RACE(2)	.597	.077	59.981	1	.000	1.816	1.561	2.112
	RACE(3)	.611	.055	123.159	1	.000	1.843	1.654	2.053
	RACE(4)	.274	.234	1.375	1	.241	1.315	.832	2.079
	SEX(1)	-1.064	.059	329.142	1	.000	.345	.308	.387
	AGERNG			172.137	5	.000			
	AGERNG(1)	1.911	.189	101.860	1	.000	6.760	4.664	9.797
	AGERNG(2)	1.421	.153	86.683	1	.000	4.140	3.070	5.584
	AGERNG(3)	1.327	.153	75.463	1	.000	3.769	2.794	5.084
	AGERNG(4)	1.190	.155	59.261	1	.000	3.286	2.427	4.449
	AGERNG(5)	.757	.165	20.998	1	.000	2.132	1.542	2.947
	MOVREG	-.345	.041	70.648	1	.000	.708	.654	.768
	twlvhrblck	.842	.041	426.998	1	.000	2.321	2.143	2.514
	Constant	-3.210	.155	427.953	1	.000	.040		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 6- Hollywood Area

Variables in the Equation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	RACE			1229.046	4	.000			
	RACE(1)	.072	.092	.610	1	.435	1.074	.898	1.285
	RACE(2)	.995	.054	335.253	1	.000	2.704	2.431	3.008
	RACE(3)	1.391	.041	1126.891	1	.000	4.017	3.704	4.357
	RACE(4)	.464	.135	11.742	1	.001	1.591	1.220	2.075
	SEX(1)	-.895	.050	323.280	1	.000	.409	.371	.450
	AGERNG			268.255	5	.000			
	AGERNG(1)	2.271	.230	97.245	1	.000	9.688	6.169	15.214
	AGERNG(2)	1.693	.180	88.018	1	.000	5.437	3.817	7.745
	AGERNG(3)	1.552	.180	74.387	1	.000	4.721	3.318	6.718
	AGERNG(4)	1.240	.183	46.134	1	.000	3.456	2.417	4.944
	AGERNG(5)	.711	.196	13.214	1	.000	2.036	1.388	2.987
	MOVREG	-.049	.035	1.924	1	.165	.952	.888	1.021
	twlvhrblck	.269	.037	53.912	1	.000	1.309	1.218	1.406
	Constant	-4.093	.181	511.110	1	.000	.017		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 7- Wilshire Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1a	RACE			1251.976	4	.000			
	RACE(1)	.200	.086	5.374	1	.020	1.222	1.031	1.448
	RACE(2)	1.707	.067	652.850	1	.000	5.514	4.837	6.285
	RACE(3)	1.767	.065	740.640	1	.000	5.851	5.152	6.644
	RACE(4)	.391	.203	3.702	1	.054	1.478	.993	2.201
	SEX(1)	-1.228	.051	574.067	1	.000	.293	.265	.324
	AGERNG			450.799	5	.000			
	AGERNG(1)	2.314	.181	164.143	1	.000	10.118	7.101	14.417
	AGERNG(2)	1.716	.137	157.726	1	.000	5.561	4.255	7.269
	AGERNG(3)	1.383	.136	102.675	1	.000	3.987	3.051	5.210
	AGERNG(4)	1.066	.139	58.712	1	.000	2.905	2.211	3.816
	AGERNG(5)	.808	.147	30.249	1	.000	2.243	1.682	2.991
	MOVREG	-.283	.034	69.537	1	.000	.753	.705	.805
	twlvhrblck	.706	.034	428.826	1	.000	2.026	1.895	2.166
	Constant	-4.400	.147	897.459	1	.000	.012		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 8- West Los Angeles Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1a	RACE			861.986	4	.000			
	RACE(1)	-.416	.160	6.731	1	.009	.660	.482	.903
	RACE(2)	1.456	.082	314.722	1	.000	4.290	3.652	5.038
	RACE(3)	1.783	.067	705.783	1	.000	5.948	5.215	6.784
	RACE(4)	.286	.177	2.619	1	.106	1.332	.941	1.884
	SEX(1)	-1.122	.080	195.218	1	.000	.326	.278	.381
	AGERNG			197.721	5	.000			
	AGERNG(1)	1.494	.264	31.945	1	.000	4.453	2.653	7.475
	AGERNG(2)	1.384	.181	58.240	1	.000	3.990	2.796	5.692
	AGERNG(3)	.966	.182	28.234	1	.000	2.627	1.840	3.751
	AGERNG(4)	.616	.188	10.803	1	.001	1.852	1.282	2.675
	AGERNG(5)	.109	.209	.272	1	.602	1.115	.740	1.681
	MOVREG	-.554	.057	95.477	1	.000	.575	.514	.642
	twlvhrblck	1.001	.057	304.061	1	.000	2.721	2.431	3.045
	Constant	-4.577	.182	634.867	1	.000	.010		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 9- Van Nuys Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1a	RACE			990.351	4	.000			
	RACE(1)	-.471	.142	11.033	1	.001	.624	.473	.824
	RACE(2)	.968	.067	207.170	1	.000	2.633	2.308	3.004
	RACE(3)	1.306	.045	836.375	1	.000	3.690	3.377	4.031
	RACE(4)	-.130	.158	.675	1	.411	.878	.645	1.197
	SEX(1)	-1.065	.052	420.795	1	.000	.345	.311	.382
	AGERNG			341.481	5	.000			
	AGERNG(1)	1.905	.180	112.462	1	.000	6.719	4.725	9.555
	AGERNG(2)	1.554	.144	115.926	1	.000	4.732	3.566	6.279
	AGERNG(3)	1.326	.144	84.303	1	.000	3.765	2.837	4.997
	AGERNG(4)	1.007	.147	46.911	1	.000	2.737	2.052	3.650
	AGERNG(5)	.535	.158	11.502	1	.001	1.708	1.253	2.327
	MOVREG	-.753	.038	388.469	1	.000	.471	.437	.508
	twlvhrblck	1.178	.037	1016.506	1	.000	3.248	3.021	3.492
	Constant	-3.914	.146	714.113	1	.000	.020		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 10- West Valley Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			1121.138	4	.000			
	RACE(1)	-.641	.126	25.735	1	.000	.527	.411	.675
	RACE(2)	.598	.067	80.865	1	.000	1.819	1.597	2.073
	RACE(3)	1.211	.039	952.148	1	.000	3.358	3.109	3.626
	RACE(4)	-.198	.128	2.404	1	.121	.820	.639	1.054
	SEX(1)	-.975	.048	411.816	1	.000	.377	.343	.414
	AGERNG			287.607	5	.000			
	AGERNG(1)	1.601	.157	103.932	1	.000	4.956	3.643	6.742
	AGERNG(2)	1.492	.124	144.615	1	.000	4.447	3.487	5.672
	AGERNG(3)	1.435	.125	132.575	1	.000	4.200	3.290	5.362
	AGERNG(4)	1.075	.127	71.305	1	.000	2.930	2.283	3.761
	AGERNG(5)	.698	.137	25.921	1	.000	2.010	1.536	2.629
	MOVREG	-.873	.036	577.318	1	.000	.418	.389	.449
	twlvhrblck	.696	.036	369.196	1	.000	2.006	1.869	2.154
	Constant	-3.287	.124	698.497	1	.000	.037		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 11- Northeast Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			684.931	4	.000			
	RACE(1)	-.141	.133	1.113	1	.291	.869	.669	1.128
	RACE(2)	.954	.111	73.408	1	.000	2.596	2.087	3.229
	RACE(3)	1.453	.064	521.897	1	.000	4.276	3.775	4.844
	RACE(4)	-.193	.231	.695	1	.404	.825	.524	1.297
	SEX(1)	-.842	.062	186.832	1	.000	.431	.382	.486
	AGERNG			250.854	5	.000			
	AGERNG(1)	2.336	.201	135.622	1	.000	10.342	6.980	15.324
	AGERNG(2)	1.434	.164	76.058	1	.000	4.197	3.041	5.794
	AGERNG(3)	1.214	.165	54.190	1	.000	3.367	2.437	4.652
	AGERNG(4)	.916	.169	29.357	1	.000	2.499	1.794	3.481
	AGERNG(5)	.684	.181	14.348	1	.000	1.982	1.391	2.823
	MOVREG	-.223	.044	25.118	1	.000	.800	.733	.873
	twlvhrblck	.490	.045	119.101	1	.000	1.632	1.495	1.782
	Constant	-4.146	.172	580.646	1	.000	.016		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 12- 77th Street Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			73.237	4	.000			
	RACE(1)	-.548	.362	2.288	1	.130	.578	.284	1.176
	RACE(2)	.944	.151	38.924	1	.000	2.571	1.911	3.459
	RACE(3)	.845	.152	30.761	1	.000	2.329	1.727	3.139
	RACE(4)	.064	.364	.030	1	.862	1.066	.522	2.176
	SEX(1)	-1.235	.041	919.096	1	.000	.291	.268	.315
	AGERNG			603.677	5	.000			
	AGERNG(1)	2.065	.126	267.142	1	.000	7.886	6.156	10.102
	AGERNG(2)	1.390	.098	201.408	1	.000	4.013	3.312	4.862
	AGERNG(3)	1.102	.098	125.406	1	.000	3.011	2.483	3.652
	AGERNG(4)	.814	.101	65.347	1	.000	2.258	1.853	2.751
	AGERNG(5)	.589	.108	29.801	1	.000	1.802	1.459	2.226
	MOVREG	-.362	.028	168.783	1	.000	.696	.659	.735
	twlvhrblck	.585	.028	438.014	1	.000	1.795	1.700	1.897
	Constant	-2.922	.177	272.028	1	.000	.054		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 13- Newton Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			203.695	4	.000			
	RACE(1)	-.803	.228	12.372	1	.000	.448	.286	.701
	RACE(2)	1.179	.123	92.349	1	.000	3.250	2.555	4.132
	RACE(3)	1.055	.121	76.544	1	.000	2.873	2.268	3.639
	RACE(4)	-1.882	.725	6.739	1	.009	.152	.037	.631
	SEX(1)	-1.098	.048	525.146	1	.000	.333	.304	.366
	AGERNG			701.131	5	.000			
	AGERNG(1)	2.693	.151	317.196	1	.000	14.774	10.985	19.870
	AGERNG(2)	1.889	.126	226.190	1	.000	6.615	5.171	8.461
	AGERNG(3)	1.567	.126	154.601	1	.000	4.794	3.744	6.137
	AGERNG(4)	1.237	.128	93.026	1	.000	3.444	2.679	4.428
	AGERNG(5)	.933	.135	47.716	1	.000	2.543	1.952	3.315
	MOVREG	-.334	.030	123.500	1	.000	.716	.675	.760
	twlvhrblck	.280	.030	86.917	1	.000	1.323	1.247	1.403
	Constant	-3.317	.171	377.992	1	.000	.036		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 14- Pacific Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			883.053	4	.000			
	RACE(1)	-.986	.110	79.815	1	.000	.373	.300	.463
	RACE(2)	1.031	.052	397.638	1	.000	2.803	2.533	3.102
	RACE(3)	1.043	.046	517.208	1	.000	2.838	2.594	3.105
	RACE(4)	-.034	.168	.040	1	.841	.967	.695	1.344
	SEX(1)	-.917	.049	345.786	1	.000	.400	.363	.440
	AGERNG			386.511	5	.000			
	AGERNG(1)	2.366	.180	173.453	1	.000	10.651	7.490	15.146
	AGERNG(2)	1.540	.124	154.994	1	.000	4.663	3.660	5.943
	AGERNG(3)	1.214	.123	96.740	1	.000	3.367	2.643	4.288
	AGERNG(4)	.984	.126	61.025	1	.000	2.675	2.090	3.424
	AGERNG(5)	.615	.135	20.677	1	.000	1.849	1.419	2.410
	MOVREG	.290	.038	58.073	1	.000	1.337	1.241	1.440
	twlvhrblck	.745	.038	382.710	1	.000	2.106	1.954	2.269
	Constant	-5.434	.124	1905.833	1	.000	.004		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 15- North Hollywood Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			666.400	4	.000			
	RACE(1)	-.427	.166	6.623	1	.010	.653	.472	.903
	RACE(2)	.758	.080	89.255	1	.000	2.133	1.823	2.496
	RACE(3)	1.167	.048	591.920	1	.000	3.211	2.923	3.527
	RACE(4)	.263	.158	2.778	1	.096	1.300	.955	1.771
	SEX(1)	-1.094	.060	334.547	1	.000	.335	.298	.377
	AGERNG			325.968	5	.000			
	AGERNG(1)	2.246	.195	132.856	1	.000	9.453	6.452	13.851
	AGERNG(2)	1.506	.163	84.857	1	.000	4.508	3.272	6.210
	AGERNG(3)	1.190	.164	52.866	1	.000	3.288	2.385	4.531
	AGERNG(4)	.965	.167	33.475	1	.000	2.624	1.893	3.639
	AGERNG(5)	.428	.182	5.545	1	.019	1.534	1.074	2.189
	MOVREG	-.432	.041	111.160	1	.000	.649	.599	.703
	twlvhrblck	.767	.041	343.673	1	.000	2.152	1.985	2.334
	Constant	-3.911	.166	556.101	1	.000	.020		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 16- Foothill Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			427.100	4	.000			
	RACE(1)	-1.143	.203	31.757	1	.000	.319	.214	.474
	RACE(2)	.936	.080	137.433	1	.000	2.551	2.181	2.983
	RACE(3)	.855	.049	301.348	1	.000	2.352	2.135	2.590
	RACE(4)	-.520	.242	4.631	1	.031	.594	.370	.955
	SEX(1)	-1.106	.053	435.133	1	.000	.331	.298	.367
	AGERNG			214.233	5	.000			
	AGERNG(1)	1.751	.172	103.843	1	.000	5.761	4.114	8.068
	AGERNG(2)	1.171	.144	65.846	1	.000	3.225	2.431	4.279
	AGERNG(3)	1.055	.145	53.103	1	.000	2.872	2.163	3.814
	AGERNG(4)	.873	.147	35.356	1	.000	2.394	1.795	3.192
	AGERNG(5)	.422	.158	7.181	1	.007	1.525	1.120	2.077
	MOVREG	-.626	.037	280.462	1	.000	.535	.497	.575
	twlvhrblck	.947	.037	645.486	1	.000	2.577	2.396	2.772
	Constant	-2.942	.148	396.095	1	.000	.053		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 17- Devonshire Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			695.002	4	.000			
	RACE(1)	-.859	.117	53.742	1	.000	.423	.337	.533
	RACE(2)	.417	.070	35.799	1	.000	1.518	1.324	1.740
	RACE(3)	.930	.042	500.303	1	.000	2.535	2.337	2.751
	RACE(4)	-.185	.166	1.238	1	.266	.831	.600	1.151
	SEX(1)	-1.083	.052	439.245	1	.000	.339	.306	.375
	AGERNG			159.692	5	.000			
	AGERNG(1)	1.712	.178	92.107	1	.000	5.540	3.905	7.858
	AGERNG(2)	1.357	.148	83.793	1	.000	3.885	2.905	5.195
	AGERNG(3)	1.339	.149	80.759	1	.000	3.814	2.848	5.107
	AGERNG(4)	1.253	.151	69.158	1	.000	3.502	2.606	4.705
	AGERNG(5)	.747	.161	21.555	1	.000	2.111	1.540	2.895
	MOVREG	-.844	.037	518.919	1	.000	.430	.400	.462
	twlvhrblck	.735	.037	404.736	1	.000	2.086	1.942	2.241
	Constant	-3.263	.149	479.322	1	.000	.038		

a. Variable(s) entered on step 1: twlvhrblck.

Policing District 18- Southeast Area

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	RACE			69.783	4	.000			
	RACE(1)	-1.264	.449	7.940	1	.005	.282	.117	.681
	RACE(2)	.914	.163	31.480	1	.000	2.495	1.813	3.434
	RACE(3)	.792	.164	23.206	1	.000	2.208	1.600	3.047
	RACE(4)	.013	.479	.001	1	.978	1.013	.396	2.592
	SEX(1)	-1.301	.051	657.756	1	.000	.272	.247	.301
	AGERNG			228.318	5	.000			
	AGERNG(1)	1.476	.151	95.276	1	.000	4.375	3.253	5.883
	AGERNG(2)	.993	.121	67.004	1	.000	2.700	2.129	3.425
	AGERNG(3)	.801	.122	43.062	1	.000	2.227	1.754	2.829
	AGERNG(4)	.539	.125	18.607	1	.000	1.714	1.342	2.190
	AGERNG(5)	.419	.132	10.049	1	.002	1.521	1.174	1.972
	MOVREG	.048	.034	2.065	1	.151	1.049	.983	1.121
	twlvhrblck	.670	.034	398.703	1	.000	1.954	1.829	2.086
	Constant	-2.782	.198	196.658	1	.000	.062		

a. Variable(s) entered on step 1: twlvhrblck.

9 Acknowledgements

We would especially like to thank our research advisor, Dr. Lily Khadjavi, for her direction, devotion, and support for the duration of this project. We would also like to thank Ms. Laura Smith for her assistance with research, computer programming knowledge, and overall contribution to our project. We would like to thank Dr. Erika Camacho, Dr. Angela Gallegos, Dr. Edward Mosteig, Dr. Stephen Wirkus, and David Uminsky for their helpful conversations. Furthermore, we greatly appreciate the assistance of Dr. Kim Weems from North Carolina State University, for helping us with the analysis of logistic regression.

This research was conducted at the Applied Mathematical Sciences Summer Institute (AMSSI) and has been partially supported by grants given by the Department of Defense (through its ASSURE program), the National Science Foundation (DMS-0453602), the National Security Agency (MSPF-06IC-022). Substantial financial and moral support was also provided by Don Straney, Dean of the College of Science at California State Polytechnic University, Pomona. Additional financial and moral support was provided by the Department of Mathematics at Loyola Marymount University and the Department of Mathematics & Statistics at California State Polytechnic University, Pomona.The authors are solely responsible for the views and opinions expressed in this research; it does not necessarily reflect the ideas and/or opinions of the funding agencies and/or LMU or Cal Poly, Pomona.

References

[Ag] Agresti, Alan, and Barbara Finlay. Statistical Methods for Social Sciences. New Jersey: Prentice Hall, 1997.

[An] Antonovics, Kate, Knight, Brian. (2004). A New Look at Racial Profiling: Evidence from the Boston Police Department. NBER Working Papers 10634, *National Bureau of Economic Research, Inc.*

[Ch] Chatterjee, Samprit, Hadi, Ali S., Price, Bertram. Regression Analysis by Example. 3rd ed. New York: Wiley-Interscience, 2000.

[Cl] Clayton, Jerry, Lamberth Karl, John Lamberth, Amy Farrel, and Jack McDevitt. Learning From Research and Practice. Lamberth Consulting, 2005.

[Du] Durose, Matthew, Smith, Erika, Langan, Patrick. Contacts Between Police and the Public, 2005. Bureau of Justice Statistics Special Report, April 2007.

[Ge] Gelman, Andrew, Fagan, Jeffrey, and Kiss, Alex. An Analysis of the NYPD's stop-and frisk policy in the context of claims of racial bias. *Journal of the American Statistical Association*, to appear.

[Kh] Khadjavi, Lily S. "Driving While Black in the City of Angels." *Chance* Vol. 19 No. 2(2006): 43-46.

[La] Lamberth, John. "Driving While Black; A Statistician Proves That Prejudice Still Rules the Road." Washington Post 16 Aug. 1998: c1.

[LAPD] "Motor Vehicle and Pedestrian Stop Field Data Report Data Collection Training, Lesson Plan." Los Angeles Police Department.

[Ra] Ramirez, Deborah, McDevitt, Jack, and Farrell, Amy (2000). *A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned.* Technical Report, U.S. Department of Justice, Washington, DC.

[Ri] Rice, Constance *et al.* Rampart Reconsidered: The Search for Real Reform Seven Years Later. Retrieved 3 Aug. 2007 <www.lapdonline.org/home/content_basic_view/32828>.

[St] Steward, Dwight and Berg, M. Douglas. (2000) A Statistical examination of Racial Profiling *Steward Research Group, Inc.*.

[Wa] Wackerly, Dennis D., Wiliam Mendenhall III, and Richard L. Scheaffer. Mathematical Statistics with Applications. 6th ed. Pacific Grove: Duxbury, 2002.

[Zp] "Zip Code Data Book." United Way LA. 2007. United Way. Retrieved 10 July 2007 ¡www.unitedwayla.org¿.