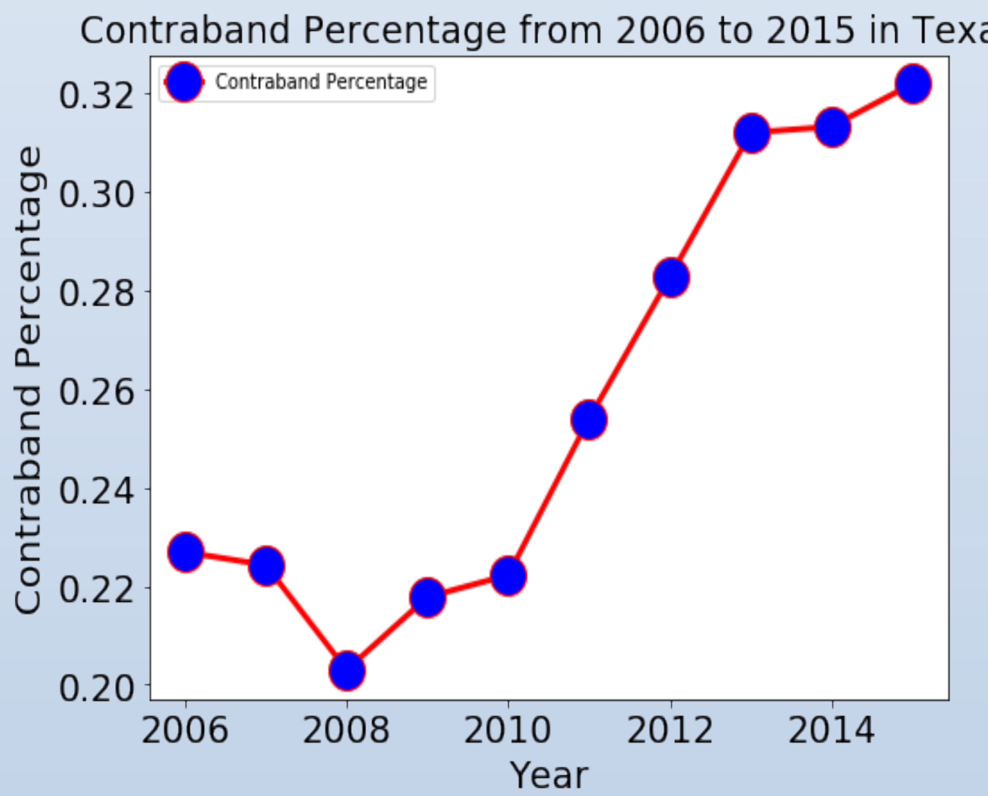
Taking Out the Guesswork: An Analytical Approach to Police Traffic Stop Contraband Searches

Tong Zhou
Anderson Summer Science Program/Data Analytics Department
Research Advisor Dr. Anthony Joseph Bonifonte

Introduction

It is very common for the police to stop a vehicle in US. However, it is relatively rare for the police to search a vehicle for contraband. Our historical data shows that only around 2% of the stopped vehicle will be searched.



Our research builds a model to predict contraband probability which can not only help police increase the chance of finding contraband in searched vehicles but also can decrease biases and social disparities among different races and gender.

Scholarly Approach/Methods

We created a logistic regression model to predict contraband in searched vehicles. Here are specific steps to build the model:

- Consider only consent searches
- Remove rows when "other" and "other (non map)" is the only violation reason
  (This mean that the police find something suspicious)
- Translate categorical variables into dummy variables
- Split training data from 2006 to 2012 and test data from 2013 to 2015
- Standardize continuous variables
- Build the logistic model

Variables we tried but not significant

- Months
- Season
- Weapon Arrests per capita by county (Replaced Outliers)
- Drug Arrests per capita by county(Replaced Outliers)

Results

Category	Variables	Coefficient	Odds ratio
	Intercept	-1.922	
Reason for Stop	DUI	0.223	1.250
	Speeding	-0.093	0.911
	Lights	-0.050	0.951
	Paperwork	-0.086	0.918
	Equipment	-0.125	0.882
	Stop sign	0.029	1.030
	Safe movement	-0.071	0.931
	Seat belt	-0.031	0.969
	License	-0.087	0.917
	Registration	-0.048	0.953
Historical Contraband Find % in the county	Historical contraband %	0.504	1.655
Time of Day	Morning (Baseline)	0	1
	Afternoon	0.125	1.133
	Evening	0.195	1.215
	Late evening	0.091	1.096
Day of Week	Weekday (Baseline)	0	1
	Weekend	0.054	1.056

- We found that drivers will tend to drive carefully if they have contraband.
- Variable equipment has the most negative coefficient.
- All predictors have p value less than 0.001.

Figure 1. Example of the police search information through database. It is convenient for the police to use our model because they have all the data needed in the system.

Results

Driver Gender	Stops	Searches	Search Percentage	Contraband Found	Contraband Found Percentage
Female	1841385	5644	0.003	1039	0.184
Male	3758805	27941	0.017	4646	0.166

* Gender is not a variable when we built the linear regression model.

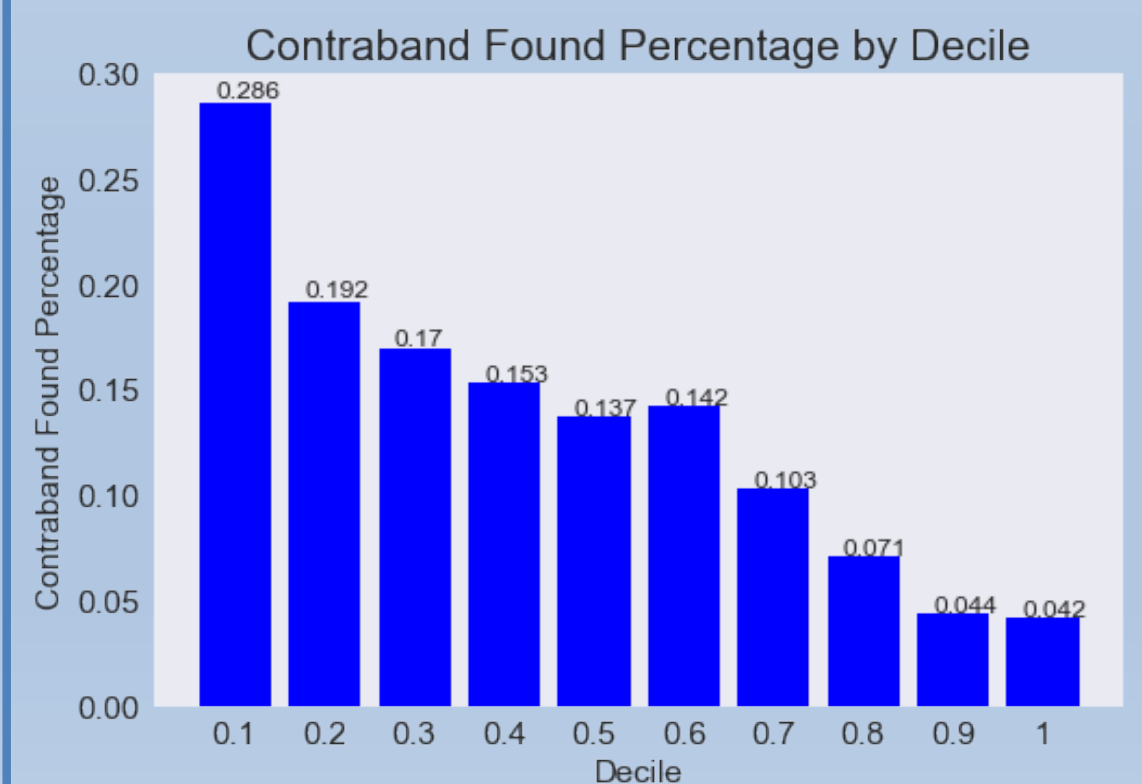
- Summary table above shows the police search and contraband found based on what we observed in the test set.

Driver Gender	Searches in Data	Model Searches	Search Percentage	Contraband Found	Contraband Found Percentage
Female	5644	817	0.145	309	0.378
Male	27941	3945	0.141	1306	0.331

- Summary table above shows the imaginary search and contraband found among test sets based on what our logistic regression model returns. (size of our test set is 33585)
- Similar demographics tables such as age and race shows the similar results.

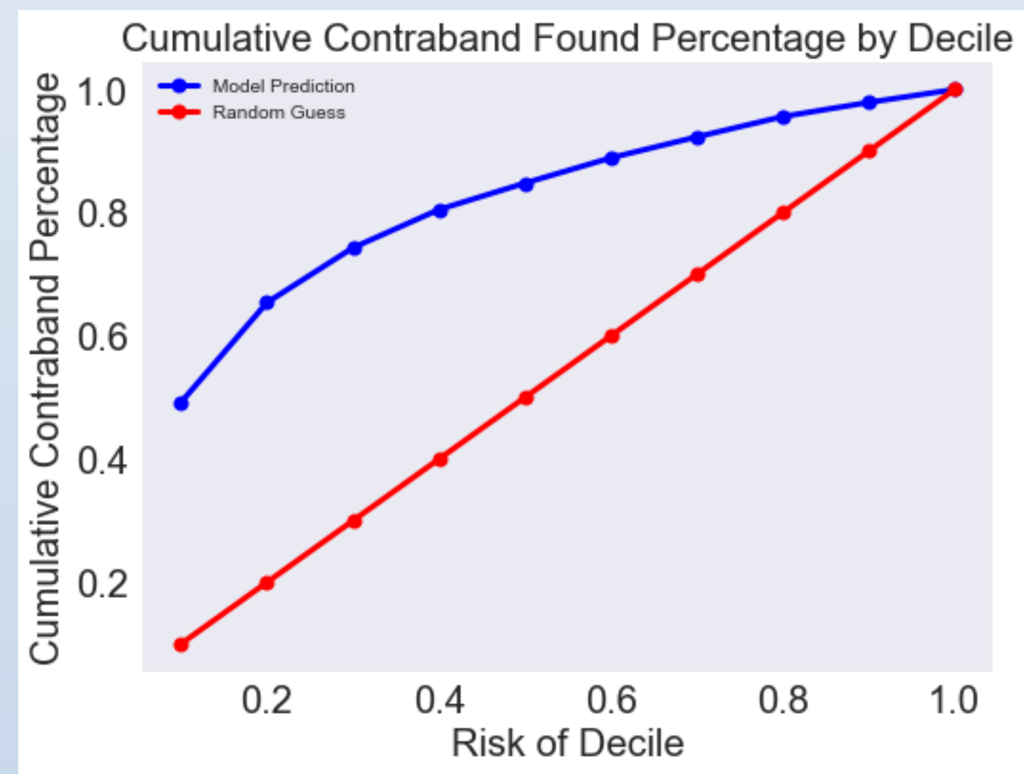
		Predicted	
Truth		No Contraband	Contraband
	No Contraband	24762	3147
	Contraband	4070	1615

- Confusion matrix table above shows the performance of our model on test data with the threshold 0.3.



- Figure above shows the decile of contraband probability and contraband found percentage.
- Accuracy of our model is 0.832.
- AUC score of our model is 0.691.

Results



- Figure above shows the relationship between risks and accumulative contraband found percentage both for our model and random guesses.

Next Steps

- A better way to analyze unsearched data and missing data.
- Find additional variables that may increase our predictive power.
- Build some additional machine learning models such as decision tree and neural networks and compare their performance.

Acknowledgements

This project was funded by the The William G. Bowen and Mary Ellen Bowen Endowed Fund at Denison University.

Works Cited

1 Epp, Charles R and Maynard-Moody, Steven and Haider-Markel, Donald P, Pulled over*: How police stops define race and citizenship*, 2014
2 E. Pierson, C. Simoiu, J. Overgoor, S. Corbett-Davies, V. Ramachandran, C. Phillips, S. Goel, *A large-scale analysis of racial disparities in police stops across the United States, 2017*
3 Susan Li, *Building A Logistic Regression in Python, Step by Step, 2017*
4 Quinton, Paul and Bland, Nick and Miller, Joel, Police stops, decision-making and practice, 2000