INF 2178H: Experimental Design for Data Science

“Model Minority Myth affecting the probability of being stripped searched”

Professor
Prof. Shion Guha

Authors
Yu Fan, and Kristina Foster

Introduction

The Toronto Police Service (TPS) Public Safety Data Portal houses a dataset called Arrests and Strip Searches (RBDC-ARR-TBL-001) that has social, cultural, and economic attributes associated with detainees as they were processed by the TPS institution. However the TPS has a proven history of violence and discrimination at least the last 30 years towards certain racial groups due to systemic and individuals' inherent bias (Tator et al., 2006, 2020) towards Black people. Another piece of literature that informs this is the "minority myth" (Tator et al., 2006, 2020), where certain groups ethnic minorities are considered "more polite" and less "uppity" within the systems of authority that rule our daily lives. This paper explores and critically evaluates the associations with perceived race as it pertains to the combination of the model minority myth (Yao, 2021) and the inherent bias towards certain races by proxy of not being in the category of a model minority and the explanatory effect on the probability of being strip searched. The structure of the analysis comprises normalizing discrete categorical variables by their marginal and joint frequencies between "Perceived Race" and "Sex", exploratory data analysis (EDA), T-tests, ANCOVA, and logistic regression. A discussion and conclusion that covers the correlations that arise from the statistical analysis of the data is at the end of the paper.

RQ1: Is there a significant difference in strip search based on perceived race?

RQ2: How does race affect the probability of being strip-searched when controlling for age?

RQ3: How does race affect the probability of being strip-searched when controlling for sex?

RQ4: Is there a significant difference in strip search between black&indigenous and other minority race?

Literature review

Strip search is a method used by the police to search a suspect thoroughly by removing the suspect's clothes and shoes, checking all items inside, and conducting a visual and physical examination of the body to determine if illegal items are hidden. It is only conducted when the police have legitimate reasons to suspect that a suspect may be carrying illegal items (Lu, 2019).

However, in recent years, there have been numerous instances of police officers requesting strip searches of suspects without corroborating evidence (Kadiri, 2022). For instance, the London Police Service strip-searched an underage girl on suspicion of possession of cannabis without evidence. Moreover, in the last five years, the Metropolitan Police have conducted nearly 10,000 strip searches of minors, including 2,360 searches of children under the age of 16 (Kadiri, 2022).

Additionally, research suggests that black girls are more likely to be subjected to the most intense strip searches than white girls (Thomas & Gidda, 2023). Across both types of strip-search for female children and teens up to the age of 19, 45% of people searched were Black (Thomas & Gidda, 2023). These investigations have raised many irregularities in the way police officers perform strip searches, which may have led to differential treatment by police for different races.

Toronto Police Service has a history of violence and discrimination (Tator et al., 2006) and has not seemed to improve in the last 20 years (2020) towards certain racial groups due to systemic and individuals' inherent bias. Inherent biases can have serious consequences such as police brutality, being profiled due to the arrestees' perceived races, and unfair discrimination towards different racial groups. In this paper, the EDA begins with several different control factors and the perceived race of the arrestee is the main explanatory variable regarding an increase in strip-searches for those who are not considered a model minority.

Exploratory Data Analysis

Overview

The possible correlated and/or explanatory attributes with discrete levels we are exploring are “Perceived Race,” which is what ethnicity the police person perceived when looking at the arrestee. “Sex” refers to an arrestee’s gender; labeled as female, male, or other. “Age” refers to several different ordinal buckets for the age of the individual arrested.

The outcome variable we are looking to explore is “Strip Search”; a binary attribute that indicates if the arrestee was strip-searched or not. After our initial review of the dataset, exploring the parameters, and removing duplicates, the dataset itself consists of 25 variables and 65267 unique arrests.

Cleaning

One of the first steps is determining if StripeSeach only contains ones and zeros. A value of 0 indicates that the suspect was not subjected to a strip-search, whereas a value of 1 indicates that the suspect was subjected to a strip-search. Race, and sex are the primary focus of our investigation.

We then found that the age column had similar data entered separately, for example 'Aged 17 years and younger' and 'Aged 17 years and under' actually described the same group. We therefore merged all data that were identical but recorded separately due to different descriptions.

Finally, we removed all rows that had “Unknown and Legacy” and “NaN” levels for the Perceived Race category.

Discrete Marginal Variables

To get a sense of the basic probability distributions for each nominal variable, we generated a probability mass function (PMF) for each category along with generating the joint distribution between each explanatory variable and the outcome variable “Strip Search”. **Figure 1**, **2**, and **3** displays the joint distributions with a heatmap that demonstrates the frequency of the joint conditional probability for each category.

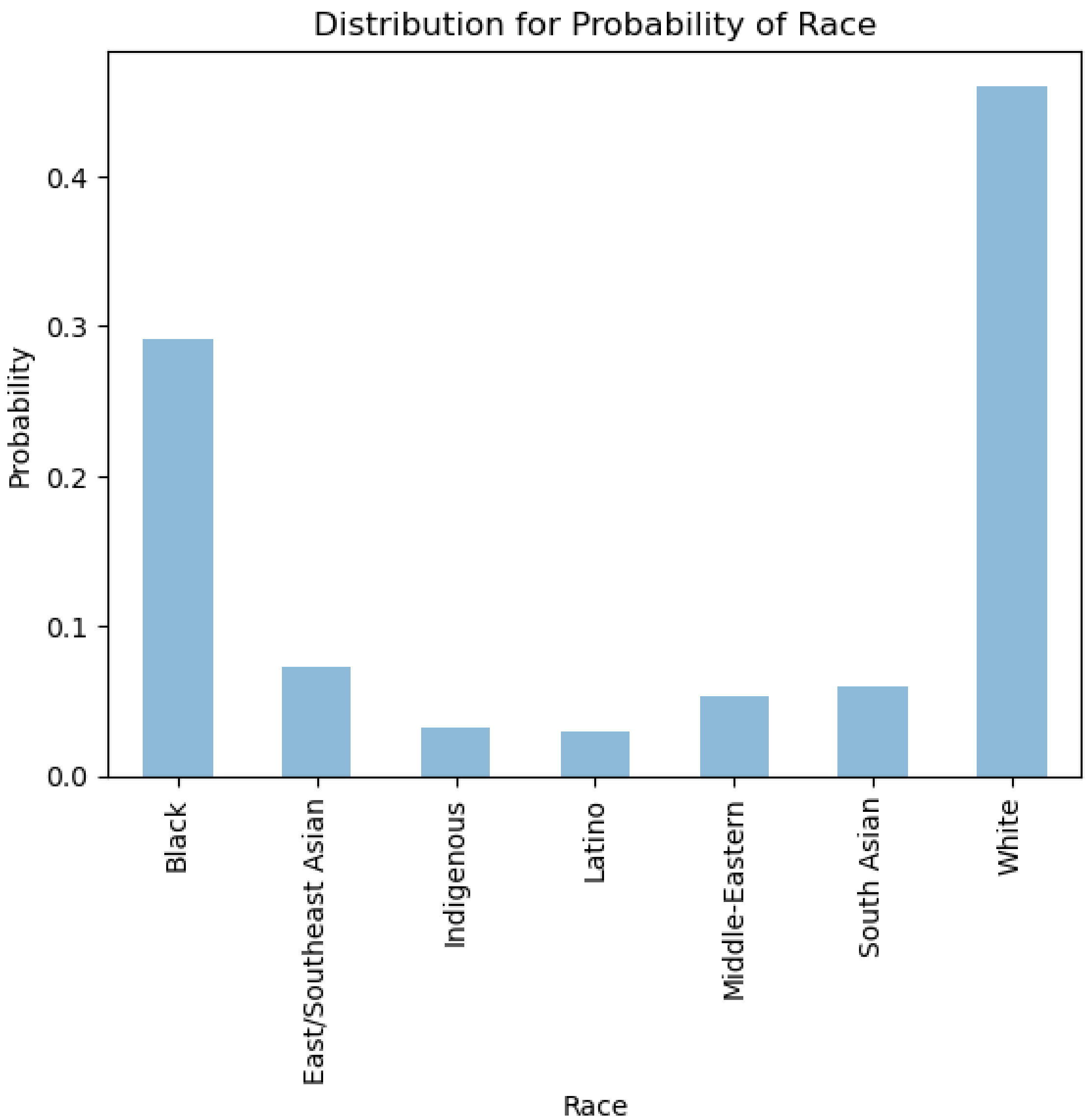

Figure 1: The distribution of different race being arrested

The figure 1 shows us the racial breakdown of arrests of suspects by Toronto police, and we can clearly see that whites and blacks are arrested in far higher numbers than others.

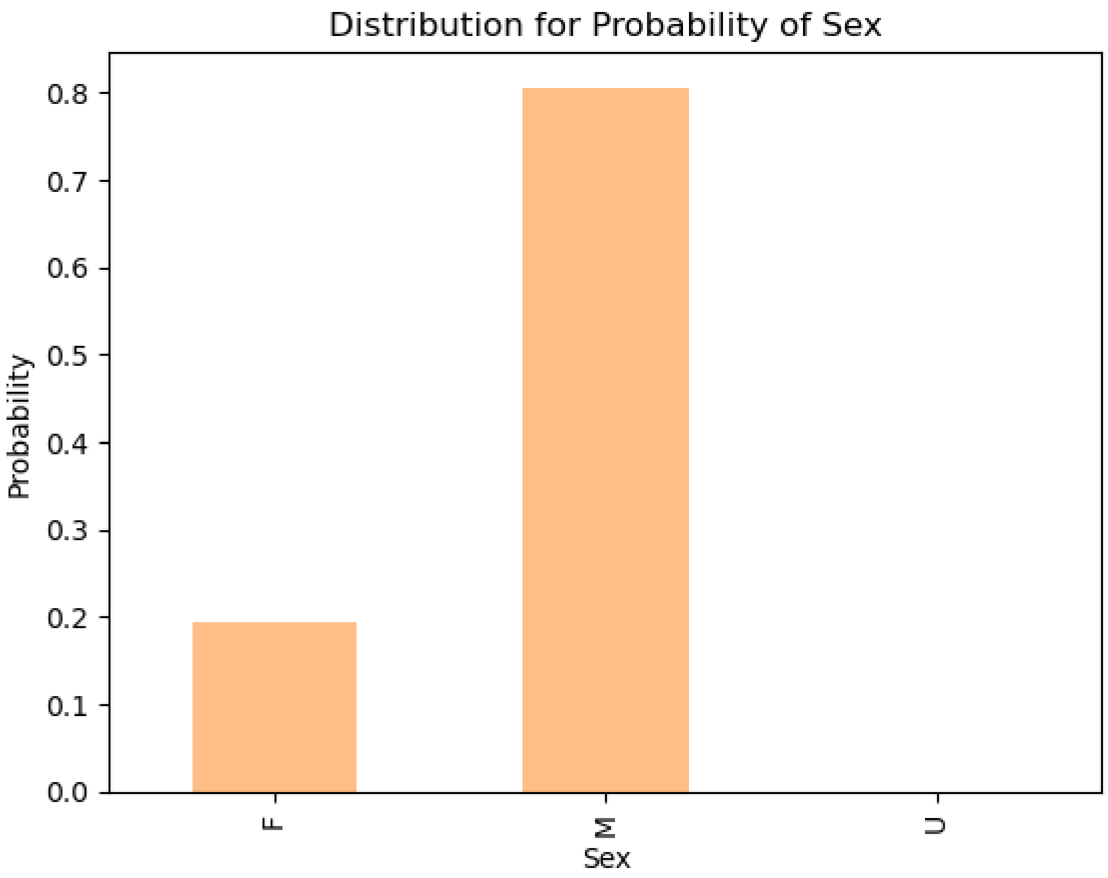

Figure 2: The distribution of different gender being arrested

The figure 2 shows us the gender breakdown of arrests of suspects by Toronto police. The number of male arrests is much larger than that of female, and at the same time makes us think, with the huge difference in the number of male and female arrests, will there be a difference in the probability of receiving a strip search by gender? We will explore the answer to this question in a follow-up study.

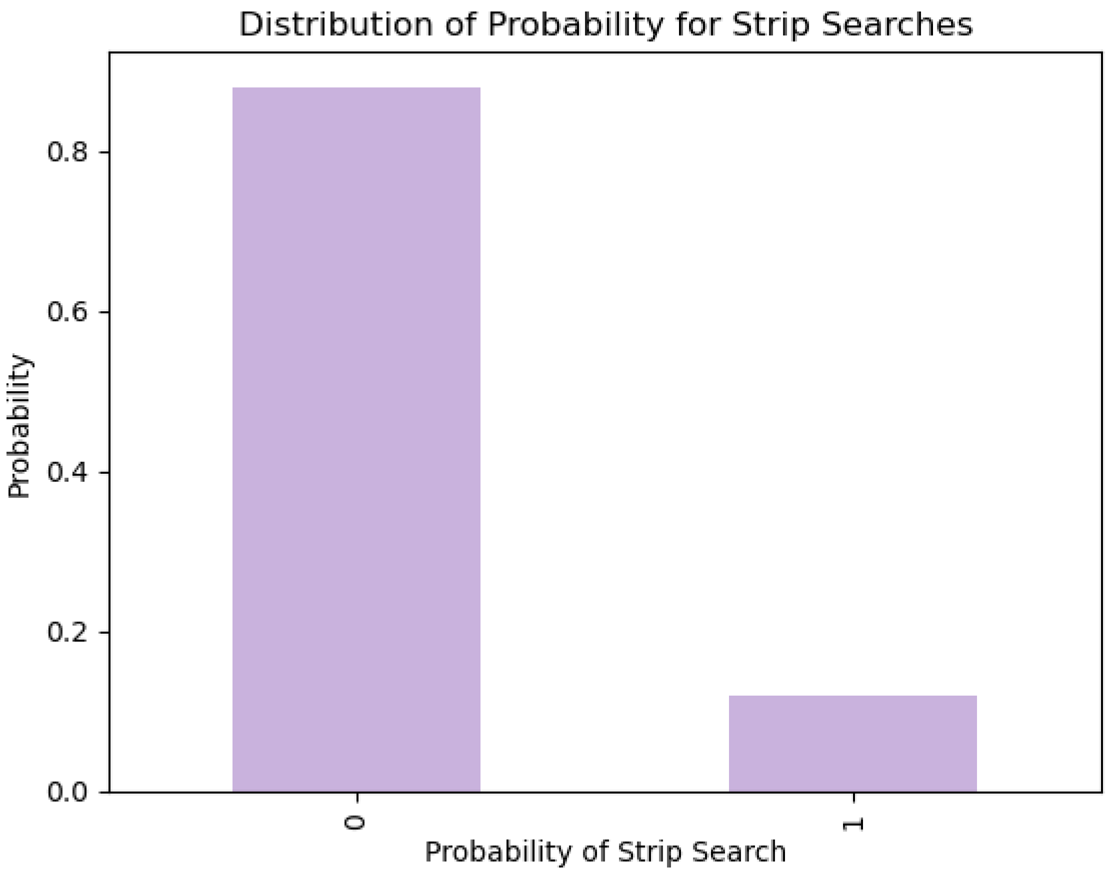

Figure 3: The distribution of whether being strip-searched

The figure 3 reveals for us the probability that Toronto police will take a strip search of a suspect. Although the probability of being strip-searched seems very low at less than 20 percent, our data contains a very large sample of more than 60,000 people, and less than 20 percent of people who are strip-searched is a very large number.

Power Calculation for Minimum Sample Required

A power analysis was conducted to determine the minimum sample size required for each group assessed for sample size. Results indicated the required sample size of each population exceeded the 80% power for detecting a large effect, at a significance criterion of α = .005, and the following population sizes were utilized for the methods used in the analysis section of this paper.

Level	Model Minority	Not Model Minority	White	All
Strip-searched	958	2740	3566	7269
Not strip-searched	12075	16720	24157	52955
All	13033	19460	27727	60220

Joint Distributions and Plots

To get a sense of the basic probability distributions for each nominal variable, we generated a probability mass function (PMF) for each category and then generated the joint distribution between each explanatory variable and the outcome variable “Strip Search”. Each calculated percentage is used to represent an estimate of the mean value for each of the combinations of the explanatory variable and the outcome variable; and the distribution count of each of those combinations acts as a continuous variable. **Figure 6**, **7**, and **8** display an estimate for the mean probability of each level of Perceived Race and being strip searched or not, the mean probability of each level for Sex and being strip searched or not, and finally the mean probability for each level of Age and being strip searched or not.

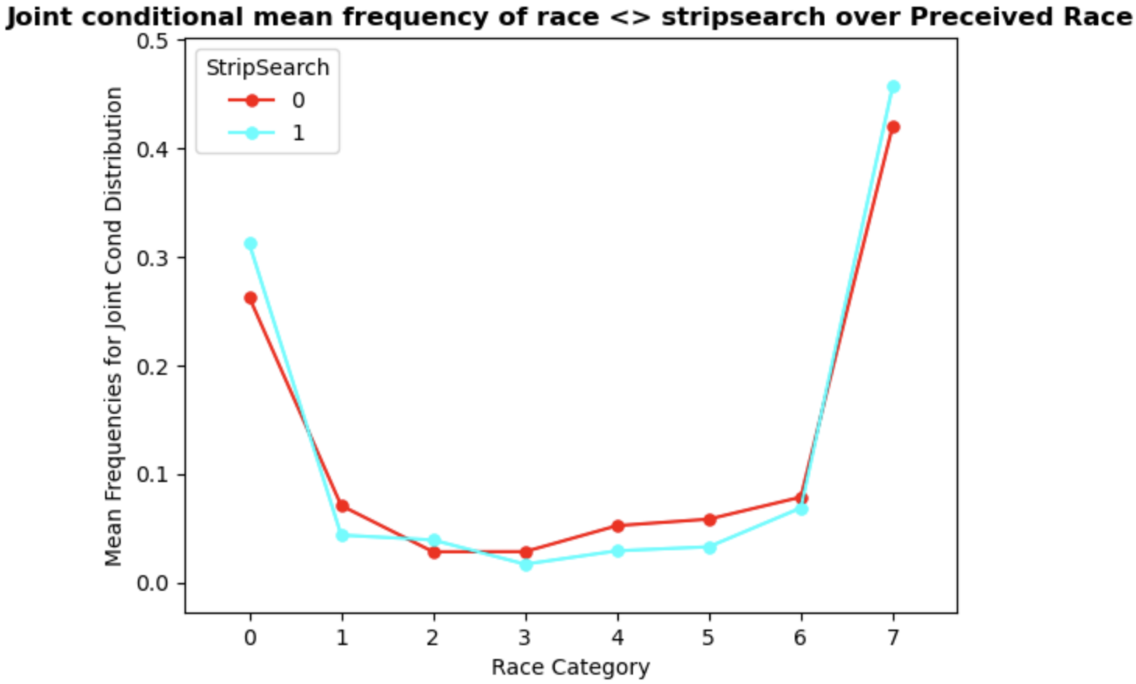

Figure 6: Plot for frequency of normalized joint distributions for Perceived Race and Strip Search.

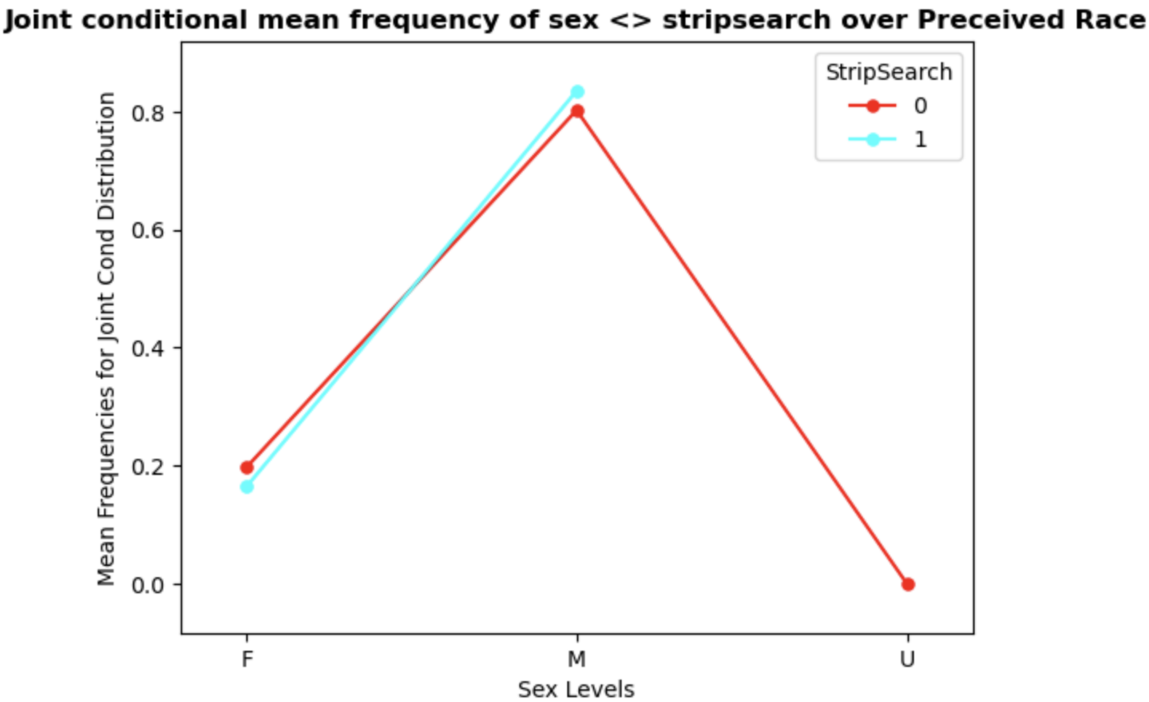

Figure 7: Plot for frequency of joint distributions for Sex and Strip Search.

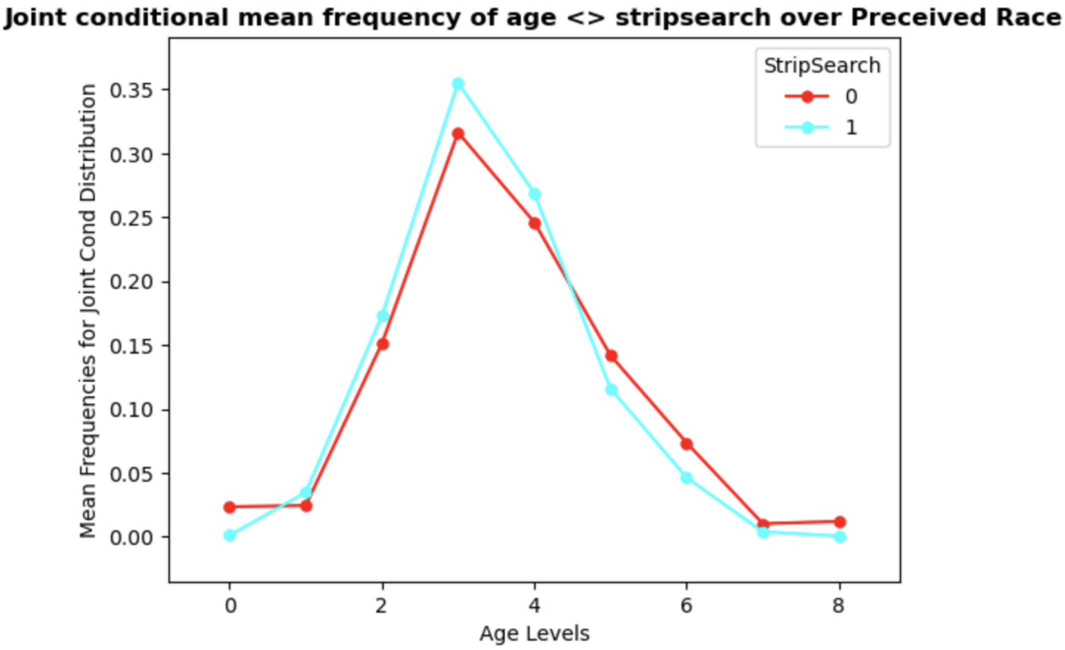

Figure 8: Plot for frequency of joint distributions for Age and Strip Search.

Figure 6 reveals several levels of perceived race and that some levels are more susceptible to the probability of being strip-searched during an arrest, specifically the "Indigenous", "black", and "white" level, which is explored further. These three levels have the opposite of all the other Perceived Race levels, those who are perceived as "dangerous" or "unruly" seem to be prone to higher levels of being strip searched compared to those who are associated with the model minority myth (Yao, 2021). We are interested in determining whether or not the difference is significant. White is an interesting case since it is considered the "default" by the institutional systems that shape our experiences, and can act as a bit of a "control" to compare "model minorities" from "unruly minorities" and how inherent bias towards the "unruly minorities" group may affect them (ie, probability of strip searches go up).

T-Test

We performed multiple t-tests to determine if there is a statistically significant difference in strip search rates based on perceived race. Specifically, we compared the probability of strip search for each race to the probability for other races to determine if there were any significant differences.

T-test 1

Black Strip Search Mean	0.1388
Non-Black Strip Search Mean	0.1123

Table 1: The probability of being searched based on black and other race.

Two-Sample T-Test	P-Value	T-Stat
Black/Other Strip Search	7.732e-19	8.8699

Table 2: P-value and t-stat for being strip-searched across race.

T-test 2

WhiteStrip Search Mean	0.1286
Non-White Strip Search Mean	0.1127

Table 3: The probability of being searched based on white and other race.

Two-Sample T-Test	P-Value	T-Stat
White/Other Strip Search	9.293e-10	6.1221

Table 4: P-value and t-stat for being strip-searched across race.

T-test 3

Indigenous Strip Search Mean	0.1582
Non-Indigenous Strip Search Mean	0.1183

Table 5: The probability of being searched based on whiteand other race.

Two-Sample T-Test for Indigenous and Other	P-value	T-stat
Indigenous and rest races	2.2002889080215823e-06	4.747807831464853

Table 6: P-value and t-stat for indigenous and other race

All of the three T-tests showed a very small p-values, which means black, white and indigenous people have a difference being strip-searched. This result helped us answered our first research question: There is a significant difference in strip search based on perceived race.

Method

ANCOVA

We wanted to know if it was the different ages that caused a higher proportion of these three races to undergo strip-searching at the time of arrest, so we decided to do an ANCOVA. The purpose of ANCOVA is to test the differences in the probability of being strip-searched between different race levels, while controlling other effect. In this case, we would like to see whether there is a difference between the probability of being strip searched in different level of races by controlling the effect of age group.

We reclassified all races into four categories: White, Black, Indigenous and Other. Then, we calculated the total number of arrests for each race and the number of people who received strip searches were then calculated separately. The probability of each race being strip-searched was then obtained by dividing the number of people who were strip-searched by the total number of people arrested for each race.

Logistic Regression

Logistic Regression is a useful tool for analyzing relationships between variables and identifying significant predictors of the outcome variable. For our study, we used logistic regression with StripSearch as our dependent variable, as it is binary (either 0 or 1), and included four race groups and sex as independent variables. By using Logistic Regression, we were able to examine the influence of sex and race on the odds of being strip-searched. This was expressed as the probability of the dependent variable (whether or not to be searched) being 1 or 0. We chose sex and race as independent variables based on the results of our ANCOVA, which indicated their statistical significance and potential importance in influencing the probability of being searched. Additionally, logistic regression can be used for predictive purposes, allowing us to infer the probability of being strip-searched based on available data (e.g., gender and race).

Results

ANCOVA

Table 7. The probability of being strip-searched among different races

Race	Get Arrested	Being Strip Searched	Prob. being seached
White	27718	3566	0.128653
Black	17526	2434	0.138879
Indigenous	1934	306	0.158221
Other	18089	1495	0.086247

Table 8. The result of ANOVA controlling for age

Source	SS	DF	F	p_unc	np2
Race	36.3	3	8.01 e+31	0.0	1.000000
Age	5.61 e-31	1	3.71	0.054	0.000057
Residual	9.87 e-27	65238			

Based on the results of the our first ANCOVA, it can be concluded that when age is controlled, race still has a significant effect on the probability of being strip-searched. The p-value of the between variable race is less than 0.05 after the inclusion of the control variable age, this means that even after controlling for age, there is still a significant difference in the probability of being searched between races. This suggests that the relationship between race and the probability of being searched is not only influenced by age, but that there are other factors that lead to differences in the probability of being searched between races. The result help us to answer the second research question, while controlling age, different race has the different probability of being strip-searched.

The p-value of the control variable age was greater than 0.05 indicating that age was not a significant factor in the probability of being searched, i.e. the relationship between age and the probability of being searched was not significant after controlling for age.

After conducting the first ANCOVA analysis, we aimed to identify a covariate that could potentially influence the observed differences in strip search probabilities across different races. Our investigation led us to consider gender as a potential variable that may exacerbate the disparities in strip search rates among ethnic groups. Therefore, we set Sex as a new controlling variable and did our second ANCOVA. The result is the following:

Table 9. The result of ANOVA controlling for sex

Source	SS	DF	F	p_unc	np2
Race	36.25	3	1.50 e+32	0.0	1.000000
Sex	1.62 e-28	1	2.02 e+03	0.0	0.029965
Residual	5.24 e-27	65262			

From the result we conclude that after controlling for the age variable, race continues to have a significant effect on the probability of taking strip-search by different race, and the probability of undergoing strip-searching continues to differ significantly between race. More than that, the F-statistic even increased compare to the first ANCOVA. This is because in ANCOVA analysis, the inclusion of a control variable reduces the variance of the error, thereby increasing the F-statistic value of the treatment factor (between variable). Therefore, the increase in the F-statistic for race with the inclusion of the control variable Sex is as expected. This suggests that the control variable sex can help us to more accurately test for significant differences in the probability of a search occurring between races. The result help us to answer the third research question, while controlling sex, different race has the different probability of being strip-searched.

The P-value for the control variable sex is less than 0.05, it means that its effect on the treatment factor race is significant, i.e. the control variable has a significant effect on the results. This also suggests that the control variable Sex is necessary in our ANCOVA model to improve the accuracy and stability of the model.

Logistic Regression

The result of our logistic regression is the following:

Table 10. The Logit Regression Results						
Dep. Variable		StripSearch		No. Obs		48947
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.5805	0.045	-57.769	0.000	-2.668	-2.493
Non Minority	0.5642	0.039	14.493	0.000	0.488	0.641

White	0.4837	0.037	13.036	0.000	0.412	0.557
Male	0.2413	0.037	6.476	0.000	0.168	0.314
Current function value:			0.3645			
Accuracy:			0.88			

This logistic regression model shows the effect of four different race and gender categories on the outcome of being strip-searched. The logistic regression model calculates the log odds of the outcome variable (strip searched) based on the predictor variables, sex and three different races. The current function value represents the value of the log-likelihood function at the estimated parameter values. A lower value indicates a better fit. In this case, the value of 0.3645 suggests that the model fits the data reasonably well. The predicte accuracy reached to 88%, which means our model is accurately identifying individuals who accepte strip search duing arrest.

According to the results, we can have the formula log(odds) as follow:

$\log(\frac{\Pi_i}{1-\Pi_i})$ = -2.5805 + 0.5642Non Minority + 0.4837White + 0.2413Male

The odds(Π) for each level:

$$e^{-2.5805 + 0.5642 Non\ Minority + 0.4837 White + 0.2413 Male}$$

Male has a positive coefficient (0.2413) and a statistically significant p-value (<0.001), which means that compared to female, the odds of being strip searched of male is 0.2413 higher. Which implies that the odds of being strip-search will rise when the people been arrested as a male.

For our race category, a positive coefficient (0.5642) on Non Minority (Black and Indigenous) means compated to minority people, the odds of non Minority people been strip-searched 0.5642 higher. Another positive (0.4837) coefficient on White indicates that their odds of being strip searched compared to Minority people will be 0.4837 higher. This result helped us to answer the third research question: Compared to indigenous people, black and indigenous people have more odds of being strip-searched. It is notable that p-values of all the variables are much smaller than the significant level (0.05), It indicates that all race people are statistically significant in predicting the outcome variable.

Discussion

Our statistical analysis revealed significant differences in strip searches based on demographic attributes of race. Our t-test indicated that three groups of people - Whites, Blacks, and Aboriginals - are more likely to be strip-searched by police in Canada than other minority groups. Notably, Black and Indigenous individuals themselves, as a minority, have a higher probability of being strip-searched than other minority races, suggesting possible racial discrimination. Furthermore, ANCOVA results revealed that gender is not the only key factor contributing to strip searches; there are many other factors at play.

While the results of our logistic regression analysis met our expectations and supported our hypotheses, there are limitations to our predictions. The dataset is composed primarily of categorical variables, making it difficult to achieve higher accuracy predictions. Our R-squared values were also low, despite high prediction values, indicating that there is still room for improvement in our prediction model. Future research could explore alternative machine learning techniques or incorporate additional variables to enhance the accuracy of predictions.

Conclusion

In this study, we utilized three distinct statistical methods: power analysis, ANCOVA, and logistic regression to investigate the relationship between perceived race and the probability of being strip-searched while under arrest. Our ANCOVA analysis revealed that race is a significant predictor of strip search probability, even when controlling for age. In addition, gender was found to be a factor affecting the probability of strip-searching, and controlling for it strengthened the relationship between race and the probability of being strip-searched.

Furthermore, our logistic regression analysis demonstrated that black and indigenous individuals were more likely to be strip-searched than other minority races even they are also minorities in Canada. While this may indicate racial discrimination, it could also be attributed to various factors such as resistance during arrest or differing reasons for arrest. Interestingly, our results also showed that Indigenous people were more likely than blacks to experience unequal treatment by the police.

It is important to note that our findings only suggest a correlation between perceived race and the probability of being strip-searched, and we cannot confirm causality. To further investigate this issue, we suggest requesting the Toronto Police Department to provide a more detailed dataset, including the identification of the police officers involved. This information can help determine whether differences in police officers lead to differing probabilities of strip-searching across different races.

Reference

Tator, C., Henry, F., Smith, C., & Brown, M. (2006). Racial profiling in Canada: Challenging the myth of "a few bad apples". Racial Profiling in Canada: Challenging the Myth of "a Few Bad Apples". University of Toronto Press. Retrieved April 16, 2023, from https://books.google.ca/books?id=n9CiinqsWIIC&dq=inherent+bias+race+police+canada&lr=&source=gbs_navlinks_s.

Ontario Human Rights Commission. (2020). (publication). Second interim report on the inquiry into racial profiling and racial discrimination of Black persons by the Toronto Police Service. Retrieved April 16, 2023, from https://www.ohrc.on.ca/sites/default/files/A%20Disparate%20Impact%20Second%20interim%20report%20on%20the%20TPS%20inquiry%20executive%20summary.pdf#overlay-context=en/disparate-impact-second-interim-report-inquiry-racial-profiling-and-racial-discrimination-black.

Yao, D. (2021, November 18). Model minority. The Canadian Encyclopedia. Retrieved April 16, 2023, from https://www.thecanadianencyclopedia.ca/en/article/model-minority

Lu, A. (2019, March 21). *When can police conduct a strip search?* FindLaw. Retrieved April 16, 2023, from https://www.findlaw.com/legalblogs/criminal-defense/when-can-police-conduct-a-strip-search/

Kadiri, H. (2022, Marth 18). *The strip search of child Q proves no child is safe from the police*. The Justice Gap. Retrieved April 16, 2023, from https://www.thejusticegap.com/the-strip-search-of-child-q-proves-no-child-is-safe-from-the-police/

Thomas, T., & Gidda, M. (2023, April 6). *Black girls three times more likely to undergo invasive strip-search by met police*. The Guardian. Retrieved April 16, 2023, from https://www.theguardian.com/uk-news/2023/apr/06/black-girls-three-times-more-likely-to-undergo-invasive-strip-search-by-met-police