The Statistical Tests of Arrests and Strip Searches Dataset for Factors Impact Arrest Cooperativeness Portion and Strip Search Proportion

Group 19:

- Teng-Chieh Chuang (Student# 1004744752)
- Danni Yang (Student# 1009535558)

Introduction

This report demonstrates the study of arrests and strip search dataset, under the rule of the 8-step data analysis cycle. Firstly, the group decided to do the research from two different perspectives to figure out which factors would impact the arrest cooperativeness and strip search proportion. Then the group members divided the work and did the data cleaning as per the requirements of each hypothesis. The paper applies power analysis, ANCOVA test and logistical regression to explore the data and finds the results comes up with the limitations. The data visualization section transfers the mathematical result to a piece of more direct and understandable information. We did the assumption check to see whether the research question was answered or not. Finally, adjustments were applied a thousand times to get the accepted results.

Research Questions

- RQ1: Whether the perceived race groups would impact the strip search proportion?
- RQ2: Can we predict if an arrested person will be cooperative based on their demographic information?

Given the data on how criminals of different backgrounds have behaved and their actions at arrest, we want to investigate any differences in cooperativeness between these criminals. Given our preliminary analysis, we want to focus on how cooperative scores may differ between criminals of different race groups. Besides, since the dataset provides detailed personal information about the crimes while the strip search does not happen every time, we would like to find out whether the perceived race information would impact the police decision-making for performing strip search or not.

Literature Review

Helmer (2001) stated that strip searches without reasonable suspicion are unconstitutional and should be prohibited since the strip search was commonly applied in cases in many American cities where the police have the permit to do so for anyone arrested for any crime. He suggested that strip searches of felony detainees are permissible only if there is reasonable suspicion that the detainee is carrying contraband or weapons and should be taken a balancing test before searching. On the other hand, Newburn, Shiner, and Hayman (2004) explored the potential racial disparities in the use of strip searches for suspects in custody in the UK and found that the strip searches are used more frequently and invasively on Black and Asian detainees compared to White detainees. They suggested that this is due to unconscious biases and stereotypes held by police officers, and the lack of clear guidelines and training on when and how to conduct strip searches. There is a need for better oversight and regulation of strip searches, as well as greater awareness and training on issues of race and unconscious bias within the police force. Simons (1979) specifically focused on the female groups that argued that strip searches are used as a method to abuse women in jail for interrogative purposes.

However, with the decades of evolution of laws and policies and the popularization of human privacy rights, strip searches have been used more prudently in recent years. We are expecting to see any positive changes based on the analysis of the arrest and strip search dataset from the Toronto Police.

Exploratory Data Analysis (EDA)

Data Cleaning for RQ1:

We checked the null values and did the value count for each race group and gender group before we started the data cleaning. We only keep the columns which are related to the research question and remove the null values in them. There are 5056 records for unknown or legacy race groups thus we removed such rows to be more focused on the explainable data. We merged the East/Southeast Asian group and the South Asian group as a new Asian group since there are no obvious differences in macroscopic features and racial information between people from these two areas. We also merge the Latino and Indigenous group since the Latinos in Canada are Canadians who are descendants of people from countries of Latin America while the indigenous in Canada refers to Indians (more commonly referred to as First Nations), Inuit and Métis are also from

America Continent. Besides, since the data between different race groups are greatly imbalanced, we believe that merge the groups mentioned above to two larger race groups would help with our observations.

```
# Recheck the count number of each value in Perceived_Race column
sa['Perceived_Race'].value_counts()
```

```
White             21898
Black             14347
Asian              7174
Indigenous         2961
Middle-Eastern     2840
Name: Perceived_Race, dtype: int64
```

Figure 1: Counted numbers for each perceived race groups.

Then we used the count function to group the Person_ID for strip search cases as the dependent variables and assaulted police cases for the covariate respectively. We then merged two counted number in one table and sorted the table as per the ascending order of strip search counted values. Table 1 shows the head of the grouped dataset.

	Arrest_Year	Arrest_Month	PersonID	Perceived_Race	Sex	StripSearch	Actions_at_arrest___Assaulted_o
0	2020	Apr-June	333666	White	M	8	0
1	2020	Jan-Mar	310131	White	M	7	1
2	2020	Jan-Mar	334680	White	M	6	0
3	2020	Apr-June	306494	White	F	6	0
4	2020	Jan-Mar	310375	Indigenous	M	6	0
5	2020	Apr-June	320543	Black	M	6	0
6	2020	July-Sept	336154	White	F	6	0

Table 1: Counted numbers for strip search cases and assaulted police cases.

Data Visualization for RQ1

We created a box plot for perceived race groups to see the big picture of the data distribution. For the strip search cases, since the data distribution is greatly imbalanced with the 57472 pieces of 0 value and 7800 pieces of 1 value, box plot does not suitable for describing this dataset that, the mean, median, and interquartile range (IQR) are all closed to 0. As shown in Figure 2, the white race group has the most significant outlier which refers to 8 strip searches for the same person. The indigenous and the black come the second place with 6 strip searches for the same person while the farthest outlier for the Asian group and the middle-eastern group is 4.

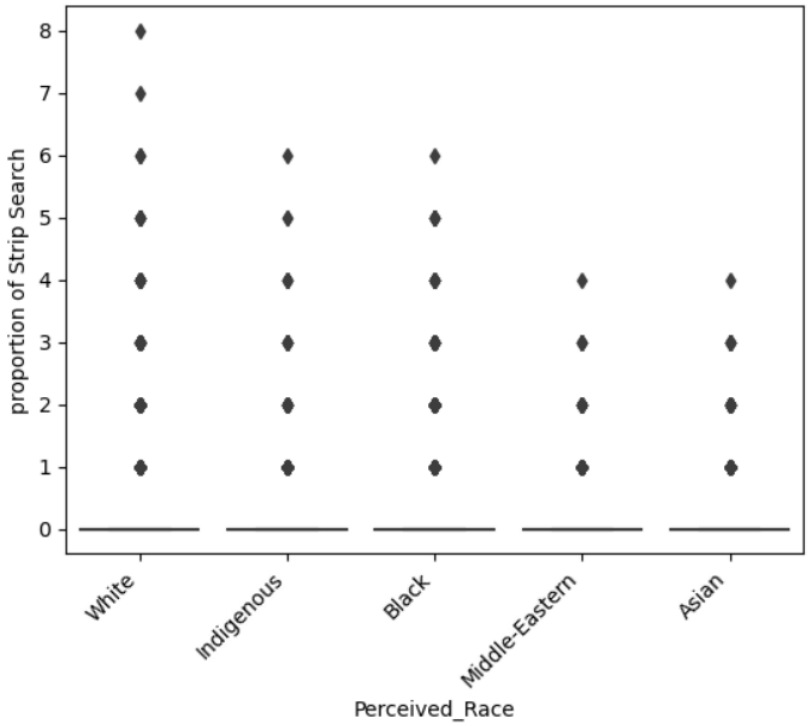


Figure 2: Box plot for perceived race group and strip search.

Figure 3 shows the distribution of strip search grouped by the perceived race groups. It is obvious to see that it's a right-skewed distribution which indicates the mode of the proportion is less than the mean value. Almost half of the data are located in the zone [0, 1]. To conclude, the statistical summary of perceived race groups and gender groups is shown in Table 1.

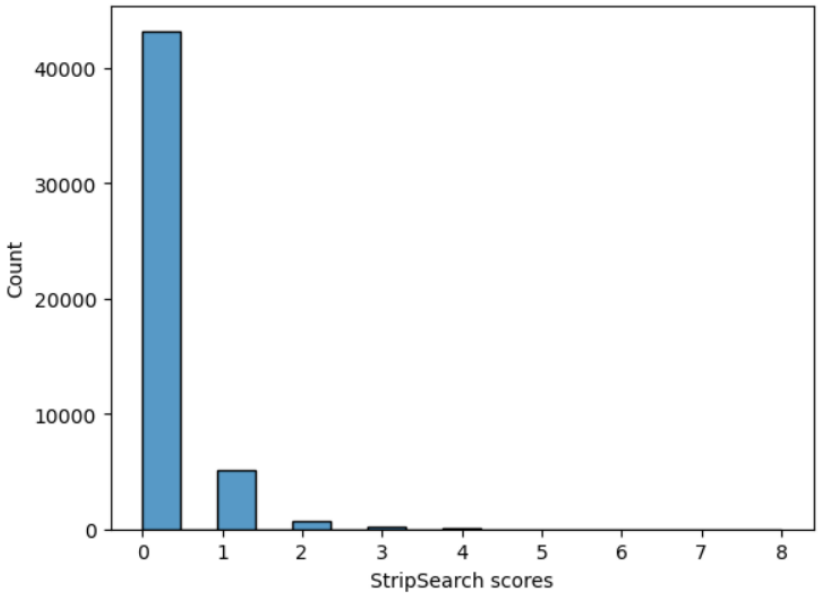


Figure 3: Distribution of strip search.

Perceived_Race	*count*	*mean*	*std*	*min*	*25%*	*50%*	*75%*	*max*
Asian	7174	0.083	0.310	0	0	0	0	4
Black	14347	0.170	0.466	0	0	0	0	6
Indigenous	2961	0.148	0.462	0	0	0	0	6
Middle-Eastern	2840	0.803	0.316	0	0	0	0	4
White	21898	0.163	0.471	0	0	0	0	8

Table 1: Statistic summary of perceived race groups and strip search.

Since we are doing ANCOVA test and we also need covariance for the model. The same steps would also be repeated for the Assaulted police cases. Figure 4 shows the box plot of the assaulted police cases which demonstrates the imbalanced data distribution with the mean, median, and interquartile range (IQR) are all closed to 0. As shown in Figure 5, the white race group has the most significant outlier which refers to 3 strip searches for the same person. The middle-eastern group has the closest outlier 1 while the outlier for the rest of race groups are located in the zone [1, 2].

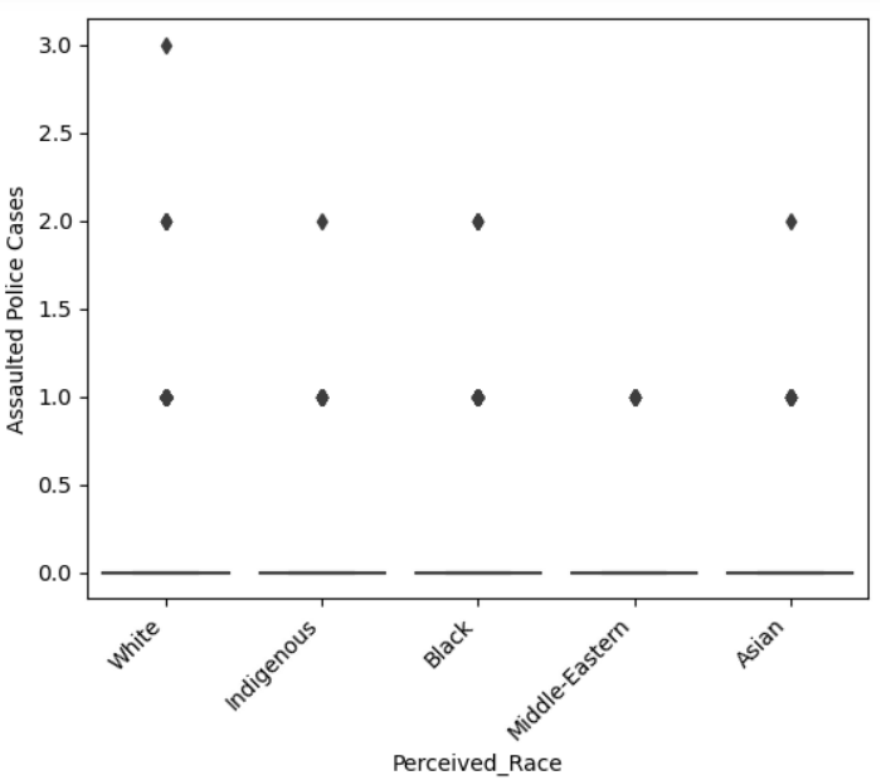


Figure 4: Box plot for perceived race group and assaulted police cases.

Figure 6 shows the distribution of assaulted police cases grouped by the perceived race groups. It is obvious to see that it's a right-skewed distribution which indicates the mode of the

proportion is less than the mean value. Almost half of the data are located in the zone [0, 1]. To conclude, the statistical summary of perceived race groups and gender groups is shown in Table 2.

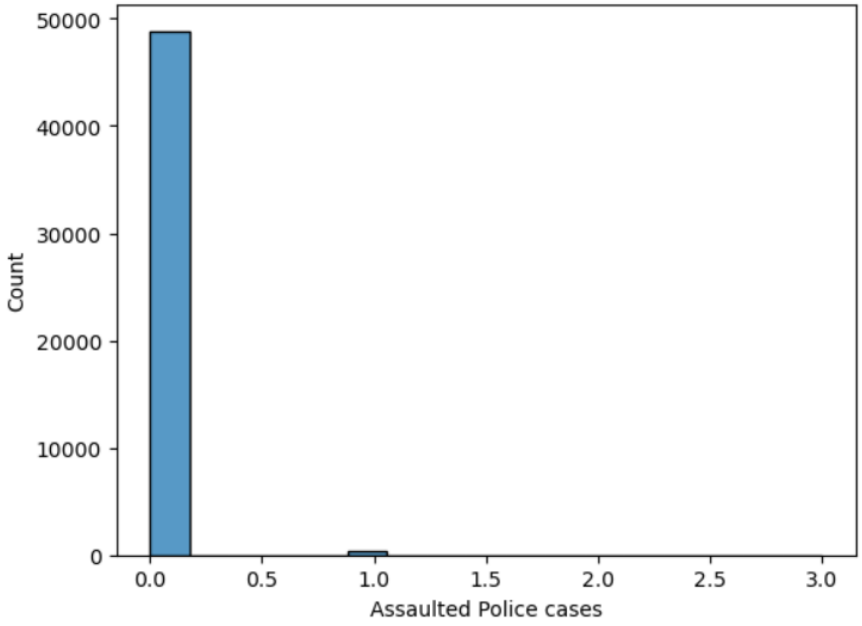


Figure 5: Distribution of assaulted police cases.

Perceived_Race	***count***	***mean***	***std***	***min***	***25%***	***50%***	***75%***	***max***
Asian	7174	0.003	0.060	0	0	0	0	2
Black	14347	0.010	0.102	0	0	0	0	2
Indigenous	2961	0.118	0.111	0	0	0	0	2
Middle-Eastern	2840	0.006	0.075	0	0	0	0	1
White	21898	0.008	0.092	0	0	0	0	3

Table 2: Statistic summary of perceived race groups and assaulted police cases.

T-test for RQ1:

Since the size of each group differs a lot, we choose the Welch t-test to check whether the means of the two populations are equal.

Indigenous and Black

- H0(Null Hypothesis): The population means of strip search of the two independent groups, Black and Indigenous, are equal.
- HA (Alternative Hypothesis): The population means of strip search of the two independent groups, Black and Indigenous, are NOT equal.

The mean value and the standard deviation of the Indigenous group (m = 0.170, SD = 0.466) are higher than Black (m = 0.150, SD = 0.462). The p-value of these two groups is 0.02, which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Indigenous and Black groups. The t-statistic value of this pair group is 2.327, which means there is not much evidence to support that there is a significant difference between these two groups.

Indigenous and White

- H0(Null Hypothesis): The population means of strip search of the two independent groups, Indigenous and White, are equal.
- HA (Alternative Hypothesis): The population means of strip search of the two independent groups, Indigenous and White, are NOT equal.

The mean value and the standard deviation of the Indigenous group (m = 0.170, SD = 0.466) are higher than White (m = 0.163, SD = 0.471). The p-value of these two groups is 0.1, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search proportion does not differ from Indigenous and White groups. The t-statistic value of this pair group is 1.647, which means there is not much evidence to support there is a significant difference between these two groups.

Indigenous and Asian

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Asian, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Asian, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.170, SD = 0.466) are higher than Asians (m = 0.083, SD = 0.310). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Indigenous and Asian groups. The t-statistic value of this pair group is -6.986, which means there is great evidence to support there are significant differences between these two groups.

Indigenous and Middle East

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Middle East, are equal.

- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Indigenous and Middle East, are not equal.

The mean value and the standard deviation of the Indigenous group (m = 0.170, SD = 0.466) are higher than the Middle East (m = 0.080, SD = 0.316). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Indigenous and Middle East groups. The t-statistic value of this pair group is -6.533, which means there is strong evidence to support there is a significant difference between these two groups.

Asian and Black

- H0(Null Hypothesis): The population means of strip search of the two independent groups, Black and Indigenous, are equal.
- HA (Alternative Hypothesis): The population means of strip search of the two independent groups, Black and Indigenous, are NOT equal.

The mean value and the standard deviation of the Asians (m = 0.083, SD = 0.310) are higher than Black groups (m = 0.150, SD = 0.462). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Indigenous and Black groups. The t-statistic value of this pair group is 16.146, which means there is strong evidence to support that there is a significant difference between these two groups.

Asian and White

- H0(Null Hypothesis): The population means of strip search of the two independent groups, Asian and White, are equal.
- HA (Alternative Hypothesis): The population means of strip search of the two independent groups, Asian and White, are NOT equal.

The mean value and the standard deviation of the Asians (m = 0.083, SD = 0.310) are higher than White (m = 0.163, SD = 0.471). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Asian and White groups. The t-statistic value of this pair group is 16.146, which means there is strong evidence to support that there is a significant difference between these two groups.

Asian and Middle East

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Asian and Middle East, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Asian and Middle East, are not equal.

The mean value and the standard deviation of the Asians (m = 0.083, SD = 0.310) are higher than the Middle East (m = 0.080, SD = 0.316). The p-value of these two groups is 0.659, which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search proportion does not differ from Asian and White groups. The t-statistic value of this pair group is -0.441, which means there is not much evidence to support there is a significant difference between these two groups.

Black and White

- H0(Null Hypothesis): The population means of strip search of the two independent groups, Black and White, are equal.
- HA (Alternative Hypothesis): The population means of strip search of the two independent groups, Black and White, are NOT equal.

The mean value and the standard deviation of the Black groups (m = 0.150, SD = 0.462) are higher than White (m = 0.163, SD = 0.471). The p-value of these two groups is 0.176 which is larger than 0.05. Thus, we could NOT reject the null hypothesis and conclude that the strip search does NOT differ from Black and White groups. The t-statistic value of this pair group is 1.354, which means there is no evidence to support that there is a significant difference between these two groups.

White and Middle East

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, White and Middle East, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, White and Middle East, are not equal.

The mean value and the standard deviation of the White (m = 0.163, SD = 0.471) are higher than the Middle East (m = 0.080, SD = 0.316). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from White and Middle-Eastern groups. The t-

statistic value of this pair group is -12.266, which means there is strong evidence to support there is a significant difference between these two groups.

Black and Middle East

- H0(Null Hypothesis): The population means of strip search proportion of the two independent groups, Black and Middle East, are equal.
- HA (Alternative Hypothesis): The population means of strip search proportion of the two independent groups, Black and Middle East, are not equal.

The mean value and the standard deviation of the Black groups (m = 0.150, SD = 0.462) are higher than the Middle East (m = 0.080, SD = 0.316). The p-value of these two groups is 0 (refers to a value closed to 0 but not 0) which is smaller than 0.05. Thus, we could reject the null hypothesis and conclude that the strip search greatly differs from Black and Middle-Eastern groups. The t-statistic value of this pair group is 12.596, which means there is strong evidence to support there is a significant difference between these two groups.

Data Cleaning for RQ2:

First, given the dataset source explanation, we change the value of booked from 0 to 1 for all rows where the value for strip search is 1, meaning that a strip search took place. Next, we recognized for the Age column, there are two columns representing people under age 17, and also there are two columns representing people aged 65 and over. We removed the redundant column by combining the two columns that represent the same group.

Table 3

	Arrest_Year	Arrest_Month	Perceived_Race	Sex	Age_group__at_arrest_	Actions_at_arrest___Cooperative	Arrests	portion
0	2020	Apr-June	Black	F	Aged 17 years and younger	16	36	0.44
1	2020	Apr-June	Black	F	Aged 18 to 24 years	30	87	0.34
2	2020	Apr-June	Black	F	Aged 25 to 34 years	23	99	0.23
3	2020	Apr-June	Black	F	Aged 35 to 44 years	22	52	0.42
4	2020	Apr-June	Black	F	Aged 45 to 54 years	6	26	0.23

To account for the unequal representation of each race in the dataset, here we first combined the data rows by five columns, “Arrest_Year”, “Arrest_Month”, “Perceived_Race”, “Sex”, and “Age_group_at_arrest_”. We summed the values (1 and 0) in the Action at arrest cooperative column as the aggregated number of cooperative instances. Next, we calculate the

cooperative scores for each race group by dividing the number of cooperative instances by the total number of arrests for each group and name the new column 'portion'.

Data Visualization for RQ2

Figure 6: *Visualise the overall pattern between Race groups and cooperativeness.*

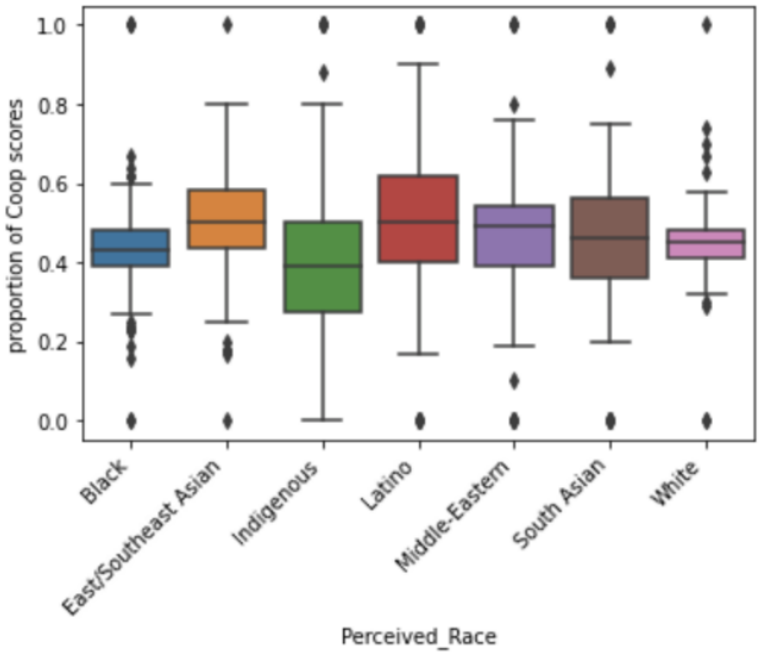
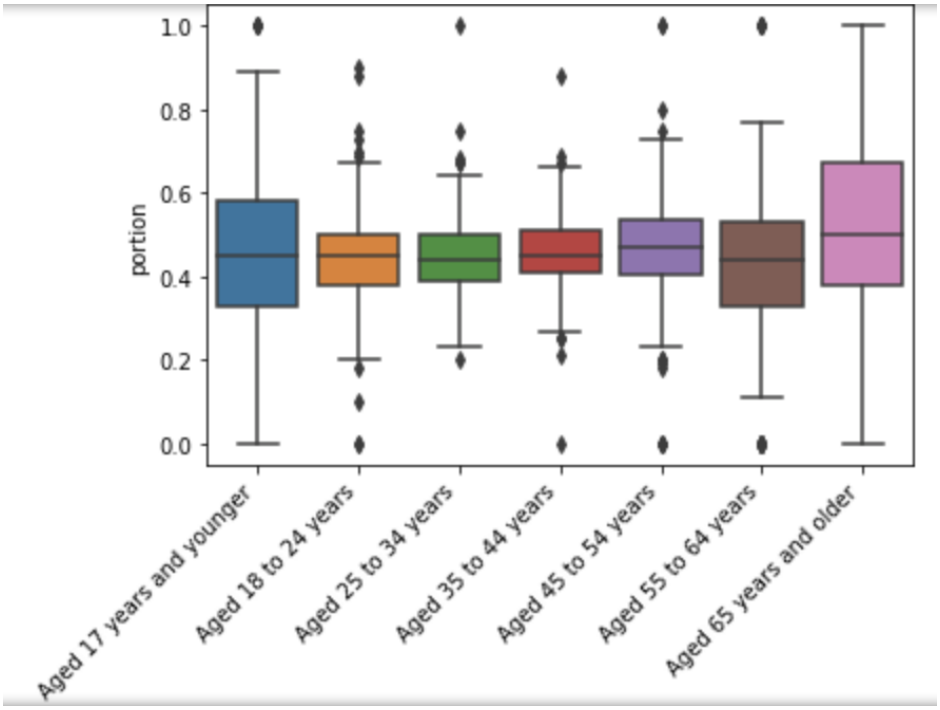


Figure 7: *Visualise the overall pattern between age group and cooperativeness.*



We created a box plot to visualize the overall pattern of the cooperative score between different race groups and age groups. First, we can see relatively consistent and uniform behaviours in the black and white groups shown by a comparatively short plot. The median mostly falls between 0.4 and 0.6, but the distribution varies. We can visually spot a difference between groups, especially between the Indigenous and Latino groups, where the upper quartile of the Latino group reaches over 0.6, and the lower quartile of the indigenous group goes as low as 0.3. From the plot, it seemed possible that there is a difference in cooperativeness between race groups. We also created a box plot of cooperative scores between different age groups, and the plot did not reveal any obvious trends or differences between groups.

Table 4: Summary statistics of cooperative scores of different race groups

Perceived_Race	count	mean	std	min	25%	50%	75%
Black	110	0.433818	0.141376	0.0	0.39	0.43	0.48
East/Southeast Asian	112	0.504196	0.142662	0.0	0.4375	0.5	0.5825
Indigenous	98	0.411939	0.248163	0.0	0.275	0.39	0.5
Latino	101	0.505545	0.242052	0.0	0.4	0.5	0.62
Middle-Eastern	107	0.464766	0.195855	0.0	0.39	0.49	0.54
South Asian	106	0.457453	0.218122	0.0	0.36	0.46	0.56
White	116	0.448621	0.105393	0.0	0.41	0.45	0.48

Again, from this table, we can see the mean cooperativeness score varies across different perceived race groups. The highest mean score is for Latinos and the lowest for Indigenous. There is also a wide variation in the standard deviation of cooperativeness scores within each perceived race group.

T-test for RQ2:

In order to answer the question, whether we can predict if an arrested person will be cooperative based on their demographic information, we ran a logistic regression which aims to investigate the relationship between race, sex and cooperation during an arrest. After performing Welch's T-test and we discovered a significant difference in mean cooperative scores between selected pairs, we proceeded to run a logistic regression. We one-hot encoded the different race groups and genders which are used as the predictors, and the dependent variable is a binary outcome, representing whether or not the individual was cooperative during the arrest.

Driven by the summary statistics of cooperative scores of different race groups, we selected the Indigenous Latino as the primary target for testing since they stood out in the boxplot, we also

incorporated Black and White groups into the tests for comparisons. More specifically, we performed Welch's t-test on the race group pairs of Indigenous versus Latino, Black and white, and Latino versus Black and white, five tests in total.

Latino and Indigenous

We computed the mean cooperative scores for the Latino group and the Indigenous group and found that the score for Latinos is much higher than for Indigenous (0.51 and 0.41, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Indigenous, are equal.
- HA (Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Indigenous, are not equal.

The results indicate that the mean cooperative score of the Latino group (M=0.51, SD=0.24) is higher than the mean cooperative score of the Indigenous group(M=0,41, SD-0,25). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.008) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and Indigenous, are not equal.

Latino and White

We computed the mean cooperative scores for the Latino group and White group and found that the score for Latinos is also higher than for White (0.51 and 0.45, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and White, are equal.
- HA(Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and White, are not equal.

The results indicate that the mean cooperative score of the Latino group(M=0.51, SD=0.24) is higher than the mean cooperative score of the white group(M=0.45, SD=0.11). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.03) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and White, are not equal.

Latino and Black

We computed the mean cooperative scores for the Latino group and the Black group and found that the score for Latinos is also higher than White (0.51 and 0.43, respectively). We conducted a Welch's T-test to analyze whether the cooperative scores (outcome variable) differ between these two race groups. The hypothesis being tested are the following:

- H0(Null Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Black, are equal.
- HA (Alternative Hypothesis): The population means of cooperative scores of the two independent groups, Latino and Black, are not equal.

The results indicate that the mean cooperative score of the Latino group (M=0.51, SD=0.14) is higher than the mean cooperative score of the White group (M=0.45, SD-0.25). With alpha established at 0.05, this is a statistically significant difference as the p-value (0.01) is less than 0.05. Therefore, we reject the null hypothesis and conclude that the population means of cooperative scores of the two independent groups, Latino and Black, are not equal.

The same Welch's T-test was performed for White vs. Indigenous and Black vs. Indigenous. However, we did not find a significant difference in cooperative scores between the indigenous group and either the Black or White groups, as indicated by the p-values of 0,443 and 0.175.
The t-tests we performed overall suggest that the proportion of cooperative individuals in the Latino group is significantly different from that in the Black, White and Indigenous group. From here, to further investigate the difference in mean cooperative scores amongst different race groups, we would proceed with a one-way ANOVA test.

Method

Dataset Description

For our project, we used the Police service dataset that's made public and available on the Public Safety Data Portal of the Toronto Police service. This dataset shows Policing information related to arrests and strip searches. The dataset also includes criminals' information such as demographic information, the occurrence of crime, and their actions at arrest. The data has 65276 records of arrests within the timeframe from 2020 to 2022. Some features of interest include gender, race groups, strip search conducted, and their actions at arrest. Note that the dataset does not

contain any continuous data but only binary data (0 and 1). For example, for the strip search column, the value of 1 indicates a strip search was conducted, and if a strip search did not take place, the value equals 0. Same for columns such as “Action_at_Arrest_Cooperative”, a value of 1 indicates the person was cooperative, and 0 if the person was not.

For the first research question, we focused on the strip search column and found that the strip search does not frequently happen that, only around 13.6% of the arrested cases applied strip searches, which is much lower than that decades ago. As for the second research question, we discovered that there are more records of people Aged between 18 to 44, with 25 to 34 being the largest age group. In addition, more than half of the total records are of the White and Black race group, with 27723 being White. We also noted that there are substantially more records of the arrest of males than females at approximately 5 to 1 ratio.

Given that the dataset lacks continuous data, we considered performing some data wrangling, including combining columns and aggregating the number of the potential explanatory variables, such as strip search.

Power Analysis for RQ1

```
Sample Size of Bkss vs Wtss: 131448 needed for Bkss
Actual size of Bkss: 14347
Sample Size of Bkss vs Wtss: 52037 needed for Wtss
Actual size of Wtss: 21898
Sample Size of Bkss vs Igss: 12703 needed for Bkss
Actual size of Bkss: 14347
Sample Size of Bkss vs Igss: 5029 needed for Igss
Actual size of Igss: 2961
Sample Size of Bkss vs Asss: 659 needed for Bkss
Actual size of Bkss: 14347
Sample Size of Bkss vs Asss: 261 needed for Asss
Actual size of Asss: 7174
Sample Size of Bkss vs Mess: 687 needed for Bkss
Actual size of Bkss: 14347
Sample Size of Bkss vs Mess: 272 needed for Igss
Actual size of Mess: 2840
Sample Size of Wtss vs Igss: 27432 needed for Wtss
Actual size of Bkss: 14347
Sample Size of Wtss vs Igss: 10860 needed for Igss
Actual size of Igss: 2961
Sample Size of Wtss vs Asss: 837 needed for Wtss
Actual size of Wtss: 21898
Sample Size of Wtss vs Asss: 331 needed for Asss
Actual size of Asss: 7174
Sample Size of Wtss vs Mess: 845 needed for Wtss
Actual size of Bkss: 14347
Sample Size of Wtss vs Mess: 334 needed for Mess
Actual size of Mess: 2840
Sample Size of Igss vs Asss: 867 needed for Igss
Actual size of Igss: 2961
Sample Size of Igss vs Asss: 343 needed for Asss
Actual size of Asss: 7174
Sample Size of Igss vs Asss: 343 needed for Asss
Actual size of Asss: 7174
Sample Size of Igss vs Mess: 956 needed for Igss
Actual size of Igss: 2961
Sample Size of Igss vs Mess: 378 needed for Mess
Actual size of Mess: 2840
Sample Size of Asss vs Mess: 284717 needed for Asss
Actual size of Bkss: 7174
Sample Size of Asss vs Mess: 112712 needed for Mess
Actual size of Mess: 2840
```

Figure 10: Power Analysis for RQ1

ANCOVA for RQ1

Figure 8 shows the result of ANCOVA test in this case. The dependent variable is strip search, the independent variable is perceived race groups, and the covariance is the assaulted police cases.

- H0(Null hypothesis): μ1 = μ2 = μ3 = μ4 = μ5. The strip search means that all the five perceived race groups are equal.
- H1(Alternative hypothesis): At least one mean is different from the others.

	Source	SS	DF	F	p-unc	np2
0	Perceived_Race	52.952050	4	68.194945	1.163452e-57	0.005512
1	Actions_at_arrest___Assaulted_o	25.987555	1	133.873563	6.385389e-31	0.002713
2	Residual	9553.428726	49214	NaN	NaN	NaN

Figure 9: ANCOVA for RQ1

Logistic Regression for RQ2

```
Optimization terminated successfully.
         Current function value: 0.685625
         Iterations 5
                                Logit Regression Results
==============================================================================================
Dep. Variable:     Actions_at_arrest___Cooperative   No. Observations:                48939
Model:                                       Logit   Df Residuals:                    48929
Method:                                        MLE   Df Model:                            9
Date:                             Sat, 15 Apr 2023   Pseudo R-squ.:               0.002099
Time:                                     02:27:03   Log-Likelihood:                -33554.
converged:                                    True   LL-Null:                       -33624.
Covariance Type:                         nonrobust   LLR p-value:                 5.899e-26
==============================================================================================
                                 coef    std err          z      P>|z|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Intercept                     -0.3864      0.026    -14.752      0.000      -0.438      -0.335
Race_East_Southeast_Asian      0.3697      0.039      9.514      0.000       0.294       0.446
Race_Indigenous               -0.0694      0.057     -1.222      0.222      -0.181       0.042
Race_Latino                    0.3245      0.057      5.659      0.000       0.212       0.437
Race_Middle_Eastern            0.1783      0.044      4.012      0.000       0.091       0.265
Race_South_Asian               0.1526      0.042      3.592      0.000       0.069       0.236
Race_Unknown_or_Legacy         0.0734      0.037      1.972      0.049       0.000       0.146
Race_White                     0.1098      0.023      4.865      0.000       0.066       0.154
Sex_M                          0.0782      0.023      3.366      0.001       0.033       0.124
Sex_U                         -0.8367      0.817     -1.024      0.306      -2.439       0.765
==============================================================================================
```

Figure 10: Logistic Regression for RQ2

Findings

ANCOVA

- Statistical interpretation

Interpretation p-unc = "uncorrected p-value" for perceived race is smaller than 0.05. Thus, we can reject the null hypothesis that each of race groups results after controlling for the assaulted police case level. The sum of squares is 52.95 for the perceived race and 25.99 for the assaulted police cases. The degree of freedom for the race group is 4 while the assaulted case is only 1.

- Practical interpretation

We hypothesized that perceived race group would be able to predict whether a person would be strip searched while arresting. From our results, we see that there is statistically significant relationship between race groups and strip search times when controlling for the assaulted police level. This raises a sad insight that we still have race discrimination and stereotype in our police system after the decades of professional training. Assaulted police is regraded as an offensive and aggressive action during the arresting process since it may cause higher danger rate for the police officers. Thus it is understandable to take strip search under such circumstance. However, as we control the assaulted level, the bias still exists. I believer we still need to take more actions to deal with the unfair judgement based on the race stereotypes.

ANCOVA Assumption Check

Figure 11, 12, 13 demonstrate the results of ANCOVA assumption check. Firstly, we converted the strip search instances and assaulted cases from the binary to the decimal continues data to fit the ANCOVA requirements. The independent variable perceived race has five groups which also fits the requirement for independent variable should be categorical data with more than 2 levels.

We checked the distribution of strip search cases and find it is almost normal distributed which satisfy the ANCOVA test requirement. Figure 12 shows the best fit line but it seems there is no point locates on the line and we cannot tell our model is accuracy and reliable. For the figure 13, the Q-Q plot for each perceived group takes very similar diagram with almost parallel best fit lines. Such information indicates that there is no interaction impact in the perceived race groups. However, since there is only a few points locate on the "best fit" line, the model accuracy is not good to go and we can't use this model to predict future behaviors which refers to

we can’t predict whether the strip search would take place or not based on people’s perceived race group.

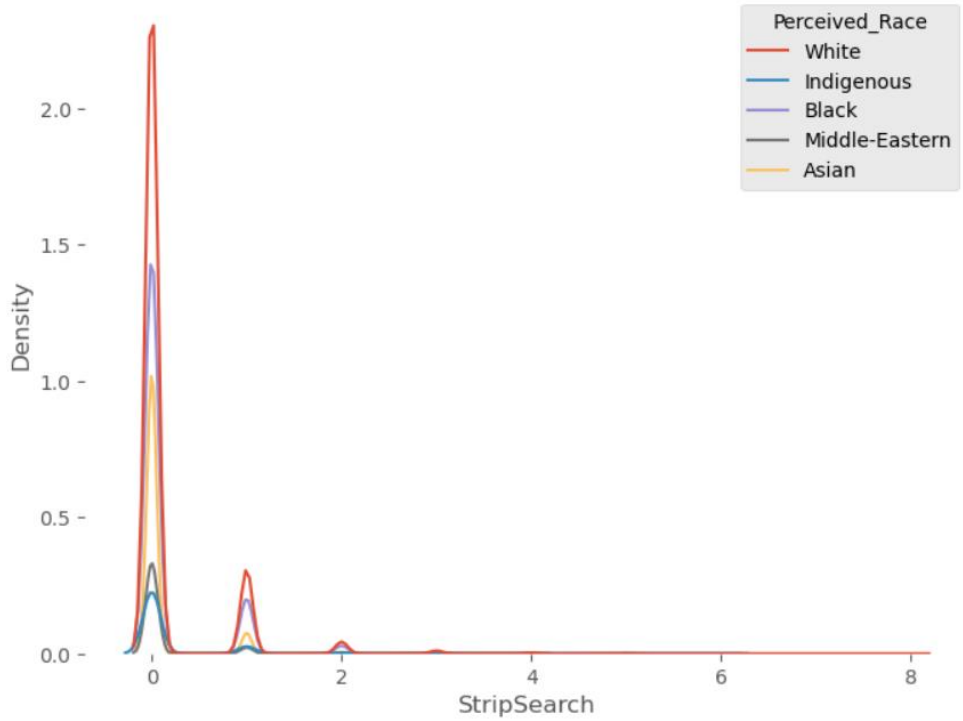


Figure 11: Distribution of strip search

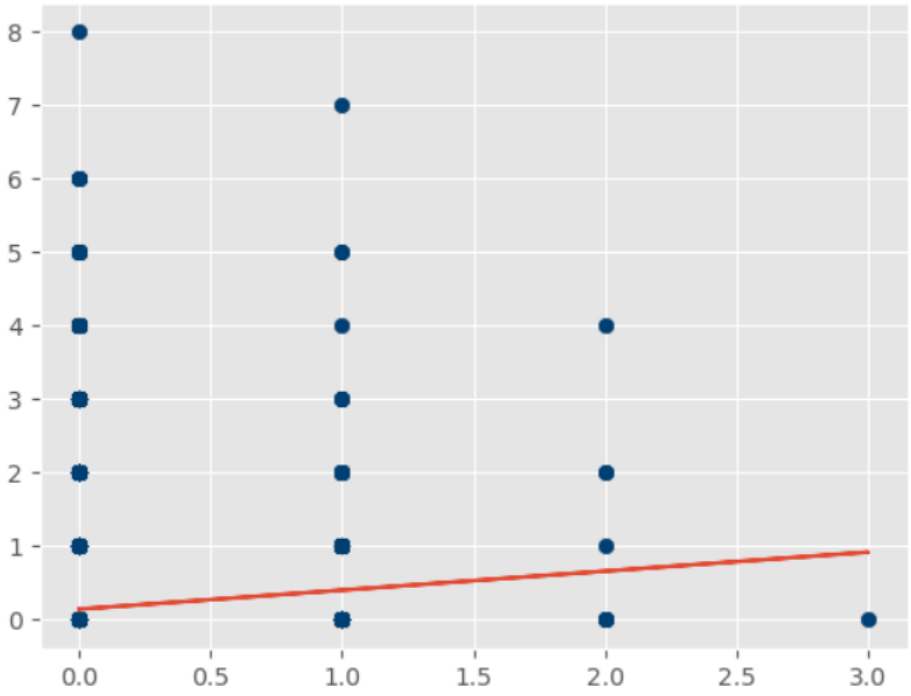


Figure 12: Scatter Plot with the best fit line for RQ1

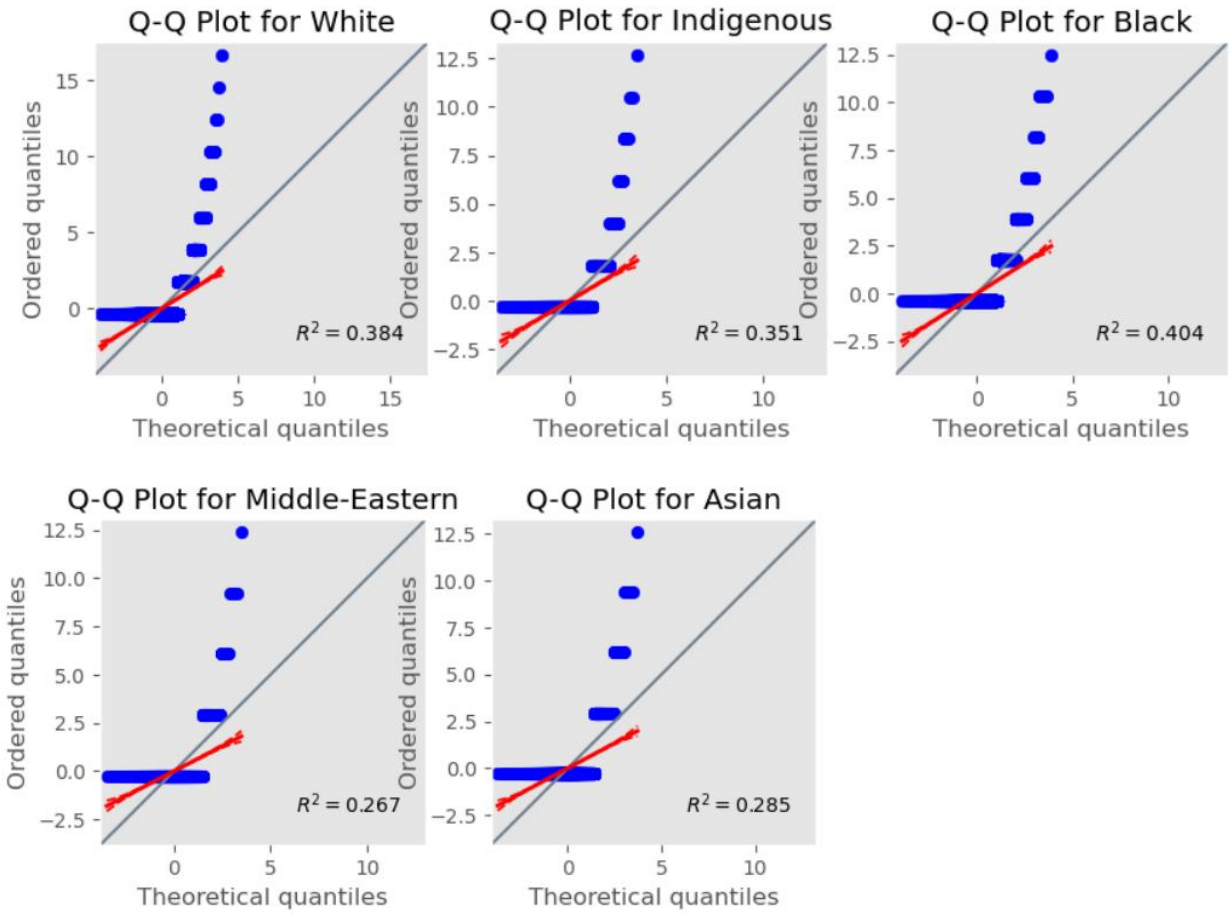


Figure 13: QQ plot for each perceived race group.

Logistic Regression

Table 5

	Variable	coef	std_err	z_value	P>|z|	95% CI Lower	95% CI Upper
0	Intercept	-0.379	0.026	-14.527	0.000	-0.431	-0.328
1	Race_East_Southeast_Asian	0.351	0.039	9.066	0.000	0.275	0.427
2	Race_Indigenous	-0.059	0.057	-1.043	0.297	-0.170	0.052
3	Race_Latino	0.347	0.057	6.079	0.000	0.235	0.458
4	Race_Middle_Eastern	0.168	0.045	3.763	0.000	0.080	0.255
5	Race_South_Asian	0.169	0.042	3.999	0.000	0.086	0.252
6	Race_Unknown_or_Legacy	0.062	0.037	1.675	0.094	-0.011	0.136
7	Race_White	0.092	0.023	4.094	0.000	0.048	0.137
8	Sex_M	0.073	0.023	3.160	0.002	0.028	0.119
9	Sex_U	-0.425	0.708	-0.601	0.548	-1.813	0.962

The coefficients represent the log-odds of being cooperative during an arrest, relative to the reference category (Race_Black). Positive coefficients indicate a higher likelihood of cooperation, while negative coefficients indicate a lower likelihood. The p-values help determine the statistical significance of each variable. We will also exponentiate the coefficient to acquire the odds ratio, which represents the change in odds of the outcome for a one-unit change in the predictor variable.

East/Southeast Asian

The positive coefficient of 0.351 and a p-value less than 0.001 indicates that compared to the reference race group, the log odds of the individuals in this racial category being cooperative upon arrest is 0.351 units higher. In terms of odds ratio(exp(0.351) = 1.42), the odds of an individual in this race group being cooperative during an arrest are 1.42 times higher than the reference group, holding other variables constant.

Indigenous

The negative coefficient of -0.059 suggests the odds of cooperation for Indigenous individuals are 0.94 times(exp(-0.059)) lower than the reference group, but the p-value of 0.297(>0.05) suggests that the difference is not statistically significant.

Latino

The positive coefficient of 0.347 and p-value less than 0.001 indicates that the log odds of Latino individuals being cooperative upon arrest is 0.347 units higher, compared to the reference group. In terms of odds ratio, Latino individuals have 1.41 times higher odds of being cooperative during an arrest than the reference group.

Middle Eastern

The positive coefficient of 0.168 and p-value less than 0.001 indicate that the log odds of middle eastern individuals being cooperative upon arrest is 0.168 units higher, compared to the reference group. In terms of odds ratio, Middle eastern have 1.18 times higher odds of being cooperative during an arrest than the reference category.

South Asian

The positive coefficient of 0.169 and p-value less than 0.001 indicate that the log odds of South Asian individuals being cooperative upon arrest is 0.169 units higher compared to the reference group. In terms of odds ratio, South Asians have 1.18 times higher odds of being cooperative during an arrest than the reference category.

White

The positive coefficient of 0.092 and p-value less than 0.001 indicate that the log odds of White individuals being cooperative upon arrest is 0.092 higher units higher compared to the reference group. In terms of odds ratio, White individuals have 1.1 times higher odds of being cooperative during an arrest than the reference category

Male

For the sex predictor, the positive coefficient of 0.073 and p-value of 0.002 indicate that the log odds of male being cooperative upon arrest is 0.073 units higher than female during an arrest. In terms of odds ratio, males have 1.08 times higher odds of being cooperative during an arrest than females.

Table 6.

Racial Category	Log-odds	Probability	Standard Error	Prediction Interval
Race_East_Southeast_Asian	-0.0281	0.4929	0.0261	(0.4418, 0.5442)
Race_Indigenous	-0.4386	0.3921	0.0388	(0.3161, 0.4680)
Race_Latino	-0.0328	0.4918	0.0568	(0.3805, 0.6031)
Race_Middle_Eastern	-0.2116	0.4473	0.0570	(0.3355, 0.5591)
Race_South_Asian	-0.2103	0.4476	0.0446	(0.3602, 0.5350)
Race_Unknown_or_Legacy	-0.3169	0.4214	0.0423	(0.3386, 0.5043)
Race_White	-0.2871	0.4287	0.0373	(0.3556, 0.5018)

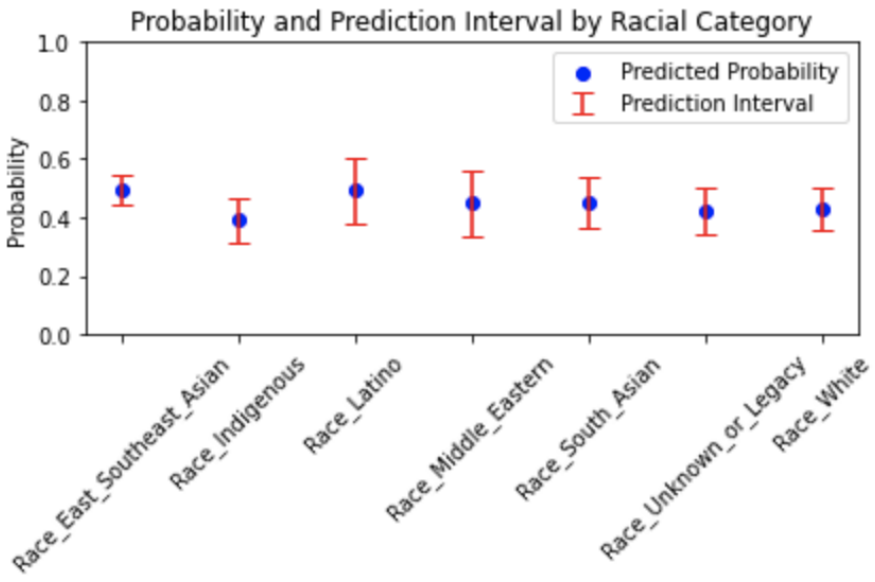


Figure 14

The prediction Intervals are calculated and shown in table 2, and also plotted in the graph. The prediction interval provides an estimate of the range within which the true probability of the outcome lies for each racial category, with a specified level of confidence (95% in this case). For our study, the outcome is whether individuals are cooperative during arrests, and the table shows the predicted probability of cooperation for each racial category.

To interpret the prediction intervals, in terms of East and Southeast Asian for example, we are 95% confident that the true probability of cooperation for this group lies between 44.18% and 54.42%. Similarly, with other groups like Latino, we are 95% confident that the true probability of cooperation for the Latino group lies between 38.05% and 60.31%.

Among the racial categories, the East/Southeast Asian and Latino group have the highest predicted probabilities of being cooperative during an arrest. The indigenous group have the lowest predicted probabilities of being cooperative during an arrest. Note that some prediction intervals are wider than others, which indicates a higher degree of uncertainty in the predicted probabilities for those racial groups. For example, the latino and Middle Eastern groups have wider prediction intervals compared to the East Southeast Asian and White groups.

Discussion

In the study of research question 1, we continued our exploration based on the mid-term project but come up with totally different findings. Previously we found that the sex information and the race information would NOT impact proportion of the strip search since the decades of unbiased training finally has shown the positive effect to the public. However, the accuracy of the previous was not accuracy thus the result may not be considered as a trustable conclusion. In this project we use ANCOVA and made changes in the data cleaning to try to make the data set be satisfied with the ANCOVA requirements and set up a predict model.

The Welch t-test shows the significant impact relations exist between most race groups, and the ANCOVA shows that when we control the assaulted level, the perceived race group greatly impacted the strip search result. However, when we turn to the prediction model, the best fit line does not make sense in this case since only a small number of points distribute on the line. We believe there should exist other factors joint influent the decision making for the strip search.

In our study of research question 2, our research question was, “Can we predict if an arrested person will be cooperative based on their demographic information?”. Before we performed our tests, we visualized the overall pattern between Race groups and their cooperatives by creating a boxplot. The boxplot for the Indigenous group and Latino group showed a big difference and fueled our further investigation. We selected the indigenous and Latino group to perform Welch’s t-test, as well as the black and white groups, which both had a narrow boxplot. We aimed to compare the means between Indigenous and the other three selected groups, and Latinos with the other groups. From our t-tests, the interesting pattern that we discovered is that the proportion of cooperative individuals in the Latino group is significantly different from all the groups that it was compared with. This further informed our research question and allowed us to move on to the next step which is to perform a logistic regression with race group being our predictor and action upon arrest being the dependent variable. While the main goal is to examine the relationship between race and cooperation, we also included gender into the model as a predictor to control for confounding variables, and also help address potential biases.

Our logistic regression results enabled us to further understand the relationship between race groups and cooperation. The analysis showed that there are significant differences in the probability of being cooperative during an arrest across different racial categories. East/Southeast Asians, Latinos, Middle Easterners, South Asians, and Whites have a higher probability of being

cooperative during an arrest compared to the reference group. These differences suggest that race may play a role in the level of cooperation during an arrest. However, we want to note that we have a low Pseudo R-Squared value, which suggests that our model has a low explanatory power in predicting the probability of being cooperative during an arrest.

In terms of the prediction intervals, they provide a range of probabilities for each racial category, accounting for the uncertainty in the logistic regression model. These intervals can be useful for understanding the variability in cooperation levels across different racial groups. Among the racial categories.

From table 5, we saw that except for Indigenous group, the coefficients for all the racial categories are positive which indicates the odds of individuals in those race groups being cooperative upon arrest is higher than the reference group, which is the black group in this case. This aligns with our finding in the EDA stage, from the boxplots and summary statistics of cooperative scores, which indicates a lower cooperative score for black individuals in comparison with all other race groups except Indigenous group. The coefficient for Indigenous group was negative which suggests the odds of cooperation for Indigenous are lower than black group, though the difference is not statistically significant.

Limitations

As we mentioned in the discussion section, the strong correlation could only encourage us that we may have found a factor greatly impacted the strip search taking decision. However, due to the low degree between the scatter and the best fit line, we still cannot predict the future actions only based on such limited information. The actions at arrested such as assaulted the police, carry the weapon or any other in-cooperativeness may also significantly influent the decision. Besides, the reason of arresting is also a good point to explore in the future research. We found that the original dataset may need a better design when collecting the future data since it is greatly imbalanced. Most of the collected data are binary thus we need to convert them to the continuous data, which may cause extra risk of inaccuracy in the prediction part. Additionally, the Occurrence_Category column has more than 30 groups with some repeated content with different write-ups. We believe a better design of classification may be helpful if we need to explore the hypothesis of whether the arrest reason would impact the strip search decision or not in the future.

For our study, we performed logistic regression to examine the relationship between race groups and the probability of being cooperative upon arrest. While we were able to detect meaningful impact of the race group on the likelihood of cooperativeness given the statistically significant coefficients for the independent variables, our model has limited explanatory power in terms of the low Pseudo R-squared value. This suggests that other factors not included in the model may play a more significant role in determining cooperation levels during arrests. Our model might not include all relevant variables that could influence the probability of being cooperative during an arrest. For example, variables related to socioeconomic status, educational level could also be important factors in cooperation levels during arrests. Furthermore, not that in our dataset, one of the race groups was classified as "Unknown or Legacy". This could indicate potential inconsistencies or inaccuracies in racial classification, which could influence the validity of our findings. We also recognize the potential for sampling bias, meaning our sample may not be representative of the population and thus making the result not generalizable. Our dataset did not provide location data which would be crucial and valuable data for our analysis. The missing location data prevent us from investigating the geographic area where the sample is drawn from. Lastly is causality, we have found a significant relationship between our predictors and the outcome, but keep in mind that these results should be interpreted cautiously, as correlation does not imply causation.

Conclusion

In conclusion, this study aimed to explore the impact of race and demographic information on police actions during arrests. Through the use of statistical analysis and modeling, we were able to uncover meaningful insights such as the significant impact that perceived race has on the strip search results and the relationship between race groups and cooperation during an arrest.

The existing limitations include the need for further research to better understand the complexities of police actions during arrests, the imbalanced nature of the original dataset, potential inaccuracies in racial classification, possible sampling bias, and the issue of causality. Further research is needed to better understand the role of other factors not included in this study, such as socioeconomic status and educational level, in determining cooperation levels during arrests. Ultimately, our study serves as a starting point for future investigations aimed at improving the fairness and impartiality of police actions during arrests.

Reference

Helmer, G. M. (2001). Strip Search and the Felony Detainee: A Case for Reasonable Suspicion. BUL Rev., 81, 239.

Newburn, T., Shiner, M., & Hayman, S. (2004). Race, Crime and Injustice? Strip Search and the Treatment of Suspects in Custody 1. British Journal of Criminology, 44(5), 677-694.

Simons, P. E. (1979). Strip-Search: The Abuse of Women in Police Stations. Barrister, 6, 8.