UNIVERSITY OF TORONTO
FACULTY OF INFORMATION
The iSchool

INF 2178H S LEC0101 20231
Experimental Design for Data Science

Final Project

The Analysis of Arrests and Strip Searches (RBDC-ARR-TBL-001)

Group 39
Miaomiao Yang
Sheng Cheng

April 14th 2023

Table of Contents

List of Figures

List of Tables

List of Equations

Abstract

This project aimed to investigate the potential influence of age groups on strip search frequency in the context of law enforcement, while controlling for the percentage of women. The study employed power analysis, ANCOVA, Kruskal-Wallis test, and logistic regression to analyze the data. The results showed that age groups may have a small to medium effect on strip search percentage, with larger sample sizes needed to achieve sufficient statistical power. The ANCOVA and logistic regression models identified age groups and sex as significant predictors of strip search likelihood. However, the model fit for logistic regression was poor, suggesting that there may be complex and non-linear relationships between the variables. The findings suggest that law enforcement agencies should consider implementing policies and guidelines that take into account the age and gender of individuals who are subject to strip searches, while also being cautious about relying solely on age and gender as factors for conducting strip searches. Collaborations between law enforcement agencies and researchers are needed to conduct larger and more rigorous studies on the impact of age, gender, and other factors on the use of strip searches in law enforcement settings.

Introduction

The necessity of continues analyzing Arrests and Strip Searches data (RBDC-ARR-TBL-001) in the Toronto area stems from the need to better understand and address potential biases in policing practices. By scrutinizing arrests and strip searches (RBDC-ARR-TBL-001) dataset, researchers can identify trends or disparities related to age groups and gender-based factors, which can in turn inform policy and training initiatives aimed at promoting more equitable policing. Such analysis can foster transparency, facilitate dialogue, and ultimately lead to improved police-community relations in Toronto area. As a diverse and multicultural city, it is crucial for Toronto to maintain a fair and impartial law enforcement system that ensures the safety and well-being of all its residents, regardless of race or gender.

The midterm project employed two-way ANOVA and Tukey's HSD tests to investigate the influence of race and occurrence category on the frequency of strip searches, as well as the impact of gender and age groups. The findings indicated variations in strip search frequency were observed among different age groups for men and women are subjected to strip searches more often than men, regardless of their age. This final project builds upon the insights gleaned from the midterm project while using a more advanced and rigorous analytical approach, power analysis, ANCOVA, and logistic regression to investigate whether age groups have an effect on the percentage of strip searches after controlling for the percentage of women. This analytical approach allows for a deeper exploration of the potential influence of age on strip search frequency and helps to clarify the role of gender in shaping the outcomes. This study holds significant potential to contribute to the ongoing discussions surrounding policing practices and their potential biases. By understanding the influence of age groups on strip search rates, we can provide valuable insights to law enforcement agencies, policymakers, and communities as they work together to create more equitable and unbiased policing strategies, ultimately aiming to contribute to a safer, more just, and equitable society for all residents of Toronto and beyond.

Background

Besides race, age bias in strip searches could be is another possible concern required identify, as it can lead to unequal treatment of individuals based on their age groups. This bias may manifest in several ways, such as disproportionate targeting of younger individuals or inconsistent application of legal guidelines across different age groups. While the available literature on age bias in strip searches is limited, it is essential to consider the potential impact of such bias in law enforcement settings.

In July 2016, a 17-year-old individual from Toronto alleged that he was subjected to carding by the Toronto police, which resulted in a wrist fracture and an unlawful strip search (Ferreira, 2016). In March 26th 2023, there was a news became a hot search "Police strip-searched children as young as eight." Both two news alarms and demonstrates strip-search getting younger and age bias become a crucial concern. New statistics reveal that in the last four years, about 3,000 young individuals have been singled out for scrutiny in England and Wales. A shocking 24% of them were between the ages of 10 and 15, with a small number even younger than that, with the youngest being just eight years old (Tapsfield, 2023).

Strip-searching individuals, especially minors, can have severe psychological and emotional consequences and may result in long-term trauma, anxiety, and loss of trust in law enforcement officials. It is crucial to address this issue and establish policies and guidelines that ensure law enforcement agencies operate within the bounds of the law and respect the rights of all individuals. The use of strip-searches should be limited and conducted only when absolutely necessary, and age-appropriate methods of investigation and interrogation should be used for minors.

Literature Review

Over the past few years, courts and watchdog agencies have sought to regulate police strip search practices and reduce their overall number (*Table 1*). In the landmark *R. v. Golden (2001) case*, the Supreme Court of Canada categorized strip searches as a distinct type of "personal search," differing from general, pat-down, or cavity searches. Strip searches involve the removal or rearrangement of clothing for a visual inspection of an individual's private areas, and are considered invasive, humiliating, and traumatic. Racialized individuals and women may equate strip searches with sexual assault.

Table 1. Trends in Strip Searches (Level 3 Search), 2013-2019

	2013	2014	2015	2016	2017	2018	2019
Counts	22,607	19,760	20,261	17,808	16,710	15,753	14,676

In the 2020 "Understanding Strip Searches Methodological Report" (Dr. Mai B. Phan, 2022) by the Toronto Police Service, it was found that 1,681 youths aged 17 and under were arrested, and 271 of them were strip searched. Youths are defined according to the Y.C.J.A, which oversees youth justice administration. Due to data protection under the Y.C.J.A, anonymized youth data was used for this analysis, following court authorization.

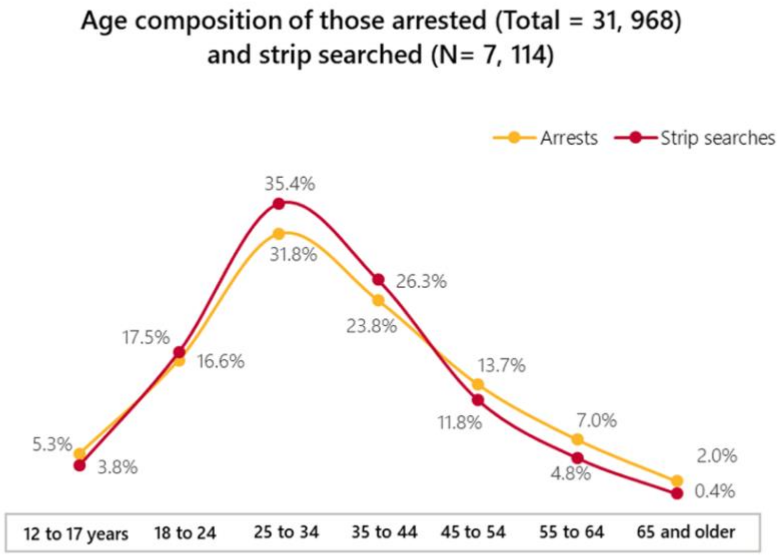

Figure 1. Proportion of age group among those arrested and strip searched, 2020.

1,681 youths aged 17 and under were arrested, and out of them, 271 were strip searched. Age trends (*Figure 1*) reveal that individuals between 12 and 17 years old accounted for 3.8% of strip searches, despite representing 5.3% of total arrests in 2020. In contrast, adults aged 18 to 44 were over-represented in strip searches compared to their proportion in arrests. This indicates that youths were less likely to undergo strip searches during arrests. Similarly, middle-aged and adults aged 54 and above were less likely to be strip searched.

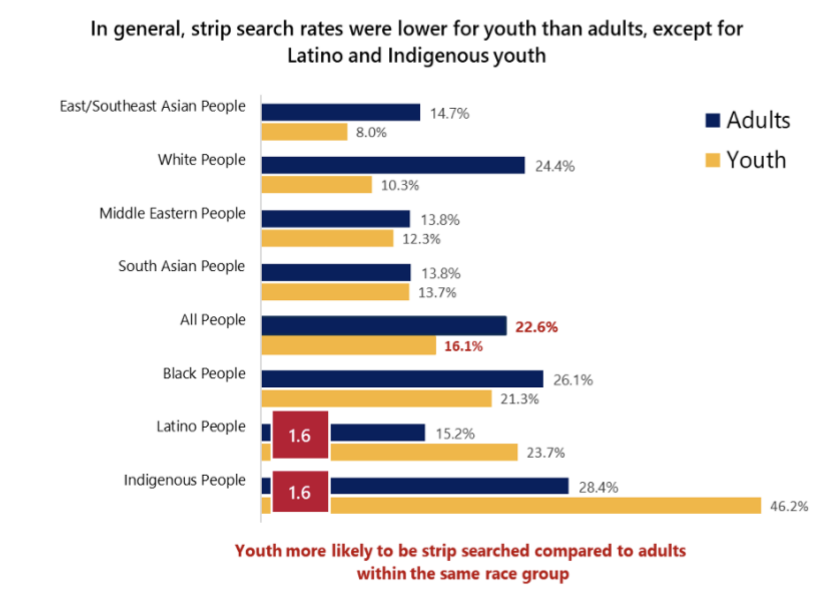

Figure 2. Comparing strip search rates for youth and adults by race group, 2020

The breakdown in strip search rates by race was provided to see if the general trend for youth is reflected across different populations. In general, it is more likely for adults to be strip searched; however, Latino and Indigenous youth are exceptions. *Figure 2* reveals that when comparing different racial groups of adults and youths, Latino and Indigenous youths experienced higher strip search rates, making them 1.6 times more likely to be strip searched compared to adults in their respective racial groups, in relation to arrest proportions. Indigenous youths had the highest strip search rate at 46%, with 6 out of 13 arrested being strip searched, while 9 out of 38 arrested Latino youths were strip searched. South Asian youths had a similar likelihood of being strip searched as adults, at 14% of arrests. The largest gap between youths and adults being strip searched was found among White and East/Southeast Asian individuals, with youths being 58% and 46% less likely to be strip searched compared to adults. To a smaller extent, Black youths were also less likely to undergo strip searches than adults.

In May 2014, the Toronto police chief reported that only 2% of strip and cavity searches uncovered items, and only a fraction posed risks. Although the record was reported in 2014, it remains a subject of concern in today. Given the infrequent discovery of dangerous items and the degrading, constitutionally infringing, and traumatizing nature of strip searches for disproportionately targeted Black and Indigenous people. The decision to persist with strip searches warrants serious contemplation (Lemke, 2022).

Exploratory Data Analysis (EDA)

Data Cleaning

Based on the specific research objective and question, the target variables `'Age_group__at_arrest_'`,`'StripSearch'`, and `'Sex'` were selected from the original dataset. Considering the need for continuous variables with as many category levels as possible, the variables `'Occurrence_Category'` and `'Perceived_Race'` were also selected and grouped together.

Table 2. New Data Frame

	Age_group__at_arrest_	Perceived_Race	Occurrence_Category	StripSearchPercentage	FemalePercentage
0	Aged 17 years and younger	Black	Assault	0.000000	34.313725
1	Aged 17 years and younger	Black	Assault & Other crimes against persons	7.608696	45.652174
2	Aged 17 years and younger	Black	Break & Enter	11.111111	14.814815
3	Aged 17 years and younger	Black	Crimes against Children	0.000000	0.000000
4	Aged 17 years and younger	Black	Drug Related	41.860465	20.930233

To prepare the data for analysis, three new columns were created. `'StripBinary'` was assigned binary values indicating whether a strip search occurred or not; column `‘Sex'` represents the sex variable as 1 for female and 0 for male; and the `'Occurrences'` column counts the total number of observations. The data was then grouped based on age group, perceived race, and occurrence category using the ‘*groupby ()*’ method and aggregations are performed on the binary strip search variable, occurrence count, and sex variable by applying *'agg ()'* function. This created a new data frame which was then used to calculate the percentage of strip searches `'StripSearchPercentage'` and the percentage of females `'FemalePercentage.'` 'Finally, the resulting data frame (*Table 2*) was cleaned by dropping unnecessary columns and resetting the index.

Descriptive Statistics

Descriptive statistics are important in analyzing and understanding a dataset, as they allow us to summarize and describe the characteristics and patterns of the data. It can be useful in two ways, one of which is to provide basic information about variables in a dataset and another of which is to highlight potential relationships between variables.

Table 3. Descriptive Statistics of Percentage of Strip Search and Females

	Strip Search Percentage	Female Percentage
count	1477.00	1477.00
mean	7.95	16.21
std	14.76	18.70
min	0.00	0.00
25%	0.00	0.00
50%	0.00	12.50
75%	11.11	23.08
max	100.00	100.00

Table 3 presents summary statistics for the variables `'StripSearchPercentage'` and `'FemalePercentage'`, which were calculated from a dataset of 1477 observations. The mean `'StripSearchPercentage'` is 7.95%, which indicates that, on average, about 7.95% of the observations in the dataset involved a strip search, while the mean `'FemalePercentage'` is 16.21%, which indicates that, on average, about 16.21% of the observations in the dataset were female. The standard deviation measures the variability or spread of the data. A larger standard deviation indicates a greater amount of variation in the data. In this case, the standard deviation for `'StripSearchPercentage'` is 14.76% and the standard deviation for `'FemalePercentage'` is 18.70%. The minimum `'StripSearchPercentage'` and `'FemalePercentage'` are both 0%, while their maximum values are 100%. The 25th percentile for both variables is 0%, indicating that 25% of the data falls at or below this value. The median `'StripSearchPercentage'` is 0% while the median `'FemalePercentage'` is 12.5%, representing the middle value that separates the lower 50% of the data from the upper 50%. The 75th percentile value for `'StripSearchPercentage'` and `'FemalePercentage'` is 11.11% and 23.08%, respectively, indicating that 75% of the data falls at or below this value.

Table 4 shows the count and percentage of observations in different age groups. The age ranges from 17 years old and younger to 65 years old and older. The highest count is in the "Aged 25 to 34 years" group with 235 observations, followed by "Aged 18 to 24 years" and "Aged 35 to 44 years" with 233 and 232 observations, respectively. "Aged 65 years and older" group has the lowest count with 157 observations. The "Aged 17 years and younger" group has the lowest percentage of observations at 10.63%, while "Aged 25 to 34 years" has the highest percentage with 15.91%. Overall, this table provides an overview of the distribution of observations across different age groups.

Table 4. Count and Percentage of Different Age

Age Groups	Count	Percentage
Aged 17 years and younger	181	10.63
Aged 25 to 34 years	235	15.91
Aged 18 to 24 years	233	15.78
Aged 35 to 44 years	232	15.71
Aged 45 to 54 years	226	14.42
Aged 55 to 64 years	213	12.25
Aged 65 years and older	157	10.63

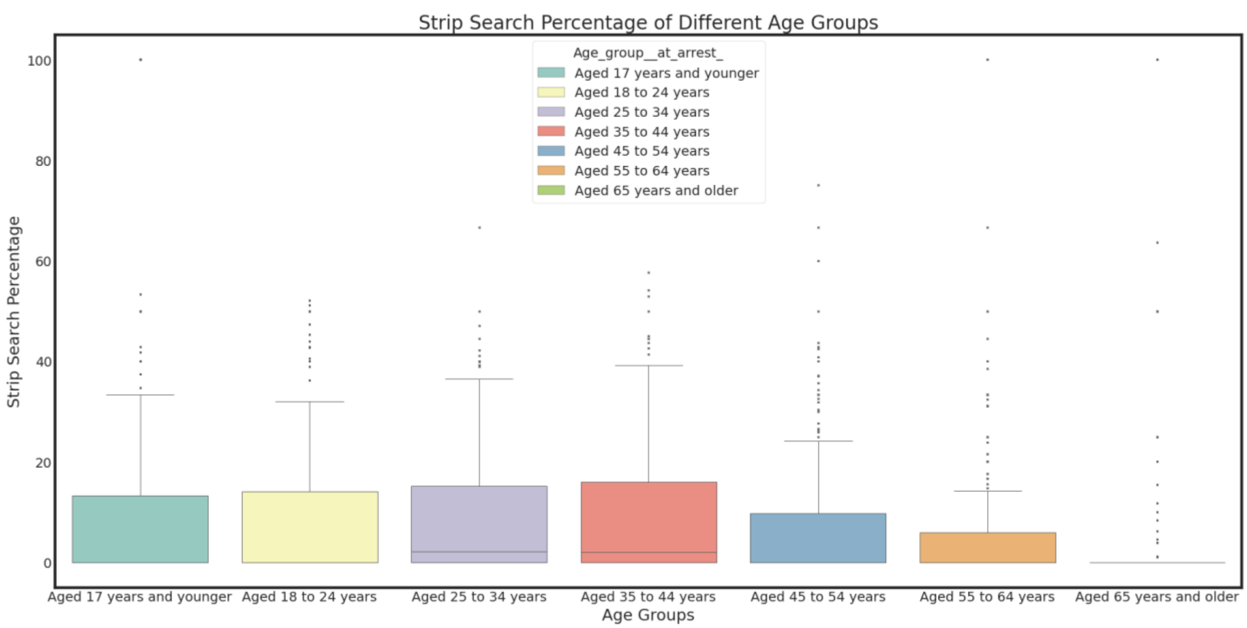

Figure 3. Strip Search Percentage of Different Age Groups – Boxplot

Figure 3 illustrates the distribution of strip search percentage over different age groups. As we can see, the median of most age groups is 0 and heavily right skewed with some outliers. The boxplots of “aged 45 to 54 years” and “aged 55 to 64 years” are comparably short to those of “aged 17 years and younger” and “aged 18 to 24 years.” The median of the boxplots of “aged 25 to 34 years” and “aged 35 to 44 years” are a little greater than 0, but also show the pattern of right skewed.

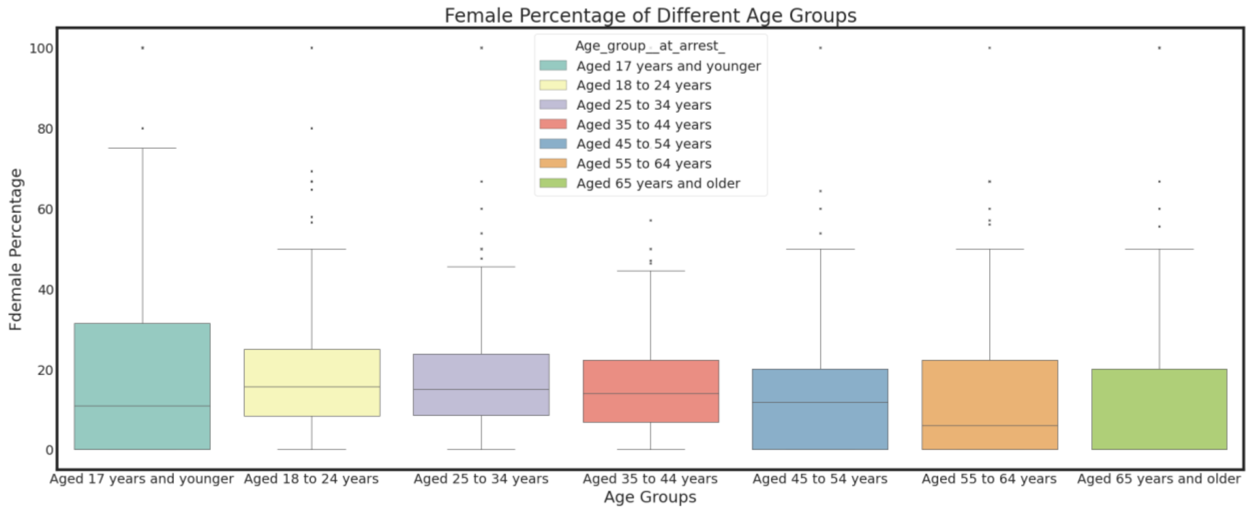

Figure 4. Female Percentage of Different Age Groups – Boxplot

The boxplots of female percentage of those (*Figure 4*) who "aged 18 to 24 years" shows similar patterns, with similar median. Compared to the previous one, the boxplot of "aged 17 and younger" has smaller median and longer interquartile ranges which implies the data is more dispersed. The median of "aged 45 to 65" and older illustrate a decreasing trend. The boxplot of "aged 45 to 54 years" shows left skewed and the boxplot of "aged 55 to 64 years" and older show right skewed.

T – Test

Before performing statistical methods, first, determine if there is a significant difference between the two groups "aged 34 years and younger" (LT35) and "aged 35 year and older" (GT 35). If there is no significant difference between the two groups, then the chosen variable may not be necessary to conduct further complex analysis. Hypothesises were raised and listed. Importing *scipy.stats* module and use *ttest_ind()* function to successfully performing the t-test.

Ho: There is no mean difference of the strip search percentage between the two groups "aged 34 years and younger" and "aged 35 year and older."

Ha: There is a mean difference of the strip search percentage between the two groups "aged 34 years and younger" and "aged 35 year and older."

Table 5. T-Test Result

T-statistic	2.5550952924
p-value	0.0107154984871

In this case, the T-statistic has a value of 2.555. This indicates that the difference between the two-sample LT35 and GT35 means is 2.555 standard deviations away from the expected value under the null hypothesis. The p-value is a measure of the evidence against the null hypothesis. The p-value is 0.0107, which is below the commonly used threshold of 0.05. This suggests that there is strong evidence against the null hypothesis and that the difference between the two-sample means is statistically significant.

Methodology

Dataset Description

The studied dataset Arrests and Strip Searches (RBDC-ARR-TBL-001) contains information on over 600,000 arrests on arrests and strip searches conducted by the Toronto Police Service between 2020 and 2021. The variables in the dataset include demographic information such as age, gender, and perceived race, as well as information on the location and reason for the arrest, actions taken by the police, and whether a strip search was conducted. Besides demographic information, it also includes variables such as the year of the arrest, event ID, arrest ID, and person ID, as well as variables related to actions taken during the arrest, such as whether the person was combative or resisted, and whether they were cooperative. The dataset also includes information on whether items were found during the search. By analyzing this dataset and answering the research questions, we can better understand potential disparities in policing practices and inform efforts to promote more equitable and just systems.

Statistical Methods

Power analysis was the first test would be conducted to determine the appropriate sample size needed to achieve the desired level of statistical power. Following the power analysis, ANCOVA was used to investigate the effect of age group on strip search percentage while controlling for the percentage of females. ANCOVA is a statistical technique that allows for the analysis of the relationship between a dependent variable and one or more independent variables while controlling for the effects of one or more covariates. In this case, age group was the independent variable, strip search percentage was the dependent variable, and the percentage of females was the covariate. However, if the homoscedasticity and normality of ANCOVA not met, it may be necessary to employ non-parametric alternatives such as the Kruskal-Wallis test. Finally, logistic regression was used to further explore the relationship between age group, the percentage of females, and the likelihood of a strip search occurring.

Power Analysis

This section consists of two parts: the calculation of *Cohen's D* and a *Power analysis.* The first part calculates the effect size across different age groups at arrest. The second part uses the effect size to perform a power analysis and to plot power curves.

Cohen's D

As there are seven age groups, the measurement of effect size to quantify the differences across seven samples was done by two functions: *'pooled_standard_deviation' ()* and *'Cohens_d' ().*

'pooled_standard_deviation' () function is defined to accept seven samples and calculate the pooled standard deviation of these seven samples. The sample sizes of each age group are calculated and stored in variables $n_1, n_2, n_3, n_4, n_5, n_6, and\ \ n_7$. The variances of each sample are determined using *NumPy's var ()* function with *ddof = 1,* and then stored in variables $var, var_2, var_3, var_4, var_5, var_6, and\ \ var_7$. The pooled standard deviation can be obtained by following equation

Equation 1. Pooled Standard Deviation

$$Pooled\ \text{Standard Deviation} = \sqrt{\frac{(n_1-1)var_1+(n_2-1)var_2+(n_3-1)var_3+(n_4-1)var_4+(n_5-1)var_5+(n_6-1)var_6++(n_7-1)var_7}{n_1+n_2+n_3+n_4+n_5+n_6+n_7-7}}$$

There is one thing should keep in mind that is the samples come from populations with equal variances.

'Cohens_d' () function then uses the pooled standard deviation and the means of each group to calculate Cohen's d for each group, which is a measure of the standardized difference between two means. The calculation of mean of each sample using the *NumPy mean ()* function. Cohen's d is calculated for each sample by subtracting the mean of the reference group (sample 7) from the mean of the sample, and then dividing this difference by the pooled standard deviation. The results are returned as a list of Cohen's d values for each sample.

Equation 2. Cohen's D for Independent Sample

$$Cohen's\ D = \frac{u_n - u_7}{Pooled\ Standard\ Deviation}, where\ n\ can\ be\ 1,2,3,4,5, and\ 6.$$

Finally, subset of the original dataset 'df_grouped', where each subset contains observations for individuals in a specific age group. These subsets are created using boolean indexing, where the *isin()*

method is used to check if the value in the `'Age_group__at_arrest_'` column matches the specified age group. The subsets are then assigned to their respective variables.

Power analysis

Before performing an ANCOVA to examine the potential differences in strip search percentage (outcome variable) across age groups (seven level explanatory variable), we computed the effect size of the explanatory variable using Cohen's D metric. There were six paired combinations of the levels in total, with level 7 representing the reference group of individuals "aged over 65." For all other groups, the effect size was calculated relative to the reference group.

Conduct a power analysis using the *TTestIndPower* class from the *statsmodels* library. It calculates the effect size (Cohen's D) across different age groups and sets the significance level (alpha) to 0.05 and desired statistical power to 0.8. The ratio of sample sizes between two groups is determined through *len()* function and the required sample sizes for groups are calculated using the *solve_power ()* method. Finally, power curve was plotted using *analysis,plot_power ()* method for different effect sizes and sample sizes to visualize how statistical power changes as a function of sample size and effect size.

ANCOVA

To address the research question, "Do age groups affect the percentage of strip searches after controlling for the percentage of women?", an Analysis of Covariance (ANCOVA) is employed.

The null hypothesis (Ho) states that the means of strip search percentages are the same across all age group levels after controlling for the female percentage, while the alternative hypothesis (Ha) posits that the means of strip search percentages differ across age groups after controlling for the female percentage.

Using the *Pingouin* library, the *ancova()* function is called with the specified dependent variable `'StripSearchPercentage'`, covariate `'FemalePercentage,'` and independent variable `'Age_group__at_arrest_'` from the *'df_grouped'* dataset. The resulting F-value, p-value, and partial eta-squared (np2) effect size are obtained to determine whether the relationships under investigation are statistically significant, ultimately answering the research question and shedding light on the validity of the null and alternative hypotheses. Finally, Checking the assumptions of the ANCOVA model and the validity of the results through boxplot, histogram, and scatterplot.

Boxplot is used to check homoscedasticity, which refers to whether the variances of the dependent variable ‘`StripSearchPercentage`’ are equal across different age groups. Homoscedasticity is important for the validity of the ANCOVA model because it assumes that the variances of the dependent variable are equal across different levels of the independent variable. The Levene's test for homogeneity of variance was conducted using the *pg.homoscedasticity ()* function on the *df_grouped* dataset, with the dependent variable as ‘`StripSearchPercentage`’ and the grouping variable as `'Age_group__at_arrest_'`.

Histogram is used to check the normality of the residuals of the model, which is important because the ANCOVA model assumes that the residuals are normally distributed. If the residuals are not normally distributed, the model may not accurately capture the relationships between the dependent and independent variables. *pg.normality ()* function was used to test the normality of the residuals from the OLS regression model.

Scatterplot is used to check the relationship between the dependent variable and the covariate, which is important because ANCOVA assumes that there is no significant correlation between the dependent variable and the covariate. If there is a significant correlation between the dependent variable and the covariate, the ANCOVA results may be biased.

Kruskal-Wallis test

As mentioned above, if the homoscedasticity and normality of ANCOVA is not met, it may be necessary to employ non-parametric alternatives such as the Kruskal-Wallis test. Such tests do not assume normality or homoscedasticity, although they have their own unique assumptions and limitations that must be taken into account.

Employing the Kruskal-Wallis test, the values of the `'Age_group__at_arrest_'` variable will be converted to numeric. The *rankdata ()* and kruskal *()* functions from the *scipy.stats* module was used to perform the Kruskal-Wallis test on the grouped data.

The *rankdata ()* function assigns a rank to each value of the 'StripSearchPercentage', 'FemalePercentage', and 'Age_group__at_arrest__numeric' variables. These rank values are then stored in new variables 'ranked_stripsearch', 'ranked_female', and 'ranked_age', respectively. The *Kruskal ()* function is used to compute the Kruskal-Wallis test statistic and associated p-value. This function takes as input the ranked values of the three variables and returns a tuple containing the test statistic and p-value.

Logistic Regression

Four major steps involve logistic regression: data split, fit the model, evaluate the model, and make predictions. By splitting the data, the model can be trained on one set of data and tested on another set of data, allowing for a more accurate evaluation of the model's performance on unseen data. Once the data is split, the logistic regression model is fit to the training data and then evaluated on the testing data to see how well it performs on new data. After the model has been evaluated, it can be used to make predictions on new data. This involves providing the model with new input data and using the trained model to predict the corresponding output.

Preparing the data for logistic regression analysis by mapping categorical variables to numerical values. First, selects the independent variables 'Age_group__at_arrest_' and 'Sex'and dependent variable 'StripSearch' from the original data frame *'df'*. Then, two dictionaries 'category_to_number1' and 'category_to_number2' are created to map the categories of 'Age_group__at_arrest_' and 'Sex' to numeric values through *map ()* method. Next, new columns 'age_group_num' and 'sex_num' are added to the independent variable data frame *'x'* with numerical values based on the dictionaries. The original columns are then dropped, and the new columns are renamed to the original column names. Finally, the data are split into training and testing datasets 不 by *train_test_split ()* function with a test size of 0.2 and a random state of 123.

Dropping the rows with missing values from the training dataset and then fits a logistic regression model using the *sm.*OLS () function from the *statsmodels* library. The *OLS.summary ()* method is used to print out the results of the regression analysis, including the coefficients, standard errors, t-values, and p-values for each predictor variable. Finally, calculate the odds ratios for each predictor variable with the *np.exp(OLS.params*).

To assess the effectiveness of a logistic regression model, the first step involves using the trained logistic regression model (OLS) to make predictions on the test data *'x_test'*. Next, a dataframe '*pred_data'* is created by concatenating the predicted values and actual values *'y_test'* of the dependent variable `'StripSearch'`. The code then replaces any missing values *NaN* in the prediction column with 0, renames the prediction column as `'pred'`, and applies the *round ()* function to the predicted values, rounding them to either 0 or 1. Lastly, the *accuracy_score ()* function from *scikit-learn* package is utilized to compute the accuracy of the model's predictions, and the *confusion_matrix ()* function is used to generate a confusion matrix that displays the number of true positive, true negative, false positive, and false negative predictions.

In order to make predictions on the model, uses the *get_prediction ()* function from the *statsmodels* library. The function generates a summary frame that contains the mean prediction, confidence interval boundaries, and prediction interval boundaries for each observation. The function is called twice, once with an alpha value of 0.05 to obtain the confidence interval boundaries and once with an alpha value of 0.1 to obtain the prediction interval boundaries. A new dataframe *'df_assess'* is then created that includes the lower and upper boundaries of the prediction intervals. By comparing the predicted values to the actual values and analyzing the prediction intervals, one can assess its usefulness for predicting future observations.

Results/Findings

Power Analysis

From Table 6, the effect sizes range from 0.16 to 0.42, indicating small to medium effect sizes. These effect sizes can be interpreted as follows: a value of 0.2 is considered a small effect size, 0.5 is considered a medium effect size, and 0.8 is considered a large effect size. Therefore, the effect sizes obtained in this analysis suggest that there may be some meaningful differences in the strip search percentage between different age groups, but the effect sizes are not large enough to be considered substantial.

Table 6. Effective Size for Each Pairwise Comparison

Group n vs. Group 7: Aged 65 years and older	Effective Size
Group 1: Aged 17 years and younger	0.41953185167744278
Group 2: Aged 18 to 24 years	0.29961997782253758
Group 3: Aged 25 to 34 years	0.33276633971789188
Group 4: Aged 35 to 44 years	0.38613208979962871
Group 5: Aged 45 to 54 years	0.22481163414242139
Group 6: Aged 55 to 64 years	0.16073260149802326

For each age group comparison, a sample size larger than the actual sample size is required to achieve the desired statistical power of 0.8 and a significance level of 0.05. For instance, to detect a significant difference between the strip search percentage of the age group "Aged 17 years and younger" (LT17) and the reference group "Aged 65years and older" (GT65) with a power of 0.8, a sample size of 97 is required for "Aged 17 years and younger" (LT17), while the actual sample size for "Aged 17 years and younger" (LT17) is 181. Similarly, for the age group B1824, a sample size of 218 is required, while the actual sample size is 233. The results indicate that larger sample sizes are required for the analysis to achieve sufficient statistical power.

Table 7. Sample Size and Actual Size of Age Group Comparison

Effect Size	Age group comparison	Sample Size	Actual Size
0.42	Aged 65 years and older	84.177	157
	Aged 17 years and younger	97.044	181
0.30	Aged 65 years and older	147.121	157
	Aged 18 to 24 years	218.339	233
0.33	Aged 65 years and older	119.009	157
	Aged 25 to 34 years	178.135	235
0.39	Aged 65 years and older	89.048	157
	Aged 35 to 44 years	131.588	232
0.22	Aged 65 years and older	263.974	157
	Aged 45 to 54 years	379.987	226
0.16	Aged 65 years and older	528.557	157
	Aged 55 to 64 years	717.087	213

We use the effect sizes of 0.2, 0.3, and 0.4 for plotting the power curves as they are commonly used benchmarks for small, medium, and large effect sizes, respectively. The power curves plotted based on these effect sizes demonstrate that when the effect size is 0.4, the power increases to nearly 1 with a sample size of approximately 200. Conversely, when the effect size is 0.2, the power increases to 0.6 with a sample size of 250. These results suggest that a larger sample size is required to achieve a high level of power when the effect size is small, and that smaller sample sizes may be sufficient to detect larger effect sizes with high power.

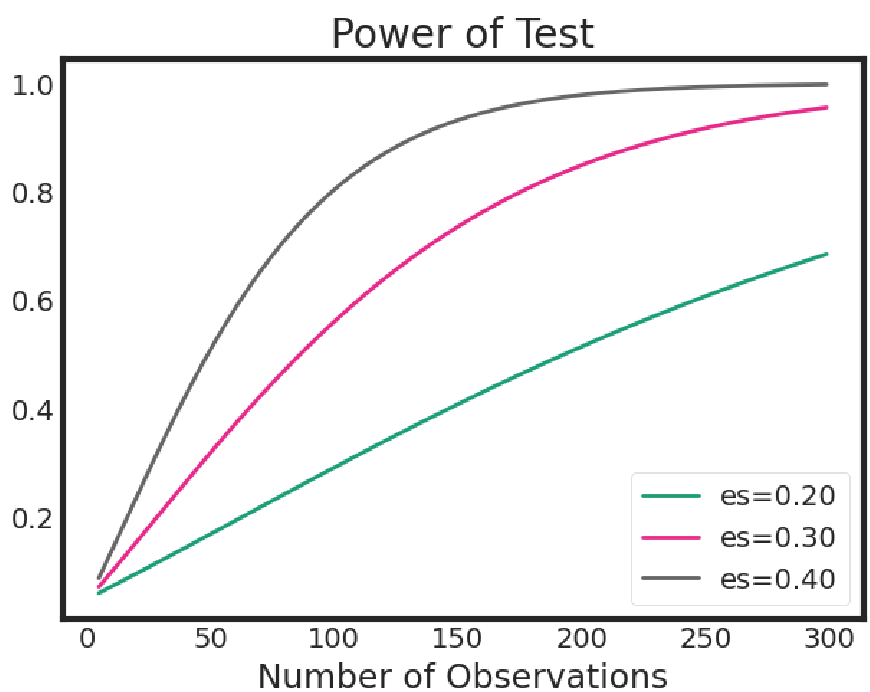

Figure 5. Power Graph

ANCOVA

The ANCOVA results are summarized in the Table (). The p-value for `'Age_group__at_arrest_'` (p-unc = 0.0005) is less than the typical significance level of 0.05. Therefore, we reject the null hypothesis (Ho) which is the means of strip search percentages are the same across all age group levels after controlling for the female percentage and conclude that there is a statistically significant difference in the means of strip search percentage across age groups, after controlling for the effect of the female percentage. The effect size (np2) for age groups is 0.0162, indicating a small effect. However, the p-value for '`FemalePercentage`' (p-unc = 0.1290) is greater than the typical significance level of 0.05. This suggests that there is not a statistically significant relationship between the percentage of women and strip search percentage after controlling for age groups. The effect size (np2) for the female percentage is 0.0016, which can be considered negligible.

Table 8. The ANCOVA Results

Source	SS	DF	F	p-unc	np2
Age_group__at_arrest	5207.1239	6	4.0354	0.0005	0.0162
FemalePercentage	496.0944	1	2.3068	0.1290	0.0016
Residual	315923.7582	1456	NaN	NaN	NaN

Based on Levene's test result, the p-value is less than 0.05, indicating that there is evidence to reject the null hypothesis that the variances of the dependent variable are equal across all levels of the independent variable. In other words, the assumption of homoscedasticity is violated in this case. This means that the variances of the dependent variable are different across different groups of the independent variable. Therefore, it may be necessary to use alternative statistical methods that do not require the assumption of homoscedasticity, such as Welch's ANOVA or the Kruskal-Wallis test.

Table 9. Levene's Test Result

	W	pval	equal_var
levene	3.8280	0.0009	False

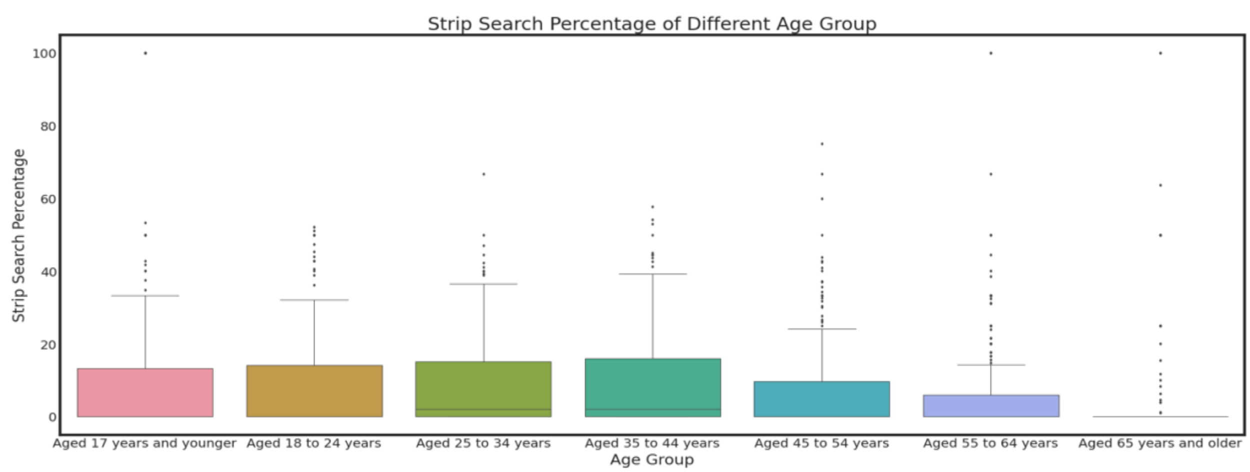

Figure 6. Strip Search Percentage of Different Age Groups by Boxplot

The output of the normality test shows the p-value is 0.0000, which is less than the typical threshold of 0.05 for rejecting the null hypothesis of normality. Therefore, the residuals are not normally distributed, as indicated by the "False" value in the "normal" column. From the Figure 8, the scatter plot does not exhibit any clear or strong relationship between these two variables.

Table 10. Test of Normality

	W	pval	equal_var
O	0.6724	0.0000	False

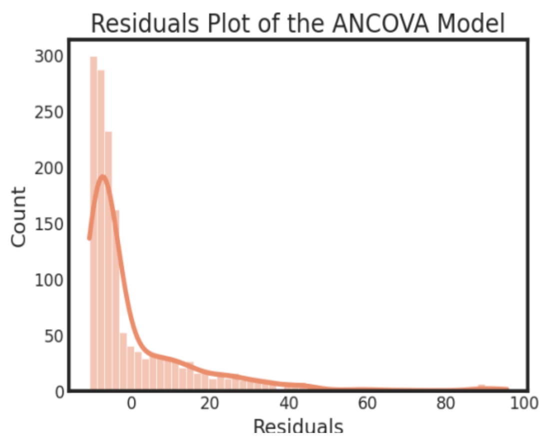

Figure 7. Residuals Plot of the ANCOVA Model.

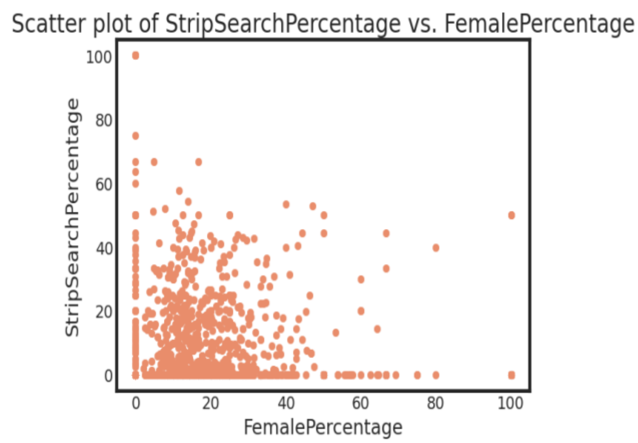

Figure 8. The Scatter Plot of Strip Search Percentage vs. Female Percentage

Overall, it appears that there are complex and possibly non-linear relationships between the variables that were not fully captured by the linear regression and ANCOVA models used in this analysis.

Kruskal-Wallis test

In this case, the Kruskal-Wallis test statistic is 0.751 and the associated p-value is 0.687. Since the p-value is greater than the typical significance level of 0.05, we fail to reject the null hypothesis. This means that we do not have sufficient evidence to conclude that there is a statistically significant difference in the median values of the `'StripSearchPercentage'` variable across different levels of the `'Age_group__at_arrest__numeric'` variable after controlling for the `'FemalePercentage'` variable.

Table 11. K-W Test Result

	Statistic	pval
Kruskal	0.750968	0.686957

Logistic Regression

The logistic regression model results show that the model is not a good fit for the data, with a low R-squared value of 0.003 and an adjusted R-squared value of 0.002. The F-statistic of 65.96 is statistically significant, with a very low p-value of 2.45e-29, indicating that at least one of the predictor variables is significant. The coefficients for 'Age_group__at_arrest_' and 'Sex' are -0.0104 and -0.0232, respectively. This means that for a one-unit increase in 'Age_group__at_arrest_' and 'Sex', the log odds of a StripSearch decreases by 0.0104 and 0.0232, respectively. The constant coefficient of 0.1834 represents the expected log odds of StripSearch when all other predictor variables are equal to zero. The standard errors of the coefficients assume that the covariance matrix of the errors is correctly specified. The Durbin-Watson statistic of 2.000 indicates that there is no significant autocorrelation among the residuals. The Jarque-Bera (JB) test statistic of 75269.259 and its associated p-value of 0.00 indicate that the errors are not normally distributed, and the skewness and kurtosis values of 2.349 and 6.540, respectively, further support this finding.

Table 12. The Logistic Regression Test Result

Dep. Variable	StripSearch	**R-squared**	0.003
Model	OLS	**Adj. R-squared**	0.002
Method	Least Squares	**F-statistic**	65.96
Prob (F-statistic)	2.45e-29	**Log-Likelihood**	15080.
No. Observations	52208	**AIC**	3.017e+04
Df Residuals	52205	**BIC**	3.019e+04
Df Model	2	**Covariance Type**	Non robust

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1834	0.006	30.529	0.00	0.172	0.195
Age_group__at_arrest_	-0.0104	0.001	-9.771	0.00	-0.012	-0.008
Sex	-0.0232	0.004	-6.481	0.00	-0.030	-0.016

Omnibus	22651.560	**Durbin-Watson**	2.000
Prob(Omnibus)	0.000	**Jarque-Bera (JB)**	75269.259
Skew	2.349	**Prob(JB)**	0.00
Kurtosis	6.540	**Cond. No.**	19.3

Also, the results of odds ratio indicate that 'Age_group__at_arrest_' and 'Sex' are significant predictors of the likelihood of a strip search. Specifically, for a one-unit increase in age group, the odds of being strip searched decrease by a factor of 0.99, holding all other variables constant. Similarly, for a one-unit increase in 'Sex' (where 1=male, 2=female, 3=unknown), the odds of being strip searched decrease by a factor of 0.98, holding all other variables constant. The test accuracy of the logistic regression model is 0.876954921803. This means that the model correctly predicted the outcome (whether a strip search occurred or not) for 87.70% of the observations in the test dataset.

Table 13. Odds Ratio

const	Age_group__at_arrest_	Sex
1.2013	0.9897	0.9771

The output shows the odds ratios for the intercept, `'Age_group__at_arrest_'` and `'Sex'`, which are 1.2013, 0.9897, and 0.9771, respectively. The odds ratio for `'Age_group__at_arrest_'` indicates that for each unit increase in the age group (e.g., from 1 to 2, or from 2 to 3), the odds of a `'stripSearch'` decrease by a factor of 0.9897, holding all other variables constant. Similarly, the odds ratio for `'Sex'` indicates that for each unit increase in the sex variable (e.g., from 1 to 2), the odds of a strip search decrease by a factor of 0.9771, holding all other variables constant. Note that because the odds ratio for `'Sex'` is less than 1, this suggests that females (coded as 2) have lower odds of being strip searched than males (coded as 1).

Table 14. Confusion Matrix

```
Confusion Matrix
 [[11439     0]
 [ 1605     0]]
```

The confusion matrix shows the number of true negative (TN) and false negative (FN) predictions in the first row, and the number of false positive (FP) and true positive (TP) predictions in the second row. In this case, the model correctly predicted 11,439 cases that did not result in a strip search (TN), but it incorrectly predicted that 1,605 cases would not result in a strip search when they actually did (FN). The model did not correctly predict any cases that resulted in a strip search (TP), but it also did not make any false positive predictions (FP), meaning that it did not predict a strip search when there was not one. The overall accuracy of the model's predictions was 87.7%.

Table 15. Table of Mean Prediction, Confidence Interval Boundaries, Prediction Interval Boundaries

	mean	mean_se	mean_ci_lower	mean_ci_upper	obs_ci_lower	obs_ci_upper
42788	0.1394	0.0023	0.1349	0.144	-0.4937	0.772
41707	0.1187	0.0016	0.1155	0.1219	-0.5144	0.7518
46368	0.1162	0.0036	0.1093	0.1232	-0.5169	0.7494
29	0.1291	0.0017	0.1257	0.1324	-0.5041	0.7622
63216	0.0955	0.0033	0.0891	0.1019	-0.5376	0.7286

The Table 15 appears to show five different observations which allowed us to make the interpretations:

- The *'mean'* column shows the central tendency of the variable being measured for each observation. For example, for the first observation (42788), the mean value of the variable is 0.1394.

- The *'mean_se'* column shows the standard error of the mean for each observation. This is a measure of the precision of the mean estimate, with smaller values indicating greater precision. For example, the second observation (41707) has a smaller *'mean_se'* value than the first observation (42788), indicating a more precise estimate of the mean for that observation.
- The *'mean_ci_lower'* and *'mean_ci_upper'* columns show the lower and upper bounds of a 95% confidence interval for the population mean for each observation. This means that there is a 95% chance that the true population mean falls within these bounds. For example, for the first observation (42788), the 95% confidence interval for the population mean ranges from 0.1349 to 0.1440.
- The *'obs_ci_lower'* and *'obs_ci_upper'* columns show the lower and upper bounds of a 95% prediction interval for an individual observation for each observation. This means that there is a 95% chance that an individual observation falls within these bounds. For example, for the first observation (42788), the 95% prediction interval for an individual observation range from -0.4937 to 0.7726.

Table 16. Prediction Intervals for Individual Observations

	obs_ci_lower	obs_ci_upper
42788	-0.3919	0.6708
41707	-0.4126	0.65
46368	-0.4151	0.6476
29	-0.4023	0.6604
63216	-0.4359	0.6268
...	...	...
40468	-0.4255	0.6372
36905	-0.4126	0.65
61457	-0.423	0.6396
50235	-0.4151	0.6476
54261	-0.3919	0.6708

Specifically, the table shows the lower and upper bounds of a 95% prediction interval for each observation, as indicated by the *'obs_ci_lower'* and *'obs_ci_upper'* columns. For each observation, the prediction interval provides a range of values within which we would expect a new, independent observation to fall with 95% confidence. For example, for the first observation (42788), we would expect a new observation to fall within the range of -0.3919 to 0.6708 with 95% confidence. The width of the prediction interval can provide information about the uncertainty or variability of the variable being measured. A wider interval indicates greater uncertainty or variability, while a narrower interval indicates more precise estimates. For example, the prediction intervals for the first and fifth observations are wider than those for the second, third, and fourth observations, suggesting that the variable may be more variable or uncertain for the first and fifth observations.

The above plot shows that the predicted values for observations can range from approximately -0.5 to 0.8, indicating that the model has limited precision in predicting the values for these observations. The wider prediction interval, in comparison to the confidence interval, is expected as it needs to account for the uncertainty in the predicted values, whereas the confidence interval only needs to account for the uncertainty in the mean target value for records.

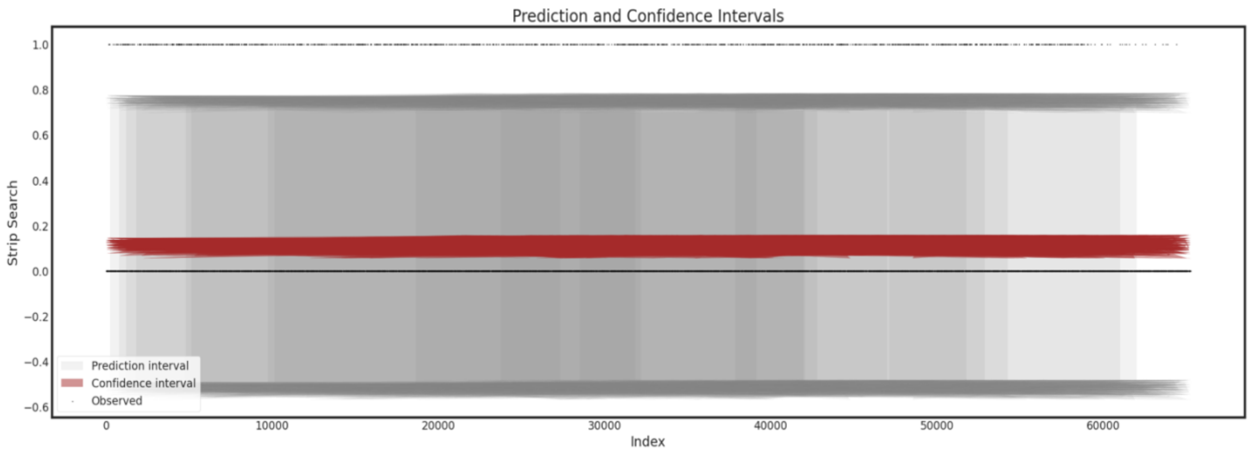

Figure 9. Plot of Prediction and Confidence Intervals

The confidence interval for the mean value of observations 0.1 to 0.15 indicates that if the experiment were repeated multiple times, the true mean target value for records would be within that range 95% of the time. However, this interval does not provide any information about the individual observations themselves but rather the mean value of the population.

Discussion

The results of the power analysis revealed that the effect sizes obtained in this study were small to medium, indicating that there may be some meaningful differences in the strip search percentage between different age groups, but the effect sizes were not large enough to be considered substantial. Furthermore, the study found that larger sample sizes were required to achieve sufficient statistical power when the effect size was small, whereas smaller sample sizes may be sufficient to detect larger effect sizes with high power.

The ANCOVA results indicated a statistically significant difference in the means of strip search percentage across age groups after controlling for the effect of the female percentage. However, the effect size for age groups was small, while the effect size for the female percentage was negligible. The assumptions of homoscedasticity and normality ware violated in this case, suggesting the need for alternative statistical methods that do not require this assumption.

The logistic regression model results indicated that age groups and sex were significant predictors of the likelihood of a strip search, while the model was not a good fit for the data. The odds ratio for 'Sex' suggested that females had lower odds of being strip searched than males.

Conclusion

The project aimed to investigate the impact of age groups on the percentage of strip searches while controlling for the percentage of women. The results suggest that age groups may have a small to medium effect on the percentage of strip searches, with larger sample sizes needed to achieve sufficient statistical power. The ANCOVA and logistic regression models showed that age groups and sex were significant predictors of the likelihood of a strip search, while the model fit for logistic regression was not good. The findings highlight the importance of considering sample size and alternative statistical methods in analyzing complex relationships between variables.

Based on the findings, several recommendations can be made regarding policies and practices related to strip searches in law enforcement setting. law enforcement agencies should consider implementing policies and guidelines that take into account the age and gender of individuals who are subject to strip searches. This may involve providing additional training and resources to officers on how to conduct strip searches in a way that is respectful and appropriate for different age groups and genders. Given the small to medium effect sizes observed in this study, law enforcement agencies should be cautious about relying solely on age and gender as factors for conducting strip searches. Other factors, such as the severity of the offense or the behavior of the individual, should also be taken into account when making decisions about conducting strip searches. To ensure that policies and practices related to strip searches are evidence-based and informed by the best available research, law enforcement agencies should collaborate with researchers and academics to conduct larger and more rigorous studies on the impact of age, gender, and other factors on the use of strip searches in law enforcement settings. This is important as larger sample sizes are needed to achieve sufficient statistical power in studies of this nature.

Reference

R. *v*. Golden, [2001] 3 S.C.R. 679, 2001 SCC 83

Dr. Mai B. Phan, M. D.-P. (2022). *Understanding Strip Searches in 2020 Methodological Report .* Toronto: Toronto Police Service.

Ferreira, V. (2016). Charges dropped against teen who says Toronto police sexually assaulted him in strip search. *NATIONAL POST*.

Garol Tator, Henry Frances . (2006). *Racial Profiling in Canada: Challenging the Myth of a Few Bad Apples.* Toronto: University of Toronto Press.

Lemke, M. (2022, June 17). Strip searches are ineffective, unnecessary and target racialized Canadians. *The Conversation.* Toronto. Retrieved from PHYS.ORG: https://phys.org/news/2022-06-ineffective-unnecessary-racialized-canadians.html

Tanovich, D. (2006). *The Colour of Justice: Policing Race in Canada.* Toronto: Irwin Law.

Tapsfield, J. (2023). Thousands of children aged as young as eight have been strip-searched by police, shock figures show in the wake of 'Child Q' scandal. *Dailymail*.