Exploring the Use of Strip Searches by the Toronto Police Service: An Analysis of Perceived Race, Sex, and Age on Arrests and Searches

Group 58: Jiayuan Guo,Yuepeng Tang

INF2178 - Experimental Design for Data Science

Instructor Shion Guha

University of Toronto, Faculty of Information

April 11, 2023

1. Introduction

1.1 Background

Toronto police officers have the power to arrest a person when they have reasonable grounds to suspect that the person has committed or is about to commit a crime. The Toronto Police Service's duty is to maintain law and public order, and as such they have the authority to detain anyone accused of committing a crime. After an arrest, police officers also have the right to conduct a strip search if they believe it is necessary to protect themselves or to find evidence. A strip search may involve the complete removal of a person's clothing as well as a visual inspection of their internal organs. This practice of arrest and search has been controversial, with some people arguing that it violates their human rights and dignity.

Strip searches have been controversial, with critics arguing that the practice can be unsettling, humiliating, and violate the privacy and dignity of the person being searched. In response to these concerns, the Toronto Police Service has established standards to control strip searches, including the requirement that they be conducted in private and by officers of the same gender as the person being searched. Despite this, strip searches are still widely used, especially when searching members of marginalized groups, such as Aboriginal people, people of color, homeless people, or people with mental health issues. As a result, the discussion of the use of strip searches continues to influence how Torontonians view policing and the upholding of human rights.

1.2 Literature Review

The use of strip searches as policing in Toronto has been controversial, with concerns about their impact on human rights and dignity. Scholars have examined the legal and ethical implications of strip searches and how they are disproportionately used to target marginalized groups. A study by Dorn et al. (2019) found that strip searches are often used against Indigenous people, people of color, and those who are homeless or have mental health issues, and lead to trauma and exacerbate inequalities. The findings are consistent with other studies highlighting the discriminatory and harmful nature of strip searches.

The article (Makin, 2001) reportedly states that the Supreme Court of Canada has banned the routine use of strip searches by police. The ruling found that such searches violated the dignity and rights of detainees and mandated that strip searches could only be conducted when there was reasonable suspicion that the person being searched was in possession of weapons, drugs, or other evidence. The case that the ruling addresses is that of a woman who was arrested and strip searched for a traffic violation, even though she did not pose any threat to the police or the public. The ruling is expected to have a significant impact on policing across Canada as it raises the bar for police conduct and protects the rights of individuals in police custody.

More academics have argued that strip searches in the context of policing are unreasonable. The Court also found that routine strip searches by police were unreasonable and violated the dignity and privacy of detainees. Henderson (2016) explores the Court's decision in detail in his paper and notes the significance of the Court in highlighting the importance of human dignity and the need to balance issues of individual rights and public safety. The decision has implications for policing practices in Canada, setting a higher standard and requiring greater accountability and oversight. The article also discusses the challenges that may be encountered in implementing the decision, particularly where there is reasonable suspicion of concealed evidence.

There may be more reasonable options for examining suspects in the future other than strip searches. In his article, Gorman (2022) examines alternative methods to strip searches, such as the use of body scanners or other technologies. The authors argue that these methods may be less invasive and may better balance individual rights and public safety concerns.

1.3 Introduction to the dataset

This dataset was provided by the Toronto Police Service. This dataset is very informative and gives us valid information to analyze and explore the phenomenon of strip surveys.

By briefly reviewing the information in the dataset, we found that the dataset has a total of 62016 records with a total of 25 fields. These fields relate to the time of arrest, personal information such as individual ID, race, gender, and various details of the arrest activity, such as whether the individual was armed, whether the inmate assisted in the escape and many other information about the arrested and strip-searched individuals. Some of the fields in the dataset contain serious missing values, so if we subsequently use the relevant missing value fields, we need to process them effectively first, depending on the actual situation. The field types in the dataset are mainly numeric, with a few fields of float type.

1.4 Research Problem

This study focuses on the effect of the variables "Perceived_Race", "Sex" and "Age" on the number of strip searches and arrests.

2. Exploratory Data Analysis

2.1 Data Visualization

2.1.1 Time of Arrest Analysis

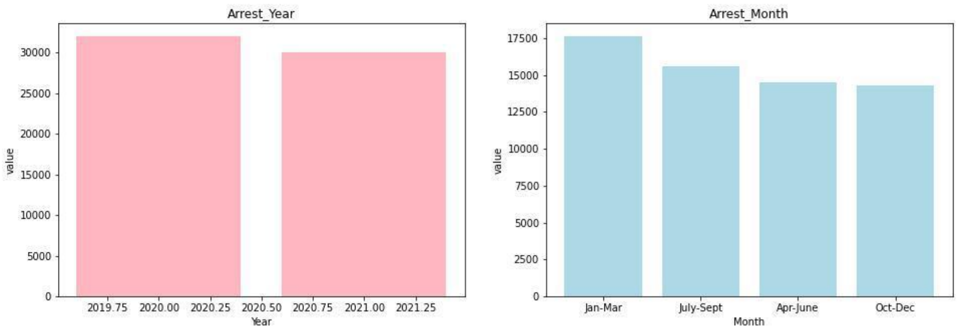


Figure 1-Time of arrest

The chart above shows when the suspects were arrested, in terms of both year and month. It can be seen that the number of arrests in 2020 and 2021 is relatively close, with slightly more in 2020. January to March is the month with the highest number of arrests, followed by July to September, then April to June, and October to December is very close to April to June. This also reflects the high incidence of crime from January to March.

2.1.2 Gender Analysis of Arrestees

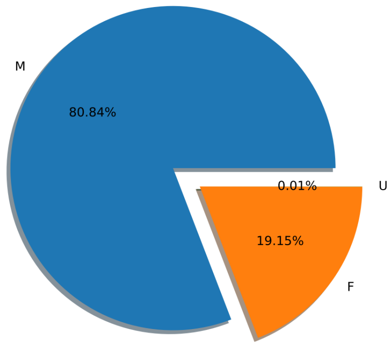
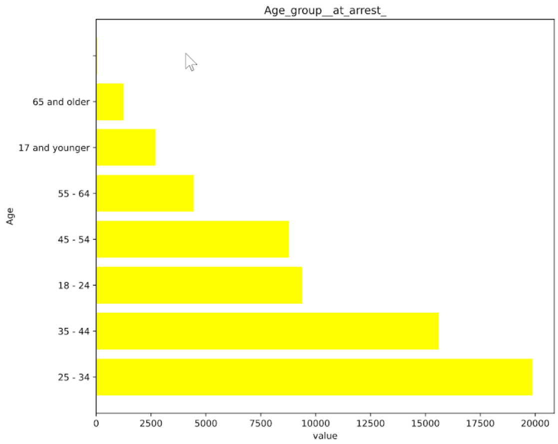


Figure 2- Gender Analysis of Arrestees

The graph above shows the number of arrests by gender. As can be seen, the proportion of males in the number of arrests far exceeds the proportion of females, with the number of males being about four times the number of females, a very large difference. In the subsequent analysis, gender is a variable that we should focus on.

2.1.3 Analysis of the Age of Arrestees



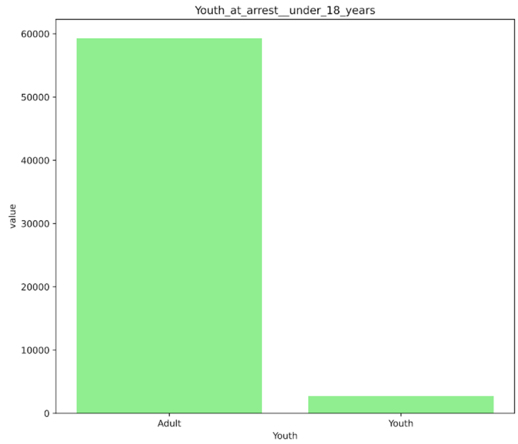


Figure 3-Age analysis of arrestees

The chart above shows the age distribution of those arrested. As can be seen, the 25 to 44 age group has the highest number of arrests and far exceeds other age groups, with the highest number of arrests in the 25 to 34 age group. 17 years old and younger and 65 years old and older have the lowest number of arrests. Therefore, the 25- to 44-year-old group should be the focus of police attention. The variable of age will also be one of the variables we will focus on as we believe there is room for more discussion because the variable of age varies so much between groups.

2.1.4 Analysis of arrestee race

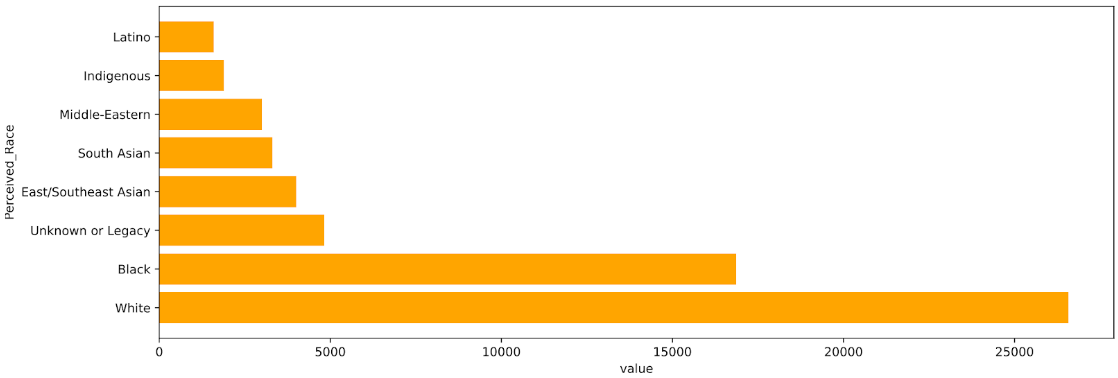


Figure 4-Race analysis of arrestees

The chart above shows the number of arrests by race. As you can see, whites have the highest number of arrests, far more than any other race, followed by blacks. Latinos have the lowest number of arrests. The differences in the number of arrests by race are significant, and the amount of change warrants an in-depth exploration later.

2.1.5 Types of Arrests

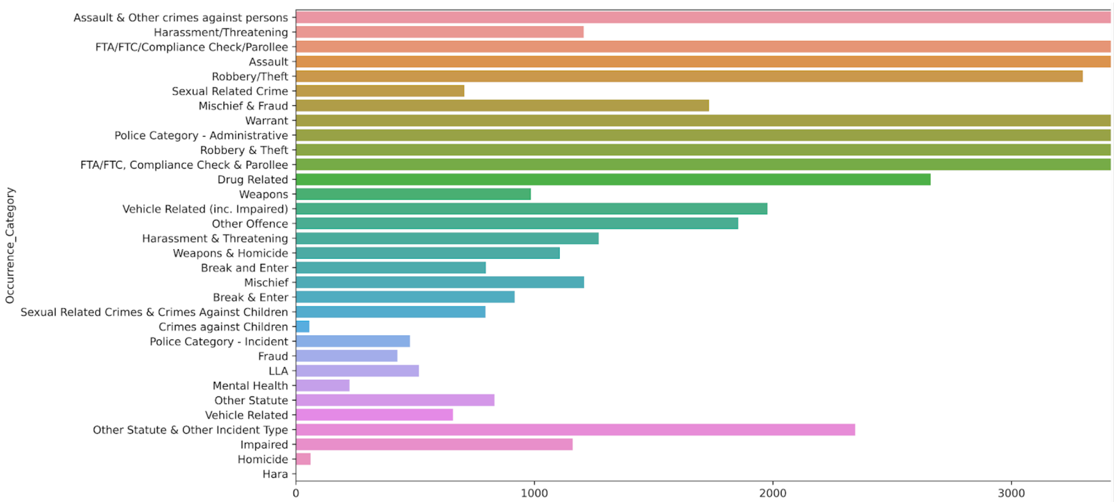


Figure 5-Types of Arrests

The figure above shows the number of arrests in different categories. It can be seen that the number of arrests for assault, theft, etc. is very high, and there are large differences in the number of arrests for some categories, but there are also some categories with very similar numbers of arrests, and if this variable is to be used as a research variable afterward, some data collation is needed for it.

2.1.6 Reasons for arrests

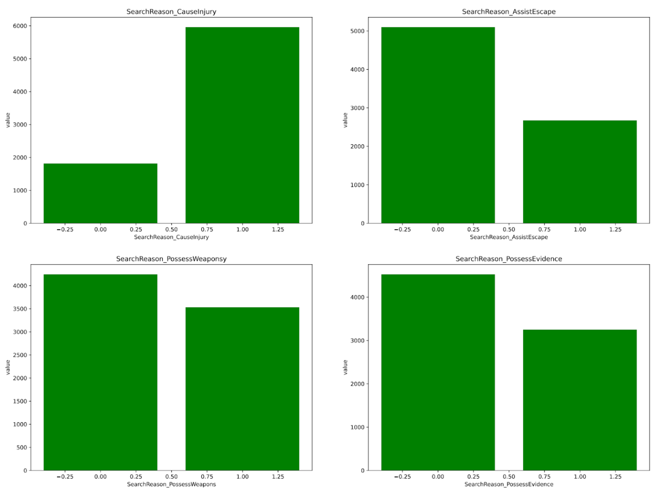


Figure 6-Reasons for arrest

The chart above shows the number of people who were strip searched for various reasons. It can be seen that the highest number of people were strip searched for causing injury and assisting in escape, with the remaining reasons accounting for a lower percentage. This indicates that there is some variation in the number of different arrests on this variable of the reason for arrest, and therefore, this verge is worth exploring in subsequent studies.

2.1.7 Arrest rate and the number of strip searches

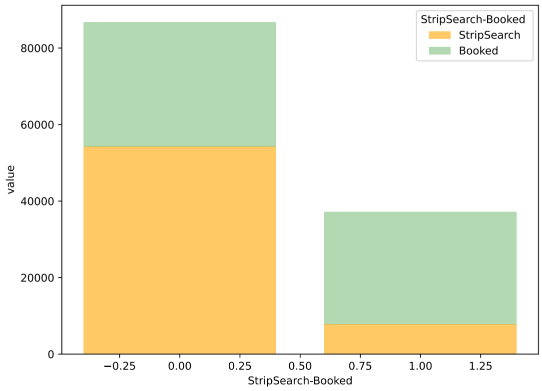


Figure 7-StripSearch and Booked

The graph above shows the number of strip searches and arrest rates for all populations. Where 0 means no strip searches and 1 means strip searches were conducted. It can be seen that there is a large difference in the values taken for whether or not they were strip searched, with the total number of strip searches being 7775. The variable Booked has a smaller difference in the values taken, with a difference of only 3168.

2.1.8 Number of strip searches by age and race

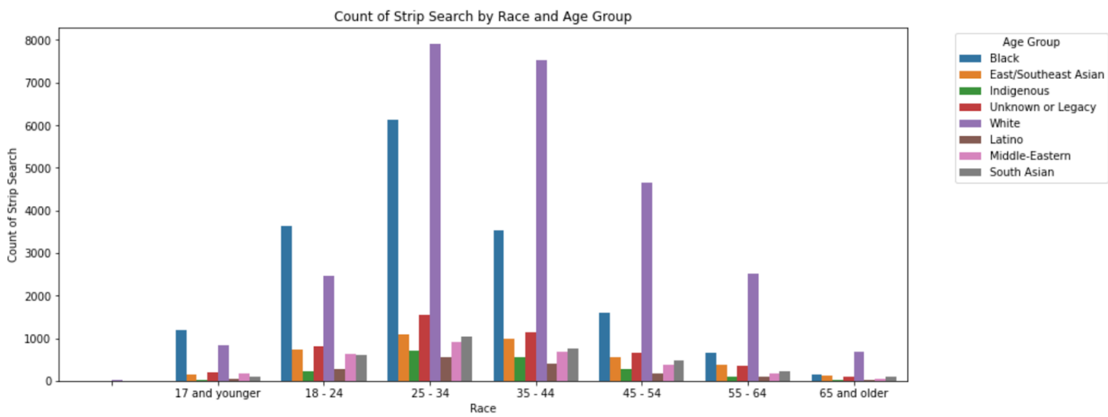


Figure 8-Number of strip searches by age and race

The above graph shows the number of strip searches by age and by race, from which we can get a number of information such as the number of strip searches for each race in different age groups and the number of strip searches for each race in different age groups. From this, it can be seen that the number of strip searches is the highest among people over 25 years old for Whites. The age of number of people who were strip searched for multiple races, including Black, White, and South Asian, are all generally concentrated in the age range of 25 to 34, which is consistent with our previous findings.

2.1.9 Occurrence category and perceived race analysis

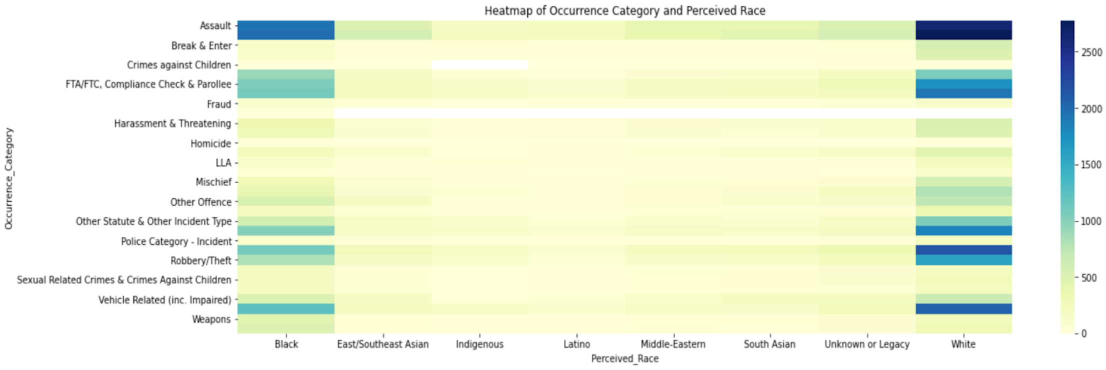


Figure 9-Heatmap of Occurrence Category and Perceived Race

The figure above plots a pivot table on occurrence category and perceived race, the color represents the magnitude of the value, the darker the color, the larger the value. The correlation between race, and occurrence category can be found from this. Overall, Assault is the main occurrence category among all races, followed by White and Black as the main arrested populations. In addition, we can also observe the specific occurrence categories for each race.

2.2 T-tests

With the visual EDA we did above, we see some interesting differences, which we next want to confirm with some t-tests.

2.2.1 Gender and number of strip searches

Null hypothesis: there is no significant difference between the means of the number of strip searches for males and females;

Alternative hypothesis: there is a significant difference between the means of the number of strip searches for males and females;

Table 1: Gender and strip searches

T-Statistic	P-Value
6.447	0.000

The t-test between the sex of the arrested population and the number of strip searches allows us to compare the differences between the two data sets of males and females. The value of the t-test statistic here is 6.447 at the 0.05 level of significance and the p-value of the t-test is 0.000. Therefore we need to reject the original hypothesis, which indicates that there is a significant difference between the means of the number of strip searches for males and females.

2.2.2 Race and the number of strip searches

Null hypothesis: there is no significant difference in the mean of the number of strip searches between whites and blacks;

Alternative hypothesis: there is a significant difference in the mean of the number of strip searches between whites and blacks;

Table 2: Race and strip searches

T-Statistic	P-Value
-2.890	0.004

The previous descriptive analysis showed that Whites and Blacks were the races with the highest number of arrests. A t-test of the means of the number of strip searches for whites and blacks showed whether there was a significant difference between the means of the number of strip searches for whites and blacks. The value of the t-test statistic was -2.890 at the 0.05 level of significance, and the p-value of the t-test result was less than 0.05, indicating that at this point we can reject the original hypothesis that there is a significant difference between the means of the number of strip searches for whites and blacks.

2.2.3 Age and the number of strip searches

Null hypothesis: there is no significant difference between the means of the number of strip searches for youth and adults;

Alternative hypothesis: there is a significant difference between the means of the number of strip searches for youth and adults;

Table 3: Age and strip search times

T-Statistic	P-Value
3.67	0.000

The t-test of the number of strip searches for youth and adults allows us to compare whether there is a significant difference between the mean of the number of strip searches for youth and adults. the value of the statistic of the t-test result is 3.67, and the p-value of the t-test is much less than 0.05, indicating that the original hypothesis can be rejected at this point, that is, there is a significant difference between the mean of the number of strip searches for youth and adults, and the number of strip searches does have a certain relationship, which is consistent with the conclusion we obtained earlier in the data visualization.

2.2.4 Gender and number of arrests

We first grouped the data by age group and calculated the number of arrests for each age group, and then formulated the associated hypotheses.

Null hypothesis: There is no significant difference between the means of the number of arrests for males and females;

Alternative hypothesis: there is a significant difference between the means of the number of arrests for males and females;

Table 4: Gender and number of arrests

T-Statistic	P-Value
4.472	0.000

By performing a t-test on the mean of the number of arrests by gender, the value of the t-test statistic is 4.472 at the 0.05 level of significance, with a p-value of 0.000, which is less than 0.05, indicating that at this point we should reject the original hypothesis that there is no difference in the mean, i.e. there is a significant difference between the mean of the number of arrests of males and females.

2.2.5 Age and number of arrests

To investigate the relationship between age and the number of arrests, we formulated the following hypotheses.

Null hypothesis: there is no significant difference between the means of the number of arrests of youth and adults;

Alternative hypothesis: there is a significant difference between the means of the number of arrests of youth and adults;

Table 5: Age and number of arrests

T-Statistic	P-Value
-14.591	0.000

It can be seen that the value of the t-test statistic my i-14.591 with a p-value of 0.000 at the 0.05 level of significance indicates that at this point we have reason to reject the original hypothesis that there is a significant difference between the means of the number of arrests of youth and adults.

2.2.6 Race and number of arrests

Here we chose whites and blacks to analyze whether there is a difference in the number of arrests by race.

Null hypothesis: There is no significant difference in the mean number of arrests between whites and blacks;

Alternative hypothesis: there is a significant difference between the means of the number of arrests of whites and blacks;

Table 6: Race and number of arrests

T-Statistic	P-Value
2.837	0.000

The positive t-statistic at the 0.05 level of significance indicates that the mean of the White group is higher than the mean of the Black group for the variable Perceived_Race. the p-value is very small (less than 0.05), indicating that we have strong evidence against the hypothesis that there is no significant difference between the means of the number of arrests of Whites and Blacks, i.e., we can conclude that the means of the number of arrests of Whites and Blacks We can conclude that there is a significant difference in the mean number of arrests between whites and blacks.

3. Methods

3.1 Data set

This dataset on arrests and strip searches was provided by the Toronto Police Service. It contains 62016 records, each with a unique personal ID. 25 attributes are included describing the individuals arrested and strip searched, as well as various types of information surrounding the arrests and strip searches. The dataset provides detailed demographic information about the individual, including information about the individual's gender, race, and age, as well as the reason for the strip search and the behavior at the time of the arrest.

3.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is the visualization and statistical analysis of data in the early stages of data analysis to discover patterns, trends, outliers, flaws, etc. EDA is an important part of data analysis that helps analysts better understand and process data, identify potential opportunities and problems, and provide support for business decisions. When performing exploratory data analysis, we must consider which attributes are most relevant to the research question and which charts can most effectively visualize those attributes. Next, when creating charts, make sure they are clear, accurate, and well-labeled. For each chart, we should finally give an explanation in relation to the actual meaning of the fields.

3.3 T-test

The t-test is a common statistical method used to compare whether the difference between two sets of data is significant or not. Specifically, the t-test can be used to compare whether the two sets of sample means are significantly different and help determine whether the two sets of sample means come from the same overall. In addition to this, the t-test can help determine whether the mean of a sample is significantly higher or lower compared to a known value. When it is necessary to compare two groups of sample variances to see if they are significantly different, the t-test can be used to determine if the two groups of sample variances come from the same overall. Here, we take samples from each group and calculate the mean and standard deviation for each group. p-values will be obtained from the t-test. p-values less than 0.05 are usually considered to be statistically significant.

3.4 Analysis of variance

Analysis of variance (ANOVA) is a statistical method used to analyze the variability between multiple sets of numerical data. It can determine whether there are significant differences in the data and also whether these differences are due to random error or to treatment factors (such as different treatments or experimental conditions). In ANOVA, the data are divided into different groups, and then the variance of each group is calculated, and then the variability between the groups is compared according to the magnitude of the variance. By calculating the ratio of the variances, it is possible to determine the significant differences between the groups of data and whether these differences are sufficient to be considered to be due to the influence of treatment factors. ANOVA has a wide range of applications. In experimental design, ANOVA is widely used to compare the efficacy of different treatments, the effect of different manipulation methods, etc.

3.5 Post hoc test

A post hoc test is a method used to test which groups are significantly different from each other after a comparative analysis of multiple groups such as ANOVA. It is often used to analyze differences between groups or conditions that have been found to be significantly different. When conducting post hoc tests, specific methods are used to control the error rate to ensure the reliability of the results. Common post hoc tests include Tukey's test, Dunnett's test, Scheffe's test, Bonferroni correction, etc. The main purpose of post hoc tests is to help researchers identify which groups are significantly different from each other and to compare and interpret these differences. However, it is important to note that post hoc tests are only a tool for analyzing multiple group comparisons and should not be used as a preliminary test to address research hypotheses.

3.6 Logistic Regression

Logistic regression is a common classification algorithm that can be used for both binary and multiclassification problems. The basic idea of logistic regression is to use a logistic function (also called a sigmoid function) to relate the input features to the probability of the output. In a binary classification problem, the logistic function maps the input features to a value between [0,1], indicating the probability that the sample belongs to a positive class. In general, when the value of the output of the sigmoid function is greater than 0.5, the sample is classified as a positive class, otherwise, it is classified as a negative class. Logistic regression has many advantages, such as simple calculation, fast, easy to interpret, etc.

3.7 Power Analysis

Power Analysis in hypothesis testing is an important statistical tool used to evaluate the effectiveness of hypothesis testing. Specifically, efficacy analysis is used to calculate the statistical efficacy of hypothesis testing, that is, the probability that a null hypothesis can be correctly rejected. Efficacy analysis can help researchers determine whether larger sample sizes are needed to increase the efficacy of hypothesis testing. Efficacy analysis plays a very important role in hypothesis testing. It can help researchers determine the minimum sample size, improve the accuracy of the test, avoid false positives and false negatives, and help select the most suitable test method.

3.8 ANCOVA

ANCOVA analysis is a statistical method that combines the techniques of analysis of variance and regression analysis to control one or more covariables of the dependent variable to determine whether the dependent variable is different between different groups. In an ANCOVA analysis, the dependent variable is the primary variable that the researcher wants to compare, while the covariates are other variables that the researcher thinks might have an effect on the dependent variable, which can be quantitative or categorical variables. By controlling the influence of covariables, ANCOVA can increase statistical power and reduce errors, and improve the explanatory power of dependent variables.

4. Results and Discussion

4.1 One-way analysis of variance

4.1.1 Gender and number of strip searches

Null hypothesis: the average number of strip searches for males and females is equal;

Alternative hypothesis: the average number of strip searches for males and females are not equal;

Table 7: One-way analysis of variance - Sex and strip search frequency

F-Statistic	P-Value
41.559	0.000

By conducting a one-way ANOVA test on gender and the number of strip searches, the results showed that the F-statistic value was 41.559, with a p-value much less than 0.05, indicating that at this point we could reject the original hypothesis and obtain the conclusion that the mean number of strip searches for males and females was not equal.

4.1.2 Age and the number of strip searches

Null hypothesis: the average number of strip searches for different age groups is equal;

Alternative hypothesis: the average number of strip searches for different age groups is not equal;

Table 8: One-way analysis of variance - Age and strip search times

F-Statistic	P-Value
13.473	0.000

A one-way ANOVA using different age groups and the number of strip searches showed that there was a statistically significant difference between the means of the different groups.The F-value was 13.473 and the p-value was 0.000, which is less than 0.05, which indicates that we can reject the original hypothesis. That is, there is a significant difference in the mean number of strip searches in at least one age group.

4.1.3 Race and the number of strip searches

Original hypothesis: the average number of strip searches is equal for different racial groups;

Alternative hypothesis: the mean number of strip searches is not equal for different racial groups;

Table 9: One-way analysis of variance - Race and strip search times

F-Statistic	P-Value
23.756	0.000

The F-value measures the ratio of the between-group variance to the within-group variance, and in this case the F-value is 23.756, which means that there is a moderate difference between the groups as far as their means are concerned. the p-value is much less than 0.05, which means that we can reject the null hypothesis and conclude that there is a statistically significant difference between the groups.

4.1.4 Gender and number of arrests

Null hypothesis: the average number of arrests in the different gender groups is equal;

Alternative hypothesis: the mean number of arrests is not equal for the different gender groups;

Table 10: One-way analysis of variance - Sex and number of arrests

F-Statistic	P-Value
20.000	0.000

In the one-way ANOVA test for gender and number of arrests, the F-test statistic was 20.000 with a p-value of 0.000, which indicates that we have reason to reject the original hypothesis that the mean number of arrests is not equal across gender groups.

4.1.5 Race and the number of arrests

Null hypothesis: the average number of arrests is equal across races;

Alternative hypothesis: the average number of arrests is not equal across races;

Table 11: One-way analysis of variance - Race and arrests

F-Statistic	P-Value
164.787	0.000

In the analysis of variance between race and number of arrests, we found statistically significant differences in at least one group. an F-value of 164.787 and a p-value of 0.000 indicate that the probability of obtaining this result by chance is less than 1%, assuming no significant differences in the mean number of arrests across races. Therefore, we reject the null hypothesis that the mean number of arrests across races is equal and conclude that the mean number of arrests across races is not equal.

4.2 Two-factor ANOVA

For the two-factor ANOVA, we can formulate the following hypothesis.

Null hypothesis: the overall means are exactly equal;

Alternative hypothesis: the overall means are not exactly equal;

4.2.1 Race age and the number of strip searches

Table 12: Two-factor analysis of variance - Race, age and strip search frequency

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	36.376257	7.0	47.668208	5.916546e-68
C(youth)	1.780478	1.0	16.332230	5.321629e-05
C(Perceived_Race):C(youth)	2.914035	7.0	3.818613	3.733767e-04
Residual	6758.679027	61997.0	-	-

For the factor C (Perceived_Race), the sum of variance is 36.376, the degree of freedom is 7, and the F-statistic value is 47.668, with a p-value much less than 0.05, indicating that the effect of Perceived_Race on the dependent variable is significant. For the factor youth, the sum of variance is 1.780, the degree of freedom is 1, and the F-statistic value is 16.332, with a p-value less than 0.05, indicating that the effect of youth on the dependent variable is significant. Also, it can be seen that the interaction effect between C(Perceived_Race):C(youth) is also significant. The residual rows show the squared error term and the degrees of freedom, which represent the transformations accounted for by the factors in the dependent variable that were not tested.

4.2.2 Gender race and the number of strip searches

Table 13: Two-factor analysis of variance - Sex, race and strip search frequency

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	28.378434	7.0	37.239834	4.917070e-24
C(Sex)	6.480971	2.0	29.766476	1.198911e-13
C(Perceived_Race):C(Sex)	11.204748	14.0	7.351761	9.400322e-12

Residual	6748.889298	61994.0	-	-

From the above table, it is clear that the effects of both factors Perceived_Race and Sex on the dependent variable are significant and the interaction effect of C(Perceived_Race):C(Sex) is also highly significant, and we should investigate further to understand the nature of the effect.

4.2.3 Gender race and the number of arrests

Table 14: Two-factor analysis of variance -- sex and race and the number of arrests

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	8.274037e+04	7.0	403.765120	8.850064e-260
C(Sex)	2.115680e+03	2.0	36.135106	2.069489e-16
C(Perceived_Race):C(Sex)	5.273246e+03	14.0	12.866469	9.273274e-23
Residual	1.814848e+06	61994.0	-	-

From the above table, it is clear that the effects of both factors Perceived_Race and Sex on the dependent variable num_arrests are significant and the interaction effect of C(Perceived_Race):C(Sex) is also very significant, and we should investigate further to understand the nature of the effect.

4.2.4 Age race and number of arrests

Table 15: Two-factor analysis of variance -- Age, race and number of arrests

	sum_sq	df	F	PR(>F)
C(Perceived_Race)	5.304204e+04	7.0	259.086519	0.000000e+00
C(youth)	5.385880e+03	1.0	184.153208	6.895681e-42
C(Perceived_Race):C(youth)	2.134663e+03	7.0	10.426871	3.739857e-13
Residual	1.813210e+06	61997.0	-	-

For the factor C (Perceived_Race), the F-statistic value is 259.087 with a p-value much less than 0.05, indicating that the effect of Perceived_Race on the dependent variable is significant. For the factor youth, the F-statistic value is 184.153, with a p-value less than 0.05, indicating that the effect of youth on the dependent variable is significant. Also, it can be seen that the interaction effect between C(Perceived_Race):C(youth) is also significant. The residual rows show the sum of squares and degrees of freedom of the error terms, which represent the transformations accounted for by the factors in the dependent variable that were not tested.

4.3 ANCOVA

4.3.1 Number of strip searches, race, and gender

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            StripSearch   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                  0.006
Method:                 Least Squares   F-statistic:                     43.50
Date:                Tue, 28 Mar 2023   Prob (F-statistic):           1.62e-78
Time:                        15:05:20   Log-Likelihood:                -19261.
No. Observations:               62016   AIC:                         3.854e+04
Df Residuals:                   62006   BIC:                         3.863e+04
Df Model:                           9
Covariance Type:            nonrobust
==========================================================================================================
                                             coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------
Intercept                                  0.1222      0.004     32.216      0.000       0.115       0.130
Perceived_Race[T.East/Southeast Asian]    -0.0596      0.006    -10.265      0.000      -0.071      -0.048
Perceived_Race[T.Indigenous]               0.0231      0.008      2.873      0.004       0.007       0.039
Perceived_Race[T.Latino]                  -0.0613      0.009     -7.074      0.000      -0.078      -0.044
Perceived_Race[T.Middle-Eastern]          -0.0692      0.007    -10.561      0.000      -0.082      -0.056
Perceived_Race[T.South Asian]             -0.0680      0.006    -10.822      0.000      -0.080      -0.056
Perceived_Race[T.Unknown or Legacy]       -0.0329      0.005     -6.105      0.000      -0.044      -0.022
Perceived_Race[T.White]                   -0.0084      0.003     -2.587      0.010      -0.015      -0.002
Sex[T.M]                                   0.0261      0.003      7.695      0.000       0.019       0.033
Sex[T.U]                                  -0.0997      0.110     -0.905      0.365      -0.315       0.116
==============================================================================
Omnibus:                    25202.388   Durbin-Watson:                   1.757
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            76287.279
Skew:                           2.240   Prob(JB):                         0.00
Kurtosis:                       6.074   Cond. No.                         115.
==============================================================================
```

Here, the number of strip searches was taken as the dependent variable, race as the independent variable, and gender as the covariable, and ANCOVA analysis was conducted to test whether the independent variable had a significant impact on the explained variable. It can be seen that the P value of the model is far less than 0.05, indicating that the model is very significant. Looking at the specific variables, only Sex[T.U] failed the significance test, and all other variables were significant.

4.3.2 Number of arrests, race, and age

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            num_arrests   R-squared:                       0.032
Model:                            OLS   Adj. R-squared:                  0.032
Method:                 Least Squares   F-statistic:                     253.9
Date:                Tue, 28 Mar 2023   Prob (F-statistic):               0.00
Time:                        15:13:17   Log-Likelihood:            -1.9270e+05
No. Observations:               62016   AIC:                         3.854e+05
Df Residuals:                   62007   BIC:                         3.855e+05
Df Model:                           8
Covariance Type:            nonrobust
==========================================================================================================
                                             coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------------------
Intercept                                  4.2020      0.042     99.230      0.000       4.119       4.285
Perceived_Race[T.East/Southeast Asian]    -1.2186      0.095    -12.798      0.000      -1.405      -1.032
Perceived_Race[T.Indigenous]               3.7567      0.132     28.513      0.000       3.498       4.015
Perceived_Race[T.Latino]                  -0.7163      0.142     -5.041      0.000      -0.995      -0.438
Perceived_Race[T.Middle-Eastern]          -0.5672      0.107     -5.285      0.000      -0.778      -0.357
Perceived_Race[T.South Asian]             -1.1188      0.103    -10.857      0.000      -1.321      -0.917
Perceived_Race[T.Unknown or Legacy]       -0.9659      0.088    -10.915      0.000      -1.139      -0.792
Perceived_Race[T.White]                    0.6537      0.053     12.234      0.000       0.549       0.758
youth[T.Youth]                            -1.4469      0.107    -13.563      0.000      -1.656      -1.238
==============================================================================
Omnibus:                    42638.792   Durbin-Watson:                   0.567
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           667257.095
Skew:                           3.176   Prob(JB):                         0.00
Kurtosis:                      17.761   Cond. No.                         8.15
==============================================================================
```

Here, ANCOVA analysis is conducted with arrest times as the dependent variable, race as the independent variable, and age (youth and adulthood) as the covariable to test whether the independent variable has a significant influence on the explained variable. It can be seen that the P value of the model is far less than 0.05, indicating that the model is very significant. Looking at specific variables, the P values of all variables tested were all less than 0.05, indicating that there were great differences between the groups, and post-hoc analysis was needed.

4.3.4 Result analysis

In the first analysis, ANCOVA was used to test the relationship between race and the number of strip searches, while controlling for the effect of gender. The results showed that the overall model was significant, with a p-value less than 0.05, indicating that the variables had a statistically significant relationship. However, only the gender variable was found to be non-significant. This suggests that race, in combination with gender, has a significant impact on the number of strip searches.

In the second analysis, ANCOVA was used to test the relationship between race, age, and the number of arrests. The results showed that the overall model was significant, with a p-value less than 0.05, indicating that the variables had a statistically significant relationship. Moreover, all specific variables were found to be significant, with p-values less than 0.05. This suggests that both race and age have a significant impact on the number of arrests and that there are significant differences between different racial and age groups.

In both cases, ANCOVA was used to control for the effect of a covariate, while examining the relationship between the independent and dependent variables. These results suggest that race, in combination with gender and age, has a significant impact on the number of strip searches and arrests, respectively. Further post-hoc analysis may be necessary to fully understand the nature of these relationships and to identify potential interventions to address disparities in policing practices.

4.4 Post hoc tests

The results of ANOVA we need to look at the p-value, p-value is less than 0.05, which indicates that there is a significant difference between treatments, and the specific which difference exists between treatments needs to be seen by post hoc test. For the post hoc analysis, we can formulate the following hypotheses.

4.4.1 Age gender and the number of strip searches

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
FAdult	FYouth	-0.0641	0.0	-0.1004	-0.0279	True
FAdult	MAdult	0.0186	0.0	0.0092	0.0281	True
FAdult	MYouth	0.0089	0.7942	-0.0127	0.0306	False
FAdult	UAdult	-0.1113	0.8515	-0.4124	0.1897	False
FYouth	MAdult	0.0828	0.0	0.0473	0.1182	True
FYouth	MYouth	0.0731	0.0	0.0326	0.1135	True
FYouth	UAdult	-0.0472	0.9932	-0.3502	0.2558	False
MAdult	MYouth	-0.0097	0.6903	-0.03	0.0106	False
MAdult	UAdult	-0.13	0.7641	-0.4309	0.171	False
MYouth	UAdult	-0.1203	0.813	-0.4219	0.1814	False

For the results of the post hoc test, we can look at the column of REJECT, and the column of REJECT means that there is a difference between the two treatments if the column is True. From this, we can see that there is a difference between the groups FAdult and FYouth, FAdult and MAdult, FYouth and MAdult, FYouth and MYouth, and there is no difference between the other groups.

4.4.2 Age gender and the number of arrests

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
AdultF	AdultM	0.2582	0.0001	0.1013	0.4152	True
AdultF	AdultU	-0.4666	0.9991	-5.4592	4.5261	False
AdultF	YouthF	-1.1256	0.0	-1.7266	-0.5247	True
AdultF	YouthM	-1.44	0.0	-1.7993	-1.0807	True
AdultM	AdultU	-0.7248	0.9948	-5.7159	4.2663	False
AdultM	YouthF	-1.3839	0.0	-1.972	-0.7958	True
AdultM	YouthM	-1.6982	0.0	-2.0356	-1.3609	True
AdultU	YouthF	-0.6591	0.9965	-5.6838	4.3657	False
AdultU	YouthM	-0.9734	0.9842	-5.9749	4.0282	False
YouthF	YouthM	-0.3143	0.705	-0.9854	0.3567	False

The post hoc tests on the three variables, age, gender and the number of arrests, revealed differences between the groups AdultF and AdultM, AdultF and YouthF, AdultF and YouthM, AdultM and YouthF, and AdultM and YouthM, and no differences between the other groups.

In addition to this, I also performed post hoc tests on the remaining variable combinations, and the test results are not shown here, but the full results can be seen in the code.

4.5 Interaction plots

4.5.1 Gender, race, and rate of strip searches

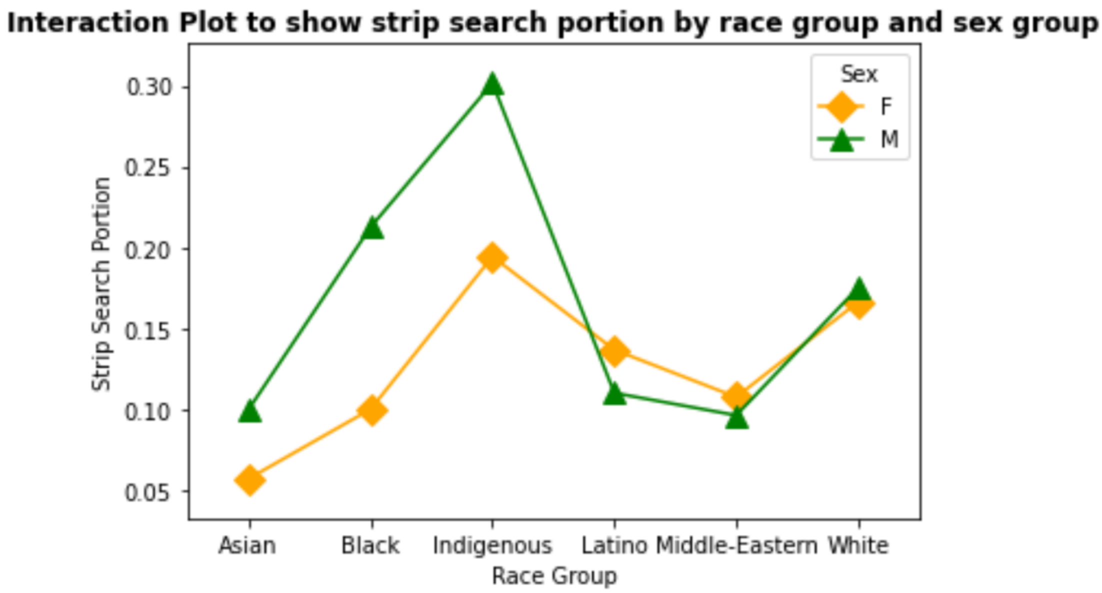


Figure 10-Interaction Plot to show strip search portion

The interaction graphs for gender, race, and strip-search rate are plotted here, containing variables with significant interaction characteristics, which are race as well as gender. The graph can be used to visualize how these two variables change as the strip-search ratio changes. It can be seen that the strip search ratio is generally greater for males than females, with Indigenous having the highest strip search ratio.

4.5.2 Gender, Race, and Number of Arrests

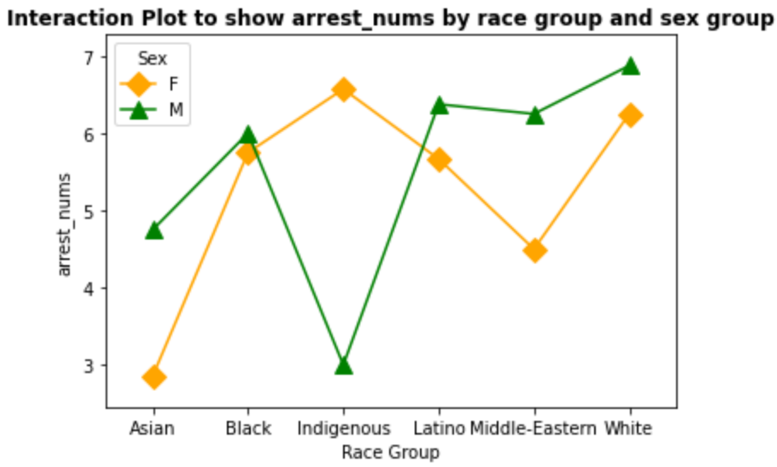


Figure 11-Interaction Plot to show arrest_nums

The interaction plot on gender, race, and number of arrests is plotted here, containing variables with significant interaction characteristics, which are race as well as gender. The graph can be used to visualize how these two variables vary with the number of arrests. It can be seen that there are large differences in the number of male and female arrests by race. For example, Indigenous women are arrested much more often than men, while the opposite trend is observed for all other races.

4.6 Logistic Regression

Here, we will develop a logistic regression model with age, gender, race, and number of arrests as independent variables and the number of strip searches as dependent variables. Firstly, the data set is divided into two parts according to the ratio of 8:2, the training set is used to train the model and the test set is used to evaluate the effect of the model. The final number of the training set is 49612 and the number of the test set is 12404. The following is the result of the model:

Table 16 Logistic regression result

Dep. Variable:	StripSearch	Prob (F-statistic):	2.03e-248
Model:	Logit	Log-Likelihood:	-18882.
Method:	Least Squares	AIC:	3.777e+04
Date:	Wed, 12 Apr 2023	BIC:	3.781e+04
Time:	16:28:03	Df Model:	3
No. Observations:	62016	Covariance Type:	nonrobust
Df Residuals:	62012		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0869	0.003	26.300	0.000	0.080	0.093
Sex	0.0211	0.003	6.295	0.000	0.015	0.028
Perceived_Race	-0.0074	0.001	-10.474	0.000	-0.009	-0.006
num_arrests	0.0075	0.000	31.289	0.000	0.007	0.008

It can be seen from the above table that all coefficients of the model pass the significance test. According to the model, the truncation of the logistic regression equation obtained is -2.226, the coefficient of the variable "Sex" is 0.173, the coefficient of the variable "Perceived_Race" is -0.079, and the coefficient of variable "num_arrests" is 0.051. This indicates that "Perceived_Race" is negatively correlated with target variables, and "Sex" and "num_arrests" are positively correlated with target variables. Finally, the arrests are 1/(1+exp(-2.226+0.173*Sex-0.079* Perceived_Race+0.051* num_arrests).

After the model was constructed, the results of prediction on the test set part of the data were as follows:

Table 17: Model prediction result

StripSearch	value
0	12389
1	15

It can be seen that the model predicted most of the StripSearch categories to be 0, with very few being 1. The calculated model accuracy was 87.62%.

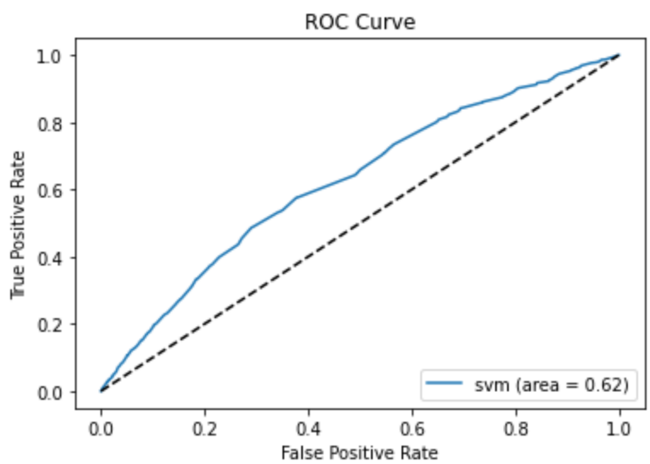


Figure 12-ROC curve

The figure above shows the ROC curve and the corresponding AUC value of the logistic regression model. The area under the ROC curve is called the AUC value. The larger the area under the curve, the higher the prediction accuracy. Here, the AUC value of the model is 0.62, and the prediction effect is reasonable.

Overall, this analysis suggests that race, gender, and number of arrests are significant factors in predicting the number of strip searches conducted, and the logistic regression model can be used to accurately predict the number of strip searches.

4.7 Power analysis

The size of the effect comparing the two groups can be quantified by the effect size measure. A common way to compare the difference between two groups of mean values is the Cohen measure. Here, we first set the significance level (α) as 5%, Cohen's d as 0.80, the default, and assumed a minimum statistical power of 80%, with a calculated minimum sample size of 25.525. We know that with the same number of samples, the larger the Effect Size, the higher the power. With the same Power, the larger the Effect Size, the smaller the sample required.

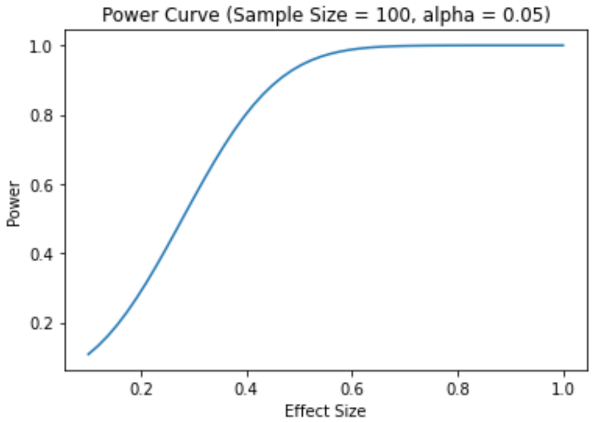


Figure 13-Power analysis

Parameters such as range of effect size, significance level, efficacy, sample ratio and sample size were set. Then, the tt_ind_solve_power function is used to calculate the statistical efficacy under each effect size, and finally, the power graph is drawn. In the power diagram, it can be seen that with the increase of the effect size, the statistical efficacy increases, until a certain effect size is reached, the efficacy becomes stable. In the course of efficacy analysis, attention should be paid to the selection of effect size. In general, the appropriate effect size should be selected according to domain knowledge, historical research and other factors.

5. Conclusion

The purpose of this study was to explore the effects of race, age, and gender on the number of strip searches and the number of arrests. We first conducted a data visualization analysis, which was used to initially explore the relationship between these variables. The results showed significant differences in strip searches and arrests by gender, with males having a much higher number of strip searches and arrests than females, reflecting the relatively higher crime rate among males. In addition, preliminary findings show that people between the ages of 25 and 44 have a higher rate of strip searches and should be a priority for police attention. In terms of race, whites had the highest number of arrests, far outpacing other races, followed by blacks, while Latinos had the lowest number of arrests. The number of arrests was very high for assault, theft, etc. There were significant differences in the number of arrests between some categories, but there were also some categories with similar numbers of arrests. In analyzing the reasons for strip searches, we found that causing injuries and assisting in the escape were the most common reasons for being strip searched. In addition to this, we conducted a series of one-way ANOVAs and post hoc analyses to further validate the data. Finally, we used a logistic regression model to predict the variable number of strip searches and to give predictive accuracy.

However, it is important to note that there are some limitations to this study. First, there may be limitations in the database that may affect the accuracy and generalizability of the results. Second, the effect of sample bias needs to be taken into account. This is because the dataset only includes arrests and searches conducted by a single law enforcement agency, so any conclusions drawn from analyzing this dataset may not be generalizable to other law enforcement agencies. Third, there is a lot of missing data in the dataset, which may limit the analysis and conclusions drawn. Missing data can affect the smooth conduct of the study because the data are incomplete and cannot be analyzed for testing. Fourth, the impact of data quality needs to be considered. The accuracy and completeness of the data may vary, so subsets of the dataset may lead to errors in the overall analysis. Finally, our own lack of practical experience and understanding of each test may lead to subjective bias in the results.

Citation

Gorman, W. K. (2022). Strip searches in Canada and the constitutional right to be free from unreasonable searches and seizures. Court Review, 58(2), 84-95.

Henderson, G. E. (2016). Could community contribution companies improve access to justice. Canadian Bar Review, 94(2), 209-240.

Makin, K. (2001, December 7). Routine strip searches by police banned. The Globe and Mail. Retrieved February 28, 2023, from https://www.theglobeandmail.com/news/national/routine-strip-searches-by-police-banned/article18420111/

Toronto Police Service. (n.d.). https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about.