UNIVERSITY OF TORONTO

Faculty of Information

INF2178 Experimental Design for Data Science

Final Report

The Influence of Age and Occurrence Categories on Arrest and Strip Search Practices in Toronto

Instructor: Shion Guha

Team 43: Yifei Chen, Jingjing Wang

April 16, 2023

Table of Contents

List of Tables

List of Figures

1. Introduction

In recent years, police enforcement practices such as arrests and strip searches have raised some concerns in North American communities, particularly in terms of their potential to cause harm or even death to individuals. These practices have impacted marginalized communities and we have seen a number of protests across the continents aiming to address this issue. A study by Chaney and Robertson (2013) discussed how strip searches could bring traumatic and unpleasant experiences to individuals and impact their well-being.

1.1 First Report Findings

This report builds on top of the first report which examined factors that influence individuals' actions at arrest and the likelihood of a strip search in Toronto using the dataset provided by the Toronto Police Service, Arrests and Strip Searches (RBDC-ARR-TBL-001). In the previous report, we used ANOVA, Tukey's HSD test, T-test, and interaction plots to examine the association between age and actions at arrest, and the association between strip searches and different types of occurrences. We found that **there is a significant difference in the actions at arrest among different age groups.** However, our findings are in contrast with those of previous studies conducted by other scholars that we did not see a trend of younger people, particularly adolescents, having a tendency of higher negative encounters with the police at arrest. Rather, our study found that individuals under 17 years of age and over 65 years of age were more likely to be cooperative and less offensive during arrests compared to middle-aged groups. Our second research finding confirmed that the **occurrence category of the arrest, specifically weapon or drug related occurrences, significantly impacts the likelihood of a strip search.**

While the first report revealed a significant association between age and actions at arrest, it did not take into account the possible impact of confounding variables. Therefore, this follow-up report aims to investigate whether controlling for such variables would affect the findings of the first report. In addition, we want to further investigate the relationship between strip searches and weapon or drug related occurrences.

1.2 Research questions and Hypotheses

The research questions are:

1. What is the relationship between age and actions at arrest, after controlling for potential confounding variables such as sex and race?
2. Do weapons or drug related occurrences significantly impact the likelihood of strip searches at arrest?

Our hypotheses are:

1. After controlling for potential confounding variables such as sex and race, age is still a significant predictor of action at arrest.
2. There is a significant positive relationship between weapons or drug related occurrences and the likelihood of strip searches at arrest.

Answering the new research questions can help to control for bias and provide a deeper understanding of the relationship between the variables being examined.

1.3 Literature Review and Related Work

We aim to investigate whether there are significant differences in the actions at arrest across different age groups. Research by Gleeson (2018) found that young people, especially adolescents, often have problematic police encounters and react negatively during arrests. However, our previous report found a contradicting conclusion. Mazerolle et al. (2013) suggest that older individuals demonstrate a greater sense of respect for authority figures, which our previous report aligns with this finding. Based on our previous study, we concluded that age is a factor during actions at arrests, but we want to understand this topic further and gain a better understanding if confounding variables impact this conclusion.

We also hope to examine the impact of occurrence categories on the likelihood of strip searches during arrests. Previous studies suggest that individuals arrested for weapons or drug related offenses were more likely to be subjected to strip searches (Chaney and Robertson, 2013; McNeilly, 2019). By gaining insights into the relationship between occurrence categories and strip searches, we can develop foundational knowledge to help future researchers determine if bias exists in this practice and if so, how could law enforcements minimize the harm to the individuals.

2. Dataset

This report uses the Arrests and Strip Searches (RBDC-ARR-TBL-001) dataset provided by the Toronto Police Service to examine the research questions (Toronto Police Service, 2022).

2.1 Size

The dataset consists of 25 columns and 65,276 observations.

2.2 Missing Values

Some columns contain missing values, they are: ArrestID, Perceived_Race, Age_group__at_arrest_, Occurrence_Category , SearchReason_CauseInjury, SearchReason_AssistEscape, SearchReason_PossessWeapons, SearchReason_PossessEvidence, ItemsFound

Handling Missing Values:

1. Variable Sex has values: M, F, U. This report assumes that U stands for “Unknown” so it is treated as missing values. Thus, 9 observations with “U” are removed from the dataset.

2. We only approached the missing values in variables: Occurrence_Category with 165 missing values and Age_group__at_arrest_ with 24 missing values, because they are relevant to our research questions. We chose to remove the rows with missing values because the number of missing values is relatively small compared to the size of the dataset and would not impact our ability to draw valid conclusions from the remaining data. In this case, it is not suitable to replace the missing values with an estimated value for these two variables due to their categorical data type.

3. The other variables with missing values remain the same because the rows contain valid data in other columns that could be useful for our analysis. And these variables are not the focus of this research.

2.3 Datatype

1. Nominal data (6 variables): Arrest_Month, Sex, Age_group__at_arrest_, ArrestLocDiv, Youth_at_arrest__under_18_years, Perceived Race, Occurrence_Category,

2. Ordinal data (1 variable): Age_group__at_arrest

3. Discrete data (18 variables): Arrest_Year, EventID, ArrestID, PersonID, StripSearch (Binary), Booked (Binary), Actions_at_arrest___Concealed_i,(Binary) Actions_at_arrest___Combative__(Binary), Actions_at_arrest___Resisted__d (Binary), Actions_at_arrest___Mental_inst(Binary), Actions_at_arrest___Assaulted_o(Binary), Actions_at_arrest___Cooperative(Binary), SearchReason_CauseInjury, SearchReason_AssistEscape, SearchReason_PossessWeapons, SearchReason_PossessEvidence, ItemsFound, ObjectId

4. Continuous data: The dataset does not contain continuous data.

2.4 Duplicates

The dataset does not have duplicate observations.

2.5 Data Inconsistencies

While inspecting the data, we noticed that within Age_group__at_arrest, "Aged 17 years and younger" and "Aged 17 years and under" refer to the same group of people; "Aged 65 years and older" and "Aged 65 and older" also represent the same age group. To avoid confusion, we combined these categories and updated them as "Aged 17 years and younger" and "Aged 65 years and older".

2.6 Data Preparation

1. We assigned unique identifiers to the age groups (Age_group__at_arrest_) which convert the categorical levels into numerical forms. A new column called age_group_id is created to store the unique identifiers.

    - Aged 17 years and younger: 1
    - Aged 18 to 24 years: 2
    - Aged 25 to 34 years: 3
    - Aged 35 to 44 years: 4
    - Aged 45 to 54 years: 5
    - Aged 55 to 64 years: 6
    - Aged 65 years and older: 7

3. Exploratory Data Analysis

3.1 Descriptive Statistics

Table 1. Descriptive Statistics

Variable	N	Mean	Std.Dev.	Min	Max
StripSearch	65276	0.119508	0.324388	0	1
Actions_at_arrest___Concealed_i	65276	0.004075	0.063706	0	1
Actions_at_arrest___Combative_	65276	0.044136	0.205398	0	1
Actions_at_arrest___Resisted__d	65276	0.03833	0.191992	0	1
Actions_at_arrest___Mental_inst	65276	0.033381	0.179632	0	1
Actions_at_arrest___Assaulted_o	65276	0.006358	0.079481	0	1
Actions_at_arrest___Cooperative	65276	0.445937	0.497072	0	1

Table 1 shows the descriptive statistics for StripSearch and Actions_at_arrest_ columns which contain binary values to indicate whether certain actions were taken at the time of the arrest. 0 refers to “No” and 1 refers to “Yes”. For example, the mean value for “Combative” is 0.04, indicating that combative behavior at arrests is rare. On the other hand, “Cooperative” has the highest mean across all actions with a mean of 0.45, indicating that most individuals are cooperative during the arrest. Here, we developed a basic understanding of the likelihood of certain actions at the time of the arrest.

3.2 Age

Table 2. Age Distribution

Age_group__at_arrest_	Count	Percentage
Aged 25 to 34 years	20945	32.09%
Aged 35 to 44 years	16241	24.88%
Aged 18 to 24 years	10039	15.38%
Aged 45 to 54 years	9066	13.89%
Aged 55 to 64 years	4588	7.03%
Aged 17 years and younger	3042	4.66%
Aged 65 years and older	1322	2.03%

Table 2 shows the count and percentage of individuals in the dataset categorized by their age group at the time of their arrest. The majority of individuals (32.09%) are between 25 to 34 years, followed by 35 to 44 years (24.88%), and 18 to 24 years (15.38%). The smallest group is aged 65 years and older, which presents only 2.03% of the dataset.

Check for normal distribution:

We used the Shapiro-Wilk test to test if a dataset follows a normal distribution.

- H0: The data is normally distributed
- H1: The data is not normally distributed.
- The result of the test is: **Statistic=0.940, p-value=0.000**

The p-value is less than the significance level of 0.05, we reject the null hypothesis and conclude that the data is not normally distributed.

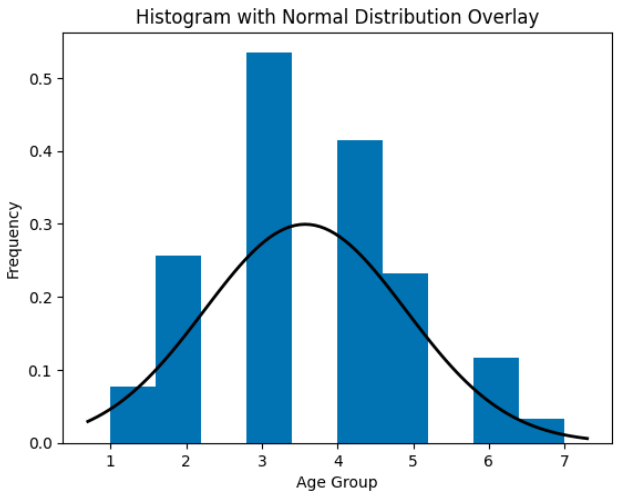

Figure 1. Age Group Histogram with Normal Distribution Overlay
(x-axis values are dummy variables representing each age group as explained in *2.6 Data Preparation*)

The histogram shows that the data is not perfectly normally distributed, as the curve does not perfectly match the shape of the histogram.

However, the curve still provides a rough approximation of the distribution, and we can see that the data is somewhat skewed to the right, with a longer tail on the right side of the distribution.

Since the data is slightly skewed, the normality assumption may be less critical, especially when our sample size of 65276 can be considered as large.

Bar plot:

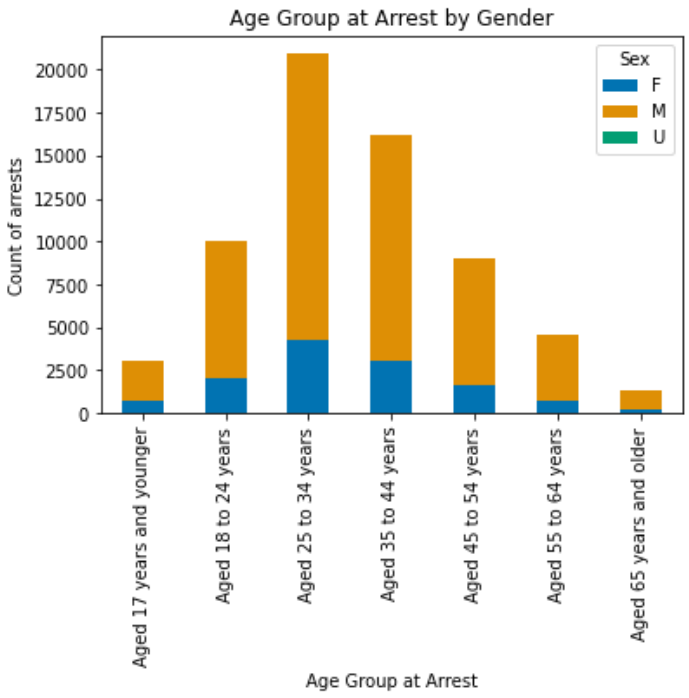

Figure 2. Age Group at Arrests by Gender

This plot shows the number of arrests across different age groups broken down by sex. The x-axis represents the age groups, and the y-axis represents the count of arrests. The bars are stacked and color-coded with blue representing male and orange representing female.

This plot shows the distribution of arrests across different age groups and genders, with most individuals in their middle age adulthood.

3.3 Occurrence Category

Table 3. Occurrence Category

Occurrence Category	**Count**	**Percentage**
Others	60290	92.395635%
Weapons or Drug Related	4962	7.604365%

The report focuses on understanding if weapons or drug related occurrences are associated with a higher likelihood of strip searches. Table 4 shows that 7.6% of the arrests are associated with weapons or drug related occurrences compared to all other occurrences in the dataset.

3.4 Strip Search

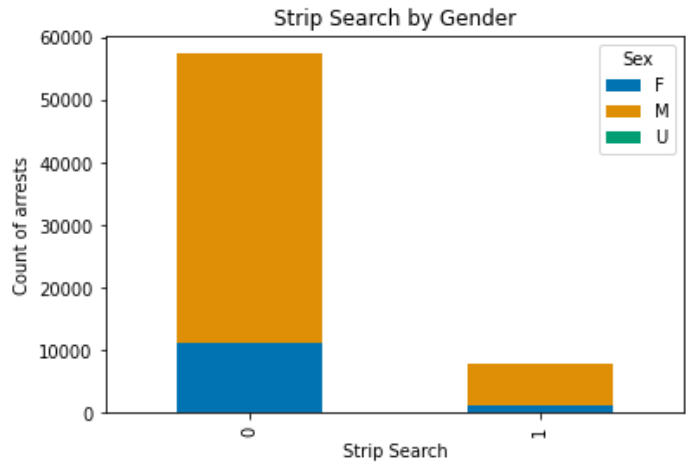

Figure 3. Strip Search by Gender

This plot shows the number of arrests in which a strip search was conducted, broken down by sex. The x-axis represents whether strip search was conducted (1) or not (0), and the y-axis represents the count of arrests. The bars are stacked and colour-coded with blue representing male and orange representing female.

There are more males than females arrested in the dataset. “U”, defined as “unknown” is later removed from the clean dataset. The majority of individuals in the dataset did not experience strip searches.

3.5 Race

Table 4. Perceived Race

Perceived Race	**Count**	**Percentage**
White	27723	42.47%
Black	17526	26.85%
Unknown or Legacy	5056	7.75%
East/Southeast Asian	4415	6.76%
South Asian	3613	5.53%
Middle-Eastern	3237	4.96%
Indigenous	1934	2.96%
Latino	1768	2.71%

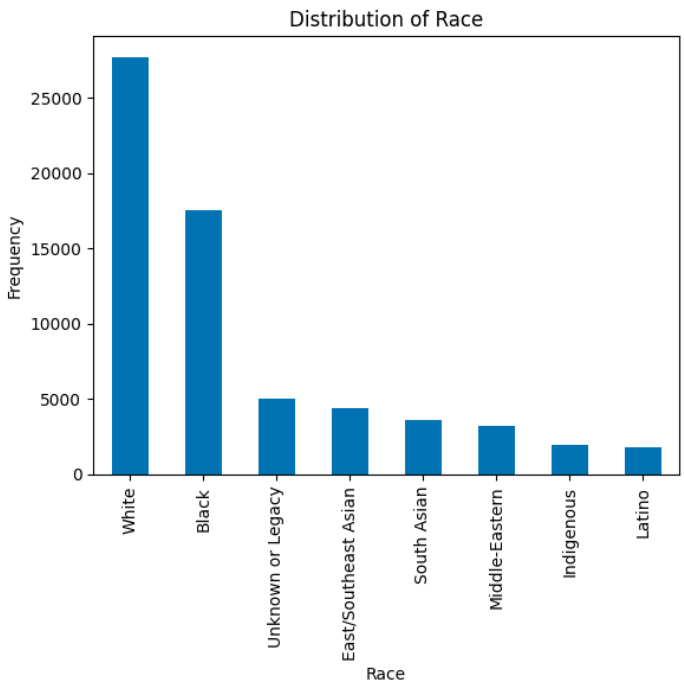

Figure 4. Distribution by Race

In this bar plot, the x-axis represents the different categories of the Perceived_Race variable, and the y-axis represents the counts.

The plot provides a visualization that allows us to easily compare the heights of the bars to see which races are more or less common in the dataset.

3.6 Hypothesis Testing

T-test 1: Age Group (IV) and Actions at Arrest (DV)

The previous report hypothesizes that there are significant differences in the actions at arrest between two age groups, 24 and younger and 25 and older. ANOVA, Tukey's HSD test, T-test, and interaction plots all indicated a significant difference in actions at arrest across age groups but due to the complex relationship between age and actions at arrest, we cannot confirm a consistent age-related trend. Based on the tests, we could only suggest that individuals under 17 years old and individuals over 65 years old are more cooperative and less offensive compared to middle-age groups. However, we cannot conclude that being cooperative has a consistent positive relationship with age, as this relationship may vary across different age groups. Similarly, being offensive does not have a consistent negative relationship with age, as the likelihood of offensive actions may vary across different age groups.

That said, this report hypothesizes individuals under 17 years old and individuals over 65 years old have similar actions at arrest and middle-age groups have similar actions. That said, our updated T-test has hypothesis:

- Null hypothesis (H0): There is no significant difference in the means of actions at arrest across different two age groups (under 17 or over 65 years old vs middle aged) .
- Alternative hypothesis (H1): There is a significant difference in the means of actions at arrest across different two age groups (under 17 or over 65 years old vs middle aged).

Outcome: The results suggest that for actions "Combative", "Resisted", "Mental_inst", and "Cooperative", the p-value is less than 0.05. We can reject the null hypothesis and conclude that age significantly impacts the likelihood of these actions at arrest.

For the categories “Concealed” and "Assaulted," the p-value is greater than 0.05, indicating that we fail to reject the null hypothesis and can conclude that age does not significantly impact the likelihood of these actions at arrest.

T-test 2: Sex (IV) and Actions at Arrest(DV)

We also explored if males and females have significantly different actions at arrests.

- Null hypothesis (H0): There is no significant difference in the means of actions at arrest between males and females.
- Alternative hypothesis (H1): There is a significant difference in the means of actions at arrests between males and females.

Outcome: the results suggest that for the category “Cooperative”, the p-value is less than 0.05. We can reject the null hypothesis and conclude that there is a significant difference between males and females when acting cooperatively at arrests.

However, for the categories “Concealed”, “Combative”, “Resisted”, “Mental_inst”, and “Assaulted”, the p-value is greater than 0.05. We fail to reject the null hypothesis that there is no significant difference between the mean actions at arrest for males and females. Thus, the act of being concealed, combative, resisted, having mental instability, or assaulted at arrest is not significantly different between males and females.

T-test 3: Sex (IV) and Strip Search (DV)

The report hypothesizes that the likelihood of a strip search is significantly impacted by factors such as sex.

- Null hypothesis (H0): There is no significant difference in the likelihood of strip searches between males and females.
- Alternative hypothesis (H1): There is a significant difference in the likelihood of strip searches between males and females.

Outcome: Based on the T-test results, the calculated t-statistic is -0.98, and the p-value is 0.43. Since the p-value is greater than the chosen significance level at 0.05, we fail to reject the null hypothesis. This means that we don't have sufficient evidence to suggest that there is a significant difference in the likelihood of strip searches between males and females.

T-test 4: Weapons or Drug Related occurrences (IV) and Strip Search (DV)

Based on the literature review, previous studies have found that weapon or drug related occurrences are associated with a higher likelihood of a strip search. To process this T-test, we created a new column in the dataframe that categorizes each occurrence as “Weapons or Drug Related” or “Others”.

- Null hypothesis (H0): There is no significant difference in the likelihood of strip searches between weapon or drug related occurrences and others.
- Alternative hypothesis (H1): There is a significant difference in the likelihood of strip searches between weapon or drug related occurrences and others.

Outcome: Based on the T-test results, the calculated t-statistic is 32.59, and the p-value is 0.0000. Since the p-value is near 0, we can reject the null hypothesis. There is sufficient evidence to suggest that there is a significant difference in the likelihood of strip searches between weapon or drug related occurrences and other occurrences.

T-test summary

1. There is a significant difference between individuals under 17 years of age or over 65 years of age and middle-aged groups when acting combative, resisted, mentally unstable, and cooperative at arrest but not other actions (Concealed,, Assaulted).
2. There is a significant difference between males and females when acting cooperatively at arrests, but not for other actions (Concealed, Combative, Resisted, Mental_inst, Assaulted).
3. There is a significant difference in the likelihood of strip searches between weapon or drug related occurrences and other occurrences.
4. There is insufficient evidence to suggest a significant difference in the likelihood of strip searches between males and females.

4. Method

In order to answer the research questions and test our hypothesis, we used Power analysis, ANCOVA, and Logistic regression.

4.1 Power Analysis

Power analysis is used to help us understand the sample size required to detect a difference in actions at arrest between the two age groups (under 17 years old /over 65 years old vs middle aged).

1. We first created a new column called “age” that contains two values: “<17 or >65” and “17<age<65”. “<17 or >65” refers to individuals under 17 years old or over 65 years old and “17<age<65” refers to middle-aged individuals.
2. We used two functions, **“pooled_standard_deviation”** and **“Cohens_d "**, to calculate the pooled standard deviation of sample and sample 2, and the effect size of the difference between the means of the two independent samples.
3. Based on the “age” column created in Step 1, we created two datasets “group1” (contains rows where the 'age' value is “<17 or >65”) and “group2” (contains rows where the 'age' value is “17<age<65”). Selecting the “action” column in “group 1” and “group 2”, we created two samples, 'g1a' and 'g2a'.

4. We calculated the **effect size (Cohen's D)** for the two groups (g1a and g2a) in relation to “actions at arrest”. The alpha level was set to 0.05 and the desired power was set to 0.8. We also calculated the ratio of the sample sizes between the two groups at this step.

5. We then calculated the **sample size** for g1a and g2a using the TTestIndPower function from the statsmodels.stats.power module.

6. In addition, we plotted a **power curve** using plot_power method from TTestIndPower to show how the power of the test changes as the sample size and effect size are varied.

4.2 ANCOVA

Once we have the effect size and the sample size for the two sample groups and understand that our actual size is higher than the required sample size, we are ready to run ANCOVA to analyze the relationship between the dependent variable and independent variable while controlling for covariates.

1. We installed the “pingouin” package to conduct the analysis.

2. The dependent variable is “action at arrest” and independent variable is “age” (grouped into group 1: <17 or >65 and group 2: 17<age<65). The defined covariates are “sex” and “perceived_race”.

3. The “Sex” column was converted to numeric values with 0 representing females and 1 representing males. The 'Perceived_Race' column was also converted to numeric values. This conversion prepares the data into a format that can be used by ANCOVA.

4.3 Logistic Regression

We used Logistic Regression to build a predictive model that can support classifying whether a person is subjected to strip search or not based on the occurrence category they belong to.

Assumption Check:

1. The dependent variable is StripSearch, which contains binary values (0 or 1) so logistic regression is an appropriate method to use.

2. Check for multicollinearity:

	StripSearch \
Occurrence_Category_Drug Related	0.175844
Occurrence_Category_Weapons	-0.024973
Occurrence_Category_Weapons & Homicide	0.120403

   a. The correlation matrix shows the pairwise correlation coefficients between variables StripSearch and Occurrence_Category_Drug Related, Occurrence_Category_Weapons, Occurrence_Category_Weapons & Homicide. The low correlation between variables indicates that there is not a strong linear relationship, which allows us to confirm the absence of multicollinearity assumption required for logistic regression.

Preparation:

1. We first selected the two relevant columns, “Occurrence_Category” and "StripSearch”, and stored them into a new Dataframe called “data_log”

2. We converted the categorical variable “Occurrence_Category” into numerical dummy variables , which is necessary for logistic regression which requires numerical inputs.

3. We separated the predictor variables (Occurrence_Category) and the response variable (StripSearch) into its own dataframe. Occurrence_Category stored in DataFrame X; StripSearch stored in DataFrame Y.

Logistic Regression:

1. The data was then split into training and testing sets using the **train_test_split()** function. One for training the model (X_train and y_train) and one for testing the model (X_test and y_test). The test set is 0.2, which is equivalent to 20% of the original data. The chosen random state is 34 to ensure that the split is reproducible.

2. We used the **fit()** method to call on the model object with the training data (X_train and y_train). This trains the logistic regression model on the training data.

3. We used the predict**()** method to call on the model object with the testing data (x_test) which generates predicted outcomes y_pred based on the training model and testing data.

4. We used the **sm.Logit** function to create a logistic regression model using the training data (y_train and X_train).

5. We evaluated the model’s performance to determine how well it is able to predict outcomes for new data by calculating the accuracy score. To further evaluate the performance of a binary classification model, we incorporated the confusion matrix.

5. Results and Findings

5.1 Age Group and Actions at Arrest

Power Analysis Output:

```
Effect size (Cohen's D) for actions at arrest: -0.049220883452969065
Sample Size of nobs1: 48436.533 needed for g2a
Actual size of g2a: 60879
Sample Size of nobs2: 3472.084 needed for g1a
Actual size of g1a: 4364
```

Interpretation:

The output of the code shows the result of the power analysis conducted on the defined two groups based on the actions at arrest variable.

1. The effect size, measured by Cohen's d, is -0.049, which indicates a very small difference between the means of the two groups being compared (g1a: <17 or >65 vs g2a: 17<age<65). The negative sign indicates that the mean value for the first group (g1a: <17 or >65) is smaller than the mean value for the second group (g2a: 17<age<65). This means that, on average, the actions taken at arrest for individuals who are either under 17 years old or over 65 years old are slightly lower than individuals aged 18 to 64.

2. For a power of 0.8 and an alpha of 0.05:

    a. Group 1 (g1a: <17 or >65) requires a sample size of 3472, while the actual sample size is 4364. If we want to detect a difference of this effect size between the two groups, we would need a sample size of at least 3472. The actual sample size is 4364, which is higher than the sample size needed for the effect size. That said, the following test has a good chance of detecting a difference in actions at arrest between the two defined age groups, given the available sample sizes.

    b. Group 2 (g2a: 17<age<65) requires a sample size of 48436, while the actual sample size for that group is 60879. If we want to detect a difference of this effect size between the two groups, we would need a sample size of at least 48436. The actual sample size is 60879, which is higher than the sample size needed for the effect size. That said, the following test has a good chance of detecting a difference in actions at arrest between the two defined age groups, given the available sample sizes.

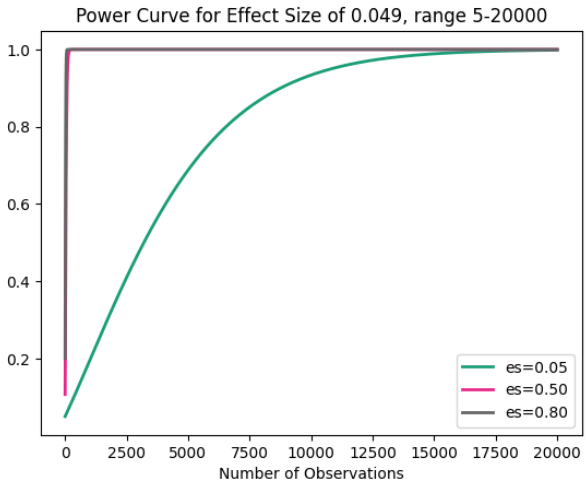

Figure 5. Power Curve for Effect Size of 0.049, range 5-20000

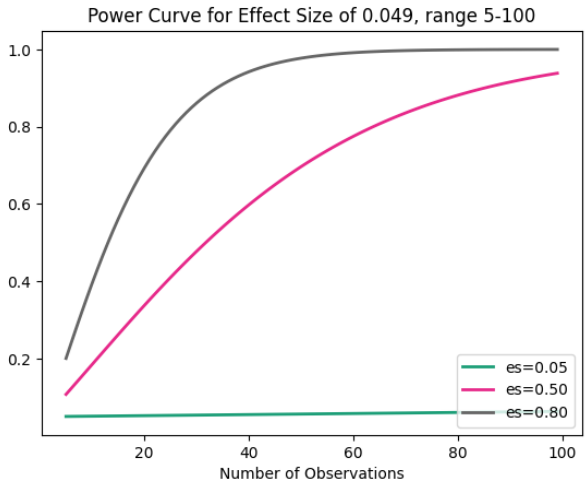

Figure 6. Power Curve for Effect Size of 0.049, range 5-100

3. Figure 5 and Figure 6 are power curve plots used to visualize the relationship between sample size, effect size, and power.

a. es=.049 is the effect size measured by Cohen's d. es=0.5 and es=0.8 are added to the plot in addition to the actual effect size of 0.049 to provide a better understanding of the relationship between effect size, sample size, and power of the test.

b. The y-axis represents the power of the test, ranging from 0 to 1.

c. The x-axis represents the range of sample sizes considered in the power analysis, from 5 to 100.

d. The power curve shows how the power of the test changes with different sample sizes (x-axis) and effect sizes (lines with different colors). When the effect size is 0.049, meaning the association between two or more groups is very low, the power of the study is relatively low unless we increase the sample size to 20000 (see Figure 5). On the other hand, when the effect size is 0.8, the power of the study is high even with a sample size of under 100 observations (see Figure 6).

ANCOVA

ANCOVA was used to analyze the relationship between the dependent variable "action at arrest" and the independent variable "Age" while controlling for the covariates "Sex" and "Perceived_Race".

Table 5. ANCOVA Output

Source	SS	DF	F	p-unc	np2
Age	3.408905	1	9.957781	0.001602	0.000153
Sex	2.353586	1	6.875079	0.008743	0.000105
Perceived_Race	0.904205	1	2.641280	0.104124	0.000040
Residual	22332.275712	65235	NaN	NaN	NaN

Interpretation:

1. The output table shows the ANCOVA output, including the source of variation, sum of squares (SS), degrees of freedom (DF), F statistic, p-value, and partial eta-squared (np2). In general, an np2 of 0.01 is considered a small effect, 0.06 a medium effect, and 0.14 a large effect.

2. Age has a significant effect on action with a low p-value of 0.001602. A small np2 of 0.000153 indicates that the effect of the independent variable (age) on the dependent variable (action at arrest) is relatively weak, after controlling for sex and perceived race.

3. Sex has a significant effect on action with a low p-value of 0.008743 but the effect on the dependent variable (action at arrest) is low with a np2 of 0.000105, after controlling for age and perceived race

4. Perceived race does not have a significant effect on action, as indicated by the high p-value of 0.104124, after controlling for age and sex.

5.3 Occurrence Category and Strip Search

We gathered information about the shape and summary statistics of the training and test data sets for the target variable StripSearch.

Output:

```
========== Training data ========
count    52194.000000
mean         0.119209
std          0.324038
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          1.000000
Name: StripSearch, dtype: float64
X_test shape: (13049, 31)
y_test shape: (13049,)
```

```
========== Test data ==========
count    13049.000000
mean         0.121005
std          0.326146
min          0.000000
25%          0.000000
50%          0.000000
75%          0.000000
max          1.000000
Name: StripSearch, dtype: float64
```

Interpretation:

1. X_train shape has 52,194 rows and 31 columns (occurrence categories).
2. y_train shape has 52,194 rows with the target variable StripSearch.
3. The StripSearch column in the training data has a mean value of 0.119, indicating that about 12% of the cases in the training set had a StripSearch.
4. X_test shape has 13,049 rows and 31 columns (occurrence categories).
5. y_test shape has 13,049 rows with the target variable StripSearch.
6. The StripSearch column in the test data has a mean value of 0.121, indicating that about 12% of the cases in the test set had a StripSearch.
7. The distribution of StripSearch is similar in the training and test data, and we do not see any imbalance issue.

Logistic Regression:

We used logistic regression to predict the likelihood of a strip search being conducted. The model has been trained on 52194 observations with 30 predictors. We set the maximum number of iterations for the algorithm as 1000.

Output:

```
                           Logit Regression Results
==============================================================================
Dep. Variable:             StripSearch   No. Observations:                52194
Model:                           Logit   Df Residuals:                    52163
Method:                            MLE   Df Model:                           30
Date:                 Thu, 13 Apr 2023   Pseudo R-squ.:                 0.2027
Time:                         22:02:52   Log-Likelihood:                -15203.
converged:                       False   LL-Null:                       -19069.
Covariance Type:             nonrobust   LLR p-value:                     0.000

==============================================================================
                                 coef    std err          t      P>|t|      [0.025      0.975]
Occurrence_Drug Related          -0.4499      0.043    -10.345      0.000      -0.535      -0.365
Occurrence——Weapons              -2.8028      0.144    -19.435      0.000      -3.085      -2.520
Occurrence_Weapons & Homicide    -0.3014      0.068     -4.460      0.000      -0.434      -0.169
```

Interpretation:

1. The model has converged as the value under the converged column is False
2. The model's pseudo R-squared value is 0.2027, which suggests that the model explains about 20% of the variance in the data.
3. The LLR p-value of 0 suggests that the model significantly outperforms the null model.
4. Some predictors have a coefficient that is statistically significant with a P> |z| less than 0.05. In the logistic regression results output shown above, we only included features that the study focused on, which are Weapon or drug related occurrences.
    a. Occurrence_Category_Drug: the coefficient of -0.4499 means that this occurrence is associated with a decrease in the log odds of the outcome by 0.4499, holding all other variables constant. This suggests a negative relationship between drug related occurrences and the number of strip tests. The standard error is 0.043, indicating a high level of precision in the estimate.
    b. Occurrence_Category_Weapons, the coefficient is -2.8028, which means that this occurrence is associated with a decrease in the log odds of the outcome by 2.8028, holding all other variables constant. The standard error is 0.144, indicating a moderate level of precision in the estimate.
    c. Occurrence_Category_Weapons & Homicide: the coefficient is -0.3014, which means that this occurrence is associated with a decrease in the log odds of the outcome by 0.3014, holding all other variables constant. The standard error is 0.068, indicating a high level of precision in the estimate.

Accuracy Score and Confusion Matrix

Table 6. Accuracy Score and Confusion Matrix

Accuracy Score	Confusion Matrix
0.878994558970036	[[11470 0] [1579 0]]

Interpretation:

1. The accuracy score of 0.8789 means that the logistic regression model correctly predicted the outcome (whether or not a strip search was conducted) in 87.89% of cases.
2. The confusion matrix shows that:
    a. True negatives: 11470 out of the 13049 cases were correctly predicted to not have a strip search conducted
    b. False positives: 1579 out of the 13049 cases were incorrectly predicted to have a strip search conducted
    c. False negatives: The model did not incorrectly predict to not have a strip search conducted when they did.
    d. True positives: The model did not correctly predict any cases where a strip search was conducted.

6. Discussion

This report has two objectives: to understand if age groups have different actions at arrest, and if occurrence categories impact the likelihood of a strip search.

6.1 Summary

Research Objective 1:

- Question: What is the relationship between age and actions at arrest, after controlling for potential confounding variables such as sex and race?
- Hypothesis: After controlling for potential confounding variables such as sex and race, age is still a significant predictor of action at arrest.
- Findings: **The ANCOVA result suggests that there is a significant relationship between age and actions at arrest after controlling for sex and perceived race. Although we accept the hypothesis, the effect size is weak, meaning age may not be the strongest predictor of actions at arrest.**

Research Objective 2:

- Question: Does weapons or drug related occurrences significantly impact the likelihood of strip searches at arrest?
- Hypothesis: There is a significant positive relationship between weapons or drug related occurrences and the likelihood of strip searches at arrest.
- Findings: **The logistic regression results suggest that weapons or drug related occurrences are negatively associated with the likelihood of a strip search, after controlling for other variables in the model. We reject our research hypothesis.**

6.2 Discussion and Limitations

Research Objective 1:

To answer our research question, we investigated the effect of age on actions at arrest with a specific focus to compare individuals under 17 years old and individuals over 65 years old to those of middle-aged individuals. This is because our previous study has suggested a difference between these two age groups.

Based on the results of the power analysis, the effect size is relatively small at -0.049, indicating that the **association between age and actions at arrest is weak.** In order to detect such a small difference between the age groups in their actions at arrest after controlling for sex and race, a large sample size is required to increase the power of the analysis. Luckily, this study uses a large dataset to compensate for an extremely small effect size.

Once we confirmed that our dataset contains enough sample size, we generated ANCOVA to further understand the relationship between age and actions at arrest.

With a low p-value, we confirmed that age has a significant effect on the dependent variable, action at arrest, but the small np2 shows that the effect is weak, after controlling for sex and perceived race. This means that age is a significant predictor of action after controlling for the covariates (Sex and Perceived_Race) but may not be the most important factor that predicts actions at arrest. Additionally, the analysis found that sex is a significant predictor on actions at arrest but also has a very weak effect, and perceived race is not a significant predictor. This finding suggests that action at arrest could be affected by other factors beyond the scope of the study that focuses on age.

Limitations:

1. While we found that age and sex impact the actions at arrest after controlling for potential confounding variables such as perceived race, we did not find a strong predictor on the actions at arrest in this study.
2. For the power analysis, we made a unique approach to split the age groups into individuals under 17 years old and individuals over 65 years old vs. those of middle-aged individuals. This approach was inspired by the outcome from our previous report that we identified a pattern in the actions at arrest between individuals under 17 and over 65. These groups are more cooperative

and less combative compared to middle-aged groups. However, other age groups might have a different effect size.

3. The study only examined a limited set of covariates (sex and perceived race), there are many other factors that we could consider to provide a more in-depth analysis.

4. Regardless of existing scholarly research done on how age influences individuals actions at arrest, our data shows a weak relationship between age and actions at arrest. While we find a significant relationship between these two variables, multiple analyses indicate that the effect is not strong.

Research Objective 2:

Our previous research used ANOVA to confirm a significant relationship between occurrences at arrest and strip searches. Specifically, Tukey's HSD indicated a higher mean frequency of strip searches for the "Weapons or Drug Related" group compared to the "Others" group. This suggests that occurrences related to weapons or drugs are more likely to result in strip searches than other types of occurrences. Building on our previous research, we wanted to explore further if the relationship between weapons or drug related occurrences and the likelihood of strip searches are significantly positive as suggested by existing scholarly research.

However, the logistic regression conducted indicates that **weapons or drugs occurrences are all negatively associated with strip searches**, which means that these occurrences are less likely to result in a strip search. This finding challenges the existing research and our previous research. The logistic regression model has an accuracy score of 0.8789, meaning it can correctly predict the outcome (whether or not a strip search was conducted) for about 88% of the cases in the test set.

To understand why we received contradicting outcomes, one reason could be that they are different statistical methods. In the previous report, the ANOVA analysis was used to compare the mean frequency of strip searches between the "Weapons or Drug Related" group and the "Others" group. Whereas the logistic regression model focused on examining the relationship between each occurrence category and the likelihood of strip searches.

Limitation:

1. The contradicting outcome of the logistic regression result could be caused by confounding variables that we did not consider controlling for the model. This could have resulted in a different relationship between the selected occurrence category and strip searches.

2. Both the DV (StripSearch) and IV (Occurrence category) are binary so we were unable to create a prediction plot to predict probabilities of a strip search against the occurrence category. The plot would have two lines and is not informative. If we worked with a continuous independent variable and a continuous dependent variable, we could generate a useful plot.

7. Conclusion

The statistical analysis (Power Analysis, ANCOVA, and Logistic Regression) conducted in this study allowed us to successfully answer the two research questions and test the hypotheses.

We first examined the relationship between age and actions at arrest after controlling for sex and perceived race. We confirmed our hypothesis that age is a significant predictor of action at arrest, but the relationship is weak. This suggests that there may be other factors that serve as a strong predictor of actions at arrest, further research are required to help us understand this.

We then investigated the relationship between strip searches and weapon or drug related occurrences. Based on the outcome, we rejected our research hypothesis that there is a positive relationship between the likelihood of a strip search and weapons or drug related occurrences. In contrast, we observed a significant negative relationship between the variables. As these occurrences increase, the likelihood of strip search being involved decreases. This finding provides important insights into the use of strip searched by police during enforcement in Toronto and suggests further research as it contradicts with existing studies on this topic.

8. Reference

Chaney, C., & Von Robertson, R. (2013). Racism and Police Brutality in America. Journal of African American Studies, 17(4), 480–505. https://doi.org/10.1007/s12111-013-9246-5

Gleeson, H. F. (2018). Exploring the meaning of young people's attitudes towards the police. A qualitative study of Irish youth. International Journal of Law Crime and Justice. https://doi.org/10.1016/j.ijlcj.2018.10.004

Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013). Shaping Citizen Perceptions of Police Legitimacy: A Randomized Field Trial of Procedural Justice. Criminology, 51(1), 33–63. https://doi.org/10.1111/j.1745-9125.2012.00289.x

McNeilly, G. (2019, March). Summary of Ontario Jurisprudence Involving Strip Searches Post R. v. Golden. OIPRD. Retrieved February 25, 2023, from https://www.oiprd.on.ca/wp-content/uploads/OIPRD-Summary-of-Ontario-Jurisprudence-Involving-Strip-Searches-Accessible.pdf

Toronto Police Service. (2022). Arrests and Strip Searches (RBDC-ARR-TBL-001) [Dataset]. https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/explore