The Study of Demographic Disparities in Strip Searches and Crime Severity with Respect to Youth Status and Gender

Group 51
Yiyang Huang 1004902901
Mengchuan Wei 1009692101

A final paper submitted in conformity with the requirements
for the course of INF2178H

Faculty of Information
University of Toronto

2023.02.28

Abstract

Our paper aims to investigate the potential biases and disparities in the use of strip searches on different demographic groups and potential difference in severity of the crimes that lead to an arrest committed by different demographic groups. Both research questions are quantitative. The study is based on 65276 arrest records containing information related to all arrests and strip searches, created by Toronto Police Service. The analysis was performed using exploratory data analysis, ANOVA tests, and post-hoc tests. The analysis conducted to answer the first research question indicates that the means of the severity score of the case involved are not the same by youth status and by gender. All results are significant and people who are under 18 years old and female are likely to involve a more severe incident. The analysis conducted to answer the second research question indicates that there exists gender and racial disparities in strip searches during the arrest process. All results are significant and people who are identified as Indigenous, Black, or White, and male, are more likely to be strip-searched.

Keywords: Strip Searches, Crime Severity, Under Age, Youth, Gender Disparity, Racial Disparities

Table of Contents

1. Introduction

The role of police officers in maintaining public order and safeguarding people's safety is critical, and this profession comes with significant power. However, it is crucial to monitor this power and ensure it is kept in institutional cages to avoid potential abuse. At the same time, it is essential to maintain a balance and not infringe upon the rights and interests of police officers, as this can demoralize the majority of police officers working on the front lines and exacerbate the profession's shortage. Today, there is a growing debate in society about police accountability and unbiased enforcement. According to *Public perceptions of the police in Canada's provinces*, published in the Canadian Centre for Justice and Community Safety Statistics, in 2019 only 42% of Canadians believe that their local police are able to treat people fairly (Dyna Ibrahim, 2020). Therefore, establishing a good interaction between police officers and citizens becomes essential for individuals and communities. According to the *Review of the RCMP's Bias-Free Policing Model Report*, the Civilian Review and Complaints Commission's review of the RCMP's bias-free policing model found policy flaws and made recommendations to strengthen the policy, such as extending the prohibition of racial profiling to all forms of bias-based profiling (CRCC, 2022).

With the increasing number of juvenile delinquency cases, the issue of low age of criminal activity has become a pressing concern in society. The harm caused by young people who engage in criminal activity extends beyond themselves to impact their families, communities, and society as a whole. Addressing youth crime has the potential to prevent future criminal behavior and reduce the social and economic costs associated with crime. According to *Youth Crime in Canada*, the juvenile crime rate is as high as 4,322 per 100,000 youth. Although youth make up only 7% of the Canadian population, they account for 13% of those charged with a crime (Mary K. Allen and Tamy Superle, 2016), which indicates that youth are overrepresented in the criminal justice system, and suggests that there may be underlying social and economic factors that contribute to delinquent behavior among young people.

Based on existing research, we are first interested in examining the correlation between the number of strip searches and race and gender, and then we would like to examine the correlation between the severity of the case involved and whether the person was under 18 years old at the time of arrest and gender. We are using the *Arrests and Strip Searches* dataset to explore both of these research questions. This dataset includes 65,276 rows of records for 27 variables. This sample size is large enough to allow for statistical testing and more complex procedures. Therefore, the following sections of this paper will include the research hypothesis, literature review, methodology, results and discussion, and conclusions to comprehensively investigate the issue.

2. Literature Review

A strip search is a procedure that is typically conducted by police officers on an arrested person, with clothing removed, to ensure that there are no weapons, drugs, or other contrabands are hidden(Lemke, 2022). This procedure can only be conducted when there is a valid suspicion that the arrested person is in possession of items that could pose a threat to staff and inmates in the correctional facility(Leach & Sabbatine, 1996). In recent years, there has been an ongoing debate about the fairness of using this extreme police power during investigations. According to race-based data released by the Toronto police in June 2022, strip searches are conducted more frequently in arrests involving certain identity groups, particularly Black and Indigenous people(Lemke, 2022). The study shows that although Black people account for 10% of the Toronto population, the likelihood of being strip-searched is over 30%(Lemke, 2022). Similarly, more than one-third of the arrested Indigenous people were subjected to strip searches(Lemke, 2022). This suggests that potential biases may be presented during the decision-making process of the strip search procedure.

According to the report produced by Toronto Police Service on the use of force and strip searches in 2020, there exists gender and race disparities in the use of strip searches by police officers(Phan, Dinca-Panaitescu, & Rebelo, 2021). In 2020, men are 3.3% more likely to be strip searched than women, with a strip search rate of 23% and 19.7% respectively(Phan et al., 2021). Study also shows that male in all perceived race group had a greater chance of being strip searched expect for white individuals. Disproportionality in Strip Search is particularly presented in Black men. Among all male arrests, 27.7% of those are black while they account for 33.4% of the male strip searches(Phan et al., 2021).

Research also shows that age can make a significant difference in criminal behaviour. According to U.S. Department of Justice, youth who ages 17 and under were involved in 15% of all crimes committed in the United States, including 26% of the robbery crimes and 16% of the violent crimes(Puzzanchera, 2010). Among all types of crime, the percentages of arson and vandalism committed by juveniles are significantly greater than other types of offences(47% and 38% respectively). In terms of all arrests related to homicide, 10% of them involved juveniles(Puzzanchera, 2010). Gender disparities also exist in juvenile crimes. Arrests of female juveniles accounted for 29% of the juvenile arrests and this percentage was lower for all types of crimes(Puzzanchera, 2010).

3. Research Hypothesis

3.1 Research Question

To address concerns about the fairness of police enforcement raised by Canadians, we are interested in investigating the potential biases and disparities in the use of strip searches on different demographic groups. More specifically, we aim to examine the role of race and sex in the frequency of strip searches during the arrest process, with a particular focus on identifying the demographic groups that may be treated unfairly during this process.

Also, we are interested in investigating the potential difference in severity of the crimes that lead to an arrest committed by different demographic groups. More specifically, we aim to examine the role of youth status(whether an individual is aged 17 and under) and sex in the severity of crimes, with a particular focus on identifying the groups of individuals that may be at a higher risk of committing severe crimes.

3.2 Research Hypothesis

1. There is a significant difference in the frequency of strip searches conducted on minority groups (Black or Indigenous) that are being arrested compared to other ethnic groups.
2. The number of strip searches conducted on black males is significantly different from people in other demographic groups.
3. The severity of crimes committed by youth aged 17 and under is significantly different from that of adults.
4. The crimes committed by male tend to have a higher severity score than those committed by female.

4. Methods

4.1 Dataset

The primary source of the research is a dataset from Toronto Police Service. This dataset contains information related to all arrests and strip searches including Person ID, Perceived Race, Sex, Age Group, Youth at Arrest (Yes/No), Arrest Location Division, Strip Search (Yes/No), Booked (Yes/No), Occurrency Category, Actions at Arrest, and Search Reason.

Each row of information in his dataset is initially collected by each police division in and out of the City of Toronto, and then compiled and published by a private member of the Toronto Police Service. The Arrests_and_Strip_Searches_(RBDC-ARR-TBL-001).csv contains the records of 65276 arrests from the first quarter of 2020 to the fourth quarter of 2021, and in this research out main variables are:

`PersonID`: The integer data (Nominal) of the identity of the arrested person.
`Perceived_Race`: The string data (Nominal) of the perceived race of the arrested person.
`Sex`: The string data (Nominal) of the gender of the arrested person.

`Youth_at_arrest__under_18_years`: The string data (Nominal) of whether the person arrested was under 18 years old.
`StripSearch`: The integer data (Nominal) of whether the arrested person was strip-searched.
`Booked`: The integer data (Nominal) of whether the arrested person was strip-searched.
`Occurrence_Category`: The string data (Nominal) of the occurrence category of the arrested person.

4.2 Measurement

4.2.1 Data Cleaning

The dataset contained a number of NaN values and does not contain the continuous variables required in this study, so we first cleaned the dataset and created the required new variables.

First, we excluded NaN values for *Sex*, *Perceive_Race*, and *Occurrency_Category*. Since there are 57,475 rows of Search Reasons (57,475 rows by 4 columns) that are all NaN values, accounting for 88.05% of the dataset, and we will not use these variables in this study, they will not be processed this time.

Second, we exclude the arrests involved unisex individuals (N = 9) from the dataset because we are only interested in studying male and female during this study.

Third, since the variable *Youth_at_arrest__under_18_years* has two values with the same meaning but different names, "Youth (aged 17 and younger)" and "Youth (aged 17 years and under)", we merge them to unify the names.

Fourth, similar to the variable *Youth_at_arrest__under_18_years*, the variable *Occurrence_Category* has some same meaningful values but different names, including "Break & Enter" and "Break and Enter", "FTA/FTC/Compliance Check/Parollee" and "FTA/FTC, Compliance Check & Parollee", "Other Statute & Other Incident Type" and "Other Statute/Other Incident Type", "Vehicle Related (inc. Impaired)" and "Vehicle Related", so we also merge them to unify the names.

Fifth, we created a numeric variable called *Score* based on *Occurrence_Category* to measure the severity of the incident that led to the arrest. We assigned a value to each type of incident that appears in *Occurrence_Category* based on the extent to which it would affect the safety of the population. Score is calculated by summing up the pre-assigned values based on the incident that led to each arrest.

Sixth, since one of the variables we wanted to study is the number of times an individual was strip-searched, which was also not available in the dataset, we calculated the number of times each arrestee was strip-searched based on the *PersonID* and *StripSearch* variables. This variable is represented in the newly created dataframe as the *Num_of_Strip_Searches*.

After data cleaning, 65074 observations are obtained.

4.2.2 Measurement

Variable	Value Category
Independent Variables	
Perceived Race of Arrested Person (*Perceived_Race*, Nominal)	White, Unknown or Legacy, Black, South Asian, Indigenous, Middle-Eastern, Latino, East/Southeast Asian
Gender of Arrested Person (*Sex*, Nominal)	M, F
Whether Arrested Person was Youth (*Youth_at_arrest__under_18_years*, Nominal)	Not a youth, Youth (aged 17 years and under)
Dependent Variables	
Severity of Cases Involved (*Score*, Ratio)	Ranging from 5 to 40 in this dataset
Number of Times Each Arrestee was Strip-searched (*Num_of_Strip_Searches*, Ratio)	Ranging from 0 to 13 in this dataset
Source: The Arrests and Strip Searches Dataset from Toronto Police Service	

Table 1. Variables and Value Categories

The first independent variable is the perceived race of the arrested person (*Perceived_Race*, Nominal). This one describes the perceived race of the arrestee. There are multiple ways to describe this variable in real life. For our dataset, the variable has 8 different values, such as Black, White, East/Southeast Asian.

The second independent variable is the gender of the arrested person (*Sex*, Nominal). This one describes the sex of the arrestee. For our dataset, the values of the variable are male (coded as M) or female (coded as F).

The last independent variable is whether the arrested person was under 18 years old or not at the time of arrest (*Youth_at_arrest__under_18_years*, Nominal). For our dataset, the values of the variable are no (coded as Not a youth) or yes (coded as Youth (aged 17 years and under)).

The first dependent variable is the severity of cases involved (*Score*, Ratio). This variable describes the severity of the incident for which the person was arrested. It is measured as a severity score and can take any non-negative number. For our dataset, the range of this variable is from 5 to 40. A higher value indicates a more serious case is involved by the arrestee.

The second dependent variable is the number of times each arrested person was strip-searched (*Num_of_Strip_Searches*, Ratio). The variable describes how many times each arrested individual was strip-searched in total. It can take any non-negative number. For our dataset, the range of this variable is from 0 to 13. The higher the value the more often the individual was strip-searched.

4.3 Data Analysis

To begin with, we used descriptive statistical analysis to show some of the basic statistics of our variables, as well as the sample in general. Next, since there were two research questions in this study, we did separate t-tests for the variables in each research question to verify whether there was a statistically significant relationship. Then, we did a two-way ANOVA for each research question separately. The first one was to determine whether there was a significant difference in the severity of the incident (*Score*) of whether or not the arrestee was under 18 years old at the time of arrest (*Youth_at_arrest__under_18_years*) with 2 different genders (*Sex*). The second one was to determine whether there was a significant difference in the number of times each arrested person was strip-searched (*Num_of_Strip_Searches*) of 8 different perceived races (*Perceived_Race*) with 2 different genders (*Sex*). In addition, we performed Tukey's HSD as post-hoc tests to determine which specific groups or combinations of groups differed significantly from one another and provided more detailed information about the effects of the independent variables on the dependent variable. All statistical analyses were performed in Python.

5. Exploratory Data Analysis

5.1 Descriptive Results

Descriptive statistics allow a basic summary of the sample and variables to be obtained as an initial basis and a part of the subsequent analysis. Looking at background characteristics, most (95%) of those arrested were adults, most (81%) of those arrested were male, and 42% of those arrested were white race. Table 2 displays the descriptive statistics for all analytic variables.

Variables	Median	Mean	SD	Range
Dependent Variables				
Score	10	11.77	7.62	5-40
Num_of_Strip_Searches	0	0.19	0.63	0-13
	Frequency	%		
Independent Variables				
Youth_at_arrest__under_18_years				—
Not a youth	62043	95.34		
Youth (aged 17 years and under)	3031	4.66		
Sex				—
M	52499	80.68		
F	12575	19.32		
Perceived_Race				—

White	27630	42.46
Black	17487	26.87
Unknown or Legacy	5041	7.75
East/Southeast Asian	4402	6.76
South Asian	3603	5.54
Middle-Eastern	3227	4.96
Indigenous	1926	2.96
Latino	1758	2.70

Table 2. Descriptive Statistics for Variables in Two-way ANOVA (N=65074).

Next we conducted Exploratory Data Analysis (EDA). EDA is very useful and there are corresponding EDA methods for different data. It allows us to become more familiar with the data, shave off obvious flaws, and understand the relationships between different variables.

Figure 1. Distribution of Sex

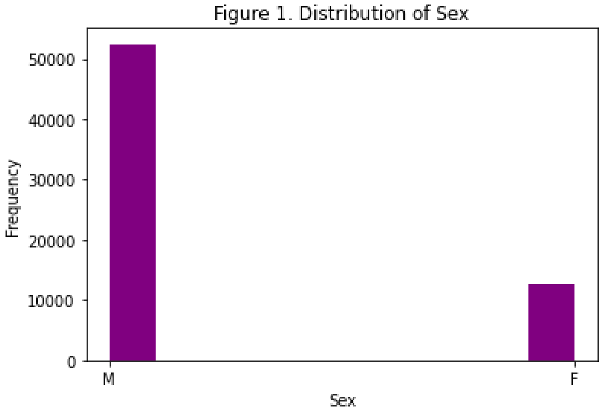

Figure 2. Distribution of Perceived Race

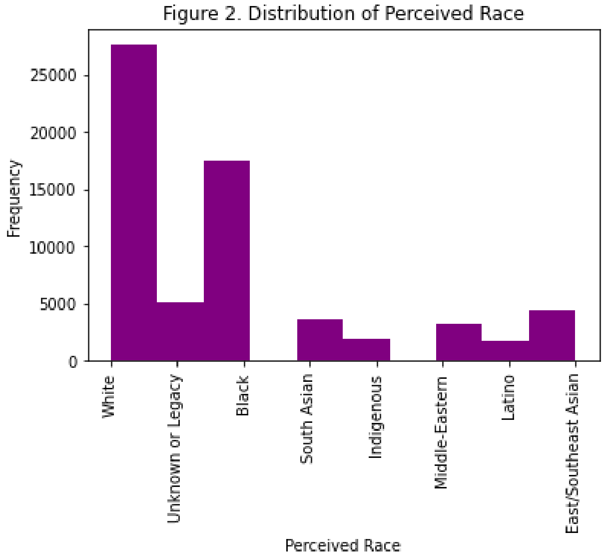

Figure 1 shows the distribution of sex, with the x-axis representing the sex and the y-axis representing the frequency. From this plot, we can tell male arrestees account for a greater proportion in all arrests compared to female arrestees. We can also tell that the data is not unbalanced. Figure 2 shows the distribution of perceived race, with the x-axis representing the perceived race and the y-axis representing the frequency. We can tell that the data is also not unbalanced.

Figure 3. Distribution of Crime Severity Scores

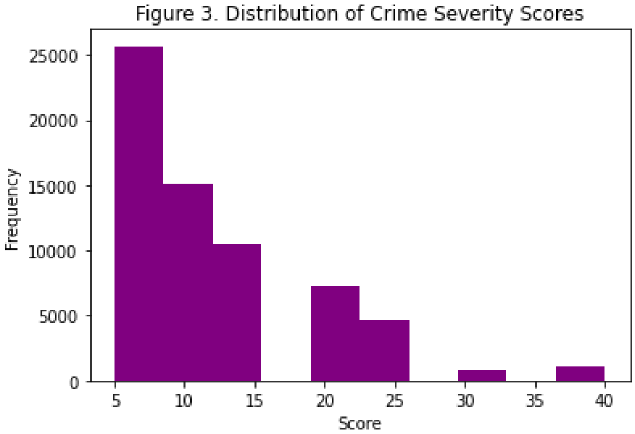

Figure 3 shows the distribution of crime severity scores, with the x-axis representing the scores and the y-axis representing the frequency. From this plot, we can tell the distribution of scores is heavily right-skewed, with more than half of the data points lying between 5 and 15. This suggests that most of the crimes committed are misdemeanors(~15) or below, such as theft and assault. The distribution of scores matches real-life scenarios, which suggests that the scoring system is valid.

Figure 4. Frequency of Arrests in Different Age Groups Grouped by Sex

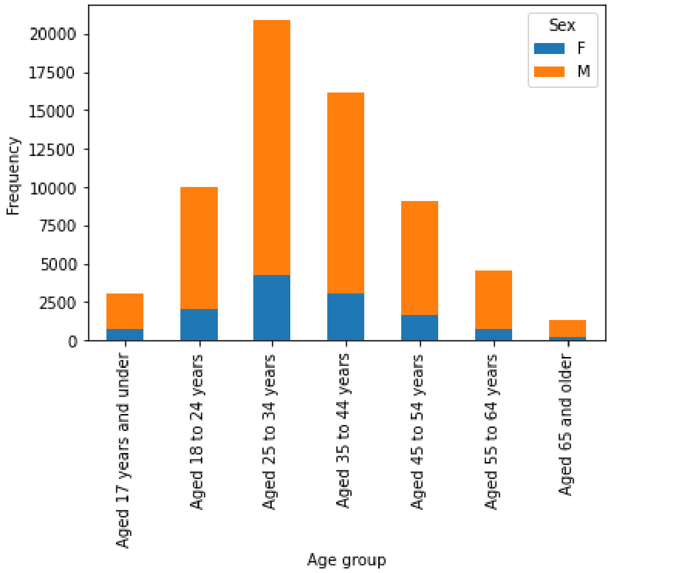

Figure 4 shows the number of arrests in each age group, grouped by sex. The bars are coloured by sex, with x-axis representing different age groups and y axis representing the frequencies. As shown in Figure 4, the number of arrests is significantly higher for males than females among all age groups. The distribution of age follows a normal distribution and the largest age group is 25-34 years old.

Sex Youth_at_arrest__under_18_years	F	M	Total
Not a youth	11843	50200	62043
Youth (aged 17 years and under)	732	2299	3031

Table 3. Contingency Table of Youth Status and Sex

Table 3 is a contingency table of youth status and sex. As shown in Table 2, 5% of all arrests involved youth who aged 17 years and under. Female youths account for a greater proportion in all youth arrests compared to female adults in all adult arrests. We can also tell that the data is highly unbalanced.

Figure 5. Score by Youth Status

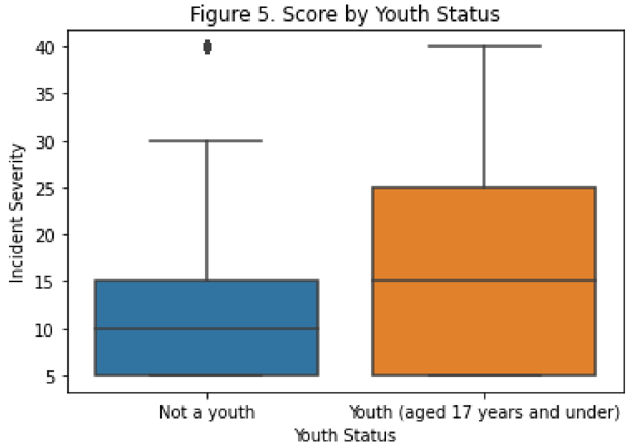

Figure 5 shows the distribution of the crime severity score grouped by youth status. As shown in this plot, the median and the third quantile of the severity score of the crimes committed by youth aged 17 or under are higher than those committed by adults. This potentially suggests that youth aged 17 or under tend to commit to more severe crimes than adults. To further investigate the difference in crime severity, we conducted a t-test and an anova test to validate this finding in later sections.

Figure 6. Distribution of Scores Grouped by Sex and Youth Status

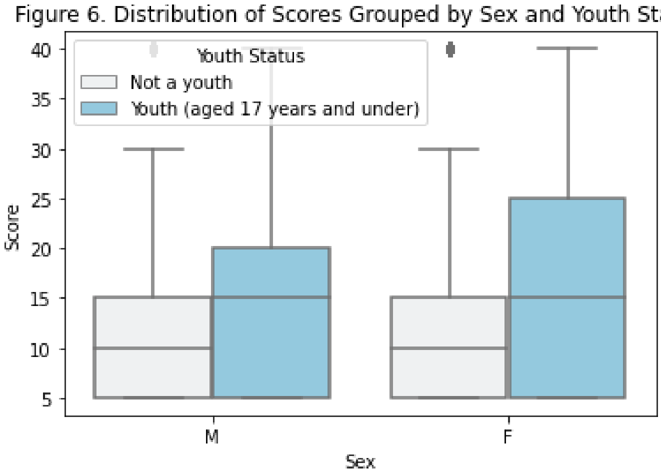

Figure 6 shows the distribution of the crime severity score grouped by sex and youth status. As shown in this plot, the distributions of the severity scores of crimes committed by male adults and female adults are similar, with a median score of 10 and some outliers around 40. The third quantile of the crime severity score for female youth is 25, which is higher than the third quantile for the other three groups. The significance of the difference in mean severity score will be further investigated by t-tests and ANOVA tests in the later section.

StripSearch Sex	0	1	Proportion
M	45983	6516	0.124117
F	11292	1283	0.102028

Table 4. Contingency Table of StripSearch and Sex

Table 4 shows the number of males and females who were strip-searched during the arrest process. As shown in this table, 12.4% of the males were strip-searched and 10.2% of the females were strip-searched. Therefore, we conducted further statistical analysis to investigate the gender disparity.

StripSearch Perceived_Race	0	1	Total Number of Arrests	Strip_Searched_Prop
Indigenous	1620	306	1926	0.158879
Black	15053	2434	17487	0.139189
White	24064	3566	27630	0.129063

Unknown or Legacy	4506	535	5041	0.106130
East/Southeast Asian	4061	341	4402	0.077465
Latino	1626	132	1758	0.075085
South Asian	3346	257	3603	0.071329
Middle-Eastern	2999	228	3227	0.070654

Table 5. Contingency Table of StripSearch and Rerceived Race

Table 5 shows the number of strip searches, the total number of arrests, and the proportion of individuals who were strip-searched during the arrest process by their perceived race. As shown in the table, Indigenous, Black, and White Individuals had the highest likelihood of being strip-searched, with proportions at 15.9%, 13.9%, and 12.9%, respectively. South Asian and Middle-Eastern individuals have a relatively low risk of being strip-searched. Thus, we decided to conduct further statistical analysis to investigate the racial disparities.

5.2 Welch's T-test

We are able to observe a difference in mean of number of strip searches and incident severity score in different demographic groups. To test the significance of differences, we ran several Welch's t-tests with the categorical features. Quantitative data are split into two groups based on the hypothesis that needs to be tested in each t-test. Before running the Welch's t-tests, we checked for the following three assumptions - Normality, Independency, and Homogeneity of variance. According to the Normal QQ-plot, the normality assumption is slightly violated, while the rest of the assumptions are fulfilled.

Sex and Strip Searches

We conducted a Welch's t-test to check the significance of the difference in mean frequency of strip search between male and female. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean frequency of strip searches between males and females.($\mu_{SSMale} = \mu_{SSFemale}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean frequency of strip searches between males and females($\mu_{SSMale} \neq \mu_{SSFemale}$)

The test statistics is 6.36, which indicates that the difference in mean frequency of the strip searches between males and females is 6.36 standard errors away from zero. The p-value is 2.02×10^{-10}, which is less than the significance level of 0.05. 95% confidence interval is [.032, .060], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean frequency of strip searches between males and females is not equal to zero.

Race(Indigenous) and Strip Searches

We conducted a Welch's t-test to check the significance of the difference in mean frequency of strip search between Indigenous people and the other arrested individuals. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean frequency of strip searches between Indigenous people and the other arrested individuals.($\mu_{SSInd} = \mu_{SSnotInd}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean frequency of strip searches between Indigenous people and the other arrested individuals.($\mu_{SSInd} \neq \mu_{SSnotInd}$)

The test statistics is 5.40, which indicates that the difference in mean frequency of the strip searches between Indigenous people and the other arrested individuals is 5.40 standard errors away from zero. The p-value is 8.87×10^{-8}, which is less than the significance level of 0.05. The 95% confidence interval is [.118, .253], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean frequency of strip searches between Indigenous people and the other arrested individuals is not equal to zero.

Race(Black) and Strip Searches

We conducted a Welch's t-test to check the significance of the difference in mean frequency of strip search between Black people and the other arrested individuals. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean frequency of strip searches between Black people and the other arrested individuals. ($\mu_{SSBlack} = \mu_{SSnotBlack}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean frequency of strip searches between Black people and the other arrested individuals. ($\mu_{SSBlack} = \mu_{SSnotBlack}$)

The test statistics is 7.78, which indicates that the difference in mean frequency of the strip searches between Black people and the other arrested individuals is 7.78 standard errors away from zero. The p-value is 7.89×10^{-15}, which is less than the significance level of 0.05. The 95% confidence interval is [.044, .074], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean frequency of strip searches between Black people and the other arrested individuals is not equal to zero.

Race(White) and Strip Searches

We conducted a Welch's t-test to check the significance of the difference in mean frequency of strip search between White people and the other arrested individuals. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean frequency of strip searches between White people and the other arrested individuals. ($\mu_{SSWhite} = \mu_{SSnotWhite}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean frequency of strip searches between White people and the other arrested individuals.($\mu_{SSWhite} \neq \mu_{SSnotWhite}$)

The test statistics is 10.05, which indicates that the difference in mean frequency of the strip searches between White people and the other arrested individuals is 10.05 standard errors away from zero. The p-value is 1.05×10^{-23}, which is less than the significance level of 0.05. The 95% confidence interval is [0.056, 0.083], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean frequency of strip searches between White people and the other arrested individuals is not equal to zero.

Youth Status and Crime Severity

We conducted a Welch's t-test to check the significance of the difference in mean severity score of crime committed by youth aged 17 or under and adult. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean severity scores of crimes committed by youth aged 17 or under and adults. ($\mu_{ScoreYouth} = \mu_{ScoreAdult}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean severity scores of crimes committed by youth aged 17 or under and adults. ($\mu_{ScoreYouth} \neq \mu_{ScoreAdult}$)

The test statistics is -20.34, which indicates that the difference in mean frequency of the strip searches between White people and the other arrested individuals is -20.34 standard errors away from zero. The p-value is 1.11×10^{-86}, which is less than the significance level of 0.05. The 95% confidence interval is [-3.731, -3.075], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean severity scores of crimes committed by youth aged 17 or under and adults is not equal to zero.

Sex and Crime Severity

We conducted a Welch's t-test to check the significance of the difference in mean severity score of crime committed by males and females. Here are the hypotheses that are being tested:

> Null Hypothesis(H_0): There is no significant difference in the mean severity scores of crimes committed by males and females. ($\mu_{ScoreMale} = \mu_{ScoreFemale}$)
> Alternative Hypothesis(H_a): There is significant difference in the mean severity scores of crimes committed by males and females. ($\mu_{ScoreMale} \neq \mu_{ScoreFemale}$)

The test statistics is -4.20, which indicates that the difference in mean frequency of the strip searches between White people and the other arrested individuals is -4.20 standard errors away from zero. The p-value is 2.67×10^{-5}, which is less than the significance level of 0.05. The 95% confidence interval is [-.460, -.167], which does not include zero. Both of the p-value and 95% CI suggest that we are able to reject the null hypothesis and state that the difference is statistically significant. We may conclude that the difference in the mean severity scores of crimes committed by male and female.

6. Two-way ANOVA Tests

Based on the p-value and the 95% Confidence Interval, we were able to reject each null hypothesis and state that the differences were statistically significant. Therefore, we performed the two-way ANOVAs to test whether there were statistically significant differences between the means of multiple different groups of the two explanatory variables on the dependent variable.

Sex, Race and Strip Searches

We conducted a Two-way ANOVA to check the significance of the difference in mean number of strip searches with respect to different races and genders. Before that, we create an interaction plot to visually assess the interaction effect between the two factors.

Figure 7. Effect of Perceived Race and Sex on Number of Strip Searches

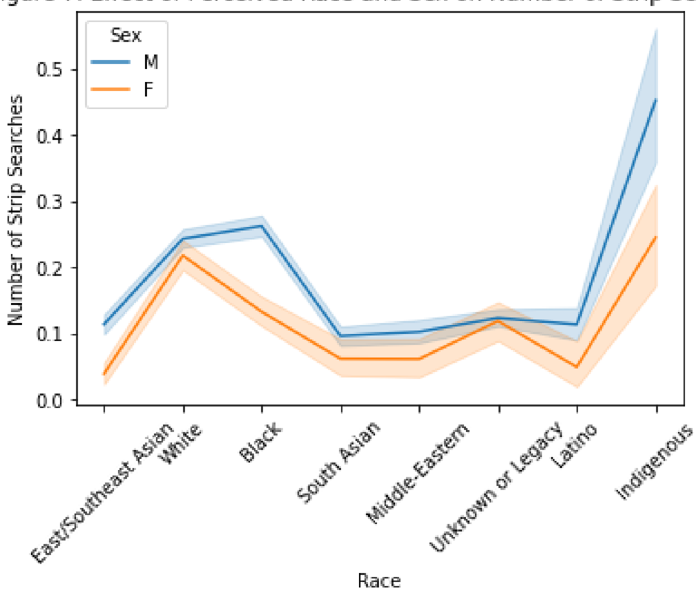

Figure 7 shows the effect of perceived race and sex on the number of strip searches. The x-axis represents the perceived race, the y-axis represents the number of strip searches, and each line on the plot represents a level of sex. The two lines are not parallel, which indicates that there is an interaction effect, and an ANOVA model with an interaction term should be used. Since the sex and perceived race are unbalanced, we should use the unbalanced design of Two-way ANOVA. Here are the hypotheses that are being tested:

Null Hypothesis (H_0):

There is no significant difference in the mean number of strip searches by different perceived races. ($\mu_{StripSearchNumber\_i} = \mu_{StripSearchNumber\_j}\ for\ any\ i{\neq}j$)

There is no significant difference in the mean number of strip searches by male and female. ($\mu_{StripSearchNumber\_Male} = \mu_{StripSearchNumber\_Female}$)

There is no significant interaction effect between perceived races and sex on the mean number of strip searches.

Alternative Hypothesis (H_a):

There is a significant difference in the mean number of strip searches by different perceived races. ($\mu_{StripSearchNumber\_i} \neq \mu_{StripSearchNumber\_j}\ for\ any\ i{\neq}j$)

There is a significant difference in the mean number of strip searches by male and female. ($\mu_{StripSearchNumber\_Male} \neq \mu_{StripSearchNumber\_Female}$)

There is a significant interaction effect between perceived races and sex on the mean number of strip searches.

	sum_sq	df	F	PR(>F)
Intercept	33.920188	1.0	86.985735	1.146767e-20
Perceived_Race	35.772018	7.0	13.104945	5.705875e-17
Sex	26.023355	1.0	66.734909	3.195420e-16
Perceived_Race:Sex	17.354405	7.0	6.357721	1.723438e-07
Residual	15709.962163	40287.0		—

Table 6. Two-way ANOVA Results for Sex, Race and Strip Searches

The F-statistics for perceived race, sex, and the interaction effect were 13.10, 66.73, 6.36, respectively, and their corresponding p-values were $5.71{\times}10^{-17}$, $3.20{\times}10^{-16}$, $1.72{\times}10^{-7}$, which indicated that we were able to reject the null hypothesis and state and the difference is statistically significant. We could conclude that there was a statistically significant difference in *Num_of_Strip_Searches* based on *Perceived_Race*, and in *Num_of_Strip_Searches* based on *Sex*, also there is a statistically significant interaction effect between *Perceived_Race* and *Sex* on *Num_of_Strip_Searches*. To determine which specific groups or combinations of groups differ significantly from one another, we would perform post-hoc tests in later sections.

Sex, Youth Status and Crime Severity

We conducted a Two-way ANOVA to check the significance of the difference in mean severity of crime with respect to different youth status and genders. Before that, we create an interaction plot to visually assess the interaction effect between the two factors.

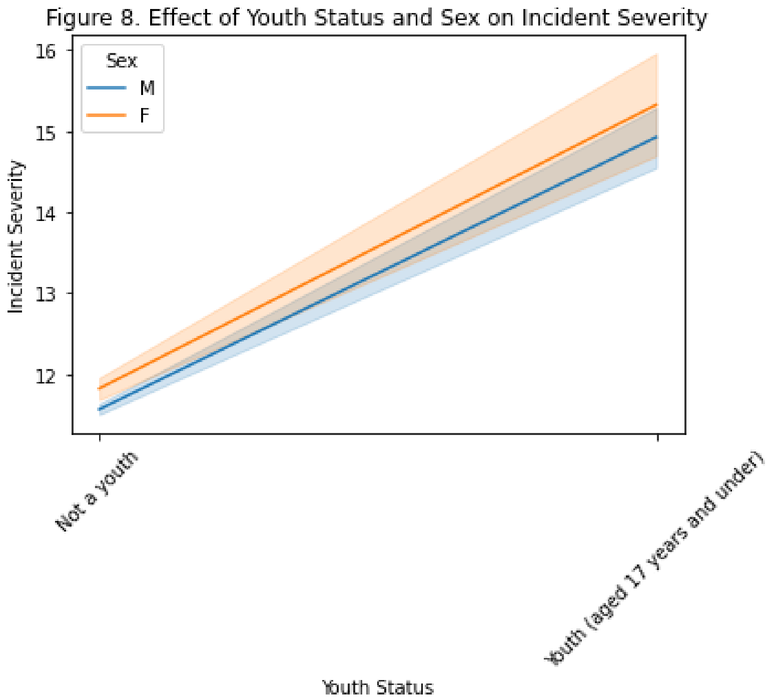

Figure 8 shows the effect of youth status and genders on the severity of crime. The x-axis represents the youth status, the y-axis represents the severity of crime, and each line on the plot represents a level of sex. The two lines are almost parallel, which indicates that there is not an interaction effect, and an ANOVA model without an interaction term should be used. Since the youth status and sex are unbalanced, we should use the unbalanced design of Two-way ANOVA. Here are the hypotheses that are being tested:

Null Hypothesis (H_0):

There is no significant difference in the mean severity of crime by different youth status. ($\mu_{SeverityCrime\_Youth} = \mu_{SeverityCrime\_NotYouth}$)

There is no significant difference in the mean severity of crime by male and female. ($\mu_{SeverityCrime\_Male} = \mu_{SeverityCrime\_Female}$)

Alternative Hypothesis (H_a):

There is a significant difference in the mean severity of crime by different youth status. ($\mu_{SeverityCrime\_Youth} \neq \mu_{SeverityCrime\_NotYouth}$)

There is a significant difference in the mean severity of crime by male and female. ($\mu_{SeverityCrime\_Male} \neq \mu_{SeverityCrime\_Female}$)

	sum_sq	df	F	PR(>F)
Intercept	1.734069e+06	1.0	30103.696831	0.000000e+00
Youth_at_arrest__under_18_years	3.317910e+04	1.0	575.993972	9.947537e-127
Sex	7.107211e+02	1.0	12.338221	4.440785e-04
Residual	3.748298e+06	65071.0		—

Table 7. Two-way ANOVA Results for Sex, Youth Status and Crime Severity

The F-statistics for youth status and sex were 575.99 and 12.34, respectively, and their corresponding p-values were 9.95×10^{-127} and 4.44×10^{-4}, which indicated that we were able to reject the null hypothesis and state and the difference is statistically significant. We could conclude that there was a statistically significant difference in *Score* based on *Youth_at_arrest__under_18_years*, and in *Score* based on *Sex*, also there is a statistically significant interaction effect between *Youth_at_arrest__under_18_years* and *Sex* on *Score*. To determine which specific groups or combinations of groups differ significantly from one another, we would perform post-hoc tests in later sections.

7. Post-hoc Tests(Tukey's HSD)

To determine which pairs of groups have means that are significantly different from each other, we carried out post-hoc analysis using Tukey's HSD.

Crime Severity Score

group1	group2	meandiff	p-adj	lower	upper	reject
Not a youth	Youth (aged 17 years and under)	3.4031	0.001	3.1263	3.6798	True
F	M	-0.3137	0.001	-0.462	-0.1653	True

Table 8. Tukey HSD for Crime Severity Score, FWER=0.05

Table 8 shows the result of the 2 Tukey's tests. The first one compares crime severity scores of two groups, youth aged under 18 and adult. The second one compares crime severity scores of male and female.

According to the result of the first Tukey's tests shown in Table X, the p-value is 0.001 and mean difference is 3.403. The 95% confidence interval for the difference in means is [3.126, 3.680], which does not include zero. Both the p-value and the confidence interval suggest that the difference in mean crime severity score between youth and adult is statistically significant. The average severity score of crimes committed by youths aged 17 and under is 3.403 higher than that of those committed by adults.

For the second Tukey's test, the p-value is 0.001 and mean difference is -0.314. The 95% confidence interval for the difference in means is [-0.462, -0.165], which does not include zero. Both the p-value and the confidence interval suggest that the difference in mean crime severity score between male and female is statistically significant. The average severity score of crimes committed by female is 0.314 higher than that of those committed by male.

Number of Strip Searches

group1	group2	meandiff	p-adj	lower	upper	reject
Black	East/Southeast Asian	-0.1383	0.001	-0.1757	-0.1008	True
Black	Indigenous	0.1374	0.001	0.0684	0.2064	True
Black	Latino	-0.1347	0.001	-0.1909	-0.0784	True
Black	Middle-Eastern	-0.1406	0.001	-0.184	-0.0972	True
Black	South Asian	-0.147	0.001	-0.1873	-0.1068	True
Black	Unknown or Legacy	-0.1154	0.001	-0.1497	-0.0812	True
Black	White	-0.001	0.9	-0.0252	0.0233	False
East/Southeast Asian	Indigenous	0.2757	0.001	0.2018	0.3496	True
East/Southeast Asian	Latino	0.0036	0.9	-0.0585	0.0658	False
East/Southeast Asian	Middle-Eastern	-0.0023	0.9	-0.0532	0.0485	False
East/Southeast Asian	South Asian	-0.0088	0.9	-0.0569	0.0394	False
East/Southeast Asian	Unknown or Legacy	0.0229	0.7226	-0.0204	0.0661	False
East/Southeast Asian	White	0.1373	0.001	0.1014	0.1732	True
Indigenous	Latino	-0.2721	0.001	-0.3571	-0.1871	True
Indigenous	Middle-Eastern	-0.278	0.001	-0.3551	-0.2009	True
Indigenous	South Asian	-0.2845	0.001	-0.3598	-0.2091	True
Indigenous	Unknown or Legacy	-0.2528	0.001	-0.3252	-0.1805	True
Indigenous	White	-0.1384	0.001	-0.2065	-0.0702	True
Latino	Middle-Eastern	-0.0059	0.9	-0.0719	0.06	False
Latino	South Asian	-0.0124	0.9	-0.0763	0.0515	False
Latino	Unknown or Legacy	0.0193	0.9	-0.041	0.0795	False
Latino	White	0.1337	0.001	0.0785	0.189	True
Middle-Eastern	South Asian	-0.0064	0.9	-0.0594	0.0465	False
Middle-Eastern	Unknown or Legacy	0.0252	0.7395	-0.0234	0.0737	False

Middle-Eastern	White	0.1396	0.001	0.0975	0.1818	True
South Asian	Unknown or Legacy	0.0316	0.4215	-0.0142	0.0774	False
South Asian	White	0.1461	0.001	0.1072	0.1849	True
Unknown or Legacy	White	0.1145	0.001	0.0819	0.147	True

Table 9. Tukey HSD for Perceived Race on Number of Strip Searches, FWER=0.05

Table 9 shows the results of the Tukey's tests conducted to identify which pairs of groups have significantly different mean strip search values. Based on this table, we can tell that the difference in the average number of strip searches between Black Individuals and any other ethnic group, except for White individuals, is statistically significant. The p-values for Black individuals versus all ethnic groups except for White is 0.001, and the mean difference is approximately -0.14. Confidence intervals do not include zero. Both the p-value and the confidence interval suggest that the difference in average number of strip searches is statistically significant. On average, the number of strip searches performed on Black Individuals is approximately 0.14 higher than for other ethnic groups, except for White.

A similar result can be seen for indigenous people. The p-value for the comparison between Indigenous and other groups is less than 0.05, indicating a statistically significant difference in the mean. The mean of differences is approximately -0.275. These results support part of our first hypothesis. However, the differences in mean between White individuals and other ethnic groups are also significant, with p-values less than 0.05, suggesting that not only minority groups but also White individuals are treated differently during the arrest process.

group1	group2	meandiff	p-adj	lower	upper	reject
F	M	0.0458	0.001	0.0305	0.0611	True

Table 10. Tukey HSD for Sex on Number of Strip Searches, FWER=0.05

In the second Tukey's test for male and female, the p-value is 0.001, and mean difference is 0.0458. The 95% confidence interval for the difference in means is [-.0305, -.0611], which does not include zero. Both the p-value and the confidence interval suggest that the difference in number of strip searches between male and female is statistically significant. The average number of strip searches performed on males is 0.0458 higher than that of those performed on females.

group1	group2	meandiff	p-adj	lower	upper	reject
Black x M	East/Southeast Asian x F	-0.2233	0.001	-0.311	-0.1357	True
Black x M	East/Southeast Asian x M	-0.1484	0.001	-0.1954	-0.1015	True
Black x M	Indigenous x F	-0.0172	0.9	-0.1426	0.1081	False
Black x M	Indigenous x M	0.19	0.001	0.0926	0.2873	True

Black x M	Latino x F	-0.213	0.001	-0.365	-0.061	True
Black x M	Latino x M	-0.1487	0.001	-0.2181	-0.0793	True
Black x M	Middle-Eastern x F	-0.2008	0.001	-0.328	-0.0737	True
Black x M	Middle-Eastern x M	-0.1597	0.001	-0.2125	-0.1069	True
Black x M	South Asian x F	-0.2004	0.001	-0.3055	-0.0952	True
Black x M	South Asian x M	-0.166	0.001	-0.2156	-0.1163	True
Black x M	Unknown or Legacy x F	-0.1435	0.001	-0.2223	-0.0646	True
Black x M	Unknown or Legacy x M	-0.1388	0.001	-0.1817	-0.0959	True
Black x M	White x F	-0.0444	0.033	-0.0872	-0.0016	True
Black x M	White x M	-0.0192	0.7121	-0.05	0.0116	False

Table 11. Tukey HSD for Interaction on Number of Strip Searches (Part), FWER=0.05

When taking a closer look into the interaction effects, we find that Black males have statistically different means compared to most other demographic groups, except for Indigenous females and White males. The p-values are less than 0.05 and confidence intervals do not include zero. On average, the number of strip searches performed on Black males are higher than that of most other demographic groups, which proves our second hypothesis.

8. Discussion

8.1 Conclusion

Based on the analyses that have been completed, we have two conclusions. The first conclusion is that the mean values of crime severity are significantly different by youth status and gender. More specifically, youth aged 17 and under had higher mean severity scores for crime than adults, and females had higher mean severity scores for crime than males. The reasons for this may be that young people may be more impulsive, less able to weigh the consequences of their actions, less mature in their decision-making skills, and more susceptible to peer pressure or negative influences.

The second finding was that the mean number of strip searches differed significantly by race and gender. More specifically, Blacks and Indigenous Peoples had significantly higher mean number of strip searches than most other races, and males had significantly higher means of being strip-searched than females. Additionally, Black males were strip-searched more often than most other population groups. The reasons for this may be biases held by law enforcement officers, differences in policing practices, and broader social issues such as racial profiling and discrimination.

Overall, these findings suggest that disparities exist in the criminal justice system based on demographic characteristics such as age, gender, and race. These disparities may have important implications for the fair and equitable treatment of individuals and warrant further investigation and potential policy changes to address these inequities.

8.2 Limitation

Firstly, the arrests and strip searches dataset is a large dataset that contains more than 60,000 observations. During the study, we used 0.05 as a threshold for significance. However, when working with large datasets, a p-value less than 0.05 may not always be an appropriate standard to determine statistical significance. As sample size increases, there is a higher chance of obtaining a significant result, which means, in some cases, we may be seeing significance by accident(Bruhn, 2021). This suggests that rejecting every null-hypothesis based on whether or not p-value is less than 0.05 may not accurately reflect the true significance of the result we obtained(Bruhn, 2021). We should adjust the significance threshold based on sample size accordingly to get more valid results.

Secondly, the normality assumption is not met before conducting t-test and ANOVA, and this could potentially affect the accuracy of the result. It would be better if considering using non-parametric tests and regression models for analysis.

Lastly, crime severity scores are assigned based on the severity of the case associated with average jail time as a reference, which could be biased. For example, different types of sexual related crimes can have different degrees of severity. Assigning a single value to all such cases may fail to accurately reflect the true severity of the incident. It would be better to have a column recording the severity of each case at the time of arrest for future studies.

References

Bruhn, A. (2021). Stop Using p<0.05. Retrieved 25th Feb, 2023 from https://towardsdatascienc e.com/stop-using-p-0-05-9743e5cddc21

Civilian Review and Complaints Commission. (2022). Review of the RCMP's Bias-Free Policing Model Report. https://www.crcc-ccetp.gc.ca/en/review-rcmps-bias-free-policing-model-report

Dyna, I. (2020). Public perceptions of the police in Canada's provinces, 2019. Retrieved 25th Feb, 2023 from https://www150.statcan.gc.ca/n1/pub/85-002-x/2020001/article/00014-eng.htm

Lemke, M. (2022). Policing Toronto: Strip Searching in a Divided City. Retrieved 25th Feb, 2023 from https://www.ojp. gov/ncjrs/virtual-library/abstracts/new-strip-search-paradigm#:~: text=Under%20the%20proposed%2 0paradigm%2C% 20a,so%20as%20to%20protect%20the

Leach, D. & Sabbatine , R. (1996). New Strip Search Paradigm. Retrieved 25th Feb, 2023 from https://www.ojp.gov/ncjrs/ virtual-library/abstracts/new-strip-searchparadigm#:~:text=Under% 20the%20proposed%20 paradigm %2C%20a,so%20as% 20to%20protect%20the

Mary K. A. & Tamy S. (2016). Youth crime in Canada, 2014. Retrieved 25th Feb, 2023 from https://www150.statcan.gc.ca/n1/pub/85-002-x/2016001/article/14309-eng.htm

Phan, M. B., Dinca-Panaitescu, M., & Rebelo, N. (2021). Understanding Strip Searches in 2020. Retrieved 25th Feb, 2023 from https://www.tps.ca/media/filer_public/e4/b1/e4b1 b125-2a2e -4d69-ad02-77ab3f3d5878/4e217e01-3cd6-4fe8-8898-39 cf8693e871.pdf

Puzzanchera, C. (2010). Juvenile Arrests (2007). Diane Publishing.

List of Appendices

Figure A1. Normal QQ Plot - Severity Score(Male)

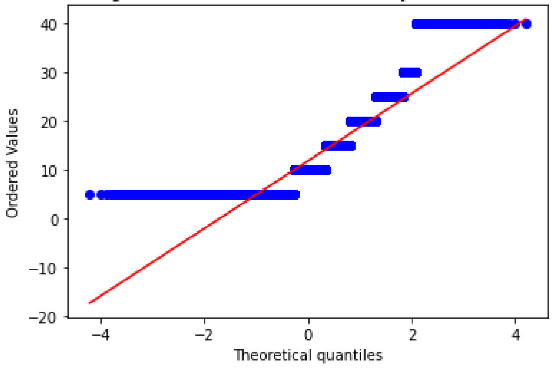

Figure A2. Normal QQ Plot - Severity Score(Female)

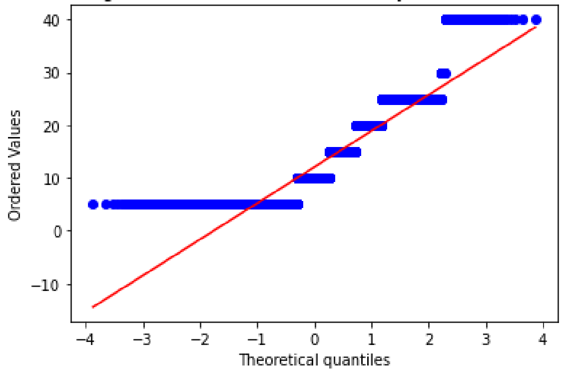

Figure A3. Normal QQ Plot - Severity Score(Youth)

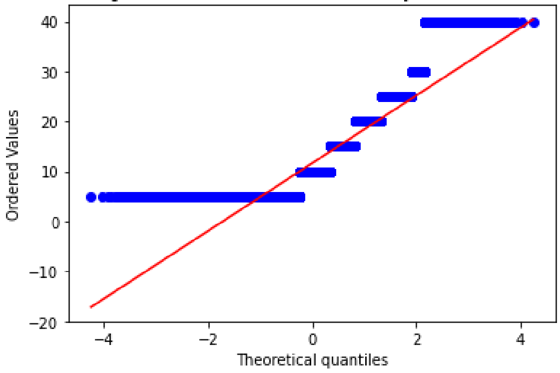

Figure A4. Normal QQ Plot - Severity Score(Adult)

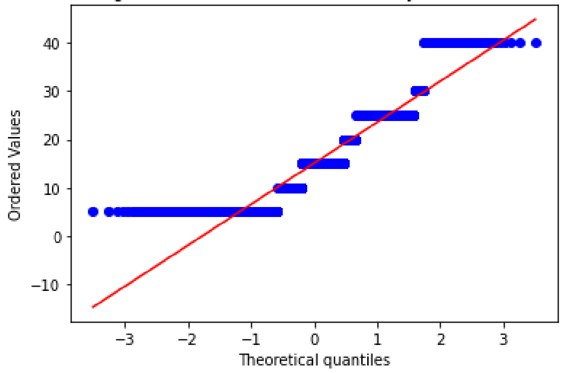

Figure A5. Variance of Score Grouped by Sex and Youth Status