

INF2178 Experimental Design for Data Science

Steve Li

Bhanvi Gupta

April 16th, 2023

Insights from Arrests Made by the Toronto Police Service, 2020-2021

Final paper submitted in conformity with the requirements for the course of INF2178H (Winter
2023)

Link to Google Colab notebook for the code:

https://colab.research.google.com/drive/1IsOO_TYh-Nj0RmggeESiuNF_V1tdIoWo?usp=sharing



UNIVERSITY OF
TORONTO

Faculty of Information
University of Toronto

Table of Contents

Table of Contents	1
1. Introduction	2
1.1 Background	2
1.2 Literature Review	2
1.2 Research Questions and Objectives	4
2. Exploratory Data Analysis (EDA)	5
2.1 Descriptive statistics	5
2.2 T-tests	8
2.2.1 T-test #1	8
2.2.2 T-test #2	8
3. Method	9
3.1 Dataset Description	9
3.2 Power Analysis	10
3.3 ANCOVA	11
3.4 Logistic Regression	13
4. Results	13
4.1 T-tests	13
4.2 ANCOVA	14
4.3 Logistic Regression	15
Discussion	15
Conclusion	17
References	18

1. Introduction

1.1 Background

Justice in policing has been a key social and ethical issue that continues to garner people's attention and spark debates over the years. As citizens with conscience and a moral compass, it is natural for us to turn our attention to injustices in policing, such as police brutality and the disproportional use of policing on certain racial or ethnic groups. While many sensational and outrageous events, such as the shooting of Michael Brown in Ferguson, Missouri in 2014 and the killing of George Floyd in Minneapolis, Minnesota in 2020, take place every few years and cause public attention and outrage, many instances of police injustice get considerably less attention and thus get overlooked. To continue the scrutiny on police forces and their use of violence and resources, we can use statistical analysis as a tool to examine for any injustices. For this project, a dataset on arrests and strip searches released by the Toronto Police Service (TPS) is used. This project aims to explore the relationship between the number of bookings and a person's demographics such as race, gender, and age group.

1.2 Literature Review

Arrests in Toronto have been the focus of a number of studies in recent years. In this literature review, we will summarize the main findings from some of the key studies in this area. One area of research has been focused on the relationship between race and arrests in Toronto. A study by Owusu-Bempah and Wortley (2011) found that black and brown people were overrepresented in Toronto's jail population, and that this was due to systemic factors such as discrimination, poverty, and biased policing. Similarly, a study by Tanovich (2016) found that racial profiling was a significant problem in Toronto, with police more likely to stop and search individuals from racialized communities.

Another area of research has explored the impact of arrests on individuals and communities in Toronto. A study by Roach (2016) found that even a brief period of detention can have a significant impact on an individual's life, including their

employment and housing opportunities. Similarly, a study by Clarke and Wortley (2019) found that the overrepresentation of black people in Toronto's criminal justice system had a negative impact on their mental health and well-being.

A number of studies have also examined the use of police discretion in arrests in Toronto. A study by Brian C. Renauer, Emma Covelli (2011) found that police officers often used their discretion when deciding whether or not to make an arrest, and that this discretion could be influenced by factors such as the offender's demeanor, the seriousness of the offense, and the officer's personal biases.

In general, younger age groups tend to have higher rates of arrests for crimes such as vandalism, drug offenses, and property crimes. However, older age groups may be more likely to be arrested for white-collar crimes, such as fraud or embezzlement.

According to data from the Toronto Police Service, in 2020, the highest number of arrests were made for individuals between the ages of 18-24, accounting for 34.7% of all arrests. The next highest age group was individuals between the ages of 25-34, who accounted for 27.5% of all arrests. This suggests that younger age groups are more likely to be involved in criminal activity than older age groups. It was also found that males were more likely to be arrested for crimes than females. Males accounted for 74.1% of all arrests, while females accounted for 25.7%. This trend is consistent with national and international patterns in which males are overrepresented in the criminal justice system.

There are also gender-specific crimes that occur in Toronto, such as sexual assault and domestic violence. In general, women are more likely to be victims of these crimes, while men are more likely to be perpetrators. The Toronto Police Service reports that in 2020, 85% of all reported sexual assaults and 80% of all reported domestic violence incidents involved female victims.

It is important to note that the overrepresentation of males in the criminal justice system and the higher rates of victimization among women are complex issues that are

influenced by a variety of social, cultural, and economic factors. It's crucial to address the root causes of these issues in order to create safer communities for all individuals.

In conclusion, arrests in Toronto have been studied from various perspectives, including the relationship between race and arrests, the impact of arrests on individuals and communities, the use of police discretion in arrests, and the use of alternatives to arrests. The findings from these studies highlight the need for ongoing research and policy interventions to address issues related to arrests in Toronto, particularly those related to racial discrimination and bias in policing.

According to the studies available on statscan, though firearms were present in a small proportion of all violent crimes, incidents where a firearm was present more commonly involved multiple victims and resulted in major injury or death compared to incidents where another type of weapon or no weapon was present.

In 2021, there were 297 victims of homicide in Canada where shooting was the primary cause of death, a rate of 0.78 per 100,000 population. This marked the highest rate since 1992.

1.2 Research Questions and Objectives

The first objective of our study is to find out the relationship between demographic factors such as gender, youth status, and race and the number of bookings. The second research objective is to find out if a person's number of arrests and sex have any influence on whether that person is booked at a police station following an arrest. Four research questions are investigated in this project.

- Is there a difference in the number of bookings between genders? (RQ1)
- Is there a difference in the number of bookings between youths and non-youths? (RQ2)
- Is there a difference in the number of bookings based on the person's perceived race (White, Black or Latino), controlled by their number of arrests (RQ3)?
- Do the number of arrests and sex (male or female) have an influence on a person being booked at a police station (yes or no) (RQ4)?

2. Exploratory Data Analysis (EDA)

2.1 Descriptive statistics

Count	64798
Mean	2.37
Standard Deviation	3.51
Minimum	0
25th Percentile	0
50th Percentile	1
75th Percentile	3
Maximum	42

Table 2.1.1 Descriptive Info for Number of Bookings

Number of Bookings	Occurrence Frequency
0	19128
1	18634
2	8308
3	5015
...	...
24	31
21	28

Table 2.1.2 Occurrence Frequency of the Number of Bookings

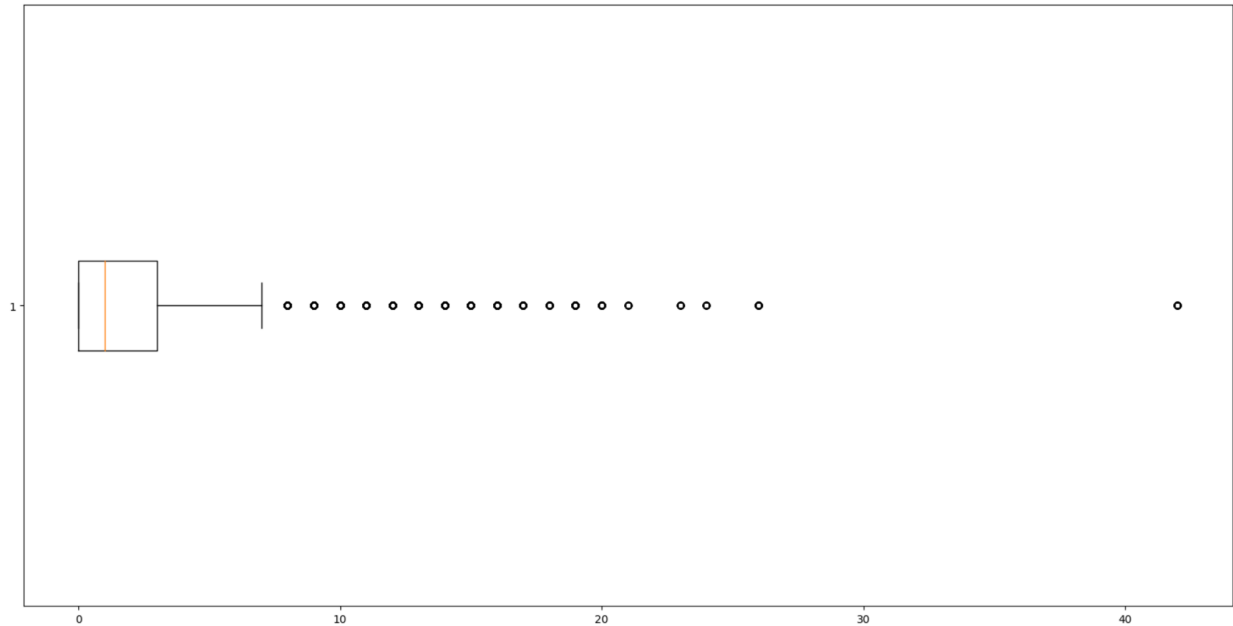


Figure 2.1.3 Horizontal Boxplot for Number of Bookings

Table 2.1.1 shows that the highest number of times that a person has been booked is 42, and the lowest number of times that a person has been booked is 0. The average number of times that a person is booked is 2.37. The minimum of 0 shows that, for the 64798 arrests made, not every arrest led to the person arrested being booked at a police station. This leads to the question: how many arrests led to no bookings?

Table 2.1.2 gave us the answer. For 19128 arrests, no bookings were made. The total number of bookings is $64798 - 19128 = 45670$

From the boxplot in Figure 2.1.3, we can see that there are a lot of outliers for the number of bookings, including some extreme outliers. The presence of extreme outliers drives up the mean and may affect the effectiveness of the t-tests. To deal with the extreme outliers, we decided to exclude the numbers of bookings that are greater than or equal to 20.

Count	64480
Mean	2.25
Standard Deviation	3.06

Minimum	0
25th Percentile	0
50th Percentile	1
75th Percentile	3
Maximum	19

Table 2.1.4 Descriptive Info for Valid Number of Bookings

Table 2.1.4 shows the descriptive statistics of the number of bookings after the extreme outliers greater than or equal to 20 have been removed. The mean number of bookings is 2.25. The standard deviation is 3.06. The 25th percentile is 0. The median is 1, and the 75th percentile is 3.

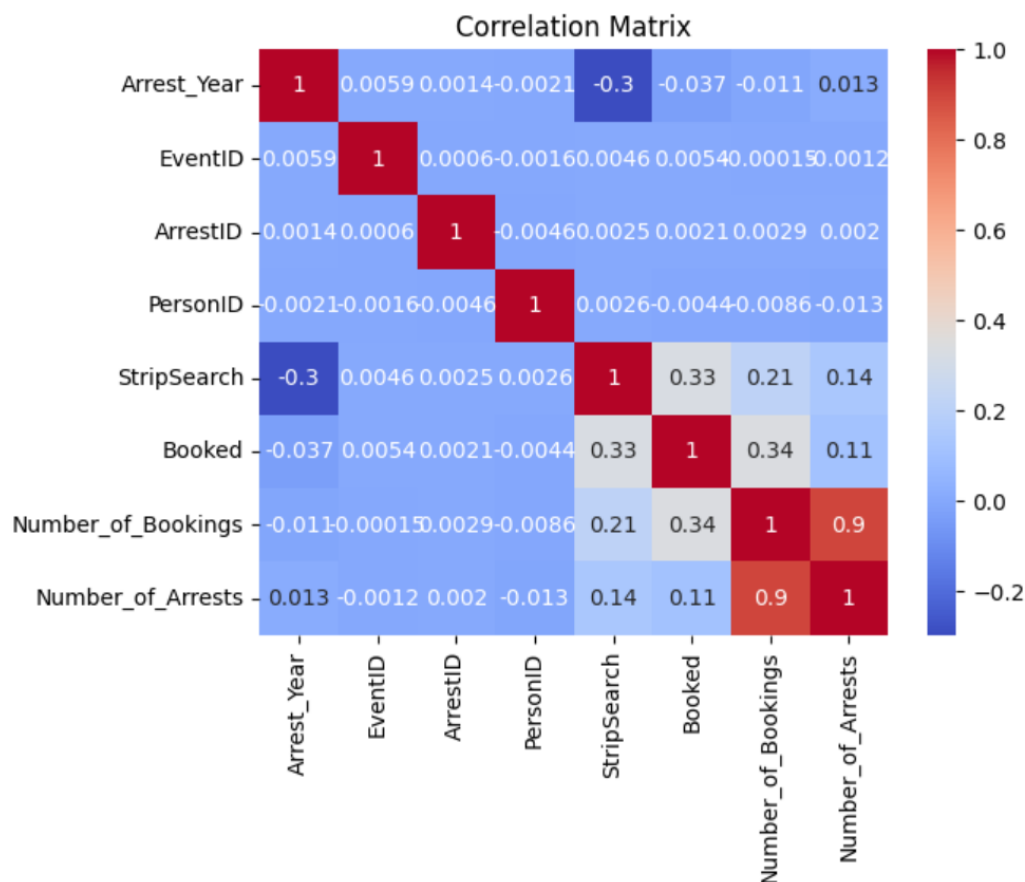


Figure 2.1.5 Correlation Matrix

From the correlation matrix in Figure 2.1.5, we can see that youth_at _arrest_under_18_years and sex are highly correlated. While this is not going to change

our estimates, it can still lead to multicollinearity. Since we want to study the relationship between age group, sex and race in further sections, we will not be dropping these variables for goodness of fit.

2.2 T-tests

Prior to running the t-tests, the assumptions for t-tests were checked. For both t-tests:

- The two samples are mutually exclusive and therefore independent,
- Neither of the variables violates the assumption of normality,
- And there is homogeneity of variances.

2.2.1 T-test #1

The purpose of the first t-test is to investigate the following research question (RQ1): Is there a difference in the number of bookings between genders?

The person's gender (male or female) is the independent variable measured with two categories, and the number of bookings is the dependent, continuous variable. In the dataset, the number of bookings for males is much higher than that of females. To ensure the accuracy of the t-test, we wanted to make sure the sample size of each group is the same. Thus, we took random samples of 10000 from each gender group for the t-test.

For the first t-test, the null and alternative hypotheses are as follows:

- Null hypothesis: On average, males and females DO NOT differ in the number of bookings. ($H_0: \mu_1 = \mu_2$)
- Alternative hypothesis: On average, males and females DO differ in the number of bookings. ($H_a: \mu_1 \neq \mu_2$)

2.2.2 T-test #2

The purpose of the second t-test is to investigate the following research question (RQ2): Is there a difference in the number of bookings between youths and non-youths?

The person's youth status (youth or not a youth) is the independent variable with two categories, and the number of bookings is the dependent, continuous variable. In

the dataset, the number of bookings for non-youths is much higher than that of youths. To ensure the accuracy of the t-test, we wanted to make sure the sample size of each group is the same. Thus, we took random samples of 10000 from each group for the t-test.

For the 2nd t-test, the null and alternative hypotheses are as follows:

- Null hypothesis: On average, non-youths and youths DO NOT differ in the number of bookings. ($H_0: \mu_1 = \mu_2$)
- Alternative hypothesis: On average, non-youths and youths DO differ in the number of bookings. ($H_a: \mu_1 \neq \mu_2$)

3. Method

3.1 Dataset Description

The dataset used for this study was released by the Toronto Police Service in November 2022. It contains information on arrests and strip searches in the years 2020 and 2021. Although not explicitly stated as so, it is inferred that the dataset includes all arrests and strip searches that took place in these two years. No exact date and time is included in the dataset – only the year (2020 and 2021) and month range (January to March, April to June, July to September, or October to December) are given. The dataset also includes the event ID and arrest ID of each arrest or strip search event. Regarding personal information, for each person involved in the arrest or strip search event, the person ID, perceived race, sex, age group at arrest, and their youth status (whether they are 17 years or younger) are recorded. Regarding the location of the arrest, a two-digit location code is included for the applicable entries.

The dataset also includes information on whether the person has undergone a strip search and whether they were booked at a police station within 24 hours of the arrest. Regarding the arrest event itself, the dataset includes the category of the arrest. The person's actions during the arrest, for instance, if they were cooperative, or if they assaulted the officer, is included. For strip searches, the reason that prompted the search was recorded.

As a part of our data preprocessing, we created a continuous variable called the 'Number_of_Arrests' which calculates the number of arrests per person using PersonID. We also created a continuous variable named "Number_of_Bookings" which shows how many times a person has been booked at a police station. This "Number_of_Bookings" variable is used as the independent variable in the t-tests and ANCOVA. We also eradicated all duplicate entries and null entries present in the dataset. Since we have a high number of entries in the dataset, we did not go for data substitution.

3.2 Power Analysis

Prior to conducting any experiment on the dataset, a power analysis was performed. In the power analysis, we calculated the sample size needed for each sample group to achieve a power of 0.8 for t-test #1. T-test #1 sought to analyze whether a students' final grade (outcome variable) differed between family sizes less than and greater than three (3) (two-level explanatory variable). In the power analysis, we calculated the effect size of the explanatory variable using Cohen's D metric, which was 0.13.

Number of Observations by Gender	Sample Sizes Needed to Achieve Power of 0.8	Actual Sample Sizes
Males	2588	50695
Females	623	12206

Table 3.2.1

After obtaining the effect size, the required sample size was computed using the obtained effect size and establishing the statistical power at 80%. As Table 3.2.1 shows, a sample size of 2588 was required for males, while a sample size of 623 was required for females. This is significant because the sample sizes provided in the dataset are 12206 and 56895 respectively. Because the actual sample sizes exceed the sample sizes needed to achieve statistical power of 0.8, we do not need to worry about low sample sizes negatively affecting the statistical power.

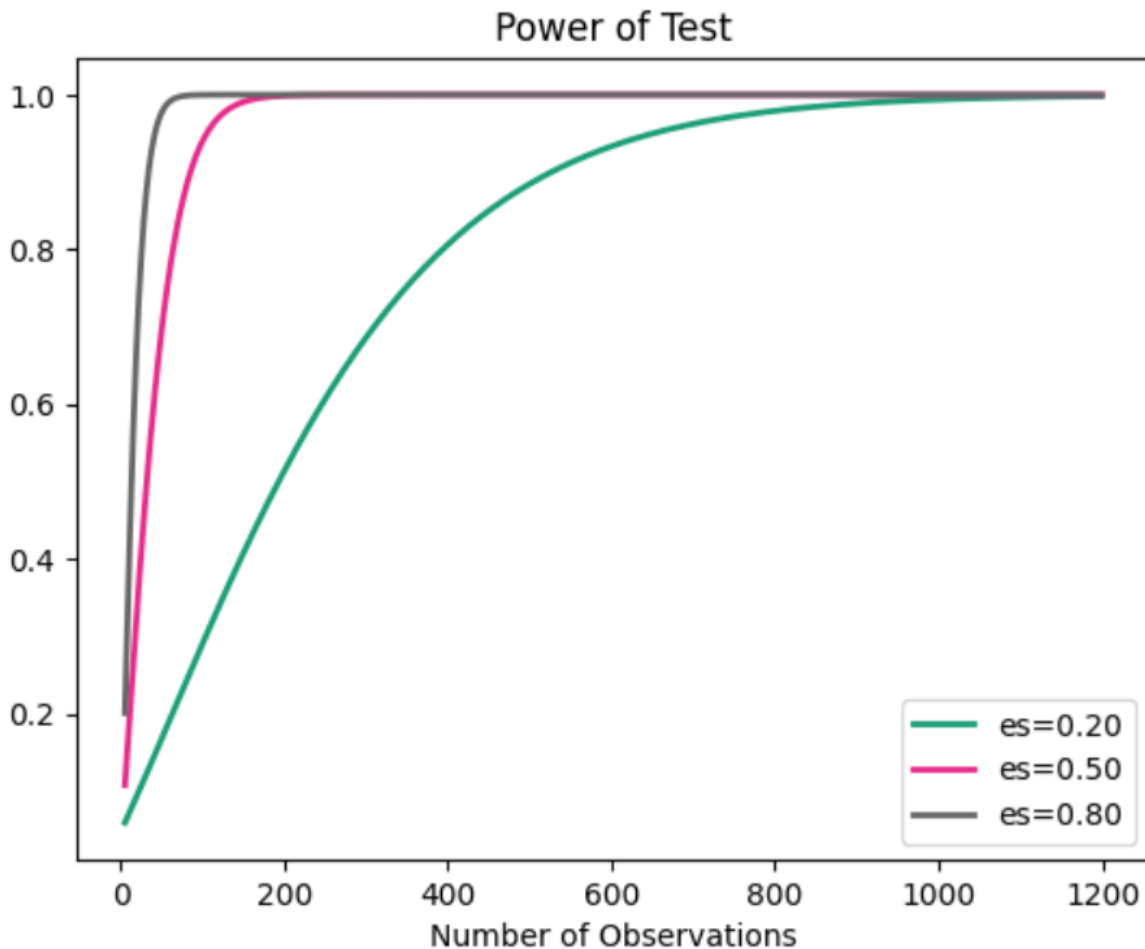


Figure 3.2.2 Power Analysis Curve

From Figure 3.2.2, which shows the power analysis curve, we can see that for all three effect sizes, as the number of observations surpasses 1000, the statistical power gets very close to 1.0. The risk of getting a Type II error gets very low.

3.3 ANCOVA

To investigate Research Question #3, We conducted an ANCOVA. The research question investigated in the first one-way ANOVA is whether there is a difference in the number of bookings based on the person's perceived race (White, Black or Latino), controlled by their number of arrests (RQ3).

The person's perceived race is the independent variable with three categorical and independent groups (White, Black or Latino), and the number of bookings is the dependent, continuous variable. The number of arrests for that person, also a continuous variable, is chosen as the covariate. This is because we want to control the effect of the number of arrests on the number of times being booked. One possible effect of the number of arrests on the number of times being booked is that, because the person has a lengthy arrest history, they are more likely to be booked following an arrest. For our analysis, we want to control for such possible effects.

Prior to conducting the ANCOVA, the following assumptions were checked.

- Residuals (experimental error) are approximately normally distributed: This was tested using the Shapiro-Wilk test. Since the p-value is less than .05, we reject the null hypothesis of the Shapiro-Wilk test. This means we have sufficient evidence to say that the sample data does not come from a normal distribution.
- Homogeneity of variances: A Levene's Test was performed to check for homogeneity of variances. Because the p-value is smaller than 0.05, we reject the null hypothesis and conclude that the variance among different groups is not equal. In other words, at least one pair the variance is not equal to others.
- Observations are sampled independently from each other: There is no relationship between the three groups as they come from three different racial groups.
- The dependent variable (Number of Bookings) is measured at the continuous level.

The null and alternative hypotheses for the one-way ANCOVA are as follows:

- Null hypothesis H_0 : $\mu_1 = \mu_2 = \mu_3$. The means of the number of bookings for the three perceived races are equal, controlling for the number of arrests.
- Alternative hypothesis H_a : at least one perceived race's mean number of bookings is different from those of the other racial groups, controlling for the number of arrests.

In addition, to ensure a balanced design, we need to make sure the sample sizes for each group are equal. Thus, we selected 1000 random samples from each group.

3.4 Logistic Regression

A logistic regression was conducted to investigate Research Question #4, which is stated as follows: Do the number of arrests and sex (male or female) have an influence on a person being booked at a police station (yes or no)?

The input variables are the number of arrests, a numerical, continuous variable, and the person's sex, with two categorical and independent groups (male or female). The response variable is whether a person was booked at a police station following an arrest, a binary variable.

4. Results

4.1 T-tests

Variable	Sample Size	Mean	Standard Deviation
Male	10000	2.01	2.45
Female	10000	1.73	2.39
Combined	20000	1.87	2.42
Degrees of Freedom	19998		
t	8.07		
p-value	0.00		

Table 4.1.1 Results for t-test #1

For T-test #1, an independent samples t-test was run to determine if there were differences in the number of bookings between males and females. The mean number of bookings for males ($M = 2.01$, $SD = 2.45$) was higher than that of females ($M = 1.73$, $SD = 2.39$), a statistically significant difference, $t(19998) = 8.07$, $p = 0.00 < 0.05$.

Variable	Sample Size	Mean	Standard Deviation
Not a Youth	10000	2.01	2.46
Youth	10000	1.18	1.67
Combined	20000	1.59	2.14
Degrees of Freedom	19998		
t	27.94		
p-value	0.00		

Table 4.2.2 Results for t-test #2

For T-test #2, an independent samples t-test was run to determine if there were differences in the number of bookings between non-youths and youths. The number of bookings of non-youths ($M = 2.01$, $SD = 2.46$) was higher than that of youths ($M = 1.18$, $SD = 1.67$), a statistically significant difference, $t(19998) = 27.94$, $p = 0.00 < 0.05$.

4.2 ANCOVA

Variable	Sum of Squares	Degrees of Freedom	F-values	Uncorrected p-value
Perceived Race	12.36	2	6.15	0.002
Number of Arrests	13485.94	1	13428.19	0.000

Table 4.2.1 Results for ANCOVA

Table 4.2.1 displays the results from the ANCOVA. From the results of the ANCOVA, we can see that the p-value for Perceived Race is smaller than 0.05. We thus

reject the null hypothesis, which states that the means of the number of bookings for the three perceived races are equal, even after controlling for the number of arrests.

4.3 Logistic Regression

	Coefficient	p-value
Intercept	-0.49	0.00
Sex	0.44	0.00
Number_of_Arrests	0.06	0.00

Table 4.3.1 Results for Logistic Regression

From Table 4.3.1 we can see that none of the p-values is larger than 0.05. Therefore, none of the features are statistically significant. In other words, a person's sex and number of arrests have no statistically significant influence on whether the person is booked at a police station.

Discussion

To fulfill the research objective of finding out the relationship between the number of bookings and demographic factors such as gender, youth status, and race, we first started with the t-tests, which examined gender and youth status. From the results of the first t-test, we were able to answer RQ1 by stating that there is no significant difference in the number of bookings between males and females; males and females are just as likely to be arrested. From the second t-test, we were able to answer RQ2 that non-youths have a significantly higher mean number of bookings than youths, that is, those who are over 17 years of age are more likely to be arrested than those under 17.

We then moved on to examine the effects of race on the number of bookings while controlling for the number of arrests through conducting an ANCOVA. The number of arrests was chosen as the control variable because we wanted to avoid any effect a person's arrest history may have on their number of bookings. One possible effect here is that a person with more arrests makes them more likely to be booked at a police station following an arrest. The results from the ANCOVA provide an answer for RQ3, indicating that there are significant differences in means of the number of bookings for the three perceived races, even after controlling for the number of arrests. This indicates that members of certain racial groups are more likely to be arrested than those in other racial groups. What our ANCOVA was not able to do, however, is to point out what racial groups experience more arrests. To do so, post-hoc tests need to be conducted.

RQ4 was formulated to investigate our second research objective, that is, to find out if a person's number of arrests and sex have any influence on whether that person is booked at a police station following an arrest. To obtain answers for RQ4, we conducted a logistic regression. The results of the logistic regression indicate that a person's sex and number of arrests have no statistically significant influence on whether the person is booked at a police station. Although the response variable is different, this result can be compared with the results from the first t-test, which revealed that males and females are just as likely to be arrested. The person's number of arrests, which was used as the control variable in the ANCOVA, is also shown to have no significant influence on whether a person is booked after an arrest. Thus, we can infer that a person's previous arrest history does not make them more likely to be booked at a police station following an arrest.

This study is limited in a few ways. Firstly, we were only able to base our study on the dataset itself. We were not able to ascertain if the dataset includes all arrests made by the Toronto Police Service in the years 2020 and 2021, thus, our findings are limited to what the dataset includes and cannot be generalized for the City of Toronto for these two years. Secondly, no post-hoc tests were conducted, meaning that we could not accurately point out what racial groups investigated in RQ3 are more likely to be

booked following an arrest. Lastly, for the ANCOVA, we only included three racial groups – white, black, and Latino. There is room for a more holistic investigation of the relationship between racial background and number of bookings.

Conclusion

This study contained two research objectives. The first research objective aimed to find out the effects of demographic factors, such as gender, youth status, race, and age, on a person's number of arrests. The second research objective was to investigate the relationship between a person's number of arrests and sex and whether they get booked at a police station following an arrest. The four research questions and their corresponding hypotheses were tested by t-tests, ANCOVA, and logistic regression. The first three research questions were aimed at fulfilling the first research objective, while the last research question was aimed at fulfilling the second research objective. Among the demographic factors that we chose, we revealed that youth status and race (while controlling for the number of arrests) had significant effects on the number of bookings, whereas sex did not. For our second research objective, we were able to conclude that a person's sex and number of arrests have no statistically significant influence on whether the person is booked at a police station. We recommend that further research can be conducted on datasets on arrests made by the TPS from other years or made by police forces in other Canadian cities to examine whether similar trends exist.

References

- *The usual suspects: police stop and search practices in Canada* by Scot Wortley and Akwasi Owusu-Bempah
(https://www.researchgate.net/publication/238046161_The_Usual_Suspects_Police_Stop_and_Search_Practices_in_Canada)
- *Environmental Criminology and Crime Analysis: Situating the Theory, Analytic Approach and Application* by Richard Wortley and Lorraine Mazerolle
(https://www.researchgate.net/publication/29468939_Environmental_Criminology_and_Crime_Analysis_Situating_the_Theory_Analytic_Approach_and_Application)
- *Annual Statistical Report* by the Toronto Police 2020
<https://data.torontopolice.on.ca/pages/asr-archives>
- *Examining the relationship between police experiences and perceptions of police bias* by Brian C. Renauer, Emma Covelli (2011) .
- *Domestic Violence Prevention and Reduction in British Columbia (2000-2010)* by Katherine R. Rossiter
(https://www.jibc.ca/sites/default/files/research/pdf/Domestic-violence-prevention-reduction_REPORT.pdf)