Project Part 3

Ajay Desai

```
## -- Attaching packages ------------------------------------------------------------

## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts --------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Background Info

Question

Is there is a difference in the age of blacks and whites who were arrested from 1997-2002 in Toronto for the possession of marijuana?

Data Description

This crime data is a sample of a larger dataset that was collected by the Toronto Star newspaper. It shows a sample of the data from 1997-2002 on arrestees for Marijuana possession in Toronto, Canada. The Toronto star newspaper collected this data to argue that blacks are treated more harshly by Toronto's police than are whites. Some of the important columns for this dataset include checks which are the number of times in which the arrestee's name appeared in other police department databases (through previous arrests, previous convictions, parole status, etc.) and shows the criminal history of an arrestee. The colour column specifies the color of the arrestee and the age column indicates the age of the arrestee. (Reference #1)

Data Relevance

This data is relevant to the question stated above, as the data was originally used to find the average number of background checks for blacks and whites in order to determine if blacks are getting treated worse than whites by the police. However, one factor that could impact the number of background checks for arrestees is their age. An individual who is younger and hasn't been involved in crime for that much time isn't likely to have many background checks while an individual who is much older and has been involved in crime for a longer period of time is likely to have more background checks than the younger person. Therefore, finding the average age of the blacks and whites that were arrested is important as the difference in the average age in the two groups could be a factor that influences the data that the Toronto Star newspaper has collected. This data contains an age column which holds the age for each and every arrestee. Therefore, the age column can be used to determine the age of individuals who are blacks and whites and therefore this data can definitely can be used to answer the stated question.

Generalization

The outcome of the statistical test will be generalized to the population of blacks and whites who were arrested in Toronto from 1997-2002 for possession of marijuana.

Hypothesis Test

Chosen Test

A two-sample t-test will be used to answer the stated question. The null hypothesis is that there isn't a difference in the age of blacks and whites that were arrested in Toronto for marijuana possession. The alternative hypothesis is that there is a difference in the age of blacks and whites that were arrested in Toronto for marijuana possession.

Test Selection Explanation

The sample size of white people is 1298 and the sample size of black people is 4099. Since the sample sizes are so large, the shape of the distributions would be normal and therefore the central limit theorem will hold and accordingly we can perform a parametric, quantitative test. Additionally, we don't know the population standard deviation of the sample and we

have to conduct the test on two samples with a sample of black people and a sample of white people, so the test could either be a paired t-test or a two-sample t-test. Because these samples are independent of each other we wouldn't conduct a paired t-test. We can then conclude that a two-sample t-test is the most appropriate test to use for answering the stated question.

Testing

```
arrestData <- data.frame(Arrests)
blackData <- data.frame(filter(arrestData, colour=="Black"))
whiteData <- data.frame(filter(arrestData, colour=="White"))
blackAge <- c(blackData$age)
whiteAge <- c(whiteData$age)
myTest <- function(){
  value <- t.test(blackAge,whiteAge,mu=0,alternative = "two.sided",
                  paired = FALSE)
  value$p.value
}
testing <- myTest()
testing
```

```
## [1] 2.063637e-06
```

Test Conclusions

Overall, there is significant evidence that there is a difference in the age of blacks and whites who were arrested in Toronto from 1997-2002 for possession of marijuana. This test result can be generalized to the entire population of blacks and whites who were arrested in Toronto from 1997-2002 for possession of marijuana.

References

1) https://www.thestar.com/news/gta/knowntopolice/2002/10/19/singled-out.html