Autothief Crime Analytics

Jaden Zhang, Kyunghyun Kim, Youn Jee Kang, Alexander Dobbin, TUT 0201, 201-1

Introduction

- In 2016, there were a total number of 3236 autothief related cases in the entire city of Toronto.
- Most automobiles are expensive, and chasing after thieves takes considerable time and resources for both the victims and the police.
- We would like to investigate in the data containing all the automobile stolen incidents happened in the past few years from the Toronto Police Service.
- We want to know if when and where the incidents are most likely to happen, and whether there is a clear trend or just due to a chance.

Objectives

- Are the auto thieves more likely to steal automobiles in a certain time interval in a day or are they equally likely to steal automobiles in anytime in a day?
- Are the auto thieves more likely to steal automobiles in a particular season or are they equally likely to steal automobiles throughout the whole year?
- Does the population density of an area have an impact on the rate of autothief related cases happened in that area? (in 2016)

Data Summary

- For objective 1, we separated a day into four equal sections: 0am to 6am, 6am to 12pm, 12pm to 6pm, and 6pm to 0am. We then found the proportions of crime for each time interval in the total crimes.
- For objective 2, we separated a year into four seasons: spring, summer, fall, and winter. Similarly, we calculated the proportions of crime occurences for each season in the total crimes.
- For objective 3, we derived the crime rate in an area by fitlering out the data in 2016 first, and then divided the number of crimes happened in a neighbourhood by its population.
- The seed for all the simulations is 101.

occurrencehour_4	num of crime	prop of crime
12am to 6am	3007	0.1654197
12pm to 6pm	4059	0.2232919
6am to 12pm	2788	0.1533722
6pm to 12am	8324	0.4579162

season	num of crime	prop of crime
Fall	5069	0.2788996
Spring	4306	0.2369188
Summer	4835	0.2660248
Winter	3965	0.2181568

5/21

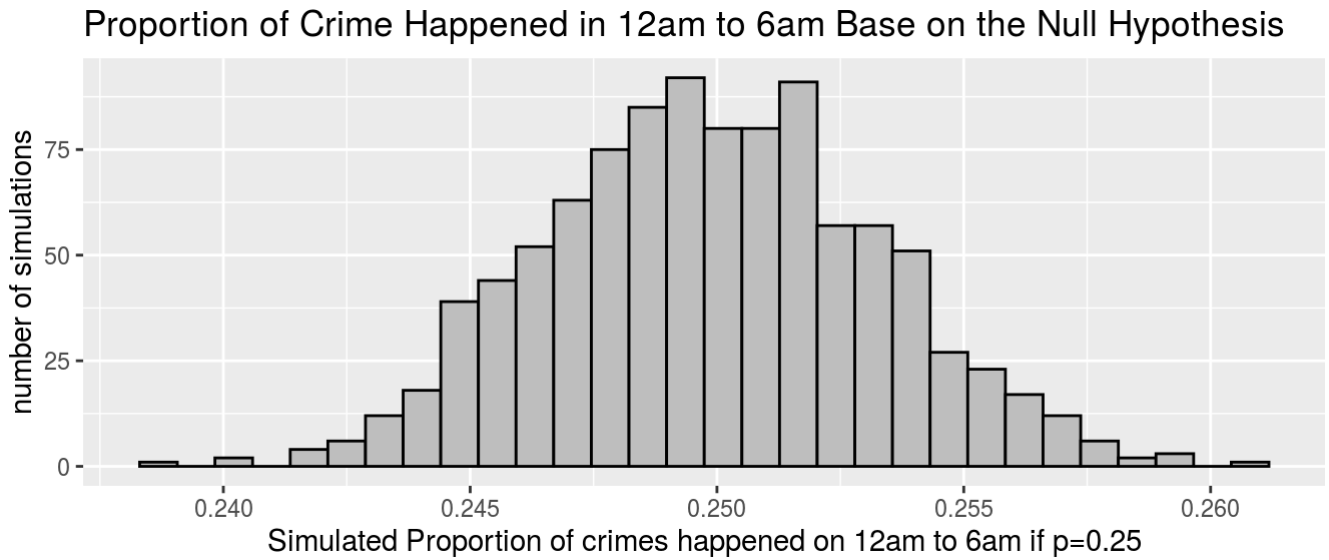
Statistic Method Used for Question 1 and 2

- We set the two hypothesis for each of the two questions.
- The null hypothesis, which is the thieves are equally likely to steal automobiles in anytime in a day, and any season in a year. (this means the probability is 0.25 for both questions' null hypothesis)
- The alternative hypothsis, which is that thieves tend to steal automobiles more often in some time of a day and in some seasons than others.
- We then simulated 1000 simulation under the above hypothesis for each section of the day and each season.
- The p-value of each simulations were calculated.
- The p-value is the probability of observing data that is at least as unusual (or at least as extreme) as the sample data, assuming that the null hypothesis is true.
- With the p-value, we accept or reject the null hypothesis.

Statistic Method Used in Question 3

- Data were divided into two groups, test and training datasets.
- A regression model were fitted to find whether there is an association between the crime rate in a neighbourhood and its population density or not.
- Then we compraed RMSE values for the training and testing data to see if it is overfitting or not.

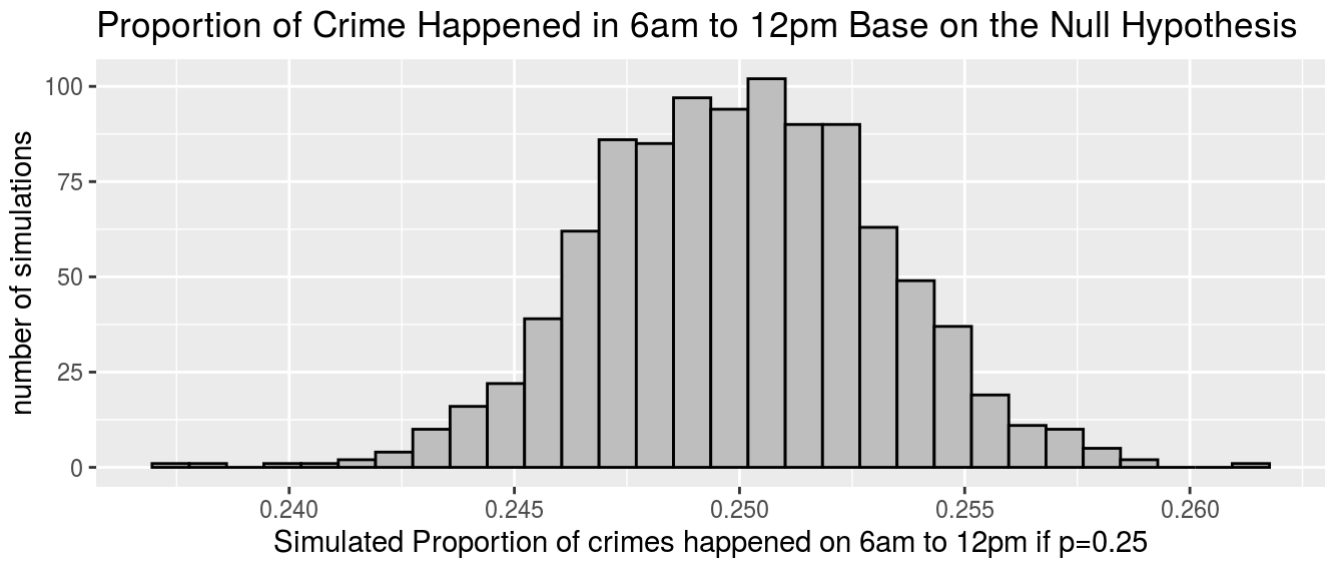
Results for Question 1



p_value

0

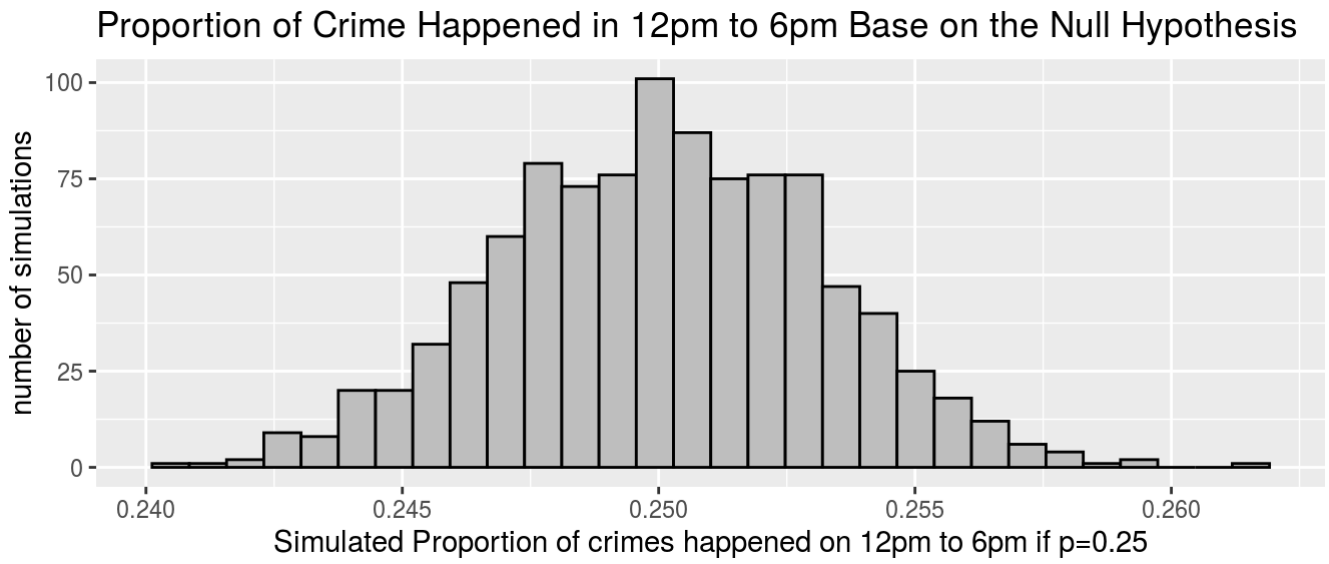
8/21


Proportion of Crime Happened in 6am to 12pm Base on the Null Hypothesis
number of simulations
100
75
50
25
0
0.240
0.245
0.250
0.255
0.260
Simulated Proportion of crimes happened on 6am to 12pm if p=0.25

p_value

0

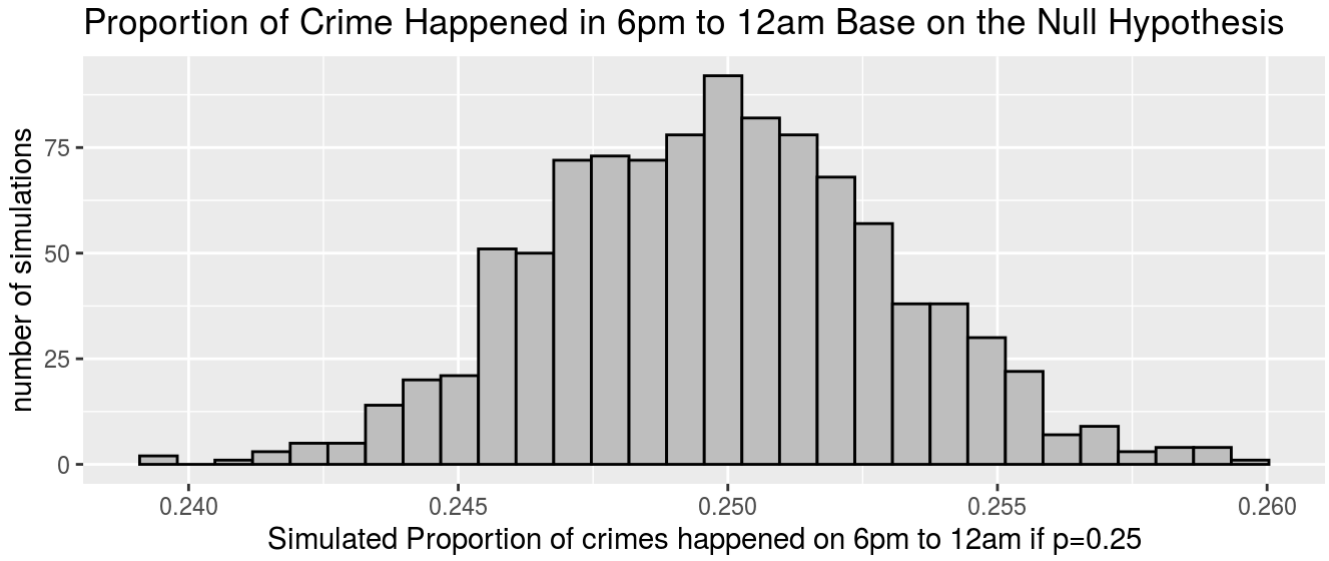
9/21



p_value

0

10/21

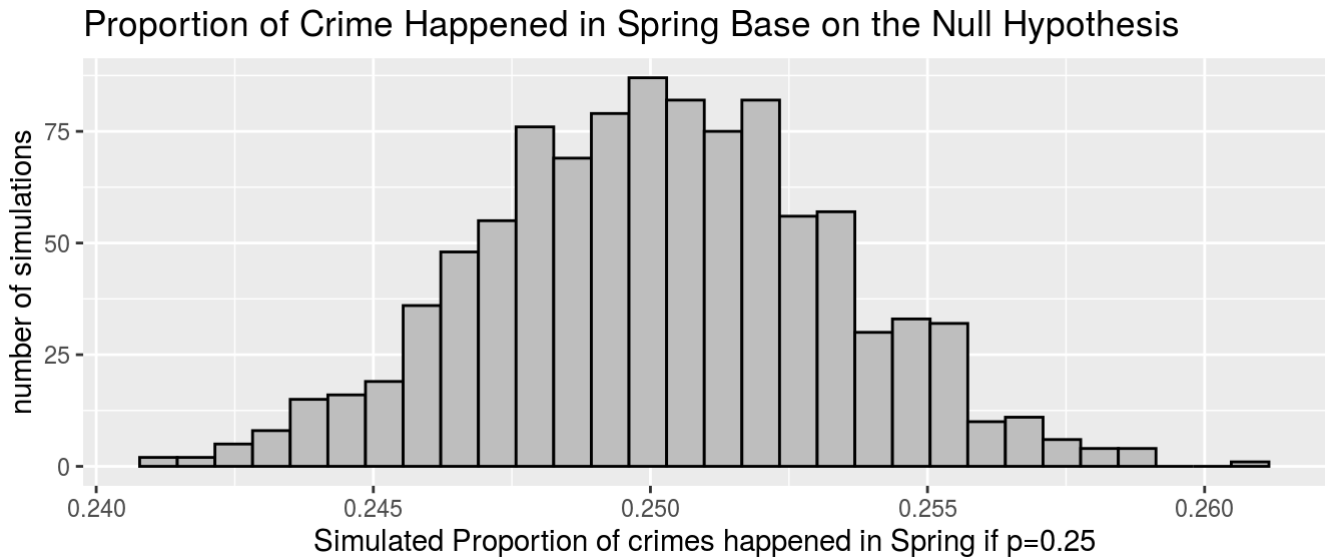


p_value

0

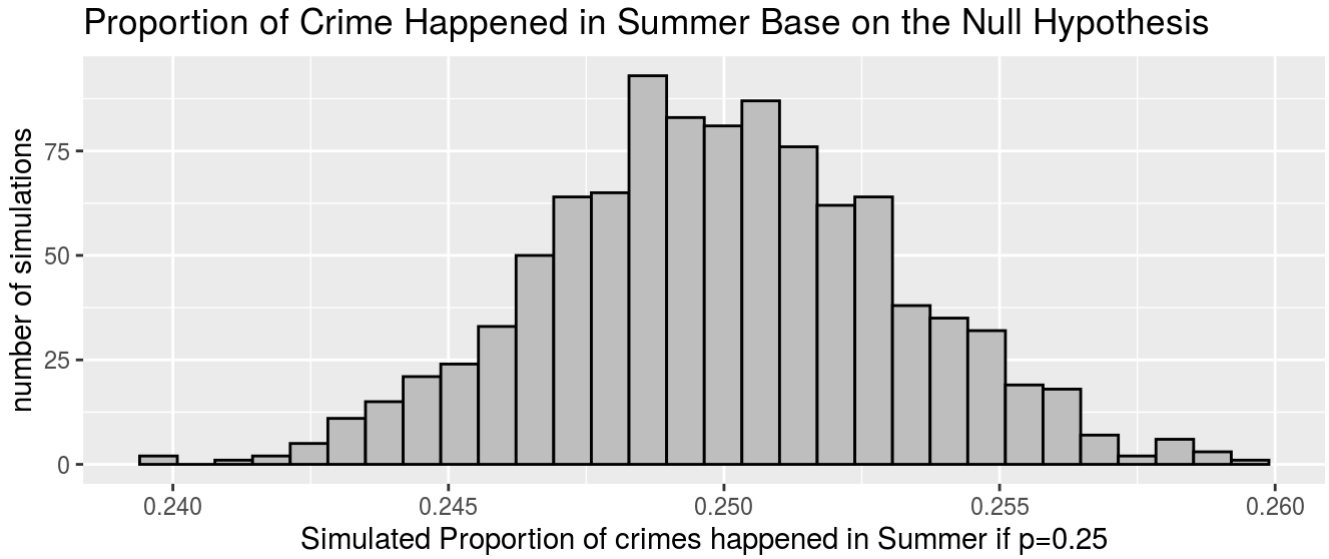
11/21

Result for Question 2



p_value

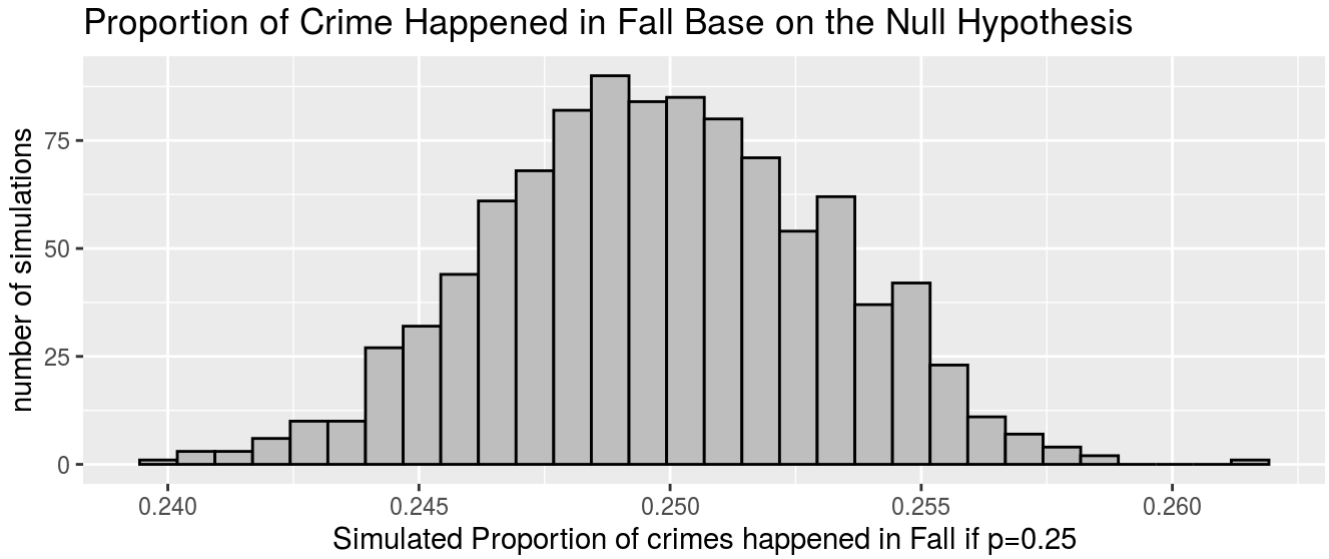
0



p_value

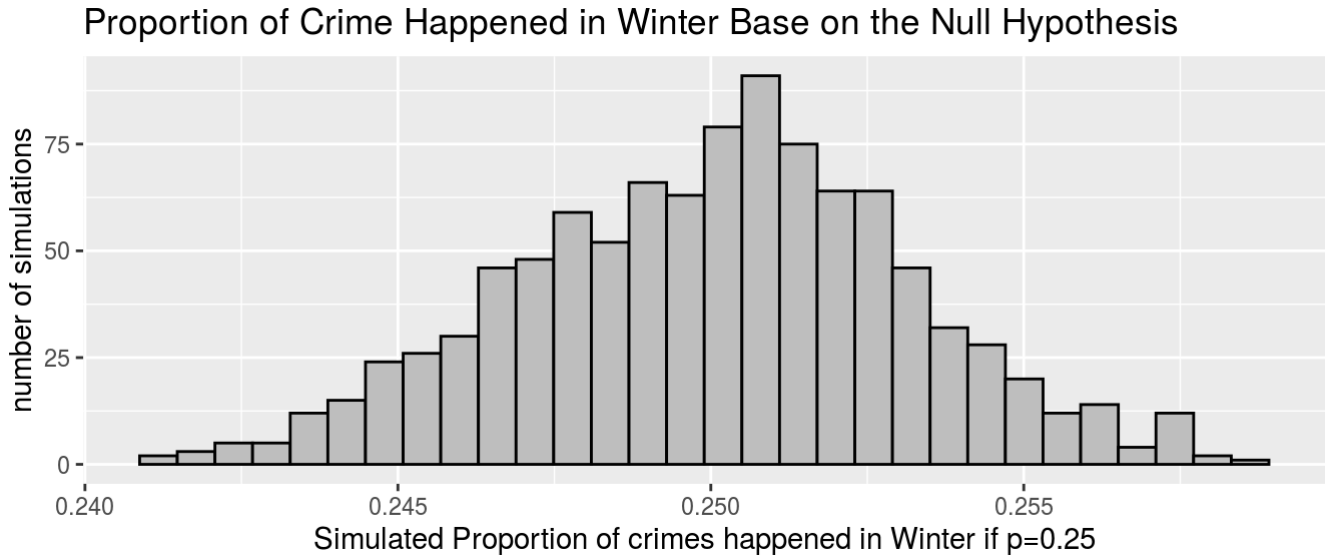
0

13/21


Proportion of Crime Happened in Fall Base on the Null Hypothesis
number of simulations
75
50
25
0
0.240
0.245
0.250
0.255
0.260
Simulated Proportion of crimes happened in Fall if p=0.25

p_value

0

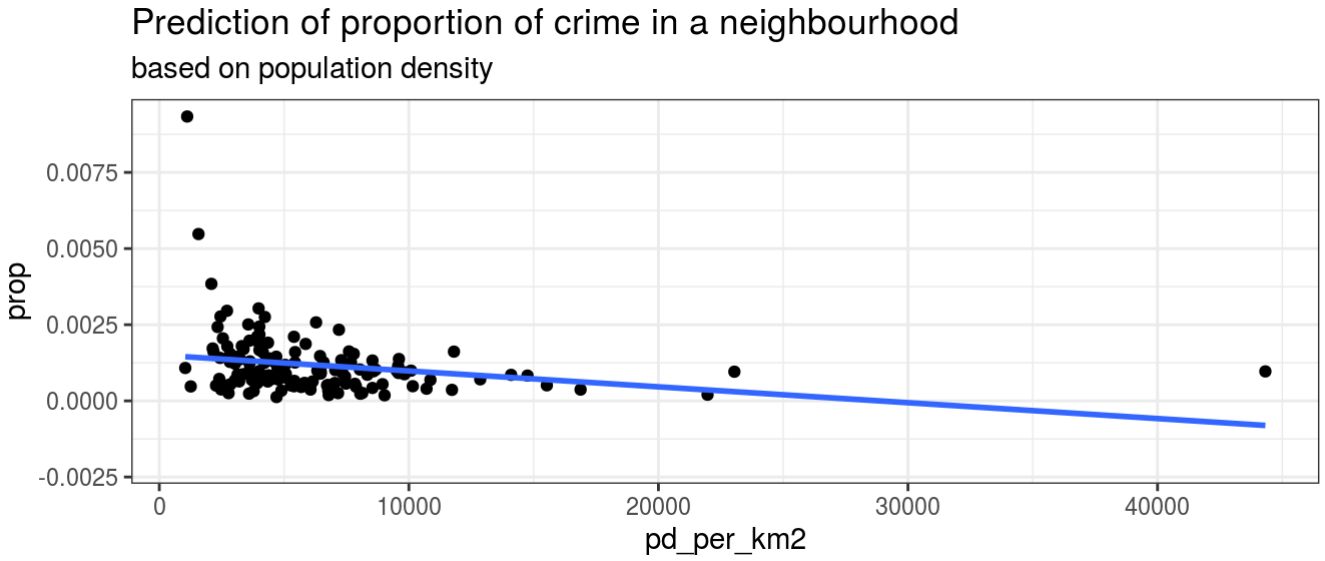

Proportion of Crime Happened in Winter Base on the Null Hypothesis
number of simulations
75
50
25
0
0.240
0.245
0.250
0.255
Simulated Proportion of crimes happened in Winter if p=0.25

p_value

0

Conclusion for Question 1 & 2

- P-values for all 8 of the simulations are 0.
- It means that based on the null hypothesis, we will never get any simulation that is as or more extreme then the observed datas.
- Since the p-values are so small, we have a strong evidance against the null hypothesis, thus we accept the alternative hypothesis.
- Autothieves are more likely to steal an automobile in evening(6pm to 12am) and afternoon(12pm to 6pm).
- Autothieves are more likely to steal an automobile in fall and summer.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0015026	0.0001398	10.747898	0.000000
pd_per_km2	-0.0000001	0.0000000	-2.945471	0.003786

rmse_test	rmse_train	ratio_of_rmses
7818.853	7924.38	1.013497

17/21

Conclusion for Question 3

- Since we obtained small p-values again, we can reject the null hypothesis and conclude that the popullation density does have an effect on the crime rate related to automobile.
- Less populated place tend to have a higher automobile crime rate compare to the more popullated place.
- RMSE for the testing data and the training data is similar, which means it is not ober fitting.

General Conclusion

- Autotheives tend to steal cars on 12pm to 12am, espesially 6pm to 12am. Also they like to steal them in the summer and fall.
- The rate of crime related to automobiles tend to be higher in the less popullated neighborhood, and lower in the neighborhood with high population density.
- Some challanges: understanding the 0 p-values.

Strength and Limitations

Strength: - For all three questions, we have extremely small p-value and a high RMSE ratio, thus we have a high confidence in our result. - From a large data set, it means that our answer will be more accurate.

Limitation: - We do not really know the reasoning behind the conclusions for question 1 and 2. - For question 3, we cannot perdict the crime rate in neighborhoods with more than 25000 population density since we do not have enough data to make the prediction accurate. - The data used for question 3 is from 2016, which might not be as useful for 2020.

Suggestions for the Police

- Schedule more police officers in the afternoon and in the evening.
- Recruit more interim polices or volunteers in the summer and fall, especially summer.
- Set more police in the areas with less popullated areas.