Examination of Physical Lifestyle Change After the COVID-19 Restriction through Survey

Ju-Eun Kim

Part 1

Goal

COVID-19 is a newly discovered infectious disease caused by a coronavirus that causes mild to moderate respiratory illness (WHO, 2021). The disease put many regions into fear and caused casualties. Many areas went through a lockdown or restrictions since early 2020 worldwide, including Canada, to prevent a further spread of the virus. (McCarthy et al., 2021). The restriction was done to keep the 2m social distance and reduce viral spread (McCarthy et al., 2021). However, the restrictions may have caused to have a significantly negative effect on physical activity. The regulations towards personal activities limit the usual lifestyle that people cannot freely move around as before. Therefore, it is essential to find the safest way to return to normal life as before the COVID-19. For that, an active and healthy lifestyle should be maintained by people to protect themselves from COVID-19 and their living health.

The goal of the survey is to examine how the physical lifestyle changed after the restriction. It is significant to consider lifestyle changes to prepare for a better life pattern in a restricted environment. The survey includes a main part of life, which is exercising routine. Also, there are some questions about the residence area, while the level of restrictions is different due to different province policies. For example, Ontario is under a strict lockdown that most in-person activities are minimal (Government of Canada, 2021). However, Alberta does not have as strong restriction and allow travelling without quarantine (Government of Canada, 2021). Therefore, it can lead to how the regulation can impact the difference in lifestyle before and after the COVID-19. The other questions are about self-diagnosis. They can think about how their lifestyles and health level has changed. The COVID-19 pandemic does not allow people to get the exact health level from the clinic. Instead, the brief questions enable the participants to think about themselves and evaluate how they have changed in the survey. The main goal is to collect and analyze how people feel that COVID-19 impacted their health physically and directly in their everyday life.

The change of environment should lead to the new method to have a healthy life. For further action, the survey will help investigate the new way to make guidelines for the people to have a healthy lifestyle in COVID-19 restriction. The result can be used as the feedback of the restriction policy. Suppose the survey indicates significant changes in the lifestyle that concerns the essential exercising to have a healthy life. In that case, the government can develop and inspire people by creating a better policy for exercising and safe outdoor activities. It will help to make everyone endure the COVID-19 pandemic more healthy and efficient way.

Procedure

First, the target population will be identified. The target population will all the University of Toronto students currently living in Canada (Caetano, 2021). However, it is inefficient and impossible to sample all the residents of Canada who are UofT students due to cost and time constriction. Therefore, the frame is created for the frame population. The frame population will be the current UofT students who live in Canada and following my (survey holder's) Instagram account (Caetano, 2021). Some UofT students who live in Canada and following my Instagram may not be available for the survey due to the internet connection or because they may refuse to participate in the study. In this process, sampling bias can occur, such as

self-selection bias that do not response. Removing all of the students who could not answer, the remaining students that attend UofT currently, living in Canda, following my Instagram account and who does the survey will be the sampled population (Caetano, 2021).

The survey will be done online through the Google Form. Google Form is an online survey platform. It allows participants to access the survey by receiving the link from the survey holder. It will be easier for people to access and participate in the study using the online survey method. It takes less time, and it is cost-efficient than a physical paper survey. However, the drawback is that the people who do not have access to the electronic device cannot participate, which may cause bias. For example, the students that do not have wifi connection in their house will not participate.

The sample will be collected by an "Instagram Direct Message." The potential survey recipients will be the UofT students living in Canada and following my social media account. Among them, 40 recipients will be selected randomly through a random generator. The sent survey will open 24 hours to collect a sample of people by providing the survey link. The sample respondents are expected to be at least 20. Probability sampling will be used. It is a sampling method where the respondent is selected involving random selection. There will be an equal chance of choosing an individual in a population. The advantage of the process is that probability sampling can make inferences regarding a total target population (Caetano, 2021). Also, the study can be replicated by other researchers (Caetano, 2021). However, it is time-intensive, costly and is subject to non-participation by potential subjects randomly selected (Caetano, 2021). It also requires a random number generator (Caetano, 2021).

The simple random sampling method will be used. It is the simplest sampling method that samples are collected from the population. The advantage would be that it will have less bias because it is collected from the population directly. However, the disadvantage is that giving every member of the population an equal chance at inclusion in a survey requires having a complete and accurate list of population members (Robinson, 2021). It is not possible when the population becomes too large (Robinson, 2021). Also, the method does not mention specific groups of people. For example, people taking summer programs may have different lifestyle patterns during the pandemic from those who do not. A simple random sample of the entire population may only include a few of the group, that it will not have analysis for the specific groups (Robinson, 2021).

Showcase of the survey

URL into survey: https://forms.gle/hoB7a2V87eHFGmmN9

3 key questions   Question 1: Is your city/town defined as a COVID-19 hotspot?

- Yes
- No
- Maybe

The question is a categorical type of question. It asks whether the city that the survey participant is currently living in is defined as a COVID-19 hotspot. For example, Ontario declared few postal codes as hotspots (Katawazi, 2021). The hotspots are the few neighbourhoods at high risk of COVID-19 transmission (Katawazi, 2021).

It contributes to the goal by comparing the people with strong and lose lockdown situations. If the participant lives in the hotspot and answers "Yes" to this question, they are more likely to have higher restrictions. It will decrease the outdoor activities more than the other region. The people who said they live in the hotspot will need further analysis if the restricted lockdown has caused more lifestyle differences than the people in the other non-hotspot area. Also, people who responded "No" are expected to have a minor difference in their lifestyle before and after the COVID-19. The further analysis will focus on the respondents who said "Yes" to the question if they have a more significant difference in their lifestyle.

The advantage of the question is that it quickly identifies the rate of people who live in the hotspot and non-hotspot area. It makes the analysis more accessible by sub-grouping the sample. The disadvantage of the question is that it does not give the details about the COVID-19 scale. Even though some neighbourhoods are identified as the hotspot, they will not have precisely the same number of patients. Some may have a

slightly lower rate of transmission of the virus. The number of patients is changing day to day, that the hotspot may change after a while. The government of the province may define the hotspot differently. They can also put different restrictions. Also, the question includes "Maybe" as an option. It can make the analysis result biased if many people do not know if they live in the hotspot area.

Question 2: How long did you exercise on average in a week *BEFORE/AFTER* the COVID-19 lockdown? (in minutes)

- short answer in numerical

The second question is a short answer type in numerical. In the survey, the question is divided into 2, before the COVID-19 lockdown and after. The question each asks the average exercise amount in the unit of a minute. According to the Department of Health and Human Services, adults are recommended to exercise at least 150 minutes per week (Laskowski, 2019). However, due to the restriction, it is likely that people cannot exercise as much as before. There may be various factors, such as lack of workout space, lack of motivation, or not feeling safe.

This question contributes to the goal of the survey by comparing the difference of exercising before and after the COVID-19. If the lockdown caused people's exercise amount to decrease, the government should look into the improvements of the environment to maintain the people's essential health through exercising. If the exercise amount has increased, the result can be used after the COVID-19 lockdown to make more people active even after the lockdown. It can be used to improve the overall workout rate in the future.

The advantage of the question is that the people can answer the question in all ranges. If it were a categorical question, it would be hard to identify the individual's exact numerical average of workout time. The question clearly identifies the unit as a minute. Therefore, there will be less confusion of units when analyzing the collected data. The disadvantage of the question is that the participants can type the value freely. However, this may cause a complexion for the analysis. For example, if the participants answered "120 minutes," the "minutes" part needs to be deleted in the cleaning process. The short answer questions may make the cleaning process more complex.

Question 3: On a scale of 1 to 10, how would you score your personal health *BEFORE/AFTER* the COVID-19 lockdown? (i.e., if you think you were very healthy, choose 10 and if you think you were not healthy at all, select 1)

- Least Healthy 1 2 3 4 5 6 7 8 9 10 Most Healthy

The last question for the showcase is the rating scale question, which is a categorical-type question. The question asks about the health level to answer on the scale. In the survey, there are two separate questions. The first question asks about the health level before the COVID-19 lockdown, and the second asks the after lockdown. After the lockdown, people are experiencing an increased risk of obesity, diabetes, and cardiovascular disease (Dunton, 2020). The disorders are caused due to the change of the lifestyle and lack of physical activity.

The question contributes to the goal of the survey by comparing the health scale before and after the COVID-19 lockdown. The main questions in the survey were based on the current physical activity. However, this question takes one step further. There might be other external factors, such as the development of diseases, due to lack of exercise. Therefore, the question indirectly asks to include all the other elements. Without stating the other specific aspect, asking the overall health helps to identify if people maintain a similar health level. Comparing the health level change, if the difference between the after and before COVID-19 is negative, implies that the health department should prepare new guidelines to help find the causation of the drop of health and improve health level. If it was positive, the survey result could be kept to use after the COVID-19 lockdown for future health maintenance.

The advantage of the question is that participants can rate the health level on a scale. It makes the analysis more straightforward, while all the results can be dropped into a category. The con is that having the result of only one question, such as having the impact of only before COVID-19, may not be helpful. Both questions should be considered together. People rate themselves by self-diagnosing in the question. The survey has limitations to make the scale with detailed health level. The question cannot have an objective measure to

scale the health. Therefore, the survey is conducted to show the difference in health level before and after the lockdown. For instance, if the health level of an individual went down, it means that the person has worse health. Also, the con is that people may be subjective. The change of difference in health does not have the guideline to explain that each individual may think on a different scale, even on a similar health level. Therefore, the question may not give an objective result. The positivity and negativity of the health change would be a more critical focus to consider. People could also lie for the personal health level when answering the question, which may cause bias.

Bibliography

1. Caetano, S. (2021) *Probability Nonprobability Sampling.* lecture in STA304, Surveys, Sampling and Observational Data, University of Toronto. (Last Accessed: May 11, 2021)

2. Caetano, S. (2021) *Sampling Definitions.* lecture in STA304, Surveys, Sampling and Observational Data, University of Toronto. (Last Accessed: May 11, 2021)

3. Caetano, S. (2021) *Sampling Techniques.* lecture in STA304, Surveys, Sampling and Observational Data, University of Toronto. (Last Accessed: May 11, 2021)

4. Dunton, G.F., Do, B. & Wang, S.D. (2020) *Early effects of the COVID-19 pandemic on physical activity and sedentary behavior in children living in the U.S..* BMC Public Health 20, 1351. https://doi.org/10.1186/s12889-020-09429-3. (Last Accessed: May 11, 2021)

5. Government of Canada. (2021). *Provincial and territorial restrictions.* https://travel.gc.ca/travel-covid/travel-restrictions/provinces. (Last Accessed: May 10, 2021)

6. Katawazi, M. (2021) *Full list of Ontario neighbourhoods where the COVID-19 vaccine will be available to those 18+.* CTV News. https://toronto.ctvnews.ca/full-list-of-ontario-neighbourhoods-where-the-covid-19-vaccine-will-be-available-to-those-18-1.5379755. (Last Accessed: May 10, 2021)

7. Laskowski, E (2019, April 27). *How much should the average adult exercise every day?.* Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/exercise/faq-20057916. (Last Accessed: May 10, 2021)

8. McCarthy, H., Potts, H., & Fisher, A. (2021). *Physical Activity Behavior Before, During, and After COVID-19 Restrictions: Longitudinal Smartphone-Tracking Study of Adults in the United Kingdom.* Journal of medical Internet research, 23(2), e23701. https://doi.org/10.2196/23701. (Last Accessed: May 10, 2021)

9. Robinson, N. (2021, May 7) *Advantages & Disadvantages of Simple Random Sampling.* sciencing.com. https://sciencing.com/advantages-disadvantages-of-simple-random-sampling-12750376.html. (Last Accessed: May 11, 2021)

10. WHO. (2021) *Coronavirus.* World Health Organization. https://www.who.int/health-topics/coronavirus#tab=tab_1. (Last Accessed: May 10, 2021)

Part 2

Data

The data was collected through the survey. It was about the relationship between the COVID-19 lockdown and the overall physical activity of the individual. The survey was done online on May 8th, 2021, from 0:00 to 23:59. The survey was opened for 24 hours and stopped receiving responses after 24 hours. The data is collected using the platform Google Form. By counting, 82 current UofT students were in Canada and were following my Instagram account. After putting the names in the google sheet, I assigned numbers 1 to 82 to each people and generated 40 random numbers using r. The students who had the generated number received the direct message through Instagram with the survey link and a brief explanation on May 8th, 2021, 0:00. Out of 40, 39 students answered. The 39 respondents are the sample population that are the current UofT students in Canada, following my Instagram account, and participating in the survey. The data were collected and automatically recorded on Google Form. The result was downloaded in the excel sheet.

There were obstacles that when identifying the population. It was hard to determine if the students were living in Canada currently after the lockdown. I had to ask if they were present in Canada before conducting the survey. Also, even it was a small population, recording all individuals with a hand manually through google sheet took a long time and was inefficient.

The data imported results from the survey, "Impact of COVID-19 Lockdown Towards the Individual's Lifestyle in Canada". There were 39 responses. The data collected on the spreadsheet was converted to CSV. File for further analysis in r. There are 13 questions in the survey: there are 14 columns in the data (1 column shows the timestamp, which was automatically created when converting the result into the spreadsheet). The columns of the data represent each question asked, and the rows represent each participant. Each cell indicates the response towards the assigned question. [2][3]

The unnecessary columns will be deleted. The variable `Timestamp` is an unnecessary variable that is created automatically when converting the collected result to the spreadsheet.

The question was directly converted to the columns' names each. However, the whole question as a column's name would be complicated to understand the data. Therefore, the name of the columns should be changed. The name of columns will be transformed into a simple word that includes the keyword. For example, if the question was "Which region/province do you live in?", it can be simplified to `province`.

A new dataset will be made in the cleaning process. One of the questions from the survey asks how they think their activity level changed after the COVID-19 restriction. The people who answered "somewhat different" or "very different" answered the last question. However, the individuals who did not choose one of those options did not answer the question. It created blanks (N/A) for the analysis. The new dataset will remove the individuals who did not answer the question. It may be helpful for further research because the individuals left will think that their activity level has changed significantly. The goal of the survey was to find how COVID-19 lockdown impacted the overall health-related to physical activity. The focus on the participants who experienced more significant differences before and after the limitation would be essential for further analysis. [4]

Key Variables

The variables `hotspot` and `vaccine` are essential as they determine the limitation towards the people's activity. The area that is defined as hotspots has more regulated rules that restrict outdoor physical activity. The places that are not hotspots will have more freedom to move or use the facilities for exercise. Then, the people in the non-hotspot area have more access to the exercising site. The variable vaccine has a similar effect. People who got vaccination are less likely to get COVID-19. Therefore, the limitations will be lower than the non-vaccinated people. Also, psychologically, people will tend to feel safe for physical activity before the vaccination. Two variables may indicate people's behaviour.

The variables `bf_out` and `af_out`, and `bf_exercise` and `af_exercise` are 2 sets of similar variables. The variables `bf_out` and `af_out` measure the numerical value using multiple choice of how many days people have gone out. Going out of the house is considered a physical activity. Getting prepared to go out, walking

a short distance, or going to the grocery can be all physical activities. The participant students answered how many days in a week on average they go out. If `af_out` is less than `bf_out`, it indicates that people are going out less. It represents the decrease in physical activity. The variables `bf_exercise` and `af_exercise` can be analyzed in a similar method. The questions collect the time of the average exercising amount in the unit of the minute. It similarly measures the activity level in numerical values but focused on exercising. It can both refer to the outdoor and indoor exercising activity. The range is not limited in time while asking for the numerical value in a short answer. Therefore, these variables can show how the mean exercise time has changed after the COVID-19 restriction.

The variable `workout` is essential data. The result is if the participants are willing to exercise more if more facilities can be used safely during the COVID-19 pandemic. If people answered `yes`, it means that people are not exercising as much as before due to the lack of space even they would like to have physical activities. Therefore, if many people answer `yes`, the government must find a better way to provide people space to work out.

The variable `cate_activitylevel` represents how people think how different their lifestyle is after the COVID-19 pandemic. If people think it is different, it indicates that they are significantly restricted from many factors. A new variable, `scale_back` was introduced for people who answered "different" for the `cate_activitylevel`. It asks if they think that their activity level will be back after the pandemic. It can be answered on a scale from 1 to 5. The closer to 5, the likely it is that people will go back to the normal lifestyle. However, if it is close to 1, it would concern while the decreased physical activity level will not go back to normal. Even though the lifestyle is very different from the normal, if they are willing to go back after the pandemic, the overall health can be maintained. However, if they have the habit of not exercising, it may cause severe diseases like obesity. Therefore, it can be used for further analysis and regulations to develop future returns.

Numerical Summary

The numerical summaries for few variables are found below. They will be helpful for further analysis. [4]

It is the first numerical result. The number of days people going out before and after the COVID-19 restriction on average per week are analyzed in the table.

```
## # A tibble: 1 x 5
##     min   max  mean   med     n
##   <dbl> <dbl> <dbl> <dbl> <int>
## 1     4     7  5.90     6    39
```

```
## # A tibble: 1 x 5
##     min   max  mean   med     n
##   <dbl> <dbl> <dbl> <dbl> <int>
## 1     0     5  2.15     2    39
```

The first numerical summaries are done for the variables `bf_out` and `af_out`. The sample data indicates that the average minimum time of going out before the COVID-19 pandemic is 4, and the maximum is 7. However, after the pandemic, the average of going out is 5, and the maximum is 5. Also, the mean decreased from 5.897 to 2.153, which became significantly low. Therefore, it indicates that physical activity decreased significantly.

The second numerical summary tables represent the exercising timing before and after the COVID-19 restriction on average per week.

```
## # A tibble: 1 x 5
##     min   max  mean   med     n
##   <dbl> <dbl> <dbl> <dbl> <int>
## 1     0   700  247.   200    39
```

```
## # A tibble: 1 x 5
##     min   max  mean   med     n
```

```
##    <dbl> <dbl> <dbl> <dbl> <int>
## 1     0   300  71.3    50    39
```

Similar to the variables `bf_out` and `af_out`, the results of `bf_exercise` and `af_exercise` are significantly different. The minimum values are the same, which are zero. However, the maximum amount of exercising before COVID-19 was 700, while after the COVID-19 became 300. Also, the mean dropped from 247.1795 to 71.282 after the regulation has started. Therefore, people may experience lots of differences in their physical activity.

Visualization

The first two graphs represent the time of exercising in the boxplot. [4]
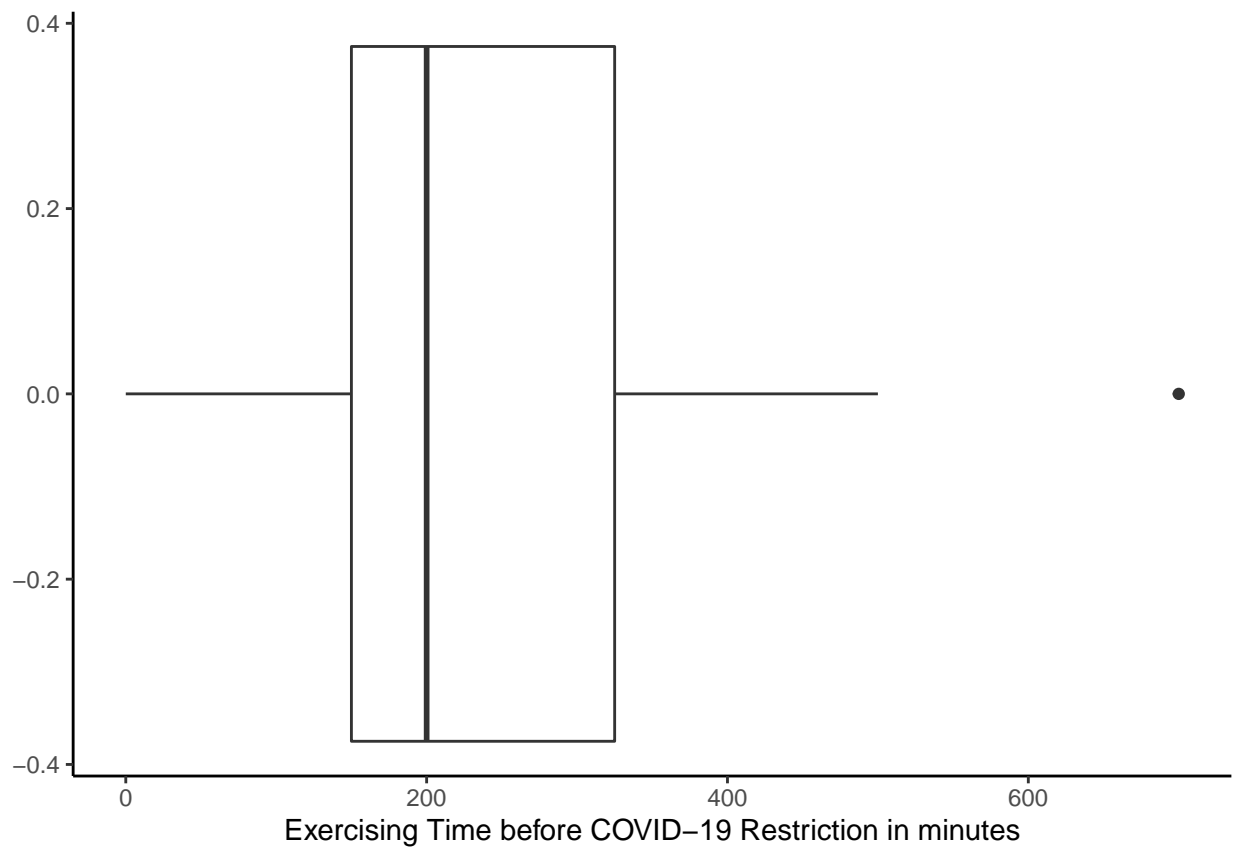
The boxplot above indicates the amount of exercise in the unit of a minute. The median is 200 minutes, and one outlier exists, which is greater than 600 minutes.
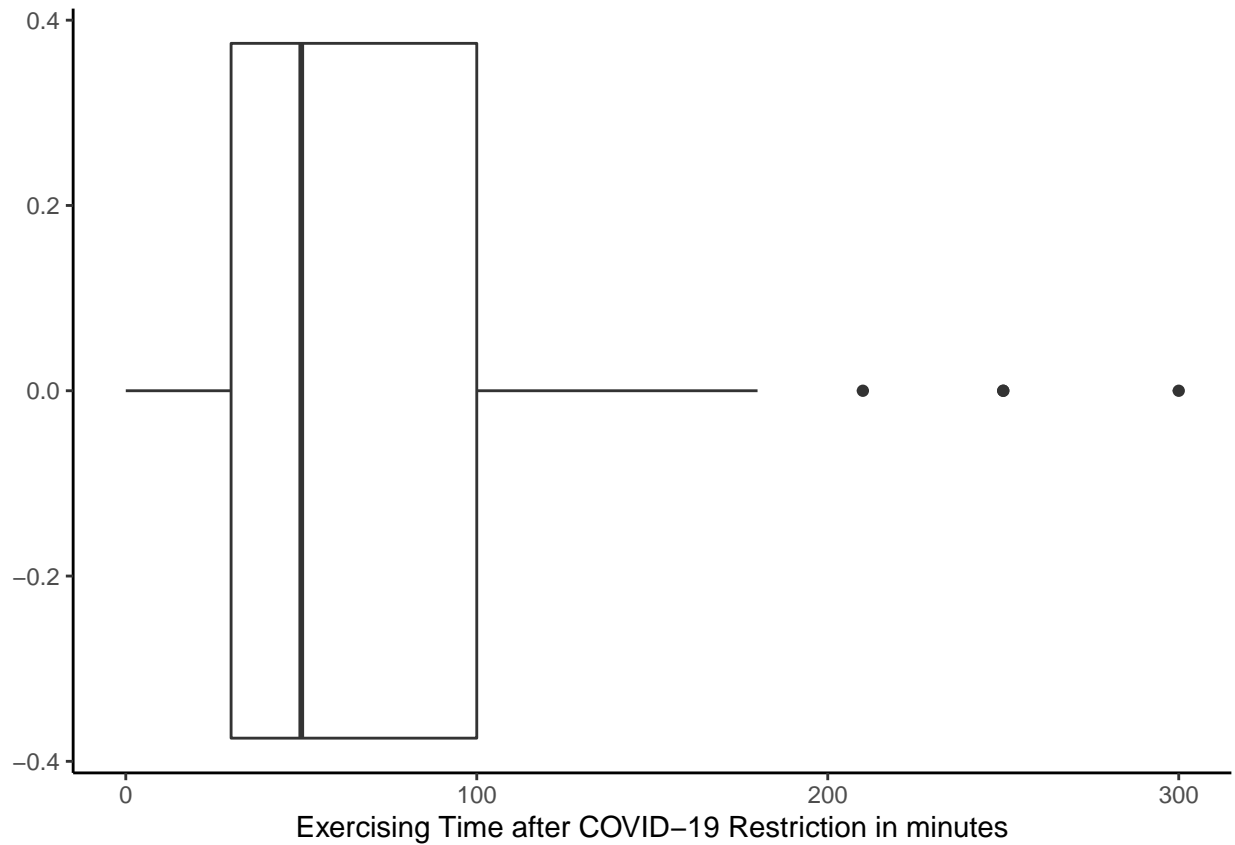
The boxplot above indicates the exercising time after the COVID-19 restriction has become active. The median line shifted left towards zero minutes compared to the before restriction. Also, the maximum exercising amount has decreased. Overall, comparing the two boxplots indicates that fewer people are likely to exercise during the COVID-19 pandemic.

The following graph is the bar graph that indicates how the lifestyle change after the COVID-19 restriction. It is represented with categorical values. One category indicates one bar. [4]

Is Lifestyle After the COVID−19 restriction different from before?
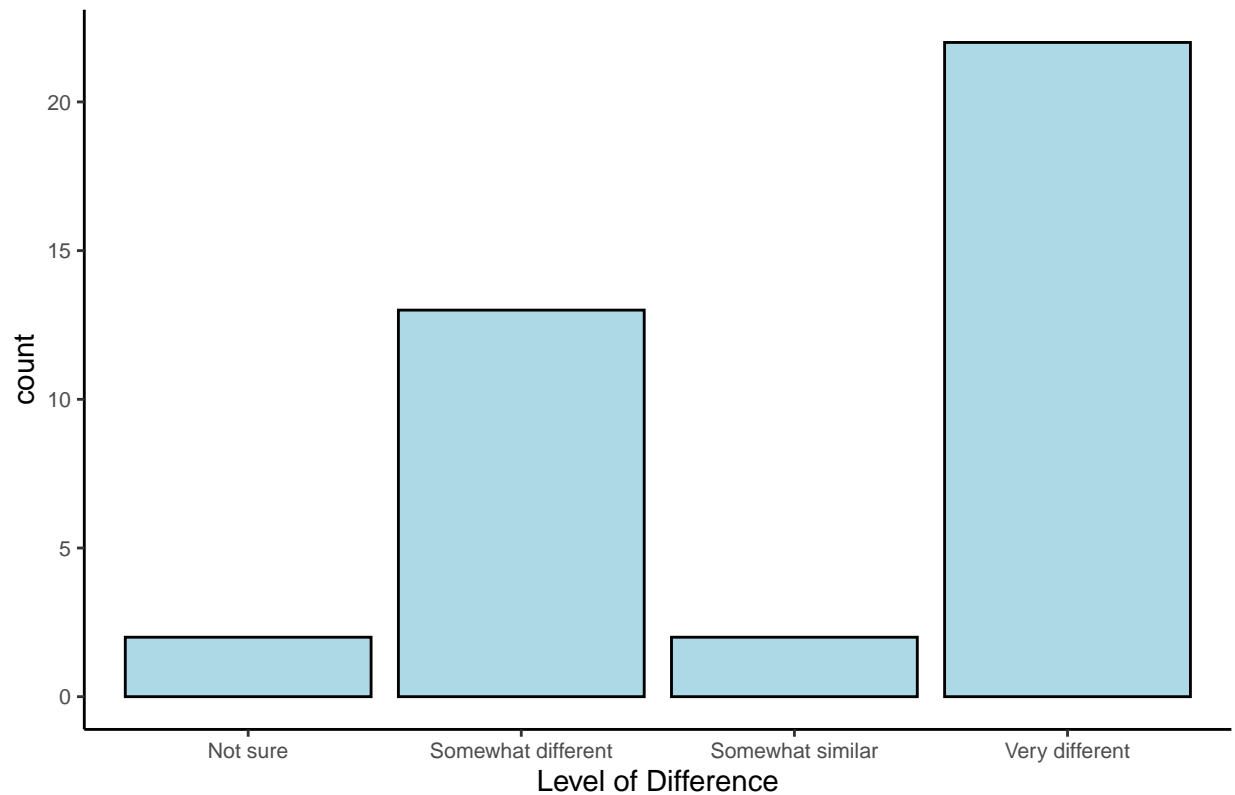
The majority of people think that their lifestyle has changed by answering either "Somewhat different" or "Very different." The participants recognize the differences in different ways. The assumption is that physical activity plays a role in the participant's decision.

The last plot represents the survey result if the physical activity level would change back to normal on a scale of 1 to 5 in the bar graph. Similar to the previous graph, it is a categorical variable. [4]

Physical activity Level will change back to Normal on a scale of 1 to 5
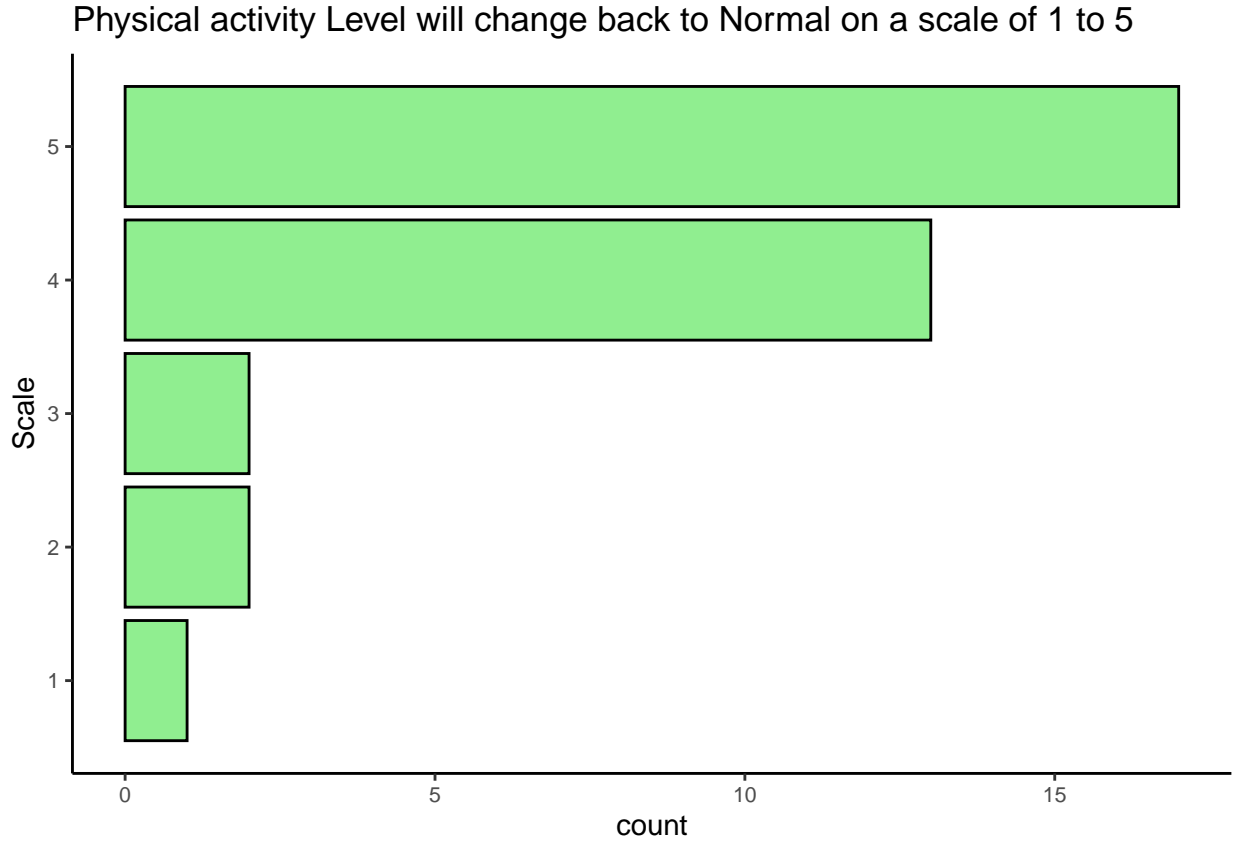
The majority of people think that they are most likely to get back to the average activity level. Only a few people answered that they would not believe they will have a normal lifestyle. The concern is that the longer-lasting COVID-19 pandemic may make people be used to the low activity level. The result of this question may lower the concern while the majority of people desire and think that they can exercise like before.

Methods

In this section, two methods will be done, the hypothesis test and the confidence interval.

The hypothesis test will focus on whether the region of the student's residence is in the hotspot. It will test if the hotspot area has a lower exercising time compared to the non-hotspot site. It will show if the hotspot regulation causes differences in people's exercising time.

The main variables to explore are the `bf_exercise` and `af_exercise` for the confidence interval. They are numerical values where the participants recorded the average amount of exercise in a minute unit. The mean range will be compared to see if the overall mean after the COVID-19 restriction has started.

Hypothesis Test

In this section, the hypothesis test of the mean will be done. The mean of the exercising time in hotspot and non-hotspot will be compared. The assumption is data collected are independently and identically distributed (Dekking, F. M., et al., 2005).

First, the null and alternative hypotheses are defined. The null hypothesis is that the mean of time of exercising in a hotspot with high COVID-19 regulation and non-hotspot with low COVID-19 restriction will be the same.

$$H_0 : \mu_h = \mu_n$$

The alternative hypothesis is that the mean of hotspot's exercising time and non-hotspot's exercising time will not be equal.

$$H_A : \mu_h \neq \mu_n$$

Then, the dataset,`af_exercise` in `health_dt`, will be used to find a test statistic and p-value, where the p-value is the probability of observing the test statistic (Dekking et al., 2005). While the test will be done using the two-tailed test, the formula for test statistic would be $\mu_n - \mu_h$

The decision for the hypothesis can be made using the calculated p-value. If the p-value is less than 0.05, the null hypothesis would be rejected because there is evidence against H_0. However, if the p-value is larger than 0.05, there is no evidence against H_0. Therefore, the null hypothesis would not be rejected.

*For note: The decision using the p-value can be an error. If the null hypothesis is rejected when H_0 is true, it will be a type 1 error (Dekking et al., 2005). However, when the null hypothesis is not rejected when H_A is true, it will be a type 2 error (Dekking et al., 2005).

Confidence Interval

The confidence interval will be calculated using the bootstrap method (Dekking et al., 2005). The sample of 39 responses may not be enough for the confidence level analysis. Therefore, by using the bootstrap method, it will be able to generate more samples by replacing the founded data. It will generate 500 bootstrap samples of size n. Then, the statistic will be calculated for each bootstrap sample.

Similar to the hypothesis test assumption, the data collected are assumed to be from an independent and identically distributed population (Dekking et al., 2005).

The empirical bootstrapping method is chosen because the dataset is given randomly in this case (Dekking et al., 2005). It is a non-parametric bootstrapping method. The bootstrap will derive the 95% confidence interval for the mean exercise time of participants.

Results

Hypothesis Test of the Mean

The mean of the exercise time, hotspot and non-hotspot regions will be compared using the hypothesis test.

The null hypothesis is the mean of exercising time in hotspot and non-hotspot are the same.

$$H_0 : \mu_h = \mu_n$$

The alternative hypothesis is that they do not equal to each other.

$$H_A : \mu_h \neq \mu_n$$

.

First, the number of observations is found. In total, 39 students participated. Then, in the total observation, the number of participants living in the hotspot and non-hotspot areas is found. Nineteen participants were living in the hotspot area, and 17 participants were not living in the hotspot.

Next, the test statistics are measured. The test statistic is the difference between the sample mean of exercising time before COVID-19 pandemic and the sample mean of time after COVID-19 pandemic. While $H_0 : \mu_h = \mu_n$, it can be also written as $\mu_h - \mu_n = 0$. The test statistics is $\bar{x}_h - \bar{x}_n$.

After the calculation, the test statistic was found to be -7.9567.

Then, the simulation will be done to test the test statistics under H_0. The 500 repetitions will be done to examine the test statistic assuming that the null hypothesis was true.

After the simulation with the actual data, the result needs to be analyzed to see if the null hypothesis was correct. It can be done visually using the histogram graph.
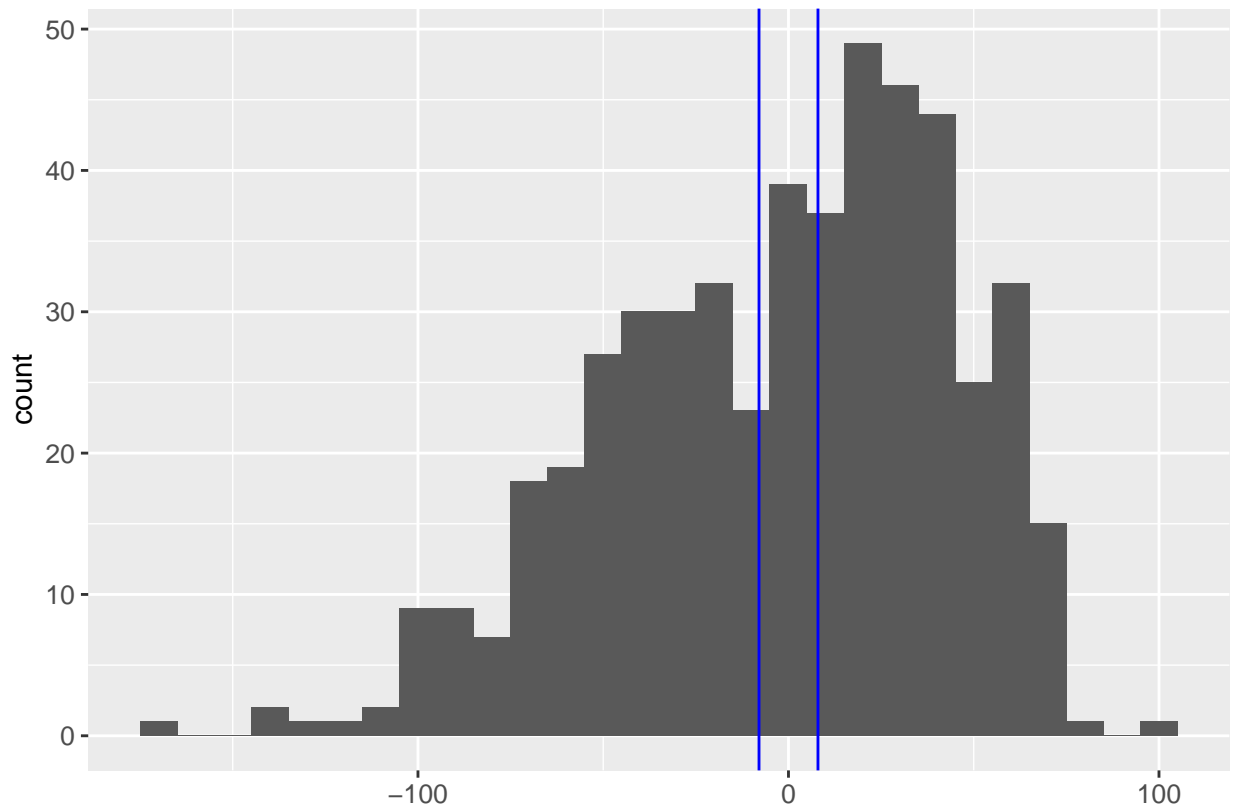
Simulated differences between before and after COVID−19 restriction, assuming no differenc

```
## # A tibble: 1 x 1
##   p_value
##     <dbl>
## 1   0.882
```

The p-value is calculated and gave the result of 0.882. It means that there is no strong evidence against the null hypothesis.

The result implies there is almost no difference between hotspot and non-hotspot exercising time after the COVID-19 restriction. The evidence is not strong to reject the null hypothesis. Even though there are fewer restrictions in the non-hotspot area, people may feel unsafe to exercise in the gym or outdoor with many people. Therefore, it seems the exercising time are similar.

If this is not true, the type 2 error occurred by not rejecting the null hypothesis when the null hypothesis is false. While the testing was done with the simulated data with a small sample, the null hypothesis may not be true.

Confidence Interval

If the student's lifestyle did not change, the exercising time per week on average should not change before and after the restriction due to COVID-19. The mean of the exercising time will be compared by generating a 95% confidence interval of the mean.

The bootstrap data will be generated first using 500 repetitions. It means that the 500 data sets will be generated from the sample.
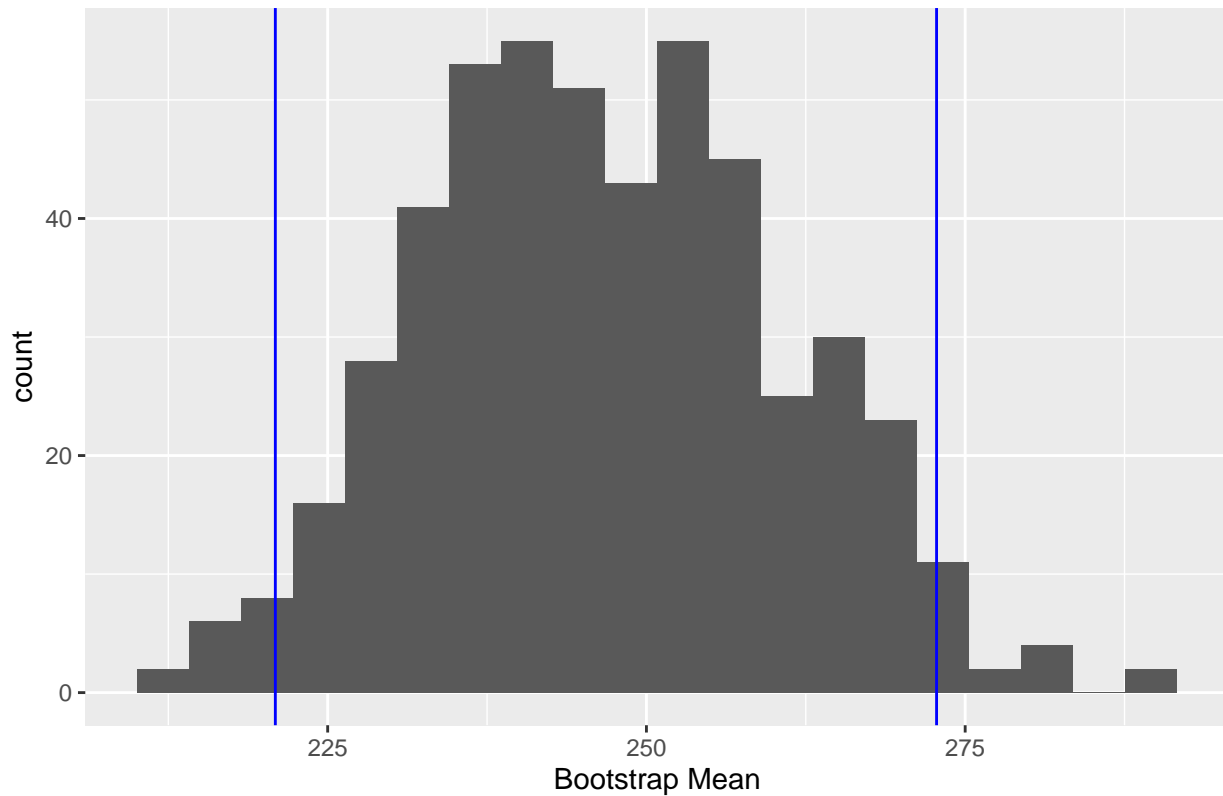
A confidence level of 95% is used. It indicates when sampling was done many times, 95% of the intervals contain the population mean (Dekking et al., 2005). A 95% confidence level for a population parameter was

calculated from the sample data. It gives a range of plausible values for the true parameter based on the limited information provided by the sample from the population.

```
##       2.5%     97.5%
## 220.8950  272.7575
```

The confidence level of 95% is between 220.8950 minutes and 272.7575 minutes. It indicates that we are 95% confident that the mean of the exercising time before the COVID-19 pandemic was between 220.8950 minutes and 272.7575 minutes.

Bootstrap distribution of Exercising Time Before COVID−19 Restriction


The exact process will be done for the exercising time after the COVID-19 restriction as well.

Again, the 95% confidence interval will be used.

```
##      2.5%    97.5%
## 57.5475  84.9100
```

The confidence level of 95% is between 54.5475 minutes and 84.9100 minutes. It indicates that we are 95% confident that the mean of the exercising time after the COVID-19 restriction is between 54.5475 minutes and 84.9100 minutes. In the graph, the blue line represents the confidence interval. [4]

Bootstrap distribution of Exercising Time After COVID−19 Restriction
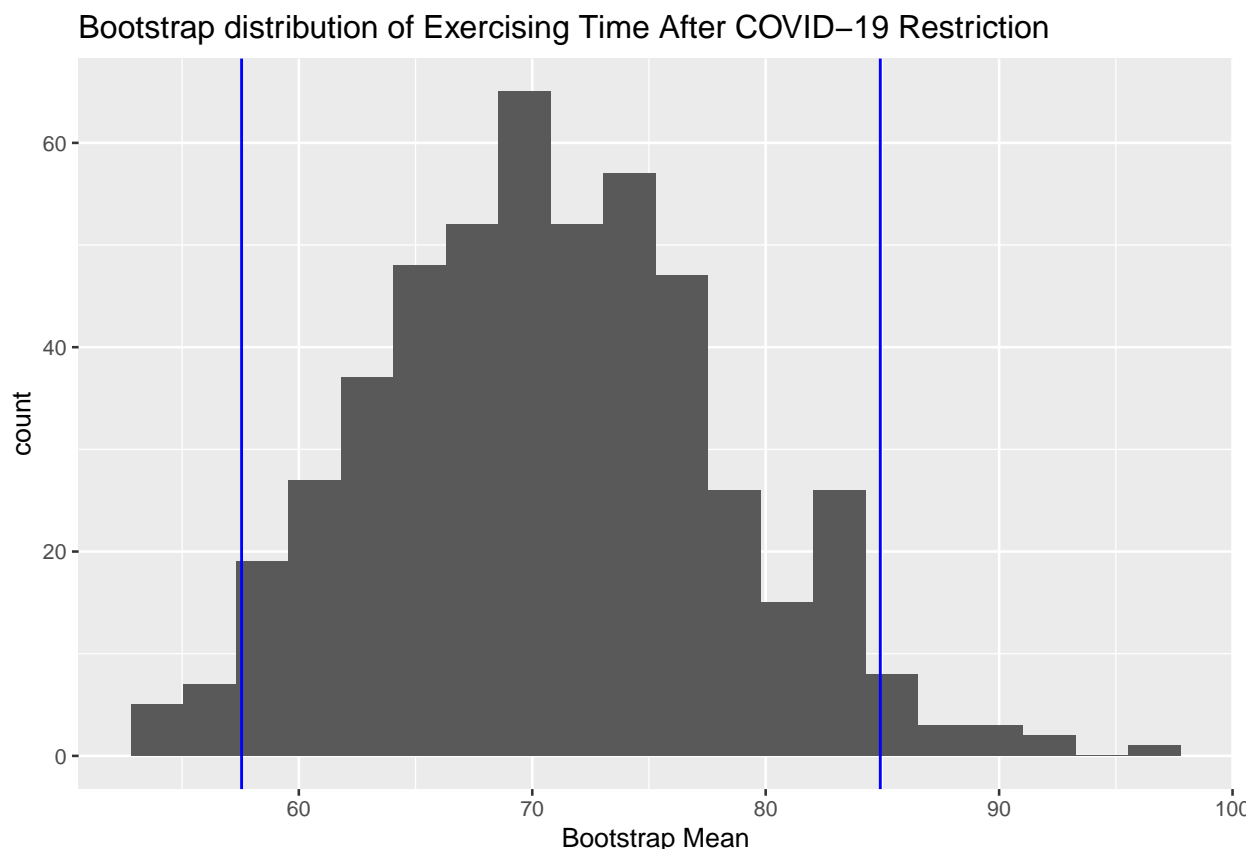
The range of the two cases is significantly different. We are 95% confident that the range of the mean of exercising time before the restriction was between 220.8950 minutes and 272.7575 minutes. However, after the pandemic, we are 95% confident the mean of the population is likely to be in the range of 54.5475 minutes and 84.9100 minutes as the average of exercising time per week. Therefore, the range has overall decreased. It implies that the overall exercising time has reduced and the lifestyle of students has changed.

As a total result, whether the residence was in hotspot or non-hotspot did not matter. There was no evidence to reject the null hypothesis of stating the non-hotspot and hotspot area students exercise the same amount after the regulation. Even though non-hotspot does not have harsh restrictions, people's behaviour was not different from people in the hotspot. Overall, the mean of the exercising time has decreased according to the result of the confidence interval. As an interpretation, the restriction may not be the simple factor that makes people exercise less. There must be other factors that reduce the exercising time. For example, people may feel nervous even with the mask to exercise in a limited area. Therefore, more space for the workout is required to stimulate the exercise time and reduce the concern of COVID-19 spread.

For further analysis, it would be better to increase the sample size. The data collected may not be representing the total population due to the small sample size. The sampling method can be kept. However, if there is more time and fewer cost constraints, it would be helpful to conduct a physical survey along with the online survey for the students who may not have access to Google Form. Also, adding more detailed questions and improving questions would be helpful for further analysis. For example, adding more questions about the participants may help because the life routine may impact exercise performance. With the more accurate result, it can be used to make better physical activity guidelines and workout facilities in the city for people to take care of the health at home during the COVID-19 restriction.

All analysis for this report was programmed using `Rstudio version 1.2.5042`, R markdown.

Bibliography

1. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

2. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: May 5, 2021)

3. Hadley Wickham and Jim Hester (2020) *readr: Read Rectangular Text Data.* https://readr.tidyverse.org

4. Wickham et al., (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686. (Last Accessed: May 13, 2021)

- The packages "tidyverse" and "readr" were used for the analysis. The tidyverse was used to clean the dataset and draw the graphs, such as `ggplot`. The readr package was used to read the CSV. data in r.