RESEARCH ON COVID-19 INFECTION RATES IN BEIJING, 2022

Jiaqi Cui(Jackie) and Jianyou Lu(Andy)

G11 AP Statistics

Beijing National Day School, Beijing, China

PROPOSAL

1. Research Question

With thousands of citizens getting infected, the COVID-19 pandemic repeatedly sprawled across Beijing in 2022. This research will be divided into two sections to analyze the data of the 2022 COVID pandemic. First, we believe that BNDS is rather easier for the virus to spread, and we will evaluate whether it is true. On a larger scale, we are also curious about if the proportion of Beijing citizens who got infected in each four seasons last year could be predicted by applying a model based on infection rates in 2020 and 2021.

2. Background Research

For the first section of our research, two articles have contributed to forming a hypothesis that infection rates in BNDS might be higher. Beaumont and Scammell reported in their article *Covid-19 may spread more easily among children than thought, report warns* that Georgia and the YMCA summer camp were shut down because of the rapid spread of COVID among teenagers. They claimed that children and teenagers tend to have "less severe symptoms" or to be asymptomatic, making the virus harder to detect and thus easier to transmit. An essay titled *Within and between classroom transmission patterns of seasonal influenza among primary school students in Matsumoto city, Japan* concludes that the probability of transmitting influenza is the highest "within class, followed by within grade and within school". Since COVID is also contagious and BNDS establishes a moving class system for students, our school is reasonable easier for transmission because students here meet different peers in each class.

For the second section of our research, two research have inspired us to utilize modeling by making the prediction of 2022 Covid-19. Since the hypothesis is that whether the number of infection reported by Beijing official account and the number of infection predicted by our models is same or not. Two models are the time series modeling, thus the related research is about *comparative study of statistical and machine learning models on near-real-time daily emissions prediction* the rapid ascent in carbon dioxide emissions is a major cause of global warming and climate change, which pose a huge threat to human survival and impose far-reaching influence on the global ecosystem. Therefore, it is very necessary to effectively control carbon dioxide emissions by accurately predicting and analyzing the change trend timely, so as to provide a reference for carbon dioxide emissions mitigation measures. This paper is aiming to select a suitable model to predict the near-real-time daily emissions based on uni variate daily time-series data from January 1st, 2020 to September 30st, 2022 of all sectors (Power, Industry, Ground Transport, Residential, Domestic Aviation, International Aviation) in China. It proposed 2 statistical model: auto regressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average with exogenous factors (SARIMAX); To evaluate the performance of these models, it must use some criteria: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) which will also be used in our models.

3. Sampling and Experimental Design

For the first section:

Variable: the COVID infection rate for BNDS students in 2022
Type of Study: this is an observational study because we will only ask whether students in BNDS have got COVID in 2022 and when did they get it.
Data collection: Data are collected through the posted questionnaire designed by the group.
Scope of inference: These data can be generalized to all BNDS students and will be compared with the official infection rate of Beijing citizens posted by the government.

For the second section:

Variable: the COVID infection rate for Beijing citizens in 2022

Type of study: this is an observational study because we use reported infection rates from 2020 and 2021 to predict that in 2022. During this process, no treatment is assigned to our subjects.

Data collection: Data are collected through online resources about infection rates in Beijing from 2020, 2021, and 2022.

Scope of inference: The data can be generalized to all Beijing citizens.

4. Explanatory Variable Analysis

2022 Spring	1903		2021 Spring	10		2020 Spring	180
2022 Summer	600		2021 Summer	58		2020 Summer	342
2022 Autumn	9861		2021 Autumn	74		2020 Autumn	16
2022 Winter	14546		2021 Winter	288		2020 Winter	98

For the first section: After collecting the data of 2022, we will conduct a chi-square Goodnes-of-Fit test for the proportion of infectors in BNDS in the four seasons, believing that the official infection rate for the four seasons in Beijing is true. The conditions checked should be random, independence(since there are about 5000 students in BNDS, we will take samples of size less than 500 to satisfy 10% condition), and a large sample size should be used(all the expected counts ≥ 5).
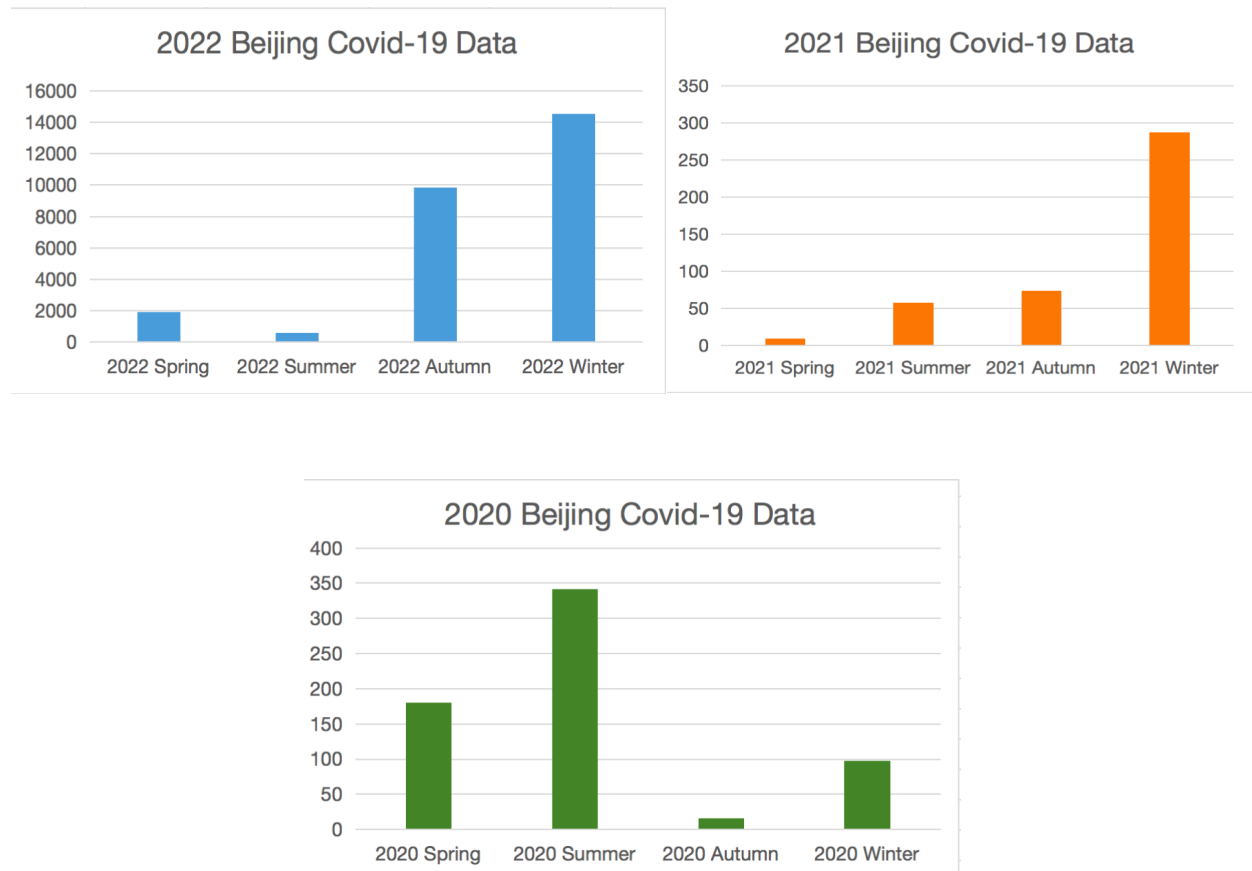
Figure 2: 3 years Beijing Covid-19 data

From the data collected, a bar graph can be conducted as shown in Figure 2

For the second section: After collecting the data of 2022, 2021, 2020. We will based on 2020 and 2021 official data of Beijing covid-19 to make the prediction of 2022 covid-19 infection rate by using the Sarimax model and Arima model. These two models need 2 years data to predict 2022's infection numbers. Sarimax and Arima need a huge amount of data to make sure the prediction is stable.

5. Group Task Assignments and Timeline

In this project, Andy and Jackie write and develop the questionnaire together. After collecting samples, Andy will conduct the Chi-Square Goodness-of-Fit test and end up with a statistically

significant answer for whether it is easier to get COVID in BNDS than in the city areas. Simultaneously, Jackie will construct a prediction model using data he collected for infection rates in 2020 and 2021.

The questionnaire will be sent in May 19, and it will stop collecting by the end of May 20. From May 21 to May 25, we will analyze the data we collected and will be continuously writing the final report. This project will be finished before May 26, and we will present our result to our class.

6. Data

Figure 3

北京	2022-11-08	38
北京	2022-11-07	36
北京	2022-11-06	44
北京	2022-11-05	49
北京	2022-11-04	41
北京	2022-11-03	38
北京	2022-11-02	34
北京	2022-11-01	33
北京	2022-10-31	33
北京	2022-10-30	18
北京	2022-10-29	17
北京	2022-10-28	25
北京	2022-10-27	11
北京	2022-10-26	15
北京	2022-10-25	26
北京	2022-10-24	19
北京	2022-10-23	14
北京	2022-10-22	11
北京	2022-10-21	25
北京	2022-10-20	18
北京	2022-10-19	19
北京	2022-10-18	37
北京	2022-10-17	17
北京	2022-10-16	19
北京	2022-10-15	28
北京	2022-10-14	22
北京	2022-10-13	18
北京	2022-10-12	14

Figure 3 shows part of the official Data

7. References

[1] ---. "Covid-19 May Spread More Easily Among Children Than Thought, Report Warns." *The Guardian*, 6 Aug. 2020, www.theguardian.com/world/2020/aug/05/covid-19-may-spread-more-easily-schools-than-thought-report-warns.

[2] Best, Alex, et al. "The Impact of Varying Class Sizes on Epidemic Spread in a University Population." *Royal Society Open Science*, vol. 8, no. 6, Royal Society, June 2021, p. 210712. https://doi.org/10.1098/rsos.210712.

[3] Smith, Amy. "Trash to Treasure: Using Text-to-image Models to Inform the Design of Physical Artefacts." *arXiv.org*, 1 Feb. 2023, arxiv.org/abs/2302.00561.