

Daniel Garcia
Tim Kelly
Parth Patel
Dixant Patel

About our data

Social justice and public health care issue:

We are studying the effect of Covid-19 on the education system and how the pandemic has impacted the quality of education in a negative way. This study will include students from various levels of education, ranging from high school to postgraduates.

Important questions and data sets:

- When will the school systems return back to the physical learning environment?
- How is the quality of education influenced by covid 19?
- How has covid-19 affected the mental health of students?

Data collection:

- Schoolsactivecovid.csv (schools active by country)
- <https://www-sciencedirect-com.subzero.lib.uoguelph.ca/science/article/pii/S0190740920310604> (quality of education data)
- Mentaldissordercovid-19datasets.csv (mental health surveys)
- Full_data.csv (case numbers)

Methods used for questions:

Question 1: When will the school system return back to the physical learning environment?

Method: Hypothesis Testing & Cluster Analysis

The method we chose to answer this question will be Hypothesis testing. Using the non-numerical data provided from our source, we plan on displaying the trend that we find from cluster analysis and analysis. Because we have three categories in our schools closed data (open, partially open, closed) we can use the k nearest neighbour algorithm to compare covid case numbers to the school status. We can then use this to say as cases change whether or not schools will likely change their status. The result provided would describe the situation of schools and whether or not it would be a good thing to return back to a physical learning environment. Looking at trends in covid case numbers we can predict the point at which they will be low enough that schools can reopen. When overall covid numbers will drop is of course a subject of much debate but our data analysis could still be useful as the situation develops to predict school changes in the near future.

Question 2: How is the quality of education influenced by Covid-19?

Method: Hypothesis Testing & Linear Regression

Our explanatory variable for the quality of education in our univariate model will be the number of covid cases. Through this testing, we can observe whether there has been no impact on the quality of education (null hypothesis) or if there has been a clear indication of impact on the quality of education (alt. hypothesis). The results of the data set provided to us is enough data to make conclusions about our hypothesis. This first requires us to come up with a single metric for quality of education to be explained. To do this we will combine factors such as “online classes attended per week”, and “Syllabus covered (%)” and use this in our first model. Then we can look at how significant of a difference these changes have been prior to COVID-19 using confidence intervals or p-values to make our conclusions. Linear regression is used when we want to predict the value of a variable based on the value of another variable. Due to COVID-19, school closure means a change in learning environment such that students now learn from home rather than in schools. In this case, we can predict the variables of “Time spending for study during the lockdown”, and “Syllabus covered (%)” and compare it with the “Online classes attended per week”. By comparing the dependent variables with the online classes attended per week, then we can determine the changes of education during COVID-19. The independent variable in these situations is the number of online classes attended per week while the other variables depend on this independent variable.

We will then create this metric with our data of “attended online classes before the outbreak of COVID-19” and compare our overall metric for quality of education from before covid. Once we know if the mean quality of education has increased or decreased and if the data is correlated to case numbers from above, we can assess the overall impact of covid on education.

Question 3: How has the pandemic affected the mental health of students

Method: Linear Regression

We will first compare our data on mental health in 2020 in the US with their covid case numbers using linear regression. The model will use increasing covid cases as an explanatory variable mental health. The null hypothesis will then be that covid cases are unrelated to low levels of mental health, and the alternative hypothesis will be that the number of covid cases are a determining factor in mental health. The alternative hypothesis can be confirmed if we get a p value of less than 0.05. As these data sets are independent and non binary values linear regression is appropriate (as opposed to something like logistic regression). We can then examine the mean levels of mental

health from our 2020 and 2019 datasets and compare. If the mean level of mental health is lower in 2020 and if we have previously concluded that covid case numbers are correlated with mental health then we can conclude that covid has had a direct negative impact on mental health.