

A5 Extension Plan

Ryan Williams

11/10/2021

How have COVID-19 infections impacted unemployment in Cuyahoga County?

Motivation

It is well documented that COVID-19 had a big impact on employment rates across the country. Many businesses were required to scale down or close entirely, leading to layoffs on a massive scale. I'm wondering how the COVID-19 infection rate directly impacted unemployment in Cuyahoga County - did more layoffs happen as infection rates climbed, or is it more likely that layoffs were based on governmental orders and not directly related to infection rates? Also, how localized was this effect - was Cuyahoga's unemployment rate more or less impacted than the other counties in the Cleveland-Elyria metropolitan area? These questions are inherently human centered, since they deal with the way people's livelihoods were impacted by COVID-19 at a local scale. I've heard a lot of statistics about unemployment in general throughout the pandemic, but this analysis would put into perspective what that impact looks like within a specific geographical area (Cuyahoga), and help us understand whether one county's unemployment can be impacted differently from others that are very close geographically.

Research questions and hypotheses

For this analysis, I have two research questions I want to answer::

- **Question 1:** Is unemployment in Cuyahoga County correlated to changes in COVID-19 infection rate?
 - **Hypothesis:** Unemployment rate and COVID-19 infection rate will have a high correlation (>0.5).
- **Question 2:** Was the unemployment rate in Cuyahoga significantly different than the unemployment rate in the other counties in the Cleveland-Elyria metropolitan area?
 - **Hypothesis:** Cuyahoga, being the most populous county in the metropolitan area, and containing by far the largest city (Cleveland) will show a statistically significant difference in changes in unemployment compared to the rest of the metropolitan area.

Data used

Monthly unemployment rates are recorded by the economic research department of the Federal Reserve Bank of St. Louis in a database called [FRED](#) (Federal Reserve Economic Data) which aggregates data from many other sources, such as the Bureau of Labor Statistics and Bureau of Economic Analysis. FRED has an extensive [terms of use](#) page, but provides a summarized statement on how FRED's data can be used: "As long as you don't engage in a prohibited/restricted use and do adhere to any applicable copyright restrictions, you are free to use FRED for your own non-commercial, educational, and personal uses". This proposed analysis doesn't include anything prohibited/restricted per the terms of use. FRED contains all the different levels of unemployment data I should need: the monthly unemployment rate for Cuyahoga County, and the unemployment rate for each of the other counties in the Cleveland-Elyria metro area. The datasets themselves are very simple, generally containing only the month and the unemployment rate (for example, this is the data on [Cuyahoga County](#)). This expands on the common analysis data by providing a new dimension (unemployment). However, this data isn't at a daily granularity, so the data from the common analysis will need to be aggregated up by month to be comparable. I don't believe there are any ethical concerns given the aggregated, anonymized, and public nature of the data.

Unknowns and dependencies

The biggest obfuscating variable in comparing infection rates to unemployment is the effect of government mandates, which aren't present in the data. Many businesses had to furlough employees as a result of government requirements - rising infection rates might signal this, but it's possible that governments were also acting on other political pressures rather than solely on infection statistics. There is also the issue of our Common Analysis infection data only existing for a specific time range (2020-04-10 to 2021-08-15) which limits the time period for which we can make inferences.

Methodology

I plan to take the following steps in this analysis:

- **Data Gathering:** I'll download the monthly unemployment rates from FRED for Cuyahoga County, and for the other 4 counties in the Cleveland-Elyria metropolitan area. Then I'll create a dataset that has monthly infection rate (aggregated from the daily Common Analysis dataset), monthly unemployment in Cuyahoga County, and monthly unemployment in each of the other Cleveland-Elyria metro area counties for comparison.
- **Analysis:** There are 2 research questions I'm looking to answer:
 - **Is unemployment in Cuyahoga County correlated to changes in COVID-19 infection rate?** For this question I'll be calculating the Pearson

correlation between unemployment rate and infection rate for the entire timespan of our infection data (2020-04-10 to 2021-08-15). Given that there may be a time lag in unemployment adapting to changes in infection rates, I'll also test whether introducing a lag of 1 month makes a difference in correlation.

- **Was the unemployment rate in Cuyahoga significantly different than that unemployment rate in the Cleveland-Elyria metropolitan area?** To answer this question I'll fit regression models to the unemployment time series for each of the counties in the Cleveland-Elyria metro area, then I'll compare for a statistically significant difference, at a 0.05 alpha level, between Cuyahoga and each of the other 4. Since I'll be doing 4 statistical tests at once, I'll use a Bonferroni correction and divide the alpha level by 4 to account for the increased possibility of type 1 error - effectively using a 0.0125 alpha for each test.
- **Visualization:** I plan to present the results using 2 visualizations and accompanying statistics:
 - The first visualization will be a line chart plotting COVID-19 infections and unemployment rate over time. This visual should help show how these two time series move together (if they do at all). I'll annotate the chart with the calculated Pearson correlation.
 - I'll supplement this visualization by either adding another series, or a separate chart, showing how COVID-19 infections with a 1-month lag compare to the unemployment rate, under the assumption that unemployment might take some time to reflect changes in infections.
 - The second visualization will show the unemployment rate for each of the 5 counties within the Cleveland-Elyria metro area, with their regression lines overlaid. For readability, this will probably be a grid of 5 separate charts. I'll annotate each county's chart with its regression coefficient and the p-value of the comparison between Cuyahoga County and the county in question.

Timeline to completion

There are 2 major deadlines for this project (the presentation and the final report) so the rest of the schedule is based on working backward from those. Dates in parenthesis are when I expect to have accomplished each milestone:

- **Collect data (11/12):** Perform the data collection mentioned above
- **Analyze question 1 (11/15):** Complete calculation of the correlation between infection rate and unemployment rate
- **Analyze question 2 (11/19):** Complete the regression modeling and hypothesis testing between models
- **Visualize the results (11/25):** Create the charts mentioned above

- **Document the process (11/30):** Document everything thus far in preparation for A7
- **Create presentation (12/07):** Format the findings for presenting per A6 instructions
- **Create final report (12/14):** Create final report per A7 instructions