Econ 388 Data Assignment 3

# Research Question

In this assignment you will use econometrics to answer the following question: what effect did county-level per-capita Covid-19 mortality rates in 2020 have on housing prices in 2021?

# Data

Your first task is to gather the data required to answer the question. You will have to gather the data from several sources and merge them into a final analysis dataset.

## County-level Covid-19 mortality data

The New York Times has compiled county-day data on Covid-19 cases and deaths, which you can read about and access here:

<https://www.nytimes.com/article/coronavirus-county-data-us.html>

Remember that you will be aggregating the daily data to mortality for the year 2020.

## County-level population data

In order to convert raw mortality numbers to per-capita mortality rates you will need data on each county’s population, which you can download from the U.S. Census bureau at the following website:

<https://www.census.gov/data.html>

## ZIP-level housing price data

The Federal Finance Housing Agency (FHFA) maintains housing price indices at the ZIP (3-digit) level over many years. You can read about and access this data from the following website:

<https://www.fhfa.gov/DataTools/Downloads/Pages/House-Price-Index-Datasets.aspx>

## ZIP-county crosswalk

Because there are often several ZIP codes in one county, you will need to aggregate the ZIP-level data to the county level. This requires knowing which ZIP codes correspond to which county, which you can find at the U.S. Department of Housing and Urban Development’s Office of Policy Development and Research’s website:

<https://www.huduser.gov/portal/datasets/usps_crosswalk.html>

# Empirical strategy

You should determine an appropriate strategy to identify and estimate the effect of interest. In the context of a regression, this means, determining what exactly your outcome variable will be, what your main regressor of interest will be, and what controls you will also include in your regression. Make these decisions with an eye to plausibly satisfying the assumptions needed for your regression to have a causal interpretation. Be prepared to explain and defend your strategy in the writeup.

# Write up

Communicate your analysis and findings in the form of a polished, professional, well-written memo. Your memo should have the following headed sections:

1. Introduction, in which you state the research question you are answering, and discuss why it is an interesting or important question
2. Data, in which you describe each of your data sources and how you chose and constructed your analysis variables. This section should also include a table of summary statistics, as well as a visualization (or visualizations) of your key variables.
3. Empirical Analysis, in which you describe your empirical strategy, give your regression specification, and clearly articulate and defend the assumptions you are making to identify the effect of interest. In this section you will also report and discuss the results of your analysis. Describe your findings in prose and also present them in table form.
4. Conclusion, in which you briefly summarize the findings, any limitations to your results, and perhaps discuss further analysis that could be done on the topic

The writing, tables, and figures should be polished and presented in a professional way. Copied-and-pasted Stata output is not appropriate here.

# Supporting materials

You should include in your submission a link to a GitHub repository, Dropbox folder, Google Drive folder, or other cloud-accessible directory that includes:

* all raw data files from which your final analysis dataset was constructed
* code (i.e., a Stata do-file) that is clearly documented, and includes the reading in of the raw data files, all manipulations needed to create your final dataset, and the econometric analysis that produces the results in your write up. I should be able to run this code using only the raw data files and produce all of the results in your write up.
* a readme file that explains what each of the files in this directory is and their sources (i.e., URLs in the case of the raw data files)