Final Exam - Take-Home Portion

EC 421: Introduction to Econometrics

Due *before* midnight on Thursday, 17 March 2022

Instructions

TAKE-HOME EXAM This document provides the *take-home* portion of the final exam for 421. It is due before midnight on Thursday, March 17th. There is also an on-Canvas portion on Tuesday, March 15th.

INTEGRITY You must work alone on this exam. Do not discuss any part of this exam with any other person. You may use any class materials and online resources (it is open book).

If you are suspected of cheating, then you will receive a zero—for the assignment and possibly for the course. We may report you to the dean. **Cheating includes copying from your classmates, from the internet, and from previous assignments.**

IMPORTANT As with all of your problem sets, you must submit **two files**:

1. your typed responses/answers to the question (in a Word file or something similar)
2. the R script you used to generate your answers. Each student must turn in her/his own answers.

If you are using RMarkdown, you can turn a single file, but it must be a `html` or `pdf` file with **both** your R code **and** your answers.

If we ask you to create a figure or run a regression, then the figure or the regression results should be in the document that you submit (not just the code—we want the actual figure or regression output with coefficients, standard errors, *etc.*).

Air quality

The dataset `data-final.csv` describes the air quality for 543 counties in the United States during 2020, along with some information on the counties' demographics (specifically, total population and racial composition).

The pollution variable is `pm_24hr`. Here, `pm` stands for (fine) *particulate matter* (PM2.5), which is one of the most damaging forms of common pollution (you can learn more here). This specific measure, the 24-hour measure, tells us about some of the highest values of PM2.5 observed in the county during 2020.

The final page describes the individual variables.

Q01 Load whichever packages you think you'll need and then load the dataset (`data-final.csv`).

Q02 Is this dataset an example of "cross-sectional data" or "time-series data"? Explain your answer.

Q03 Create histograms for the variables `pop_total` (total population), `pct_nonwhite` (the percent of the population that does not identify as 'white'), and `pm_24hr` (measured PM pollution).

Make sure the histograms are well labeled.

Q04 Create a scatterplot where the percent of the population who does not identify as white is on the x axis (`pct_nonwhite`) and the **log** of PM pollution is on the y axis (`log(pm_24hr)`).

Note: You will either need to create this logged pollution variable or you can use the `log()` function inside your `ggplot` command.

Q05 Regress the **log** of PM pollution (the log of `pm_24hr`) on the non-white percent of the population (`pct_nonwhite`). Report your results (no interpretation needed).

Q06 Interpret the coefficient on `pct_nonwhite`.

Q07 How would measurement error in `pct_nonwhite` affect the estimated coefficient on `pct_nonwhite`?

Q08 Suppose urban areas have (1) more pollution and (2) larger non-white population percentages. Will omitting the variable "urban" from our regression cause us to over-estimate, under-estimate, or have no effect on the coefficient on `pct_nonwhite`? Explain your answer.

Q09 Regress the **log** of PM pollution (the log of `pm_24hr`) on the non-white percent of the population (`pct_nonwhite`) and the **log** of total population (`pop_total`). Report your results (no interpretation needed).

Q10 Interpret the coefficient on `log(pop_total)`.

Q11 Based on the change in the coefficient on `pct_nonwhite` between the regressions in **Q05** and **Q10**, do you think omitting population was causing bias? Explain your answer.

Q12 Should we be concerned about autocorrelation in this setting? Briefly explain your answer.

Q13 Should we be concerned heteroskedasticity and/or correlated disturbances? Briefly explain your answer.

Description of variables and names

Variable	Description
county	County name
county_code	County code (FIPS)
state	State name
pop_total	Total population (2010)
pct_aian	Percent of population identifying as American Indian or Alaska Native
pct_asian	Percent of population identifying as Asian
pct_black	Percent of population identifying as Black
pct_nonwhite	Percent of population identifying as any non-White race
pct_white	Percent of population identifying as White
pm_24hr	Pollution: Fine particulate matter level (24-hour standard)