Midterm: COVID-19 Case Study

WRITE YOUR NAME HERE

March 31, 2025

Contents

Instruction

All the teaching team members will be available from 7:00 - 9:10 PM. The submission will be closed sharp at 9:10PM.

Instruction: This midterm requires the use of R. You are allowed to refer to lecture notes, homework; however, any use of the internet or large language models (e.g. ChatGPT) is strictly prohibited. Write your answers using this .rmd file and knitr it into the html file. Show your codes, plots or R-output when needed. You always need to show your code with `echo = TRUE` which is the default setup for this file. If you have trouble formatting the plots, don't worry about it. We are not looking for pretty solutions, rather to see if you are able to make sense out of the data. Make sure the compiled html (and/or pdf) file shows your answers completely and that they are not cut-off. Throughout the exam, you do not need to use any LaTeX or mathematical equations. **Whenever we ask for test at some significant level, assume all the model assumptions are satisfied.**

All the answers should be clearly supported by relevant R code or based on the R output. There are many ways to provide answers.

There are 4 questions with various parts:

DO NOT spend too much time on a single question. Come back to where you stuck after you have tried all the questions.

Files needed for the midterm:

- `Midterm/midterm.Rmd`
- `Midterm/midterm_25.csv`

Electronic Submission: Two files needed: your `.rmd` file and a compiled html file. If you have trouble submitting the files to Canvas, email them to lzhao@wharton.upenn.edu, dongwooo@wharton.upenn.edu and neil.fasching@asc.upenn.edu.

Label them with your full name. In the `Assignments` section, go to the `Midterm` assignment and upload your completed files.

The submission folder will be closed sharp at 9:15PM.

On Site Help: We will answer any clarification questions. We may also help out with some minor code issues. We will, however, not provide any answers as to what functions to use for example.

Raise your hand if you want to talk to one of us.

In case of emergency, here is Linda's cell: 6106590187 (text or call her)

Background

The outbreak of the novel Corona virus disease 2019 (COVID-19) was declared a public health emergency of international concern by the World Health Organization (WHO) on January 30, 2020. Upwards of 112 million cases have been confirmed worldwide, with nearly 2.5 million associated deaths. Within the US alone, there have been over 500,000 deaths and upwards of 28 million cases reported. Governments around the world have implemented and suggested a number of policies to lessen the spread of the pandemic, including mask-wearing requirements, travel restrictions, business and school closures, and even stay-at-home orders. The global pandemic has impacted the lives of individuals in countless ways, and though many countries have begun vaccinating individuals, the long-term impact of the virus remains unclear.

The impact of COVID-19 on a given segment of the population appears to vary drastically based on the socioeconomic characteristics of the segment. In particular, differing rates of infection and fatalities have been reported among different racial groups, age groups, and socioeconomic groups. One of the most important metrics for determining the impact of the pandemic is the death rate, which is the proportion of people within the total population that die due to the the disease.

==================== **End of the Midterm** ====================

Appendix

Data Summary

The data comes from several different sources:

1. County-level infection and fatality data - This dataset gives daily cumulative numbers on infection and fatality for each county.
   - NYC data
2. County-level socioeconomic data - The following are the four relevant datasets from this site.
   i. Income - Poverty level and household income.
   ii. Jobs - Employment type, rate, and change.
   iii. People - Population size, density, education level, race, age, household size, and migration rates.
   iv. County Classifications - Type of county (rural or urban on a rural-urban continuum scale).