

# Principles of Empirical Analysis

## Final exam

February 21, 2022

Tuukka Saarimaa and Matti Sarvimäki

Please answer all questions. A perfect answer is complete, but short and focused entirely on the question asked. Please write clearly, and refrain from unsubstantiated assertions and guesses. You may use any reference material. Submit your answers through MyCourses in a single pdf-file by **12:00** (the same way as you did with the course exercises). Please note that you can update your answers until 12:00. Include your name and student ID in the file (top left corner). If for some reason you have trouble submitting through MyCourses, you can send your answer through email to [tuukka.saarimaa@aalto.fi](mailto:tuukka.saarimaa@aalto.fi). Also, if you have any questions during the exam, send your question through email to [tuukka.saarimaa@aalto.fi](mailto:tuukka.saarimaa@aalto.fi).

### Question 1

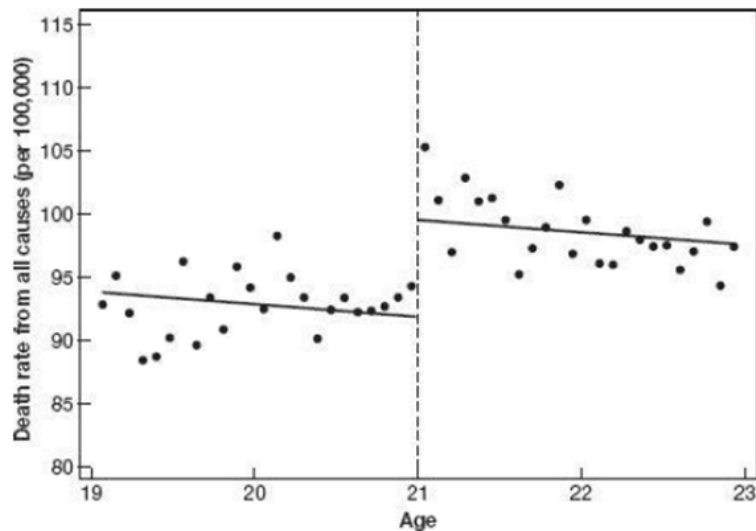
In order to decrease differences in learning outcomes and educational inequality, the Finnish government has decided to examine the possibility of providing pre-primary education over a two-year period (Programme of Prime Minister Sanna Marin's Government 2019, p. 176). For background, in the current system children have one year of pre-primary education, typically at the year they turn six years old, after which they enter the first grade of primary school. Thus, here the treatment is adding an extra year to pre-primary education, i.e. to start it already at the year children turn five years old.

Imagine that the government has asked you to design an experiment that will determine whether the government will implement a reform that would provide mandatory two-year pre-primary education for all children living in Finland. You are given free hands except that (a) you cannot force parents to enroll their children in the experiment and (b) your budget covers the cost of enrolling 15,000 children for an additional year of pre-primary education. You can assume that collecting data does not incur any additional cost and that there are in total 35,000 children in your potential target population.

In broad terms, please discuss how you would design such an experiment. Please make sure that you clearly explain the reasons behind your choices. A perfect answer will also include a discussion on the implications of possible non-compliance and other possible limitations for your experiment. When discussing minimum detectable effect size, you can assume that your chosen outcome measure has a standard deviation of one.

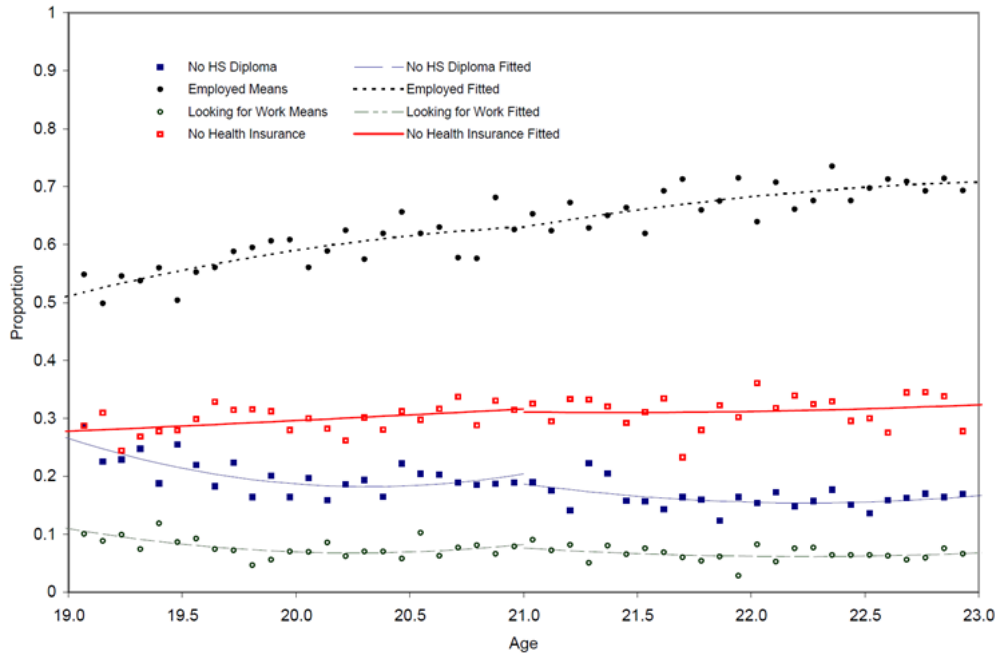
## Question 2

Below is a figure from the lectures relating death rate and age.



*Notes:* This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

- What is the research design called that this figure is related to?
- What are the three fundamental components related to this research design? What are they specifically in this case?
- What is the underlying assumption needed for this design to work?
- Below is another figure from the lectures. Explain why this figure enhances the plausibility of the research design.



### Question 3

You are interested in the effect of mandated mask wearing on the spread of the coronavirus. From January 1 2021 onwards, some Finnish municipalities mandated mask use in all public venues, while other municipalities did not. You are given weekly data on municipal level infection rates from November 1 2020 to March 31 2021 for all municipalities.

- (a) How would you test whether mandating masks has affected the spread of the coronavirus using this data? Illustrate your strategy using a figure. What is your research design called?
- (b) What are the key assumptions for your research design to work? Can you directly test these assumptions? How plausible do you think these assumptions are in this particular case?