Principles of Empirical Analysis: Practice questions

2024

Exam instruction text: 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. (note that the real exam will only have 3 questions and you should answer all 3)

Question 1

In our fifth lecture, we discussed a paper by Chattopadhyay and Duflo (2004). To remind you what it is about, here is the abstract:

> This paper uses political reservations for women in India to study the impact of women's leadership on policy decisions. Since the mid-1990's, one third of Village Council head positions in India have been randomly reserved for a woman: In these councils only women could be elected to the position of head. Village Councils are responsible for the provision of many local public goods in rural areas. Using a dataset we collected on 265 Village Councils in West Bengal and Rajasthan, we compare the type of public goods provided in reserved and unreserved Village Councils. We show that the reservation of a council seat affects the types of public goods provided. Specifically, leaders invest more in infrastructure that is directly relevant to the needs of their own genders.

We also spent a lot of time looking at this extract from one of their tables:

	West Bengal		
	Mean, Reserved GP	Mean, Unreserved GP	Difference
Dependent Variables	(1)	(2)	(3)
A. *Village Level*			
Number of Drinking Water Facilities	23.83	14.74	9.09
Newly Built or Repaired	(5.00)	(1.44)	(4.02)

As you may recall from the lecture, it reports the mean number of drinking facilities newly built or repaired (henceforth the "outcome") in reserved and unreserved villages. Using the information above, please answer the following questions:

(a) Why is it important for the research that the villages reserved for female leaders were selected randomly?

(b) The third column reports two numbers (9.09 and 4.02). Please explain what these numbers are and briefly discuss their interpretation.

(c) Is there a statistically significant difference at 5% level in the mean outcome between reserved and unreserved villages? How can you tell? That is, please include the definition of statistical significance in your answer.

(d) What do Type I and Type II errors mean in the context of this study?

(e) How would increasing the sample size affect the likelihood of finding a false positive? How would it affect the likelihood of finding a false negative?

Question 2

In order to decrease differences in learning outcomes and educational inequality, the outgoing Finnish government decided to "examine the possibility of providing pre-primary education over a two-year period." For background, in the old 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., starting it 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 you cannot force parents to enroll their children in the experiment. In broad terms, please discuss how you would design such an experiment. (Hint: a perfect answer includes discussing sampling concerns, the implications of possible non-compliance and other possible limitations for your experiment.)
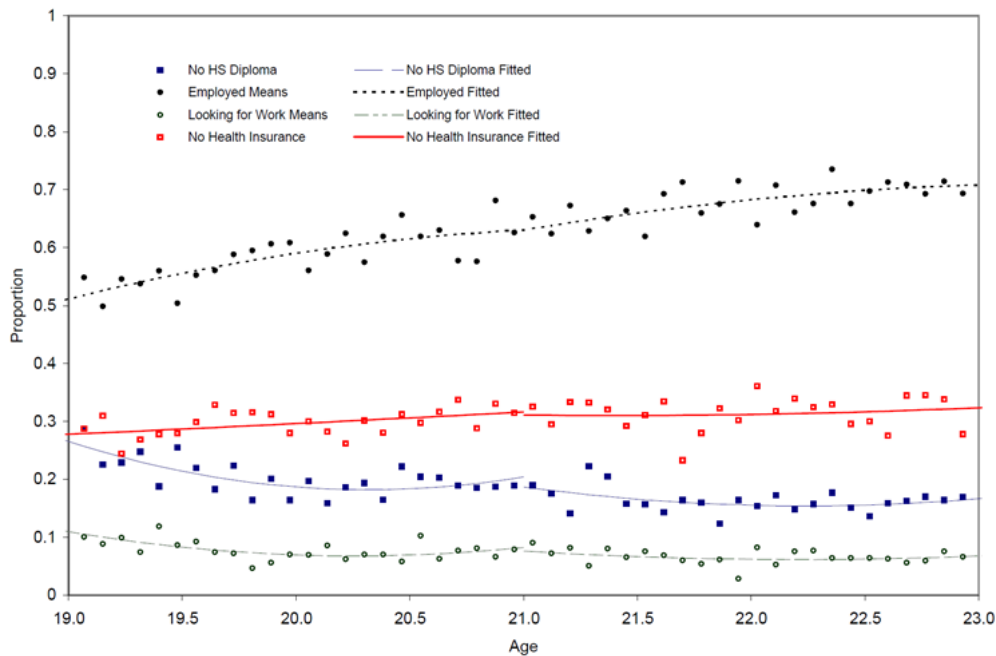
Question 3

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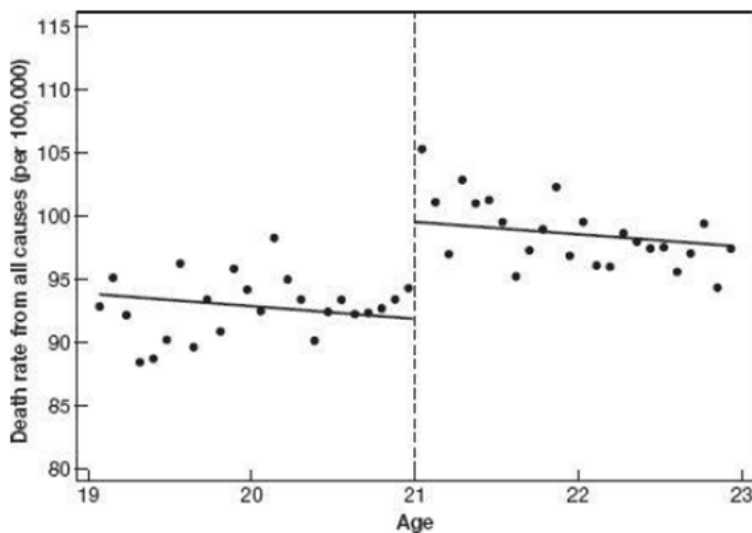
Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

(a) What is the research design called that this figure is related to?

(b) What are the three fundamental components related to this research design? What are they specifically in this case?

(c) What is the underlying assumption needed for this design to work?

(d) Below is another figure from the lectures. Explain why this figure enhances the plausibility of the research design.

Question 4

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?