

# ECO 251: Take-Home R Exam

Sounak Thakur

April 12, 2024

**Due: Wednesday 20<sup>th</sup> April 12:30 PM**

## Instructions:

You are free to consult your lecture notes, textbooks or any other resource (offline/online/AI). You will be on your honor to not seek or receive any human assistance. Submit your solution on MooKIT. Your submission should contain an RScript file and a .pdf file containing a write-up in response to each question. In your write-up, you should present relevant output (summary stats, regression tables, etc.) from R to support your answers. If you use a regression, you should always write down the specification.

## Background

You are provided with an extract from the India Human Development Survey (IHDS in short). The IHDS is a nationally representative panel dataset containing rich information on demographic and individual characteristics of residents in surveyed households. In this exercise, we will investigate the association between a mother's widowhood and children's educational attainments. Typically, households headed by widowed mothers are poorer, but these mothers have more autonomy. Many studies indicate that higher maternal autonomy translates into better performance of children in school. Based on these considerations, the association between widowhood and children's educational attainment appears ambiguous *a priori*. Therefore, it is an empirical question worth exploring.

You may access the data extract at this [link](#). The data extract from the IHDS data is a two-wave panel dataset. Thus, each child is observed twice, once at wave 1 (2005) and once at wave 2 (2012). Each row in the dataset is a child observed in a given year. The outcome variable of interest is the child's standardized years of education ("std-edu"). The variable "widow" is an indicator for widowhood status of the child's mother in the corresponding wave. All other variables have been assigned fairly descriptive labels which should be self-explanatory. Based on these pieces of information, answer the questions that follow:

## Questions:

[20 points]

- (i) Which variable(s) is/are the unique identifier in your dataset? How did you confirm this using R? [2 points]
- (ii) Using R, create a table of summary statistics for key variables in your analysis. Does the rate of widowhood look reasonable? [3 points]
- (iii) Is it true the households headed by widowed mothers are poorer (in the sense that they earn less)? [3 points]  
[Hint: We do not have a direct measure of income, but consumption expenditure per capita is a good proxy for income.]
- (iv) In studying the association between widowhood and educational attainments, why does it make sense to use standardized years of education rather than completed years of education? [2 points]
- (v) Comment on the association between widowhood and children's educational attainment. What background characteristics of the household would you like to control for? [5 points]
- (vi) Does widowhood appear to affect girls differently than it affects boys? [5 points]