

Research Associate Assignment

Data Analysis

You have been provided with a data set that contains scores of children during baseline and endline periods from control and intervention locations. In the annexure, the "Scoring guidelines" details the maximum score for each question. You are requested to do the following data analysis in Stata/ R:

1. Import the data into Stata/ R
2. Merge the baseline and endline data to make one data set.
3. Assign labels/codes for each question and their option.
4. Clean data where necessary. Stata should have text detailing the cleaning done.
5. Where you have doubt, make the necessary assumption to complete the task. Document the assumptions made in the do file, next the syntax/command used.
6. Save the data in a ". dta" format
7. Generate data tables reporting summary statistics (across gender, age, caste) for students who scored more than average total score vs those who scored less than average total score.
8. Generate summary statistics tables comparing student scores at baseline and endline for those in intervention and control groups.
9. Generate summary statistics tables (age, gender, caste) for all the districts
10. use the necessary tests of significance, where necessary, and provide an explanation for the same
 - a. Run a difference-in-difference (DID) analysis model to show the difference in results/scores between students in intervention and control groups. Include all the appropriate independent variables in the DID model.

Deliverables from the data analysis:

- *Do file with the analysis*
- *A two-page note with the analysis and interpretation of the results*

Writing

A 1-2 page note on any of the following topics:

- Growth opportunities of EdTech for the low-income segment in India
- Potential for reversible methods of contraception in India
- Gender integration in the food fortification supply chain