Training for Success? Training Centers and Indian Engineering Entrance Exams
Faculty Sponsor: Dr. Ina Ganguli
Undergraduate Research Assistant: Vikram Sablani
University of Massachusetts Amherst, Department of Economics

Introduction

To enter various engineering colleges in India, including the elite Indian Institutes of Technology (IIT), students must take an entrance exam, called the Joint Entrance Examination (JEE). This project is focused on understanding the effectiveness of a set of training centers that prepare students from disadvantaged backgrounds for the JEE. The research team has access to student-level data on participants in several training centers throughout India, including information on demographics, parental wealth, and test scores prior to taking the JEE, called the admission data set. The team also has a data set with students actual scores on the JEE and the college they attended, called the test data set.

Research Questions

Who participates in a training center program? How does participating in a training program affect someone's JEE score?

Role of the Research Assistant

My role in this research project was to assist with construction of the dataset that will be used for the analysis. My main task was to understand why the two data sets that we had acquired had a lot of unmatched data when we merged the two in STATA. I proceeded to clean the data sets in STATA to reduce the number of mismatches so that the merged data set would be more accurate.

Data and Methods

The first task was to identify why the merged data sets had unmatched data. I did this by writing code in STATA to export only the data points that were unmatched to two different Excel sheets, one for each data set.

In Excel, I used the highlight duplicates functions to see which data points were duplicates or had some form of similarities between them. I then exported this information to a third excel sheet where I went through the sheet to see why the duplicates didn't match. I found that there was misspelling of names and center names in the data set. For example, in the admission data set one of the centers was spelled as Sivsagar while in the test data set the same center was spelled as Sivasagar which lead to data not being matched.

In STATA, I wrote code to replace names of unmatched students and center names in each data set. This resulted in the number of matched observations to increase from 499 to 607.

Results and Discussion

Figure 1 shows the average test scores of students prior to taking the JEE on the X-axis and their ranking nationwide on the JEE on the Y-axis. It shows the scores by one's caste. It also includes a fitted line for SC, which represents Scheduled Caste, the disadvantaged group, and GEN, which represents the
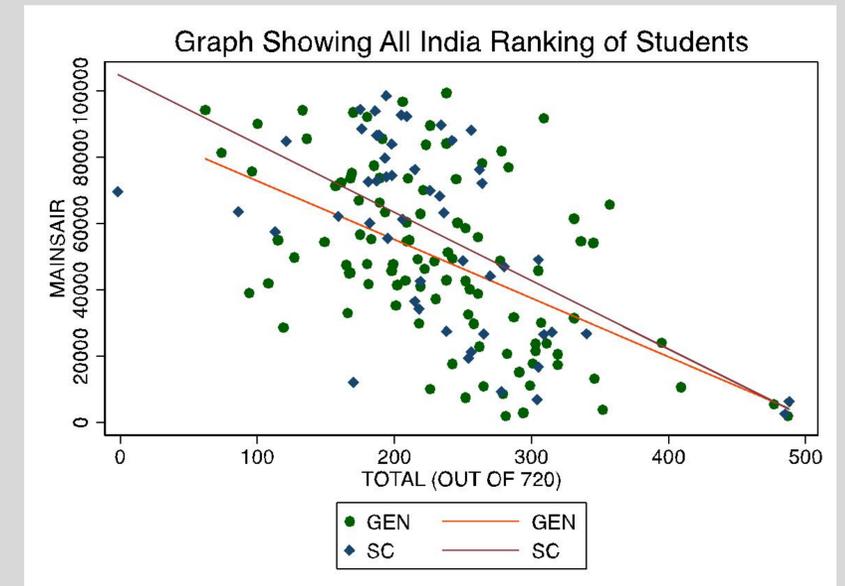

Figure 1

general caste. The graph shows that scores on the tests prior to the JEE are correlated with performance on the JEE (higher rank). There does not seem to be much difference between the SC and GE in this correlation, but SCs appear to do slightly worse on the JEE relative to the practice test scores at the lower practice scores.

Conclusion

People from the Scheduled Caste are more likely to participate in the training center program. Even though the data shows that people from the Scheduled Caste have lower scores then people from the general caste, future research will investigate how participating training center program affects someone's JEE score particularly by caste.