

MIT Blueprint Labs

Data Task

Thanks for your interest in working with MIT Blueprint Labs. As the next step in your application process, please complete the data task below. If you have questions, please consult our [Data Task FAQ](#).

Overview

Government agencies and private foundations distribute billions of dollars in financial aid to college-goers in hopes of boosting educational attainment. Tennessee, for example, guarantees no-cost tuition for all high school graduates at in-state community and technical colleges. In this exercise, you'll investigate the impact of the Tennessee Promise on college enrollment. You'll assemble a clean data set and use it to answer a few questions. When you're finished, please combine the following files in a single, zipped folder:

- raw data files in their original format, including those we have provided
- your clean data
- any programs you wrote to produce your analysis
- a 3- to 4-page PDF (single spaced 12-point font with 1" margins) answering the 4 questions below.
- a readme text file describing the contents of the folder

The enclosed PDF contains instructions for submitting your zip file on Canvas. Please use Stata to perform the analysis. In case you do not have access to Stata, please let us know. You're also welcome to use another software, preferably R. Your goal is to produce files that other members of your research team could use without further explanation. Please keep that in mind as you organize and document your work. If any issues or questions arise while you're working, please note them in your write up. This task will help us assess your current qualitative and quantitative research skills, but a "perfect" score is by no means required for the job. With that in mind, please do this work yourself without input from other people.

A. Assembling the Data

The first step is to assemble a data set tracking enrollment and aid awards at Tennessee colleges over time. The U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS) provides the data we need. Your data folder contains two types of extracts:

- "Directory Information" files from the "Institutional Characteristics" survey
- "Student Financial Aid and Net Price" files from the survey by the same name

Please use these raw files to create your panel data set. Your cleaned data should contain the following variables:

- ID_IPEDS: a unique identifier for each institution
- year: the four-digit academic year, where "2010-11" is coded as 2010, "2011-12" is 2011, and so on
- degree_bach: a dummy variable that identifies bachelor's degree-granting institutions

- public: a dummy variable that identifies public institutions
- enroll_ftug: the total number of first-time, full-time degree/certificate-seeking (henceforth, “FTFT”) undergraduates
- grant_state: total amount of state and local grant aid awarded to FTFT undergraduates
- grant_federal: total amount of federal grant aid awarded to FTFT undergraduates

Some additional notes:

- Your panel should cover the academic years 2010-11 through 2015-16. We'll use a strongly balanced panel, so include only the colleges in Tennessee for which enrollment and grant aid information is available in all years.
- Restrict the sample to undergraduate institutions, which may offer bachelor's and associate degrees, as well as diplomas and certificates. Institutions that only offer graduate degrees should be excluded from your sample.
- Include both public and private institutions in your sample.

B. Analysis

The Tennessee Promise guarantees no-cost tuition for all Tennessee high school graduates who attend the state's public community and technical colleges. For the sake of concision, we'll refer to schools that offer bachelor's degrees as “four-year colleges” and schools that don't as “two-year colleges.” (We recognize that, in practice, time to degree completion varies widely across students and credentials). We'll also distinguish between public and private schools, so your analysis will compare four groups:

- public, two-year colleges
- public, four-year colleges
- private, two-year colleges
- private, four-year colleges

Given these categories, community and technical colleges will fall in the “public, two-year” group.

Please answer the following 5 questions. Throughout your answers, please be brief. We value clarity more than length.

1. Create a table of summary statistics of the data, broken down by school year:
 - a. Number of schools that are bachelor's degree-granting institutions, by year
 - b. Number of public institutions, by year
 - c. The mean total number (to the nearest student) of FTFT undergraduates, by year
 - d. The mean total amount (to the nearest dollar) of state and local grant aid awarded to FTFT undergraduate, by year
 - e. The mean total amount (to the nearest dollar) of federal grant aid awarded to FTFT undergraduates, by year
2. Tennessee Promise scholarships were first offered to students in the 2015-16 academic year. Create two graphs:

- a. one that compares average school-level state plus local grant aid across the four types of institutions during the sample period
- b. one that compares average school-level enrollment of first-time, full-time undergraduates across the four groups during the sample period

Describe your findings for the first graph.

3. Specify a regression model of your choice that you believe will generate the most accurate numeric estimate of the causal effect of the Tennessee Promise program on enrollment at public, two-year colleges. Explain how the graphs you created in question 1 inform your model choice. Present your results in a table.
4. Do you think your regression model provides an unbiased estimate of the program's causal effect on enrollment? Describe any assumptions required for causal interpretation of these estimates. Explain any reservations you may have about the results.
5. How would you design an experiment to test the effectiveness of no-cost tuition on students' college outcomes? What outcomes beyond enrollment would you include? What types of data would you need and what are some potential data sources? Please keep this answer short – we are not expecting a fully-fledged research proposal with data permissions and power calculations here! Please keep this answer to no more than half a page.