

This read.me provides information for replicating results of *Semesters or quarters? The effect of the academic calendar on postsecondary student outcomes*, written by Valerie Bostwick, Stefanie Fischer, and Matthew Lang and published in 2022 in American Economic Journal: Economic Policy. This replication adds one year (2017) of data to the analysis.

Email Raines Lucas at rlucas6@wisc.edu for further questions and assistance.

For a general idea of replicating results, readers are directed to the README_original file, the original read.me written by the paper's authors. This read.me functions as a supplement to that document, describing how to overcome issues that will arise and how to add another year of data.

Here are the two main issues that will arise when following the original read.me:

Issues

Issue 1: Downloading Appropriate Data

Not only is the raw data not given, but the naming in the subsequent do-files was arbitrarily generated by the authors, so the replicator has no idea what datasets they used. For example, they'll call on a csv called "ic1998_stata.csv", but that name does not exist anywhere on the IPEDS dataset. In fact, the appropriate file to use was "ic9798_hdr.csv". Moreover, these naming conventions do not remain constant over time. For example, when downloading institutional characteristics, there are up to 6 possible datasets to choose from *for each year*. In 1990, the appropriate dataset was "ic_1990abce". In 1991, "ic_1991c". In 1992, "ic_1992b". Clearly the IPEDS database did not have a common naming system. Each of these data frames represents the same information with different file names. In the do-files, however, one naming system (chosen by the authors) was used across all years, despite the changes in the names of the raw files.

Solution: After navigating to the IPEDS [website](#), use the filters above to select the data you are looking for. As mentioned above, there will be several datasets for each year and each type of data. The data for the year 2016 will match the naming convention in the appropriate 2016 do-file. From there, you must work backwards. Each year, find the dataset that has the same data as the 2016 dataset. Rename that year's dataset so that it matches the convention of the 2016 one.

Issue 2: Bugs in STATA code

Some of the 189 do-files run properly, but most do not. Most either call on a package that does not exist, use deprecated code, or have some erroneous error. The most common error was labeling certain non-string data as strings. This was likely due to a change in the data after publication. This meant that almost every one of the 180 yearly do-files required manually inspecting and deleting hundreds of dysfunctional lines of code.

Solution: The bugs will primarily be in the years 2004-2007. Open the invalid do-files and find which line is causing the string error (STATA will tell you). Delete that line. Repeat until there are no more errors with the do-file.

Adding 2017 to Analysis

Fortunately, adding another year is the easier part of analysis. For each intermediate “create” do file, add a line that calls in a 2017 year-specific do-file. Create this 2017 do-file by copying the 2016 do-file and changing any calls to “2016” to read “2017” instead.