Graduating Midwest Liberal Arts Colleges:

The influence of student characteristics on graduation rates

Sneha Verma

Department of Economics, Accounting, and Management

ECON 493: Senior Honors Project

Professor Bird

17 May 2022

My research studies the influence of student characteristics such as gender, race, and financial status on the probability of graduating a private liberal arts college in the Midwest. This document explains how the files in the folders can be used to replicate the study and reproduce the results.

Contents and Structure

Main folder: "**Verma-493-Project**"

➔ Final Research Paper

Provides the introduction to the topic and data, and the results of the analysis with a discussion about its implications.

➔ Read Me File

Provides the information required to replicate the study.

➔ Original Data Folder

o Original and Importable Data folders

The data files obtained from the sources and the datasets used for processing, respectively.

o Metadata subfolder

▪ Metadata file

A guide on how to obtain the original datasets from its sources and their descriptions.

➔ Command Files folder

o processing.do

consists of all the steps involved in transforming importable data files into full cleaned and processed data files that will be used to generate analysis data.

o summary.do

consists of all the steps involved in generating summary statistics about the data.

o analysis.do

consists of all the steps involved in generating the regression analysis results.

➜ Analysis Data folder

consists the datasets that are used to produce the regression results from analysis.do.

➜ Documents folder

o Final Research Paper and Read Me files

the Word documents of the final research paper and read me files. PDF versions are in the main folder.

o Data Appendix file

statistical descriptions of the variables in each of the analysis datasets.

Creation of Importable Data Files

The original datasets were obtained from the respective colleges' websites. Refer to the Metadata Guide on further instructions on downloading these datasets.

Since all the original data files are in a pdf format, the following method was used to create the importable version from the original data files:

1) Open the pdf file in Acrobat DC.

2) Click on "Export PDF" in the right pane.

3) Choose "spreadsheet" as the export format and then select "Microsoft Excel Workbook."

4) Click "Export."

5) Save the converted file by clicking the "Save" button.

Through this process the following original files were converted into importable excel documents:

Original File	Importable File
4-yrGraduationBySex&Race.pdf	4-yrGraduationBySex&Race.xlsx
graduationByPell.pdf	graduationByPell.xlsx
CC_2010_20_graduationByPell.pdf	CC_2010_20_graduationByPell.xlsx
GC_graduationByPell.pdf	GC_graduationByPell.xlsx
GC_graduationRates.pdf	GC_graduationRates.xlsx
LU_2003_14_graduationBySex.pdf	LU_2003_14_graduationBySex.xlsx
LU_2005_14_graduationByPell.pdf	LU_2005_14_graduationByPell.xlsx
LU_2005_14_graduationByRace.pdf	LU_2005_14_graduationByRace.xlsx
MC_2012_21_graduation.pdf	MC_2012_21_graduation.xlsx

Replicating the Study

To replicate the research, use R Studio and the R programming language. Installation of various packages will be required; they are addressed within the command files, therefore, please read the command files to understand when and how to do the required installations.

Copy the main folder ("Verma-493-Project") onto your computer with the same layout as described above. The working directory should be the main folder- a command is run in each command file to ensure that the correct working directory is set to ensure the proper execution of the code. Please ensure that the path to the working directory in the command files corresponds with the path in your computer.

First, run the processing.do command file to ensure that the importable data files are wrangled to usable forms for the statistical and regression analysis. The outputs from this command file are three datasets (racec.xlsx, sex.xlsx, pell.xlsx) that will be saved in the 'Analysis Data' folder. Second, the summary.do command file should be run to produce frequency tables and histograms that describe each variable. There are no external outputs from this command file. Lastly, the analysis.do command file should be run to retrieve the regression results. No external output will be generated from this command file.

Please note that each of these command files can be "knitted" to produce Word documents with the comments, code, and results. To do so, click the 'Knit' button at the top center of the screen.

To convert the Rmd files into R files, run the command 'knitr:purl('Command Files', documentation = 0) in the console.