PHP 2410E - Assignment 2

Kevin W. McConeghy

Compiled: 2019-09-21

Contents

Introduction

This is the completed assignment 2 for the 'Medicare data' course at Brown University. All code is stored in a Github repository, https://github.com/kmcconeghy/PHP2410E

Statement of work

This document was created solely by the author, guidance in the homework solutions was driven by class instruction, materials or prior experience. The solutions were not shared with anyone else.

Note on R markdown

This report was generated using R markdown, LaTEX, and several non-base R packages (e.g. tidyverse).

```
## Non-base packages loaded:  Scotty tidyverse rJava kableExtra
```

Assignment as written:

- Data Assignment #2
- Working with Medicare Public Use Files
- Due October 2nd, 2019

Assignment Overview

There are two data sets each containing identifying information on Medicare beneficiaries. The information includes date of birth, gender and race. The two sources of data are:

1) the Minimum Data Set, a clinical patient record assessment that is completed each time a patient is admitted to a nursing home in the US that is certified by Medicare/Medicaid. The MDS data file (in either STATA or SAS format) is likely to include multiple records per ID (unique individual) for many individuals;

2) The Medicare Enrollment Record is the individual identifier of every Medicare beneficiary. The overall file includes detailed data from Social Security as well as whether and when the beneficiary had joined a Medicare Advantage Plan and when. The current file includes only identifying information such as date of birth, gender and race. There is only one record per person per year. The ID number on the Medicare Enrollment Record is the same as that on the MDS data file.

There are two parts to this data assignment. Each is described below. There are various ways to complete these components. It is up to the student to choose how s/he goes about completing the assignment.

- Using the MDS data determine how consistent the identifying data are across records for the same persons who have multiple records. This means, among those with more than one record, what is the rate of inconsistency for date of birth, gender and race. DoB, gender and race have multiple ways in which they could disagree, in addition to calculating the rate of inconsistency, characterize the different ways (e.g. day, month or year of birth; missing categories of information in which this inconsistency manifests itself). (Note:

The Codebook for the MDS2.0 data can be found here: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/downloads/MDS20MDSAllForms.pdf)

- Match the MDS and Medicare enrollment records based upon patient ID, using the first MDS record to match for those cases that have multiple MDS records. Next, compare the DoB, gender and Race from the two different data sources. Note that DoB is necessarily measured the same way across the two data sources as is gender (although there may be differing levels of missing data). The race variables across the two data sets are coded differently so it may be important to separately estimate the degree of agreement across the categories that are comparably labeled (e.g. white vs. white; black vs. black) since the Medicare Enrollment record is known to underestimate the number of Hispanic and Asian Americans.

Session Info

Thank you for taking the time to review my work!

```
##   setting  value
##   version  R version 3.6.1 (2019-07-05)
##   os       Windows 10 x64
##   system   x86_64, mingw32
##   ui       RTerm
##   language (EN)
##   collate  English_United States.1252
##   ctype    English_United States.1252
##   tz       America/New_York
##   date     2019-09-21
```