Exercise 1

1-10-2019

Goals:

* Combine claims data with enrollment and area-level data to create an analytic dataset.
* Examine demographics and unadjusted costs
* Run a simple linear regression.

Analysis Plan:

* Cohort Definition:
  + Members included in 2012 enrollment file
  + Had >= 1 month of enrollment in the diabetes program in 2012
* Outcome: Annual total cost of office visits with any diabetes diagnosis that occurred while enrolled in the diabetes management program
  + Claims included must meet the following criteria:
    - Office visits
    - >=1 line with an ICD9 diabetes diagnosis (250.x)
    - Claim date of service during enrollment in the diabetes program
  + Calculate the individual total annual cost of claims meeting above criteria (include all claim lines within claim, not just the line with the diabetes diagnosis)
  + Calculate the annual costs of included claims
* Predictor: Area level indicator of lower SES
  + Zip code file includes median income for each zip code
  + Create indicator of lower SES
    - Definition of Lower SES: zip code median income < median income across zip codes
    - Indicator = 0 if not lower SES, =1 if lower SES
* Covariates:
  + Age (treat as a continuous variable)
  + Indicator for gender (make the reference group = male)
  + Months of enrollment in the diabetes management program in 2012 (treat as a continuous variable)
  + Indicators for race (make the reference group = white)
* Examine summary statistics for all variables (mean, SD, median, min, max)
  + Outcome
  + Predictor
  + Covariates
* Model:
  + Linear regression