1 Cohort selection

We sampled 18.60% of the *Régie de l'assurance maladie du Québec* (RAMQ) insured population that whose home address had a postal code within the Montreal 2006 Census Metropolitan Area (CMA) over the period of 1998-2006. The vast majority of the Montreal population is insured by RAMQ. We did so by randomly sampling 18.60% of the population that was ever insured in 1998. We then sampled 18.60% of patients that became newly insured in 1999 (they were born in 1999 in Montreal, they became RAMQ insured in 1999, or they were previously RAMQ insured but recently moved to Montreal). We repeated this for every year up to 2006, retaining each sampled person in the cohort until they no longer had an address in Montreal or they died. This sampling scheme resulted in an open, dynamic cohort, that was a random sample of 18.60% of the Montreal population during any year.

2 Introduction

3 Methods

3.1 Data

3.2 Software

We implemented the random forest using the "bigr" [1] package . We implemented the GLM fitting using coordinate descent using the "glmnet" package.

4 Results

5 References

References

[1] Aloysius Lim, Leo Breiman, and Adele Cutler. *bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets*. R package version 0.1-9. 2014. URL: `https://github.com/aloysius-lim/bigrf`.