

PS7

Bradley Rann

March 2023

1 Introduction

Output: ===== Model 1 Model 2 -----
----- const nan nan (nan) (nan) hgc nan (nan) tenure nan (nan) N 2229 2229 R2
nan nan ===== Standard errors in parentheses.
* p<.1, ** p<.05, ***p<.01

This is most likely MNAR due to the fact that the the other two are extremely rare.

Among the four different models we can see the differences between the beta hats. We can see that as each model runs then it gets closer and closer to the true value as the imputation methods become more and more robust. The last two B1's are essentially just looking at better ways to impute by running regressions on the existing data we do have to fill in the missing data.

So far I have been collecting many different kinds of medical data to run my tests on, some from python packages but I'm also finding that many health organizations have very good API's that interface with Python very well. I haven't put much thought into what specific models I will use but will start thinking about it.