within income decile include vs exclude income<$10000

Chongjun Liao

11/29/2021

The second analysis is similar to the first one, instead of fitting the regression model on the whole dataset, we fit a regression model within each household income decile and see whether including applicants who reported household income below $10,000 would change the regression results. In figure 1, we can see the regression on the lowest decile explains more variation in household income by SAT scores after including applicants. While the results of regressions on other deciles have no significant change.

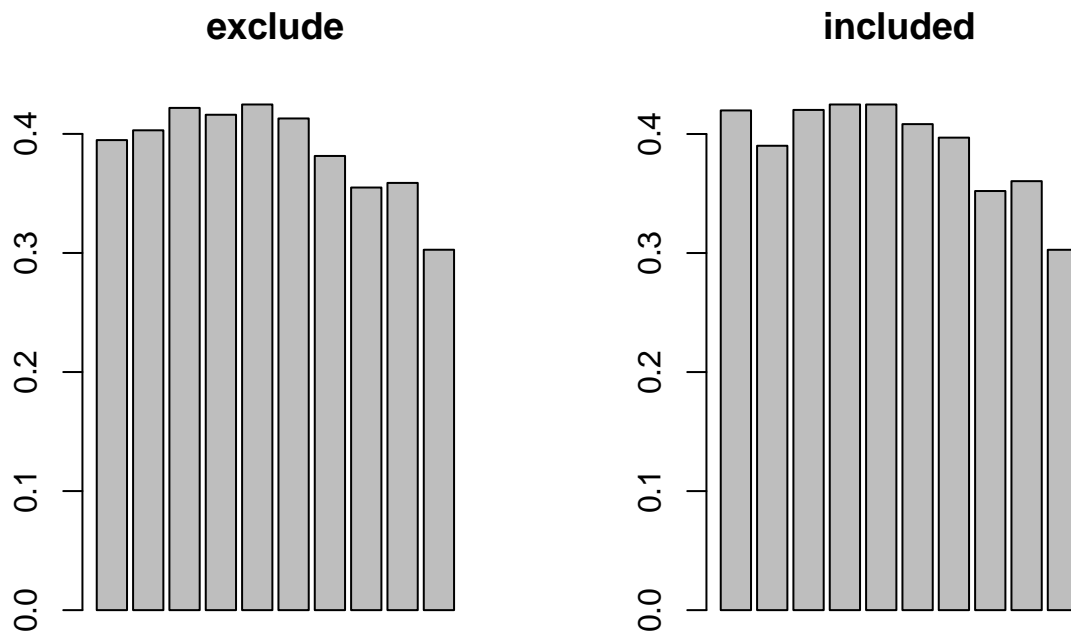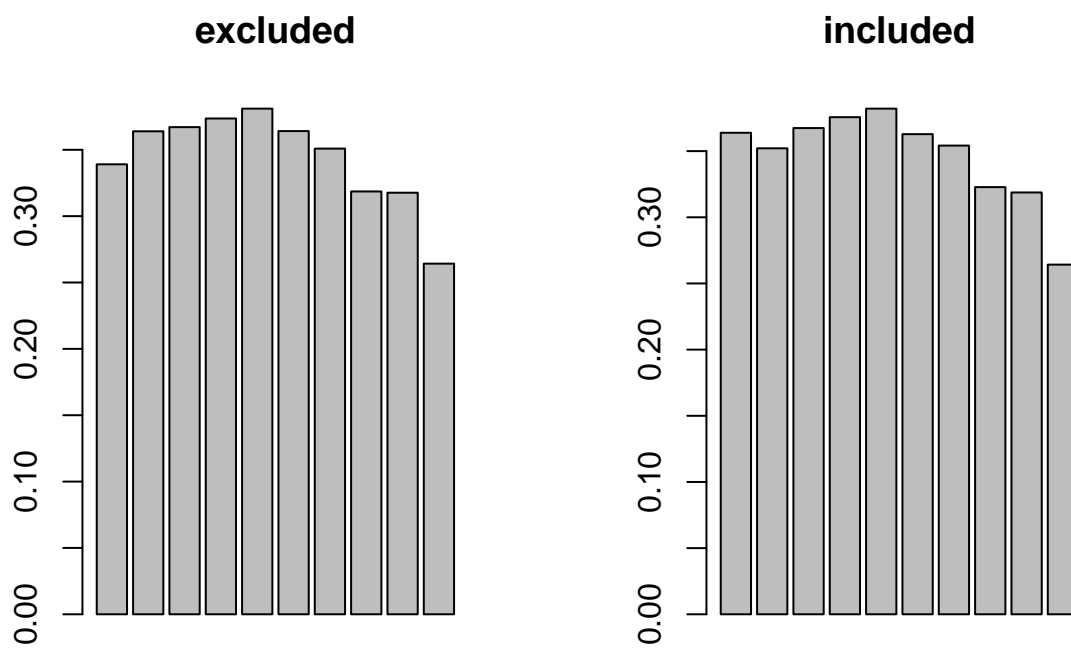In figure 2, we also find the regression on the lowest decile performs better.

Figure 1: topics on income

Figure 2: dictionary patterns on income