**PUBH 7485/8485**

**Homework 5**

The data for this assignment, as in the first three assignments, come from the obstetrics and periodontal therapy (OPT) study which tested whether or not treatment of maternal periodontal disease could reduce pre-term birth and improve other birth outcomes.

*Mediation analysis:*

1. Consider the change from baseline to visit 5 in the proportion of sites that bleed on probing (*BL..BOP - V5..BOP*) as a potential mediator of the effect of periodontal therapy on birthweight. I have updated the dataset to include *V5..BOP* in the dataset.
   1. Fit a mediation model (i.e., model for which the mediator is the outcome and treatment and other covariates are predictors). Note: as before, please use an ad hoc imputation method (e.g., median for continuous covariates or most frequent value for categorical covariates) for any regression covariates and the mediator. You can consider the mediator as a continuous outcome.
   2. Using the outcome model from previous homework assignments, add in the mediator as a covariate (note: you should consider treatment by mediator interaction).
   3. Estimate the controlled direct effect and the natural indirect effect for periodontal therapy. Please include standard errors and 95% confidence intervals.

*(Optimal) treatment rules:*

1. Fit flexible regression models for the birthweight and whether or not the pregnancy ended before 37 weeks outcomes in the OPT study, and find the estimated optimal treatment rule. A “flexible” regression model will vary depending on your level of experience with modeling approaches. Those familiar with some machine learning approaches could consider random forests, neural networks, super learner (with a several flexible methods in the library), generalized additive models, or other approach. For those without that experience, fitting a model with treatment by covariate interactions should suffice. In the latter case, the optimal treatment rule would exist in close form (although might be somewhat complex) and you can write out that estimated treatment rule. In the former, the estimated optimal rule would not exist in closed form. Please summarize the proportion who would benefit from treatment and write a simple piece of code which would return the optimal treatment decision for a given set of covariates.
2. Taking a classification-learning (C-learning) perspective, estimate an optimal “decision list” treatment rule for each outcome in the OPT study.

**PubH 8485 students**

1. Consider the change from baseline to visit 5 in the proportion of sites that bleed on probing (*BL..BOP - V5..BOP*) as a potential mediator of the effect of periodontal therapy on whether or not pregnancy ended before 37 weeks (a binary outcome). Assuming that we using a logistic model for outcome models (with and without the mediator), the controlled direct effect and natural indirect effect cannot be expressed as a straightforward expression of the regression parameters. Propose estimators for the controlled direct effect and natural indirect effect. Argue that these estimators are consistent and asymptotically normal. Estimate the controlled direct effect and natural indirect effect and obtain standard errors and 95% confidence intervals.