**PUBH 7485/8485**

**Homework 3**

The data for this assignment, as in the first two 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.

1. We will consider several different models for the propensity score and its effect on achieving covariate balance. Create a line plot of the weighted standardized mean difference of each of the covariates using the propensity score model from (a) – (d) below as well as the unweighted standardized mean difference. Comment on the balance that weighting was able to achieve.
   1. The logistic regression model considered in homework #2 with just main effects for the subset of covariates described in homework #1.
   2. The same set of covariates as in (a) but allow for nonlinear terms in the continuous covariates.
   3. The same set of covariates as in (a) but allowing for all pairwise interactions and performing some form of variable selection (e.g., forward selection using AIC).
   4. A logistic regression model considering all available baseline covariates and some form of variable selection. Note that *Birth.outcome* and *GA.at.outcome* are additional outcome variables and *PID* is the patient identifier none of which should not be included in the regression models.
   5. BONUS: A flexible regression technique considering all covariates such as random forest.
2. Using the outcome and treatment models you used in the previous assignments, find a “doubly robust” estimator of the ATE (as well as a standard error and confidence interval). How does the point estimate and standard error of this approach compare to the regression adjustment and IPW estimators? Please use both outcomes (whether or not pregnancy ended before 37 weeks and birthweight) considered in the first two assignments.