**Homework 6**

**PubH 7485/8485**

The data for this assignment concern (a) the HIV example presented in class and (b) longitudinal data from patients with primary biliary cirrhosis (PBC), a rare autoimmune liver disease, at the Mayo Clinic.

For the HIV example, I have posted this dataset to canvas. Please answer the following questions.

1. Using the gfoRmula package, implement the g-computation algorithm to estimate the anticipated response and standard error for the following treatment combinations. Note that the point estimates should be similar to what we obtained in class.
2. Using the ipw package, implement inverse probability weighted estimators for the anticipated response and standard error for the following treatment combinations. Note that the point estimates should be similar to what we obtained in class.

Concerning the PBC dataset, for patients with liver failure, liver transplantation is one possible therapy and the treatment of interest here. The goal of this assignment is to estimate the anticipated survival under different treatment regimes. The data are structured as follows. Patients return to the clinic at baseline (0 years), 1, 2, 3, 4, 5, and 6 years (variable *year\_round*) assuming that they are still alive. Covariates collected include the serum bilirubin (mg/dl, *serBilir*), *albumin* (mg/dl), *prothrombin* time (seconds), and *edema* (a factor with levels *No edema*, *edema no diuretics*, i.e., edema present without diuretics, or edema resolved by diuretics, and *edema despite diuretics*). *Age* and *sex* are measured at baseline. After data collection at a specific timepoint, the patient/provider decide whether or not to pursue transplantation. In this study, a patient can only be transplanted once and after transplantation they remain “transplanted” (i.e., in the treatment condition). The variable *dead* indicates whether or not the patient died after the current year and before the next (1 = dead, 0 = alive). Once a patient has died, there are no more rows included in the dataset. The primary outcome is whether or not the patient is alive at 7 years (i.e., if *dead* = 0 at 6 years). Obviously, if a patient died between 3 years, for example, s/he remains dead at 7 years. The temporal ordering of the covariates at any given time point are covariates, then *transplanted*, and then *dead*.

The data for this study are from the R package JM although a lot of data wrangling has been done to put the data in the structure needed for this assignment. As such, the “true” data generating model is not known.

1. We want to consider estimating the mean response of treatment regimes by the use of inverse propensity score weighted estimators. Since the treatment assignments were not part of a planned intervention study, the propensity score needs to be estimated. You should assume that the probability of treating at any point in time follows a logistic regression model.
   1. Using the data provided to you, estimate the parameters in the logistic regression model. Note that you should come up with a single set of parameter estimates. Note that once a person had died, we know that they will not receive treatment, so once a person dies, they should not contribute to the propensity score model.
   2. Using inverse propensity score weighted estimators, find the estimators for the mean response (indicator for being alive at 7 years) for the eight treatment regimes of the following form “Do not transplant for the first *x* years, if still alive then transplant” for or never transplanted. Note that once a person dies, they are always compliant with the treatment regimes thereafter.
   3. Using bootstrap methods, estimate the standard error for the estimators in 3b.