Practical 4: G-computation, IPW, and AIPW

Summer 2025
Instructors: Ting Ye & Qingyuan Zhao

We will revisit the NHANES dataset nhanesi_class_dataset.csv used for Practical 2 and use the g-computation, IPW, and AIPW methods we learned, also to answer the question "Does being physically active cause you to live longer?"

The following package is needed: CausalGAM

1. Create a dataset after applying steps 1-2 in Practical 2 (also listed below):

   (1) income.poverty.ratio and dietary.adequacy have missing values (indicated by NA). Create indicator variables for whether income.poverty.ratio and dietary.adequacy have missing values and fill in the missing values with the mean of the observed values. [Note that education has a few missing values but Missing is already coded as a category for education].

   (2) Fit a propensity score model adjusting for the confounders and the two missing indicators in Q1. To find a subset of the units with overlap, follow the procedure of Dehejia and Wahba (1999, Journal of American Statistical Association): exclude from further analysis any treated unit whose propensity score is greater than the maximum propensity score of the control units and exclude any control unit whose propensity score is less than the minimum propensity score of the treated units. How many (if any) units are excluded by this procedure?

2. Apply three alternative methods on the data without matching: g-computation, IPW, and AIPW. These three estimators and their standard errors can be directly obtained using the estimate.ATE function in the R package CausalGAM. An example of code is as follows:

```
library(CausalGAM)

ps.formula=physically.inactive~sex+smoking.status+income.poverty.ratio+
age.at.interview+race+education+working.last.three.months+married+
alcohol.consumption+dietary.adequacy+income.poverty.ratio.missingind+
dietary.adequacy.missingind

out1.formula<-years.lived.since.1971.up.to.1992~sex+smoking.status+income.poverty.ratio+
age.at.interview+race+education+working.last.three.months+married+
alcohol.consumption+dietary.adequacy+income.poverty.ratio.missingind+
dietary.adequacy.missingind

ATE.out<-estimate.ATE(pscore.formula=ps.formula,pscore.family = binomial,
outcome.formula.t=out1.formula,outcome.formula.c=out1.formula,
outcome.family=gaussian,
treatment.var="physically.inactive",data=df,nboot=0)

print(ATE.out)
```

3. Interpret your results. In addition, compare the point estimators and standard errors of these three methods to the matching method.